การเปรียบเทียบผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ระหว่างวิธี IRT-LR กับวิธี SIBTEST

พิจักษณา กาวี

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรมหาบัณฑิต
สาขาวิชาการวิจัยและสถิติทางวิทยาการปัญญา
วิทยาลัยวิทยาการวิจัยและวิทยาการปัญญา มหาวิทยาลัยบูรพา
พฤษภาคม 2561
ลิขสิทธิ์เป็นของมหาวิทยาลัยบูรพา

คณะกรรมการควบคุมวิทยานิพนธ์และคณะกรรมการสอบวิทยานิพนธ์ ได้พิจารณาวิทยานิพนธ์ของ พิจักษณา กาวี ฉบับนี้แล้ว เห็นสมควรรับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาการวิจัยและสถิติทางวิทยาการปัญญา ของมหาวิทยาลัยบูรพาได้

คณะกรรมการควบคุมวิทยานิพนธ์

.....อาจารย์ที่ปรึกษาหลัก
(ดร.ปิยะทิพย์ ประดุจพรม)

คณะกรรมการสอบวิทยานิพนธ์

.....ประธาน
(รองศาสตราจารย์ ดร.เสรี ชัดแช้ม)

.....กรรมการ
(ดร.ปิยะทิพย์ ประดุจพรม)

.....กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.ภัทราวดี มากมี)

.....กรรมการ
(ดร.กนก พานทอง)

วิทยาลัยวิทยาการวิจัยและวิทยาการปัญญาอนุมัติให้รับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาการวิจัยและสถิติทางวิทยาการปัญญา ของมหาวิทยาลัยบูรพา

.....คณบดีวิทยาลัยวิทยาการวิจัย
(ผู้ช่วยศาสตราจารย์ ดร.สุชาดา กรเพชรปาณี) และวิทยาการปัญญา
วันที่ 11 เดือน พฤษภาคม พ.ศ. 2561

การวิจัยนี้ได้รับทุนอุดหนุนการวิจัย ประเภทบัณฑิตศึกษา
จากสำนักงานคณะกรรมการวิจัยแห่งชาติ ประจำปี 2559

ประกาศคุณูปการ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงได้ด้วยความกรุณาจาก ดร.ปิยะทิพย์ ประดุจพรม อาจารย์ที่ปรึกษาหลัก ได้ถ่ายทอดความรู้ ประสิทธิ์ประสาทวิชา รวมทั้งคำแนะนำที่หลากหลายเป็นอย่างดี ทั้งในส่วนของเนื้อหาวิทยานิพนธ์ การวางแผนการทำวิทยานิพนธ์ การใช้โปรแกรมต่าง ๆ ในการวิเคราะห์ข้อมูล พร้อมทั้งให้กำลังใจ ตลอดเวลาที่ดำเนินการทำวิจัย ทำให้วิจัยฝ่าฟันอุปสรรคต่าง ๆ มาได้ด้วยดี ผู้วิจัยรู้สึกซาบซึ้งในพระคุณเป็นอย่างยิ่ง จึงขอกราบขอบพระคุณเป็นอย่างสูงไว้ ณ ที่นี้

ขอขอบพระคุณ Prof. Dr. Larry R. Nelson สำหรับการให้ความช่วยเหลือและให้ความสนับสนุน อีกทั้งยังอนุเคราะห์โปรแกรม LERTAP ให้ใช้สำหรับการวิเคราะห์ข้อมูล ทั้งนี้ ขอขอบพระคุณ Prof. Dr. Louis Roussos สำหรับการให้ความช่วยเหลือและให้ความสนับสนุน อีกทั้งยังอนุเคราะห์โปรแกรม SIBTEST ให้ใช้สำหรับการวิเคราะห์ข้อมูล ขอขอบคุณสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) หรือ สทศ. ที่ให้ความอนุเคราะห์ข้อมูลผลการตอบแบบทดสอบ O-NET ทั้ง 8 กลุ่มสาระการเรียนรู้ การทำงานวิจัยครั้งนี้ด้วย นอกจากนี้ ขอขอบพระคุณ คณาจารย์ทุกท่าน บุคลากรทุกคนและหน่วยงานที่เกี่ยวข้องทุกแห่งสำหรับการอำนวยความสะดวกตลอดการดำเนินการวิจัยในครั้งนี้

ขอกราบขอบพระคุณครอบครัวหันทะยุง และครอบครัวกาวีทุกคน ที่เป็นทั้งต้นแบบของความเพียร ความพยายาม ความอดทน ความมุ่งมั่น ตั้งใจ ทั้งยังเป็นกำลังใจที่ดี ทำให้ผู้วิจัยไม่ย่อท้อต่ออุปสรรคต่าง ๆ ที่เข้ามามากมาย นอกจากนี้ ขอขอบคุณมิตรภาพของเพื่อนปริญญาโทวิทยาเขตสระแก้วทุกคน สำหรับการช่วยเหลือ คำแนะนำ การสนับสนุนและกระตุ้นให้ผู้วิจัยมีแรงกายแรงใจมากเพียงพอที่จะดำเนินการวิจัยจนผ่านพ้นไปได้ด้วยดี ผู้วิจัยขอขอบคุณในมิตรภาพที่มอบให้เสมอมา

ขอขอบคุณสำนักงานคณะกรรมการวิจัยแห่งชาติ ที่ให้ทุนอุดหนุนงานวิจัยส่วนหนึ่งในการทำวิจัยครั้งนี้ คุณค่าและประโยชน์ของวิทยานิพนธ์ฉบับนี้ ผู้วิจัยขอมอบแด่บุพการี บูรพาจารย์ และผู้มีพระคุณทุกท่านทั้งในอดีตและปัจจุบัน ที่ทำให้ข้าพเจ้าเป็นผู้มีการศึกษา และประสบความสำเร็จมาจนตราบเท่าทุกวันนี้

พิจักษณา กาวี

56910392: สาขาวิชา: การวิจัยและสถิติทางวิทยาการปัญญา;
วท.ม. (การวิจัยและสถิติทางวิทยาการปัญญา)
คำสำคัญ: การทำหน้าที่ต่างกันของข้อสอบ/ ทฤษฎีการตอบสนองข้อสอบ/ แบบทดสอบ O-NET/ วิธี IRT-LR/ วิธี SIBTEST

พิจักษณา กาวี: การเปรียบเทียบผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ระหว่างวิธี IRT-LR และวิธี SIBTEST (A COMPARISON OF DETECTING DIFFERENTIAL ITEM FUNCTIONING IN GRADE 12 O-NET RESULTS BETWEEN THE IRT-LR AND SIBTEST METHODS) คณะกรรมการควบคุมวิทยานิพนธ์: ปิยะทิพย์ ประดุจพรม, Ph.D., 263 หน้า. ปี พ.ศ. 2561.

การทำหน้าที่ต่างกันของข้อสอบ (DIF) กับความลำเอียงของข้อสอบ (Item Bias) มีความแตกต่างกัน การทำหน้าที่ต่างกันของข้อสอบเป็นกระบวนการที่ใช้วิธีการทางสถิติ ในการตรวจสอบส่วนความลำเอียง ของข้อสอบเป็นกระบวนการตัดสินความยุติธรรมของข้อสอบ โดยนำผลการทำหน้าที่ต่างกันของข้อสอบนำมาวิเคราะห์เชิงเนื้อหาโดยผู้เชี่ยวชาญ การวิจัยนี้มีวัตถุประสงค์เพื่อวิเคราะห์คุณภาพของข้อสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2556 จำนวน 8 กลุ่มสาระการเรียนรู้ ทั้งฉบับก่อนและฉบับหลังตัด DIF ตรวจสอบความเที่ยง ความตรงเชิงโครงสร้างของแบบทดสอบ O-NET ทั้งฉบับก่อนและฉบับหลังตัด DIF ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธี IRT-LR กับวิธี SIBTEST และเปรียบเทียบอัตรา ความคลาดเคลื่อนประเภทที่ 1 และ 2 ของผลการตรวจ DIF ใช้ข้อมูลทุติยภูมิ จากสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) วิเคราะห์คุณภาพของข้อสอบตามหลักการของทฤษฎีการตอบสนองข้อสอบ แบบ 3 พารามิเตอร์ ด้วยโปรแกรม Xcalibre Version 4.2.2

ผลการวิจัยปรากฏว่า

1) แบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 จำนวน 8 กลุ่มสาระการเรียนรู้ทั้งฉบับก่อนและฉบับหลังตัด DIF มีค่าอำนาจจำแนกของข้อสอบ (a) และค่าความยากของข้อสอบ (b) แตกต่างกัน ส่วนค่าโอกาสการเดาของข้อสอบ (c) ไม่เกิน 0.3 โดยมีค่าความเที่ยงของแบบทดสอบฉบับก่อนและหลังตัด DIF แตกต่างกัน

2) การตรวจสอบความตรงเชิงโครงสร้างของแบบทดสอบ O-NET ทั้ง 8 กลุ่มสาระการเรียนรู้ ฉบับก่อนและฉบับหลังตัด DIF มี 2 กลุ่มสาระการเรียนรู้ ไม่แตกต่างกัน ซึ่งความสอดคล้องกับข้อมูลเชิงประจักษ์

3) การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ O-NET ทั้ง 8 กลุ่มสาระการเรียนรู้ พบว่า วิธี IRT-LR ตรวจพบ DIF มากกว่าวิธี SIBTEST คิดเป็นร้อยละ 41.86 และทั้ง 2 วิธี พบ DIF ตรงกัน จำนวน 65 ข้อ คิดเป็นร้อยละ 15.12 ($p < 0.05$)

4) การเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 และ 2 พบว่า มี 2 กลุ่มสาระการเรียนรู้ที่มีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำ คือ สาระการเรียนรู้ภาษาอังกฤษ และคณิตศาสตร์ โดย มี 2 กลุ่มสาระการเรียนรู้ ที่มีอัตราความคลาดเคลื่อนประเภทที่ 2 สูง คือ สาระการเรียนรู้ภาษาต่างประเทศ (ภาษาอังกฤษ) และการงานอาชีพและเทคโนโลยี

56910392: MAJOR: RESEARCH AND STATISTICS IN COGNITIVE SCIENCE;
M.Sc. (RESEARCH AND STATISTICS IN COGNITIVE SCIENCE)
KEYWORDS: DIFFERENTIAL ITEM FUNCTIONING/ ITEM RESPONSE THEORY/ ORDINARY NATIONAL EDUCATION TEST/ IRT METHOD/ SIBTEST METHOD

PIJAGSANA GAWEE: A COMPARISON OF DETECTING DIFFERENTIAL ITEM FUNCTIONING IN GRADE 12 O-NET RESULTS BETWEEN THE IRT-LR AND SIBTEST METHODS. ADVISORY COMMITTEE: PIYATHIP PRADUJPROM, Ph.D., 263 P. 2018.

Differential Item Functioning (DIF) and Item Bias possess a different idea. Differential Item Functioning is a process that uses statistical methods to validate. Whilst, Item Bias is a process that justifies the item fairness using the results of the Differential Item Functioning for content analysis and the experts are assigned to validate. The objectives of this research were (1) to analyze the quality of O-NET test items for grade twelve in the academic year 2556 across eight groups of curriculum implementation for both before and after DIF item elimination in terms of reliability index and construct validity and (2), to compare IRT-LR and SIBTEST type I and II error rates. Secondary data were obtained from the National Institute of Educational Testing Service. The three parameter logistic model of item response theory was used to assess item quality. Statistical analysis was performed using Xcalibre Version 4.2.2.

Results were shown as follows:

1) The O-NET results for grade twelve on eight groups of curriculum implementations for both before and after DIF item elimination indicated differences among item-discrimination values and item-difficulty values. The item guessing parameter did not exceed 0.30. The reliability index was significantly different after DIF-item deletion.

2) The construct validity of the O-NET test before and after DIF-item deletion across eight groups of curriculum implementations indicated that two groups of curriculum implementations did not differ, and were consistent with empirical data.

3) Comparing the results revealed that IRT-LR method was better at detecting DIF than the SIBTEST method.

4) Comparing the result of the type I and II error rates revealed that there was a low level of Type I error on two groups of curriculum implementation (English and Mathematics), but there was a high level of Type II error on two groups of curriculum implementation (English, and Careers and Technology).

สารบัญ

หน้า
บทคัดย่อภาษาไทย..... จ
บทคัดย่อภาษาอังกฤษ..... ฉ
สารบัญ..... ช
สารบัญตาราง..... ฎ
สารบัญภาพ..... ฒ
บทที่
1 บทนำ..... 1
ความเป็นมาและความสำคัญของปัญหา..... 1
วัตถุประสงค์ของการวิจัย..... 4
กรอบแนวคิดการวิจัย..... 5
สมมติฐานของการวิจัย..... 7
ประโยชน์ที่คาดว่าจะได้รับจากการวิจัย..... 8
ขอบเขตของการวิจัย..... 8
นิยามศัพท์เฉพาะ..... 9
2 เอกสารและงานวิจัยที่เกี่ยวข้อง..... 12
ตอนที่ 1 การทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน (O-NET) และงานวิจัย ที่เกี่ยวข้อง..... 12
ตอนที่ 2 ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT) และงานวิจัยที่เกี่ยวข้อง..... 21
ตอนที่ 3 การวิเคราะห์ความตรงเชิงโครงสร้าง (Construction Validity) และงานวิจัยที่เกี่ยวข้อง..... 34
ตอนที่ 4 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) และงานวิจัย ที่เกี่ยวข้อง..... 40
ตอนที่ 5 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ด้วยวิธี IRT-LR และงานวิจัยที่เกี่ยวข้อง..... 52
ตอนที่ 6 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ด้วยวิธี SIBTEST และวิจัยที่เกี่ยวข้อง..... 61
3 วิธีดำเนินการวิจัย..... 74
ระยะที่ 1 การวิเคราะห์คุณภาพของข้อสอบ (a, b และ c) และค่าความเที่ยงของ แบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 จำนวน 8 กลุ่มสาระการเรียนรู้ ทั้งฉบับก่อนและฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกัน ตามหลักทฤษฎี การตอบสนองข้อสอบ แบบ 3 พารามิเตอร์..... 76

สารบัญ (ต่อ)

บทที่ หน้า

ระยะที่ 2 การตรวจสอบความตรงเชิงโครงสร้างของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 จำนวน 8 กลุ่มสาระการเรียนรู้ ทั้งฉบับก่อนและฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกัน..... 78

ระยะที่ 3 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ทั้ง 8 กลุ่มสาระการเรียนรู้ ด้วยวิธี IRT-LR และวิธี SIBTEST ของแบบทดสอบฉบับก่อนตัดข้อสอบที่ทำหน้าที่ต่างกัน..... 86

ระยะที่ 4 การเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 และ 2 ของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 จำนวน 8 กลุ่มสาระการเรียนรู้ ระหว่างวิธี IRT-LR กับวิธี SIBTEST..... 90

4 ผลการวิจัย..... 92

ตอนที่ 1 การวิเคราะห์คุณภาพของข้อสอบ (a, b และ c) และค่าความเที่ยงของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 จำนวน 8 กลุ่มสาระการเรียนรู้ ทั้งฉบับก่อนและฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกัน ตามหลักทฤษฎีการตอบสนองข้อสอบ แบบ 3 พารามิเตอร์..... 95

ตอนที่ 2 การตรวจสอบความตรงเชิงโครงสร้างของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 จำนวน 8 กลุ่มสาระการเรียนรู้ ทั้งฉบับก่อนและฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกัน..... 117

ตอนที่ 3 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ทั้ง 8 กลุ่มสาระการเรียนรู้ ด้วยวิธี IRT-LR และวิธี SIBTEST ของแบบทดสอบฉบับก่อนตัดข้อสอบที่ทำหน้าที่ต่างกัน 133

ตอนที่ 4 การเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 และ 2 ของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 จำนวน 8 กลุ่มสาระการเรียนรู้ ระหว่างวิธี IRT-LR กับวิธี SIBTEST 145

5 สรุปและอภิปรายผล..... 147

สรุปผลการวิจัย..... 147

อภิปรายผล..... 152

สารบัญ (ต่อ)

หน้า

ข้อเสนอแนะในการนำผลการวิจัยไปใช้..... 154
ข้อเสนอแนะสำหรับการวิจัยต่อไป..... 154
บรรณานุกรม..... 155
ภาคผนวก..... 165
ภาคผนวก ก หนังสือขอความอนุเคราะห์ขอข้อมูลเพื่อการวิจัย..... 166
ภาคผนวก ข บันทึกข้อตกลงว่าด้วยการขอใช้ประโยชน์จากข้อสอบในการทดสอบ
ทางการศึกษาระดับชาติขั้นพื้นฐาน (O-NET)..... 168
ภาคผนวก ค แบบรายงานผลการพิจารณาจริยธรรมการวิจัยในมนุษย์
วิทยาลัยวิทยาการวิจัยและวิทยาการปัญญา มหาวิทยาลัยบูรพา..... 171
ภาคผนวก ง ตัวอย่าง Print Out ผลการวิเคราะห์คุณภาพของข้อสอบ
และค่าความเที่ยงของแบบทดสอบ O-NET ฉบับก่อนและหลังตัดข้อสอบ
ที่ทำหน้าที่ต่างกันออกจากแบบทดสอบ..... 173
ภาคผนวก จ ตัวอย่าง Print Out ผลการตรวจสอบความตรงเชิงโครงสร้าง (CFA)
ในแบบทดสอบ O-NET ฉบับก่อนและหลังตัดข้อสอบที่ทำหน้าที่
ต่างกัน..... 181
ภาคผนวก ฉ ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบ O-NET
ด้วยวิธี IRT-LR และวิธี SIBTEST..... 202
ภาคผนวก ช ตัวอย่าง Print Out ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ
ในแบบทดสอบ O-NET ด้วยวิธี IRT-LR โดยใช้โปรแกรม IRTPRO..... 215
ภาคผนวก ซ ตัวอย่าง Print Out ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ
ในแบบทดสอบ O-NET ด้วยวิธี SIBTEST โดยใช้โปรแกรม SIBTEST..... 231
ภาคผนวก ฌ ตัวอย่าง Print Out ผลการคำนวณอัตราความคลาดเคลื่อน
ของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ
ในแบบทดสอบ O-NET ระหว่างวิธี IRT-LR กับวิธี SIBTEST..... 249
ภาคผนวก ญ ผลการคำนวณอัตราความคลาดเคลื่อนประเภทที่ 1 และ 2
ของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ..... 258
ประวัติย่อของผู้วิจัย..... 263

สารบัญตาราง

ตารางที่ หน้า

2-1 การปรับสัดส่วนการใช้ผลการทดสอบระดับชาติขั้นพื้นฐาน เป็นองค์ประกอบหนึ่งของการตัดสินผลการเรียนของผู้ที่จบการศึกษาตามหลักสูตรการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 17

2-2 ดัชนีที่ใช้ในการพิจารณาความสอดคล้องของโมเดลตามสมมติฐานกับข้อมูลเชิงประจักษ์ 37

2-3 วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการตรวจให้คะแนนแบบทวิภาค (Dichotomous DIF) และพหุวิภาค (Polytomous DIF) 48

3-1 โครงสร้างองค์ประกอบของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ฉบับก่อนและฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกัน 79

4-1 ผลการวิเคราะห์คุณภาพของข้อสอบ (a, b และ c) ฉบับก่อนและฉบับหลังตัดการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบภาษาไทย ชั้นมัธยมศึกษาปีที่ 6 จำนวน 70 ข้อ 96

4-2 ผลการวิเคราะห์คุณภาพของข้อสอบ (a, b และ c) ฉบับก่อนและฉบับหลังตัดการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบสังคมศึกษา ศาสนาและวัฒนธรรม ชั้นมัธยมศึกษาปีที่ 6 จำนวน 65 ข้อ 99

4-3 ผลการวิเคราะห์คุณภาพของข้อสอบ (a, b และ c) ฉบับก่อนและฉบับหลังตัดการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบภาษาอังกฤษ ชั้นมัธยมศึกษาปีที่ 6 จำนวน 80 ข้อ 102

4-4 ผลการวิเคราะห์คุณภาพของข้อสอบ (a, b และ c) ฉบับก่อนและฉบับหลังตัดการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 6 จำนวน 32 ข้อ 105

4-5 ผลการวิเคราะห์คุณภาพของข้อสอบ (a, b และ c) ฉบับก่อนและฉบับหลังตัดการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบวิทยาศาสตร์ ชั้นมัธยมศึกษาปีที่ 6 จำนวน 80 ข้อ 107

4-6 ผลการวิเคราะห์คุณภาพของข้อสอบ (a, b และ c) ฉบับก่อนและฉบับหลังตัดการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบสุขศึกษา และพลศึกษา ชั้นมัธยมศึกษาปีที่ 6 จำนวน 32 ข้อ 110

4-7 ผลการวิเคราะห์คุณภาพของข้อสอบ (a, b และ c) ฉบับก่อนและฉบับหลังตัดการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบศิลปะ ชั้นมัธยมศึกษาปีที่ 6 จำนวน 30 ข้อ 112

4-8 ผลการวิเคราะห์คุณภาพของข้อสอบ (a, b และ c) ฉบับก่อนและฉบับหลังตัดการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบการงานอาชีพ และเทคโนโลยี ชั้นมัธยมศึกษาปีที่ 6 จำนวน 41 ข้อ 114

สารบัญตาราง (ต่อ)

ตารางที่ หน้า

4-9 ผลการตรวจสอบค่าความเที่ยงของข้อสอบฉบับก่อนและฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกัน ตามหลักทฤษฎีการตอบสนองข้อสอบ (IRT) แบบ 3 พารามิเตอร์..... 116

4-10 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสองของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ฉบับก่อนตัดตัด DIF ของแบบทดสอบคณิตศาสตร์..... 119

4-11 ค่าสถิติความสอดคล้องของโมเดลความสัมพันธ์ตามสมมติฐานกับข้อมูลเชิงประจักษ์องค์ประกอบเชิงยืนยันอันดับสองของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ฉบับก่อนตัด DIF ของแบบทดสอบคณิตศาสตร์..... 120

4-12 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสองของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ฉบับก่อนตัดตัด DIF ของแบบทดสอบศิลปะ..... 123

4-13 ค่าสถิติความสอดคล้องของโมเดลความสัมพันธ์ตามสมมติฐานกับข้อมูลเชิงประจักษ์องค์ประกอบเชิงยืนยันอันดับสองของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ฉบับก่อนตัดตัด DIF ของแบบทดสอบศิลปะ..... 124

4-14 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสองของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ฉบับหลังตัด DIF ของแบบทดสอบคณิตศาสตร์..... 127

4-15 ค่าสถิติความสอดคล้องของโมเดลความสัมพันธ์ตามสมมติฐานกับข้อมูลเชิงประจักษ์องค์ประกอบเชิงยืนยันอันดับสองของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ฉบับหลังตัดตัด DIF ของแบบทดสอบคณิตศาสตร์..... 128

4-16 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสองของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ฉบับก่อนตัดตัด DIF ของแบบทดสอบศิลปะ..... 131

4-17 ค่าสถิติความสอดคล้องของโมเดลความสัมพันธ์ตามสมมติฐานกับข้อมูลเชิงประจักษ์องค์ประกอบเชิงยืนยันอันดับสองของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ฉบับก่อนตัดตัด DIF ของแบบทดสอบศิลปะ..... 132

4-18 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบภาษาไทย ด้วยวิธี IRT-LR และวิธี SIBTEST จำนวน 70 ข้อ..... 133

4-19 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบสังคมศึกษา ศาสนาและวัฒนธรรม ด้วยวิธี IRT-LR และวิธี SIBTEST จำนวน 65 ข้อ..... 135

4-20 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบภาษาอังกฤษ ด้วยวิธี IRT-LR และวิธี SIBTEST จำนวน 80 ข้อ..... 136

4-21 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบคณิตศาสตร์ ด้วยวิธี IRT-LR และวิธี SIBTEST จำนวน 32 ข้อ..... 138

สารบัญตาราง (ต่อ)

ตารางที่ หน้า
4-22 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบวิทยาศาสตร์ ด้วยวิธี IRT-LR และวิธี SIBTEST จำนวน 80 ข้อ..... 139
4-23 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบสุขศึกษา และพลศึกษา ด้วยวิธี IRT-LR และวิธี SIBTEST จำนวน 32 ข้อ..... 140
4-24 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบศิลปะ ด้วยวิธี IRT-LR และวิธี SIBTEST จำนวน 30 ข้อ..... 142
4-25 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบการงานอาชีพ และเทคโนโลยี ด้วยวิธี IRT-LR และวิธี SIBTEST จำนวน 41 ข้อ..... 143
4-26 ผลการคำนวณอัตราความคลาดเคลื่อนประเภทที่ 1 และ 2 ของผลการตรวจสอบ การทำหน้าที่ต่างกันของข้อสอบ ทั้ง 8 กลุ่มสาระการเรียนรู้..... 145
4-27 ผลการเปรียบเทียบการทดสอบความแตกต่างระหว่างวิธี IRT-LR กับวิธี SIBTEST..... 146
ฉ-1 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ด้วยวิธี IRT-LR และวิธี SIBTEST ของแบบทดสอบภาษาไทย..... 203
ฉ-2 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ด้วยวิธี IRT-LR และวิธี SIBTEST ของแบบทดสอบสังคมศึกษา ศาสนาและวัฒนธรรม..... 205
ฉ-3 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ด้วยวิธี IRT-LR และวิธี SIBTEST ของแบบทดสอบภาษาอังกฤษ..... 207
ฉ-4 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ด้วยวิธี IRT-LR และวิธี SIBTEST ของแบบทดสอบคณิตศาสตร์..... 209
ฉ-5 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ด้วยวิธี IRT-LR และวิธี SIBTEST ของแบบทดสอบวิทยาศาสตร์..... 210
ฉ-6 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ด้วยวิธี IRT-LR และวิธี SIBTEST ของแบบทดสอบสุขศึกษาและพลศึกษา..... 212
ฉ-7 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ด้วยวิธี IRT-LR และวิธี SIBTEST ของแบบทดสอบศิลปะ..... 213
ฉ-8 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ด้วยวิธี IRT-LR และวิธี SIBTEST ของแบบทดสอบการงานอาชีพ และเทคโนโลยี..... 214

สารบัญตาราง (ต่อ)

ตารางที่ หน้า

ฌ-1 ผลการคำนวณอัตราความคลาดเคลื่อนประเภทที่ 1 และ 2 ของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ในแบบทดสอบ O-NET ระหว่างวิธี IRT-LR กับวิธี SIBTEST แบบทดสอบภาษาไทย..... 250

ฌ-2 ผลการคำนวณอัตราความคลาดเคลื่อนประเภทที่ 1 และ 2 ของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ในแบบทดสอบ O-NET ระหว่างวิธี IRT-LR กับวิธี SIBTEST แบบทดสอบสังคมศึกษา ศาสนา และวัฒนธรรม..... 251

ฌ-3 ผลการคำนวณอัตราความคลาดเคลื่อนประเภทที่ 1 และ 2 ของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ในแบบทดสอบ O-NET ระหว่างวิธี IRT-LR กับวิธี SIBTEST แบบทดสอบภาษาอังกฤษ..... 252

ฌ-4 ผลการคำนวณอัตราความคลาดเคลื่อนประเภทที่ 1 และ 2 ของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบ O-NET ระหว่างวิธี IRT-LR กับวิธี SIBTEST แบบทดสอบคณิตศาสตร์..... 253

ฌ-5 ผลการคำนวณอัตราความคลาดเคลื่อนประเภทที่ 1 และ 2 ของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ในแบบทดสอบ O-NET ระหว่างวิธี IRT-LR กับวิธี SIBTEST แบบทดสอบวิทยาศาสตร์..... 254

ฌ-6 ผลการคำนวณอัตราความคลาดเคลื่อนประเภทที่ 1 และ 2 ของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ในแบบทดสอบ O-NET ระหว่างวิธี IRT-LR กับวิธี SIBTEST แบบทดสอบสุขศึกษา และพลศึกษา..... 255

ฌ-7 ผลการคำนวณอัตราความคลาดเคลื่อนประเภทที่ 1 และ 2 ของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ในแบบทดสอบ O-NET ระหว่างวิธี IRT-LR กับวิธี SIBTEST แบบทดสอบศิลปะ..... 256

ฌ-8 ผลการคำนวณอัตราความคลาดเคลื่อนประเภทที่ 1 และ 2 ของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ในแบบทดสอบ O-NET ระหว่างวิธี IRT-LR กับวิธี SIBTEST แบบทดสอบการงานอาชีพ และเทคโนโลยี..... 257

สารบัญภาพ

ภาพที่ หน้า
1-1 กรอบแนวคิดการเปรียบเทียบผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ ในแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ระหว่างวิธี IRT-LR กับวิธี SIBTEST..... 6
2-1 ระบบการทดสอบทางการศึกษาของประเทศไทย..... 13
2-2 สาระการเรียนรู้ในหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พ.ศ. 2551..... 15
2-3 โค้งคุณลักษณะของข้อสอบ แบบ 1 พารามิเตอร์..... 24
2-4 โค้งคุณลักษณะของข้อสอบ แบบ 2 พารามิเตอร์..... 25
2-5 โค้งคุณลักษณะของข้อสอบ แบบ 3 พารามิเตอร์..... 26
2-6 แท็บระบุไฟล์ข้อมูลเมทริกซ์..... 27
2-7 แท็บสำหรับป้อนข้อมูลเพื่อระบุคอลัมน์ของเมทริกซ์..... 28
2-8 แท็บ IRT โมเดล..... 28
2-9 แท็บการเปรียบเทียบ..... 29
2-10 แท็บการประเมินความยากของข้อสอบ..... 29
2-11 แท็บตัวเลือกระบุผลลัพธ์..... 30
2-12 ขั้นตอนการวิเคราะห์องค์ประกอบเชิงยืนยันลำดับที่สอง (2^{nd} order CFA)..... 38
2-13 ข้อสอบทำหน้าที่ต่างกันแบบเอกรูป (Uniform DIF)..... 42
2-14 ข้อสอบทำหน้าที่ต่างกันแบบไม่มีทิศทาง (Non-unidirectional DIF)..... 43
2-15 ข้อสอบทำหน้าที่ต่างกันแบบมีทิศทางเดียวกัน (Unidirectional DIF)..... 44
2-16 การเพิ่มข้อมูล..... 55
2-17 การเพิ่มเติมชื่อและแสดงความคิดเห็น..... 56
2-18 การจัดกลุ่มตัวแปร..... 56
2-19 การเลือกรายการที่จะวิเคราะห์..... 57
2-20 การแสดงค่าเริ่มต้นใน 2PL..... 57
2-21 การนำเข้าสู่การวิเคราะห์ DIF โดยการกำหนดเงื่อนไขเพิ่มเติม..... 58
2-22 ตัวอย่างผลลัพธ์จากการวิเคราะห์ DIF..... 58
2-23 กราฟการวิเคราะห์ DIF..... 59
2-24 หน้าจอเริ่มต้นของโปรแกรม SIBTEST..... 66
2-25 ขั้นตอนการนำเข้าไฟล์ข้อมูล..... 67
2-26 การใช้ข้อมูลของแบบทดสอบพหุมิติ..... 67
2-27 การเลือกไฟล์ข้อมูลกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ..... 68
2-28 การเลือกคำนวณทางสถิติ..... 69
2-29 การวิเคราะห์ผลในโปรแกรม SIBTEST..... 70
3-1 ขั้นตอนการดำเนินการวิจัย..... 75

สารบัญภาพ (ต่อ)

ภาพที่ หน้า

3-2 ขั้นตอนวิเคราะห์คุณภาพและตรวจสอบค่าความเที่ยงของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ทั้งฉบับก่อนและฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกันออกจากแบบทดสอบ ตามหลักทฤษฎีการตอบสนองข้อสอบ แบบ 3 พารามิเตอร์..... 76

3-3 ขั้นตอนการการตรวจสอบความตรงเชิงโครงสร้างของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 จำนวน 8 กลุ่มสาระการเรียนรู้ ทั้งฉบับก่อนและฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกัน..... 78

3-4 ตัวอย่างการเตรียมไฟล์ข้อมูล วิเคราะห์ความองค์ประกอบเชิงยืนยันลำดับที่สอง (CFA) ทั้งฉบับก่อนและฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกันออกจากแบบทดสอบ..... 84

3-5 แสดงขั้นตอนการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ทั้ง 8 กลุ่มสาระการเรียนรู้ ด้วยวิธี IRT-LR และวิธี SIBTEST ฉบับก่อนตัดข้อสอบที่ทำหน้าที่ต่างกัน 86

3-6 ตัวอย่างการจัดไฟล์ข้อมูลสำหรับวิเคราะห์ด้วยวิธี IRT-LR ในรูปแบบไฟล์ .sav..... 87

3-7 ตัวอย่างการจัดไฟล์ข้อมูลสำหรับการวิเคราะห์ด้วยวิธี SIBTEST ในรูปไฟล์ .dat..... 88

3-8 ขั้นตอนการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 และ 2 ของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ระหว่างวิธี IRT-LR กับวิธีSIBTEST..... 90

4-1 โมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสองของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ฉบับก่อนตัด DIF ของแบบทดสอบคณิตศาสตร์..... 118

4-2 โมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสองของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ฉบับก่อนตัด DIF ของแบบทดสอบศิลปะ..... 122

4-3 โมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสองของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ฉบับหลังตัด DIF ของแบบทดสอบคณิตศาสตร์..... 126

4-4 โมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสองของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ฉบับหลังตัดตัด DIF ของแบบทดสอบศิลปะ..... 130

4-5 จำนวนข้อสอบที่พบว่าทำหน้าที่ต่างกันในแบบทดสอบภาษาไทย ด้วยวิธี IRT-LR และวิธี SIBTEST..... 134

4-6 จำนวนข้อสอบที่พบว่าทำหน้าที่ต่างกันในแบบทดสอบสังคมศึกษา ศาสนา และวัฒนธรรม ด้วยวิธี IRT-LR และวิธี SIBTEST 141

4-7 จำนวนข้อสอบที่พบว่าทำหน้าที่ต่างกันในแบบทดสอบภาษาอังกฤษ ด้วยวิธี IRT-LR และวิธี SIBTEST..... 137

4-8 จำนวนข้อสอบที่พบว่าทำหน้าที่ต่างกันในแบบทดสอบคณิตศาสตร์ ด้วยวิธี IRT-LR และวิธี SIBTEST..... 138

สารบัญภาพ (ต่อ)

ภาพที่ หน้า

4-9 จำนวนข้อสอบที่พบว่าทำหน้าที่ต่างกันในแบบทดสอบวิทยาศาสตร์ ด้วยวิธี IRT-LR และวิธี SIBTEST 140

4-10 จำนวนข้อสอบที่พบว่าทำหน้าที่ต่างกันในแบบทดสอบสุขศึกษา และพลศึกษา ด้วยวิธี IRT-LR และวิธี SIBTEST..... 141

4-11 จำนวนข้อสอบที่พบว่าทำหน้าที่ต่างกัน ในแบบทดสอบศิลปะ ด้วยวิธี IRT-LR และวิธี SIBTEST..... 143

4-12 จำนวนข้อสอบที่พบว่าทำหน้าที่ต่างกันในแบบทดสอบการงานอาชีพและเทคโนโลยี ด้วยวิธี IRT-LR และวิธี SIBTEST..... 144

บทที่ 1
บทนำ

ความเป็นมาและความสำคัญของปัญหา

ปัจจุบันผู้สอนต้องให้ความสำคัญกับการจัดการศึกษาในด้านต่าง ๆ เช่น การวัดผล และการทดสอบผลสัมฤทธิ์ทางการเรียนของผู้เรียนให้มีความน่าเชื่อถือ และเป็นไปตามมาตรฐานสากล โดยการวัดผลนั้นเป็นกระบวนการที่ทำให้เราทราบถึงลักษณะในตัวบุคคลหนึ่ง ซึ่งเป็นไปตามเกณฑ์ที่ได้ตั้งไว้ ส่วนการทดสอบนั้นเป็นกระบวนการสังเกตพฤติกรรมอย่างมีระบบใน เชิงปริมาณ (เยาวดี รางชัยกุล วิบูลย์ศรี, 2556, หน้า 12) โดยระบบการวัดผลและการทดสอบผลสัมฤทธิ์ทางการเรียนของประเทศไทยมีอยู่ด้วยกัน 5 ระดับ ได้แก่ ระดับชั้นเรียน ระดับสถานศึกษา ระดับเขตพื้นที่ ระดับชาติ และระดับนานาชาติ ซึ่งทุกระดับมีความสัมพันธ์เชื่อมโยงกัน และยังสะท้อนให้เห็นถึงคุณภาพของผู้เรียนในแต่ละระดับ จะเห็นได้ว่าข้อมูลผลสัมฤทธิ์ในการทดสอบทางการศึกษาแต่ละระดับสามารถนำไปเป็นแนวทางในพัฒนาการจัดการศึกษาได้อย่างต่อเนื่อง (สัมพันธ์ พันธุ์พฤกษ์, 2557, หน้า 6)

การทดสอบระดับชาติขั้นพื้นฐาน (Ordinary National Educational Test: O-NET) มีความสำคัญมาก เป็นการทดสอบความคิดรวบยอดของนักเรียนชั้นประถมศึกษาปีที่ 6 มัธยมศึกษาปีที่ 3 และมัธยมศึกษาปีที่ 6 ตามมาตรฐานการเรียนรู้ของหลักสูตรการศึกษาขั้นพื้นฐาน โดยทดสอบความรู้ทั้ง 8 กลุ่มสาระการเรียนรู้ ได้แก่ ภาษาไทย คณิตศาสตร์ วิทยาศาสตร์ สังคมศึกษา ศาสนาและวัฒนธรรม สุขศึกษาและพลศึกษา ศิลปะ การงานอาชีพและเทคโนโลยี และภาษาต่างประเทศ ซึ่งจัดสอบโดยสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) หรือ สทศ. (National Institute of Educational Testing Service: NIETS) โดยมีวัตถุประสงค์เพื่อวัดผลสัมฤทธิ์ทางการเรียนของผู้เรียนแต่ละช่วงชั้น และประเมินคุณภาพการศึกษาระดับชาติตามหลักสูตรการศึกษาขั้นพื้นฐานของนักเรียนแต่ละโรงเรียน แล้วนำข้อมูลมาทำแผนพัฒนานักเรียน โดยการทดสอบ O-NET นั้น มีประโยชน์ต่อการศึกษา และต่อบุคคลที่เกี่ยวข้องเป็นอย่างมาก เช่น โรงเรียน ครูผู้สอน นักเรียน และผู้ปกครอง เป็นต้น โดยเฉพาะอย่างยิ่งตัวนักเรียนที่สามารถใช้ประโยชน์จากผลคะแนนที่นักเรียนได้รับ เป็นหลักฐานสำคัญในการศึกษาต่อระดับสูงขึ้นไป (สัมพันธ์ พันธุ์พฤกษ์, 2557, หน้า 7)

การทดสอบ O-NET มีความสำคัญเป็นอย่างยิ่ง จึงทำให้ต้องมีการตรวจสอบคุณภาพของแบบทดสอบที่นำมาใช้ในการทดสอบ ซึ่งต้องอาศัยทฤษฎีในการทดสอบ แบ่งเป็น 2 ประเภท ได้แก่ ทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory: CTT) ทฤษฎีนี้มุ่งตรวจสอบความสัมพันธ์ระหว่างคะแนนที่สังเกตได้กับคะแนนที่แท้จริง และวิเคราะห์คุณภาพโดยรวมของข้อสอบและแบบสอบ โดยตั้งอยู่บนข้อตกลงเบื้องต้นที่ว่า “ความคลาดเคลื่อนที่เกิดจากการวัดมีแบบแผนคงที่เหมือนกันสำหรับทุกกลุ่มบุคคลที่ตอบข้อสอบ” และทฤษฎีการทดสอบแบบใหม่ (Modern Test Theory) สามารถจำแนกออกเป็น 2 แนวทาง ได้แก่ ทฤษฎีการอ้างอิงทางการทดสอบ (Generalizability Theory: G-Theory) ซึ่งจะมุ่งศึกษาความเที่ยงทั่วไปของแบบสอบ และทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT) พยายามวัดคุณลักษณะภายในหรือวัดความสามารถที่แท้จริงของแต่ละบุคคล (ศิริชัย กาญจนวาสี, 2556, หน้า 38) ในการวิจัยครั้งนี้ผู้วิจัย

สนใจที่จะใช้หลักการของทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT) มาใช้เป็นแนวทางในการวิเคราะห์คุณภาพของข้อสอบ O-NET ซึ่งบางส่วนของทฤษฎีการตอบสนอง (IRT) ข้อสอบมีข้อได้เปรียบมากกว่าทฤษฎีการทดสอบแบบดั้งเดิม (CTT) (Gulliksen, 1950) ในการศึกษาเกี่ยวกับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ (CAT) (Hambleton et al., 1991) ทฤษฎีการทดสอบแบบดั้งเดิมยังใช้อยู่ในปัจจุบัน แต่มักจะมีการเชื่อมโยงทฤษฎีการตอบสนองข้อสอบเข้ามาเกี่ยวข้องด้วย (Dimitrov, 2003) แต่ทฤษฎีการทดสอบแนวใหม่ได้รับความนิยมมากปัจจุบัน (Nunnally & Bernstein, 1994) ในส่วนของโมเดลโลจิสติกในทฤษฎีการตอบสนองข้อสอบนั้นมีค่าพารามิเตอร์อธิบายในรูปแบบโค้งคุณลักษณะของข้อสอบ (Item Characteristic Curve: ICC) (Lord & Novick, 1968) ค่าพารามิเตอร์ของข้อสอบจะอยู่ในรูปแบบที่สามารถอธิบายข้อมูลภายใต้เงื่อนไขขนาดของกลุ่มตัวอย่างที่ต้องใช้ในการประมาณค่าพารามิเตอร์ได้อย่างถูกต้อง (Ellis & Mead, 2002) ในการวิเคราะห์ค่าพารามิเตอร์แต่ละครั้งสามารถที่จะใช้กลุ่มตัวอย่างขั้นต่ำที่มีขนาด 200 คน ได้ในบางโปรแกรม เช่น โปรแกรม BILOG เป็นต้น (Ellis & Mead, 2002) การประมาณค่าพารามิเตอร์ ตามหลักทฤษฎีการตอบสนองข้อสอบจะประมาณค่าพารามิเตอร์ ในกลุ่มตัวอย่างที่มีขนาดเล็กที่สุดประมาณ 100 คน (Maydeu-Olivares, 2001) โมเดลทฤษฎีการตอบสนองข้อสอบสามารถประมาณค่าแบบ 1 พารามิเตอร์ แบบ 2 พารามิเตอร์ และแบบ 3 พารามิเตอร์ได้

การวิเคราะห์คุณภาพของข้อสอบนั้น สิ่งที่เราต้องคำนึงถึง ได้แก่ ความเที่ยง (Reliability) ของแบบทดสอบ ความเที่ยง คือ ความคงเส้นคงวาของคะแนนที่ได้จากการทดสอบ ซึ่งสามารถจำแนกได้ 4 ประเภท ได้แก่ ความเที่ยงแบบความคงที่ ความเที่ยงแบบความสมมูล ความเที่ยงแบบความคงที่และสมมูล และความเที่ยงแบบความสอดคล้องภายใน (ศิริชัย กาญจนวาสี, 2556, หน้า 59) ซึ่งการวิจัยครั้งนี้ผู้วิจัยสนใจที่จะตรวจสอบความเที่ยงแบบความสอดคล้องภายใน ที่สามารถประมาณค่าได้โดยวิธีการหาค่าสัมประสิทธิ์แอลฟ่าของครอนบาค เนื่องจากโปรแกรมที่ใช้ในการวิเคราะห์คุณภาพข้อสอบสามารถทำการประมาณค่าในส่วนนี้ได้เลย ทำให้สะดวกในการตรวจสอบความเที่ยงของแบบทดสอบ ในการตรวจสอบคุณภาพของข้อสอบนั้น ถ้าตรวจสอบค่าความเที่ยงอย่างเดียวอาจจะไม่เพียงพอ ผู้วิจัยสนใจที่จะนำแบบทดสอบมาตรวจสอบความตรง (Validity) ของแบบทดสอบ ซึ่งเป็นคุณสมบัติที่สำคัญอีกอย่างหนึ่งของแบบทดสอบนั่นก็หมายความว่าแบบทดสอบฉบับนั้นมีความถูกต้องแม่นยำในสิ่งที่ต้องการจะวัด จำแนกได้ 3 ประเภท คือ ความตรงตามเนื้อหา (Content Validity) ความตรงตามเกณฑ์สัมพันธ์ (Criterion-related Validity) และความตรงเชิงโครงสร้าง (Construct Validity) ซึ่งผู้วิจัยสนใจที่จะตรวจสอบความตรงเชิงโครงสร้าง (Construct Validity)ด้วยการวิเคราะห์องค์ประกอบเชิงยืนยัน (Confirmatory Factor Analysis: CFA) การวัดจะตรงตามคุณลักษณะที่มุ่งวัด และมีความสอดคล้องกับโครงสร้าง (ศิริชัย กาญจนวาสี, 2556, หน้า 103) ในการตรวจสอบความตรงเชิงโครงสร้างนั้นต้องวิเคราะห์โครงสร้างให้สอดคล้องกับโมเดลที่ใช้ในการวัด (Loevinger, 1957) ถึงแม้ว่าการทดสอบความตรงเชิงโครงสร้างนั้นจะอยู่ภายใต้เงื่อนไขของความแตกต่างทางด้านประชากรก็ตาม ในการวิเคราะห์ความตรงเชิงโครงสร้างที่รู้จักกันอีกวิธีการหนึ่งคือการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ (Differential Item Functioning: DIF) (Mellenbergh, 1982)

การทำหน้าที่ต่างกันของข้อสอบ (DIF) นั้นเป็นการเปรียบเทียบผลการตอบข้อสอบระหว่างกลุ่มผู้สอบอย่างน้อย 2 กลุ่มขึ้นไป ปกตินิยมทำการเปรียบเทียบ 2 กลุ่ม ซึ่งกลุ่มแรกเรียกว่า กลุ่มเปรียบเทียบ (Focal Group: F) เป็นกลุ่มที่สนใจศึกษาและคาดว่าจะเป็นกลุ่มที่เสียเปรียบในการตอบข้อสอบ และกลุ่มที่สองเรียกว่า กลุ่มอ้างอิง (Reference group: R) เป็นกลุ่มที่คาดว่าจะได้เปรียบในการตอบข้อสอบได้สอบได้ถูกต้อง (ศิริชัย กาญจนวาสี, 2555, หน้า 120) การทำหน้าที่ต่างกันของข้อสอบ แบ่งเป็น 2 ประเภท ได้แก่ ข้อสอบทำหน้าที่ต่างกันแบบเอกรูป (Uniform DIF) เป็นข้อสอบที่ไม่มีปฏิสัมพันธ์ (Interaction) ระหว่างความสามารถของผู้เข้าสอบกับการเป็นสมาชิกในกลุ่ม โดยโอกาสในการตอบข้อสอบได้ถูกต้องของผู้สอบกลุ่มย่อยกลุ่มหนึ่งสูงกว่ากลุ่มย่อยอีกกลุ่มหนึ่ง และข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป (Nonuniform DIF) เป็นข้อสอบที่มีปฏิสัมพันธ์ระหว่างระดับความสามารถของผู้สอบกับการเป็นสมาชิกกลุ่มย่อย โดยโอกาสของการตอบข้อสอบได้ถูกต้องของผู้สอบกลุ่มย่อยกลุ่มหนึ่งสูงกว่าผู้สอบกลุ่มย่อยอีกกลุ่มหนึ่งไม่ตลอดช่วงความสามารถ (Mellenbergh, 1982)

วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบมีหลายวิธี วิธีที่น่าสนใจวิธีหนึ่งคือ วิธี IRT-LR เป็นวิธีที่ใช้ในการตรวจสอบบนพื้นฐานทฤษฎีการตอบสนองข้อสอบทั้งยังสามารถตรวจพบอัตราความถูกต้องและอัตราความคลาดเคลื่อนได้อย่างมีประสิทธิภาพ (Elosua & Wells, 2013) การประมาณค่าพารามิเตอร์ทั้งในแบบเอกมิติและพหุมิติมีประสิทธิภาพ (Paek & Han, 2012) เมื่อจำแนกขนาดของกลุ่มตัวอย่างพบว่า วิธี IRT-LR สามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบและตรวจพบอัตราความคลาเคลื่อนได้ดีเมื่อเทียบกับที่สุด เมื่อเปรียบเทียบกับวิธีโครงสร้างความแปรปรวน (Mean and Covariance Structures Method: MACS) และวิธีการถดถอยโลจิสติก (Logistic Regression Analysis Method) โปรแกรมที่ใช้ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้แก่ โปรแกรม IRTPRO มีประสิทธิภาพและประสิทธิผลในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบและการคำนวณอัตราความคลาดเคลื่อนสูง ซึ่งทำให้มีความน่าเชื่อถือมากและมีความน่าสนใจเป็นอย่างยิ่ง (Paek & Han, 2012) วิธี SIBTEST เป็นใช้สถิติทดสอบแบบนันพาราเมทริก (Non-Parametric) ซึ่งพัฒนาบนพื้นฐานทฤษฎีการตอบสนองข้อสอบชนิดพหุมิติ ไม่ต้องใช้ฟังก์ชันการตอบสนองข้อสอบหรือการประมาณค่าความสามารถแฝง วิธี SIBTEST ได้รับความนิยมในการประเมินความสามารถของผู้สอบจากคะแนนจริงและมีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในกลุ่มตัวอย่างที่มีขนาดใหญ่ (Gierl, Gotzmann & Boughton, 2004) วิธี SIBTEST มีประสิทธิภาพในการตรวจสอบอัตราความคลาดเคลื่อนสูง เมื่อเปรียบเทียบกับวิธี IRT Likelihood Ratio (IRT-LR) และวิธี Mantel-Haenszel (MH) จุดเด่นของวิธี SIBTEST คือ สามารถคำนวณได้ง่ายไม่ซับซ้อน ประหยัดค่าใช้จ่ายและไม่จำเป็นต้องใช้กลุ่มตัวอย่างที่มีขนาดใหญ่ ทั้งยังใช้สถิติทดสอบ (Narayanan & Swminathan, 1996) แม้ว่าทั้งสองวิธีนี้อาจมีความแตกต่างกันในแง่ของวิธีที่ใช้สถิติในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ซึ่ง วิธี SIBTEST เป็นวิธีที่ใช้สถิตินันพาราเมทริก ส่วนวิธี IRT เป็นวิธีที่ใช้สถิติพาราเมทริก ซึ่งเป็นมิติของสถิติวิเคราะห์เมื่อแบ่งกลุ่มตามมิติลักษณะของตัวแปร จะพบว่าวิธี SIBTEST จะอยู่ในกลุ่มที่มีคุณลักษณะแฝงซึ่งวิเคราะห์บนพื้นฐานของทฤษฎีการตอบสนองข้อสอบ โดยการวิเคราะห์ทั้งแบบพาราเมทริกและนันพาราเมทริกไม่มีความแตกต่างกันภายใต้เงื่อนไขที่แตกต่าง แสดงให้เห็นว่าการตรวจสอบ

การทำหน้าที่ต่างกันของข้อสอบตามหลักการทฤษฎีการตอบสนองข้อสอบเป็นเทคนิควิธีที่กำลังได้รับความนิยม (Basokcu & Ogretmen, 2014) โดยมีประสิทธิภาพของการประมาณค่า และโปรแกรมใช้งานง่าย เป็นที่สนใจของผู้ต้องการใช้มาก

จากเหตุผลที่กล่าวมาผู้วิจัยสนใจที่จะนำแบบทดสอบ O-NET มาตรวจสอบคุณภาพของข้อสอบ และตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) โดยใช้ข้อมูลทุติยภูมิ (Secondary Data) ที่จากสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) หรือ สทศ. เป็นผลการตอบข้อสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ทั้ง 8 กลุ่มสาระการเรียนรู้ ได้แก่ 1) วิชาภาษาไทย 2) วิชาคณิตศาสตร์ 3) วิชาวิทยาศาสตร์ 4) วิชาสังคมศึกษา ศาสนาและวัฒนธรรม 5) วิชาภาษาอังกฤษ 6) วิชาสุขศึกษาและพลศึกษา 7) วิชาศิลปะ และ 8) วิชาการงานพื้นฐานอาชีพและเทคโนโลยี เนื่องจากทั้ง 8 กลุ่มสาระการเรียนรู้เป็นวิชาที่สำคัญในปัจจุบันเป็นอย่างมาก ผู้วิจัยจึงสนใจที่จะศึกษาเพื่อเป็นแนวทางในการพัฒนาแบบทดสอบ และเพื่อเป็นการยกระดับมาตรฐานการจัดการศึกษาของสถาบันการศึกษาต่างๆ ให้มีคุณภาพ และมีประสิทธิผลในการบริหารจัดการศึกษาที่เป็นมาตรฐานเท่าเทียมกันและเป็นข้อมูลให้กับสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) หรือผู้ที่มีส่วนเกี่ยวข้องในการสร้างแบบทดสอบระดับชาติขั้นพื้นฐาน ในระดับอื่น ๆ ได้นำไปใช้พัฒนาแบบทดสอบให้มีความยุติธรรมต่อไป

ดังนั้นผู้วิจัยสนใจที่จะนำแบบทดสอบ O-NET มาวิเคราะห์เพื่อทำการหาคุณภาพของข้อสอบ นอกจากนี้ยังมีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ผู้วิจัยสนใจใช้วิธี SIBTEST เป็นวิธีเกณฑ์ เนื่องจากในการศึกษาระดับสถานศึกษานั้นมีค่าเฉลี่ยของปัจจัยทางเศรษฐกิจและสังคมเข้ามาเกี่ยวข้องเป็นอย่างมาก ปัจจัยในส่วนนี้เป็นส่วนที่ทำให้ผลสัมฤทธิ์ทางการเรียนของผู้เรียนมีความแปรปรวนและเกิดความลำเอียงเนื่องจากอคติทางวัฒนธรรมต่าง ๆ ในระดับโรงเรียน ด้วยสาเหตุนี้เองจึงทำให้ต้องมีการทดสอบเพื่อวัดผลสัมฤทธิ์ทางการเรียน โดยการทดสอบในแต่ละระดับนั้นต้องมีคุณภาพ น่าเชื่อถือ และเป็นไปในมาตรฐานเดียวกัน (Jehangir et al., 2015) โดยผู้วิจัยสนใจที่จะศึกษาข้อมูลจากผลการตอบแบบทดสอบ O-NET ทั้ง 8 กลุ่มสาระการเรียนรู้ ซึ่งเป็นข้อมูลทุติยภูมิจากสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) หรือ สทศ. เนื่องจาก ทั้ง 8 กลุ่มสาระการเรียนรู้เป็นวิชาที่สำคัญในปัจจุบันเป็นอย่างมาก จึงมีความสนใจที่จะศึกษาเพื่อเป็นแนวทางในการพัฒนาและวางนโยบายเพื่อยกระดับมาตรฐานการจัดการศึกษาของสถาบันการศึกษาต่างๆให้มีคุณภาพและประสิทธิผลในการบริหารจัดการศึกษาที่เป็นมาตรฐานเท่าเทียมกันและเป็นข้อมูลให้กับสถาบันทดสอบ หรือผู้ที่มีส่วนเกี่ยวข้องในการสร้างแบบทดสอบระดับชาติขั้นพื้นฐาน ในระดับอื่น ๆ ได้นำไปใช้พัฒนาแบบทดสอบให้มีความยุติธรรมต่อไป

วัตถุประสงค์ของการวิจัย

1. เพื่อวิเคราะห์คุณภาพของข้อสอบ (a, b และ c) ในแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 จำนวน 8 กลุ่มสาระการเรียนรู้ ทั้งฉบับก่อนและฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกันออกจากแบบทดสอบ ตามหลักทฤษฎีการตอบสนองข้อสอบ แบบ 3 พารามิเตอร์

2. เพื่อตรวจสอบความเที่ยงของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 จำนวน 8 กลุ่มสาระการเรียนรู้ ทั้งฉบับก่อนและฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกันออกจากแบบทดสอบ

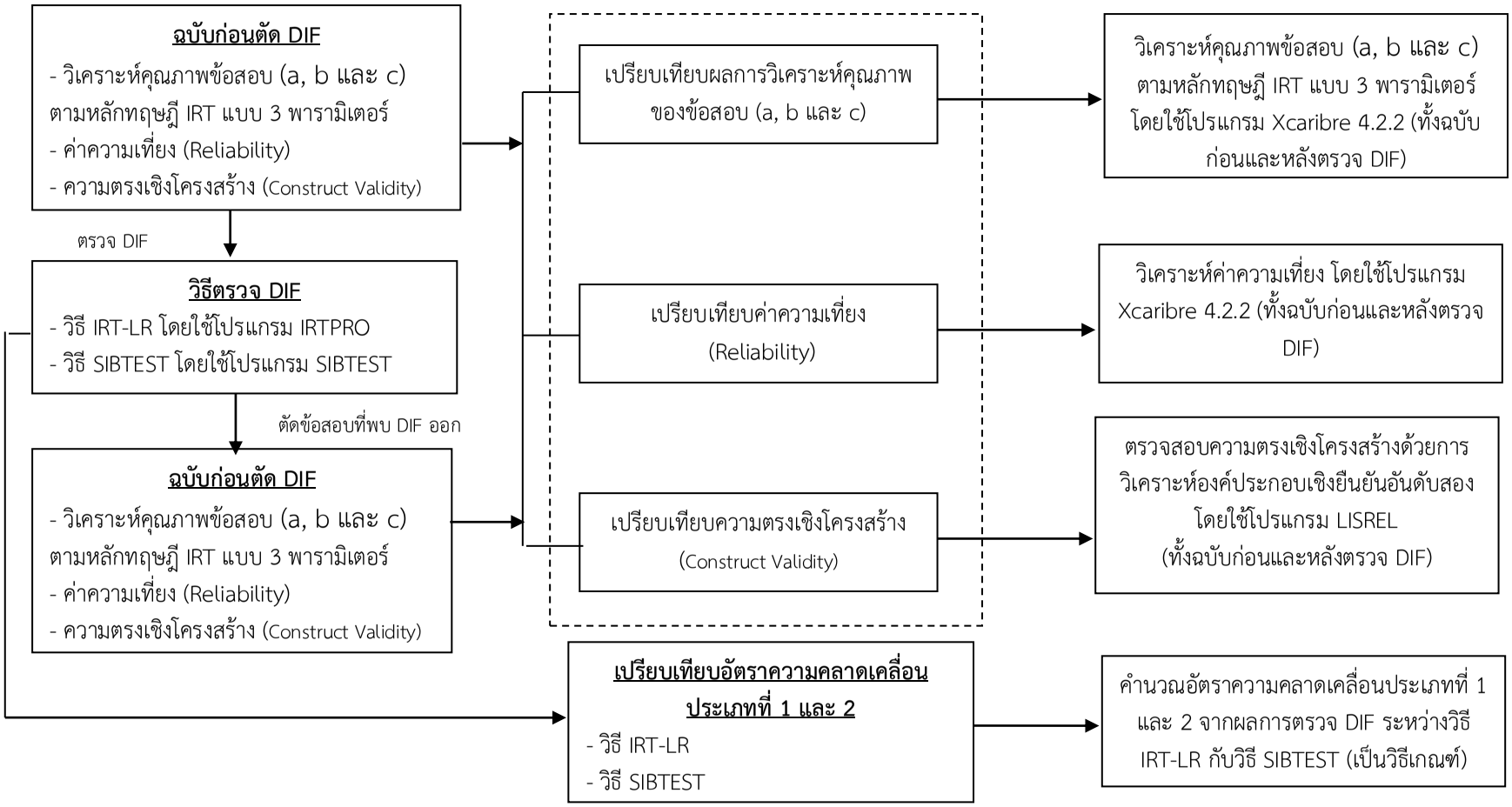
3. เพื่อตรวจสอบความตรงเชิงโครงสร้างของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 จำนวน 8 กลุ่มสาระการเรียนรู้ ทั้งฉบับก่อนและฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกันออกจากแบบทดสอบ

4. เพื่อตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ระหว่างวิธี IRT-LR กับวิธี SIBTEST ฉบับก่อนตัดข้อสอบที่ทำหน้าที่ต่างกันออกจากแบบทดสอบ

5. เพื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 และ 2 ของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ระหว่างวิธี IRT-LR กับวิธี SIBTEST

กรอบแนวคิดการวิจัย

การเปรียบเทียบผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ระหว่างวิธี IRT-LR กับวิธี SIBTEST ครั้งนี้ ผู้วิจัยได้นำผลการตอบข้อสอบ O-NET ทั้ง 8 กลุ่มสาระการเรียนรู้ มาวิเคราะห์คุณภาพของข้อสอบ (a, b และ c) ตามหลักทฤษฎีการตอบสนองข้อสอบ (IRT) แบบ 3 พารามิเตอร์ พร้อมทั้งหาค่าความเที่ยง ความตรงเชิงโครงสร้าง ด้วยการวิเคราะห์องค์ประกอบเชิงยืนยัน ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วย วิธี IRT-LR กับวิธี SIBTEST และเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 และ 2 ของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ดังภาพที่ 1-1



ภาพที่ 1-1 กรอบแนวคิดการเปรียบเทียบผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ระหว่างวิธี IRT-LR กับวิธี SIBTEST

สมมติฐานของการวิจัย

การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบระหว่างวิธี IRT-LR กับวิธี SIBTEST ทั้งสองวิธีอยู่บนพื้นฐานทฤษฎีการตอบสนองข้อสอบ โดยการวิเคราะห์คุณภาพข้อสอบ (a, b และ c) ตามหลักทฤษฎีการตอบสนองข้อสอบ แบบ 3 พารามิเตอร์ ค่าความยากของแบบทดสอบมีแนวโน้มลดลง แต่ค่าอำนาจจำแนก ค่าการเดา ส่วนใหญ่มีแนวโน้มสูงขึ้น (สุนทร เทียนงาม, 2553, หน้า 248) ค่าความเที่ยงของแบบทดสอบมีค่าแตกต่างกันไม่สูงมากนัก ในการตรวจสอบความเที่ยงของแบบทดสอบอาจจะไม่เพียงพอต่อการสนับสนุนว่าแบบทดสอบนั้นมีคุณภาพและเหมาะสมที่จะนำมาใช้ในการทดสอบ (ณรงค์ จันทรมหา, 2554, หน้า 68) การตรวจสอบความตรงโครงสร้างที่ได้จากการวิเคราะห์องค์ประกอบเชิงยืนยันเหมาะสมที่จะเป็นโครงสร้างเกณฑ์ในการทดสอบ เนื่องจากวิธีการวิเคราะห์องค์ประกอบเชิงยืนยัน (CFA) เป็นโครงสร้างที่มีความถูกต้องและสามารถใช้ในการทดสอบได้ ในส่วนการทำหน้าที่ต่างกันของข้อสอบ (Goh & Aryadoust, 2010) วิธี IRT-LR เป็นวิธีที่ใช้สถิติพาราเมทริก แบบ 3 พารามิเตอร์ มีคุณภาพกว่าวิธีเชิงเส้น (จิระนาฎ ฉวีพัฒน์, บัณฑิต อินสมบัติ และ ธีรยุทธ ภูเขา, 2554, หน้า 79) และประสิทธิภาพของการประมาณค่าและการดำเนินงานของโปรแกรมน่าสนใจต่อผู้ใช้ มีประสิทธิภาพและประสิทธิผลของโปรแกรม IRTPRO 2.1 สูง (Paek & Han, 2012) โปรแกรม IRTPRO มีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ เพราะผลของมันคู่ขนานกับโปรแกรม IRTLRDIF และ BILOG-MG 3 (Ong, 2014) ส่วนวิธี SIBTEST เป็นวิธีที่ใช้สถิตินันพาราเมทริก สามารถตรวจข้อสอบที่ทำหน้าที่ต่างกันได้เพียงพอกับขนาดกลุ่มตัวอย่าง 1,000 คน มีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ดี โดยโปรแกรมที่ใช้ในการตรวจสอบคือโปรแกรม SIBTEST ซึ่งมีการพัฒนาให้มีความสะดวกและไม่ซับซ้อนในการใช้งาน (Mark, Andrea & Keith, 2004) เมื่อทดสอบอัตราความคลาดเคลื่อนของวิธี SIBTEST และวิธีแมนเทล-แฮนเซล นั้นมีประสิทธิภาพที่ใกล้เคียงกัน (Roussos & Stout, 1996) ในส่วนของวิธี IRT-LR นั้นตรวจพบอัตราความคลาดเคลื่อนได้ดีที่สุดเมื่อเปรียบเทียบกับวิธีโครงสร้างความแปรปรวน (MACS) และวิธีการถดถอยโลจิสติก (Elosua & Wells, 2013) ดังนั้นผู้วิจัยจึงกำหนดสมมติฐานการวิจัย ดังนี้

1. ผลการวิเคราะห์คุณภาพของข้อสอบ (a, b และ c) ในแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 จำนวน 8 กลุ่มสาระการเรียนรู้ ทั้งฉบับก่อนและฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกันออกจากแบบทดสอบแตกต่างกัน

2. ความเที่ยงของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 จำนวน 8 กลุ่มสาระการเรียนรู้ ทั้งฉบับก่อนและฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกันออกจากแบบทดสอบแตกต่างกัน

3. ความตรงเชิงโครงสร้างของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ทั้งฉบับก่อนและฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกันออกจากแบบทดสอบไม่แตกต่างกัน

4. ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 จำนวน 8 กลุ่มสาระการเรียนรู้ ด้วยวิธี IRT-LR และวิธี SIBTEST มีจำนวนข้อที่พบข้อสอบที่ทำหน้าที่ต่างกันแตกต่างกัน

5. อัตราความคลาดเคลื่อนประเภทที่ 1 และ 2 ของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ระหว่างวิธี IRT-LR กับวิธี SIBTEST ไม่แตกต่างกัน

ประโยชน์ที่คาดว่าจะได้รับจากการวิจัย

1. สถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) หรือ สทศ. นำผลที่ได้จากการวิเคราะห์คุณภาพข้อสอบ (a, b, และ c) ในแบบทดสอบ O-NET ทั้ง 8 กลุ่มสาระการเรียนรู้ ทั้งค่าความเที่ยงของแบบทดสอบทั้งฉบับ ความตรงเชิงโครงสร้าง ว่าข้อสอบทั้ง 8 กลุ่มสาระการเรียนรู้มีความคงที่หรือไม่เมื่อตัด DIF ออกจากแบบทดสอบ มาใช้ในการจัดทำแบบทดสอบ O-NET ครั้งต่อไป เพื่อการวัดความรู้ความสามารถของผู้สอบตามจริง

2. นักวัดผลประเมินผล นำผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ มาเป็นแนวทางในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบและปรับปรุงข้อสอบในแต่ละกลุ่มสาระการเรียนรู้ ในการทดสอบระดับต่างๆ

3. นักวัดผลประเมินผล นำผลการเปรียบเทียบการตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน ว่าวิธีใดตรวจพบได้มากกว่า มีประสิทธิภาพ สะดวก รวดเร็วในการตรวจสอบเพิ่มขึ้น และเป็นทางเลือกในการเลือกใช้วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

ขอบเขตของการวิจัย

การวิจัยนี้มีวัตถุประสงค์เพื่อวิเคราะห์คุณภาพของข้อสอบ (a, b และ c) ตรวจสอบความเที่ยง ความตรงเชิงโครงสร้าง ฉบับก่อนตัดกับฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกันออกจากแบบทดสอบ ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ และคำนวณอัตราความคลาดเคลื่อนประเภทที่ 1 และ 2 ของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ในแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ระหว่างวิธี IRT-LR กับวิธี SIBTEST ซึ่งใช้ข้อมูลทุติยภูมิ (Secondary Data) ที่ได้มาจากสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) โดยมีขอบเขตของการวิจัย ดังนี้

ประชากร เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 6 ซึ่งตอบแบบทดสอบ O-NET ทั้ง 8 กลุ่มสาระการเรียนรู้ ปีการศึกษา 2556 ของสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) จำนวน 413,459 คน

1. เนื้อหา

การวิจัยนี้ใช้ผลการตอบแบบทดสอบ O-NET ทั้ง 8 กลุ่มสาระการเรียนรู้ ชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2556 ซึ่งใช้ข้อมูลทุติยภูมิ (Secondary Data) โดยขอความร่วมมือจาก สถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) หรือ สทศ.

2. ตัวแปรที่ศึกษา

2.1 ตัวแปรต้น ได้แก่

2.1.1 วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ 2 วิธี คือ

2.1.1.1 วิธี IRT-LR

2.1.1.2 วิธี SIBTEST

2.1.2 ชนิดของแบบทดสอบมี 2 ลักษณะ คือ

2.1.2.1 แบบทดสอบฉบับก่อนตัดข้อสอบที่ทำหน้าที่ต่างกันของข้อสอบ (DIF)

2.1.2.2 แบบทดสอบฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกันของข้อสอบ (DIF)

2.2 ตัวแปรตาม ได้แก่

2.2.1 ผลการวิเคราะห์คุณภาพของข้อสอบ (a, b และ c) ฉบับก่อนและหลังตัดข้อสอบที่ทำหน้าที่ต่างกันออกจากแบบทดสอบ

2.2.2 ค่าความเที่ยงฉบับก่อนและหลังตัดข้อสอบที่ทำหน้าที่ต่างกันออกจากแบบทดสอบ

2.2.3 ความตรงเชิงโครงสร้างฉบับก่อนและหลังตัดข้อสอบที่ทำหน้าที่ต่างกันออกจากแบบทดสอบ

2.2.4 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยพิจารณาจากอัตราความคลาดเคลื่อนประเภทที่ 1 และ 2 ของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

นิยามศัพท์เฉพาะ

การวิเคราะห์คุณภาพของข้อสอบ (Quality Analysis of the Test) หมายถึง การตรวจสอบคุณภาพของข้อสอบรายข้อด้วยการประมาณค่าพารามิเตอร์ เพื่อชี้ให้เห็นคุณภาพของข้อสอบที่สำคัญ

ความเที่ยง (Reliabilty) หมายถึง ความคงเส้นคงวาของคะแนนที่ได้จากการทดสอบ ซึ่งเป็นการตรวจสอบค่าความเที่ยงแบบความสอดคล้องภายใน ที่สามารถประมาณค่าได้โดยวิธีการหาค่าสัมประสิทธิ์แอลฟาของครอนบาค โดยได้จากการคำนวณจากโปรแกรมการวิเคราะห์คุณภาพข้อสอบ

ความตรงเชิงโครงสร้าง (Construct Validity) หมายถึง การศึกษาความสัมพันธ์ระหว่างผลการตอบแบบทดสอบของผู้ตอบแบบทดสอบกับโครงสร้าง ซึ่งจะต้องมีความสอดคล้องกันตามความหมายของทฤษฎีที่จะมุ่งวัด การวิจัยในครั้งนี้ใช้วิธีวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสอง (Second-Order Confirmatory Factor Analysis)

การทำหน้าที่ต่างกันของข้อสอบ (Differential Item Functioning: DIF) หมายถึง การที่ข้อสอบที่ทำให้ผู้ตอบที่มีคุณลักษณะเหมือนกันแต่ ลักษณะของกลุ่มย่อยแตกต่างกัน เช่น เพศ เชื้อชาติ ศาสนา วัฒนธรรม สภาพภูมิศาสตร์ เศรษฐกิจ เป็นต้น มีโอกาสที่จะตอบข้อสอบได้ถูกต้องแตกต่างกัน เกิดการได้เปรียบเสียเปรียบกันในระหว่างกลุ่มผู้สอบ โดยกลุ่มที่คาดว่าจะเป็นกลุ่มได้เปรียบเรียกว่า กลุ่มอ้างอิง (Reference Group: R) และกลุ่มที่คาดว่าจะเป็นกลุ่มที่เสียบเปรียบเรียกว่า (Focal Group: F)

การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (The Detecting of the Differential Item Functioning) หมายถึง การวิเคราะห์ข้อมูลด้วยกระบวนการทางสถิติ เพื่อให้ทราบถึงการทำหน้าที่ต่างกันของข้อสอบโดยใช้กลุ่มผู้ตอบแบบทดสอบที่มีความสามารถในระดับเดียวกัน แต่กลุ่มคุณลักษณะของกลุ่มแตกต่างกัน ซึ่งงานวิจัยในครั้งนี้ใช้วิธีการตรวจสอบแบบสองค่า

คุณภาพของการทำหน้าที่ต่างกันของข้อสอบ (The Quality of the Differential Item Functioning) หมายถึง ความถูกต้องของการระบุข้อสอบที่ทำหน้าที่ต่างกันโดยการวิเคราะห์วิธี IRT-LR ซึ่งพิจารณาได้จากอัตราความคลาดเคลื่อนประเภทที่ 1 และ 2 เมื่อเปรียบเทียบกับผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี SIBTEST

วิธี IRT- LR (Linear Regression in Item Response Theory Method) หมายถึง วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบซึ่งอยู่บนพื้นฐานทฤษฎีการตอบสนองข้อสอบเป็นการวัดที่อธิบายความสัมพันธ์ระหว่างคุณลักษณะภายในหรือความสามารถที่มีอยู่ภายในตัวบุคคลกับพฤติกรรมการตอบสนองข้อสอบของบุคคลนั้นว่ามีโอกาสตอบข้อสอบถูกมากน้อยเพียงใด โดยอธิบายความสัมพันธ์ดังกล่าวในรูปของฟังก์ชันคณิตศาสตร์หรือโมเดลที่แสดงความสัมพันธ์ระหว่างระดับความสามารถ คุณลักษณะของข้อสอบ และโอกาสของการตอบข้อสอบได้ถูก

วิธี SIBTEST (SIBTEST Method) หมายถึง วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยใช้คะแนนจริงจากผู้ตอบแบบทดสอบ โดยคำนวณจากสัดส่วนของผู้ที่ตอบแบบทดสอบถูกระหว่างกลุ่มเปรียบเทียบและกลุ่มอ้างอิงที่ใช้แบบทดสอบเดียวกัน การทดสอบจะมีนัยสำคัญด้วยค่าสถิติ

อัตราความคลาดเคลื่อน (Error Rate) หมายถึง ความคลาดเคลื่อนที่เกิดจากการตัดสินใจที่จะปฏิเสธหรือยอมรับสมมุติฐานทางสถิติ โดยเป็นสัดส่วนของจำนวนข้อสอบที่ระบุ ผิดพลาด มี 2 ประเภท คือ อัตราความคลาดเคลื่อนประเภทที่ 1 และอัตราความคลาดเคลื่อนประเภทที่ 2

อัตราความคลาดเคลื่อนประเภทที่ 1 (Type I Error) หมายถึง การระบุความผิดพลาดของข้อสอบที่ทำหน้าที่ต่างกันทั้ง ๆ ที่ในความเป็นจริงข้อสอบทำหน้าที่ไม่ต่างกัน ในการวิจัยนี้คำนวณจากสัดส่วนของจำนวนข้อสอบที่ระบุผิดพลาดว่าทำหน้าที่ต่างกันต่อจำนวนข้อสอบที่ทำหน้าที่ไม่ต่างกันทั้งหมดในแบบทดสอบ

อัตราความคลาดเคลื่อนประเภทที่ 2 (Type II Error) หมายถึง การระบุความผิดพลาดของข้อสอบที่ทำหน้าที่ต่างกันทั้ง ๆ ที่ในความเป็นจริงข้อสอบทำหน้าที่ไม่ต่างกัน ในการวิจัยนี้คำนวณจากสัดส่วนของจำนวนข้อสอบที่ระบุผิดพลาดว่าทำหน้าที่ไม่ต่างกันต่อจำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งหมดในแบบทดสอบ

ค่าอำนาจจำแนกของข้อสอบ (Discrimination Parameter: a) หมายถึง ค่าที่เป็นสัดส่วนโดยตรงกับความชันของโค้งคุณลักษณะของข้อสอบ ณ จุดเปลี่ยนโค้ง มีตั้งแต่-α ถึง α แต่ในทางปฏิบัติ มีค่าตั้งแต่ 0.5 -2.5 เพราะค่าอำนาจจำแนกที่เป็นลบแสดงว่าข่อสอบข้อนั้นไม่ดีต้องตัดทิ้ง ค่าต่ำกว่า 0.5 แสดงว่าข้อสอบไม่มีอำนาจจำแนก ค่า 2.5 แสดงว่าข้อสอบมีค่าอำนาจจำแนกสูงในการคัดเลือกข้อสอบ ข้อสอบที่คัดไว้จะมีค่าอำนาจจำแนกตั้งแต่ 0.5 ขึ้นไป

ค่าความยากของข้อสอบ (Difficulty Parameter: b) หมายถึง สัดส่วนของคนที่ทำข้อสอบนั้นได้ถูกหรือหมายถึงค่าที่แสดงถึงระดับความสามารถของผู้สอบ (θ) ที่จุดโค้งลักษณะข้อสอบมีความชัน มากที่สุดมีค่าตั้งแต่-α ถึง α แต่ในทางปฏิบัติจะอยู่ระหว่าง -2.5 ถึง 2.5 ค่า -2.5 บ่งบอกว่าข้อสอบนั้นง่ายมาก และ ค่า 2.5 บ่งบอกว่าข้อสอบนั้นยากมาก

ค่าโอกาสการเดาของข้อสอบ (Guessing Parameter: c) หมายถึง ความน่าจะเป็นของบุคคลหนึ่งที่ไม่มีความสามารถในการตอบข้อสอบนั้นได้ถูกต้อง เป็นค่าที่แสดงถึงโอกาสการตอบข้อสอบถูก โดยไม่มีความรู้ในเรื่องนั้นๆ มีค่าจาก 0 ถึง 1 จะคัดเลือกเอาข้อสอบที่มีค่าโอกาสการเดาต่ำกว่า 0.3

กลุ่มอ้างอิง (Reference Group: R) หมายถึง กลุ่มผู้สอบที่คาดว่าจะเป็นกลุ่มที่ได้เปรียบในการตอบข้อสอบ คือกลุ่มที่มีความน่าจะเป็นในการตอบข้อสอบได้สูงกว่าผู้สอบอีกกลุ่มหนึ่ง

กลุ่มเปรียบเทียบ (Focal Group: F) หมายถึง กลุ่มผู้สอบที่คาดว่าจะเป็นกลุ่มที่เสียเปรียบในการตอบข้อสอบ คือกลุ่มที่มีความน่าจะเป็นในการตอบข้อสอบได้ต่ำกว่าอีกกลุ่มหนึ่ง

การทดสอบระดับชาติขั้นพื้นฐาน (Ordinary National Education Test: O-NET) หมายถึง การประเมินคุณภาพการศึกษาตามตัวชี้วัดขั้นพื้นฐาน โดยมีหน่วยงานที่ดำเนินการคือ สถาบันทดสอบทางการศึกษาแห่งชาติ(องค์การมหาชน) โดยใช้แบบทดสอบ O-NET ซึ่งวัดกระบวนการคิดรวบยอดของผู้เรียน ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พ.ศ. 2551 ในการ วิจัยครั้งนี้ ใช้แบบทดสอบ O-NET ของชั้นมัธยมศึกษาปีที่ 6 ทั้ง 8 กลุ่มสาระการเรียนรู้ ปีการศึกษา 2556

แบบทดสอบ O-NET หมายถึง แบบทดสอบที่ใช้วัดผลสัมฤทธิ์ทางการศึกษาของโรงเรียนในสังกัดต่างๆ ให้เป็นมาตรฐานเดียวกันทั่วประเทศ ที่จัดสอบโดย สถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) หรือ สทศ. ซึ่ง ชั้นมัธยมศึกษาปีที่ 6 ประกอบด้วย 8 กลุ่มสาระการเรียนรู้ ได้แก่ 1) ภาษาไทย 2) สังคมศึกษา ศาสนาและวัฒนธรรม 3) ภาษาอังกฤษ 4) คณิตศาสตร์ 5) วิทยาศาสตร์ 6) สุขศึกษาและพลศึกษา 7) ศิลปะ 8) การงานอาชีพและเทคโนโลยี

แบบทดสอบฉบับก่อนตัด DIF หมายถึง แบบทดสอบ O-NET ที่จัดสอบโดย สถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) หรือ สทศ. ชั้นมัธยมศึกษาปีที่ 6 จำนวน 8 กลุ่มสาระการเรียนรู้

แบบทดสอบฉบับหลังตัด DIF หมายถึง แบบทดสอบ O-NET ที่ตัดข้อที่ตรวจพบ DIF ออกจากการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธี IRT-LR และวิธี SIBTEST

บทที่ 2
เอกสารและงานวิจัยที่เกี่ยวข้อง

การวิจัยนี้มีวัตถุประสงค์เพื่อวิเคราะห์คุณภาพของข้อสอบ (a, b และ c) ในแบบทดสอบ O-NET ตรวจสอบความเที่ยง ความตรงเชิงโครงสร้าง ฉบับก่อนและฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกันออกจากแบบทดสอบ ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบและเปรียบเทียบอัตราความคลาดเคลื่อนของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ระหว่างวิธี IRT-LR กับวิธี SIBTEST ผู้วิจัยแบ่งการนำเสนอเป็น 6 ตอน ดังนี้

ตอนที่ 1 การทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน (O-NET) และงานวิจัยที่เกี่ยวข้อง

ตอนที่ 2 ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT) และงานวิจัยที่เกี่ยวข้อง

ตอนที่ 3 การวิเคราะห์ความตรงเชิงโครงสร้าง (Construct Validity) และงานวิจัยที่เกี่ยวข้อง

ตอนที่ 4 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) และงานวิจัยที่เกี่ยวข้อง

ตอนที่ 5 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ด้วยวิธี IRT-LR และงานวิจัยที่เกี่ยวข้อง

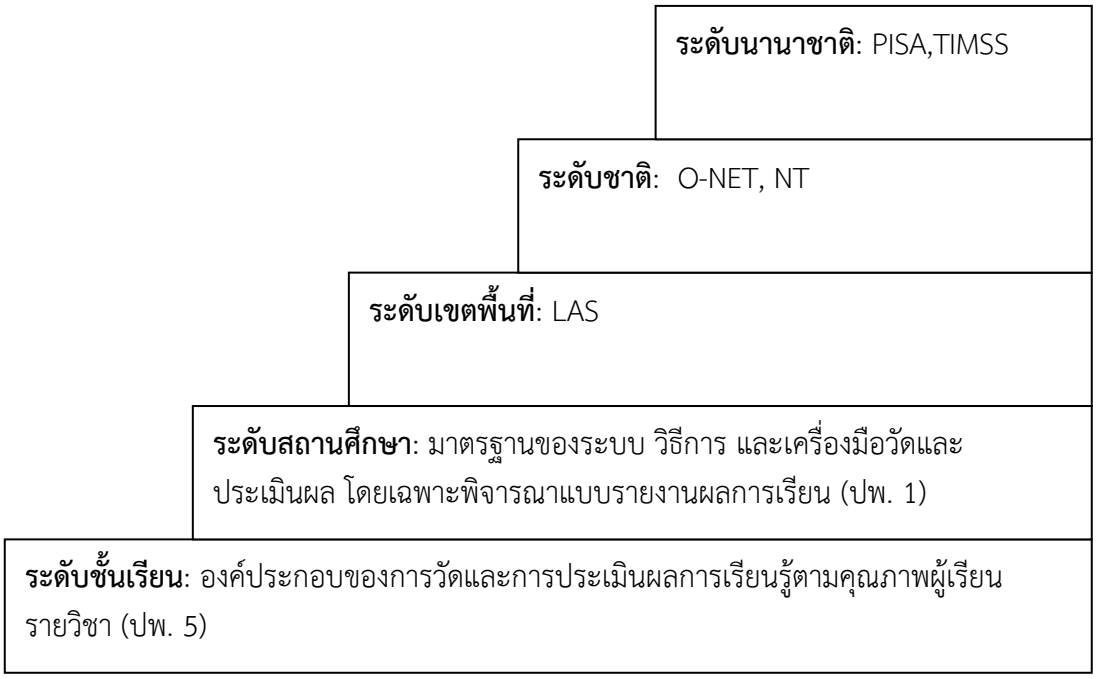
ตอนที่ 6 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ด้วยวิธี SIBTEST และงานวิจัยที่เกี่ยวข้อง

ตอนที่ 1 การทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน (O-NET) และงานวิจัยที่เกี่ยวข้อง

1. ความเป็นมาของการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน

สัมพันธ์ พันธ์พฤกษ์ (2557) ได้กล่าวถึงความเป็นมาของการทดสอบระดับชาติไว้ดังนี้ การทดสอบระดับชาติเริ่มมาตั้งแต่ พ.ศ. 2478 ถึง พ.ศ. 2520 ถือว่าเป็นการกระจายอำนาจให้โรงเรียน โดยมีระบบ Accountability กล่าวคือ ผู้จบชั้นประโยค ต้องสอบได้คะแนนทดสอบระดับชาติไม่ต่ำกว่าร้อยละ 50 แต่เมื่อปีการศึกษา 2521 จนถึงปัจจุบัน ได้มอบให้โรงเรียนแต่ละโรงเรียนทำหน้าที่ในการทดสอบตัวประโยคแทนการสอบกลางระดับชาติ โดยให้คณะกรรมการการศึกษาจังหวัดพิจารณาอนุญาตให้โรงเรียนที่อยู่ขั้นมาตรฐานให้มีการเลื่อนชั้นอัตโนมัติ ซึ่งเป็นการกระจายอำนาจแบบไม่มี Accountability และไม่สามารถเทียบเคียงคุณภาพผู้เรียนของแต่ละสถานศึกษาได้ เนื่องจากแต่ละสถานศึกษาใช้การทดสอบต่างกัน (สัมพันธ์ พันธุ์พฤกษ์, 2557, หน้า 9) ซึ่งในปีการศึกษา 2554 สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน ได้กำหนดให้สำนักงานเขตพื้นที่การศึกษาใช้ผลสอบ O-NET เป็นเครื่องมือการยกระดับคุณภาพการศึกษาทั้งระบบ ส่งผลทำให้มีการพัฒนาผู้เรียนด้านผลสัมฤทธิ์ทางการเรียนอย่างเป็นรูปธรรม ทั้งนี้ คะแนนสอบ O-NET สามารถนำไปใช้ในการทำนายผลคะแนนความรู้ด้านวิทยาศาสตร์ คณิตศาสตร์ เทคโนโลยี และการอ่านจากการทดสอบตามโครงการ PISA ได้

ระบบการทดสอบทางการศึกษาของไทยและนานาชาติ มีอยู่ด้วยกันทั้งหมด 5 ระดับ คือ ระดับชั้นเรียน เป็นการประเมินคุณภาพผู้เรียนตามมาตรฐาน ตัวชี้วัด ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 ระดับต่อมาคือ ระดับสถานศึกษา เป็นการประเมินผลการเรียนของผู้เรียนเป็นรายภาค หรือรายปี โดยสถานศึกษาจะรายงานผลการเรียนนักเรียนรายบุคคลเมื่อผู้เรียนจบการศึกษาแต่ละระดับชั้นหรือเมื่อนักเรียนย้ายสถานศึกษาด้วยระเบียนแสดงผลการเรียน (ปพ.1) เพื่อใช้แสดงผลการเรียน ระดับต่อมาเป็น ระดับเขตพื้นที่ เป็นการประเมินคุณภาพผู้เรียนโดยเขตพื้นที่การศึกษาตามมาตรฐานการเรียนรู้ของหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน เพื่อใช้เป็นข้อมูลพื้นฐานในการพัฒนา คุณภาพการศึกษาของเขตพื้นที่การศึกษา อีกระดับ คือ ระดับชาติ เป็นการประเมินผู้เรียนตามมาตรฐานและตัวชี้วัดตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน ที่ดำเนินการโดยหน่วยงานภายในประเทศ 2 หน่วยงาน ได้แก่ สถานบันทดสอบทางการศึกษาแห่งชาติ(องค์การมหาชน) ซึ่งหน่วยงานนี้ได้ ดำเนินการทดสอบระดับชาติขั้นพื้นฐาน (Ordinary National Education Test: O-NET) และอีกหน่วยงานคือ สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน (สพฐ.) กระทรวงศึกษาธิการ ดำเนินการประเมินคุณภาพผู้เรียน ตามมาตรฐานและตัวชี้วัดตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน เรียกว่าการทดสอบระดับชาติ (National Testing: NT) การทดสอบระดับสุดท้ายคือ ระดับนานาชาติ เป็นการประเมินเพื่อเทียบเคียงคุณภาพของผู้เรียนระดับนานาชาติ ได้แก่ การประเมินของ PISA (Programme for International Student Assessment) และการประเมินของ TIMSS (Trends in International Mathematics and Science Study) กล่าวคือการประเมินผลทุกระดับเป็นการตรวจสอบความก้าวหน้าทางการเรียนรู้ของนักเรียนเพื่อนำไปใช้ในการประเมินผลเพื่อการประกันและพัฒนาคุณภาพการศึกษาให้เกิดความต่อเนื่องและเชื่อมโยงในทุกระดับ ดังภาพในแผนภาพที่ 2-1

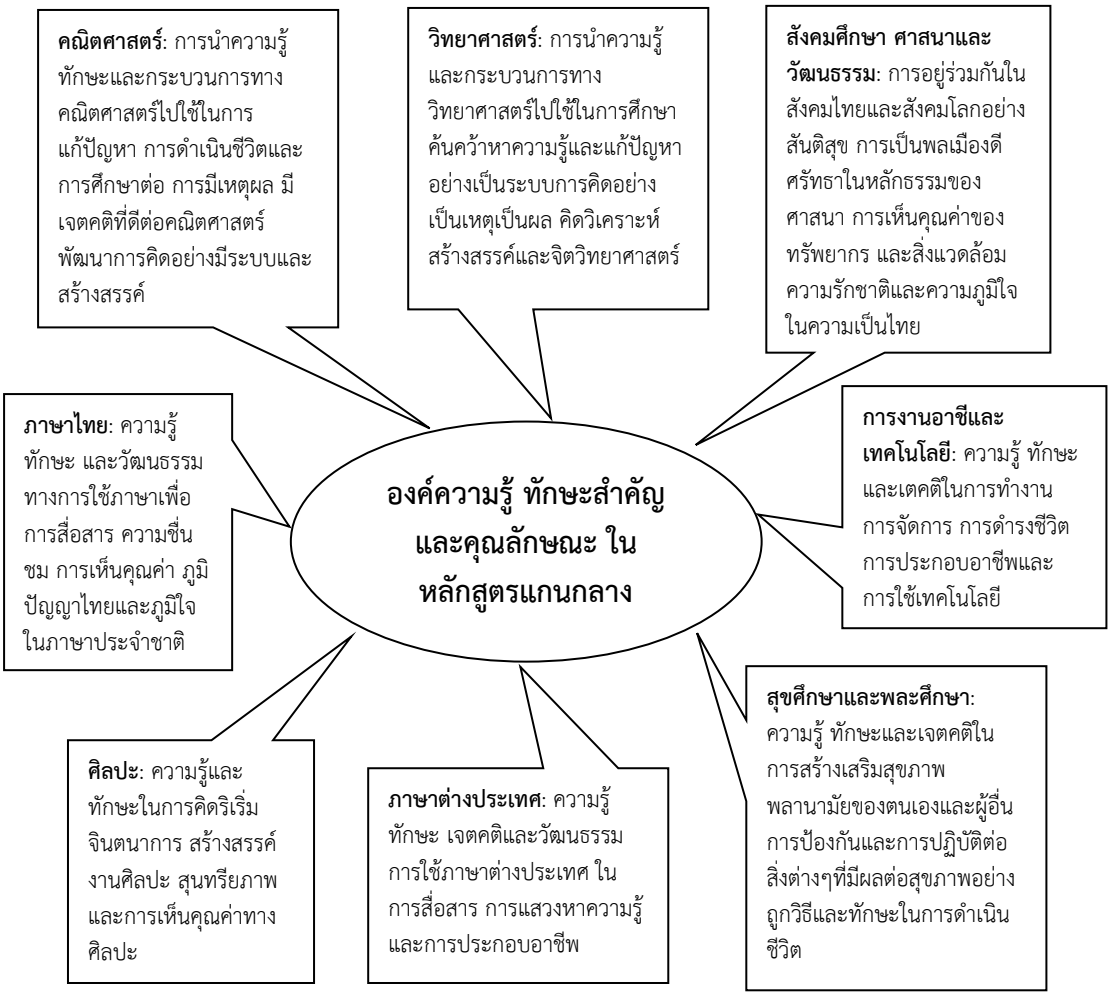


ภาพที่ 2-1 ระบบการทดสอบทางการศึกษาของประเทศไทย (สัมพันธ์ พันธุ์พฤกษ์, 2557, หน้า 6)

ในการวิจัยครั้งนี้ผู้วิจัยได้ใช้การทดสอบระดับชาติขั้นพื้นฐานในการศึกษาซึ่งการทดสอบ (O-NET) ถือว่ามีความสำคัญมากกับทั้ง ผู้เรียน ครู ผู้บริหาร และสถานศึกษา เป็นต้น O–NET ย่อมาจาก Ordinary National Educational Test คือ แบบสอบทางการศึกษาแห่งชาติขั้นพื้นฐาน เป็นการวัดผลการจัดการศึกษาขั้นพื้นฐาน จัดสอบ 8 กลุ่มสาระการเรียนรู้ ได้แก่ ภาษาไทย สังคมศึกษา ศาสนาและวัฒนธรรม ภาษาต่างประเทศ (ภาษาอังกฤษ) คณิตศาสตร์ วิทยาศาสตร์ สุขศึกษา ศิลปะ และ การงานอาชีพและเทคโนโลยี ส่วนลักษณะข้อสอบและการประเมินผล O–NET ประกอบด้วยแบบทดสอบจะมีทั้งปรนัย และอัตนัย ในอัตราส่วนระหว่าง 80% – 90% : 10% – 20% ข้อสอบแบบปรนัยจะเป็นข้อสอบแบบ 4 ตัวเลือก สำหรับข้อสอบอัตนัยจะเป็นข้อสอบแบบเขียนคำตอบสั้นๆ (Short Answer) เวลาในการทำข้อสอบวิชาละ 2 ชั่วโมง ข้อสอบแต่ละข้อ คะแนนอาจจะไม่เท่ากันขึ้นอยู่กับความยากง่ายของข้อสอบข้อสอบครอบคลุมสาระและทักษะสำคัญของ 5 กลุ่มสาระการเรียนรู้และการสอบเป็นบริการของรัฐให้แก่นักเรียนทุกคนโดยไม่ต้องเสียค่าใช้จ่ายใด ๆ (สถาบันทดสอบทางการศึกษาแห่งชาติ, 2557) ในการทดสอบ O-NET นั้น ช่วยยกระดับคุณภาพผู้เรียนโดยเฉพาะชั้นประถมศึกษาปีที่ 6 ชั้นมัธยมศึกษาปีที่ 3 และชั้นมัธยมศึกษาปีที่ 6 เพราะมีการนำผลการทดสอบ O-NET ไปใช้เป็นส่วนหนึ่งของการคัดเลือกเข้าศึกษาต่อในระดับชั้นมัธยมศึกษาปีที่ 1 และ ชั้นมัธยมศึกษาปีที่ 4 และในสถาบันอุดมศึกษาตามลำดับ

2. สาระการเรียนรู้ในหลักสูตรแกนกลางการศึกษาระดับชาติขั้นพื้นฐาน พ.ศ. 2551

การะบวนการจัดการศึกษา ประกอบไปด้วย 3 องค์ประกอบ ที่มีความสำคัญเชื่อมโยงกัน คือ จุดมุ่งหมายการศึกษา การจัดกระบวนการเรียนรู้ และ การวัดผลประเมินผลการศึกษา โดยใช้ทั้ง 3 องค์ประกอบ เป็นข้อมูลสารสนเทศ ที่ได้จากการวัดและประเมินผลซึ่งสะท้อนให้เห็นถึง คุณภาพจากการจัดการเรียนรู้ที่เป็นเป้าหมายและความเหมาะสมของหลักสูตร จึงนำไปสู่การปฏิรูปหลักสูตรและการจัดการเรียนรู้ ก่อนการจัดการเรียนรู้ ครูจะต้องทำการวิเคราะห์หลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พ.ศ. 2551 โดยเฉพาะอย่างยิ่ง วัตถุประสงค์ และเป้าหมายในหลักสูตร จากตัวชี้วัดและสาระการเรียนรู้แกนกลางของแต่ละกลุ่มสาระการเรียนรู้ พร้อมทั้งดูคุณภาพของผู้เรียนในแต่ละสาระ ซึ่งผลการวิเคราะห์ดังกล่าวนั้นนำไปสู่การออกแบบการเรียนรู้ การเขียนแผนการสอน และเป้าหมายของครูผู้สอนที่ชัดเจนมากยิ่งขึ้น โดยการพัฒนาคุณภาพผู้เรียนตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน มี 4 องค์ประกอบ ได้แก่ การเรียนรู้ 8 กลุ่มสาระการเรียนรู้ คุณลักษณะอันพึงประสงค์ การอ่านคิดวิเคราะห์และเขียน และกิจกรรมพัฒนาผู้เรียน ซึ่งสาระการเรียนรู้ในหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พ.ศ. 2551 ได้กำหนดให้ผู้เรียนทุกคนในระดับการศึกษาต้องเรียนรู้แบ่งเป็น 8 กลุ่มสาระการเรียนรู้ โดยประกอบด้วย องค์ความรู้ ทักษะ หรือกระบวนการเรียนรู้และคุณลักษณะอันพึงประสงค์ ดังภาพที่ 2-2



ภาพที่ 2-2 สาระการเรียนรู้ในหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พ.ศ. 2551
(สัมพันธ์ พันธุ์พฤกษ์, 2557, หน้า 4)

หลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พ.ศ. 2551 ได้กำหนดวิสัยทัศน์ของหลักสูตรไว้ว่า มุ่งพัฒนาผู้เรียน ให้เป็นมนุษย์โดยสมบูรณ์ทั้งด้านร่างกาย ความรู้ คุณธรรม มีจิตสำนึกในความเป็นพลเมืองไทยและเป็นพลโลกโดยมุ่งเน้นที่ผู้เรียนเป็นสำคัญบนพื้นฐานความเชื่อว่า ทุกคนสามารถพัฒนาตนเองได้อย่างเต็มศักยภาพ (กระทรวงศึกษาธิการ, 2551) สำหรับการวัดผลและการประเมินผลสาระการเรียนรู้ทั้ง 8 กลุ่มสาระการเรียนรู้นั้น หลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พ.ศ. 2551 ได้กำหนดไว้ว่า ผู้เรียนจะต้องผ่านครบทุกตัวชี้วัดในแต่ละระดับชั้นโดยกลุ่มสาระการเรียนรู้ภาษาไทย มีทั้งหมด 5 สาระ ได้แก่ สาระที่ 1 การอ่าน สาระที่ 2 การเขียน สาระที่ 3 การฟัง การดูและการพูด สาระที่ 4 หลักหารใช้ภาษาไทย และสาระที่ 5 วรรณคดีและวรรณกรรม ซึ่งในระดับชั้น ม. 6 มีทั้งหมด 36 ตัวชี้วัด ในการตัดสินผล กลุ่มสาระการเรียนรู้คณิตศาสตร์ มีทั้งหมด 6 สาระ ได้แก่ สาระที่ 1 จำนวนและการดำเนินการ สาระที่ 2 การวัด สาระที่ 3 เรขาคณิต สาระที่ 4

พิชคณิต สาระที่ 5 การวิเคราะห์ข้อมูลและความน่าจะเป็น สาระที่ 6 ทักษะและกระบวนการทางคณิตศาสตร์ ซึ่งในระดับชั้น ม.6 มีทั้งหมด 32 ตัวชี้วัด กลุ่มสาระการเรียนรู้วิทยาศาสตร์ มีทั้งหมด 8 สาระ ได้แก่ สาระที่ 1 สิ่งมีชีวิตกับกระบวนการดำรงชีวิต สาระที่ 2 ชีวิตกับสิ่งแวดล้อม สาระที่ 3 สาร และสมบัติของสาร สาระที่ 4 แรง และการเคลื่อนที่ สาระที่ 5 พลังงาน สาระที่ 6 กระบวนการเปลี่ยนแปลงของโลก สาระที่ 7 ดาราศาสตร์ และอวกาศ สาระที่ 8 ธรรมชาติของวิทยาศาสตร์และเทคโนโลยีซึ่งในระดับชั้น ม.6 มีทั้งหมด 67 ตัวชี้วัด กลุ่มสาระการเรียนรู้สังคมศึกษา ศาสนาและวัฒนธรรม มีทั้งหมด 5 สาระได้แก่ สาระที่ 1 ศาสนา ศิลธรรม จริยธรรม สาระที่ 2 หน้าที่พลเมือง วัฒนธรรมและการดำเนินชีวิตในสังคม สาระที่ 3 เศรษฐศาสตร์ สาระที่ 4 ประวัติศาสตร์ และสาระที่ 5 ภูมิศาสตร์ ซึ่งในระดับชั้น ม.6 มีทั้งหมด 63 ตัวชี้วัด กลุ่มสาระการเรียนรู้สุขศึกษา และพลศึกษา มีทั้งหมด 5 สาระได้แก่ สาระที่ 1 การเจริญเติบโตและพัฒนาการของมนุษย์ สาระที่ 2 ชีวิตและครอบครัว สาระที่ 3 การเคลื่อนไหว การออกกำลังกาย การเล่นเกม กีฬาไทยและกีฬาสากล สาระที่ 4 การสร้างเสริมสุขภาพ สมรรถภาพและการป้องกันโรค และสาระที่ 5 ความปลอดภัยในชีวิต ซึ่งในระดับชั้น ม.6 มีทั้งหมด 29 ตัวชี้วัด กลุ่มสาระการเรียนรู้ศิลปะ มีทั้งหมด 3 สาระได้แก่ สาระที่ 1ทัศนศิลป์ สาระที่ 2 ดนตรี สาระที่ 3 นาฏศิลป์ ซึ่งในระดับชั้น ม. 6 มีทั้งหมด 41 ตัวชี้วัด กลุ่มสาระการเรียนรู้การงานอาชีพและเทคโนโลยี มีทั้งหมด 4 สาระได้แก่ สาระที่ 1การดำรงชีวิตและครอบครัว สาระที่ 2 การออกแบบและเทคโนโลยี สาระที่ 3 เทคโนโลยีสาระสนเทศและการสื่อสาร และสาระที่ 4 การอาชีพ ซึ่งในระดับชั้น ม.6 มีทั้งหมด 29 ตัวชี้วัด กลุ่มสาระการเรียนรู้ภาษาต่างประเทศ มีทั้งหมด 4 สาระได้แก่ สาระที่ 1 ภาษาเพื่อการสื่อสาร สาระที่ 2 ภาษาและวัฒนธรรม สาระที่ 3 ภาษากับความสัมพันธ์กับกลุ่มสาระการเรียนรู้อื่น และสาระที่ 4 ภาษากับความสัมพันธ์กับชุมชนและโลก ซึ่งในระดับชั้น ม.6 มีทั้งหมด 21 ตัวชี้วัด ในแต่ละกลุ่มสาระการเรียนรู้ครูผู้สอนจะต้องเข้าใจหลักสูตรอย่างถ่องแท้ซึ่งสอดคล้องกับผลการวิจัยของ วีรวรรณ โฆษิตสุคต (2557) พบว่าการที่จะพัฒนาแบบทดสอบให้มีคุณภาพนั้นครูผู้สอนจะต้องศึกษาหลักสูตรให้เข้าใจอย่างถ่องแท้เพื่อให้ข้อสอบนั้นมีคุณภาพและสามารถวัดผลสัมฤทธิ์ทางการเรียนตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พ.ศ. 2551 อย่างแท้จริง

โดยการวัดและการประเมินผลคุณภาพของผู้เรียนได้ถูกกำหนดไว้ในหลักสูตรแต่ละกลุ่มสาระการเรียนรู้ที่ต้องสอดคล้องกับการเรียนการสอนและวัตถุประสงค์ของหลักสูตร และวัตถุประสงค์ของกลุ่มสาระการเรียนรู้ที่เป็นเป้าหมายคุณภาพของผู้เรียนด้วย

3. การใช้ผลการทดสอบระดับชาติขั้นพื้นฐาน (O-NET)

การวัดและประเมินผลการศึกษามีอยู่ด้วยกันหลายระดับดังที่ได้กล่าวไปแล้วในข้างต้น ดังนั้นจากฐานความคิดของฟังก์ชั่นการผลิตทางการศึกษา (Educational Production Function) ที่คำนึงถึงความสัมพันธ์ระหว่างปัจจัยนำเข้าและผลผลิตทางการศึกษา สามารถนำมาใช้ในการวางแผนเพื่อพัฒนาคุณภาพการศึกษาได้ ด้วยการวิเคราะห์ผลการทดสอบที่ผ่านมาเทียบกับเกณฑ์เพื่อนำมากำหนดเป้าหมายในการพัฒนา การวิเคราะห์ปัจจัยที่ส่งผลต่อคะแนนผลสัมฤทธิ์ทางการเรียนของผู้เรียน เพื่อนำมาวางแผนการจัดการเรียนรู้การวัดและการประเมินผลเพื่อให้เกิดการพัฒนาผู้เรียนได้อย่างยั่งยืนผลการทดสอบของสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) เป็นข้อมูลและสารสนเทศที่สำคัญให้คนไทยและผู้รับผิดชอบการศึกษาของชาติ นำไปใช้ใน

การปรับปรุงและพัฒนาการเรียนการสอน การส่งเสริมและพัฒนาให้การวัดและประเมินผลการเรียนรู้ของโรงเรียนมีมาตรฐานนำไปประยุกต์ใช้ทดสอบในชั้นเรียน การส่งเสริมและพัฒนาให้การวัดและประเมินผลการเรียนรู้ของโรงเรียนมีมาตรฐานนำไปประยุกต์ใช้ทดสอบในชั้นเรียน เพื่อใช้ในการปรับปรุงคุณภาพการเรียนการสอนให้ได้มาตรฐานอย่างต่อเนื่อง การนำผลการทดสอบ O-NET ไปใช้ในการพัฒนาคุณภาพการศึกษา เพื่อให้เกิดการพัฒนาอย่างยั่งยืน ควรเริ่มจากการสร้างความรู้ความเข้าใจต่อผู้มีส่วนได้ส่วนเสีย (Stakeholder) เกี่ยวกับสารสนเทศผลการทดสอบของผู้เรียน ได้แก่ ค่าสถิติผลการทดสอบ การจัดอันดับผลการทดสอบ เป็นต้น จากนั้นจึงสร้างความตระหนักความรับผิดชอบต่อผลการทดสอบ และเปิดโอกาสให้ผู้มีส่วนได้ส่วนเสียมีส่วนร่วมต่อการพัฒนาผู้เรียนอย่างกัลยาณมิตรและให้ผู้เกี่ยวข้องได้รับประโยชน์จากผลการสอบอีกด้วย O-NET เป็นส่วนหนึ่งของการจบการศึกษาในปัจจุบัน ซึ่งเป็นนโยบายของรัฐมนตรีว่าการกระทรวงศึกษาธิการ กำหนดว่าการกระจายอำนาจต้องมีความรับผิดชอบ โดยเฉพาะด้านคุณภาพผู้เรียน ดังนั้นจึงต้องนำผลการทดสอบทางการศึกษาระดับชาติไปใช้เป็นส่วนหนึ่งของการจบการศึกษาระดับชั้น ประถมศึกษาปีที่ 6 ระดับมัธยมศึกษาปีที่ 3 และระดับมัธยมศึกษาปีที่ 6 ซึ่งเป็นไปตามนโยบายของรัฐมนตรี หลักกฎหมาย หลักวิชาการ หลักการอ้างอิงนานาชาติและหลักบูรณาการโดยมีการวางแผนในการใช้ผล O-NET ดังนี้

ตารางที่ 2-1 การปรับสัดส่วนการใช้ผลการทดสอบระดับชาติขั้นพื้นฐาน เป็นองค์ประกอบหนึ่งของการตัดสินผลการเรียนของผู้ที่จบการศึกษาตามหลักสูตรการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 (สัมพันธ์ พันธุ์พฤกษ์, 2557, หน้า 14)

ปีการศึกษา	สัดส่วนคะแนนของผลการประเมินระดับโรงเรียนต่อคะแนนของผลการประเมินระดับชาติ
2555	80:20
2556	80:20
2557	70:30
2558 เป็นต้นไป	50:50

โดยผลการทดสอบนั้นสามารถทำได้โดยครู และผู้บริหารนำผลคะแนนไปใช้ในการปรับปรุงและพัฒนาการจัดการเรียนรู้ การบริหาร การประกันคุณภาพการศึกษา (พรบ. การศึกษาแห่งชาติ มาตรา9 (3) ,47, 48 และ 49) เป็นส่วนหนึ่งของการคัดเลือกเข้าศึกษาต่อในระดับสูงขึ้น (พรบ. การศึกษาแห่งชาติ มาตรา 26 วรรค 2) เป็นองค์ประอบหนึ่งในการตัดสินผลการเรียนของผู้เรียนที่จบหลักสูตรการศึกษาขั้นพื้นบาน พ.ศ. 2551 และนำไปใช้เพื่อวัตถุประสงค์อื่นๆ เช่น การทำวิจัย ให้รางวัลเป็นต้น ครูเป็นบุคคลที่หนึ่งที่ต้องนำผลการสอบ O-NET ไปใช้เพื่อปรับปรุงและพัฒนาการจัดการเรียนรู้และประเมินสมรรถนะของครูในด้านการวัดและการประเมินผลการเรียนรู้ของผู้เรียน ต้องสอดคล้องกับการเรียนการสอนและวัตถุประสงค์ของการเรียนการสอนตามลำดับตามองค์ประกอบของกระบวนการจัดการศึกษา เพื่อนำไปสู่การพัฒนาผู้เรียนที่ยั่งยืนต่อไป

การทดสอบ O–NET นั้น นอกจากจะเป็นการประกันคุณภาพการศึกษาตามหลักสูตรแกนกลางการศึกษาระดับชาติขั้นพื้นฐาน พ.ศ. 2551 แล้ว ผลการทดสอบยังสามารถนำมาใช้ในการวางแผนการจัดการศึกษาเพื่อพัฒนาคุณภาพการศึกษาต่อไปในอนาคตด้วย ซึ่งสอดคล้องกับผลการวิจัยของ อิทธิพัทธ์ สุวทันพรกุล (2558) พบว่าโรงเรียนที่ได้รับผลการสอบ O-NET แล้วจะมีการประชุมชี้แจงให้ครูทุกกลุ่มสาระทราบพร้อมทั้งวิเคราะห์ลักษณะของข้อสอบ แนวโน้ม และทิศทางของการออกข้อสอบในแต่ละกลุ่มสาระการเรียนรู้ เพื่อนำไปพัฒนาการเรียนการสอน โดยมีการกำหนดแผนงาน โครงการ ด้านหลักสูตรและการบริหารหลักสูตร รวมทั้งการจัดเตรียมบุคลากร การจัดครูและตารางสอนอย่างเหมาะสม ซึ่งในการกำกับติดตามการจัดการเรียนการสอนนั้น พบว่า โรงเรียนได้มีการจัดนิเทศภายใน เพื่อนิเทศ กำกับ ติดตาม การใช้หลักสูตรและเป็นผู้นำในการจัดทำหลักสูตร โดยร่วมกันประสานงานกับบุคลากรทุกฝ่ายและส่งเสริมงานด้านการนิเทศภายในและส่งเสริมให้ครูมีความรู้ความเข้าใจในการจัดการเรียนการสอนอย่างแท้จริง

งานวิจัยที่เกี่ยวข้องกับการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน (O-NET) มีดังนี้

สำราญ มีแจ้ง ประภัสสร วงษ์ดี และยุพิน โกณฑา (2552) ได้ศึกษาการปรับเทียบคะแนน O-NET ระหว่างปีโดยวิธีการเทียบมาตราเดียวกันกับโดยใช้ทฤษฎีการตอบสนองข้อสอบ ระหว่างปีการศึกษา 2550 และ 2551 ระดับชั้นมัธยมศึกษาปีที่ 6 โดยวิธีการปรับเทียบคะแนนให้เป็นมาตราเดียวกันนั้นใช้วิธีการปรับเทียบคะแนนเชิงเส้นโดยการแปลงให้เป็นคะแนนมาตรฐาน Z ส่วนวิธีการปรับเทียบที่ใช้ทฤษฎีการตอบสนองข้อสอบ โดยใช้ โมเดล 3 พารามิเตอร์ ผลการปรับเทียบคะแนนพบว่า การปรับเทียบคะแนนโดยวิธีการปรับเทียบให้เป็นมาตรเดียวกัน ค่าเฉลี่ยคะแนน O-NET เดิมกับค่าเฉลี่ยที่ได้จากการปรับเทียบโดยใช้วิธีการปรับเทียบมาตราเดียวกันไม่แตกต่างกันทั้ง 2 ปีการศึกษา และผลการเปรียบเทียบคะแนน O-NET เดิมและคะแนนแปลง O-NET ที่ปรับเทียบโดยใช้วิธีทฤษฎีการตอบสนองข้อสอบ แบบ 3 พารามิเตอร์ มีค่าคะแนนจริงที่ได้จากการเทียบแตกต่างกัน โดยผู้สอบมีระดับความสามารถเดียวกันซึ่งค่าคะแนนจริง จากแบบทดสอบ O-NET ปีการศึกษา 2550 สูงกว่าปีการศึกษา 2551 ค่าคะแนน O-NET ที่ได้จากทั้ง 2 วิธีมีความสัมพันธ์กับคะแนน O-NET เดิมสูงในทางบวก ทั้ง 2 ปีการศึกษา เมื่อเปรียบเทียบคุณภาพของทั้ง 2 วิธีพบว่า ค่าสัมประสิทธิ์สหสัมพันธ์ที่ได้จากทั้ง 2 วิธี ไม่แตกต่างกัน

กนกกร ศิริสุข อิศรัฏฐ์ รินไธสง และจรัส อติวิทยาภรณ์ (2555) ได้ศึกษาพัฒนาการคุณภาพการศึกษาไทยโดยผ่านโมเดลโค้งพัฒนาการในการศึกษาการเปลี่ยนแปลงระยะยาวของผลคะแนนการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน (O-NET) ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 เครื่องมือที่ใช้ในการวิจัยคือแบบทดสอบ 5 กลุ่มสาระการเรียนรู้ ได้แก่ ภาษาไทย ภาษาอังกฤษ คณิตศาสตร์ วิทยาศาสตร์ และสังคมศึกษาศาสนาและวัฒนาธรรม จำนวน 2,775 โรงเรียน จำแนกเป็น ภูมิภาค สังกัด เขตชนบท/เขตเมืองและขนาดของโรงเรียน จากการศึกษาพบว่า โมเดลโค้งพัฒนาการผลคะแนนที่ไม่มีตัวแปรร่วมในวิชาภาษาไทย ภาษาอังกฤษ คณิตศาสตร์ วิทยาศาสตร์ มีอัตราคะแนนต่อปีมีแนวโน้มลดลง ภาษาอังกฤษลดลงมากที่สุด ส่วนสังคมศึกษา ศาสนาและวัฒนธรรม มีแนวโน้มสูงขึ้น สิ่งที่ควรพิจารณานอกจากจะปรับปรุงคุณภาพการศึกษาแล้วยังจำเป็นต้องให้ความเสมอภาคในเรื่องการศึกษาในบริบทต่างๆ ทั้งภูมิภาค สังกัด เขตชนบท/เขตเมืองและขนาดของโรงเรียน เพื่อให้เกิดการพัฒนาการศึกษาที่มาคุณภาพทั้งระบบควบคู่ไปกับการพัฒนา

ผลสัมฤทธิ์ทางการเรียนให้บรรลุตามวัตถุประสงค์ที่กำหนดไว้เพื่อการปฏิรูปการศึกษาในทศวรรษที่สอง

มนัส สายแก้ว และนิตยา ภัสสรศิริ (2555) ได้ศึกษาความสัมพันธ์ระหว่างสุขภาพองค์การกับคะแนนการทดสอบการศึกษาระดับชาติขั้นพื้นฐาน (O-NET) ของนักเรียนระดับมัธยมศึกษาในจังหวัดสุราษฎร์ธานี มีวัตถุประสงค์เพื่อเปรียบเทียบสุขภาพองค์การ และคะแนน O-NET ตามเขตพื้นที่การศึกษา และประเภทโรงเรียน และ วิเคราะห์ความสัมพันธ์ระหว่างสุขภาพองค์การกับคะแนน O-NET จำแนกตามกลุ่มสาระการเรียนรู้ ในระดับมัธยมศึกษาปีที่ 3 และมัธยมศึกษาปีที่ 6 กลุ่มตัวอย่าง ได้แก่ โรงเรียนมัธยมศึกษา 44 โรง ในจังหวัดสุราษฎร์ธานี ประกอบด้วยโรงเรียนมัธยมศึกษาประเภทโรงเรียนต้นแบบโรงเรียนในฝัน 18 โรง และโรงเรียนมัธยมศึกษาทั่วไป 26 โรง นักเรียนชั้นมัธยมศึกษาปี ที่ 3 และ มัธยมศึกษาปีที่ 6 จำนวน 12,824 คน พบว่า สุขภาพองค์การของโรงเรียนมัธยมศึกษาโดยภาพรวม และด้านต่างๆ อยู่ในระดับปานกลาง เมื่อเปรียบเทียบตามเขตพื้นที่การศึกษาและประเภทโรงเรียน พบว่าไม่แตกต่างกัน คะแนน O-NET เปรียบเทียบตามเขตพื้นที่การศึกษาพบว่า ระดับมัธยมศึกษาปีที่ 6 แตกต่างกันในวิชาคณิตศาสตร์และศิลปะ เมื่อวิเคราะห์ตามประเภทโรงเรียน พบว่า ระดับมัธยมศึกษาปีที่ 3 แตกต่างกันในวิชาภาษาไทย วิทยาศาสตร์ ศิลปะ และสุขศึกษาและพลศึกษา โดยโรงเรียนต้นแบบโรงเรียนในฝันมีคะแนนสูงกว่าหนึ่งวิชาคือ วิชาสุขศึกษาและพลศึกษา ส่วนในระดับมัธยมศึกษาปีที่ 6 แตกต่างกันในวิชาศิลปะ สุขศึกษาและพลศึกษา และสังคมศึกษา ศาสนา และวัฒนธรรม โดยโรงเรียนต้นแบบโรงเรียนในฝันมีคะแนนสูงกว่าหนึ่งวิชา คือ วิชาศิลปะ และความสัมพันธ์ระหว่างสุขภาพองค์การกับคะแนน O-NET พบว่า ระดับมัธยมศึกษาปีที่ 3 มิติความเข้มแข็งขององค์การสัมพันธ์กับคะแนนรวมและสัมพันธ์กับวิชาสังคมศึกษา ศาสนาและวัฒนธรรม สุขศึกษาและพลศึกษา และคณิตศาสตร์ สำหรับระดับมัธยมศึกษาปีที่ 6 พบว่า มิติความเข้มแข็งขององค์การสัมพันธ์กับวิชาภาษาไทย และศิลปะ มิติอิทธิพลของผู้บริหารโรงเรียน สัมพันธ์กับวิชาสังคมศึกษา ศาสนาและวัฒนธรรม และสุขศึกษาและ พลศึกษา ส่วนมิติการบริหารแบบมิตรสัมพันธ์ สัมพันธ์ กับวิชาภาษาไทย และศิลปะ

นุภาพรรณ ปลื้มใจ ปิยะทิพย์ ตินวร และโสพส สุขานนท์สวัสดิ์ (2558) ได้ศึกษาการพัฒนาโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์สำหรับการจัดสอบ O-NET ระดับชั้นมัธยมศึกษาปีที่ 6 จำนวน 61 คน จาก 2 โรงเรียน จำนวน 8 กลุ่มสาระการเรียนรู้ จากผลการศึกษาพบว่า จากการวิเคราะห์คุณภาพข้อสอบด้วยทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT) แบบ 3 พารามิเตอร์ ในคลังข้อสอบ O-NET มีความยากของข้อสอบ (b) ในระดับค่อนข้างยาก ข้อสอบที่ผ่านจำนวน 1,197 ข้อ คลังข้อสอบ O-NET สามารถบรรจุข้อสอบแบบหลายตัวเลือก (Multiple Choice) ชนิด 4 ตัวเลือก ได้โดยไม่จากัด โปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ มีความเหมาะสมของโปรแกรมอยู่ในระดับมากที่สุด เป็นที่ยอมรับของผู้เชี่ยวชาญและนักเรียนกลุ่มตัวอย่างที่ทดลองใช้โปรแกรม มีความคิดเห็นอยู่ในเกณฑ์ดี

จารุจิตร สิทธิปรุ ปิยะทิพย์ ตินวร และโสพส สุขานนท์สวัสดิ์ (2559) ศึกษาการพัฒนาโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ สำหรับการจัดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 3 เพื่อวิเคราะห์คุณภาพข้อสอบ O-NET จัดทำคลังข้อสอบ O-NET และพัฒนาโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ สำหรับการจัดสอบ O-NET จำนวน 8 กลุ่มสาระการเรียนรู้

ระหว่างปี พ.ศ. 2551–2553 ในรูปแบบของ Web Application วิเคราะห์คุณภาพข้อสอบโดยใช้ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT) แบบ 3 พารามิเตอร์ โดยใช้โปรแกรมสำเร็จรูป Xcalibre Version 4.1.7 กลุ่มตัวอย่างเป็นนักเรียนระดับชั้นมัธยมศึกษาปีที่ 3 โรงเรียนอ่างศิลาพิทยาคม จังหวัดชลบุรี จำนวน 30 คน พบว่า การวิเคราะห์คุณภาพข้อสอบ O-NET 8 กลุ่มสาระการเรียนรู้ แสดงให้เห็นว่า ข้อสอบ O-NETระดับชั้นมัธยมศึกษาปีที่ 3 มีความยากของข้อสอบ (b) เฉลี่ยค่อนข้างยาก คลังข้อสอบ O-NET สามารถบรรจุข้อสอบแบบหลายตัวเลือก (Multiple Choice) ได้ไม่จำกัด ทั้งนี้ขึ้นอยู่กับขนาดของ Server และโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ สำหรับการจัดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 3 มีประสิทธิภาพอยู่ในเกณฑ์ดี เป็นที่ยอมรับของผู้เชี่ยวชาญและนักเรียนกลุ่มตัวอย่างที่ทดลองใช้โปรแกรม ทั้งในด้านลักษณะทั่วไปของโปรแกรม และด้านความสะดวกในการใช้โปรแกรม

วัชรินทร์ อิ่มอาเทศ และภัทราวุฒิ แสงศิริ (2559) ได้ศึกษาการพยากรณ์ค่าคะแนนสอบ O-NET ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 โดยใช้เทคนิคเหมืองข้อมูลมีวัตถุประสงค์เพื่อคัดเลือกรายวิชาที่ส่งผลต่อคะแนน O-NET และนำมาสร้างโมเดลการพยากรณ์คะแนนสอบ พบว่าคะแนนสอบ O-NET ของนักเรียนส่วนใหญ่มีค่าต่ำกว่าเกณฑ์ โดยเฉพาะนักเรียนชั้นมัธยมศึกษาปีที่ 6 นักเรียนสามารถทราบว่ารายวิชาใดมีผลอย่างยิ่งต่อคะแนน O-NET และทราบถึงแนวโน้มผลคะแนนสอบในอนาคตล่วงหน้า ก็จะก่อให้เกิดความกระตือรือร้นในการเตรียมความพร้อมการทบทวนรายวิชานั้น ทั้งยังสามารถพัฒนาคะแนนสอบให้ดียิ่งขึ้น

สุชาดา กรเพชรปาณี ปิยะทิพย์ ตินวร และโสพส สุขานนท์สวัสดิ์ (2559) ได้ศึกษาการพัฒนาโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์สาหรับการจัดสอบ O-NET ระดับชั้นประถมศึกษาปีที่ 6 ชั้นมัธยมศึกษาปีที่ 3 และชั้นมัธยมศึกษาปีที่ 6 ระดับชั้นละ 8 กลุ่มสาระการเรียนรู้ การจัดทำคลังข้อสอบใช้ฐานข้อมูล MySQL และคัดเลือกข้อสอบ O-NETของสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) ระหว่างปี พ.ศ. 2551-2553 ที่ผ่านการวิเคราะห์ข้อสอบตามทฤษฎีการตอบสนองข้อสอบ แบบ 3 พารามิเตอร์ ในระดับชั้นประถมศึกษาปีที่ 6 จำนวน 258 ข้อชั้นมัธยมศึกษาปีที่ 3 จำนวน 469 ข้อ และชั้นมัธยมศึกษาปีที่ 6 จำนวน 1,197 ข้อ โดยข้อสอบในคลังข้อสอบ O-NET อยู่ในระดับค่อนข้างยาก พัฒนาโปรแกรมในรูปแบบของ Web Application โดยใช้ภาษา PHP พัฒนาโปรแกรมผู้ใช้สามารถเข้าใช้โปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ สำหรับการจัดสอบ O-NET ใน website://www.onetcat.net พบว่าการใช้งานของโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ สำหรับการจัดสอบ O-NET อยู่ในเกณฑ์ดี เป็นที่พึงพอใจของนักเรียน

Na Wichian Wongwanich and Saengsiri (2014) ได้ทำการศึกษาการจัดทำเหมืองข้อมูลเพื่อพยากรณ์คะแนนเฉลี่ยของสถานศึกษาจากผลการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน (O-NET) ช่วงชั้นที่ 2 ระดับชั้นประถมศึกษาปีที่ 6 กลุ่มตัวอย่างเป็นข้อมูลคะแนนเฉลี่ยของสถานศึกษาในเขตกรุงเทพมหานคร จำนวน 859 คน พบว่า การวิเคราะห์การจัดกลุ่มข้อมูลการเรียนรู้แบบเพิ่มขึ้น (ILM) มีประสิทธิภาพในการการเป็นโมเดลพยากรณ์ได้ดีกว่าวิธีอื่น ๆ โดยมีค่าเฉลี่ยของความถูกต้อง (Accuracy Rate) เท่ากับร้อยละ 92.19 ขณะที่วิธีการพยากรณ์วิธีอื่น ๆมีค่าความถูกต้องน้อยกว่าร้อยละ 70 สำหรับประสิทธิภาพในการจัดกลุ่มตามเกณฑ์ที่กำหนดไว้เบื้องต้น

จำนวน 3 กลุ่มพบว่าวิธี ILM และ K-Means เท่านั้น ที่สามารถจัดกลุ่มข้อมูลออกเป็น 3 กลุ่ม ขณะที่วิธี COBWEB จัดได้มากกว่า 400 กลุ่มและมีค่าความถูกต้องน้อยกว่าร้อยละ 7 แสดงให้เห็นว่าการประเมินความสามารถของระบบการพยากรณ์น่าพอใจ การทดสอบความต้องการใช้งานและการทดสอบความสามารถในการใช้งานพบว่าประสิทธิภาพของระบบดีขึ้น

จากการศึกษางานวิจัยที่เกี่ยวข้องกับการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน (O-NET) ยังเป็นปัจจัยสำคัญที่มีผลกระทบต่อการจัดการศึกษาในประเทศไทยเป็นอย่างมาก ทั้งด้านตัวผู้ตอบแบบทดสอบ และปัจจัยแวดล้อมที่มีผลกระทบ ทำให้ข้อสอบหลายกลุ่มสาระการเรียนรู้มีค่าความยากอยู่ในระดับยาก การพยากรณ์ค่าคะแนนทำให้ผู้เรียนมีความกระตือรือร้นในการเตรียมความพร้อมการทบทวนรายวิชานั้น ทั้งยังสามารถพัฒนา คะแนนสอบให้ดียิ่งขึ้น

ตอนที่ 2 ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT) และงานวิจัยที่เกี่ยวข้อง

ทฤษฎีการทดสอบเป็นองค์ความรู้ที่สำคัญในการสร้างและพัฒนาแบบทดสอบให้มีคุณภาพและแปลความหมายในการวัดผลได้อย่างถูกต้อง ซึ่งได้มีการจัดประเภทของทฤษฎีออกเป็น 2 แนวทาง ได้แก่ ทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory: CTT) และทฤษฎีการทดสอบแนวใหม่ (Moderm Test Theory: MTT) โดยในทฤษฎีการทดสอบแนวใหม่ได้แบ่งการพัฒนาการที่สำคัญเป็น 2 แนวทาง คือ ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของการวัดผล (Generalizability Theory: G Theory) และทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT) (ศิริชัย กาญจนวาสี, 2556) ซึ่งในการวิจัยครั้งนี้ ขอนำเสนอเฉพาะในส่วนของทฤษฎีการตอบสนองข้อสอบ (IRT) โดยมีรายละเอียดดังนี้ ทฤษฎีการตอบสนองข้อสอบ (IRT) เป็นทฤษฎีการวัดที่อธิบายถึงความสัมพันธ์ระหว่างคุณลักษณะภายในหรือความสามารถที่มีอยู่ภายในตัวบุคคลกับพฤติกรรมการตอบสนองข้อสอบของบุคคลนั้น โดยมีพื้นฐานความเชื่อที่ว่าพฤติกรรมการตอบสนองของผู้สอบเป็นสิ่งที่เห็นได้โดยตรงว่าถูกหรือผิด จะถูกกำหนดโดยคุณลักษณะภายใน (Trait) หรือความสามารถ (Ability) ที่อยู่ภายในตัวบุคคลซึ่งเป็นสิ่งที่ไม่สามารถสังเกตได้โดยตรง (Hambleton, Swaminathan, & Rogers, 1991) และสามารถอธิบายความสัมพันธ์ในรูปแบบของฟังก์ชั่นคณิตศาสตร์หรือโมเดลที่แสดงความสัมพันธ์ระหว่างระดับความสามารถ คุณลักษณะของข้อสอบและโอกาสของการตอบข้อสอบได้ถูก เรียกว่า ฟังก์ชั่นการตอบสนองข้อสอบ (Item Response Function: IRF) สามารถนำมาใช้ในการศึกษาความสัมพันธ์ระหว่างความน่าจะเป็นในการตอบข้อสอบแต่ละข้อได้ถูกต้อง $[\mathrm{P}i(\theta)]$ กับระดับความสามารถของผู้สอบที่วัดได้โดยแบบทดสอบฉบับนั้น (θ) เมื่อนำมาเขียนเป็นกราฟได้ “ โค้งลักษณะข้อสอบ” (Item Characteristic Curve: ICC) มีได้หลายลักษณะขึ้นอยู่กับโมเดล ซึ่งสามารถจำแนกเป็น 3 ประเภทใหญ่ ๆ คือ โมเดลการตอบสนองข้อสอบที่ใช้กับคะแนนรายข้อแบบ 2 ค่า (Dichotomous) เป็นข้อสอบที่ตรวจให้คะแนน 0,1 (ตอบผิดได้ 0 คะแนน, ตอบถูกได้ 1 คะแนน) ส่วนที่สองคือ โมเดลการตอบสนองข้อสอบที่ใช้กับคะแนนรายข้อแบบมากว่า 2 ค่า (Polytomous) เป็นข้อสอบที่ตรวจให้คะแนน 0, 1 (ตอบผิดได้ 0 คะแนน, ตอบถูกได้ 1 คะแนน) และ โมเดลการตอบสนองข้อสอบที่ใช้กับคะแนนรายข้อ

แบบต่อเนื่อง (Continuous) ในงานวิจัยนี้ ขอกล่าวถึงเฉพาะในส่วนของทฤษฎีการตอบสนองข้อสอบแบบตรวจให้คะแนน 2 ค่า (Dichotomous IRT) เท่านั้น

1. ข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบ

1.1 ความเป็นเอกมิติ (Unidimensionality) คือการมุ่งวัดเพียงคุณลักษณะหรือความสามารถเดียวโดยใช้เทคนิคการวิเคราะห์ทางสถิติ

1.2 ความเป็นอิสระ (Local Independence) คือ ความเป็นอิสระระหว่างข้อสอบและผู้สอบ โดยพิจารณาเมตริกซ์ ความแปรปรวนและความแปรปรวนร่วม หรือเมตริกซ์สหสัมพันธ์ของคะแนนคำตอบรายข้อ

1.3 โมเดลการตอบสนองข้อสอบ (Item Response Model) คือ ฟังก์ชั่นหรือโค้งลักษณะของข้อสอบ (ICC) ที่สะท้อนให้เห็นถึงความสัมพันธ์จริงระหว่างความสามารถของผู้สอบกับลักษณะของข้อสอบซึ่งโมเดลที่ใช้กันแพร่หลายในส่วนของโมเดลการตอบสนองข้อสอบที่ใช้กับคะแนนรายข้อแบบ 2 ค่า (Dichotomous Model) ได้แก่ โมเดลการตอบสนองข้อสอบแบบ 1 พารามิเตอร์ ซึ่งมีข้อตกลงเบื้องต้นว่า ข้อสอบแต่ละข้อมีค่าพารามิเตอร์ c เท่ากับ 0 และค่าพารามิเตอร์ a เท่ากัน แต่มีความแตกต่างเฉพาะพารามิเตอร์ b เท่านั้น โมเดลนี้จึงเหมาะสำหรับใช้กับข้อสอบอิงเกณฑ์ที่ไม่สลับซับซ้อน ข้อสอบที่ค่อนข้างเรียบง่ายสำหรับพัฒนาคลังข้อสอบที่มีความเป็นเอกพันธ์ ประเภทที่ 2 คือ โมเดลการตอบสนองข้อสอบแบบ 2 พารามิเตอร์ มีข้อตกลงเบื้องต้นว่าข้อสอบแต่ละข้อมีค่าพารามิเตอร์ c เท่ากับ 0 และมีความแตกต่างกันของพารามิเตอร์ a และ b โมเดลนี้เหมาะสำหรับใช้กับข้อสอบที่ต้องเติมคำตอบหรือข้อสอบแบบเลือกตอบที่ไม่ยากมากนักและกลุ่มผู้สอบมีความพร้อมในการตอบ และโมเดลประเภทที่ 3 คือ โมเดลการตอบสนองข้อสอบแบบ 3 พารามิเตอร์ มีข้อตกลงเบื้องต้นว่า ข้อสอบแต่ละข้อมีความแตกต่างกันได้ทั้ง พารามิเตอร์ a, b และ c โมเดลนี้เหมาะสำหรับใช้กับข้อสอบแบบเลือกตอบทั่วไป ข้อสอบแบบหลายตัวเลือก เนื่องจากผู้ตอบสามารถเดาคำตอบได้

1.4 การสอบที่ไม่แข่งขันด้านเวลา (Nonspeeded Test Administration) กล่าวคือความเร็วต่อการตอบจะต้องไม่มีอิทธิพลต่อผลการตอบ ในการตัดสอบนั้นผู้สอบต้องอยู่ในสถานการณ์ที่มีเวลาเพียงพอต่อการทำข้อสอบ

2. โมเดลการวัดภายใต้แนวคิดพื้นฐานของทฤษฎีการตอบสนองข้อสอบ (IRT)

โมเดลการตอบสนองข้อสอบประกอบด้วยค่าพารามิเตอร์ของข้อสอบ (Item parameter) และค่าคงที่ แต่ละตัวนั้นมีความหมาย (ศิริชัย กาญจนวาสี, 2555, หน้า 54-55) ดังนี้

2.1 ค่าพารามิเตอร์ของผู้สอบ สัญลักษณ์ คือ θ หมายถึง ความสามารถของผู้สอบ นิยมปรับให้เป็นคะแนนมาตรฐานที่มีค่าเฉลี่ยเป็น 0 และส่วนเบี่ยงเบนมาตรฐานเป็น 1 ซึ่งค่าความสามารถของผู้สอบ (θ) มีพิสัยอยู่ระหว่าง $-\infty$ และ ∞ ผลการวิเคราะห์ส่วนใหญ่มักจะมีค่าอยู่ระหว่าง -3 ถึง +3

2.2 ค่าพารามิเตอร์ของข้อสอบ แบ่งเป็น 3 พารามิเตอร์ ได้แก่

2.2.1 ค่าความยาก (Difficulty Parameter) สัญลักษณ์คือ b หมายถึง สัดส่วนของคนที่ทำข้อสอบข้อนั้นถูกหรือค่าที่แสดงถึงระดับความสามารถของผู้สอบ (θ) ที่จุดโค้งลักษณะข้อสอบ

ที่มีความชันมากที่สุด มีค่าตั้งแต่ -∞ และ ∞ แต่ในทางปฏิบัติมักจะมีค่าอยู่ระหว่าง -2.50 ถึง 2.50 ค่า -2.50 แสดงว่าข้อสอบนั้นง่ายมากและค่า 2.50 แสดงว่าข้อสอบนั้นยากมาก

2.2.2 ค่าอำนาจจำแนก (Discrimination Parameter) สัญลักษณ์คือ a หมายถึง ความสามารถของข้อสอบที่แยกเด็กออกเป็น 2 พวก คือพวกตอบถูกกับพวกตอบผิด ค่าที่เป็นสัดส่วนโดยตรงกับความชันของโค้งคุณลักษณะของข้อสอบ ณ จุดเปลี่ยนโค้ง มีตั้งแต่ -∞ และ ∞ แต่ในทางปฏิบัติจะมีค่าเป็นบวก มีค่าตั้งแต่ 0.5 -2.5 เพราะค่าอำนาจจำแนกที่เป็นลบแสดงว่าข่อสอบข้อนั้นไม่ดีต้องตัดทิ้ง ค่าต่ำกว่า 0.5 แสดงว่าข้อสอบไม่มีอำนาจจำแนก ค่า 2.5 แสดงว่าข้อสอบมีค่าอำนาจจำแนกสูงในการคัดเลือกข้อสอบ ข้อสอบที่คัดไว้จะมีค่าอำนาจจำแนกตั้งแต่ 0.5 ขึ้นไป

2.2.3 ค่าโอกาสการเดา (Guessing Parameter) สัญลักษณ์คือ c หมายถึง ความน่าจะเป็นที่บุคคลหนึ่งที่มีความสามารถต่ำสุดจะตอบข้อสอบนั้นได้ถูกต้อง เป็นค่าที่แสดงถึงโอกาสการตอบข้อสอบถูก โดยไม่มีความรู้ในเรื่องนั้นๆ มีค่าจาก 0 ถึง 1 จะคัดเลือกเอาข้อสอบที่มีค่าโอกาสการเดา ต่ำกว่า 0.3

2.3 ค่าคงที่ แบ่งออกเป็น 2 ประเภทคือ ค่าคงที่ของลอกาลิทึมธรรมชาติ (Natural log) มีค่า e เท่ากับ 2.71828 และ ค่าองค์ประกอบของการปรับสเกล (Scaling Factor) มีค่า D เท่ากับ 1.70

ในโมเดลการตอบสนองข้อสอบแบบ 1, 2 และ 3 พารามิเตอร์นั้นมีโค้งลักษณะข้อสอบที่เขียนด้วยฟังก์ชั่นโลจิสติกดังนี้

2.3.1 โมเดลการตอบสนองข้อสอบแบบ 1พารามิเตอร์ (One – Parameter Model) เป็นโมเดลที่อธิบายผลการวิเคราะห์ข้อสอบด้วยค่าความยากของข้อสอบ (b) เรียกอีกอย่างหนึ่งว่า “Rasch Model” มีโค้งลักษณะข้อสอบที่สำคัญคือ ค่าความยากของข้อสอบ (b) มีค่าแปรเปลี่ยนตามลักษณะของข้อสอบแต่ละข้อ ส่วนค่าอำนาจจำแนกของข้อสอบ (a) มีค่าคงที่ และค่าการเดาของข้อสอบ (c) มีค่าเป็น 0 ซึ่งเขียนด้วยฟังก์ชั่นโลจิสติกดังสมการ

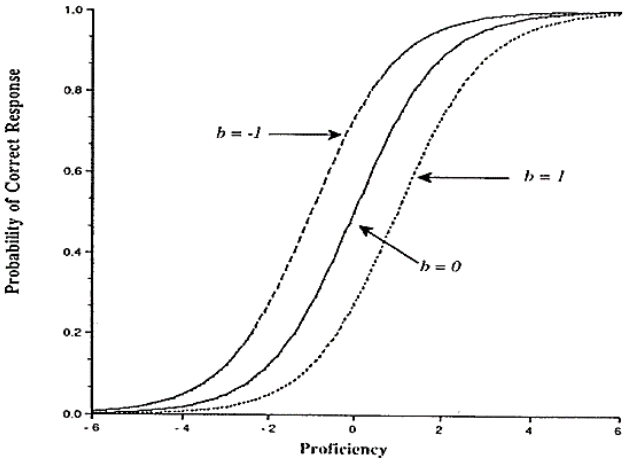
$$Pi(\theta) = \frac{1}{1+e^{-(\theta-bi)}} \quad (1)$$

เมื่อ

$Pi(\theta)$ คือ ความน่าจะเป็นของผู้เข้าสอบที่มีความสามารถ จะตอบข้อสอบข้อที่ i ได้ถูกต้อง

bi คือ ค่าพารามิเตอร์ความยากของข้อสอบข้อที่ i

e คือ 2.718



ภาพที่ 2-3 โค้งคุณลักษณะของข้อสอบแบบ 1 พารามิเตอร์ (Wainer, 2000, p. 68)

2.3.2 โมเดลการตอบสนองข้อสอบแบบ 2 พารามิเตอร์ (Two – Parameter Model) เป็นโมเดลการวิเคราะห์ข้อสอบที่ใช้ค่าพารามิเตอร์แบบ 2 พารามิเตอร์ คือ ค่าความยากของข้อสอบ (b) และค่าอำนาจจำแนกของข้อสอบ (a) มีโค้งลักษณะข้อสอบที่สำคัญคือ ค่าความยากของข้อสอบ (b)และค่าอำนาจจำแนกของข้อสอบ (a) มีค่าแปรเปลี่ยนตามลักษณะของข้อสอบแต่ละข้อ และค่าการเดาของข้อสอบ (c) มีค่าเป็น 0 ซึ่งเขียนด้วยฟังก์ชั่นโลจิสติกดังสมการ

$$Pi(\theta) = \frac{1}{1+e^{-Dai(\theta-bi)}} \quad (2)$$

เมื่อ

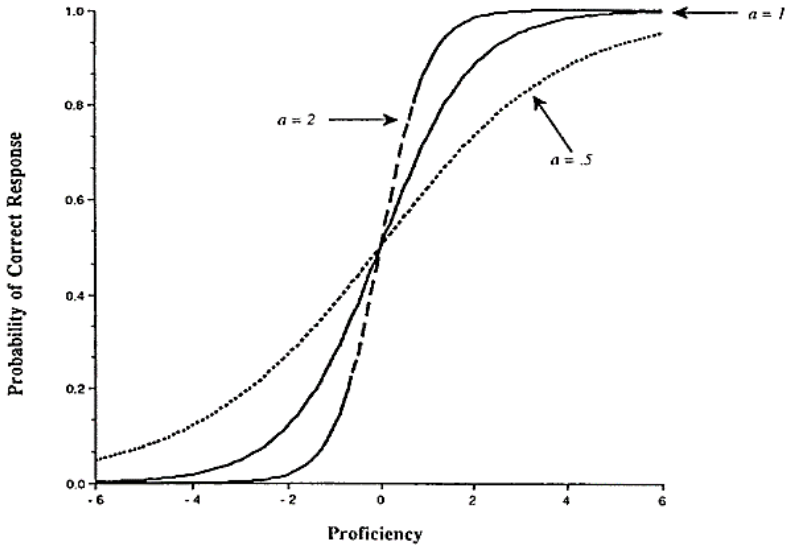
$Pi(\theta)$ คือ ความน่าจะเป็นของผู้เข้าสอบที่มีความสามารถ จะตอบข้อสอบข้อที่ I ได้ถูกต้อง

bi คือ ค่าพารามิเตอร์ความยากของข้อสอบข้อที่ i

ai คือ ค่าพารามิเตอร์อำนาจจำแนกของข้อสอบข้อที่ i

D คือ 1.70

e คือ 2.718



ภาพที่ 2-4 โค้งคุณลักษณะของข้อสอบแบบ 2 พารามิเตอร์ (Wainer, 2000, p. 70)

2.3.3 โมเดลการตอบสนองข้อสอบแบบ 3 พารามิเตอร์ (Three – Parameter Model) เป็นโมเดลพัฒนามาจากโมเดลการวิเคราะห์ข้อสอบที่ใช้ค่าพารามิเตอร์ 3 พารามิเตอร์ คือ ค่าความยากของข้อสอบ (b) ค่าอำนาจจำแนกของข้อสอบ (a) และค่าการเดาของข้อสอบ (c) มีโค้งลักษณะข้อสอบที่สำคัญคือ ค่าความยากของข้อสอบ (b) ค่าอำนาจจำแนกของข้อสอบ (a) และค่าการเดาของข้อสอบ (c) มีค่าแปรเปลี่ยนตามลักษณะของข้อสอบแต่ละข้อ ซึ่งเขียนด้วยฟังก์ชั่นโลจิสติกดังสมการ

$$Pi(\theta) = ci + \frac{(1-ci)}{1+e^{-Dai(\theta-bi)}} \quad (3)$$

เมื่อ

$Pi(\theta)$ คือ ความน่าจะเป็นของผู้เข้าสอบที่มีความสามารถ จะตอบข้อสอบข้อที่ i ได้ถูกต้อง

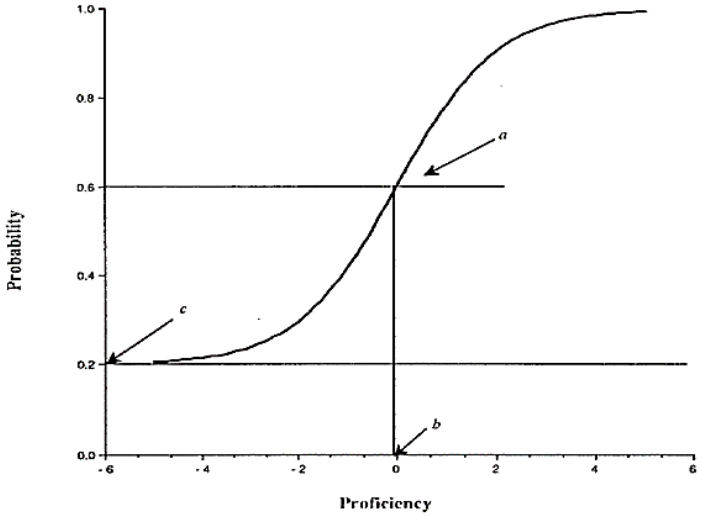
bi คือ ค่าพารามิเตอร์ความยากของข้อสอบข้อที่ i

ai คือ ค่าพารามิเตอร์อำนาจจำแนกของข้อสอบข้อที่ i

ci คือ ค่าพารามิเตอร์การเดาของข้อสอบข้อที่ i

D คือ 1.70

e คือ 2.718



ภาพที่ 2-5 โค้งคุณลักษณะของข้อสอบแบบ 3 พารามิเตอร์ (Wainer, 2000, p. 71)

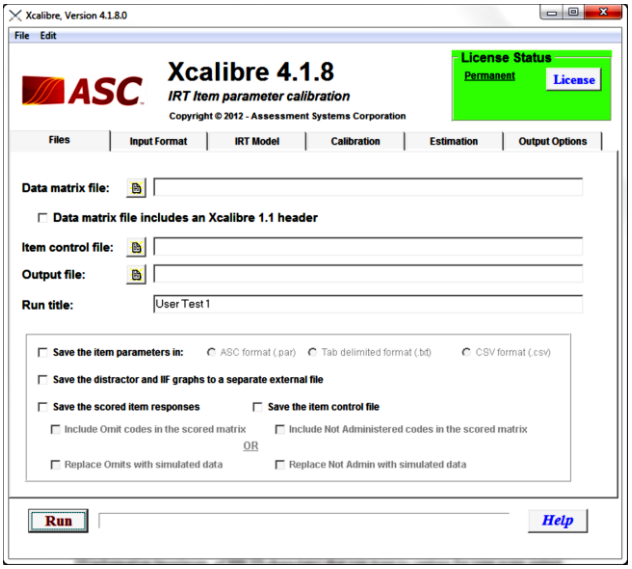
ทฤษฎีการตอบสนองข้อสอบนั้นถือเป็นทฤษฎีที่ทำให้การวัดผลแบบทดสอบมีคุณภาพน่าเชื่อถือ เมื่อเปรียบเทียบกับวิธีอื่นซึ่งสอดคล้องกับงานวิจัยของ Elosua and Wells (2013) พบว่า พบว่าอัตราความคลาดเคลื่อนสูงสำหรับวิธีทฤษฎีการตอบสนองข้อสอบเมื่อเทียบกับวิธีการถดถอยโลจิสติก และวิธีโครงสร้างความแปรปรวน (MACS) เมื่อบางส่วนของแบบทดสอบที่มีอยู่ข้อสอบที่ทำหน้าที่ต่างกัน ทั้ง 3 วิธีแสดงอัตราความถูกต้องคล้ายกันซึ่งวิธีการถดถอยโลจิสติกแสดงอัตราความถูกต้องสูงขึ้นเพียงเล็กน้อยเมื่อเทียบกับอีกสองวิธี สำหรับกลุ่มตัวอย่างที่มีขนาดเล็ก การทำหน้าที่ต่างกันของข้อสอบจะมีความไม่สม่ำเสมอ วิธีทฤษฎีการตอบสนองข้อสอบจะแสดงอัตราความถูกต้องได้ดีที่สุดเมื่อเทียบกับวิธีโครงสร้างความแปรปรวน (MACS) และวิธีถดถอยโลจิสติก เมื่อพูดถึงการศึกษาภายใต้ขนาดของกลุ่มตัวอย่างก็พบว่าทฤษฎีการตอบสนองข้อสอบนั้น ตามการทดสอบประเภทอัตราอัตราความคลาดเคลื่อน สูงขึ้นสำหรับ IRT และการถดถอยโลจิสติกตามลำดับเมื่อบางส่วนของข้อสอบที่มีอยู่ในการทำหน้าที่ต่างกันของข้อสอบ สำหรับข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูปนั้น MACS และ IRT แสดงอัตราความถูกต้องที่คล้ายกัน แต่วิธีถดถอยโลจิสติแสดงอัตราความถูกต้องที่สูงขึ้นเล็กน้อย เมื่อเทียบกับอีกสองวิธีสำหรับขนาดของกลุ่มตัวอย่างที่มีขนาดเล็ก ข้อสอบที่มีการทำหน้าที่ต่างกันแบบอเนกรูป IRT แสดงอัตราความถูกต้องมากขึ้นเมื่อเทียบกับ MACS และถดถอยโลจิสติก (Tihomir & Bengt, 2012) ซึ่งสอดคล้องกับการศึกษาของ จิระนาฎ ฉวีพัฒน์ และคณะ (2554) พบว่า ความคลาดเคลื่อนของวิธีทฤษฎีการตอบสนองข้อสอบแบบสามพารามิเตอร์มีค่าน้อยกว่าวิธีเชิงเส้น ส่วนการทดสอบความพอเพียงปรากฏว่าวิธีทฤษฎีการตอบสนองข้อสอบแบบสามพารามิเตอร์มีความพอเพียงในระดับความน่าพอใจกว่าวิธีการปรับเทียบเชิงเส้นตรง ซึ่งทำให้ทราบว่าวิธีทฤษฎีการตอบสนองข้อสอบแบบสามพารามิเตอร์มีคุณภาพกว่าวิธีเชิงเส้นตรง การเปรียบเทียบค่าพารามิเตอร์ของข้อสอบและผู้สอบตามทฤษฎีการตอบสนองข้อสอบ เมื่อกลุ่มตัวอย่างมีขนาดต่างกันทำให้ มีค่าอำนาจจำแนก ค่าความยาก ค่าการเดาข้อสอบและค่าความสามารถของผู้สอบ

แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ สรุปได้ว่า ทฤษฎีการตอบสนองข้อสอบมีคุณภาพและน่าเชื่อถือเมื่อนำไปใช้ในการตรวจสอบคุณภาพของแบบทดสอบ

3. การวิเคราะห์คุณภาพของข้อสอบ โดยใช้โปรแกรม Xcalibre

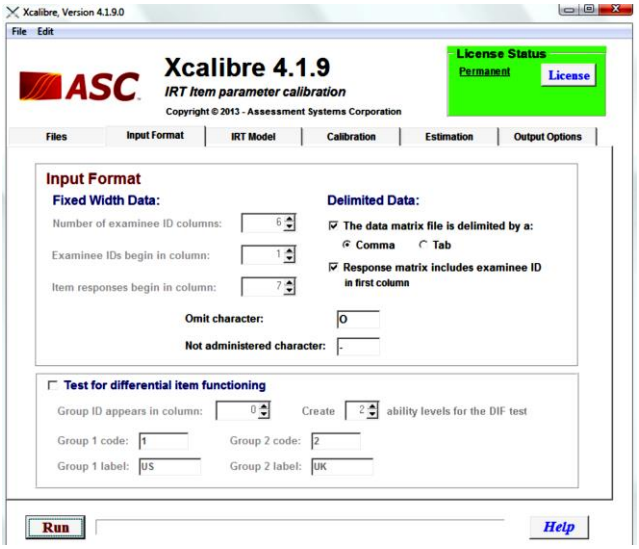
Xcalibre เป็นโปรแกรมที่ออกแบบมาเพื่อดำเนินการประมาณตามทฤษฎีการตอบสนองข้อสอบ (IRT) พารามิเตอร์ที่มีการใช้งานง่าย มีวัตถุประสงค์เพื่อช่วยให้โปรแกรมการวิเคราะห์คุณภาพข้อสอบซึ่งตรวจสอบลักษณะทางจิตวิทยาของผู้สอบ โดยใช้ทฤษฎีการตอบสนองข้อสอบ แบบ 3 พารามิเตอร์ ที่มีความจำเป็นสำหรับคะแนนสอบกับ ทฤษฎีการตอบสนองข้อสอบ โดยเฉพาะอย่างยิ่งกับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ (CAT) โปรแกรมในเวอร์ชั่นนี้มีข้อได้เปรียบหลายอย่างเมื่อเปรียบเทียบกับเวอร์ชั่นเดิม (Xcalibre 1.1) คือ การเพิ่มเติมของกราฟิก สามารถเทียบพารามิเตอร์ตามทฤษฎีการตอบสนองข้อสอบได้ (Guyer & Thompson, 2014) หน้าจอมีสถิติแบบดั้งเดิมที่ยอมรับก่อนที่จะดำเนินการเปรียบเทียบกับทฤษฎีการตอบสนองข้อสอบโดยการประมาณค่าพารามิเตอร์ที่สามารถระบุได้ นอกจากนี้ยังสามารถปรับแต่งแต่ละส่วนได้ การแทนที่จะไฟล์ค่อนข้างง่าย เป็นไปตามรูปแบบที่เตรียมไว้เพื่อรายงานอย่างเป็นทางการและยังอยู่ในค่าที่สามารถที่จัดการได้อย่างสะดวก คะแนนที่ได้สามารถประมาณค่าตามทฤษฎีการตอบสนองข้อสอบได้หลากหลายโดเมนเช่นเดียวกับการทดสอบแบบเต็มรูป คะแนนสามารถแบ่งได้เป็นสองกลุ่ม สามารถระบุจุดตัดของคะแนนซึ่งผู้ใช้สามารถกำหนดได้ พร้อมทั้งสามารถวิเคราะห์คะแนนสูงสุดของแบบทดสอบ ได้ถึง 1500 คะแนน และทั้งยังสามารถอนุญาตให้เรียกใช้ข้อมูลหลายชุดโดยไม่ต้องใช้อินเตอร์เฟซ สำหรับไฟล์ที่สร้างขึ้นนอกโปรแกรมสามารแก้ไขข้อความหรือการโต้ตอบภายใน Xcalibre ซึ่ง Xcalibre มีอินเตอร์เฟซกับผู้ใช้แบบกราฟิกทำให้มันง่ายต่อการเรียกใช้โปรแกรมแม้ว่าคุณไม่ได้เป็นผู้เชี่ยวชาญในทฤษฎีการตอบสนองข้อสอบ หน้าจอจัดเป็น 6 แท็บ คือ

3.1 แท็บระบุไฟล์ข้อมูลเมทริกซ์ที่จะใช้การควบคุมและการส่งออก ไฟล์ที่ส่งออกใด ๆ ที่ไม่จำเป็นจะถูกระบุไว้ ดังภาพที่ 2-6



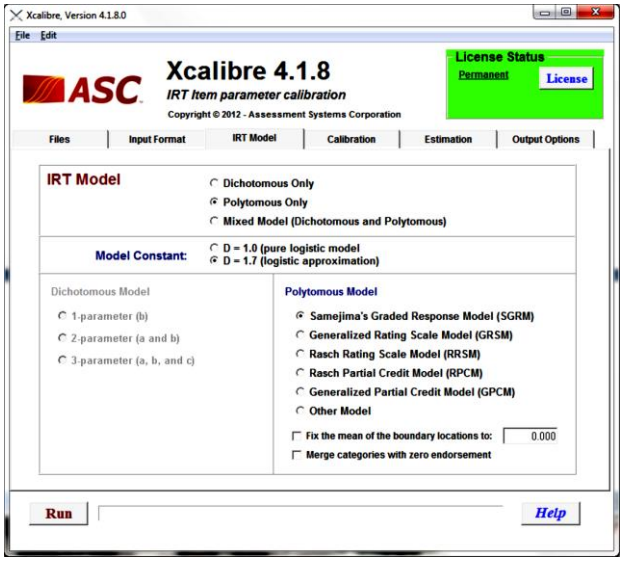
ภาพที่ 2-6 แท็บระบุไฟล์ข้อมูลเมทริกซ์ (Guyer & Thompson, 2014, p. 11)

3.2 แท็บสำหรับป้อนข้อมูลเพื่อระบุคอลัมน์ของเมทริกซ์ ไฟล์ข้อมูลสำหรับลงรหัสและการตอบสนองข้อสอบทั้งยังสามารถระบุรหัสตัวอักษรที่ใช้ในข้อมูลเมทริกซ์ นอกจากนี้ยังสามารถขอและตั้งค่าการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบได้ ดังภาพที่ 2-7



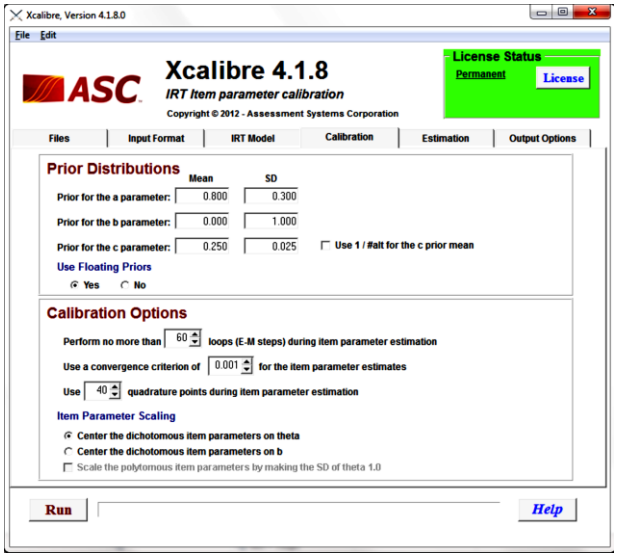
ภาพที่ 2-7 แท็บสำหรับป้อนข้อมูลเพื่อระบุคอลัมน์ของเมทริกซ์ (Guyer & Thompson, 2014, p. 12)

3.3 แท็บ IRT โมเดล ระบุโมเดลทฤษฎีการตอบสนองข้อสอบ ใช้สำหรับการประมาณค่าพารามิเตอร์ของข้อสอบ ดังภาพที่ 2-8



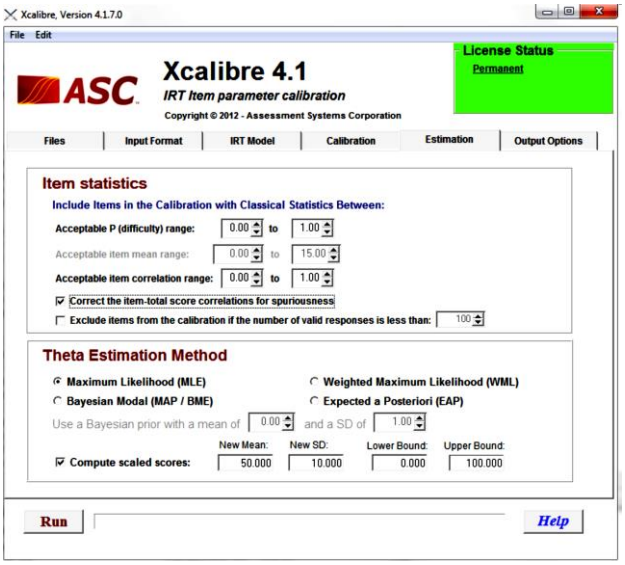
ภาพที่ 2-8 แท็บ IRT โมเดล (Guyer & Thompson, 2014, p. 14)

3.4 แท็บการเปรียบเทียบที่สามารถระบุวิธีการก่อนและหลังกระบวนการประมาณค่าพารามิเตอร์ นอกจากนี้ยังสามารถระบุตัวเลือกที่จะใช้ในระหว่างการเปรียบเทียบข้อสอบในทฤษฎีการตอบสนองข้อสอบ แบบ 3 พารามิเตอร์ ดังภาพที่ 2-9



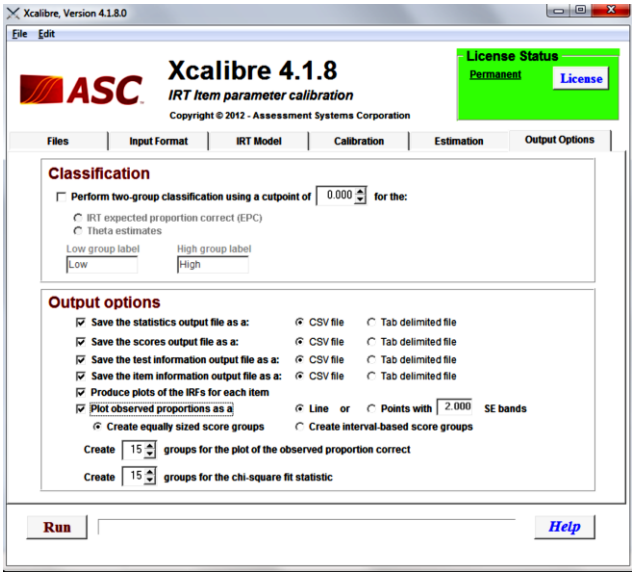
ภาพที่ 2-9 แท็บการเปรียบเทียบ (Guyer & Thompson, 2014, p. 16)

3.5 แท็บการประเมินความยากของข้อสอบที่ได้รับการยอมรับและมีความสัมพันธ์ในช่วงของการสอบเทียบ นอกจากนี้สามารถระบุวิธีการประมาณค่า θ คือการถูกนำมาใช้สำหรับการให้คะแนน IRT ดังภาพที่ 2-10



ภาพที่ 2-10 แท็บการประเมินความยากของข้อสอบ (Guyer & Thompson, 2014, p. 18)

3.6 แท็บตัวเลือกระบุผลลัพธ์ ว่าจะได้รับการดำเนินการจัดหมวดหมู่เช่นเดียวกับตัวเลือกการส่งออกที่เพิ่มขึ้น ดังภาพที่ 2-11



ภาพที่ 2-11 แท็บตัวเลือกระบุผลลัพธ์ (Guyer & Thompson, 2014, p. 19)

การป้อนข้อมูลของโปแกรม Xcalibre 4.2 ต้องใส่ 2 ไฟล์ ในการวิเคราะห์ข้อมูล ได้แก่ ไฟล์ข้อมูลเมทริกซ์ และไฟล์ควบคุมข้อสอบ รูปแบบของแฟ้มเหล่านี้จะอธิบายไว้ในส่วนต่อไปนี้ อาจทำงาน Xcalibre 4.2 โดยใช้ไฟล์ข้อมูลเข้าในรูปแบบที่ใช้โดยรุ่นก่อนหน้านี้ ที่รวมแฟ้มข้อมูลที่มีข้อมูลการควบคุมข้อสอบในไฟล์เดียว; โดยใช้ตัวเลือกนี้ แต่ไม่ได้ช่วยให้สามารถเข้าถึงคุณลักษณะทั้งหมดของโปรแกรม Xcalibre 4.2

4. ความเที่ยงและวิธีการประมาณค่า

ในการใช้โปรแกรม Xcalibre ในโปรแกรมยังสามารถประมาณค่าความเที่ยง ซึ่งความเที่ยงนั้น เยาวดี รางชัยกุล วิบูลย์ศรี (2556) ได้ให้ความหมายของความเที่ยงไว้ 2 ประการ คือ ประการแรก หมายถึงความคงที่ของคะแนนที่ได้ของผู้ตอบข้อสอบ ย่อมบ่งชี้ถึงระดับความสามารถที่กระทำได้ของผู้สอบ ประการที่สองคือ ค่าความเที่ยงจะช่วยให้เราสามารถประมาณคะแนนจริง (True Score) ของผู้สอบซึ่งแสดงถึงความสามารถที่แท้จริงของผู้สอบได้โดยการหาสัมประสิทธิ์ความเที่ยงเครื่องมือที่ใช้ในการวัดเป็นสิ่งสำคัญอย่างหนึ่ง เนื่องจากเครื่องมือการวัดใดที่มีความเที่ยงในการวัดสูง เมื่อเรานำเครื่องมือนั้นไปใช้วัดกี่ครั้งก็จะได้ผลคงเดิมหรือคงที่หมายความว่า แบบสอบใดที่ให้ผลการสอบที่แสดงถึงคุณสมบัติของสิ่งที่วัดได้อย่างคงที่และถูกต้องแม่นยำแล้ว แบบทดสอบนั้นย่อมมีความเที่ยงสูงโดยปกติแล้วค่าความเที่ยงมักจะอยู่ในรูปของค่าสัมประสิทธิ์สหสัมพันธ์มักจะอยู่ระหว่าง -1 ถึง +1 แต่ค่าสัมประสิทธิ์ความเที่ยงมีค่าอยู่ระหว่าง 0 ถึง 1 เท่านั้น โดยองค์ประกอบที่มีอิทธิพลต่อความเที่ยง คือ จำนวนข้อคำถาม ยิ่งมาก ยิ่งมีความเที่ยงสูง แต่ถ้ามากเกินไป จะทำให้ผู้ตอบไม่ตั้งใจตอบ องค์ประกอบต่อมาคือ ความยากง่ายของข้อคำถาม ถ้ายากเกินไปหรือง่ายเกินไป จะทำให้คะแนนที่ได้มีค่าไม่แตกต่างกันมาก หรือมีความแปรปรวนน้อย ซึ่งจะส่งผลให้มีค่าความเที่ยงต่ำ องค์ประกอบที่

สามคือ ความชัดเจนของข้อคำถาม ถ้ามีลักษณะกำกวม ไม่ชัดเจน ตีความได้หลากหลาย หรือขาดความเป็นปรนัยแล้ว จะทำให้เครื่องมือมีค่าความเที่ยงต่ำ เนื่องจากความแปรปรวนของคะแนนความคลาดเคลื่อนสูง (คะแนนที่ได้ไม่ใช่คะแนนที่แท้จริง แต่เป็นคะแนนที่เกิดจากการเดา) และความแปรปรวนของคะแนนที่วัดได้มีค่าต่ำ (เพราะทุกคนมั่วตอบ) องค์ประกอบที่สี่ คือ ความเป็นเอกพันธ์ของกลุ่ม ถ้ากลุ่มที่แสดงความคิดเห็น หรือทำแบบทดสอบมีความสามารถใกล้เคียงกัน หรือมีความคิดเห็นไปในทำนองเดียวกันทั้งหมด จะทำให้คะแนนที่สังเกตได้มีค่าไม่แตกต่างกันมาก หรือมีความแปรปรวนต่ำ ส่งผลให้ค่าความเที่ยงต่ำไปด้วย และองค์ประกอบสุดท้ายคือจำนวนตัวเลือกตอบ ถ้าน้อย เช่น ถูก-ผิด หรือ ให้เลือกตอบ 3 ข้อ มีผลทำให้คะแนนที่วัดได้มีความแปรปรวนต่ำ ซึ่งส่งผลต่อค่าความเที่ยงต่ำไปด้วย ด้วยเหตุนี้แบบสอบถามที่เป็นมาตราวัดเรียงลำดับ 5 ระดับจะดีกว่า 3 ลำดับ แต่ถ้ามีมากไป เช่น 7 ระดับ ผู้ตอบจะเกิดความยากลำบากในการตัดสินใจตอบ การประมาณค่าความเที่ยงแบ่งออกเป็น 4 ประเภท ได้แก่

4.1 ความเที่ยงแบบคงที่ (Measure of Stability) เป็นความคงเส้นคงวาของคะแนนจากการวัดในช่วงเวลาที่ต่างกันโดยใช้วิธีสอบซ้ำ (Test-retest Method)

4.2 ความเที่ยงแบบความสมมูล (Measure of Equivalence) เป็นความสอดคล้องของคะแนนจากการวัดในช่วงเวลาเดียวกัน โดยใช้เครื่องมือ 2 ฉบับที่ทัดเทียมกัน

4.3 ความเที่ยงแบบคงที่และสมมูล (Measure of Stability and Equivalence) เป็นความสอดคล้องกันของคะแนนในช่วงเวลาที่ต่างกัน ใช้วิธีการทดสอบซ้ำ โดยใช้เครื่องมือ 2 ฉบับที่ทัดเทียมกัน

4.4 ความเที่ยงแบบความสอดคล้องภายใน (Measure of Internal Consistency) เป็นความสอดคล้องกันระหว่างรายข้อหรือความเป็นเอกพันธ์ของเนื้อหาอันเป็นคุณลักษณะเด่นเดียวกันกับสิ่งที่ต้องการวัดด้วยวิธีการหลากหลายวิธี ได้แก่ วิธีแบ่งครึ่ง (Split-half Method) เป็นการคำนวณสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนที่วัดได้จากการแบ่งครึ่งที่สมมูลกัน ส่วนวิธีของคูเดอร์-ริชาร์ดสัน (Kuder- Richartson method) เป็นการคำนวณค่าสถิติของคะแนนรายข้อ ซึ่งให้คะแนนแบบ 0, 1 และคะแนนรวม (Kuder & Richardson, 1937) วิธีวิเคราะห์ความแปรปรวนของฮอยส์ (Hoyt's Analysis of Variance Method) เป็นการวิเคราะห์ความแปรปรวน (Hoyt, 1941) และ วิธีสัมประสิทธิ์แอลฟาของครอนบาค (Cronbach's Alpha Method) เป็นการคำนวนค่าสถิติรายข้อและคะแนนรวม (Conbach, 1951) โดยการวิจัยในครั้งนี้ใช้วิธีสัมประสิทธิ์แอลฟ่าของครอนบาค (Cronbach's Alpha Method) เนื่องจากโปรแกรม Xcaribre สามารถตรวจสอบค่าความเที่ยงโดยใช้วิธีนี้ ซึ่งเป็นวิธีที่นิยมใช้กันมากที่สุด เครื่องมือดังกล่าวเป็นแบบทดสอบแบบถูกเป็น 1 ผิดเป็น 0 หรือจะให้เป็นคะแนนอื่นๆ นอกเหนือไปจาก 0 กับ 1 หรือจะเป็นแบบประเมินค่า (Rating Scale) แต่ละข้อคำถามไม่จำเป็นต้องมีความยาก ง่ายเท่ากัน กรณีเป็นแบบทดสอบที่มีคำตอบเป็น 1 กับ 0 สูตรนี้จะให้ผลเหมือนกับสูตร KR_{20} ซึ่งจะไม่ขอกล่าวในที่นี้

สูตร $\alpha = \frac{\mathrm{K}}{\mathrm{K}-1}(1 - \frac{\sum S_i^2}{S_t^2})$ (4)

โดยที่

α คือ สัมประสิทธิ์ความเที่ยงของแบบทดสอบ

K คือ จำนวนข้อสอบ

$\sum S_i^2$ คือ ผลรวมของความแปรปรวนของคะแนนแต่ละข้อ

S_t^2 คือ ความแปรปรวนของคะแนนรวม

การหาสัมประสิทธิ์แอลฟาของครอนบาคจะให้ค่าความเชื่อมั่นของเครื่องมือวิจัยได้ดีก็ต่อเมื่อเครื่องมือ วิจัยชุดนั้นได้วัดคุณลักษณะเพียงคุณลักษณะเดียวเท่านั้น และจำนวนข้อหรือองค์ประกอบแต่ละ องค์ประกอบในฉบับมีความเท่าเทียมกันจะทาให้ได้ค่าความเชื่อมั่นที่ใกล้เคียงกับความเชื่อมั่นที่แท้จริง ของเครื่องมือการวิจัย และเป็นวิธีการที่ได้รับความนิยมใช้กันอย่างแพร่หลายสำหรับการประมาณค่าความเที่ยงสอดคล้องภายใน เพราะสะดวกในการนำไปใช้เนื่องจากเก็บข้อมูลกลุ่มผู้ให้ข้อมูลครั้งเดียว และใช้ได้อย่างหลากหลายทั้งเครื่องมือที่ให้คะแนนแบบ 0,1 หรือแบบถ่วงน้ำหนัก หรือแบบมาตรส่วน ประมาณค่า หรือแบบทดสอบแบบอัตนัย

งานวิจัยที่เกี่ยวข้องกับทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT) มีดังนี้

สุนทร เทียนงาม ศิริชัย กาญจนวาสี และดิเรก ศรีสุโข (2553) ได้ศึกษาผลความไม่เป็นอิสระของข้อสอบที่มีต่อความเที่ยง ค่าพารามิเตอร์ ค่าความสามารถของผู้สอบและค่าสาระสนเทศของแบบทดสอบเมื่อมีเงื่อนไขการทดสอบแตกต่างกัน โดยมีวัตถุประสงค์เพื่อ วิเคราะห์ผลของระดับความเป็นอิสระของข้อสอบต่อความเที่ยงของแบบสอบ ค่าพารามิเตอร์ของข้อสอบ ค่าความสามารถของผู้สอบและสารสนเทศในแบบสอบด้วยโมเดลทฤษฎีการตอบสนองข้อสอบ ที่มีผลต่อจำนวนข้อสอบและขนาดของกลุ่มตัวอย่างที่แตกต่างกัน พร้อมด้วยวิเคราะห์ลักษณะความไม่เป็นอิสระของข้อสอบจากแบบสอบจริงในวิชาวิทยาศาสตร์และภาษาอังกฤษ จำนวนข้อแบ่งเป็น 3 กลุ่ม (30, 50, 80) ขนาดของกลุ่มตัวอย่าง แบ่งเป็น 3 กลุ่ม (400, 800, 1200) โดยผลการวิจัยพบว่า เมื่อระดับความไม่เป็นอิสระของข้อสอบเพิ่มสูงขึ้น ค่าความเที่ยง (KR20) จะเพิ่มสูงขึ้นทุกเงื่อนไขของการทดสอบ เมื่อวิเคราะห์ข้อสอบตามทฤษฎีการตอบสนองข้อสอบปรากฏว่า เมื่อใช้โมเดล 1 พารามิเตอร์ ระดับความไม่เป็นอิสระสูงขึ้น ค่าความยากและค่าความสามารถไม่แตกต่างกัน ค่าสารสนเทศของแบบทดสอบคงที่ เมื่อใช้โมเดล 2 พารามิเตอร์ พบว่า ค่าความยากลดลง แต่ค่าความสามารถและค่าสารสนเทศของแบบทดสอบมีแนวโน้มสูงขึ้นและเมื่อใช้โมเดล 3 พารามิเตอร์ พบว่า ค่าความยาก ค่าความสามารถ และค่าสารสนเทศของแบบทดสอบมีแนวโน้มลดลงแต่ค่าอำนาจจำแนก ค่าการเดา ส่วนใหญ่มีแนวโน้มสูงขึ้น ส่วนการวิเคราะห์ความเป็นอิสระของข้อสอบทั้งวิชาวิทยาศาสตร์และวิชาภาษาอังกฤษ พบว่าข้อสอบส่วนใหญ่มีความไม่เป็นอิสระของข้อสอบเพียงเล็กน้อย

จิระนาฏ ฉวีพัฒน์ บัณฑิต อินสมบัติ และธีรยุทธ ภูเขา (2554) ได้ศึกษาเกี่ยวกับการเปรียบเทียบคุณภาพของการปรับเทียบคะแนนตามแนวตั้งโดยการใช้แบบทดสอบร่วมภายในระหว่างวิธีเชิงเส้นตรงกับวิธีทฤษฎีการตอบสนองข้อสอบแบบสามพารามิเตอร์ โดยมีวัตถุประสงค์เพื่อศึกษา

คุณภาพของการปรับเทียบคะแนนตามแนวตั้งโดยการใช้แบบทดสอบร่วมภายใน โดยการเปรียบเทียบความคลาดเคลื่อนและความพอเพียงระหว่างวิธีเชิงเส้นตรงกับวิธีทฤษฎีการตอบสนองข้อสอบแบบสามพารามิเตอร์ โดยใช้กลุ่มตัวอย่างจำนวน 1,080 คน เครื่องมือที่ใช้ในการทดสอบคือแบบทดสอบวิชาคณิตศาสตร์ จำนวน 4 ฉบับ พบว่า ความคลาดเคลื่อนของวิธีทฤษฎีการตอบสนองข้อสอบแบบสามพารามิเตอร์มีค่าน้อยกว่าวิธีเชิงเส้น ส่วนการทดสอบความพอเพียง ปรากฏว่า วิธีทฤษฎีการตอบสนองข้อสอบแบบสามพารามิเตอร์มีความพอเพียงในระดับความน่าพอใจกว่าวิธีการปรับเทียบเชิงเส้นตรง ซึ่งทำให้ทราบว่าวิธีทฤษฎีการตอบสนองข้อสอบแบบสามพารามิเตอร์มีคุณภาพกว่าวิธีเชิงเส้นตรง

ณรงค์ จันทรมหา (2554) ได้ศึกษาการเปรียบเทียบค่าความเที่ยงของแบบทดสอบผลสัมฤทธิ์ทางการเรียนที่มีจำนวนข้อสอบทำหน้าที่ต่างกันแตกต่างกัน มีวัตถุประสงค์เพื่อเปรียบเทียบค่าความเที่ยงของแบบทดสอบผลสัมฤทธิ์ทางการเรียนที่มีจำนวนข้อสอบทำหน้าที่แตกต่างกัน 7 เงื่อนไข กลุ่มตัวอย่าง จำนวน 2,000 คน โดยใช้แบบทดสอบทางการศึกษาระดับชาติ วิชาภาษาไทย จำนวน 40 ข้อ ซึ่งวิเคราะห์ค่าสถิติพื้นฐานโดยใช้โปรแกรม SPSS และตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยใช้โปรแกรม SIBTEST พร้อมทั้งทดสอบความแตกต่างของค่าความเที่ยงของแบบทดสอบ โดยใช้การทดสอบสถิติซี ผลการศึกษาพบว่าการเปรียบเทียบของแบบทดสอบระหว่างฉบับที่ไม่มีข้อสอบทำหน้าที่ต่างกันกับฉบับที่มีข้อสอบที่ทำหน้าที่ต่างกัน ปรากฏว่า ค่าความเที่ยงของแบบทดสอบมีค่าไม่สูงมากนักและมีค่าไม่แตกต่างกัน

Moghadamzadeh, Salehi and Khodaie (2011) ได้ศึกษาการเปรียบเทียบฟังก์ชั่นของข้อสอบและแบบทดสอบในโมเดล 1, 2, 3 พาราเมทริก ของทฤษฎีการตอบสนองข้อสอบ (IRT) มีการกำหนดรูปแบบข้อมูลบนรากฐานทางทฤษฎีและเชิงจิตวิทยา ข้อมูลเป็นผลการตอบแบบทดสอบของนักเรียนอิหร่านที่สอบเข้าเรียนมหาวิทยาลัยในสาขาคณิตศาสตร์-ฟิสิกส์ ปีการศึกษา 2009 ขององค์การทดสอบทางการศึกษาแห่งชาติ (NOET) กลุ่มตัวอย่างจำนวน 2,000 คน ใช้โปรแกรม SPSS และ BILOG ในการวิเคราะห์ข้อมูล พบว่า ค่าของฟังก์ชั่นของข้อสอบและแบบทดสอบของโมเดล 2 พามิเตอร์ มีค่าสูงกว่า โมเดล 1 และ 3 พารามิเตอร์ นอกจากนี้ยังพบว่าค่าของฟังก์ชั่นของข้อสอบและแบบทดสอบ ของโมเดล 1 พารามิเตอร์ไม่แตกต่างกันอย่างมีนัยสำคัญจากค่าที่สอดคล้องกันทั้ง 3 พารามิเตอร์ เมื่อพิจารณาถึงค่าประมาณในโมเดลทั้ง 3 การศึกษาครั้งนี้มีความหมายสำหรับการวัดและทดสอบทางการศึกษาโดยเฉพาะอย่างยิ่งสำหรับการทดสอบแบบปรนัยที่มีมาตรฐานพบดัชนีความน่าเชื่อถือที่คำนวณจากระดับความสามารถแตกต่างกัน ในทฤษฎีการตอบสนองข้อสอบคือเป็นไปได้ที่จะสร้างแบบทดสอบที่ตรงกับวัตถุประสงค์ และสามารถตรวจสอบความถูกต้องของการทดสอบโดยใช้ฟังก์ชั่นข้อมูลได้

Kose and Demirtasli (2012) ได้ศึกษาการเปรียบเทียบรูปแบบการวัดมิติเดียวและการวัดหลายมิติบนพื้นฐานทฤษฎีการตอบสนองข้อสอบ (IRT) ในแง่ของ 2 ตัวแปรได้แก่ ความยาวของแบบทดสอบและขนาดของกลุ่มตัวอย่าง ข้อมูลเป็นผลการตอบการทดสอบภาษาตุรกี จากนักเรียนเกรด 8 จำนวน 1516 คน พบว่า ค่าพารามิเตอร์ความสามารถที่ประมาณค่าภายใต้ทฤษฎีการตอบสนองข้อสอบ (IRT) มีอัตราความคลาดเคลื่อนน้อยลงและมีอัตราความถูกต้องเพิ่มมากขึ้นในการวัดหลายมิติ ค่าที่ได้เหมาะสำหรับทฤษฎีการตอบสนองข้อสอบ (IRT) แบบการวัดหลายมิติ

นอกจากนี้การควบคุมขนาดตัวอย่างและความยาวของแบบทดสอบยังไม่มีผลในเชิงบวกบนความพอดีของข้อมูลในทฤษฎีการตอบสนองข้อสอบ (IRT) แบบการวัดมิติเดียว แต่ข้อสอบที่มากขึ้นและขนาดกลุ่มตัวอย่างที่ใหญ่ขึ้นสามารถลดการประมาณค่าความคลาดเคลื่อนและรายละเอียดเพิ่มมากภายใต้การวัดผลทฤษฎีการตอบสนองข้อสอบ (IRT) แบบการวัดหลายมิติ

Basokcu and Ogretmen (2014) ได้ศึกษาการเปรียบเทียบเทคนิคการตอบสนองข้อสอบแบบพาราเมทริกและนันพาราเมทริกในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในข้อสอบแบบตรวจให้คะแนนมากกว่า 2 ค่า โดยมีวัตถุประสงค์เพื่อเปรียบเทียบการตอบสนองข้อสอบแบบพาราเมทริกและนันพาราเมทริกบนพื้นฐานทฤษฎีการตอบสนองข้อสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในข้อสอบแบบตรวจให้คะแนนมากกว่า 2 ค่า การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบนั้นอยู่บนพื้นฐานของทฤษฎีการตอบสนองข้อสอบแบบพาราเมทริกได้ดำเนินการโดยใช้วิธีการเปรียบเทียบพารามิเตอร์ สำหรับการวิเคราะห์ทฤษฎีการตอบสนองข้อสอบแบบนันพารา-เมทริกนั้น การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบจะถูกกำหนดโดยการเปรียบเทียบดัชนีพื้นที่ที่เกี่ยวข้องกับไอซีซีได้รับสำหรับกลุ่มอ้างอิงและกลุ่มเปรียบเทียบโดยกลุ่มตัวอย่างคือนักเรียนเกรด 8 ที่เป็นนักเรียนจากประเทศตุรกีและเกาหลีใต้ เพื่อให้ได้กลุ่มตัวอย่างภายใต้เงื่อนไขประเทศและเพศ โดยดำเนินการในชุดข้อมูลจาก TIMSS 2011 ในหัวข้อ"ทัศนคติต่อวิชาคณิตศาสตร์" จากการสังเกตุพบว่าทั้งแบบพาราเมทริกและนันพาราเมทริกในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบภายใต้เงื่อนไขประเทศมีผลคล้ายกัน อย่างไรก็ตามผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบสำหรับประเทศภายใต้เงื่อนไขเพศที่แตกต่างกันตามเทคนิคการขึ้นอยู่กับทฤษฎีการตอบสนองแบบพาราเมทริกและนันพาราเมทริก ผลการศึกษาแสดงให้เห็นว่าข้อสอบที่ผสมผสานการทำหน้าที่ต่างกันของข้อสอบ เป็นเทคนิคที่ต้องการ

จากการศึกษางานวิจัยที่เกี่ยวข้องกับทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT) แบบ 3 พารามิเตอร์ ค่าความยากของแบบทดสอบมีแนวโน้มลดลงแต่ค่าอำนาจจำแนก ค่าโอกาสการเดา ส่วนใหญ่มีแนวโน้มสูงขึ้น ความคลาดเคลื่อนของวิธีทฤษฎีการตอบสนองข้อสอบแบบ 3 พารามิเตอร์มีค่าน้อยกว่าวิธีเชิงเส้น กลุ่มตัวอย่างที่มีขนาดแตกต่างกัน ทำให้มีค่าอำนาจจำแนก ค่าความยาก และค่าโอกาสการเดาข้อสอบแตกต่างกัน

ตอนที่ 3 การวิเคราะห์ความตรงเชิงโครงสร้าง (Construct Validity) และงานวิจัยที่เกี่ยวข้อง

1. ความเป็นมาของการตรวจสอบความตรง

การตรวจสอบความตรงเชิงโครงสร้างนั้นในงานวิจัยนี้ใช้การวิเคราะห์องค์ประกอบเชิงยืนยันในการตรวจสอบ (Confirmatory Factory Analysis: CFA) ในปี ค.ศ. 1901 เพียรสันเป็นผู้เสนอวิธีการสร้างแกนมุขสำคัญ (Principle Axes) จากนั้นก็มีนักจิตวิทยาหลายคนได้นำทฤษฎีมาประยุกต์ใช้ ต่อมา Spearman (1904) ได้เสนอโมเดลสององค์ประกอบในการศึกษาทางสติปัญญาว่าแบบทดสอบแต่ละฉบับวัดองค์ประกอบร่วมของสถิติปัญญาส่วนหนึ่งและอีกส่วนหนึ่งวัดองค์ประกอบเฉพาะ Holzinger (1937) เป็นผู้เสนอแนวคิดว่าโมเดลสององค์ประกอบอาจมีองค์ประกอบทั่วไปเป็นองค์ประกอบร่วมของคะแนนจากแบบทดสอบแต่ละฉบับได้มากกว่า 1 องค์ประกอบและเสนอทฤษฎี

ทวิองค์ประกอบ (Bi-factor Theory) ช่วงปี ค.ศ. 1931-1947 Thurstone (1947) ได้พัฒนาวิธีการวิเคราะห์ที่ว่าการวิเคราะห์องค์ประกอบพหุคูณ (Multiple-factor Analysis) ร่วมกับวิธีการสร้างส่วนประกอบมุขสำคัญ (Principle Component) ซึ่งเป็นแนวคิดของ Hotelling (1993) โดย Lawley (1940) เริ่มคิดวิธีการประมาณค่าพารามิเตอร์ด้วยไลค์ลีฮูดสูงสุด (Maximum Likelihood) จากนั้นได้มีนักสถิติหลายท่านพยายามพัฒนาวิธีการวิเคราะห์องค์ประกอบเชิงยืนยันให้ดีขึ้น โดยบุคคลที่มีส่วนอย่างมากในการพัฒนา โดยในปี ค.ศ. 1966 Bocdk และ Bargmann (1966) ได้เสนอวิธีการทดสอบสมมติฐานทางสถิติเกี่ยวกับพารามิเตอร์ในการวิเคราะห์องค์ประกอบ และในระหว่างปี ค.ศ. 1966-1970 เริ่มพัฒนาวิธีการคำนวณและโปรแกรมคอมพิวเตอร์สำหรับการวิเคราะห์องค์ประกอบ โดย Joreskog ซึ่งได้พัฒนาเป็นโมเดลลิสเรลที่ใช้โปรแกรม LISREL ซึ่งสามารถใช้ได้ถึงปัจจุบัน

2. ประเภทของการตรวจสอบความตรง

ศิริชัย กาญจนวาสี (2556) ได้อธิบาย ความตรง (Validity) หรือความถูกต้องเป็นคุณสมบัติที่สำคัญประการหนึ่งของเครื่องมือวิจัยกล่าวคือ เป็นเครื่องมือที่สามารถวัดได้ตรงกับที่ต้องการจะวัด การตรวจสอบความตรงทำได้หลายวิธี ดังนี้

2.1 ความตรงตามเนื้อหา (Content Validity) เป็นการมองโดยส่วนรวมว่าเครื่องมือหรือชุดของคำถามหรือแบบวัดนั้น ครอบคลุมเนื้อหาที่จะวัดได้ถูกต้องครบถ้วนหรือไม่ เช่น แบบประเมินคุณภาพของการพยาบาล ผู้วิจัยสร้างไว้จำนวน 30-40 ข้อ ย่อมต้องการการตรวจสอบ 30-40 ข้อนั้น เป็นตัวแทนของพฤติกรรมการพยาบาลทั้งหมดที่ต้องการประเมินแล้วหรือยัง การตรวจสอบต้องอาศัยผู้เชี่ยวชาญด้านเนื้อหาของเครื่องมือนั้น ๆ ในทางปฏิบัติต้องการผู้เชี่ยวชาญตรวจสอบ ครั้งละ 3-5 ท่าน ภายหลังการตรวจสอบผู้สร้างเครื่องมือจะนำข้อแนะนำที่ได้รับมาแก้ไขปรับปรุงจนกว่าจะได้รับการยอมรับจากผู้เชี่ยวชาญความเห็นพ้องต้องกันของผู้เชี่ยวชาญ แสดงถึงการมีความตรงตามเนื้อหาของเครื่องมือนั้น คิดเป็นร้อยละ 80

2.2 ความตรงตามเกณฑ์สัมพันธ์ (Criterion-Related Validity) เป็นการประเมินความตรงตามเกณฑ์ที่ได้มาตรฐานการตรวจสอบที่มุ่งหาความสัมพันธ์ระหว่างค่าที่วัดได้จากเครื่องมือที่สร้าง กับค่าที่วัดได้จากเกณฑ์ ความสำคัญอยู่ที่เกณฑ์ที่ผู้วิจัยเลือกว่าถูกต้องตามหลักทฤษฎีได้มาตรฐานเป็นที่ยอมรับโดยทั่วไปหรือไม่ เกณฑ์ที่เลือกใช้มี 2 ลักษณะ ขึ้นอยู่กับช่วงเวลาที่ต้องการให้เครื่องมือนั้นวัดได้ตรงตามเกณฑ์ กล่าวคือ

2.2.1 ตรงตามสภาพ (Concurrent Validity) หมายถึง ลักษณะที่เครื่องมือวัดได้มีความตรงตามสภาพความเป็นจริงโดยทั่ว ๆ ไปในเวลานั้น

2.2.2 ความตรงตามทำนาย (Predictive Validity) หมายถึง ลักษณะที่เครื่องมือวัดได้มีความตรงตามความจริงที่จะเกิดขึ้นตามมาภายหลัง หรือในอนาคต ซึ่งสามารถทำนายได้

2.3 ความตรงตามโครงสร้าง (Construct Validity) เป็นการมองความตรงของเครื่องมือวิจัยในแง่ที่จะบอกว่าสิ่งที่ได้มาจากการวัดนั้นมีความตรงตามแนวคิดเชิงทฤษฎีอย่างไร ซึ่งเป็นการยากที่จะตรวจสอบเนื่องจากแนวคิดเชิงทฤษฎีนี้มักอยู่ในรูปนามธรรม ความตรงตามโครงสร้างมีความสำคัญในแง่ของการเชื่อมโยงระหว่างทฤษฎีและสิ่งที่วัดได้จริงจากการปฏิบัติ เช่น ในเรื่องสติปัญญา บุคลิกภาพ ความถนัด เป็นต้น เรื่องต่าง ๆ เหล่านี้มีโครงสร้างทางทฤษฎีอย่างไร

เมื่อนำมาถ่ายทอดเป็นข้อความในเครื่องมือวิจัย องค์ประกอบและรายละเอียดต่าง ๆ ในเครื่องมือนั้นตรงตามทฤษฎีมากน้อยเพียงใด ถ้าวัดได้ครบถ้วนถือว่ามีความถูกต้องตามโครงสร้างการตรวจสอบความตรงเชิงโครงสร้างอาจใช้กลุ่มผู้เชี่ยวชาญ หรือใช้ Known Group Technique ขึ้นอยู่กับว่าวิธีใดจะเป็นไปได้ (Cronbach & Meenl, 1955)

3. การวิเคราะห์องค์ประกอบเชิงยืนยัน (Confirmatory Factor Analysis: CFA)

เป็นการวิเคราะห์ทางสถิติชั้นสูงที่มีความสลับซับซ้อนในขั้นตอนการดำเนินงาน ซึ่งในที่นี้จะขอกล่าวในส่วนที่เกี่ยวข้องกับสิ่งที่ผู้วิจัยได้ศึกษา ดังนี้

3.1 โมเดลตัวประกอบเชิงยืนยัน (Confirmatory Factory Model) เป็นการวิเคราะห์และการตรวจสอบความสอดคล้องระหว่างข้อมูลกับโมเดลซึ่งใช้โปรแกรมคอมพิวเตอร์ชื่อ LISREL (Linear Structure Relationship) (Joreskog & Sorbom, 1985) โปรแกรมได้นำเสนอความสอดคล้องของข้อมูลเชิงประจักษ์ วิเคราะห์รายงานผลด้วยค่าดัชนีเป็นภาพรวมโมเดลสมการโครงสร้าง ค่าดัชนีต่างๆ ถูกใช้ในแต่ละกรณี ดังตารางที่ 2-2

ตารางที่ 2-2 ดัชนีที่ใช้ในการพิจารณาความสอดคล้องของโมเดลตามสมมติฐานกับข้อมูลเชิงประจักษ์ (พูลพงศ์ สุขสว่าง, 2556, หน้า 15-16)

ดัชนีตรวจสอบความสอดคล้อง	ค่าที่เป็นไปได้	เกณฑ์การพิจารณา
ค่าสถิติทดสอบไค-สแควร์ (Chi-square)	0 (Perfect Fit) to Positive Value (Poor Fit)	ค่าไค-แสควร์ที่คำนวณได้น้อยกว่าไค-แสควร์ตารางหรือพิจารณาค่า p-value ต้องมากกว่า 0.05 มีค่าน้อยกว่า 2.00
ค่าไค-แสควร์สัมพันธ์ ($\chi^2_{model}/df_{model}$)	0 (Perfect Fit) to Positive Value (Poor Fit)	
Goodness of Fit Index (GFI)	0 (No Fit) to 1 (Perfect Fit)	มีค่ามากกว่า 0.95
Adjusted Goodness to Fit Index (AGFI)	0 (No Fit) to 1 (Perfect Fit)	มีค่ามากกว่า 0.95
Comparative Fit Index (CFI)	0 (No Fit) to 1 (Perfect Fit)	มีค่ามากกว่า 0.95
Tucker-Lewis Index (TLI) หรือ Non Norm Fit Index (NNFI)	0 (No Fit) to 1 (Perfect Fit)	มีค่ามากกว่า 0.95
Norm Fit Index (NFI)	0 (No Fit) to 1 (Perfect Fit)	มีค่ามากกว่า 0.95
Root Mean Square Residual (RMR)	0 (Perfect Fit) to Positive Value (Poor Fit)	มีค่าเข้าใกล้ 0 (ขึ้นอยู่กับระดับที่นักวิจัยกำหนด)
Standardized RMR (SRMR)	0 (Perfect Fit) to Positive Value (Poor Fit)	น้อยกว่า 0.05
Root Mean Square Error of Approximation (RMSEA)	0 (Perfect Fit) to Positive Value (Poor Fit)	มีค่าน้อยกว่า 0.05 หรือ 0.08
Parsimony Normed Fit Index (PNFI)	0 (No Fit) to 1 (Perfect Fit)	ใช้เปรียบเทียบโมเดลทางเลือก (Alternative Model) โมเดลในมีค่า PNFI สูงกว่าจะเป็นโมเดลที่ดี ใช้เปรียบเทียบโมเดลทางเลือก
Akaike Information Criterion (ACI)	0 (Perfect Fit) to Positive Value (Poor Fit)	(Alternative Model) โมเดลใดมีค่า AIC น้อยกว่าจะเป็น โมเดลที่ดีกว่า

3.2 ข้อตกลงเบื้องต้น คือ โมเดลสามารถยืดหยุ่นได้ สามารถเพิ่มหรือลดข้อจำกัดได้

3.3 ขั้นตอนการวิเคราะห์ตัวประกอบเชิงยืนยัน มีขั้นตอนพอสรุปได้ดังนี้

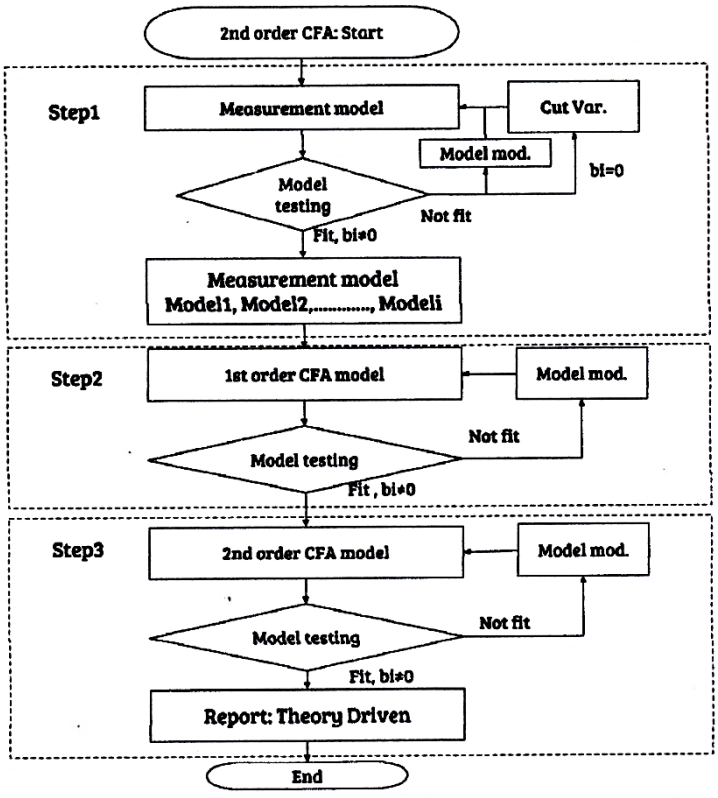
3.3.1 กำหนดรูปแบบของโมเดลตัวประกอบ (Specification of the Confirmatory Factor Model) ซึ่งต้องกำหนดรายละเอียด ได้แก่ จำนวนของตัวประกอบร่วม ความสัมพันธ์ระหว่างตัวประกอบร่วมกับตัวแปร ความแปรปรวนและความแปรปรวนร่วมระหว่างตัวประกอบร่วม และความแปรปรวนและความแปรปรวนร่วมระหว่างตัวประกอบส่วนที่เหลือ

3.3.2 ศึกษาคุณสมบัติที่จำเป็นสำหรับการประมาณค่าพารามิเตอร์ของโมเดล (Identification of the Confirmatory Factor Model) โดยเริ่มจากเงื่อนไขที่จำเป็นสำหรับโครงสร้างของโมเดล และความเพียงพอสำหรับการประมาณค่าพารามิเตอร์ของโมเดล

3.3.3 ทำการประมาณค่าพารามิเตอร์ของโมเดล (Estimation of the Confirmatory Factor Model) โดยการใช้โปรแกรมคอมพิวเตอร์ชื่อ LISREL (Joreskog & Sorborn, 1985)

4. การวิเคราะห์องค์ประกอบเชิงยืนยันลำดับที่สอง (2nd order CFA)

มีวัตถุประสงค์ เพื่อยืนยันตัวแปรที่ศึกษาว่ามีองค์ประกอบตรงตามทฤษฎีที่กำหนดจริง เพื่อตรวจสอบว่าภายในแต่ละองค์ประกอบค่าน้ำหนักของตัวแปรตัวใดมีน้ำหนักหรือความสำคัญมากกว่ากันในบริบทที่ศึกษาและส่วนที่เพิ่มเติม (พูลพงศ์ สุขสว่าง, 2561, หน้า 160) โดยสิ่งที่ต้องพิจารณา ในการวิเคราะห์องค์ประกอบเชิงยืนยันลำดับที่สองสามารถสรุปขั้นตอน ดังภาพที่ 2-12



ภาพที่ 2-12 ขั้นตอนการวิเคราะห์องค์ประกอบเชิงยืนยันลำดับที่สอง (2nd order CFA) (พูลพงศ์ สุขสว่าง, 2561, หน้า 165)

5. การปรับโมเดลการวัด

การปรับปรุงโมเดลเป็นขั้นตอนที่กระทำต่อเที่อค่าพารามิเตอร์บางค่าไม่แตกต่าง จาก 0 (ltl>1.96) หรือมีทิศทางของค่าพารามิเตอร์ไม่ตรงกับทฤษฎีที่กำหนดไว้ หรืออาจเกิดปัญหาทั้ง 2 อย่าง สามารถแยกการปรับโมเดลได้เป็น 2 ประเด็น ได้แก่ 1) การปรับโมเดลสมการโครงสร้างในส่วนที่เป็นความคลาดเคลื่อน 2) การปรับโมเดลสมการโครงสร้างในส่วนที่เป็นโมเดลการวัด

(พูลพงศ์ สุขสว่าง, 2556, หน้า 9) โดยนำผลการวิเคราะห์มาเป็นแนวทางในการปรับ โดยใช้ค่าดังต่อไปนี้

5.1 ค่า Standardized residuals เป็นค่าที่แสดงความแตกต่างระหว่าง observed covariance และ estimated covariance ที่มีทั้งค่า + และ – พิจารณาจากค่าที่อยู่ระหว่าง I2.5I ถึง I4.0I ค่าที่ต้องพิจารณาตัดข้อคำถามออกคือค่าที่เกิน I4.0I

5.2 ค่า Modification residuals เป็นการคำนวณค่า Covariance ที่เป็นไปได้ทั้งหมดในโมเดล ค่าที่ต้องพิจารณาคือค่าตั้งแต่ 4 ขึ้นไป การปรับโดมเดลการวัดจะต้องมีการอ้างอิงด้วยทฤษฎีและงานวิจัยที่เกี่ยวข้อง

สิ่งที่ควรระวังในการปรับโมเดล ได้แก่ การปรับโมเดลการวัดยังต้องการอ้างอิงด้วยทฤษฎีและงานวิจัยที่เกี่ยวข้อง เพราะการวิเคราะห์องค์ประกอบเชิงยืนยันเป็นการวิเคราะห์เพื่อยืนยันโมเดลตามทฤษฎี การปรับแก้อาจส่งผลต่อความตรงเชิงเนื้อหาได้ และการประมาณค่าพารามิเตอร์ แต่ละค่ามีผลมาจากการประมาณค่าพารามิเตอร์ตัวอื่นๆ ดังนั้นการแก้ที่ค่าใดค่าหนึ่งอาจไม่สามารถแก้ปัญหาที่เกิดขึ้นได้ (สุวิมล ติรกานันท์, 2555, หน้า 255)

การตรวจสอบความตรงเชิงโครงสร้างด้วยการวิเคราะห์องค์ประกอบเชิงยืนยันนั้นมีประสิทธิภาพในการวิเคราะห์ข้อมูลสอดคล้องกับงานวิจัย ของ ชัยยศ ชาวระนอง (2553) พบว่า การวิเคราะห์องค์ประกอบยืนยันอันดับสองแฝงภายในมีประสิทธิภาพมากที่สุด มีความสอดคล้องกับข้อมูลเชิงประจักษ์ เมื่อไม่มีการปรับโมเดลอยู่ในระดับดี ค่าดัชนีไค-สแควร์ ไม่แปรเปลี่ยนไปตามกลุ่มตัวอย่าง

งานวิจัยที่เกี่ยวข้องของการวิเคราะห์ความตรงเชิงโครงสร้าง (Confirmatory Factor Analysis: CFA) มีดังนี้

ชัยยศ ชาวระนอง ไพรัตน์ วงษ์นาม สมศักดิ์ ลิลา และสุรีพร อนุศาสนนันท์ (2553) ได้ทำการศึกษาประสิทธิภาพของโมเดลและการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบพหุมิติโดยวิธีการวิเคราะห์องค์ประกอบเชิงยืนยันแฝงภายในมีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของโมเดลการวิเคราะห์องค์ประกอบยืนยันอันดับหนึ่ง, องค์ประกอบยืนยันอันดับสอง, องค์ประกอบยืนยันอันดับหนึ่งแฝงภายใน, องค์ประกอบยืนยันอันดับสองแฝงภายใน ภายใต้เงื่อนไขขนาดของกลุ่มตัวอย่าง,จำนวนข้อสอบต่อองค์ประกอบและขนาดกลุ่มตัวอย่างในการทำหน้าที่ต่างกันของข้อสอบ พบว่า การวิเคราะห์องค์ประกอบยืนยันอันดับสองแฝงภายในมีประสิทธิภาพมากที่สุด มีความสอดคล้องกับข้อมูลเชิงประจักษ์ เมื่อไม่มีการปรับโมเดลอยู่ในระดับดี ค่าดัชนีไค-สแควร์ ไม่แปรเปลี่ยนไปตามกลุ่มตัวอย่าง แต่แปลเปลี่ยนตามจำนวนข้อสอบต่อองค์ประกอบเมื่อนำข้อสอบมาตรวจตรวจสอบการทำหน้าที่ต่างกันโดยใช้วิธีการวิเคราะห์องค์ประกอบยืนยันอันดับสองแฝงภายใน พบว่ามีข้อสอบที่ทำหน้าที่ต่างกัน 5 ข้อ และเมื่อตัดข้อสอบที่ทำหน้าที่ต่างกันออก พบว่ามีค่าสัมประสิทธิ์ความเชื่อมั่นสูง

วลัยพร สุขปลั่ง และบรรพต วิรุณราช (2558) ได้วิเคราะห์องค์ประกอบเชิงยืนยันอันดับที่สองของโมเดลสมรรถนะบุคลากรสายผู้สอนสาขาวิชาการบริหารทรัพยากรมนุษย์ สังกัดมหาวิทยาลัยราชภัฏ บุคคลากรสายผู้สอน 3 ด้านคือ สมรรถนะหลัก สมรรถนะทั่วไป และสมรรถนะประจำสายงาน เพื่อรองรับการเข้าสู่ประชาคมอาเซียนจำนวน 150 ท่าน โดยใช้แบบสอบถามเป็นเครื่องมือใน

การวิจัย วิเคราะห์ข้อมูลด้วยการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับที่สอง พบว่า โมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับที่สอง มีความสอดคล้องกับข้อมูลเชิงประจักษ์ พบว่า สมรรถนะหลักที่มีอิทธิพลต่อบุคลากรสายผู้สอนสูงสุดได้แก่ สมรรถนะด้านการมุ่งผลสัมฤทธิ์ในการปฏิบัติงาน

Goh and Aryadoust (2010) ได้ตรวจสอบความตรงเชิงโครงสร้างของ แบบทดสอบทางด้านการฟังภาษาอังกฤษ ในรูปแบบราชโมเดลและโมเดลความสัมพันธ์ที่เป็นเอกลักษณ์ (CUM) โดยใช้การวิเคราะห์องค์ประกอบเชิงยืนยัน (CFA) เป็นเกณฑ์ในการทดสอบเนื่องจากการวิเคราะห์องค์ประกอบเชิงยืนยัน (CFA) เนื่องจากโครงสร้างมีความถูกต้องและสามารถเป็นเกณฑ์ในการทดสอบได้ ซึ่งกลุ่มตัวอย่างจำนวน 916 คน ผู้วิจัยได้ทำการประเมิน 3 องค์ประกอบ คือ เข้าใจในบริบท การปฏิสัมพันธ์สั้นและยาว ซึ่งเป็นองค์ประกอบที่เหมาะสมและเพียงพอต่อการวิเคราะห์ข้อมูล พบว่าการทดสอบค่อนข้างง่ายสำหรับกลุ่มตัวอย่าง ซึ่งมีความสามารถในการฟังที่หลากหลาย กลุ่มตัวอย่างมีทักษะการฟังที่สูง การทดสอบโดยใช้ราชโมเดลถือว่ามีทดทดสอบที่ดีและมีโครงสร้างใกล้เคียงกับโมเดลที่เป็นเกณฑ์ ข้อมูลที่ได้รับสนับสนุนความตรงเชิงโครงสร้างของการทดสอบ

Tan ม.ร.ว. สมพร สุทัศนีย์ และสุพิมพ์ ศรีพันธ์วรสกุล (2013) ได้พัฒนามาตรวัดแรงจูงใจในการเรียนภาษาอังกฤษสำหรับนักเรียนชั้นมัธยมศึกษาตอนปลาย ในจังหวัดกระแจะ ราชอาณาจักรกัมพูชา กลุ่มตัวอย่างคือนักเรียนที่เรียนในระดับชั้น 10 ถึงระดับชั้น 12 จำนวน 1,619 คน ประกอบด้วย 4 องค์ประกอบได้แก่ ได้แก่ ความคิดที่ส่งเสริม พฤติกรรมที่ส่งเสริม ความคิดที่ขัดขวาง และพฤติกรรมที่ขัดขวาง มาตรวัดทั้งฉบับมีค่าความเที่ยง เท่ากับ .94 มาตรวัดแรงจูงใจในการเรียนภาษาอังกฤษสาหรับนักเรียนชั้นมัธยมศึกษาตอนปลาย มีความตรงเชิงโครงสร้าง จากการวิเคราะห์องค์ประกอบเชิงยืนยันอยู่ในเกณฑ์ดีและโมเดลมีความสอดคล้องกับข้อมูลเชิงประจักษ์

จากการศึกษางานวิจัยที่เกี่ยวข้องการตรวจสอบความตรงเชิงโครงสร้าง (Confirmatory Factory Analysis: CFA) ด้วยการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับที่สอง เป็นวิธีที่ได้รับความนิยมเป็นอย่างมากในการวิเคราะห์ข้อมูลทีมีความสอดคล้องกับข้อมูลเชิงประจักษ์

ตอนที่ 4 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) และงานวิจัยที่เกี่ยวข้อง

1. ความเป็นมาของการทำหน้าที่ต่างกันของข้อสอบ

การศึกษาถึงคุณภาพของข้อสอบจากผลการตรวจข้อสอบของผู้สอบกลุ่มต่างๆ ในประชากร มีมานานแล้ว แต่การศึกษาคุณภาพด้านความยุติธรรมของข้อสอบหรือแบบสอบระหว่างผู้สอบกลุ่มต่างๆ เริ่มศึกษากันอย่างจริงจังในช่วงปลายทศวรรษของ ปี ค.ศ. 1960 มีการเสนอวิธีการต่างๆ เพื่อตรวจสอบความลำเอียงของข้อสอบ (Item Bias) ความลำเอียงของแบบสอบ (Test Bias) และความลำเอียงในการคัดเลือก (Selection Bias) โดยในความลำเอียงว่าเป็น ความคลาดเคลื่อนอย่างเป็นระบบ (Systematic Bias) ที่เกิดจากการวัด ความพยายามของการตรวจสอบความลำเอียงดังกล่าวดำเนินไปเพื่อจำแนกข้อสอบที่ทำหน้าที่ไม่เหมาะสมหรือความไม่ยุติธรรมสำหรับปรับปรุงหรือตัดข้อสอบข้อนั้นออกจากแบบสอบ เป็นการขจัดข้อสอบที่ทำให้เกิดปัญหา ความยุติธรรมระหว่างกลุ่มต่างๆที่มีลักษณะบางอย่าง แตกต่างกัน เช่น เชื้อชาติ ศาสนา วัฒนธรรม ภูมิลำเนา สังคม เพศ

ภาษา อายุ และประสบการณ์ เป็นต้น เพื่อพัฒนาแบบสอบให้มีคุณภาพเหมาะสมสำหรับการนำไปใช้ต่อไป (ศิริชัย กาญจนวาสี, 2555, หน้า 115)

Holland and Thayer, 1988, pp 129-145; Holland and Wainer, 1993; ศิริชัย กาญจนวาสี, 2555, หน้า 116) ได้เสนอแนวคิดเกี่ยวกับการใช้คำและความหมาย มีประเด็นโต้แย้งกันว่าความลำเอียงของข้อสอบเป็นผลการตัดสินว่าข้อสอบมีความยุติธรรมหรือไม่ อันส่งผลต่อการบรรลุจุดมุ่งหมายของการใช้แบบสอบหรือความลำเอียงของข้อสอบ เป็นสารสนเทศทางสถิติที่ได้จากข้อสอบเกี่ยวกับความสัมพันธ์ระหว่างคุณลักษณะที่ข้อสอบมุ่งวัดกับประสบการของผู้สอบกลุ่มต่างๆ ที่ทำการสอบ เมื่อกลุ่มผู้สอบต่างกลุ่มกับตอบสนองข้อสอบเดียวกัน ความต่างที่เกิดขึ้นอาจมาจากความไม่เหมาะสมของข้อคำถาม ซึ่งสามารถเกิดขึ้นได้หลายลักษณะ หรือประสบการณ์ของผู้สอบซึ่งอาจมีลักษณะพื้นฐานเดิม แตกต่างกันในหลายสถานการณ์จึงไม่เหมาะสมที่จะใช้คำว่า ข้อสอบลำเอียง (Biased Item) เนื่องจากเป็นภาษาที่มีความหมายในเชิงลบ ประกอบกับเกณฑ์ที่ใช้สำหรับตัดสินความลำเอียงยังมีความคลุมเครือและค่อนข้างสับสน ดังนั้นจึงควรเปลี่ยนมาใช้คำว่า การทำหน้าที่ต่างกันของข้อสอบ (Differential Item Function: DIF) ซึ่งเป็นคำที่มีความหมายกลางและเหมาะสมกว่า

การทำหน้าที่ต่างกันของข้อสอบ (DIF) กับความลำเอียงของข้อสอบ (Item Bias) มีแนวคิดที่แตกต่างกัน สำหรับการทำหน้าที่ต่างกันของข้อยสอบ เป็นกระบวนการที่เน้นการใช้วิธีการทางสถิติสำหรับการตรวจสอบ เพื่อให้ได้สารสนเทศเกี่ยวกับการทำหน้าที่ต่างกันของข้อสอบสำหรับผู้สอบกลุ่มย่อยที่มีลักษณะบางอย่างแตกต่างกัน ส่วนความลำเอียงของข้อสอบ เป็นกระบวนการตัดสินความยุติธรรมของข้อสอบ โดยนำสารสนเทศการทำหน้าที่ต่างกันของข้อสอบมาวิเคราะห์เชิงตรรกะ (Logical Analysis) โดยผู้เชี่ยวชาญพิจารณาถึงการเขียนข้อสอบ เนื้อหาสาระของข้อสอบและจุดมุ่งหมายของการวัด เพื่อระบุว่าข้อสอบนั้นลำเอียงเข้าข้างกลุ่มใดหรือไม่ เพราะเหตุใดจึงเป็นการตัดสินความลำเอียงของข้อสอบ (Camilli & Shapard, 1994; ศิริชัย กาญจนวาสี, 2555, หน้า 116)

2. ความหมายการทำหน้าที่ต่างกันของข้อสอบ

นักวิจัยทางการวัดผลหลายท่านได้ให้ความหมายของการทำหน้าที่ต่างกันของข้อสอบไว้ดังนี้

Camilli and Shapard (1994) ได้ให้ความหมายของการทำหน้าที่ต่างกันของข้อสอบว่า หมายถึง ความเป็นพหุมิติในการวัดที่ได้จาก การแจกแจงความสามารถหลัก (Primary Ability) ของกลุ่มผู้สอบตั้งแต่ 2 กลุ่มขึ้นไปมีความเท่าเทียมกันแต่มีการแจกแจงความสามารถรอง (Secondary Ability) ที่แตกต่างกัน

Narayanan and Swaminathan (1996) ได้ให้ความหมายของการทำหน้าที่ต่างกันของข้อสอบว่า หมายถึง ฟังก์ชันการตอบสนองข้อสอบที่คำนวณได้จากกลุ่มผู้สอบกลุ่มย่อยที่ต่างกัน มีค่าแตกต่างกัน

ศิริชัย กาญจนวาสี (2555) ได้ให้ความหมายของการทำหน้าที่ต่างกันของข้อสอบว่า หมายถึง การที่ข้อสอบทำให้ผู้สอบจากต่างกลุ่มกันที่มีความสามารถหรือคุณลักษณะที่มุ่งวัดเท่ากันมีโอกาสในการตอบข้อสอบได้ถูกต้องแตกต่างกัน หรือมีฟังก์ชันการตอบสนองแตกต่างกัน

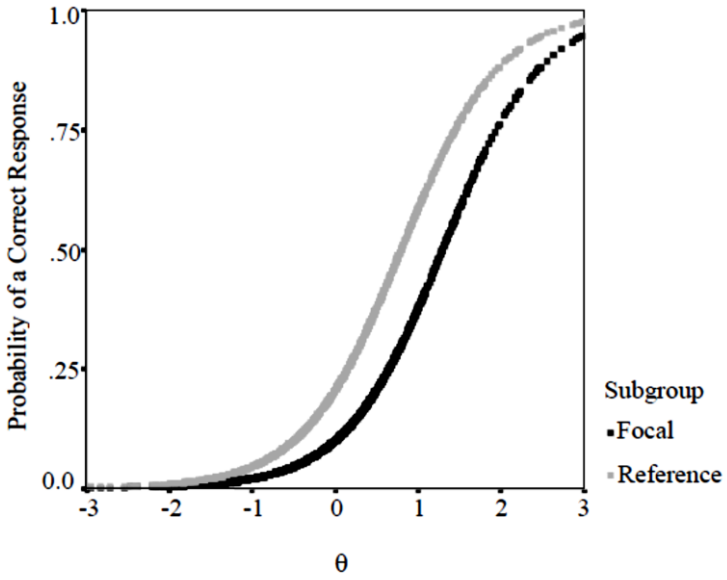
การทำหน้าที่ต่างกันของข้อสอบ เกิดขึ้นเมื่อนำข้อสอบไปทดสอบกับผู้สอบกลุ่มย่อยต่างกัน ที่มีความสามารถหลัก (Primary Ability) ระดับเดียวกันหรือคุณลักษณะแฝง (Latent Trait) ที่ต้องการวัดเท่ากัน แต่มีความสามารถรอง (Secondary Ability) แตกต่างกัน ทำให้ผู้สอบต่างกลุ่มกันที่นำมาจับคู่เปรียบเทียบมีโอกาสตอบสนองข้อสอบถูกแตกต่างกัน

ดังนั้นจึงสรุปได้ว่า การทำหน้าที่ต่างกันของข้อสอบ (DIF) หมายถึง ข้อสอบที่ทำให้ผู้ตอบที่มีคุณลักษณะเหมือนกัน แต่ลักษณะของกลุ่มย่อยแตกต่างกัน เช่น เพศ เชื้อชาติ ศาสนา วัฒนธรรม สภาพภูมิศาสตร์ และเศรษฐกิจ เป็นต้น มีโอกาสที่จะตอบข้อสอบได้ถูกต้องแตกต่างกัน เกิดการได้เปรียบเสียเปรียบกันในระหว่างกลุ่มผู้สอบ ทำให้ต้องตรวจสอบการทำหน้าที่ต่างกันของข้อสอบเพื่อพัฒนาประสิทธิภาพตามโครงสร้างและมาตรฐาน

3. ประเภทของการทำหน้าที่ต่างกันของข้อสอบ

Mellenbergh (1982) พบว่าข้อสอบสามารถทำหน้าที่แตกต่างกันแบ่งเป็น 2 ประเภท ดังนี้

3.1 ข้อสอบทำหน้าที่ต่างกันแบบเอกรูป (Uniform DIF) หมายถึง ข้อสอบที่ทำให้ผู้สอบกลุ่มหนึ่งมีโอกาสในการตอบข้อสอบถูกมากกว่าผู้สอบอีกกลุ่มหนึ่งอย่างสม่ำเสมอกันในทุกระดับความสามารถเมื่อพิจารณาโค้งคุณลักษณะข้อสอบของผู้สอบ 2 กลุ่ม จะพบว่าไม่มีปฏิสัมพันธ์ระหว่างระดับความสามารถของผู้สอบกับการเป็นสมาชิกของกลุ่ม (Group Membership) ดังภาพที่ 2-13



ภาพที่ 2-13 ข้อสอบทำหน้าที่ต่างกันแบบเอกรูป (Uniform DIF) (Mellenbergh, 1982; ศิริชัย กาญจนวาสี, 2555, หน้า 118)

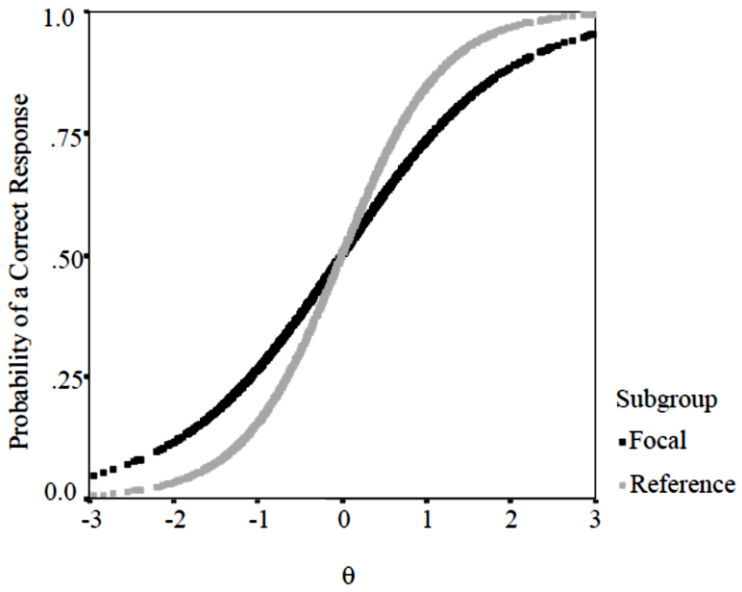
ตามทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT) สามารถพิจารณา "ปฏิสัมพันธ์" ดังกล่าวได้จากความแตกต่างระหว่างของค่าพารามิเตอร์อำนาจจำแนกของข้อสอบระหว่างผู้สอบกลุ่มย่อยสองกลุ่ม กล่าวคือ ถ้าข้อสอบทำหน้าที่ต่างกันแบบเอกรูป แล้วโค้งลักษณะ

ข้อสอบ (Item Characteristic Curves: ICCs) ระหว่างผู้สอบกลุ่มย่อยสองกลุ่มจะขนานกันหรือมีฟังก์ชันการตอบสนองข้อสอบ (Item Response Functions: IRFs) เหมือนกัน แต่ถ้าข้อสอบทำหน้าที่ต่างกันแบบอเนกรูปแล้วโค้งลักษณะข้อสอบระหว่างผู้สอบกลุ่มย่อยสองกลุ่มจะไม่ขนานกันหรือมีฟังก์ชันการตอบสนองข้อสอบ ต่างกันดังนั้นความแตกต่างระหว่างโค้งลักษณะข้อสอบทั้งสองแบบจะบ่งบอกถึงขนาดและทิศทางของข้อสอบที่ทำหน้าที่ต่างกันซึ่งสามารถคำนวณได้โดยใช้สูตรการคำนวณพื้นที่ของ Raju (1990)

3.2 ข้อสอบทำหน้าที่ต่างกันแบบอเนกรูป (Nonuniform DIF) หมายถึง ข้อสอบที่ทำโอกาสในการตอบถูกต้องของผู้สอบระหว่างกลุ่มแตกต่างกันอย่างไม่สม่ำเสมอกันในทุกระดับความสามารถ เพื่อพิจารณาโค้งคุณลักษณะข้อสอบของผู้สอบ 2 กลุ่ม พบว่ามีปฏิสัมพันธ์ร่วมกันระหว่างระดับความสามารถของผู้สอบ กับการเป็นสมาชิกของกลุ่ม เช่นที่ระดับความสามารถหนึ่งกลุ่มผู้สอบกลุ่ม R มีโอกาสในการตอบข้อสอบถูกมากกว่ากลุ่มผู้สอบกลุ่ม F แต่ที่ระดับความสามารถอีกระดับหนึ่งกลุ่มผู้สอบกลุ่ม F มีโอกาสในการตอบข้อสอบถูกมากกว่ากลุ่มผู้สอบกลุ่ม R

Swaminathan and Rogers (1990) ได้จำแนกข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูปเป็น 2 ลักษณะ ดังนี้

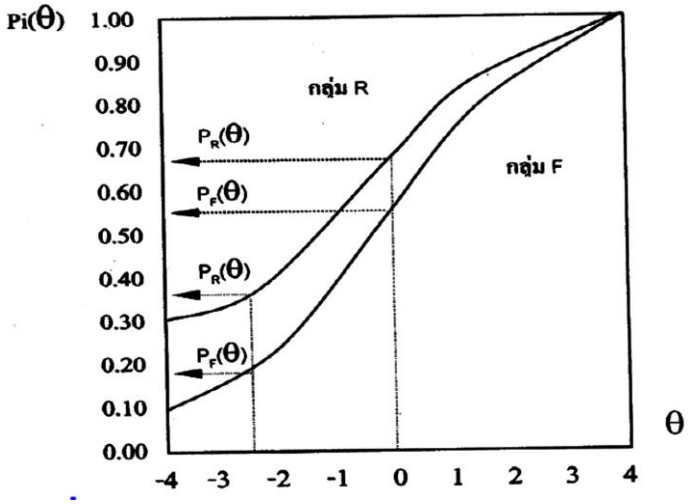
3.2.1 ข้อสอบทำหน้าที่ต่างกันแบบอเนกรูปโดยมีปฏิสัมพันธ์ไม่เป็นลำดับ (Disordinal Interaction) เป็นการทำหน้าที่ต่างกันสำหรับกลุ่มผู้สอบซึ่งเกิดขึ้น เมื่อโค้งลักษณะข้อสอบตัดกันระหว่างช่วงความสามารถของผู้สอบหรือเรียกว่าข้อสอบทำหน้าที่ต่างกันแบบไม่มีทิศทาง (Non-Directional DIF) ดังภาพที่ 2-14



ภาพที่ 2-14 ข้อสอบทำหน้าที่ต่างกันแบบไม่มีทิศทาง (Non - Directional DIF)
(Mellenbergh, 1982; ศิริชัย กาญจนวาสี, 2555, หน้า 119)

3.2.2 ข้อสอบทำหน้าที่ต่างกันแบบอเนกรูป โดยมีการปฏิสัมพันธ์เป็นลำดับ (Ordinal Interaction) เป็นการทำหน้าที่ต่างกัน สำหรับกลุ่มผู้สอบซึ่งเกิดขึ้น เมื่อโค้งลักษณะ

ข้อสอบต่างกันอย่างไม่สม่ำเสมอ แต่ไม่ตัดกัน หรืออาจตัดกันนอกช่วง ความสามารถของผู้สอบตรงปลายสุดของช่วงความสามารถต่ำหรือสูง อาจเรียกข้อสอบลักษณะนี้ว่าข้อสอบทำหน้าที่ต่างกันแบบมีทิศทางเดียว (Unidirectional DIF) ดังภาพที่ 2-15



ภาพที่ 2-15 ข้อสอบทำหน้าที่ต่างกันแบบมีทิศทางเดียวกัน (Unidirectional DIF)
(Mellenbergh, 1982; ศิริชัย กาญจนวาสี, 2555, หน้า 120)

4. การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

การตรวจการทำหน้าที่ต่างกันของข้อสอบ (DIF Detection) เป็นการเปรียบเทียบผลการตอบสนองของข้อสอบเป็นรายข้อระหว่างกลุ่มผู้สอบอย่างน้อย 2 กลุ่ม ที่มีความสามารถหลัก (Primary Ability) ที่มุ่งวัดเท่ากัน แต่คาดว่าจะไม่มีความได้เปรียบหรือเสียเปรียบกัน โดยกลุ่มหนึ่งถือเป็น กลุ่มอ้างอิง (Reference Group) ซึ่งคาดว่าน่าจะได้เปรียบในการตอบข้อสอบข้อนั้น หรือมีโอกาสตอบข้อสอบได้ถูกต้องมากกว่า ส่วนอีกกลุ่มคือ กลุ่มเปรียบเทียบ (Focal Group) ซึ่งเป็นกลุ่มสนใจศึกษาและค่าว่าน่าจะได้เป็นกลุ่มที่เสียเปรียบ (ศิริชัย กาญจนวาสี, 2555, หน้า 120)

ในการเปรียบเทียบผลการตอบข้อสอบระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบจำเป็นต้องจับคู่ (Matching) ผู้สอบตามความสามารถซึ่งเป็นเงื่อนไขสำคัญของการตรวจการทำหน้าที่ต่างกันของข้อสอบ เกณฑ์การจับคู่ (Matching Criteria) ที่นิยมใช้กันมี 2 วิธีดังนี้

4.1 เกณฑ์ภายนอก (External Criterion)

การวิเคราะห์การทำหน้าที่ต่างกัน โดยใช้เกณฑ์ภายนอกนี้สามารถนำไปใช้ได้ทั้งข้อสอบรายข้อแบบสอบทั้งฉบับ โดยการใช้คะแนนจากแบบสอบอื่นเป็นเกณฑ์ภายนอกแล้วใช้เทคนิคการการถดถอย (Regression Analysis) เพื่อทำการเปรียบเทียบเส้นกราฟความสัมพันธ์ระหว่างตัวแปรเกณฑ์กับตัวแปรทำนายระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบหลักการและจุดมุ่งหมายนี้เพื่อสร้างสมการทำนายตัวแปรเกณฑ์ ซึ่งเป็นคะแนนของแบบสอบอื่นจากตัวแปรทำนายซึ่งเป็นคะแนนรายข้อ หรือคะแนนแบบสอบระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบในการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ จะใช้คะแนนรายข้อเป็นตัวแปรทำนาย แต่ถ้าเป็นการวิเคราะห์การทำหน้าที่

ต่างกันของแบบทดสอบจะใช้คะแนนรวมของแบบสอบทั้งสองฉบับเป็นตัวแปรทำนาย สำหรับตัวแปรเกณฑ์ที่ใช้เป็นเกณฑ์ในการภายนอก อาจใช้คะแนนรวมทั้งฉบับ หรือเกรดเฉลี่ย หรือผลสัมฤทธิ์ในงานที่เกี่ยวข้องของผู้สอบ (Cronbach & Furby, 1970) สามารถทำนาย สำหรับกลุ่มอ้างอิงและกลุ่มเปรียบเทียบแสดงผลได้ดังนี้

กลุ่มอ้างอิง (R) $Yi = AR + BR\,Xi$ (5)

กลุ่มเปรียบเทียบ (F) $YF = AF + BF\,Xi$ (6)

เมื่อ Yi = คะแนนของตัวแปรเกณฑ์

Xi = คะแนนของตัวแปรทำนาย

A = ค่าคงที่หรือค่าตัดแกน (Intercept)

B = ค่าความชัน (Slope)

จากกราฟของฟังก์ชันการทำนายทั้ง 2 สมการ สามารถเปรียบเทียบค่าตัดแกน (A) และค่าความชันของ (B) ของเส้นกราฟระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบได้ ถ้าเส้นกราฟดังกล่าวมีค่าความชัน หรือค่าตัดแกนแตกต่างกัน สำหรับข้อสอบใดหรือแบบทดสอบใดแสดงว่าข้อสอบหรือแบบสอบนั้นมีการทำหน้าที่ต่างกัน โดยเข้าข้างกลุ่มผู้สอบที่มีค่าตัดแกนหรือค่าความชันสูงกว่า การใช้เกณฑ์ภายนอกมีข้อดี คือเกณฑ์ที่ใช้มีความเป็นอิสระจากข้อสอบ และแบบสอบที่ดีต้องมีการตรวจสอบ แต่มีจุดอ่อนตรงที่ความเหมาะสมของเกณฑ์ที่จำนำมาใช้ ในทางปฏิบัติเป็นการยากที่จะหาตัวแปรเกณฑ์ภายอกจากแบบทดสอบฉบับอื่นที่มีความตรงเชิงทำนายและมีความยุติธรรมสำหรับกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ ถ้าตัวแปรเกณฑ์ภายนอกขาดคุณสมบัติดังกล่าวจะทำให้ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบหรือแบบทดสอบขาดความแม่นยำและความสมบูรณ์

4.2 เกณฑ์ภายใน (Internal Criterion)

การวิเคราะห์การทำหน้าที่ต่างกัน โดยใช้เกณฑ์ภายในเป็นการนำวิธีทางสถิติมาตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ หรือแบบทดสอบ หรือแบบสอบ โดยเน้นการพิจารณาจากโครงสร้างภายในของแบบทดสอบหลักด้วยการวิเคราะห์ผลการตอบข้อสอบและความสามารถหรือคะแนนจริงของผู้สอบจากแบบทดสอบฉบับนั้น เพื่อนำมาเปรียบเทียบระหว่างผู้สอบจากกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ ที่มีความสามารถหรือคะแนนจริงเท่ากัน ว่าจะมีผลการตอบหรือโอกาสในการตอบข้อสอบได้ถูกต้องแตกต่างกันหรือไม่ เพื่อบ่งชี้ถึงการทำหน้าที่ต่างกันของข้อสอบ ค่าสถิติทดสอบที่นิยมนำมาพอสรุปได้ดังนี้

4.2.1 การทดสอบปฏิสัมพันธ์ (Interaction)

ในระยะเริ่มแรกของการศึกษาความลำเอียงของข้อสอบ มีการใช้สถิติทดสอบเอฟ (F – test) จากการวิเคราะห์ความแปรปรวน (ANOVA) เพื่อทดสอบปฏิสัมพันธ์ระหว่างกลุ่มผู้สอบกับข้อสอบ ถ้าการทดสอบมีนัยสำคัญเป็นสัญญาณของการทำหน้าที่ต่างกันของข้อสอบ (Cleary & Hiltion, 1968; Jensen, 1974) จากนั้นจึงทำการวิเคราะห์ต่อด้วยวิธีการ Post Hoc เพื่อระบุข้อสอบที่มีผลต่อการปฏิสัมพันธ์ ซึ่งเป็นข้อที่ทำหน้าที่ต่างกัน

วิธีการนี้มีข้อดีที่สามารถศึกษาผู้ตอบหลายๆ กลุ่มได้สะดวก แต่มีจุดอ่อนในเรื่องการควบคุมกลุ่มต่างๆ ให้มีความสามารถที่ทัดเทียมกัน ขนาดกลุ่มตัวอย่างของกลุ่มต่างๆ และอัตราการคลาดเคลื่อนประเภทที่ 1 จะสูงขึ้น ถ้าจำนวนข้อสอบเพิ่มมากขึ้น

4.2.2 การวัดความเบี่ยงเบนสัมพัทธ์ (Relative Deviation)

การคำนวณค่าความยากของข้อสอบ เช่น P และ b เป็นต้น เมื่อคำนวณแยกระหว่างกลุ่มและแปลงให้เป็นค่าความยากมาตรฐานสามารถนำมาพล๊อตเปรียบเทียบรายข้อ ถ้าข้อใดเบี่ยงเบนไปจากแกนหลักที่คาดหมาย หรือเบี่ยงเบนเกินจากความคลาดเคลื่อนมาตรฐานของค่าความยากที่กำหนด ย่อมแสดงถึงการทำหน้าที่ต่างกันของข้อสอบ (Cleary & Hilton, 1968; Angoff & Ford, 1973) รวมทั้งสามารถคำนวณค่าสหสัมพันธ์ระหว่างค่าความยากง่ายรายข้อระหว่างกลุ่มเพื่อแสดงถึงการทำหน้าที่ต่างกันของข้อสอบ ถ้าสหสัมพันธ์เข้าใกล้ 1.00 แสดงว่าค่าความยากสัมพัทธ์ของข้อสอบมีค่าใกล้เคียงกันระหว่างกลุ่มดังนั้นแบบสอบวัดคุณลักษณะคล้ายกันระหว่างกลุ่ม วิธีการนี้มีข้อดีและข้อเสียคล้ายการทดสอบปฏิสัมพันธ์ นอกจากนี้ค่าความยากง่ายของข้อสอบ (P) มิใช่ตัวแทนของค่าความจริงของข้อสอบ และได้รับอิทธิพลจากค่าแทรกซ้อนอื่น ได้แก่ ค่าอำนาจจำแนกและความสามารถของผู้สอบ

4.2.3 การเปรียบเทียบน้ำหนักตัวประกอบ (Factor Analysis)

เป็นเทคนิคทางสถิติที่นิยมใช้ในการตรวจสอบความตรงเชิงทฤษฎีหรือโครงสร้าง (Construct Validity) เมื่อนำมาวิเคราะห์ตัวประกอบมาใช้ในการวิเคราะห์โครงสร้างแบบทดสอบแยกตามกลุ่มผู้สอบ ความไม่สอดคล้องกันระหว่างน้ำหนักตัวประกอบบนคุณลักษณะสำคัญที่มุ่งวัด หรือความแตกต่างของค่าเฉลี่ยคะแนนตัวประกอบ (Factor Score) ระหว่างกลุ่มผู้สอบ ย่อมสะท้อนการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ

การใช้เทคนิคการวิเคราะห์ตัวประกอบเชิงสำรวจ (Exploratory Factor Analysis: EFA) สำหรับศึกษาการทำหน้าที่ต่างกัน จะมีจุดอ่อนในเรื่องความไม่สอดคล้อง ระหว่างน้ำหนักตัวประกอบอาจเกิดจากความแตกต่างของความสามารถกลุ่มก็ได้ แนวทางที่เหมาะสมควรใช้เทคนิคการวิเคราะห์ตัวประกอบเชิงยืนยัน (Confirmatory Factor Analysis: CFA) นอกจากนี้ยังสมารถใช้ CFA สำหรับตรวจสอบความแตกต่างระหว่างกลุ่ม ในด้านคุณลักษณะหรือความสามารถรองได้อีกด้วย (Camili & Shepard, 1994)

4.2.4 การเปรียบเทียบโอกาสตอบข้อสอบถูก

การวิเคราะห์โอกาสตอบข้อสอบถูกของผู้สอบจากกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่มีความสามารถเท่ากัน เป็นแนวทางสำคัญที่นิยมใช้กันและเป็นที่ยอมรับกันในปัจจุบัน สำหรับบ่งชี้การทำหน้าที่ต่างกันของข้อสอบ มีการคำนวณค่าสถิติ 2 แนวทางดังนี้ เปรียบเทียบค่าสัดส่วนหรือความน่าจะเป็นในการตอบข้อสอบถูกของผู้สอบต่างกลุ่มที่มีระดับความสามารถเท่ากัน เช่น วิธีแมนเทล–แฮนส์เซล (HM) เป็นต้น เปรียบเทียบค่าฟังก์ชันการตอบสนองข้อสอบ หรือโค้งลักษณะข้อสอบระหว่างกลุ่มที่มีระดับความสามารถเท่ากัน เป็นวิธีที่อยู่บนพื้นฐานของทฤษฎี IRT เช่นวิธีวัดความแตกต่างของพื้นที่ วิธีวัดความแตกต่างของค่าพารามิเตอร์ความยาก วิธีทดสอบไค–สแควร์ของลอร์ด (Lords' x^2 - test) เป็นต้น

วิธีการนี้มีข้อดีที่สำคัญได้แก่ การคำนวณค่าทางสถิติของข้อสอบมีความน่าเชื่อถือมีกลไกการควบคุมความสามารถเท่ากัน จึงเป็นวิธีการยอมรับกันทั่วไป แต่มีข้อจำกัดในด้านความความสลับซับซ้อนของแนวคิดพื้นฐาน และการวิเคราะห์มีความจำเป็นต้องใช้โปรแกรมคอมพิวเตอร์โดยเฉพาะ

5. วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

Potenza and Doran (1995) ได้นำเสนอวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF Detection) จำแนกตามลักษณะการตรวจให้คะแนนได้เป็น 2 ประเภทคือ ข้อสอบที่มีการให้คะแนนแบบบททวิภาค หรือสองค่า (Dichotomous Scoring) และข้อสอบที่มีการให้คะแนนแบบพหุวิภาค หรือหลายค่า (Polytomous Scoring) วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแต่ละประเภท ยังสามารถจำแนกได้อีก 2 มิติได้แก่ มิติลักษณะของตัวแปรเกณฑ์ ซึ่งแบ่งเป็นกลุ่มวิธีที่ใช้คะแนนสังเกตได้ (Observed Score) และกลุ่มวิธีที่ใช้คะแนนสังเกตไม่ได้หรือคะแนนของตัวแปรแฝง (Latent Variable) และมิติลักษณะของสถิติวิเคราะห์ ซึ่งบ่างเป็นกลุ่มวิธีที่ใช้สถิติพาราเมตริก (Parametric Approach) และกลุ่มวิธีที่ใช้สถิตินันพาราเมตริก (Nonparametric Approach) รายชื่อวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่สำคัญ ดังแสดงในตารางที่ 2-3

ตารางที่ 2-3 วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการตรวจให้คะแนนแบบทวิภาค (Dichotomous DIF) และพหุวิภาค (Polytomous DIF) (ศิริชัย กาญจนวาศรี, 2555, หน้า 124)

ประเภทตัวแปรเกณฑ์	พาราเมตริก	นันพาราเมตริก
1. DIF แบบทวิภาค		
1.1 คะแนนที่สังเกตุได้	ANOVA	TID (Transformed Item Difficulty)
(Observed Score)	LR (Logistic Regression)	MH (Mantel – Haenzel)
		STND (Standardization)
1.2 คุณลักษณะ/ตัวแปรแฝง	IRT-D2	SIBTEST
(Latent Variable)	Lord's x^2	
	General IRTLR	
	Loglinear IRTLR	
2 DIF แบบพหุวิภาค		
2.1 คะแนนที่สังเกตได้	ANOVA	
(Observed Score)	Polytomous Logistic	Polytomous STND
	Regression	GHM (General Mantel – Haenzel)
2.2 คุณลักษณะ/ตัวแปรแฝง	General IRTLR	Polytomous SIBTEST
	PCM (Partial Credit	GPCM (Generalized
(Latent Variable)	Model)	partial Credit Model)

5.1 วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ให้คะแนนแบบทวิภาค

5.1.1 กลุ่มวิธีที่ใช้คะแนนที่สังเกตได้

วิธีในกลุ่มนี้มักวิเคราะห์ตามทฤษฎีการทดสอบแบบดั้งเดิม (CTT) หรือกลุ่มที่ไม่ใช้ทฤษฎีการตอบสนองข้อสอบ (Non – IRT Approach) โดยใช้คะแนนรวมของผู้สอบเป็นเกณฑ์การจับคู่ของกลุ่มผู้สอบ วิธีการตรวจสอบที่สำคัญในกลุ่มนี้ ได้แก่

5.1.1.1 การวิเคราะห์ความแปรปรวน (ANOVA) (Clearyn & Hiltion, 1968)

5.1.1.2 วิธีการวิเคราะห์การถดถอยโลจิสติก (Logistic Regression: LR) (Swaminathan & Rogers, 1990)

5.1.1.3 วิธีแปลงค่าความยากของข้อสอบ (Transformed Item Difficulty: TID) (Cleary & Hilton, 1968; Angoff & Ford, 1973)

5.1.1.4 วิธีแมนเทล – แฮนส์เซล (Mantel – Haenzel: MH) (Holland & Thayer, 1988)

5.1.1.5 วิธีวัดดัชนีมาตรฐาน (Standardization: STND) การปรับให้เป็นมาตรฐานด้วยน้ำหนักประกอบ (Dorans & Kulick, 1986)

5.1.2 กลุ่มที่ใช้คุณลักษณะแฝง

วิธีในกลุ่มนี้ใช้คุณลักษณะหรือตัวแปรแฝง ซึ่งวิเคราะห์บนพื้นฐานของทฤษฎีการตอบสนองข้อสอบ (IRT) สำหรับเกณฑ์การจุบคู่กลุ่มผู้สอบ วิธีการตรวจสอบที่สำคัญในกลุ่มนี้ได้แก่

5.1.2.1 วิธีวัดพื้นที่ความแตกต่างระหว่างโค้งการตอบสนอง (IRT –D2) (Linn et al, 1981; Shepard et al., 1984; Raju, 1990; Kin& Cohen, 1991)

5.1.2.2 วิธีไค – สแควร์ของลอร์ด์ (Lord's $\boldsymbol{\chi}^2$) (Lord,1980)

5.1.2.3 วิธีอัตราส่วนไลค์ลิฮูดทั่วไป (Gerneral IRT Likelihood Ratio) (Thissen, Steinberg & Wainer, 1993)

5.1.2.4 วิธีซิปเทสท์ (SIBTEST) (Shealy & Stout, 1993)

5.2 วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ให้คะแนนแบบพหุวิภาค

5.2.1 กลุ่มวิธีใช้คะแนนที่ สังเกตได้

5.2.1.1 วิธีการวิเคราะห์ความแปรปรวน (ANOVA) (Cleary & Hilton, 1968)

5.2.1.2 วิธีการวิเคราะห์การถดถอยโลจิสติกพหุวิภาค (Polytomous Logistic Regression) (Swaminathan & Rogers, 1990)

5.2.1.3 วิธีแมนเทล – แฮนส์เซลทั่วไป (General Mantel – Haenzel: GMH) (Holland & Thayer, 1988)

5.2.2 กลุ่มวิธีที่ใช้คุณลักษณะแฝง

5.2.2.1 วิธีอัตราส่วนไลค์ลิฮูดในรูปทั่วไป (Gerneral IRT Likelihood Ratio) (Thissen, Steinberg, & wainer, 1993)

5.2.2.2 วิธีการให้คะแนนบางส่วน (Partial Credit Model: PCM) (Master, 1982)

5.2.2.3 วิธีซิปเทสท์พหุวิภาค (Polytomous SIBTEST) (Shealy & Stout, 1993)

5.2.2.4 วิธีการให้คะแนนบางส่วนทั่วไป (Generalized Partial Credit Model; GPCM) (Muraki, 1992)

งานวิจัยที่เกี่ยวข้องกับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) มีดังนี้

ชัยวัฒน์ หฤทัยพันธ์ (2558) ได้ศึกษาพัฒนาวิธีการสำหรับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยผู้เชี่ยวชาญและเปรียบเทียบประสิทธิภาพการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้านอัตราความถูกต้องและอัตราความคลาดเคลื่อนการทดลองใช้การวิธีวิเคราะห์ข้อสอบโดยผู้เชี่ยวชาญ วิธีการประยุกต์ใช้เทคนิคการประชุมแบบเดลฟายจากกลุ่มผู้เชี่ยวชาญและวิธีการประยุกต์ใช้เทคนิคโปรโตคอลอะลาวด์ กลุ่มตัวอย่างคือ ผู้เชี่ยวชาญจำนวน 21 คนและนักเรียนชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2556 จำนวน 139 คน ใช้วิธีการเลือกกลุ่มตัวอย่างแบบเจาะจง

เครื่องมือที่ใช้ประกอบด้วยแบบวินิจฉัยและแบบยืนยันการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบจากผู้เชี่ยวชาญ แบบทดสอบสุขศึกษาและพละศึกษาที่ผ่านการตรวจสอบจากผู้เชี่ยวชาญและตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีแมนเทล – แฮนส์เซล ด้วยโปรแกรม DDFS 1.0 และโปรแกรม DIFAS 5.0 พบว่า วิธีแบบวินิจฉัยโดยผู้เชี่ยวชาญมีอัตราความถูกต้องคิดเป็นร้อยละ 50 และมีอัตราความคลาดเคลื่อนคิดเป็นร้อยละ 50 วิธีการประยุกต์ใช้เทคนิคการประชุมแบบเดลฟายจากกลุ่มผู้เชี่ยวชาญ ตามฉันทามติจากกลุ่มผู้เชี่ยวชาญคิดเป็นร้อยละ 0 และมีอัตราความคลาดเคลื่อนคิดเป็นร้อยละ 100 วิธีวิธีการประยุกต์ใช้เทคนิคโปรโตคอลอะลาวด์ มีอัตราความคลาดเคลื่อนคิดเป็นร้อยละ 25 และมีอัตราความคลาดเคลื่อนคิดเป็นร้อยละ 75

พิชชา สุริอาจ และประกฤติยา ทักษิโณ (2559) ได้พัฒนาแบบวัดความตระหนักต่อโลกในยุคศตวรรษที่ 21 ของนักเรียนมัธยมศึกษาตอนต้น โดยใช้แบบวัดเชิงสถานการณ์ กลุ่มตัวอย่างคือ นักเรียนชั้นมัธยมศึกษาปีที่ 1-3 สังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน จังหวัดนครราชสีมา จำนวน 1,200 คน ได้มาจากการสุ่มแบบหลายขั้นตอน เครื่องมือที่ใช้ในการวิจัย คือ แบบวัดความตระหนักต่อโลกในยุคศตวรรษที่ 21 วิเคราะห์ค่าอำนาจจำแนกของข้อสอบตามทฤษฎีแบบดั้งเดิมวิเคราะห์ค่าความเที่ยงโดยใช้โปรแกรม SPSS for windows วิเคราะห์ค่าอำนาจจำแนกตามทฤษฎีการตอบสนองข้อสอบโดยใช้โปรแกรม Multilog วิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธี Polytomous SIBTEST โดยใช้โปรแกรม DIFPACK 1.7 วิเคราะห์องค์ประกอบเชิงยืนยันอันดับที่สองโดยใช้โปรแกรม Mplus จากการศึกษาพบว่า ค่าอำนาจจำแนกของข้อสอบตามทฤษฎีแบบดั้งเดิม มีค่าระหว่าง 0.20 ถึง 0.81 ค่าอำนาจจำแนกตามทฤษฎีการตอบสนองข้อสอบมีค่าระหว่าง 0.15-3.22 ค่าความเที่ยง เท่ากับ 0.08 การทำหน้าที่ต่างกันของข้อสอบ พบ DIF จำนวน 3 ข้อ และการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับที่สองหลังการปรับโครงสร้างแบบวัดหลังการตัด DIF มีความสอดคล้องกับข้อมูลเชิงประจักษ์

Chanpleng, Lawthong, and Ngudgratoke (2015) ได้ศึกษาการวิเคราะห์เปรียบเทียบโมเดลประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์: การประยุกต์ใช้โมเดลมูลค่าเพิ่มพหุระดับที่มีการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบ มีวัตถุประสงค์1) เพื่อศึกษาผลการทำหน้าที่ต่างกันของข้อสอบ (DIF) ที่มีการตรวจให้คะแนนทวิวิภาคและพหุวิภาค และการทำหน้าที่ต่างกันของแบบทดสอบ (DTF) ใน แบบทดสอบการประเมินการรู้เรื่องวิทยาศาสตร์ 2) เพื่อเปรียบเทียบประสิทธิภาพของโมเดลประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์ ใช้โมเดลมูลค่าเพิ่มพหุระดับ 2 โมเดล คือ โมเดลที่ไม่ได้พิจารณาการทำหน้าที่ต่างกันของข้อสอบและแบบสอบกับโมเดลที่มีการพิจารณาการทำหน้าที่ต่างกันของข้อสอบและแบบสอบโดยการตัดข้อสอบที่มีการทำหน้าที่ต่างกันของข้อสอบและการทำหน้าที่ต่างกันเป็นขั้นที่แตกต่างกัน ข้อมูลเป็นข้อมูลทุติยภูมิจากโครงการประเมินผลนักเรียนนานาชาติเกี่ยวกับการรู้เรื่องวิทยาศาสตร์ (PISA) ปี ค.ศ. 2009 กลุ่มตัวอย่าง ได้แก่ นักเรียนจำนวน 4,292 คน จากสถานศึกษา 230 แห่ง พบว่า 1) การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ มีข้อสอบที่เข้าข้างนักเรียนหญิง (3 ข้อ), นักเรียนที่เรียนพิเศษวิทยาศาสตร์นอกสถานศึกษา (1 ข้อ), นักเรียนที่มาจากครอบครัวที่มีเศรษฐานะต่ำ (2 ข้อ), นักเรียนที่มาจากครอบครัวที่มีแหล่งทรัพยากรที่บ้านสูง (5 ข้อ), และนักเรียนที่มาจากครอบครัวที่มีความมั่งคั่งสูง (2 ข้อ) ค่าความเที่ยงของแบบทดสอบในโมเดลที่ 2 ลดลงเมื่อตัดข้อสอบที่ทำหน้าที่ต่างกันออก ขนาด

อิทธิพลการทำหน้าที่ต่างกันของแบบสอบ 2 ฉบับ ลดลงหลังจากตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันออกไป ส่วนแบบสอบอีก 7 ฉบับมีขนาดอิทธิพลการทำหน้าที่ต่างกันของแบบสอบไม่เปลี่ยนแปลง และหลังจากตัดข้อสอบบางข้อที่ตรวจพบการทำหน้าที่ต่างกันออกไปในโมเดลที่ 2 พบว่า ผลการวิเคราะห์องค์ประกอบเชิงยืนยันทั้งแบบรวมและแยกรายฉบับมีความตรงเชิงโครงสร้างที่ยอมรับได้

Öğretmen (2015) ได้ศึกษาวิเคราะห์ DIF จำแนกเพศสำหรับความเข้าใจการอ่านภาษาอังกฤษเป็นภาษาต่างประเทศภายใต้ทฤษฎีการตอบสนองข้อสอบ (IRT) ผลการตอบแบบทดสอบรวบรวมจากการวัดผลสัมฤทธิ์ทางการเรียนด้านความเข้าใจการอ่านภาษาอังกฤษกับนักศึกษาระดับมหาวิทยาลัยในประเทศตุรกี กลุ่มตัวอย่างจำนวน 2,117 คน (เพศชาย 1,011 คน และเพศหญิง 1,116 คน) ผลการศึกษาชี้ให้เห็นว่า มีความไม่สม่ำเสมอในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ในการทดสอบความเข้าใจในการอ่านภาษาอังกฤษและมีความสัมพันธ์กันอย่างมีนัยสำคัญ ในข้อสอบทั้งหมด 21 ข้อ เพศหญิงเป็นกลุ่มได้เปรียบ ซึ่งการศึกษาครั้งนี้ ครูสามารถใช้การจัดการเรียนการสอนเชิงรุกเพื่อลดความแตกต่างทางเพศในระดับสูงกับระดับทักษะการคิดที่มีความสำคัญต่อความเข้าใจในการอ่านที่มีประสิทธิภาพ

Özdemir (2015) ได้เปรียบเทียบวิธีทฤษฎีการตอบสนองข้อสอบ (IRT) สำหรับการทำหน้าที่ต่างกันของข้อสอบ ในแบบทดสอบคณิตศาสตร์ TIMSS ปีการศึกษา 2011 ค่าของการทำหน้าที่ต่างกันของข้อสอบ (DIF) ได้จากการทดสอบโดยวิธี Lord’s Chi-Square วิธี Raju’s Area และวิธี Likelihood-Ratio โดยกำหนดให้เพศชายเป็นกลุ่มอ้างอิงและเพศหญิงเป็นกลุ่มเปรียบเทียบ การเปรียบเทียบเหล่านี้สามารถเป็นหลักฐานในการพิจารณาแบบจำลองที่ดีที่สุดสำหรับการตรวจจับการทำหน้าที่ต่างกันของข้อสอบ ผลการวิจัยชี้ให้เห็นว่าวิธีทฤษฎีการตอบสนองข้อสอบ (IRT) แบบ 2PL มีความพอดีกับข้อมูลของทั้ง 3 วิธี แม้ว่าจำนวนการตรวจจับการทำหน้าที่ต่างกันของข้อสอบจะต่างกันพบว่า มีจำนวน 2 ข้อ จากจำนวนข้อสอบ 22 ข้อ มีค่าคงที่ระหว่าง 3 วิธี ซึ่งมีแนวโน้มที่จะตอบได้อย่างถูกต้องโดยเพศชายเป็นกลุ่มได้เปรียบ

Nenty, Kgosidialwa, and Batsalelwang (2017) ได้ศึกษาเพศและความสามารถทางการทดสอบคณิตศาสตร์ JSC ปีการศึกษา 2015 ของนักเรียนในโกโบโรเน ซึ่งมีความแตกต่างทั้งด้านสภาพแวดล้อม ครอบครัวและความคาดหวังระหว่างเพศชายและหญิงจากพ่อแม่ ศึกษาโดยใช้ข้อมูลจากนักเรียน 735 คน (นักเรียนชาย 298 คนชายและหญิง 437 คน) ที่กำลังเรียนอยู่ สอบ JSC ปีการศึกษา 2015 ครั้งสุดท้ายจากโรงเรียนมัธยมศึกษาตอนต้น ที่คัดเลือกแล้ว 5 แห่ง พบว่าเพศมีอิทธิพลการการทดสอบคณิตศาสตร์อย่างมีนัยสำคัญ

Aryadoust (2018) ได้ใช้วิธี Rasch Model เพื่อวิเคราะห์ความเข้าใจในการอ่านภาษาที่สอง กลุ่มตัวอย่างจำนวน 1550 คน ข้อมูลคือความรู้ด้านคำศัพท์และไวยากรณ์และเพศของผู้สอบ ซึ่งค่าความยากของข้อสอบแตกต่างกันอย่างมีนัยสำคัญ การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบพบว่ามีเพียง 204 คน (13.16%) ผู้สอบที่มีคะแนนสูงมาก คะแนนไวยากรณ์ได้รับผลกระทบจากการจำแนกเพศแสดงให้เห็นว่า การทำหน้าที่ต่างกันของข้อสอบเกิดจากตัวแปรที่มีอิทธิพลจากข้อมูลเชิงประจักษ์ เฉพาะกลุ่มย่อยบางกลุ่มของผู้สอบที่มีความสามารถเฉพาะเจาะจงจึงสร้างเครือข่าย

ความสัมพันธ์ที่ซับซ้อน ระหว่างตัวแปรที่เกี่ยวข้องกับตัวแปรที่ไม่เกี่ยวข้อง นวัตกรรมเหล่านี้อำนวยความสะดวกในการสำรวจข้อมูลเชิงประจักษ์และ ช่วยเพิ่มความแม่นยำของผลลัพธ์

จากการศึกษางานวิจัยที่เกี่ยวข้องกับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) พบว่า เป็นกระบวนการที่เน้นการใช้วิธีการทางสถิติสำหรับการตรวจสอบ เพื่อให้ได้สารสนเทศเกี่ยวกับการทำหน้าที่ต่างกันของข้อสอบสำหรับผู้สอบกลุ่มย่อยที่มีลักษณะบางอย่างแตกต่างกัน ข้อสอบที่ทำให้ผู้ตอบที่มีคุณลักษณะเหมือนกัน แต่ลักษณะของกลุ่มย่อยแตกต่างกัน มีโอกาสที่จะตอบข้อสอบได้ถูกต้องแตกต่างกัน เกิดการได้เปรียบเสียเปรียบกันในระหว่างกลุ่มผู้สอบ ทำให้ต้องตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ เพื่อพัฒนาประสิทธิภาพตามโครงสร้างและมาตรฐาน

ตอนที่ 5 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ด้วยวิธี IRT-LR และงานวิจัยที่เกี่ยวข้อง

การทำหน้าที่ต่างกันของข้อสอบด้วยวิธี IRT-LR โมเดล ประกอบด้วย 2 ขั้นตอน ที่สัมพันธ์กัน ได้แก่ขั้นตอนแรกทำการวัดขนาดของการทำหน้าที่ต่างกันของข้อสอบ (Measurement of DIF) และขั้นตอนที่สอง ทำการทดสอบทางสถิติของการทำหน้าที่ต่างกันของข้อสอบ (Statistical test of DIF) คำถามหลักของการตรวจสอบด้วย DIF ด้วย IRT โมเดล คือ " โค้งลักษณะข้อสอบของประชากรต่างกลุ่มมีความแตกต่างกันหรือไม่" จึงต้องวัดขนาดของความแตกต่างและทดสอบทางสถิติว่ามีความแตกต่างอย่างมีนัยสำคัญระหว่างกลุ่มหรือไม่

วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วย IRT โมเดล ที่นิยมใช้กันมีดังนี้

1. วิธีวัดความแตกต่างของพื้นที่ (Area Measures: AREA)

1.1 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการวัดพื้นที่จะเปรียบเทียบฟังก์ชันการตอบข้อสอบระหว่างกลุ่ม โดยการคำนวณค่าพื้นที่ระหว่างโค้งลักษณะข้อสอบ จากกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่มีความสามารถระดับเดียวกัน ซึ่งมีสูตรทั่วไปที่ใช้ในการคำนวณพื้นที่ดังนี้ (Millsap & Everson, 1993)

$$A = fs[P_R(\theta - P_F)(\theta)] \quad (7)$$

เมื่อ

A = ดัชนีพื้นที่ $S = (\theta_{L,}\theta_H)$

fs = ฟังก์ชัน f ในช่วง s ซึ่ง$S = (\theta_{L,}\theta_H)$ โดยที่ L และ H เป็นค่าต่ำสุดและสูงสุดตามลำดับ

$P_R(\theta)$ = โอกาสของการตอบข้อสอบถูก ที่ระดับความสามารถ จากกลุ่มผู้สอบกลุ่มอ้างอิง

$P_F(\theta)$ = โอกาสของการตอบข้อสอบถูกที่ระดับความสามารถ จากกลุ่มผู้สอบกลุ่มเปรียบเทียบ

ในการคำนวณพื้นที่สามารถเลือกคำนวณได้ทั้งชนิดคิดเครื่องหมาย (Signed Area) หรือชนิดไม่คิดเครื่องหมาย (Unsigned Area) และคำนวณพื้นที่แบบต่อเนื่อง (Continuous Integration) หรือการประมาณค่าแบบไม่ต่อเนื่อง (Discrete Approximation)

การคำนวณในระยะแรกเป็นแบบไม่ต่อเนื่องและไม่คิดเครื่องหมาย Rudner (1997) ได้เสนอสูตรดังนี้

$$R = \sum_{j=3}^{3} |D_j| \Delta \quad (8)$$

เมื่อ $D_j = P_{R(\theta)} - P_{F(\theta)}$

$\Delta = .005$

ดัชนี R นี้คำนวณแบบไม่คิดเครื่องหมายเมื่อนำค่าสัมบูรณ์ออกก็จะเปลี่ยนเป็นดัชนีชนิดคิดเครื่องหมาย ซึ่งอาจจะมีเครื่องหมายเป็นบวกหรือลบก็ได้ ต่อมา Linn et al (1981) ได้เสนอดัชนีชนิดไม่คิดเครื่องหมาย เรียกว่า รากกำลังสองของความแตกต่างเฉลี่ย
(Root Mean Squared Different: RMSD) ระหว่างฟังก์ชันการตอบสนองข้อสอบโดยแบ่งช่วงระดับความสามารถระหว่าง – 3 ถึง + 3 ออกเป็น 600 ช่วง ดังนี้

$$RMSD = \sqrt{\frac{1}{600}\sum_{j=1}^{n}\left[P_{R(\theta)} - P_{F(\theta)}\right]^2} \quad (9)$$

รูดเดอร์ เกทสัน และไนท์ (Ruder, Getson & Knight, 1980) ได้เสนอสูตรการคำนวณพื้นที่แบบต่อเนื่องชนิดคิดเครื่องหมาย (SIGNED – AREA) และไม่คิดเครื่องหมาย
(UNSIGNED - AREA) ดังนี้

$$SIGNED - AREA = \int \left[P_{R(\theta)} - P_{F(\theta)}\right] d\theta \quad (10)$$

$$UNSIGNED - AREA = \sqrt{\int \left[P_{R(\theta)} - P_{F(\theta)}\right]^2 d\theta} \quad (11)$$

ดัชนี (SIGNED – AREA) ถ้ามีเครื่องหมายเป็น + แสดงว่ากลุ่มอ้างอิงทำข้อสอบได้ดีกว่า แต่ถ้าเครื่องหมายเป็น - แสดงว่ากลุ่มเปรียบเทียบทำข้อสอบได้ดีกว่า สำหรับดัชนีบวก (UNSIGNED - AREA) ที่มีค่ามากกว่า (SIGNED - AREA) เป็นสัญญาณแสดงว่าโค้งลักษณะข้อสอบของกลุ่มทั้งสองตัดกัน

วิธีการวัดความแตกต่างของพื้นที่เป็นวิธีที่ทำความเข้าใจได้ง่ายสามารถวาดภาพแสดงได้ชัดเจน แต่มีจุดอ่อนที่มิได้สนใจระบุช่วง ที่มีความแตกต่างกันมากและมีปัญหาด้านความน่าเชื่อถือของค่าที่คำนวณได้ ถ้ากลุ่มทั้งสองมีค่าพารามิเตอร์ C แตกต่างกัน วิธีวัดความแตกต่างของพื้นที่นิยมใช้กัน 2 วิธี คือวิธีการวัดพื้นที่ของราจู (Raju, 1990) และวิธีการวัดพื้นที่ของคิมและโคเฮน
(Kim & Cohen,1991) ซึ่งมีรายละเอียดดังนี้

วิธีการวัดพื้นที่ของ (Raju, 1990) ได้เสนอสูตรการคำนวณพื้นที่ในช่วงเปิดของ ชนิดคิดเครื่องหมาย(Open Interval Signed Area or Exact Signed Area: ESA)และพื้นที่ชนิดไม่คิดเครื่อง วิธีการวัด (Open Interval Unsigned Area or Exact Signed Area: UEA) เพื่อใช้

ตรวจสอบการทำหน้าที่ต่างกันของข้อสอยภายใต้ทฤษฎีIRT โมเดลโลจิสติก แบบ 1, 2 และ 3 พารามิเตอร์ สูตรทั่วไปที่ใช้ในการคำนวณพื้นที่ทั้งสองชนิดมีลักษณะ ดังนี้

$$\mathrm{ESA} = \int_{-\infty}^{+\infty}\left[\mathrm{P_{R(\theta)}} - \mathrm{P_{F(\theta)}}\right]\mathrm{d\theta} \tag{12}$$

$$\mathrm{EUA} = \int_{-\infty}^{+\infty}\left|\mathrm{P_{R(\theta)}} - \mathrm{P_{F(\theta)}}\right|\mathrm{d\theta} \tag{13}$$

ในการทดสอบว่าข้อสอบที่นำมาตรวจสอบการทำหน้าที่ต่างกันหรือไม่ ราจู (Raju, 1990) ได้เสนอให้นำดัชนีการวัดพื้นที่มาทดสอบนัยสำคัญโดยใช้สถิติ Z ภายใต้สมมติฐานการแจกแจงแบบปกติ ในการทดสอบดังกล่าวแบ่งออกเป็น 2 ลักษณะดังนี้

(1) การทดสอบนัยสำคัญของ ESA นำดัชนี ESA ของข้อสอบข้อที่ i ไปทดสอบความแตกต่างกับ 0 โดย ใช้สถิติทดสอบ Z ดังนี้

$$\mathrm{Z_i(ESA)} = \frac{\mathrm{\hat{b}_{iF} - \hat{b}_{iR}}}{\sqrt{\mathrm{Var(\hat{b}_{iF}) + var\hat{b}_{iR}}}} \tag{14}$$

(2) การทดสอบนัยสำคัญของ EUA ในการทดสอบนัยสำคัญของดัชนี EUA สามารถแบ่งออกเป็น 2 กรณีคือกรณีที่ 1 เมื่อ $a_{iR} = a_{iF}$ จะทดสอบเหมือนกับดัชนี ESA สำหรับกรณีที่ 2 $a_{iR} \neq a_{iF}$ จะนำดัชนี EUA ของข้อสอบที่ i ไปทดสอบความแตกต่างกับ 0 โดยใช้สถิตทดสอบดังนี้

$$\mathrm{Z_1(H)} = \frac{\mathrm{H_1}}{\sqrt{\mathrm{Var(H_1)}}} \tag{15}$$

1.2 การวัดพื้นที่ของ Kim และ Cohen

คิมและโคเฮน (Kim & Cohen ,1991) ได้พัฒนาสูตรการคำนวณพื้นที่ในช่วงปิดชนิดคิดเครื่องหมาย (Close - interval Singed Area; CSA) และพื้นที่ชนิดไม่คิดเครื่องหมาย (Close- interval UnSinged Area; CUA) โดยคำนวณพื้นที่ระหว่างฟังก์ชันการตอบสนองข้อสอบ (IRF_s) บนช่วงความสามารถ$[\theta_1\theta_2]$ซึ่งมีสูตรในรูปทั่วไปดังนี้

$$CSA = \int_{\theta_1}^{\theta_2}\left[P_R(\theta) - P_{F(\theta)}\right]d\theta \tag{16}$$

$$= S_{R(\theta_1,\theta_2)} - S_{F(\theta_1,\theta_2)} \tag{17}$$

$$\mathrm{CUA} = \int_{\theta_1}^{\theta_2}\left|\mathrm{P_R(\theta) - P_{F(\theta)}}\right|\mathrm{d\theta} \tag{18}$$

$$= \left|\mathrm{S_{R(\theta_1,\theta_2)} - S_{F(\theta_1,\theta_2)}}\right| \tag{19}$$

สำหรับการทดสอบการทำหน้าที่ต่างกันของข้อสอยด้วยวิธีการวัดพื้นที่ของคิมและโคเฮน (1991) ไม่ได้เสนอวิธีการทดสอบนัยสำคัญ แต่จะนำขนาดของพื้นที่ไปเปรียบเทียบกับเกณฑ์ที่กำหนดขึ้น

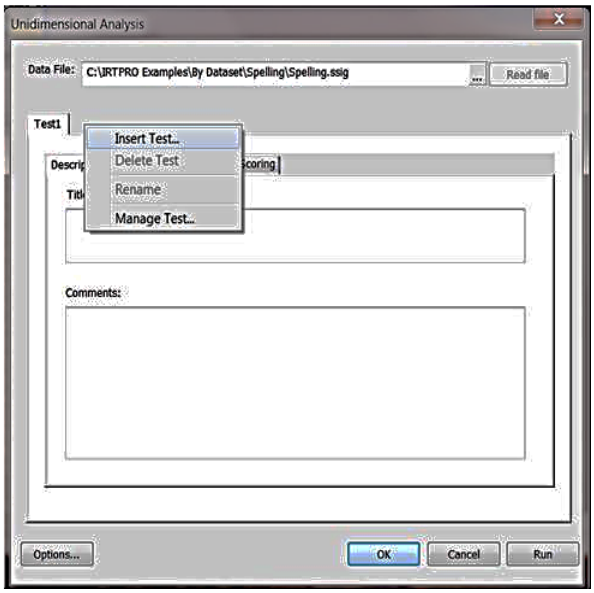
1. การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ โดยใช้โปรแกรม IRTPRO

IRTPRO เป็นโปรแกรมใหม่ในการทดสอบโดยใช้การวิเคราะห์ข้อมูลบนพื้นฐานทฤษฎีการตอบสนองข้อสอบ (IRT) IRTPRO ใช้อัลกอลิทึมในการประมาณค่าความน่าจะเป็นสูงสุดของพารามิเตอร์ ซึ่งสามารถวัดได้ทั้งแบบเอกมิติและพหุมิติ แม้กระทั่งแบบทวิภาคและพหุวิภาค นอกจากนี้โปรแกรมยังคำนวณดัชนีความพอดีทางสถิติ สำหรับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ และปรับขนาดในการตอบสนองข้อสอบ (Thissen, 2009)

1.1 ขั้นตอนการวิเคราะห์ วิธีการตอบสนองข้อสอบจากโปแกรม IRTPRO มีดังต่อไปนี้

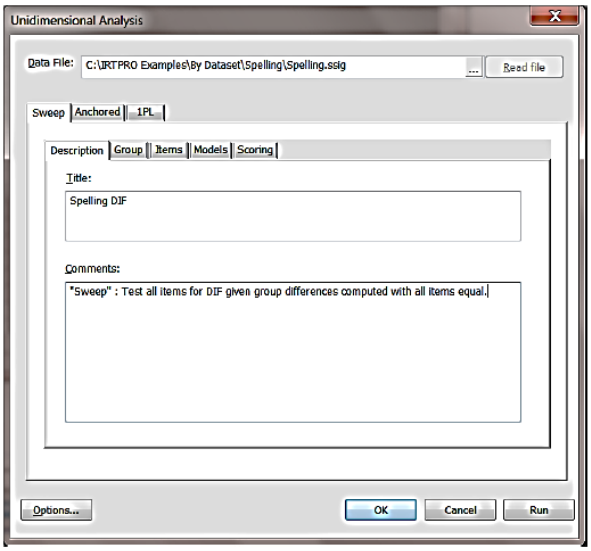
1.1.1 เริ่มต้นด้วยการเพิ่ม 2 แบบทดสอบ ที่ใช้ข้อมูลเดียวกันทั้งหมด

1.1.2 คลิกขวาบนแท็บใน test 1 ในการป้อนข้อมูล ลบการทดสอบหรือการเปลี่ยนชื่อการทดสอบที่มีอยู่ชื่อการทดสอบ ดังภาพที่ 2-16



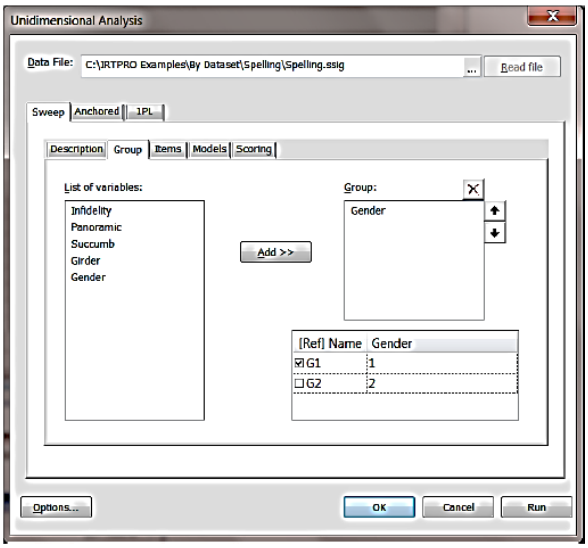
ภาพที่ 2-16 การเพิ่มข้อมูล (Thissen, 2009, p. 100)

1.1.3 คลิกแท็บคำอธิบาย เพื่อเพิ่มชื่อและแสดงความคิดเห็น ดังภาพที่ 2-17



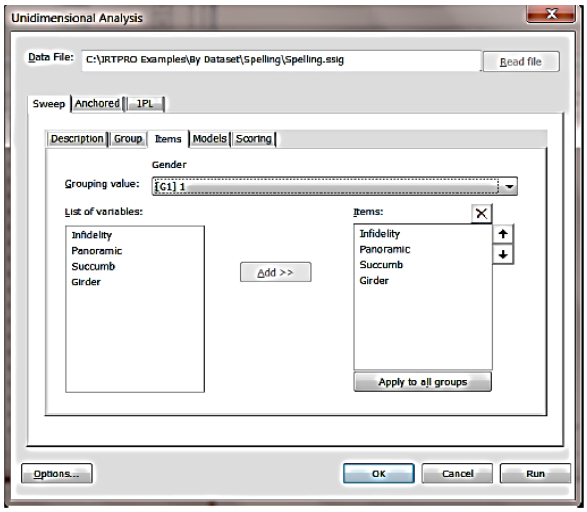
ภาพที่ 2-17 การเพิ่มเติมชื่อและแสดงความคิดเห็น (Thissen, 2009, p. 101)

1.1.4 คลิกที่แท็บเลือกกลุ่มและเพศ เป็นต้น เพื่อเป็นตัวแปรการจัดกลุ่ม กลุ่มอ้างอิงคือกลุ่มแรกที่แสดงด้านล่าง ดังภาพที่ 2-18



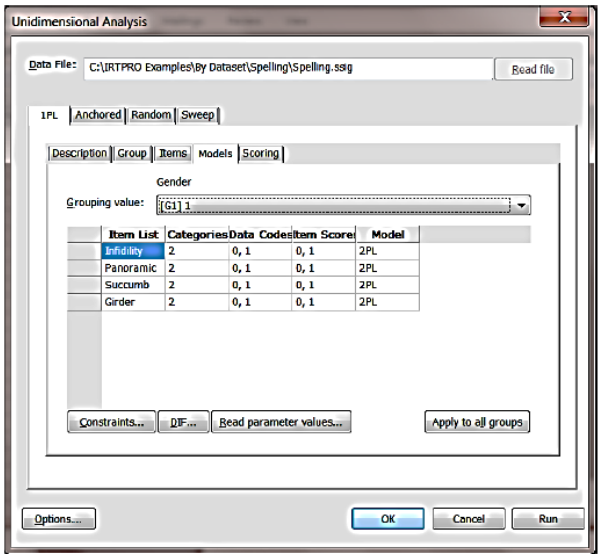
ภาพที่ 2-18 การจัดกลุ่มตัวแปร (Thissen, 2009, p. 102)

1.1.5 ใช้แท็บการวิเคราะห์เอกมิติ เลือกรายการที่จะได้รับการวิเคราะห์และคลิกที่ทุกปุ่มของกลุ่ม เพื่อเลือกรายการเหล่านั้นสำหรับทั้ง 2 กลุ่ม ดังภาพที่ 2-19



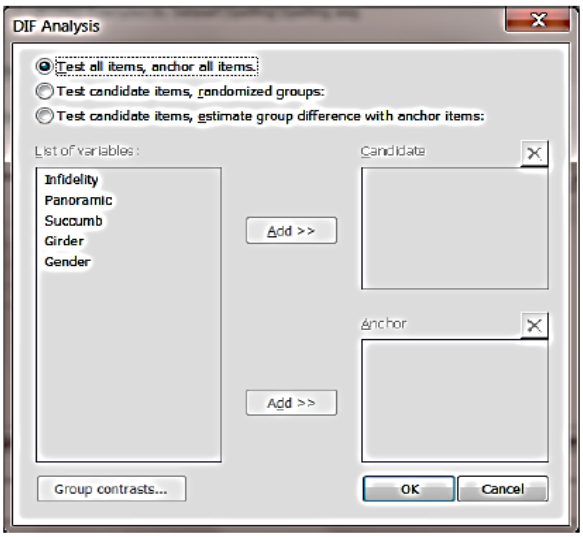
ภาพที่ 2-19 การเลือกรายการที่จะวิเคราะห์ (Thissen, 2009, p. 103)

1.1.6 เมื่อการดำเนินการเริ่มขึ้น ต่อไปจะเห็นได้ว่ามี 2 ประเภทและรูปแบบที่ 2PL จะแสดงเป็นค่าเริ่มต้นสำหรับแต่ละรายการ ดังภาพที่ 2-20



ภาพที่ 2-20 การแสดงค่าเริ่มต้นใน 2PL (Thissen, 2009, p. 105)

1.1.7 คลิกแท็บ การวิเคราะห์ DIF ขั้นตอนนี้จะนำเข้าสู่การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบโดยการกำหนดเงื่อนไขเพิ่มเติมต่อไป คลิกปุ่ม OK และคลิกที่ปุ่ม Run เพื่อเริ่มต้นการวิเคราะห์ ดังภาพที่ 2-21



ภาพที่ 2-21 การนำเข้าสู่การวิเคราะห์ DIF โดยการกำหนดเงื่อนไขเพิ่มเติม (Thissen, 2009, p. 105)

1.1.8 ตัวอย่างของผลลัพธ์ที่เกิดจากการวิเคราะห์ข้อมูล ดังภาพที่ 2-22

2PL Model Item Parameter Estimates for Group 1, logit: $a\theta + c$ or $a(\theta - b)$ (Back to TOC)

Item	Label	a	s.e.	c	s.e.	b	s.e.
1	Infidelity	[2] 0.88	0.26	[1] 1.30	0.18	-1.48	0.37
2	Panoramic	[4] 1.48	0.35	[3] 0.78	0.19	-0.52	0.14
3	Succumb	[6] 1.86	0.54	[5] -1.19	0.27	0.64	0.13
4	Girder	[8] 1.44	0.36	[7] 0.72	0.19	-0.50	0.14

2PL Model Item Parameter Estimates for Group 2, logit: $a\theta + c$ or $a(\theta - b)$ (Back to TOC)

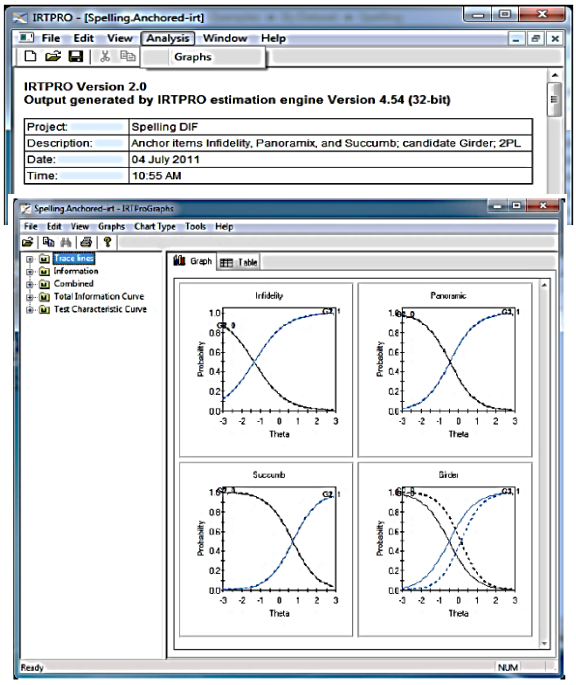
Item	Label	a	s.e.	c	s.e.	b	s.e.
1	Infidelity	[10] 1.62	0.43	[9] 2.29	0.34	-1.41	0.22
2	Panoramic	[12] 1.51	0.37	[11] 0.94	0.19	-0.62	0.11
3	Succumb	[14] 1.14	0.29	[13] -0.78	0.13	0.69	0.21
4	Girder	[16] 1.64	0.43	[15] 0.14	0.16	-0.09	0.09

DIF Statistics for Graded Items (Back to TOC)

Item numbers in:										
Group 1	Group 2	Total X^2	d.f.	p	X^2_a	d.f.	p	$X^2_{c\|a}$	d.f.	p
1	1	7.5	2	0.0235	2.2	1	0.1348	5.3	1	0.0219
2	2	0.5	2	0.7765	0.0	1	0.9516	0.5	1	0.4787
3	3	2.1	2	0.3544	1.4	1	0.2393	0.7	1	0.4069
4	4	8.2	2	0.0167	0.1	1	0.7164	8.1	1	0.0045

ภาพที่ 2-22 ตัวอย่างผลลัพธ์จากการวิเคราะห์ DIF (Thissen, 2009, p. 106)

1.1.9 ในขณะที่การส่งออกข้อมูลจะปรากฏกราฟนี้อยู่ภายใต้การวิเคราะห์รายการบนแถบเมนูหลัก ดังภาพที่ 2-23



ภาพที่ 2-23 กราฟการวิเคราะห์ DIF (Thissen, 2009, pp. 106-107)

การศึกษาแบบจำลองเพื่อเปรียบเทียบวิธีการหลายวิธีสำหรับการวิเคราะห์ข้อมูลเชิงสำรวจและข้อมูลเชิงประจักษ์ โดยใช้โปรแกรม IRTPRO เมื่อจำนวนของปัจจัยที่มีขนาดใหญ่ โดยวิธีการรวมตัวเลขที่ใช้สำหรับการคำนวณประมาณค่าสูงสุดเป็นยาก เนื่องจากเป็นข้อมูลจำนวนมาก จึงใช้การคำนวณความน่าจะเป็น (Scott et al., 2009) โดย Ong (2014) ได้การทำหน้าที่ต่างกันของข้อสอบ (DIF) วัตถุประสงค์แรกของการศึกษานี้คือเพื่อเปรียบเทียบข้อมูลเชิงประจักษ์ด้วย โปรแกรม IRTPRO, BILOG-MG 3 และ IRTLRDIF และในการตรวจสอบ DIF ในสองกลุ่มตัวอย่างที่มีรูปแบบ IRT, 1PL, 2PL และ 3PL วัตถุประสงค์ที่สองคือเพื่อตรวจสอบ ประสิทธิภาพของโปรแกรม IRTPRO ในการตรวจสอบ DIF และสุดท้ายที่จะพิจารณาว่าการทำหน้าที่ต่างกันของข้อสอบที่อยู่ในแบบทดสอบ GHSGPT สำหรับคุณลักษณะที่แตกต่างกันทางเชื้อชาติโดยเฉพาะในวิชาสังคมศึกษา แบบทดสอบ GHSGPT ใช้ทำนายในอนาคตนักเรียนของเกรด 11 ที่จบการศึกษาโรงเรียนมัธยมศึกษาจอร์เจีย แบบทดสอบ ประกอบด้วย 79 ข้อ เป็นแบบทดสอบประเภททวิภาค ผลการศึกษา พบว่า แบบทดสอบ GHSGPT มีการทำหน้าที่ต่างกันของข้อสอบหลายข้อ ยกตัวอย่างเช่นทั้งสามโปรแกรมมีการทำหน้าที่ต่างกันของข้อสอบ 13 ข้อ และเป็นประโยชน์ต่อคนผิวขาว นอกจากนี้ โปรแกรม IRTPRO มีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ เพราะผลของมันคู่ขนานโปรแกรม IRTLRDIF และ BILOG-MG 3 สอดคล้องกับงานวิจัยของ Paek and Han (2012) ได้ทำการศึกษาการใช้โปรแกรม IRTPRO 2.1 โดยใช้ทฤษฎีการตอบสนองข้อสอบสำหรับการรายงานการรักษาผู้ป่วยซึ่งโปรแกรมมีประสิทธิภาพในการประมาณค่าพารามิเตอร์ทั้งในแบบเอกมิติและพหุมิติ พบว่าประสิทธิภาพของการประมาณค่าและการดำเนินงานของโปรแกรมน่าสนใจต่อผู้ใช้

ซึ่งทั้งประสิทธิภาพและประสิทธิผลของโปรแกรม IRTPRO 2.1 สูง ซึ่งการศึกษาดังกล่าวทำให้ทราบว่า โปรแกรม IRTPRO มีความน่าสนใจเป็นอย่างยิ่ง

งานวิจัยที่เกี่ยวข้องกับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธี IRT-LR มีดังนี้

Paek and Han (2012) ได้ทำการศึกษาการใช้โปรแกรม IRTPRO 2.1 ตามหลักทฤษฎีการตอบสนองข้อสอบสำหรับการรายงานการรักษาผู้ป่วยซึ่งโปรแกรมมีประสิทธิภาพในการประมาณค่าพารามิเตอร์ทั้งในแบบเอกมิติและพหุมิติ พบว่าประสิทธิภาพของการประมาณค่าและการดำเนินงานของโปรแกรมน่าสนใจต่อผู้ใช้ซึ่งทั้งประสิทธิภาพและประสิทธิผลของโปรแกรม IRTPRO 2.1 สูง

Tihomir and Bengt (2012) ได้ศึกษาการเปรียบเทียบอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ที่ใช้ค่าเฉลี่ยและโครงสร้างความแปรปรวน (MACS) และทฤษฎีการตอบสนองข้อสอบ (IRT) และคะแนนที่สังเกตได้ตามวิธีถดถอยโลจิสติก สำหรับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ในแบบทดสอบพหุวิภาค การศึกษาเป็นแบบจำลองการทำงานซึ่งเป็นข้อมูลพหุวิภาคกับห้าประเภทที่ถูกสร้างขึ้นโดยใช้รูปแบบระดับของการตอบสนองของ Samejima ภายใต้ 3 ปัจจัย : ขนาดของกลุ่มตัวอย่าง (300, 500 และ 1,000 คน) ประเภทของการทำหน้าที่ต่างกันของข้อสอบและขนาดของการทำหน้าที่ต่างกันของข้อสอบ (ขนาดเล็กและขนาดใหญ่ของการทหน้าที่ต่างกันของข้อสอบ) ตามการทดสอบ ประเภทอัตราอัตราความคลาดเคลื่อน สูงขึ้นสำหรับ IRT และการถดถอยโลจิสติกตามลำดับเมื่อบางส่วนของข้อสอบที่มีอยู่ในการทำหน้าที่ต่างกันของข้อสอบ สำหรับข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูปนั้น MACS และ IRT แสดงอัตราความถูกต้องที่คล้ายกัน แต่วิธีถดถอยโลจิสติแสดงอัตราความถูกต้องที่สูงขึ้นเล็กน้อย เมื่อเทียบกับอีกสองวิธีสำหรับขนาดของกลุ่มตัวอย่างที่มีขนาดเล็ก ข้อสอบที่มีการทำหน้าที่ต่างกันแบบอเนกรูป IRT แสดงอัตราความถูกต้องมากขึ้นเมื่อเทียบกับ วิธี MACS และวิธีถดถอยโลจิสติก

Elosua and Wells (2013) ศึกษาการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบพหุมิติใช้วิธีโครงสร้างความแปรปรวน (MACS) และทฤษฎีการตอบสนองข้อสอบ (IRT) และวิธีถดถอยโลจิสติก โดยมีวัตถุประสงค์เพื่อเปรียบเทียบอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 การศึกษาการจำลองการทำงานในข้อมูลพหุมิติ ห้าประเภทที่ สร้างขึ้นโดยใช้ รูปแบบการตอบสนองของ Samejima ภายใต้ 3 ปัจจัย คือขนาดของกลุ่มตัวอย่าง (300, 500 และ 1,000) ประเภทของการทำหน้าที่ต่างกันของข้อสอบ (พารามิเตอร์ a, b, c) และขนาดของการทำหน้าที่ต่างกันของข้อสอบ (ขนาดเล็กและขนาดใหญ่) พบว่าอัตราความคลาดเคลื่อนสูงสำหรับวิธีทฤษฎีการตอบสนองข้อสอบและการถดถอยโลจิสติกตามลำดับ เมื่อบางส่วนของแบบทดสอบที่มีอยู่ข้อสอบที่ทำหน้าที่ต่างกัน ทั้ง 3 วิธีแสดงอัตราความถูกต้องคล้ายกันซึ่งวิธีการถดถอยโลจิสติกแสดงอัตราความถูกต้องสูงขึ้นเพียงเล็กน้อยเมื่อเทียบกับอีกสองวิธี สำหรับกลุ่มตัวอย่างที่มีขนาดเล็ก การทำหน้าที่ต่างกันของข้อสอบจะมีความไม่สม่ำเสมอ วิธีทฤษฎีการตอบสนองข้อสอบจะแสดงอัตราความถูกต้องได้ดีที่สุดเมื่อเทียบกับวิธีโครงสร้างความแปรปรวน (MACS) และวิธีถดถอยโลจิสติก

Ling (2014) ได้การทำหน้าที่ต่างกันของข้อสอบ (DIF) วัตถุประสงค์แรกของการศึกษานี้คือเพื่อเปรียบเทียบข้อมูลเชิงประจักษ์ด้วย โปรแกรม IRTPRO, BILOG-MG 3 และ IRTLRDIF และใน

การตรวจสอบ DIF ในสองกลุ่มตัวอย่างที่มีรูปแบบ IRT, 1PL, 2PL และ 3PL วัตถุประสงค์ที่สองคือ เพื่อตรวจสอบ ประสิทธิภาพของโปรแกรม IRTPRO ในการตรวจสอบ DIF, และสุดท้ายที่จะพิจารณาว่าการทำหน้าที่ต่างกันของข้อสอบที่อยู่ในแบบทดสอบ GHSGPT สำหรับคุณลักษณะที่แตกต่างกันทางเชื้อชาติโดยเฉพาะในวิชาสังคมศึกษา แบบทดสอบ GHSGPT ใช้ทำนายในอนาคตนักเรียนของเกรด 11 ที่จบการศึกษาโรงเรียนมัธยมศึกษาจอร์เจีย แบบทดสอบประกอบด้วย 79 ข้อ เป็นแบบทดสอบประเภททวิภาค ผลการศึกษาพบว่าแบบทดสอบ GHSGPT มีการทำหน้าที่ต่างกันของข้อสอบหลายข้อ ทั้งสามโปรแกรมมีการทำหน้าที่ต่างกันของข้อสอบ 13 ข้อ และเป็นประโยชน์ต่อคนผิวขาว นอกจากนี้ โปรแกรม IRTPRO มีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ เพราะผลของมันคู่ขนานกับโปรแกรม IRTLRDIF และ BILOG-MG 3 ซึ่งการศึกษาดังกล่าวทำให้ทราบว่า โปรแกรม IRTPRO มีความน่าสนใจเป็นอย่างยิ่ง

จากการศึกษางานวิจัยที่เกี่ยวข้องกับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ด้วยวิธี IRT-LR โดยใช้โปรแกรม IRTPRO เป็นโปรแกรมที่น่าสนใจเป็นอย่างยิ่ง วิธี IRT-LR จะแสดงอัตราความถูกต้องได้ดีที่สุดเมื่อเทียบกับวิธีโครงสร้างความแปรปรวน (MACS)และวิธีถดถอยโลจิสติก และมีประสิทธิภาพของการประมาณค่าและการดำเนินงานของโปรแกรมน่าสนใจต่อผู้ใช้ซึ่งทั้งประสิทธิภาพและประสิทธิผลของโปรแกรม IRTPRO 2.1 สูง

ตอนที่ 6 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ด้วยวิธี SIBTEST และงานวิจัยที่เกี่ยวข้อง

1. ความเป็นมาของวิธี SIBTEST

วิธี SIBTEST (Simulatataneous Item Bias Test: SIBTEST) โดย Shealy และ Stout (1993) ใช้ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) การทำหน้าที่ต่างกันของแบบทดสอบ (DIF) และการทำหน้าที่ต่างกันของกลุ่มข้อสอบ (Differential Bundle Functioning: DBF) วิธีนี้สามารถวิเคราะห์ได้ทั้งแบบสอบเอกมิติ (Unidimensional Test) และแบบสอบพหุมิติ (Multidimensional Test) (Stout et al., 1997) วิธี SIBTEST ใช้สถิติทดสอบแบบนันพาราเมตริก (Nonparmetric) ซึ่งมาพัฒนาบนพื้นฐานทฤษฎี IRT ชนิดพหุมิติไม่ต้องใช้ฟังก์ชันการตอบสนองข้อสอบหรือการประมาณค่าความสามารถแฝง วิธี SIBTEST ถูกออกแบบมาสำหรับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบมีทิศทางเดียว (Unidirectional DIF) ดังนั้นจึงไม่มีความไวในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบมีทิศทางเดียว (Nondirectional DIF) (Li & Stout, 1996) จุดเด่นของวิธี SIBTEST คือสามารถคำนวณได้ง่ายไม่ซับซ้อน ประหยัดค่าใช้จ่ายและไม่จำเป็นต้องใช้กลุ่มตัวอย่างที่มีขนาดใหญ่ ทั้งยังใช้สถิติทดสอบ นัยสำคัญ (Narayanan & Swminathan, 1996) นอกจากนี้ ยังสามารถนำไปประยุกต์ใช้กับการตรวจสบการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบพหุวิภาค (Polytomous DIF) (Chang, Mazzeo, & Roussos, 1996)

ในการศึกษาการทำหน้าที่ต่างกันของข้อสอบของแบบสอบเอกมิติของวิธี SIBTEST ถือว่าข้อสอบในแบบสอบจะต้องมุ่งวัดคุณลักษณะหรือความสามารถแฝงเพียงลักษณะเดียว ความสามารถแฝงจำแนกเป็นความสามารถเป้าหมายที่ต้องการวัด (Target Ability: θ) กับความสามารถแทรกซ้อน

ที่ไม่ใช่เป้าหมายของการวัด (Nuisanee Ability: η) ตัวอย่างเช่น แบบสอบคำศัพท์ในวิชาภาษาอังกฤษ ข้อสอบบางข้ออาจถามความรู้สำหรับผู้ชายเป็นพิเศษ เช่น ความรู้เกี่ยวกับกีฬา ในขณะที่ข้อสอบบางข้ออาจถามความรู้เกี่ยวกับคำศัพท์ในวิชาภาษาอังกฤษ เป็นความสามารถเป้าหมายที่ต้องการวัด (θ) ส่วนความรู้ด้านกีฬาและงานในบ้านเป็นความสามารถแทรกซ้อน ที่ไม่ใช่เป้าหมายของการวัด (η_1 และ η_2) ข้อสอบทุกข้อในแบบทดสอบจะวัดความสามารถเป้าหมาย ส่วนข้อสอบบางข้อทำหน้าที่ต่างกันจะวัดทั้งความสามารถเป้าหมาย และความสามารถพร้อมแทรกซ้อน (Nadakumar, 1993)

ถ้าให้ฟังก์ชันการตอบสนองข้อสอบ (IRF) ข้อที่ i ซึ่งขึ้นอยู่กับความสามารถ θ เพียงอย่างเดียวแทนด้วย $P_{i(\theta)}$ ส่วน IRF ข้อที่ i ขึ้นอยู่กับความสามารถ ทั้ง θ และ η แทนด้วย $P_i(\theta, n)$ ฟังก์ชันการตอบข้อสนองข้อสอบของข้อสอบดังกล่าวแบบ 3 พารามิเตอร์จะเป็นดังนี้ (Shealy & Stout, 1993)

$$P_{i(\theta)} = C_{i+\frac{\langle 1-c \rangle}{1+xep\left[-1.7\partial_{i\theta}\left(\theta-b_{ig}\right)\right]}} \tag{20}$$

$$P_i(\theta, n) = C_{i+\frac{1-c}{1+xep\{-1.7\left[\partial_{i\theta}\left(\theta-b_{ig}\right)+\partial_{in}(n-b_{in})\right]}} \tag{21}$$

ดังนั้นฟังก์ชันความน่าจะเป็นอย่างมีเงื่อนไขของแบบแผนการตอบสนองข้อสอบทั้งฉบับเป็นดังนี้

$$P[u/(\emptyset = \theta, n = n)] = \widehat{\cap_{i=1}} P_i(\theta, n)^{ui}[1 - P(\theta, n)]^{1-ui} \tag{22}$$

เชลลี และสเตาท์ (Shealy & Stout, 1993) ได้ใช้ Marginal IRF$_s$ อธิบายการทำหน้าที่ต่างกันของข้อสอบดังนี้

$$M_{ig(\theta)} = \int n P_{i(\theta,n)} \int g\left({}^{n}/_{\theta}\right) dn \tag{23}$$

เมื่อ

$M_{ig(\theta)}$ = Marginal IRF สำหรับความสามารถเป้าหมายที่ต้องการวัด θ ของผู้สอบกลุ่มอ้างอิงหรือกลุ่มเปรียบเทียบ

$P_{i(\theta,n)}$ = IRF ของข้อสอบข้อที่ i

$\int g\left({}^{n}/_{\theta}\right)$ = การแจกแจงแบบมีเงื่อนไขของกลุ่มผู้สอบ

การเปรียบเทียบ Marginal IRF$_s$ ระหว่างกลุ่มอ้างอิง (R) กับกลุ่มเปรียบเทียบ (F) จะทำให้ทราบถึงทิศทางของการได้เปรียบหรือเสียเปรียบกล่าวคือ ถ้า $M_{if}(\theta) < M_{irR}(\theta)$ ทุกค่าของ θ แสดงว่าข้อสอบทำหน้าที่ต่างกันแบบมีทิศทางเดียวโดยข้อสอบจะเข้าข้างผู้สอบกลุ่มอ้างอิงและ

ถ้า $M_{if}(\theta) > M_{ir}(\theta)$ ทุกค่าของ θ แสดงว่าข้อสอบทำหน้าที่ต่างกันแบบมีทิศทางเดียวกันอาจเรียกอีกอย่างหนึ่งว่า "การทำหน้าที่ต่างกันแบบไม่ตัดกัน" (Noncrossing DIF)

ในการเปรียบเทียบผลการตอบข้อสอบระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบตามวิธี SIBTEST จะแบ่งข้อสอบออกเป็น 2 ชุดย่อย (Subtests) คือ (1) ชุดแบบสอบที่มีความตรง (Valid Subtests) หรือชุดแบบสอบที่ใช้จับคู่เปรียบเทียบ (Matching Subtests) แบบสอบชุดนี้ประกอบด้วยข้อสอบที่ไม่ทำหน้าที่ต่างกันและ (2) ชุดแบบสอบที่ต้องการศึกษา (Studied Subtests) ประกอบด้วยข้อสอบที่สงสัยว่าทำหน้าที่ต่างกัน ถ้าแบบสอบชุดแรกมีจำนวน n ข้อ (ข้อที่ 1 ถึง n) แล้วแบบสอบชุดที่สองจะมีจำนวน N - n ข้อ (ข้อที่ n +1 ถึง N) เมื่อ N เป็นจำนวนข้อสอบทั้งหมด

ฟังก์ชันการตอบสนองข้อสอบของแบบสอบที่ต้องการศึกษา จากผู้สอบกลุ่มอ้างอิงกับกลุ่มเปรียบเทียบ กำหนดในรูปฟังก์ชัน Marginal ดังนี้

$$M_{SR(\theta)} = \sum_{i=n+1}^{N} M_{iR(\theta)} \quad (24)$$

$$M_{SF(\theta)} = \sum_{i=n+1}^{N} M_{iR(\theta)} \quad (25)$$

เมื่อ

$M_{SR(\theta)}$ = ผลรวมของ marginal IRF_s ของข้อสอบที่ต้องการศึกษาจากกลุ่มผู้สอบกลุ่มอ้างอิง ณ ระดับความสามารถ θ

$M_{SF(\theta)}$ = ผลรวมของ marginal IRF_s ของข้อสอบที่ต้องการศึกษาจากกลุ่มผู้สอบ เปรียบเทียบ ณ ระดับความสามารถ θ

ขนาดของความแตกต่างระหว่าง $M_{SR(\theta)}$และ $M_{SF(\theta)}$ แสดงถึงปริมาณความเข้มของการทดสอบ การทำหน้าที่ต่างกันของข้อสอบแบบมีทิศทางเดียวหรือการทำหน้าที่ต่างกันแบบไม่ตัดกันจากชุดแบบสอบที่ต้องการศึกษา ณ ระดับความสามารถθ ของผู้สอบ ด้วยการอินทิเกรทดังนี้

$$\beta_{uni} = \int \theta \left[M_{SR(\theta)} - M_{SF(\theta)}\right] \int p(\theta)^{d\theta} \quad (26)$$

เมื่อ

β_{uni} = ดัชนีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบมีทิศทางเดียว

$\int p(\theta)$ = ฟังก์ชันของโอกาสการแจกแจงความสามารถของผู้สอบทั้งสองกลุ่ม

ดัชนี ที่คำนวณได้จากสูตรดังกล่าว นำมาทดสอบสมมติฐานของการทำหน้าที่ต่างกันของข้อสอบแบบมีทิศทางเดียวดังนี้

$$H_{0:\beta_{uni}=0}$$

$$H_{1:\beta_{uni}>0}$$

สมมติฐานอื่น (H_1) มีลักษณะทิศทางเดียว ซึ่งใช้การทดสอบการทำหน้าที่ต่างกันของข้อสอบที่เข้าข้างผู้สอบกลุ่มอ้างอิง สำหรับค่าประมาของ β_{uni} คำนวณได้จากคะแนนรวมของชุดแบบสอบที่มีความตรงและชุดแบบสอบที่ต้องการศึกษา ซึ่งกำหนดสัญลักษณ์ดังนี้

เมื่อ X = คะแนนรวมของชุดแบบสอบที่มีความตรง

Y = คะแนนรวมของชุดแบบสอบที่ต้องการศึกษา

U_i = ผลการตอบข้อสอบชุดที่ i (ตอบถูกได้ 1 คะแนน และตอบผิดได้ 0 คะแนน)

นำคะแนนของชุดแบบสอบที่มีความตรง (X) เป็นเกณฑ์ในการจับคู่ผู้ที่มีความสามารถระดับเดียวกันแล้วคำนวณคะแนนเฉลี่ยจากผลการตอบชุดแบบสอบที่ต้องการศึกษาระหว่างผู้สอบกลุ่มอ้างอิง และกลุ่มเปรียบเทียบที่มีความสามารถระดับเดียวกันมาจับคู่เปรียบเทียบกัน ณ $X = K$ โดยเขียนในรูป

สัญลักษณ์ได้ดังนี้ $\mathrm{Y_{RK} - Y_{FK}}$

เมื่อ

Y_{RK} = ค่าเฉลี่ยของคะแนนรายข้อ จากชุดแบบสอบที่ต้องการศึกษาของกลุ่มผู้สอบกลุ่มอ้างอิงซึ่งได้ คะแนน $X = K$

Y_{FK} = ค่าเฉลี่ยของคะแนนรายข้อ จากชุดแบบสอบที่ต้องการศึกษาของกลุ่มเปรียบเทียบซึ่งได้ คะแนน $X = K$

K = คะแนนรวมจากชุดแบบสอบที่มีความตรง

ค่า$Y_{RK} - Y_{FK}$ ดังกล่าวเป็นความแตกต่างของผลการตอบข้อสอบข้อสอบในชุดแบบสอบที่ต้องการศึกษา ระหว่างกลุ่มผู้สอบกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่มีความสามารถระดับเดียวกัน

ถ้า $Y_{RK} - Y_{FK} = 0$ ทุกคะแนน K แสดงว่าข้อสอบที่ต้องการศึกษาทำหน้าที่ไม่ต่างกัน แต่ถ้า $Y_{RK} - Y_{FK} > 0$ ทุกคะแนน K แสดงว่าข้อสอบที่ต้องการศึกษาทำหน้าที่ต่างกันแบบมีทิศทางเดียว โดยจะเข้าข้างผู้สอบกลุ่มอ้างอิง ค่าความแตกต่างของผลการตอบข้อสอบ ประมาณค่าในรูป $\hat{\beta}_{uni}$ ได้ดังนี้

$$\hat{\beta}_{uni} = \sum_{k=0}^{n} \hat{P}_k (\bar{Y}_{RK} - \bar{Y}_{FK}) \quad (27)$$

โดยที่ $$\hat{P}_k = \frac{(J_{Rk}+J_{Fk})}{\sum_{k=0}^{n}(J_{Rk}+J_{Fk})} \quad (28)$$

เมื่อ

$\hat{P}_k$ = สัดส่วนของจำนวนผู้สอบกลุ่มเปรียบเทียบและกลุ่มอ้างอิงที่ได้คะแนนรวม $X = k$ จากจำนวนผู้สอบ

J_{Fk} = จำนวนผู้สอบกลุ่มเปรียบเทียบที่ได้คะแนนรวม X = k

J_{Rk} = แทนจำนวนผู้สอบกลุ่มอ้างอิงที่ได้คะแนนรวม X = k

สำหรับการทดสอบสมมุติฐานศูนย์*ของ no DIF* ใช้สถิติ β_{uni} ดังนี้

$$\beta_{uni} = \frac{\hat{\beta}_{uni}}{\hat{\sigma}\,(\hat{\beta}_{uni})} \tag{29}$$

โดยที่ $\hat{\rho}(\hat{\beta}_{uni}) = \sqrt{\sum_{k=0}^{n} \widehat{Pk}^2 \left[\frac{1}{J_{Rk}} \hat{\rho}^2(Y[k,R) + \frac{1}{J_{Fk}} \hat{\rho}^2(Y[k,F)\right]}$ (30)

เมื่อ

$\hat{\rho}(\hat{\beta}_{uni})$ = ค่าประมาณความคลาดเคลื่อนมาตรฐาน

$\hat{\rho}^2(Y[k,g)$ = ค่าประมาณความแปรปรวนของคะแนนจากชุดแบบสอบที่ต้องการศึกษาสำหรับผู้สอบกลุ่ม g (*R* หรือ *F*) ซึ่งมีคะแนนรวมเท่ากับ *k*

J_{gk} = จำนวนผู้สอบกลุ่ม g (*R* หรือ *F*) ซึ่งตอบแบบทดสอบชุดที่มีความตรงแล้วได้คะแนนรวม X = K

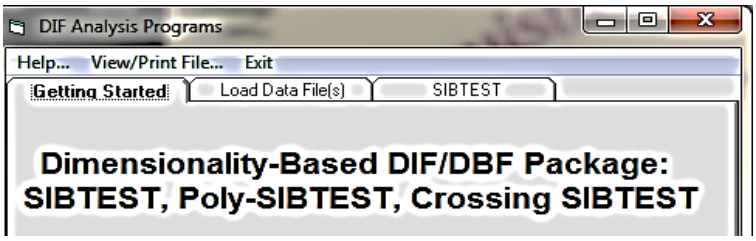
β_{uni} = มีการแจกแจงในลักษณะปกติมาตรฐานจึงสามารถทดสอบนัยสำคัญด้วยสถิติทดสอบ Z ถ้าผลการทดสอบปรากฏว่า

$\beta_{Uni} > Z_{\alpha}$ แสดงว่า การทดสอบมีนัยสำคัญจึงปฏิเสธ H_0 นั่นคือ ข้อสอบที่นำมาตรวจสอบการทำหน้าที่ต่างกัน โดยจะเข้าข้างผู้สอบกลุ่มอ้างอิง เมื่อβ_{Uni} มีค่าเป็นบวกและจะเข้าข้างผู้สอบกลุ่มเปรียบเทียบเมื่อ β_{Uni} มีค่าเป็นลบ สถิติที่ใช้ในการทดสอบสำหรับสรุปอ้างอิงการทำหน้าที่ต่างกันของข้อสอบดังกล่าวมักจะมีปัญหาในกรณีที่มีความแตกต่างของการแจกแจงความสามารถเป้าหมายระหว่างกลุ่มผู้สอบ กล่าวคือ ถ้ากลุ่มผู้สอบอ้างอิงมีความสามารถเป้าหมายสูงกว่าผู้สอบกลุ่มเปรียบเทียบจะเกิดผลกระทบทำให้สถิติ β_{Uni} มีค่าเฟ้อ (Inflate) หรือมีค่าสูงผิดปกติ ถึงแม้ว่าในความเป็นจริงข้อสอบทำหน้าที่ไม่ต่างกัน จึงควรแก้ไขความแตกต่างของการแจกแจงความสามารถเป้าหมาย ด้วยการปรับแก้ค่าถดถอย เพื่อขจัดผลกระทบดังกล่าวโดยปรับแก้ค่า $\bar{Y}_{Rk}$ และ $\bar{Y}_{Fk}$ เป็นรายคู่ก่อนการคำนวณ β_{Uni}

2. การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ (DIF) โดยใช้โปรแกรม SIBTEST

วิธี SIBTEST ใช้โปรแกรม SIBTEST ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบทั้งแบบเอกมิติและพหุมิติ ระหว่างกลุ่มผู้สอบ 2 กลุ่ม คือ กลุ่มอ้างอิง (Reference Group: R) และกลุ่มเปรียบเทียบ (Focal Group: F) รวมทั้งตรวจสอบสมมติฐานและการประมาณค่าในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

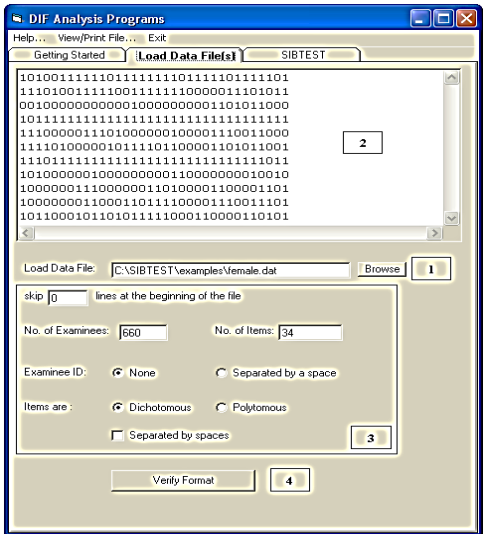
โปรแกรม SIBTEST มีองค์ประกอบ 3 เมนู ได้แก่ เมนูแนะนำซึ่งเป็นตัวช่วยเหลือในการรันโปรแกรม ข้อมูลสารสนเทศในนั้นสามารถสามารถบ่งบอกถึงเทคนิคการใช้โปแกรม ต่อมาคือไฟล์มุมมองในการพิมพ์ เป็นส่วนที่ผู้ใช้สามารถมองเห็นข้อมูลที่ต้องการจะพิมพ์และข้อมูลที่ส่งออกมาจากโปรแกรมที่เรากำหนด สุดท้ายคือเมนูออกจาโปรแกรม เป็นเมนูที่ใช้ออกจากโปรแกรม ซึ่งโปรแกรมจะมีโฟลเดอร์ในการทำงาน 3 โฟลเดอร์ คือ โฟลเดอร์เริ่มต้นการทำงาน ซึ่งเป็นตัวเชื่อมต่อกับข้อมูลสารสนเทศในโปรแกรมนี้และเป็นที่ซึ่งช่วยเหลือทางเทคนิคแก่ผู้ใช้ โฟลเดอร์ต่อมาคือไฟล์นำเข้าข้อมูล โปรแกรมนี้จะนำเข้าข้อมูลเป็นกลุ่ม การนำเข้าข้อมูลหนึ่งครั้งสามารถรันโปรแกรมได้ทั้ง SIBTEST, Poly-SIBTEST และ Crossing SIBTEST โฟลเดอร์สุดท้ายคือ SIBTEST เป็นการรันโปรแกรม ซึ่งจะเชื่อมต่อไปถึง Poly-SIBTEST และ Crossing SIBTEST ในการรันโปรแกรมด้วย ดังภาพที่ 2-24



ภาพที่ 2-24 หน้าจอเริ่มต้นของโปรแกรม SIBTEST (Stout & Roussos, 1999, p. 1)

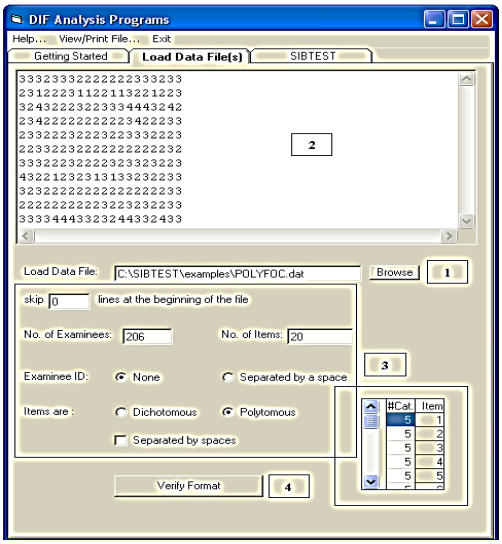
ขั้นตอนการนำเข้าข้อมูล มีดังต่อไปนี้

เริ่มต้นด้วยการคลิกที่กล่องสีขาวที่เขียนว่า Load Data File หน้าต่างจะปรากฏให้คุณเลือกไฟล์ในคอมพิวเตอร์ของคุณเมื่อคุณเลือกข้อมูลแล้วกดปุ่ม OK โปรแกรมจะอ่านไฟล์ข้อมูลของคุณ ต่อจากนั้นให้ตรวจสอบเนื้อหาในไฟล์ข้อมูล พร้อมทั้งตรวจสอบคุณลักษณะของข้อมูลและทำการเปลี่ยนแปลงเท่าที่จำเป็น สุดท้ายให้กดปุ่ม Verify Format ตามขั้นตอน ดังภาพที่ 2-25



ภาพที่ 2-25 ขั้นตอนการนำเข้าไฟล์ข้อมูล (Stout & Roussos, 1999, p. 4)

ในกรณีที่เป็นแบบทดสอบแบบเอกมิติ ให้ทำตามขั้นตอนข้างบน ส่วนในแบบทดสอบที่เป็นพหุมิตินั้นจะต้องเพิ่มตัวเลขของประเภทต่อข้อสอบโดยการเพิ่มค่าของมันออกเป็น 4 ขั้นตอนโดยสมบูรณ์ ถ้าข้อสอบแบบพหุมิติ มีคะแนนจาก 1 ถึง M ผู้ใช้ต้องเพิ่มตัวเลขเป็น M + 1 เนื่องจากว่าโปรแกรม SIBTEST สมมุติให้ค่าต่ำสุดของแบบทดสอบ คือ 0 ตั้งแต่โปรแกรมยอมให้วิเคราะห์ข้อมูลทั้งแบบเอกมิติและพหุมิติ คุณสามารถตั้งค่าข้อมูลต่ำสุดในการรับโปรแกรมโดยใช้แบบทดสอบพหุมิติได้ ดังภาพที่ 2-26



ภาพที่ 2-26 การใช้ข้อมูลของแบบทดสอบพหุมิติ (Stout & Roussos, 1999, p. 5)

ขั้นตอนการวิเคราะห์โปรแกรม SIBTEST มีดังต่อไปนี้

1.เลือกไฟล์ข้อมูลกลุ่มอ้างอิง

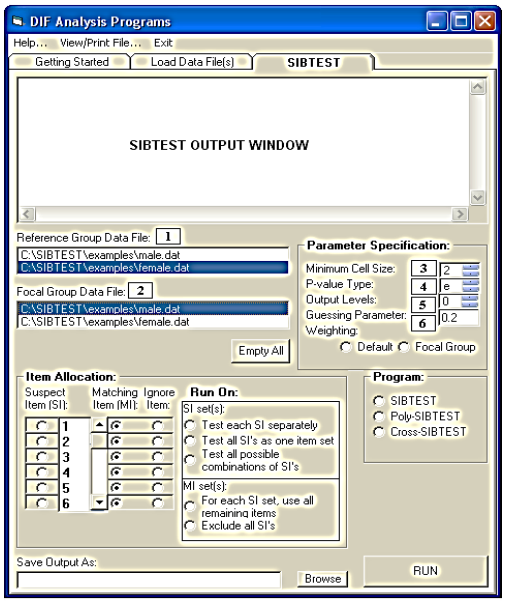
2. เลือกไฟล์ข้อมูลกลุ่มเปรียบเทียบ โดยกลุ่มอ้างอิงและกลุ่มเปรียบเทียบต้องมีแบบทดสอบเดียวกันทั้งรายข้อและรายหมวด และต้องเป็นข้อสอบประเภทเดียวกันไม่ว่าจะเป็นเอกมิติ หรือพหุมิติ ถ้าข้อมูลมีคุณสมบัติต่างกัน จะได้รับข้อความที่คลาดเคลื่อน

3. เลือกค่าต่ำสุดของเซลที่ใช้ในการคำนวณทางสถิติของการทำหน้าที่ต่างกันของข้อสอบ โดยเลือกจาก 2 ถึง 20

4. เลือกระดับนัยสำคัญ (*p*-value) โดยต้องเลือกว่าจะให้วิเคราะห์รายข้อหรือทั้งแบบทดสอบ เพื่อให้ขัดแย้งกับกลุ่มสองกลุ่ม กลุ่มอ้างอิงหรือกลุ่มเปรียบเทียบ

5. เลือกระดับผลการวิเคราะห์ข้อมูล โดยต้องเลือกจากการสรุปผล (0) ผลที่เป็นมาตรฐาน (1) ไปสู่ ผลที่ลงรายละเอียด (2)

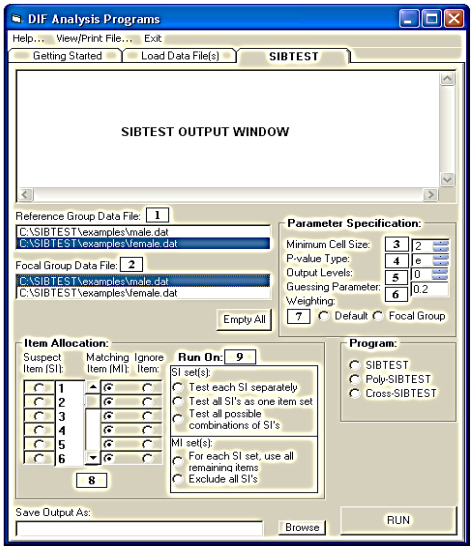
6. เลือกพารามิเตอร์การเดาสำหรับข้อสอบทั้งหมด สิ่งนี้เป็นค่าประมาณสำหรับพารามิเตอร์การเดา ค่าประมาณทั้งหมดเป็นการจับคู่แบบทดสอบ รวมทั้งแบบทดสอบพหุมิติที่มี c=0 ดังภาพที่ 2-27



ภาพที่ 2-27 การเลือกไฟล์ข้อมูลกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ (Stout & Roussos, 1999, p. 14)

7. เลือกน้ำหนักที่ใช้คำนวณทางสถิติ โดยต้องเลือกตามขนาดของกลุ่มเปรียบเทียบ (Focal Group) หรือขนาดของกลุ่มเปรียบเทียบรวมกับขนาดของกลุ่มอ้างอิง (Default)

8. เลือกข้อสอบที่สงสัยว่า (Suspected Item) มีผลต่อการทดสอบ ไม่ว่าจะแยกข้อหรือรวมทั้งหมด อีกแบบหนึ่งคือการจับคู่แบบทดสอบย่อย (Matching Item) เป็นการเลือกข้อสอบที่สงสัยทั้งรายข้อและทั้งแบบทดสอบที่มีผลต่อการทดสอบ การจับคู่แบบทดสอบย่อยจะเป็นค่าบวกทั้งหมดแต่ไม่ได้มีการทดสอบสมมติฐานเฉพาะในการเลือกข้อสอบที่ต้องสงสัย ดังภาพที่ 2-28



ภาพที่ 2-28 การเลือกคำนวณทางสถิติ (Stout & Roussos, 1999, p. 15)

9. เลือกวิธีการรันโปแกรมจาก การรันกลุ่มย่อย (All Subtest) ข้อสอบรายข้อทั้งหมดแยกเป็นข้อ (All Single Item) หรือข้อสอบรายหมวดทีละหมวด (Single Run)

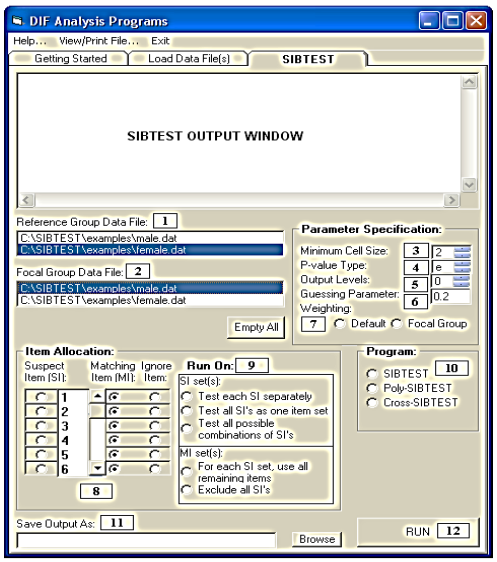
10. เลือกโปรแกรม SIBTEST

11. บันทึกผลการวิเคราะห์ข้อมูลที่ Save Output As เพื่อกำหนดชื่อข้อมูลที่ได้

12. กดปุ่ม RUN SIBTEST เพื่อรันข้อมูล

13. ถ้าต้องการต้องการลบข้อมูลทั้งหมดจาก SIBTEST ให้กดปุ่ม Empty All

การทำงานของ SIBTEST ในแต่ละครั้งผลการวิเคราะห์ข้อมูลจะส่งออกไปปรากฏที่ SIBTEST OUTPUT WINDOW อาจจะเลื่อน Scoll Bar ดูผลการวิเคราะห์ข้อมูลในโปรแกรมหรืออาจจะในเมนู View/Print File ในการแสดงผลการวิเคราะห์ข้อมูล ดังภาพที่ 2-29



ภาพที่ 2-29 การวิเคราะห์ผลในโปรแกรม SIBTEST (Stout & Roussos, 1999, p. 16)

3. ความคลาดเคลื่อนในการทดสอบสมมติฐาน

เมื่อมีการทดสอบสมมติฐานทางสถิติ จะต้องมีการตัดสินใจว่า จะยอมรับหรือปฏิเสธสมมติฐานไร้นัยสำคัญทางสถิติ (H_0) ที่ตั้งเอาไว้ ไม่ว่าเราจะยอมรับหรือปฏิเสธ ก็อาจให้เกิดความคลาดเคลื่อนในการตัดสินใจได้ ซึ่งการตัดสินใจดังกล่าว จะมีความคลาดเคลื่อนได้ 2 ประเภท คือ

3.1 ความคลาดเคลื่อนประเภท ที่ 1 (Type I Error) หมายถึง ความคลาดเคลื่อนที่เกิดจากการปฏิเสธ H_0 ทั้งที่ H_0 เป็นจริง ความน่าจะเป็นในการเกิดความคลาดเคลื่อนประเภทที่ 1 มีค่าเท่ากับ α (อัลฟา) (Freund & Walpole, 1980)

ค่า α คือ ค่าความน่าจะเป็นซึ่งก็คือ ค่าระดับนัยสำคัญทางสถิติ ที่ผู้วิจัยตั้งไว้ก่อนทำการทดสอบสมมติฐานนั้นเอง

3.2 ความคลาดเคลื่อนประเภทที่ 2 (Type II Error) หมายถึง ความคลาดเคลื่อนที่เกิดจากการยอมรับ H_0ทั้งที่ H_0 เป็นเท็จ ความน่าจะเป็นในการเกิดความคลาดเคลื่อนประเภทที่ 2 มีค่าเท่ากับ β (เบตา) (Freund & Walpole, 1980)

จากการศึกษาเอกสารงานวิจัยที่เกี่ยวข้องทำให้ทราบว่าวิธี SIBTEST เป็นวิธีที่มีความน่าเชื่อถืออีกวิธีหนึ่งซึ่งสอดคล้องกับงานวิจัยของ Roussos and Stout (1996) ได้ศึกษาผลกระทบของกลุ่มตัวอย่างขนาดเล็กและการศึกษาแบบพารามิเตอร์ โดยใช้วิธี SIBTEST และวิธีแมนเทล - แฮนด์เซล ในการทดสอบอัตราความคลาดเคลื่อนประเภทที่ 1 โดยมีวัตถุประสงค์เพื่อตรวจสอบประสิทธิภาพอัตราความคลาดเคลื่อนประเภทที่ 1 ของทั้งสองวิธีสำหรับการตรวจสอบการทำงานที่แตกต่างกันของข้อสอบ (DIF) ผลปรากฏว่าการทดสอบอัตราความคลาดเคลื่อนประเภทที่ 1 ไม่มีนัยสำคัญทางสถิติ ทั้งสองวิธีมีการแสดงประสิทธิภาพที่ใกล้เคียงกัน ถึงแม้ว่า วิธี SIBTEST จะมีประสิทธิภาพดีกว่า วิธีแมนเทล-แฮนด์เซล เพียงเล็กน้อยในทางสถิติ วิธี SIBTEST ให้ผลเพียงพอในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบภายใต้เงื่อนไขของกลุ่มตัวอย่าง (Mark, Andrea &

Keith, 2004) สอดคล้องกับการศึกษาของ พีรญา สูงเนิน, เสรี ชัดแช้ม และสมโภชน์ อเนกสุข (2551) พบว่ากลุ่มตัวอย่างที่มีขนาดใหญ่จะตรวจพบข้อสอบที่ทำหน้าที่ต่างกันของข้อสอบในกลุ่มตัวอย่างขนาดใหญ่คือเป็นร้อยละ 20 ซึ่งสรุปได้ว่า วิธี SIBTEST มีประสิทธิภาพเพียงพอในการวิเคราะห์ข้อมูล

งานวิจัยที่เกี่ยวข้องกับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธี SIBTEST มีดังนี้

พีรญา สูงเนิน, เสรี ชัดแช้ม และสมโภชน์ อเนกสุข (2552) ทำการศึกษาการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบพหุมิติ: การเปรียบเทียบระหว่างข้อสอบรายข้อกับหมวดข้อสอบโดยใช้วิธี SIBTEST โดยมีวัตถุประสงค์เพื่อเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบพหุมิติ: การเปรียบเทียบระหว่างข้อสอบรายข้อกับหมวดข้อสอบโดยใช้วิธี SIBTEST ภายใต้เงื่อนไขขนาดกลุ่มตัวอย่าง พบว่ากลุ่มตัวอย่างที่มีขนาดใหญ่จะตรวจพบข้อสอบที่ทำหน้าที่ต่างกันของข้อสอบในกลุ่มตัวอย่างขนาดใหญ่คือเป็นร้อยละ 20

Roussos and Stout (1996) ได้ศึกษาผลกระทบของกลุ่มตัวอย่างขนาดเล็กและการศึกษาแบบพารามิเตอร์ โดยใช้วิธี SIBTEST และวิธี Mantel-Hansel (MH) ในการทดสอบอัตราความคลาดเคลื่อนประเภทที่ 1 โดยมีวัตถุประสงค์เพื่อตรวจสอบประสิทธิภาพอัตราความคลาดเคลื่อนประเภทที่ 1 ของทั้งสองวิธีสำหรับการตรวจสอบการทำงานที่แตกต่างกันของข้อสอบ (DIF) ปรากฏว่าการทดสอบอัตราความคลาดเคลื่อนประเภทที่ 1 ไม่มีนัยสำคัญทางสถิติ ทั้งสองวิธีมีการแสดงประสิทธิภาพที่ใกล้เคียงกัน ถึงแม้ว่า วิธี SIBTEST จะมีประสิทธิภาพดีกว่า วิธี Mantel-Hansel (MH) เพียงเล็กน้อยในทางสถิติ

Gierl, Gotzmann, and Boughton (2004) ได้ศึกษาประสิทธิภาพของ วิธี SIBTEST เมื่อร้อยละของการทำหน้าที่ต่างกันของข้อสอบ (DIF) ในกลุ่มตัวอย่างขนาดใหญ่ โดยการทำหน้าที่ต่างกันของข้อสอบถูกใช้เพื่อระบุความแตกต่างระหว่างกลุ่มสองกลุ่มที่มีการควบคุมความสามารถ วิธี SIBTEST ได้รับความนิยมในการประเมินความสามารถของผู้สอบจากคะแนนจริง อย่างไรก็ตาม ในการทดสอบบางอย่างกลุ่มตัวอย่างอาจมีขนาดใหญ่ ในการศึกษานี้มีประสิทธิภาพของ วิธี SIBTEST อาจไม่ได้รับการประเมินอย่างละเอียด มี 4 ตัวแปรที่ใช้ในการศึกษา ได้แก่ จำนวนข้อสอบ 40 ข้อ ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (20%, 40% และ 60% ของข้อสอบที่มีการทำหน้าที่ต่างกันทั้งขนาดกลางและขนาดใหญ่) ทิศทางของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (ข้อสอบที่ทำหน้าที่ต่างกันอย่างสมดุลและไม่สมดุล) ขนาดของกลุ่มตัวอย่าง (500, 1,000, 1,500 และ 2,000 คน) และการกระจายความสามารถความแตกต่างระหว่างกลุ่ม (เท่าเทียมกันและไม่เท่ากัน) พบว่า วิธี SIBTEST มีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบอย่างเพียงพอแม้ในขณะข้อสอบ 60% มีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบให้การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ มีความสมดุลระหว่าง กลุ่มอ้างอิงและกลุ่มเปรียบเทียบ และกลุ่มตัวอย่างมีผู้สอบอย่างน้อย 1,000 คนต่อกลุ่ม วิธี SIBTEST มีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่เพียงพอในเงื่อนไข การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ไม่สมดุล 20% ด้วยตัวอย่าง 1,000 คนต่อกลุ่ม อย่างไรก็ตาม วิธี SIBTEST จะมีอัตราความ

คลาดเคลื่อนทุกเงื่อนไขในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ไม่สมดุล 40% และ 60% งานวิจัยนี้สามารถกำหนด ทิศทางการวิจัยในอนาคตต่อไปได้

Kabasakal, Arsan, Gök, and Kelecioğlu. (2014) ได้ศึกษาการเปรียบเทียบประสิทธิภาพ (อัตราความถูกต้องและอัตราความคลาดเคลื่อน) ด้วยวิธี IRT Likelihood Ratio (IRT-LR) วิธี SIBTEST และวิธี Mantel-Haenszel (MH) ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ภายใต้ขนาดของกลุ่มตัวอย่าง ความแตกต่างระหว่างกลุ่ม ความยาวของแบบทดสอบ ค่าเฉลี่ยของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) และรูปแบบพื้นฐานในการวิเคราะห์ข้อมูล พบว่า วิธี SIBTEST มีอัตราความคลาดเคลื่อนสูงที่สุดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบอย่างสม่ำเสมอ วิธี Mantel-Hansel (MH) มีอำนาจในการตรวจจับสูงในทุกสภาวะ โมเดลพื้นฐานมีผลต่ออัตราความคลาดเคลื่อนประเภทที่ 1 ความแตกต่างของความสามารถของการทำหน้าที่ต่างกันของข้อสอบไม่มีผลต่อประสิทธิภาพของของวิธี Mantel-Hansel (MH) และวิธี SIBTEST แต่มีผลต่ออัตราการใช้ของวิธี IRT Likelihood Ratio (IRT-LT)

Apinyapibal, Lawthong, and Kanjanawasee (2015) ได้ศึกษาการวิเคราะห์เปรียบเทียบประสิทธิภาพของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบสอบที่มีการตรวจให้คะแนนแบบทวิวิภาค ระหว่างวิธีการถดถอยโลจิสติก วิธี SIBTEST และ วิธี Raschtree มีวัตถุประสงค์เพื่อเปรียบเทียบอำนาจการทดสอบและ อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสำหรับแบบสอบที่มีการตรวจให้คะแนนแบบทวิวิภาค ระหว่างวิธีการถดถอยโลจิสติก วิธี SIBTEST และ วิธี Raschtree โดยการทดสอบระดับนัยสำคัญและการวัดขนาดอิทธิพล ข้อมูลที่ใช้ศึกษาจำลองภายใต้ทฤษฎีการตอบสนองข้อสอบ โมเดล 1 พารามิเตอร์ จัดทำข้อมูลตาม 4 ปัจจัย คือ (1) ค่าความยากของข้อสอบ 3 ระดับ (2) ความยาวของแบบสอบ 2 ขนาด (3) สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 2 ขนาด และ (4) ขนาดกลุ่มตัวอย่าง 3 ขนาด จำแนกตัวแปร 3 ตัว คือ เพศ สังกัดของโรงเรียน และภูมิภาค แล้ววิเคราะห์ข้อมูลด้วยวิธีการถดถอยโลจิสติก วิธี SIBTEST และ วิธี Raschtree พบว่า 1) อำนาจในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยใช้ระดับนัยสำคัญของวิธี SIBTEST มีค่าสูงกว่าวิธีการถดถอยโลจิสติกและวิธี Raschtree ภายใต้เกือบทุกเงื่อนไข และอำนาจการตรวจสอบของวิธี Raschtree มีค่าสูงกว่าวิธีการถดถอยโลจิสติกภายใต้เกือบทุกเงื่อนไข ส่วนอำนาจการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการวัดขนาดอิทธิพลของวิธี SIBTEST มีค่าสูงกว่าวิธีการถดถอยโลจิสติกภายใต้เกือบทุกเงื่อนไข 2) อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการทดสอบระดับนัยสำคัญของวิธี Raschtree มีค่าต่ำกว่าวิธีการถดถอยโลจิสติก และวิธี SIBTEST ภายใต้เกือบทุกเงื่อนไข และอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธี SIBTEST มีค่าต่ำกว่าวิธีการถดถอยโลจิสติกภายใต้เกือบทุกเงื่อนไข ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการวัดขนาดอิทธิพลของวิธีการถดถอยโลจิสติกมีค่าต่ำกว่าวิธี SIBTEST ภายใต้เกือบทุกเงื่อนไข 3) ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสำหรับข้อมูลเชิงประจักษ์ด้วยวิธี Raschtree มีอำนาจการทดสอบสูงกว่าวิธีการถดถอยโลจิสติก และวิธี SIBTEST แต่มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าเกณฑ์ที่กำหนด ที่ระดับ 10%

จากการศึกษางานวิจัยที่เกี่ยวข้องกับกับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี SIBTEST ให้ผลเพียงพอในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ อัตราการตรวจสอบขณะที่ 60% ของข้อสอบที่มีอยู่ การทำหน้าที่ต่างกันของข้อสอบได้รับการปรับให้สมดุลระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ และขนาดตัวอย่างน้อยกว่า 1,000 คน วิธี SIBTEST ยังมีอัตราการตรวจจับที่เพียงพอ การทดสอบอัตราความคลาดเคลื่อนประเภทที่ 1 วิธี SIBTEST จะมีประสิทธิภาพดีกว่า วิธี Mantel-Hansel (MH) เพียงเล็กน้อยในทางสถิติ

บทที่ 3
วิธีดำเนินการวิจัย

การวิจัยนี้มีวัตถุประสงค์เพื่อวิเคราะห์คุณภาพของข้อสอบ (a, b และ c) ตรวจสอบความเที่ยง ความตรงเชิงโครงสร้าง ฉบับก่อนและฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกันออกจากแบบทดสอบ ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ และเปรียบเทียบอัตราความคลาดเคลื่อนของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ในแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 จำนวน 8 กลุ่มสาระการเรียนรู้ ระหว่างวิธี IRT-LR กับวิธี SIBTEST ซึ่งใช้ข้อมูลทุติยภูมิ (Secondary Data) ที่ได้มาจากสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) ซึ่งผู้วิจัยแบ่งการนำเสนอเป็น 4 ระยะ ดังนี้

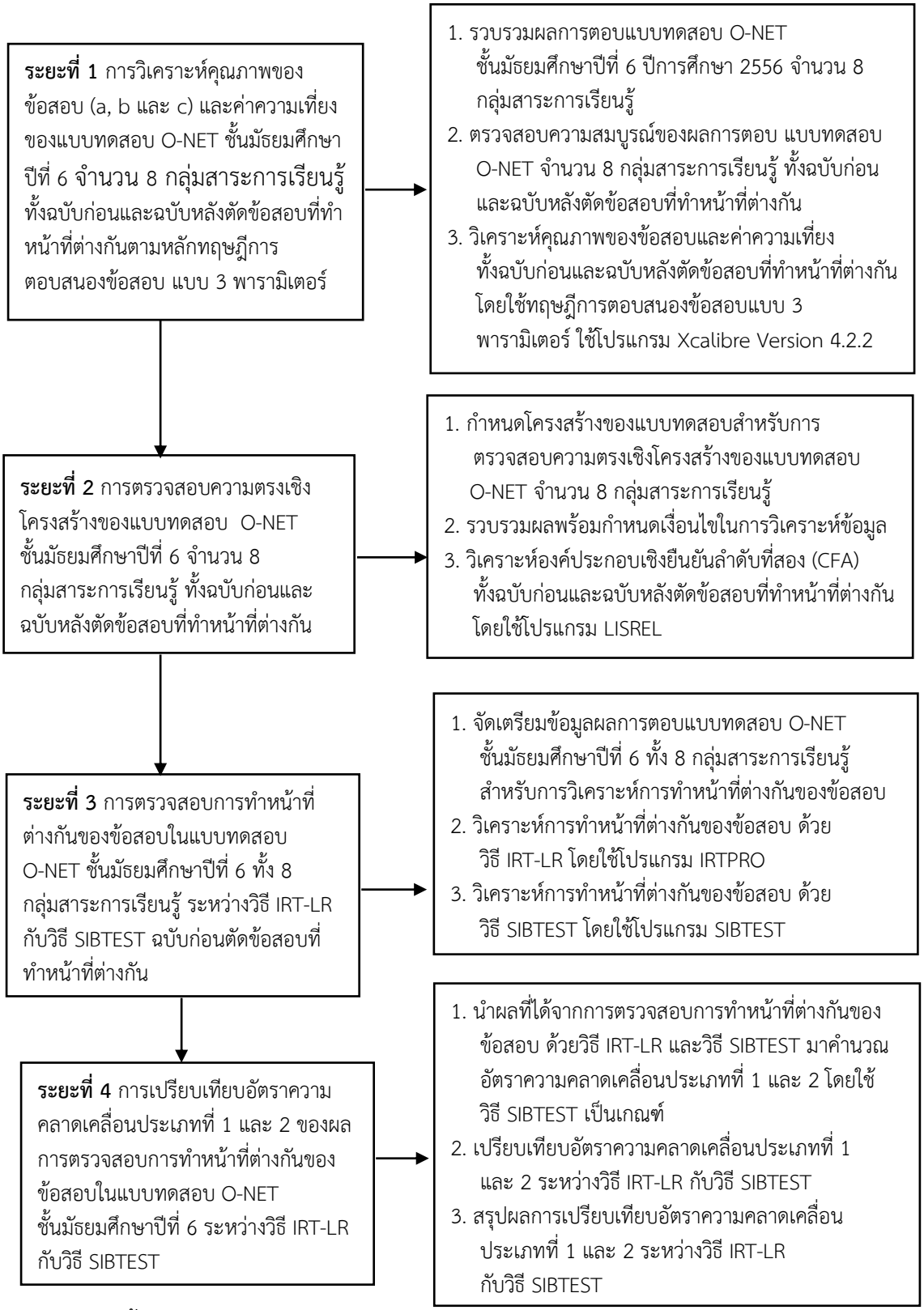
ระยะที่ 1 การวิเคราะห์คุณภาพของข้อสอบ (a, b และ c) และค่าความเที่ยงของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 จำนวน 8 กลุ่มสาระการเรียนรู้ ทั้งฉบับก่อนและฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกัน ตามหลักทฤษฎีการตอบสนองข้อสอบ แบบ 3 พารามิเตอร์

ระยะที่ 2 การตรวจสอบความตรงเชิงโครงสร้างของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 จำนวน 8 กลุ่มสาระการเรียนรู้ ทั้งฉบับก่อนและฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกัน

ระยะที่ 3 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ทั้ง 8 กลุ่มสาระการเรียนรู้ ด้วยวิธี IRT-LR และวิธี SIBTEST ฉบับก่อนตัดข้อสอบที่ทำหน้าที่ต่างกัน

ระยะที่ 4 การเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 และ 2 ของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ระหว่างวิธี IRT-LR กับวิธี SIBTEST

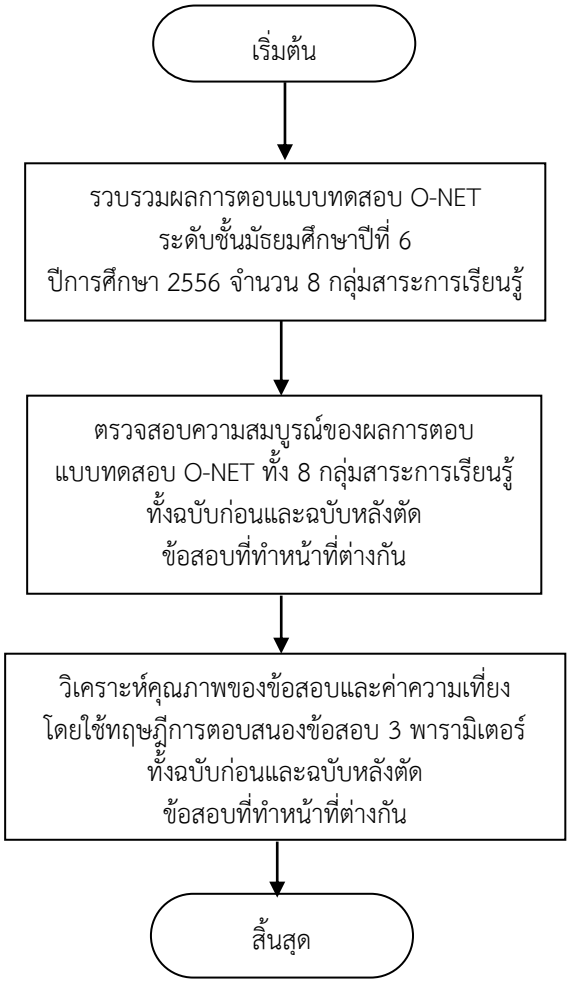
โดยมีขั้นตอนในการดำเนินการวิจัยเป็น 4 ระยะ ดังภาพที่ 3-1



ภาพที่ 3-1 ขั้นตอนการดำเนินการวิจัย

ระยะที่ 1 การวิเคราะห์คุณภาพของข้อสอบ (a, b และ c) และค่าความเที่ยงของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 จำนวน 8 กลุ่มสาระการเรียนรู้ ทั้งฉบับก่อนและฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกัน ตามหลักทฤษฎีการตอบสนองข้อสอบแบบ 3 พารามิเตอร์

การวิเคราะห์คุณภาพของข้อสอบและค่าความเที่ยงของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ฉบับก่อนและฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกันออกจากแบบทดสอบ ดังภาพที่ 3-2



ภาพที่ 3-2 ขั้นตอนวิเคราะห์คุณภาพของข้อสอบและค่าความเที่ยงของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ทั้งฉบับก่อนและฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกัน ตามหลักทฤษฎีการตอบสนองข้อสอบ แบบ 3 พารามิเตอร์

จากภาพที่ 3-2 แสดงขั้นตอนการวิเคราะห์คุณภาพของข้อสอบและค่าความเที่ยงของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ทั้งฉบับก่อนและฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกัน ดังนี้

1. ผู้วิจัยดำเนินการขอหนังสือขอความอนุเคราะห์ขอข้อมูลเพื่อการวิจัย เพื่อขอผลการตอบแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2556 ทั้ง 8 กลุ่มสาระการเรียนรู้จากสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) หรือ สทศ. มีขั้นตอนดังนี้

1.1 แบบฟอร์มขอออกหนังสือขอความอนุเคราะห์ขอข้อมูลเพื่อการวิจัยกับทางวิทยาลัยวิทยาการวิจัยและวิทยาการปัญญา มหาวิทยาลัยบูรพาเพื่อทำหนังสือขอผลการตอบแบบแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2556 ทั้ง 8 กลุ่มสาระการเรียนรู้จากสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) หรือ สทศ. โดยแนบเอกสารเอกสารแนบ เค้าโครงฉบับย่อและแบบรายงานการผ่านจริยธรรมการวิจัยประกอบการขอออกหนังสือ

1.2 นำหนังสือขอความอนุเคราะห์เพื่อการวิจัยส่งสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) หรือ สทศ. ด้วยตนเอง

2. ตรวจสอบความสมบูรณ์ของผลการตอบข้อสอบ O-NET ที่ได้รับจากสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) หรือ สทศ. ทั้ง 8 กลุ่มสาระการเรียนรู้ มีขั้นตอนดังนี้

2.1 ตรวจสอบความสมบูรณ์ของข้อมูลที่ได้รับจาก ที่ได้รับจากสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) หรือ สทศ. ทั้ง 8 กลุ่มสาระการเรียนรู้ ฉบับก่อนตัดการทำหน้าที่ต่างกัน

2.2 ทำการสุ่มกลุ่มตัวอย่างด้วยคอมพิวเตอร์ให้ได้กลุ่มตัวอย่างจำนวน 2,000 คน โดยจำแนกเพศ เป็นชายจำนวน 1,000 คน และเพศหญิงจำนวน 1,000 คน ทั้ง 8 กลุ่มสาระการเรียนรู้

3. วิเคราะห์คุณภาพของข้อสอบ (a, b และ c) และค่าความเที่ยงของแบบทดสอบ O-NET ตามหลักการของทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT) แบบ 3 พารามิเตอร์ ประกอบด้วยค่าอำนาจจำแนกของข้อสอบ (a) ค่าความยากของข้อสอบ (b) และ ค่าโอกาสการเดาของข้อสอบ (c) โดยใช้โปรแกรม Xcalibre Version 4.2.2 โดยข้อสอบจะถูกนำมาวิเคราะห์คุณภาพของข้อสอบและค่าความเที่ยงทั้งฉบับก่อนตัดกับฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกัน มีขั้นตอนดังนี้

3.1 ทำการจัดรูปแบบของข้อมูลทั้ง 8 กลุ่มสาระการเรียนรู้เพื่อวิเคราะห์ข้อมูลด้วยโปรแกรม LERTAP ในแบบทดสอบ O-NET ฉบับก่อนตัดการทำหน้าที่ต่างกัน เมื่อวิเคราะห์ข้อมูลจะได้ไฟล์ข้อมูล 2 ไฟล์ ในการวิเคราะห์ Xcaliber 4.2.2 ได้แก่ ไฟล์ข้อมูลเมทริกซ์ และไฟล์ควบคุมข้อสอบ

3.2 นำไฟล์ข้อมูลเมทริกซ์ และ ไฟล์ควบคุมข้อสอบ ของแบบทดสอบ O-NET ทั้ง 8 กลุ่มสาระการเรียนรู้ ฉบับก่อนตัดการทำหน้าที่ต่างกัน มาวิเคราะห์คุณภาพของข้อสอบ (a, b และ c) และค่าความเที่ยง ด้วยโปรแกรม Xcalibre Version 4.2.2 ด้วยการระบุไฟล์ข้อมูลเมทริกซ์ ป้อนข้อมูลเพื่อระบุคอลัมน์ของเมทริกซ์ ระบุโมเดลทฤษฎีการตอบสนองข้อสอบ ที่ใช้สำหรับการประมาณค่าพารามิเตอร์ ผู้วิจัยใช้ข้อมูลผลการตอบแบบ Dichotomous จากนั้นจึงเริ่มการวิเคราะห์ข้อมูล

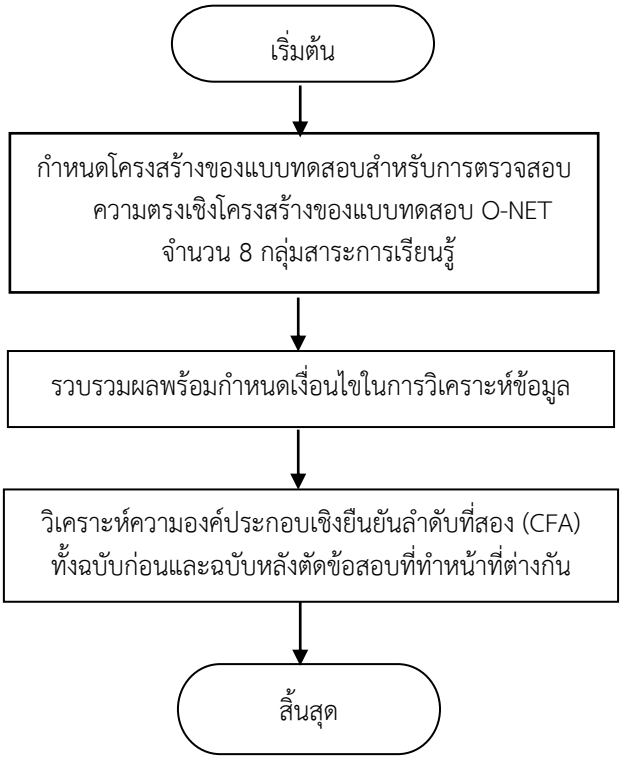
3.3 เมื่อทำการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ระหว่าง วิธี IRT-LR และวิธี SIBTEST และตัดข้อสอบที่ทำหน้าที่ต่างกันออกทั้ง 2 วิธี ได้นำข้อสอบฉบับหลังตัดการทำหน้าที่ต่างกัน มาจัดเรียงข้อมูลและทำการวิเคราะห์ข้อมูล ด้วยโปรแกรม LERTAP ในแบบทดสอบ O-NET

ฉบับก่อนตัดการทำหน้าที่ต่างกัน เมื่อวิเคราะห์ข้อมูลจะได้ไฟล์ข้อมูล 2 ไฟล์ ในการวิเคราะห์ Xcaliber 4.2.2 ได้แก่ ไฟล์ข้อมูลเมทริกซ์ และไฟล์ควบคุมข้อสอบ

3.4 นำไฟล์ข้อมูลเมทริกซ์ และไฟล์ควบคุมข้อสอบ ของแบบทดสอบ O-NET ทั้ง 8 กลุ่มสาระการเรียนรู้ ฉบับหลังตัดการทำหน้าที่ต่างกัน มาวิเคราะห์คุณภาพของข้อสอบ (a, b, c) และค่าความเที่ยง ด้วยโปรแกรม Xcalibre Version 4.2.2 ด้วยการระบุไฟล์ข้อมูลเมทริกซ์ ป้อนข้อมูลเพื่อระบุคอลัมน์ของเมทริกซ์ ระบุโมเดลทฤษฎีการตอบสนองข้อสอบ ที่ใช้สำหรับการประมาณค่าพารามิเตอร์ ผู้วิจัยใช้ข้อมูลผลการตอบแบบ Dichotomous จากนั้นจึงเริ่มการวิเคราะห์

ระยะที่ 2 การตรวจสอบความตรงเชิงโครงสร้างของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 จำนวน 8 กลุ่มสาระการเรียนรู้ ทั้งฉบับก่อนและฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกัน

การตรวจสอบความตรงเชิงโครงสร้างของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2556 ผู้วิจัยทราบลักษณะโครงสร้างของแบบทดสอบตามสาระการเรียนรู้ของทั้ง 8 กลุ่มสาระการเรียนรู้ จากทางสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) หรือ สทศ. ผู้วิจัยนำมาตรวจสอบความตรงเชิงโครงสร้างทั้งฉบับก่อนและฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกัน ดังภาพที่ 3-3



ภาพที่ 3-3 ขั้นตอนการการตรวจสอบความตรงเชิงโครงสร้างของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 จำนวน 8 กลุ่มสาระการเรียนรู้ ทั้งฉบับก่อนและฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกัน

จากภาพที่ 3-3 แสดงขั้นตอนการตรวจสอบความตรงเชิงโครงสร้างของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ทั้งฉบับก่อนและฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกัน ดังนี้

1. ผู้วิจัยกำหนดโครงสร้างของแบบทดสอบ ตามสาระการเรียนรู้ทั้ง 8 กลุ่มสาระการเรียนรู้ตามโครงสร้าง ดังตารางที่ 3-1

ตารางที่ 3-1 โครงสร้างองค์ประกอบของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ฉบับก่อนและฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกัน

องค์ประกอบ	องค์ประกอบย่อย	ก่อนตัด DIF	หลังตัด DIF
		จำนวน (ข้อ)	จำนวน (ข้อ)
1. ภาษาไทย	1.1 การอ่าน	20 ข้อ (1-20)	16 ข้อ (1-16)
	1.2 การเขียน	20 ข้อ (21-33)	16 ข้อ (17-28)
	1.3 การฟัง การดูและการพูด	2 ข้อ (34,35)	2 ข้อ (29-30)
	1.4 หลักหารใช้ภาษา	13 ข้อ (35-55)	12 ข้อ (31-46)
	1.5 วรรณคดีและวรรณกรรม	15 ข้อ (56-70)	11 ข้อ (47-57)
	รวม	70 ข้อ	57 ข้อ
2. สังคมศึกษา ศาสนาและวัฒนาธรรม	2.1 ศาสนา ศิลธรรม จริยธรรม	13 ข้อ (1-13)	10 ข้อ (1-10)
	2.2 หน้าที่พลเมือง วัฒนธรรมและการดำเนินชีวิตวนสังคม	13 ข้อ (14-26)	8 ข้อ (11-18)
	2.3 เศรษฐศาสตร์	13 ข้อ (27-39)	11 ข้อ (19-29)
	2.4 ประวัติศาสตร์	13 ข้อ (40-52)	12 ข้อ (30-41)
	2.5 ภูมิศาสตร์	13 ข้อ (53-65)	11 ข้อ (42-52)
	รวม	65 ข้อ	52 ข้อ

ตารางที่ 3-1 (ต่อ)

องค์ประกอบ	องค์ประกอบย่อย	ก่อนตัด DIF	หลังตัด DIF
		จำนวน (ข้อ)	จำนวน (ข้อ)
3. ภาษาอังกฤษ	3.1 ภาษาเพื่อการสื่อสาร	21 ข้อ (1-21)	21 ข้อ (1-21)
	3.2 ภาษาและวัฒนธรรม	18 ข้อ (22-39)	17 ข้อ (22-38)
	3.3 ภาษากับความสัมพันธ์กับกลุ่มสาระการเรียนรู้อื่น	20 ข้อ (40-59)	19 ข้อ (39-57)
	3.4 ภาษากับความสัมพันธ์กับชุมชนและโลก	21 ข้อ (60-80)	20 ข้อ (58-77)
	รวม	80 ข้อ	77 ข้อ
4. คณิตศาสตร์	4.1 จำนวนและการดำเนินการ	7 ข้อ (1-7)	6 ข้อ (1-6)
	4.2 การวัด	4 ข้อ (8-11)	4 ข้อ (7-10)
	4.3 พิชคณิต	15 ข้อ (12-26)	14 ข้อ (11-24)
	4.4 การวิเคราะห์ข้อมูลความน่าจะเป็น	6 ข้อ (27-32)	6 ข้อ (25-30)
	รวม	32 ข้อ	30 ข้อ
5. วิทยาศาสตร์	5.1 สิ่งมีชีวิตกับกระบวนการดำรงชีวิต	11 ข้อ (1-11)	9 ข้อ (1-9)
	5.2 ชีวิตกับสิ่งแวดล้อม	7 ข้อ (12-18)	6 ข้อ (10-15)
	5.3 สารและสมบัติของสาร	18 ข้อ (19-36)	16 ข้อ (16-31)
	5.4 แรงและการเคลื่อนที่	10 ข้อ (37-46)	8 ข้อ (32-39)

ตารางที่ 3-1 (ต่อ)

องค์ประกอบ	องค์ประกอบย่อย	ก่อนตัด DIF	หลังตัด DIF
		จำนวน (ข้อ)	จำนวน (ข้อ)
	5.5 พลังงาน	8 ข้อ (47-54)	8 ข้อ (40-47)
	5.6 กระบวนการเปลี่ยนแปลงของโลก	9 ข้อ (55-63)	8 ข้อ (48-55)
	5.7 ดาราศาสตร์และอวกาศ	9 ข้อ (64-72)	9 ข้อ (56-64)
	5.8 ธรรมชาติของวิทยาศาสตร์และเทคโนโลยี	8 ข้อ (73-80)	5 ข้อ (65-69)
	รวม	80 ข้อ	69 ข้อ
6. สุขศึกษาและพลศึกษา	6.1 การเจริญเติบโตและพัฒนาการของมนุษย์	6 ข้อ (1-6)	5 ข้อ (1-5)
	6.2 ชีวิตและครอบครัว	4 ข้อ (7-10)	3 ข้อ (6-8)
	6.3 การเคลื่อนไหว การออกกำลังกาย การเล่นเกม กีฬาไทย และกีฬาสากล	4 ข้อ (11-14)	2 ข้อ (9-10)
	6.4 การสร้างเสริมสุขภาพ สมรรถภาพและการป้องกันโรค	4 ข้อ (15-18)	3 ข้อ (11-13)
	6.5 ความปลอดภัยในชีวิต	14 ข้อ (19-32)	8 ข้อ (14-21)
	รวม	32 ข้อ	21 ข้อ
7. ศิลปะ	7.1 ทัศนศิลป์	10 ข้อ (1-10)	8 ข้อ (1-8)
	7.2 ดนตรี	10 ข้อ (11-20)	8 ข้อ (9-16)

ตารางที่ 3-1 (ต่อ)

องค์ประกอบ	องค์ประกอบย่อย	ก่อนตัด DIF	หลังตัด DIF
		จำนวน (ข้อ)	จำนวน (ข้อ)
	7.3 นาฏศิลป์	10 ข้อ (21-30)	8 ข้อ (17-24)
	รวม	30 ข้อ	24 ข้อ
8. การงานอาชีพและเทคโนโลยี	8.1 การดำรงชีวิตและครอบครัว	22 ข้อ (1-22)	18 ข้อ (1-18)
	8.2 การออกแบบและเทคโนโลยี	10 ข้อ (23-32)	10 ข้อ (19-28)
	8.3 เทคโนโลยีสารสนเทศและการสื่อสาร	6 ข้อ (33-38)	4 ข้อ (29-32)
	8.4 การอาชีพ	3 ข้อ (39-41)	3 ข้อ (33-35)
	รวม	41 ข้อ	35 ข้อ

จากตารางที่ 3-1 ได้กำหนดโครงสร้างของแบบทดสอบแต่ละกลุ่มสาระการเรียนรู้ได้จัดโครงสร้างตามกลุ่มสาระการเรียนรู้ของทั้ง 8 กลุ่มสาระการเรียนรู้ โดยจำนวนข้อของแต่ละองค์ประกอบย่อย กำหนดตามสาระการเรียนรู้จากทางสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) หรือ สทศ. เป็นผู้ให้ข้อมูล ดังต่อไปนี้

1) วิชาภาษาไทย ประกอบด้วย 5 สาระการเรียนรู้ จำนวน 70 ข้อ ได้แก่

1.1) การอ่าน จำนวน 20 ข้อ
1.2) การเขียน จำนวน 20 ข้อ
1.3) การฟัง การดูและการอ่าน จำนวน 2 ข้อ
1.4) หลักการใช้ภาษา จำนวน 13 ข้อ
1.5) วรรณคดีและวรรณกรรม จำนวน 15 ข้อ

2) วิชาสังคมศึกษา ศาสนา และวัฒนธรรม ประกอบด้วย 5 สาระการเรียนรู้ จำนวน 65 ข้อ ได้แก่

2.1) ศาสนา ศิลธรรม จริยธรรม จำนวน 13 ข้อ
2.2) หน้าที่พลเมือง วัฒนธรรม และการดำเนินชีวิตในสังคม จำนวน 13 ข้อ
2.3) เศรษฐศาสตร์ จำนวน 13 ข้อ
2.4) ประวัติศาสตร์ จำนวน 13 ข้อ

2.5) ภูมิศาสตร์ จำนวน 13 ข้อ

3) วิชาภาษาอังกฤษ ประกอบด้วย 4 สาระการเรียนรู้ จำนวน 80 ข้อ ได้แก่

3.1) ภาษาเพื่อการสื่อสาร จำนวน 21 ข้อ

3.2) ภาษาและวัฒนธรรม จำนวน 18 ข้อ

3.3) ภาษากับความสัมพันธ์กับกลุ่มสาระการเรียนรู้อื่น จำนวน 20 ข้อ

3.4) ภาษากับความสัมพันธ์กับชุมชนและโลก จำนวน 21 ข้อ

4) วิชาคณิตศาสตร์ ประกอบด้วย 6 สาระการเรียนรู้ ใช้ในแบบทดสอบ 4 สาระการเรียนรู้ จำนวน 32 ข้อ ได้แก่

4.1) จำนวนและการดำเนินการ จำนวน 7 ข้อ

4.2) การวัด จำนวน 4 ข้อ

4.3) พิชคณิต จำนวน 15 ข้อ

4.4) การวิเคราะห์ข้อมูลและความน่าจะเป็น จำนวน 6 ข้อ

5) วิชาวิทยาศาสตร์ ประกอบด้วย 8 สาระการเรียนรู้ จำนวน 80 ข้อ ได้แก่

5.1) สิ่งมีชีวิตกับกระบวนการดำรงชีวิต จำนวน 11 ข้อ

5.2) ชีวิตกับสิ่งแวดล้อม จำนวน 7 ข้อ

5.3) สารและสมบัติของสาร จำนวน 18 ข้อ

5.4) แรงและการเคลื่อนที่ จำนวน 10 ข้อ

5.5) พลังงาน จำนวน 8 ข้อ

5.6) กระบวนการเปลี่ยนแปลงของโลก จำนวน 9 ข้อ

5.7) ดาราศาสตร์และอวกาศ จำนวน 9 ข้อ

5.8) ธรรมชาติของวิทยาศาสตร์และเทคโนโลยี จำนวน 8 ข้อ

6) วิชาสุขศึกษาและพลศึกษา ประกอบด้วย 5 สาระการเรียนรู้ จำนวน 32 ข้อ ได้แก่

6.1) การเจริญเติบโตและพัฒนาการของมนุษย์ จำนวน 6 ข้อ

6.2) ชีวิตและครอบครัว จำนวน 4 ข้อ

6.3) การเคลื่อนไหว การออกกำลังกาย การเล่นเกม กีฬาไทยและกีฬาสากล จำนวน 4 ข้อ

6.4) การสร้างเสริมสุขภาพ สมรรถภาพและการป้องกันโรค จำนวน 4 ข้อ

6.5) ความปลอดภัยในชีวิต จำนวน 14 ข้อ

7) วิชาศิลปะ ประกอบด้วย 3 สาระการเรียนรู้ จำนวน 30 ข้อ ได้แก่

7.1) ทัศนศิลป์ จำนวน 10 ข้อ

7.2) ดนตรี จำนวน 10 ข้อ

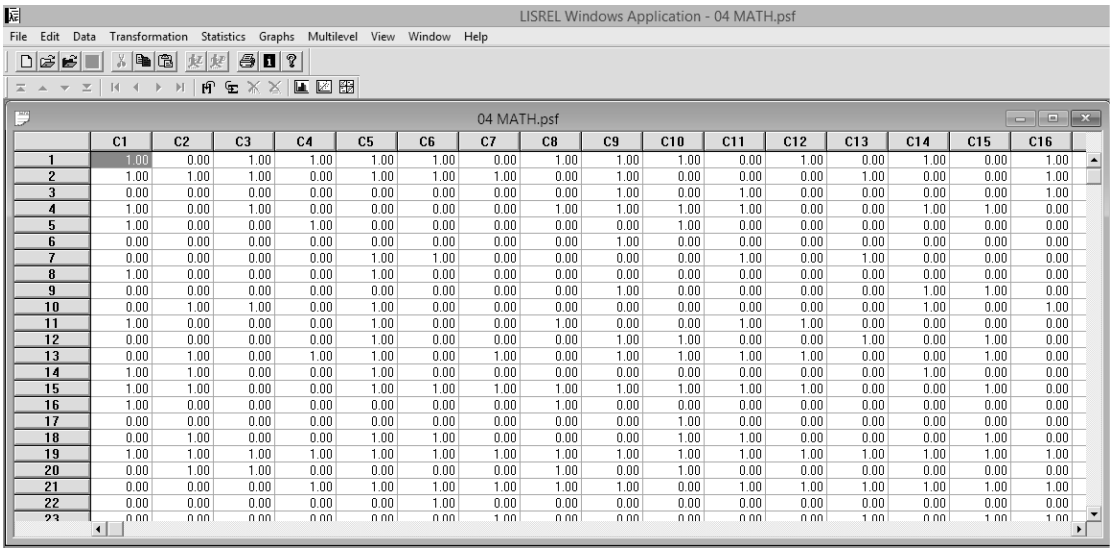
7.3) นาฏศิลป์ จำนวน 10 ข้อ

8) วิชาการงานอาชีพและเทคโนโลยี ประกอบด้วย 4 สาระการเรียนรู้ จำนวน 41 ข้อได้แก่

8.1) การดำรงชีวิตและครอบครัว	จำนวน	22	ข้อ
8.2) การออกแบบและเทคโนโลยี	จำนวน	10	ข้อ
8.3) เทคโนโลยีสาระสนเทศและการสื่อสาร	จำนวน	6	ข้อ
8.4) การอาชีพ	จำนวน	3	ข้อ

2. รวบรวมผลพร้อมกำหนดเงื่อนไขในการวิเคราะห์ข้อมูล แบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ทั้ง 8 กลุ่มสาระการเรียนรู้ ฉบับก่อนตัดข้อสอบที่ทำหน้าที่ต่างกัน ดังนี้

2.1 ผู้วิจัยเตรียมข้อมูลให้อยู่ในรูปแบบการนำเข้าข้อมูลของข้อสอบในแต่ละข้อ โดยใช้โปรแกรม SPSS เพื่อนำเข้าข้อมูลให้อยู่ในรูป .psf เมื่อนำเข้าข้อมูลโดยใช้ โปรแกรม LISREL ดังภาพที่ 3-4



	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16
1	1.00	0.00	1.00	1.00	1.00	1.00	0.00	1.00	1.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00
2	1.00	1.00	1.00	0.00	1.00	1.00	1.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	1.00
3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00
4	1.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00	1.00	0.00
5	1.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
7	0.00	0.00	0.00	0.00	1.00	1.00	0.00	0.00	0.00	0.00	1.00	0.00	1.00	0.00	0.00	0.00
8	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00	1.00	0.00
10	0.00	1.00	1.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	1.00
11	1.00	0.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	1.00	0.00	0.00	0.00	0.00
12	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	1.00	0.00	0.00	1.00	0.00	1.00	0.00
13	0.00	1.00	0.00	1.00	1.00	0.00	1.00	0.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00	0.00
14	1.00	1.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00
15	1.00	1.00	0.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00	0.00
16	1.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
17	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
18	0.00	1.00	0.00	0.00	1.00	1.00	0.00	0.00	0.00	1.00	1.00	0.00	0.00	0.00	1.00	0.00
19	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
20	0.00	1.00	1.00	0.00	0.00	0.00	0.00	1.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
21	0.00	0.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00
22	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
23	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	1.00	1.00

ภาพที่ 3-4 ตัวอย่างการเตรียมไฟล์ข้อมูล วิเคราะห์ความองค์ประกอบเชิงยืนยันลำดับที่สอง (CFA) ทั้งฉบับก่อนตัดกับฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกัน

2.2 กำหนดคำสั่งที่ใช้ในการวิเคราะห์ข้อมูลของแบบทดสอบ แบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ทั้ง 8 กลุ่มสาระการเรียนรู้ ฉบับก่อนตัดข้อสอบที่ทำหน้าที่ต่างกัน

3. วิเคราะห์ความตรงเชิงโครงสร้างด้วยวิธีการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสอง (Second-Order Confirmatory Factor Analysis :CFA) โดยใช้โปรแกรม LISREL โดยค่าสถิติที่ใช้ตรวจสอบความสอดคล้องของโมเดลองค์ประกอบกับข้อมูลผลการตอบแบบทดสอบของผู้สอบ คือ ค่าสถิติ ไค-แสควร์ (Chi-square: x^2) ค่าดัชนีวัดระดับความกลมกลืน (Goodness of Fit Index: GFI) ค่าดัชนีวัดระดับความกลมกลืนที่ปรับแก้แล้ว (Adjust Goodness of Fit Index: AGFI) ค่าดัชนีวัดระดับความกลมกลืนเปรียบเทียบ(Comparative Fit Index: CFI) ค่าดัชนีวัดความสอดคล้องปกติ

(Normed Fit Index: NFI) และค่ารากของค่าเฉลี่ยกำหลังสองของความคลาดเคลื่อนโดยประมาณ (Root Mean Squared Error of Approximation: RMSEA)

3.1 แบบทดสอบ O-NET ทั้ง 8 กลุ่มสาระการเรียนรู้ฉบับก่อนตัดข้อสอบที่ทำหน้าที่ต่างกัน มีค่าพารามิเตอร์บางค่าที่ไม่แตกต่างจากศูนย์หรือมีทิศทางไม่ตรงกับทฤษฎีที่กำหนดไว้จึงจำเป็นต้องปรับโมเดล

4. เมื่อตัดข้อสอบที่ทำหน้าที่ต่างกันออกจากแบบทดสอบ ผู้วิจัยจึงนำมาตรวจสอบความตรงเชิงโครงสร้างอีกครั้ง

4.1 กำหนดโครงสร้างของแบบทดสอบแต่ละกลุ่มสาระการเรียนรู้ได้จัดโครงสร้างตามกลุ่มสาระการเรียนรู้ของทั้ง 8 กลุ่มสาระการเรียนรู้ ฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกัน โดยจำนวนข้อของแต่ละองค์ประกอบย่อย กำหนดตามสาระการเรียนรู้จากทางสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) หรือ สทศ. เป็นผู้ให้ข้อมูล

5. รวบรวมผลพร้อมกำหนดเงื่อนไขในการวิเคราะห์ข้อมูล แบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ทั้ง 8 กลุ่มสาระการเรียนรู้ ฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกัน ดังนี้

5.1 ผู้วิจัยเตรียมข้อมูลให้อยู่ในรูปแบบการนำเข้าข้อมูลของข้อสอบในแต่ละข้อ โดยใช้โปรแกรม SPSS เพื่อนำเข้าข้อมูลให้อยู่ในรูป .psf เมื่อนำเข้าข้อมูลโดยใช้ โปรแกรม LISREL

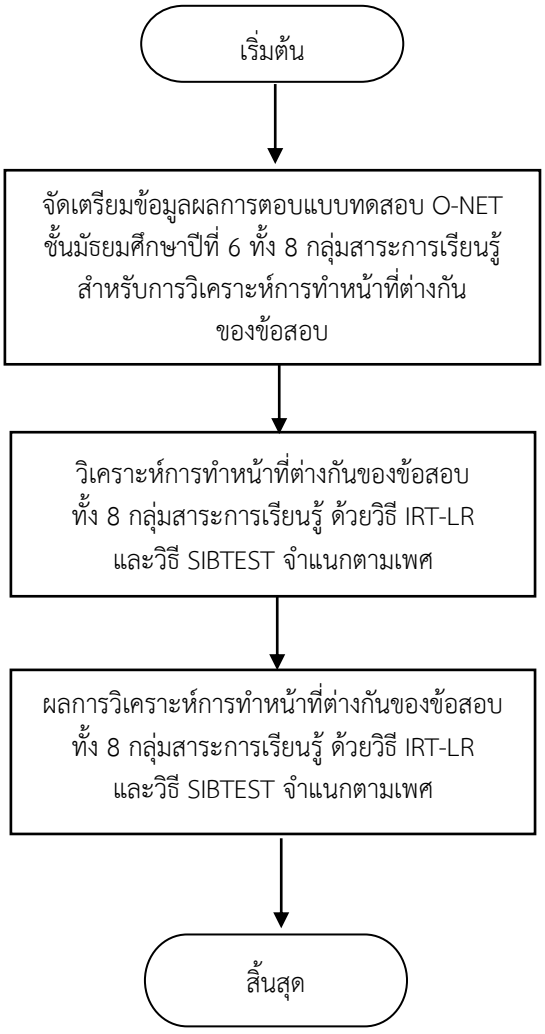
5.2 กำหนดคำสั่งที่ใช้ในการวิเคราะห์ข้อมูลของแบบทดสอบ แบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ทั้ง 8 กลุ่มสาระการเรียนรู้ ฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกัน

6. วิเคราะห์ความตรงเชิงโครงสร้างด้วยวิธีการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสอง (Second-Order Confirmatory Factor Analysis :CFA) โดยใช้โปรแกรม LISREL โดยค่าสถิติที่ใช้ตรวจสอบความสอดคล้องของโมเดลองค์ประกอบกับข้อมูลผลการตอบแบบทดสอบของผู้สอบ คือ ค่าสถิติ ไค-แสควร์ (Chi-square: $\boldsymbol{x}^2$) ค่าดัชนีวัดระดับความกลมกลืน (Goodness of Fit Index: GFI) ค่าดัชนีวัดระดับความกลมกลืนที่ปรับแก้แล้ว (Adjust Goodness of Fit Index: AGFI) ค่าดัชนีวัดระดับความกลมกลืนเปรียบเทียบ(Comparative Fit Index: CFI) ค่าดัชนีวัดความสอดคล้องปกติ (Normed Fit Index: NFI) และค่ารากของค่าเฉลี่ยกำลังสองของความคลาดเคลื่อนโดยประมาณ (Root Mean Squared Error of Approximation: RMSEA)

6.1 แบบทดสอบ O-NET ทั้ง 8 กลุ่มสาระการเรียนรู้ฉบับก่อนตัดข้อสอบที่ทำหน้าที่ต่างกัน มีค่าพารามิเตอร์บางค่าที่ไม่แตกต่างจากศูนย์หรือมีทิศทางไม่ตรงกับทฤษฎีที่กำหนดไว้จึงจำเป็นต้องปรับโมเดล

ระยะที่ 3 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ทั้ง 8 กลุ่มสาระการเรียนรู้ ด้วยวิธี IRT-LR และวิธี SIBTEST ฉบับก่อนตัดข้อสอบที่ทำหน้าที่ต่างกัน

การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ระหว่างวิธี IRT-LR กับ วิธี SIBTEST ดังภาพที่ 3-5



ภาพที่ 3-5 แสดงขั้นตอนการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ทั้ง 8 กลุ่มสาระการเรียนรู้ ด้วยวิธี IRT-LR และวิธี SIBTEST ฉบับก่อนตัดข้อสอบที่ทำหน้าที่ต่างกัน

จากภาพที่ 3-5 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ทั้ง 8 กลุ่มสาระการเรียนรู้ ด้วยวิธี IRT-LR และวิธี SIBTEST ฉบับก่อนตัดข้อสอบที่ทำหน้าที่ต่างกัน จำแนกตามเพศ ดังนี้

1. จัดเตรียมข้อมูลผลการตอบแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ทั้ง 8 กลุ่มสาระการเรียนรู้ สำหรับการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ

ผู้วิจัยได้กำหนดกลุ่มตัวอย่างซึ่งกลุ่มตัวอย่างที่ใช้ในงานวิจัยคือ นักเรียนชั้นมัธยมศึกษาปีที่ 6 ซึ่งตอบแบบทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน (O-NET) ประจำปีการศึกษา 2556 ได้มาโดยวิธีการสุ่มอย่างง่ายโดยใช้คอมพิวเตอร์ในการสุ่ม โดยจำแนกเป็นนักเรียนชายและนักเรียนหญิงกลุ่มละเท่าๆ กัน จำนวน 2,000 คน เนื่องจากเป็นกลุ่มตัวอย่างที่สามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้อย่างถูกต้องและมีความพอเพียงสำหรับการประมาณค่าพารามิเตอร์ได้อย่างถูกต้องเหมาะสม (Mark, Andrea & Keith, 2004) ซึ่งกำหนดให้ตัวแปรเพศชายมีค่าเท่ากับ 1 และตัวแปรเพศหญิงมีค่าเท่ากับ 2 ทั้ง 8 กลุ่มสาระการเรียนรู้ โดยการเปรียบเทียบความแตกต่างของพารามิเตอร์ในกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ มีวิธีการดำเนินการตามขั้นตอน ดังนี้

1.1 เตรียมไฟล์ข้อมูลสำหรับการวิเคราะห์ ด้วยวิธี IRT-LR

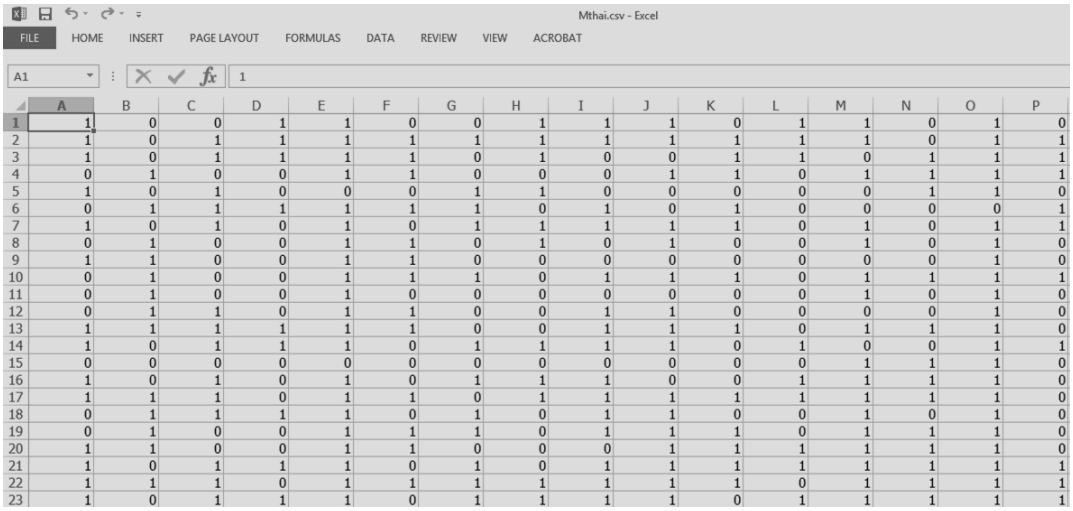
การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธี IRT ผู้วิจัยได้เตรียมข้อมูลในรูปแบบของไฟล์ .sav เพื่อวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยโปรแกรม IRTPRO ในการวิเคราะห์ข้อมูล ข้อมูลป้อนเข้า คือ ข้อมูลที่ให้ผลคะแนนแบบ 0, 1 ประกอบด้วย เพศ (Gender) และผลการตอบแบบทดสอบของผู้สอบ (Response) ดังภาพที่ 3-6

	GENDER	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14
1	1	1	0	1	0	1	1	1	1	0	1	0	1	0	0
2	1	0	1	1	1	1	0	1	1	1	0	0	1	0	0
3	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0
4	1	0	0	1	0	1	0	0	0	0	1	1	0	0	0
5	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0
6	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1
7	1	0	1	0	0	0	0	1	1	0	0	0	0	1	0
8	1	0	0	1	0	0	0	1	0	0	0	0	0	0	0
9	1	0	0	0	0	0	0	0	0	0	1	1	0	0	0
10	1	0	0	0	1	1	0	1	0	0	1	0	1	0	0
11	1	1	0	1	0	0	0	1	0	0	0	0	0	1	0
12	1	0	1	0	0	0	0	1	0	0	0	1	0	0	0
13	1	1	0	0	1	0	1	1	0	1	0	1	0	0	0
14	1	0	0	1	1	0	0	1	0	0	1	0	0	0	0
15	1	1	0	1	1	0	0	1	1	1	0	1	0	1	1
16	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0
17	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	1	0	0	0	1	0	0	1	1	0	0	1	0	0	0
19	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
20	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0

ภาพที่ 3-6 ตัวอย่างการจัดไฟล์ข้อมูลสำหรับวิเคราะห์ด้วยวิธี IRT-LR ในรูปแบบไฟล์ .sav

1.2 เตรียมไฟล์ข้อมูลสำหรับการวิเคราะห์ ด้วยวิธี SIBTEST

การตรวจสอบทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธี SIBTEST ผู้วิจัยได้เตรียมข้อมูลให้อยู่ในรูปแบบไฟล์ .cvs เพื่อแปลงไฟล์ให้อยู่ในรูป .dat เพื่อใช้วิเคราะห์ข้อมูลในโปรแกรม SIBTEST โดยแยกเพศชายและเพศหญิง ผลการตอบแบบทดสอบเป็นแบบเอกมิติ (Unidimensional) ดังภาพที่ 3-7



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	1	0	0	1	1	0	0	1	1	1	0	1	1	0	1	0
2	1	0	1	1	1	1	1	1	1	1	1	1	1	0	1	1
3	1	0	1	1	1	1	0	1	0	0	1	1	0	1	1	1
4	0	1	0	0	1	1	0	0	0	1	1	0	1	1	1	1
5	1	0	1	0	0	0	1	1	0	0	0	0	0	1	1	0
6	0	1	1	1	1	1	1	0	1	0	1	0	0	0	0	1
7	1	0	1	0	1	0	1	1	1	1	1	0	1	0	1	1
8	0	1	0	0	1	1	0	1	0	1	0	0	1	0	1	0
9	1	1	0	0	1	1	0	0	0	0	0	0	0	0	1	0
10	0	1	0	0	1	1	1	0	1	1	1	0	1	1	1	1
11	0	1	0	0	1	0	0	0	0	0	0	0	1	0	1	0
12	0	1	1	0	1	1	0	0	1	1	0	0	0	0	1	0
13	1	1	1	1	1	1	0	0	1	1	1	0	1	1	1	0
14	1	0	1	1	1	0	1	1	1	1	0	1	0	0	1	1
15	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0
16	1	0	1	0	1	0	1	1	1	0	0	1	1	1	1	0
17	1	1	1	0	1	1	0	1	1	1	1	1	1	1	1	0
18	0	1	1	1	1	0	1	0	1	1	0	0	1	0	1	0
19	0	1	0	0	1	1	1	0	1	1	1	0	1	1	1	0
20	1	1	0	0	1	1	0	0	0	1	1	1	1	1	1	0
21	1	0	1	1	1	0	1	0	1	1	1	1	1	1	1	1
22	1	1	1	0	1	1	1	1	1	1	1	0	1	1	1	1
23	1	0	1	1	1	0	1	1	1	1	0	1	1	1	1	1

ภาพที่ 3-7 ตัวอย่างการจัดไฟล์ข้อมูลสำหรับการวิเคราะห์ด้วยวิธี SIBTEST ในรูปไฟล์ .dat

2. วิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ ทั้ง 8 กลุ่มสาระการเรียนรู้ ด้วยวิธี IRT-LR และวิธี SIBTEST จำแนกตามเพศ

2.1 การทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธี IRT-LR

การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ด้วยวิธี IRT-LR จะประเมินความสำคัญระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ ในความแตกต่างของแต่ละพารามิเตอร์ ดังนี้

1. เลือกวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบตามทฤษฎีการตอบสนองข้อสอบแบบเอกมิติ (Unidimensional IRT)

2. กำหนดตัวแปรเพศ ให้กับแบบทดสอบ O-NET ทั้ง 8 กลุ่มสาระการเรียนรู้ ดังนี้

2.1 แบบทดสอบภาษาไทย ให้เพศหญิงเป็นกลุ่มอ้างอิง (Reference Group: R) และเพศชายเป็นกลุ่มเปรียบเทียบ (Focal Group: F) (สุพัฒนา หอมบุปผา, ไพรัตน์ วงษ์นาม และสมพงษ์ ปั้นหุ่น, 2556, หน้า 21)

2.2 แบบทดสอบสังคมศึกษา ศาสนาและวัฒนธรรม ให้เพศชายเป็นกลุ่มอ้างอิง (Reference Group: R) และเพศหญิงเป็นกลุ่มเปรียบเทียบ (Focal Group: F) (ACAR, 2011, p. 285)

2.3 แบบทดสอบภาษาอังกฤษ ให้เพศหญิงเป็นกลุ่มอ้างอิง (Reference Group: R) และเพศชายเป็นกลุ่มเปรียบเทียบ (Focal Group: F) (Öğretmen, 2015, p. 1510)

2.4 แบบทดสอบคณิตศาสตร์ ให้เพศชายเป็นกลุ่มอ้างอิง (Reference Group: R) และเพศหญิงเป็นกลุ่มเปรียบเทียบ (Focal Group: F) (Sokuler & Justman, 2016, p. 230)

2.5 แบบทดสอบวิทยาศาสตร์ ให้เพศชายเป็นกลุ่มอ้างอิง (Reference Group: R) และเพศหญิงเป็นกลุ่มเปรียบเทียบ (Focal Group: F) (Jansen, Schroeders, & Lüdtke, 2014, p. 11)

2.6 แบบทดสอบสุขศึกษาและพลศึกษา ให้เพศชายเป็นกลุ่มอ้างอิง (Reference Group: R) และเพศหญิงเป็นกลุ่มเปรียบเทียบ (Focal Group: F) (Abd-El-Fattah, Al-Sinani, Shourbagi & Fakhroo, 2014, p. 29)

2.7 แบบทดสอบศิลปะ ให้เพศหญิงเป็นกลุ่มอ้างอิง (Reference Group: R) และเพศชายเป็นกลุ่มเปรียบเทียบ (Focal Group: F) (Caballero, Casares & Caballero, 2017, p. 529)

2.8 แบบทดสอบการงานอาชีพและเทคโนโลยี ให้เพศหญิงเป็นกลุ่มอ้างอิง (Reference Group: R) และเพศชายเป็นกลุ่มเปรียบเทียบ (Focal Group: F) (Nenty, 2015, p. 84)

3. ลงรายละเอียดในเมนู Description

4. เลือกโมเดลและตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ เพื่อวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ โดยกำหนดให้ทำการวิเคราะห์ข้อสอบทุกข้อเป็นข้อสอบของแบบทดสอบจากนั้นพิจารณาว่าข้อสอบข้อใด เป็นข้อสอบที่ทำหน้าที่ต่างกัน โดยดูจากค่า *p*-value ที่มีนัยสำคัญทางสถิติที่ระดับ .05

2.2 การทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธี SIBTEST

การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ด้วยวิธี SIBTEST จะประเมินความสำคัญระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ ในความแตกต่างของแต่ละพารามิเตอร์ ได้ดังนี้

1. เลือกวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบตามทฤษฎีการตอบสนองข้อสอบแบบเอกมิติ (Unidimensional)

2. กำหนดตัวแปรเพศ ให้กับแบบทดสอบ O-NET ทั้ง 8 กลุ่มสาระการเรียนรู้

3. กำหนดค่า Mininum Cell Size

4. เลือกระดับผลการวิเคราะห์ข้อมูล

5. เลือก Common Guessing Parameter

6. เลือกค่าน้ำหนักในการวิเคราะห์ข้อมูลของวิธี SIBTEST

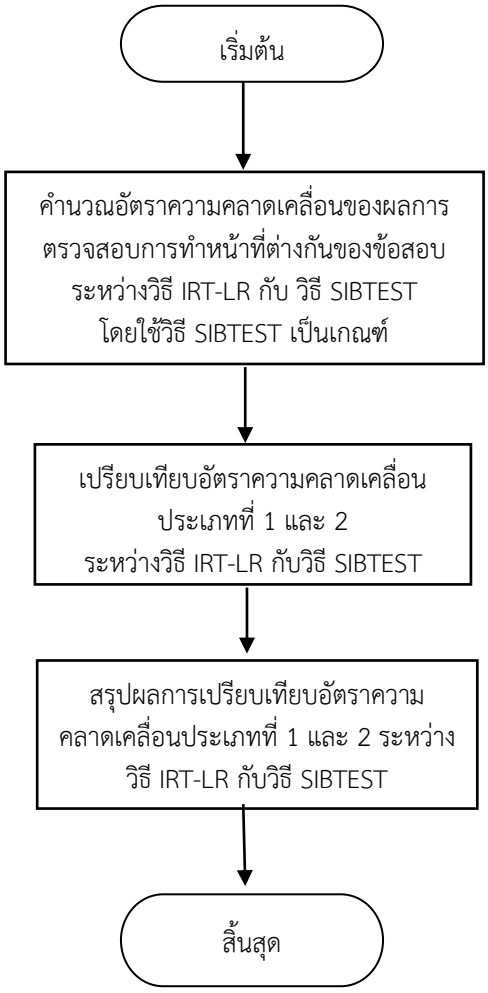
7. ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ เพื่อวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ โดยกำหนดให้ทำการวิเคราะห์ข้อสอบทุกข้อเป็นข้อสอบของแบบทดสอบจากนั้นพิจารณาว่าข้อสอบข้อใด เป็นข้อสอบที่ทำหน้าที่ต่างกัน โดยดูจากค่า *p*-value ที่มีนัยสำคัญทางสถิติที่ระดับ .05

3. ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ ทั้ง 8 กลุ่มสาระการเรียนรู้ ด้วยวิธี IRT-LR และวิธี SIBTEST จำแนกตามเพศ

จากการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ ผู้วิจัยได้รวบรวมผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิธี IRT-LR และ วิธี SIBTEST จากนั้นตัดข้อสอบที่พบว่าทำหน้าที่ต่างกัน โดยข้อสอบที่ถูกตัดสินว่าทำหน้าที่ต่างกันและถูกตัดออกจะไม่นำมาคิดคะแนนให้กับผู้สอบได้มาจากวิธีตรวจสอบทั้ง 2 วิธี ที่พบ DIF ตรงกัน

ระยะที่ 4 การเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 และ 2 ของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 จำนวน 8 กลุ่มสาระการเรียนรู้ ระหว่างวิธี IRT-LR กับวิธี SIBTEST

การคำนวณอัตราความคลาดเคลื่อนในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธี IRT-LR กับวิธี SIBTEST มีขั้นตอนการดำเนินการ ดังนี้



ภาพที่ 3-8 ขั้นตอนการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 และ 2 ของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธี IRT-LR กับวิธี SIBTEST

ผู้วิจัยทำการคำนวณอัตราความคลาดเคลื่อนของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบ O-NET ระดับชั้นมัธยมศึกษาปีที่ 6 ดังนี้

1. นำผลที่ได้จากการวิเคราะห์ด้วยวิธี IRT-LR มาเปรียบเทียบกับผลการตรวจสอบด้วยวิธี SIBTEST โดยมีเกณฑ์ว่า ข้อสอบที่ถูกตรวจพบว่าทำหน้าที่ต่างกันด้วยวิธี SIBTEST ถือว่าเป็นข้อสอบ

ที่ทำหน้าที่ต่างกันหากวิธี IRT-LR พบข้อสอบที่ทำหน้าที่ต่างกันตรงกับข้อที่ตรวจพบด้วยวิธี SIBTEST ถือว่าวิธีนั้นมีความถูกต้องในการตรวจสอบ แต่ถ้าพบว่าข้อสอบข้อใดทำหน้าที่ต่างกันไม่ตรงกับข้อที่ตรวจพบด้วยวิธี SIBTEST ถือว่ามีความคลาดเคลื่อนในการตรวจสอบ

2. การคำนวณอัตราความคลาดเคลื่อนประเภทที่ 1 และ 2 โดยการนำผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี IRT-LR เมื่อใช้เกณฑ์มาเปรียบเทียบกับผลการตรวจสอบด้วยวิธี SIBTEST ถ้าข้อสอบข้อใดที่วิเคราะห์ด้วยวิธี IRT-LR ตรวจพบว่าทำหน้าที่ต่างกัน แต่วิธี SIBTEST ตรวจพบว่าเป็นข้อสอบที่ทำหน้าที่ไม่ต่างกัน ถือว่าวิธีนั้นมีการระบุผิดพลาดทำให้เกิดความคลาดเคลื่อนประเภทที่ 1 และถ้าวิธี IRT-LR ตรวจสอบพบว่าข้อสอบข้อใดทำหน้าที่ไม่ต่างกัน แต่วิธี SIBTEST ตรวจพบว่าเป็นข้อสอบที่ทำหน้าที่ต่างกันถือว่าวิธีนั้น มีการระบุผิดพลาดซึ่งทำให้เกิดความคลาดเคลื่อนประเภทที่ 2 การคำนวณอัตราความคลาดเคลื่อนในการตรวจสอบประเภทที่ 2 (Freund & Walpole, 1980) ดังนี้

อัตราความคลาดเคลื่อนประเภทที่ 1

$$= \frac{\text{จำนวนข้อสอบที่วิธี IRT-LR ระบุผิดพลาดว่าทำหน้าที่ต่างกัน}}{\text{จำนวนข้อสอบที่วิธี SIBTEST ระบุว่าทำหน้าที่ไม่ต่างกันทั้งหมดในแบบทดสอบ}}$$

อัตราความคลาดเคลื่อนประเภทที่ 2

$$= \frac{\text{จำนวนข้อสอบที่วิธี IRT-LR ระบุผิดพลาดว่าทำหน้าที่ไม่ต่างกัน}}{\text{จำนวนข้อสอบที่วิธี SIBTEST ระบุว่าทำหน้าที่ต่างกันทั้งหมดในแบบทดสอบ}}$$

ดังตัวอย่างต่อไปนี้

<u>ตัวอย่าง</u> การคำนวณอัตราความคลาดเคลื่อนประเภทที่ 1 และ 2 ของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ในแบบทดสอบคณิตศาสตร์ ระหว่างวิธี IRT-LR กับวิธี SIBTEST

กลุ่มตัวอย่าง 2,000 คน จำนวน 32 ข้อ

ตรวจสอบ DIF ด้วยวิธี IRT-LR โดยใช้วิธี SIBTEST เป็นเกณฑ์

		วิธี SIBTEST		
		NO DIF	DIF	
วิธี IRT-LR	NO DIF	26	0	26
	DIF	4	2	6
		30	2	32

อัตราความถูกต้องประเภทที่ 1 $= \frac{26}{30} = 0.867$

อัตราความถูกต้องประเภทที่ 2 $= \frac{2}{2} = 1.000$

อัตราความคลาดเคลื่อนประเภทที่ 1 $= \frac{4}{30} = 0.133$

อัตราความคลาดเคลื่อนประเภทที่ 2 $= \frac{0}{2} = 0.000$

บทที่ 4
ผลการวิจัย

การวิจัยนี้มีวัตถุประสงค์เพื่อวิเคราะห์คุณภาพของข้อสอบ (a, b และ c) ตรวจสอบความเที่ยง ความตรงเชิงโครงสร้าง ฉบับก่อนและฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกันออกจากแบบทดสอบ ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ และเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 และ 2 ของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 จำนวน 8 กลุ่มสาระการเรียนรู้ ระหว่างวิธี IRT-LR กับวิธี SIBTEST ซึ่งข้อมูลที่ใช้เป็นข้อมูลทุติยภูมิ (Secondary Data) ที่ได้มาจากสถาบันทดสอบทางการศึกษาแห่งชาติ(องค์การมหาชน) ผู้วิจัยนำเสนอผลการวิจัยเป็น 4 ตอน ดังนี้

ตอนที่ 1 การวิเคราะห์คุณภาพของข้อสอบ (a, b และ c) และค่าความเที่ยงของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 จำนวน 8 กลุ่มสาระการเรียนรู้ ทั้งฉบับก่อนและฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกัน ตามหลักทฤษฎีการตอบสนองข้อสอบ แบบ 3 พารามิเตอร์

ตอนที่ 2 การตรวจสอบความตรงเชิงโครงสร้างของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 จำนวน 8 กลุ่มสาระการเรียนรู้ ทั้งฉบับก่อนและฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกัน

ตอนที่ 3 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ทั้ง 8 กลุ่มสาระการเรียนรู้ ด้วยวิธี IRT-LR และวิธี SIBTEST ฉบับก่อนตัดข้อสอบที่ทำหน้าที่ต่างกัน

ตอนที่ 4 การเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 และ 2 ของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 จำนวน 8 กลุ่มสาระการเรียนรู้ ระหว่างวิธี IRT-LR กับวิธี SIBTEST

ผู้วิจัยได้กำหนดสัญลักษณ์ที่ใช้ในการวิจัย ดังนี้

M	หมายถึง	ค่าเฉลี่ย
SD	หมายถึง	ส่วนเบี่ยงเบนมาตรฐาน
n	หมายถึง	จำนวนกลุ่มตัวอย่าง
a	หมายถึง	ค่าอำนาจจำแนกของข้อสอบ
b	หมายถึง	ค่าความยากของข้อสอบ
c	หมายถึง	ค่าการเดาของข้อสอบ
p	หมายถึง	ความน่าจะเป็นในการทดสอบสมมติฐาน
Low Point-Biserial	หมายถึง	ค่าสัมประสิทธิ์ต่ำสุดในการประมาณค่าความสัมพันธ์ระหว่างข้อสอบรายข้อกับทั้งฉบับ
Cut DIF	หมายถึง	ข้อสอบที่ถูกตัดเมื่อตรวจพบการทำหน้าที่ต่างกันของข้อสอบทั้ง 2 วิธี
S.E.	หมายถึง	ค่าความคลาดเคลื่อนมาตรฐาน (Standard Error)
b_i	หมายถึง	ค่าสัมประสิทธิ์คะแนนมาตรฐาน

r^2	หมายถึง	ค่าสัมประสิทธิ์พยากรณ์ (Coefficient of Determination)
x^2	หมายถึง	ค่าสถิติไค-สแควร์ (Chi-square)
df	หมายถึง	จำนวนองศาความเป็นอิสระ (Degree of freedom)
GFI, AGFI, NFI, NNFI	หมายถึง	ดัชนีที่บ่ง บอกถึงความกลมกลืนของข้อมูลเชิงประจักษ์กับโมเดลที่กำหนดขึ้น ประกอบด้วย ดัชนี GFI (Goodness of fit), ดัชนี AGFI (Adjust goodness of fit index) ดัชนี NFI (Normal fit index), ดัชนี NNFI (Non Norm Fit Index)
RMR	หมายถึง	ดัชนีที่วัดค่าเฉลี่ยส่วนที่เหลือจากการเปรียบเทียบขนาดของความแปรปรวนและความแปรปรวนร่วมระหว่างตัวแปรของประชากร (Root mean square residual)
SRMR	หมายถึง	ดัชนีรากของกำลังสองเฉลี่ยของส่วนที่เหลือ ในรูปคะแนนมาตรฐาน (Standardized Root Mean Square Residual)
RMSEA	หมายถึง	ดัชนีรากที่สองของความคลาดเคลื่อนในการประมาณค่า (Root Mean Square Error of Approximation)
CFI	หมายถึง	ดัชนีวัดวัดระดับความเหมาะสมพอดีเชิงเปรียบเทียบ (Comparative Fit Index)
2^{nd} CFA	หมายถึง	การวิเคราะห์องค์ประกอบเชิงยืนยันสองอันดับ
Thai	หมายถึง	แบบทดสอบวิชาภาษาไทย (ตัวแปรแฝง)
Read	หมายถึง	องค์ประกอบด้านการอ่าน
Write	หมายถึง	องค์ประกอบด้านการเขียน
Speak	หมายถึง	องค์ประกอบด้านการฟัง การดูและการพูด
Used	หมายถึง	องค์ประกอบด้านหลักการใช้ภาษา
Liter	หมายถึง	องค์ประกอบด้านวรรณคดีและวรรณกรรม
Social	หมายถึง	แบบทดสอบวิชาสังคมศึกษา ศาสนาและวัฒนธรรม(ตัวแปรแฝง)
Relig	หมายถึง	องค์ประกอบด้านศาสนา ศิลธรรม จริยธรรม
Civic	หมายถึง	องค์ประกอบด้านหน้าที่พลเมือง วัฒนธรรมและการดำเนินชีวิตในสังคม
Econo	หมายถึง	องค์ประกอบด้านเศรษฐศาสตร์
Histo	หมายถึง	องค์ประกอบด้านประวัติศาสตร์
Geo	หมายถึง	องค์ประกอบด้านภูมิศาสตร์
English	หมายถึง	แบบทดสอบวิชาภาษาอังกฤษ (ตัวแปรแฝง)
Comca	หมายถึง	องค์ประกอบด้านภาษาเพื่อการสื่อสาร

Culture	หมายถึง	องค์ประกอบด้านภาษาและวัฒนธรรม
Other	หมายถึง	องค์ประกอบด้านภาษากับความสัมพันธ์กับกลุ่มสาระการเรียนรู้อื่น
Commu	หมายถึง	องค์ประกอบด้านภาษาความสัมพันธ์กับชุมชนโลก
Math	หมายถึง	แบบทดสอบวิชาคณิตศาสตร์ (ตัวแปรแฝง)
Number	หมายถึง	องค์ประกอบด้านจำนวนและการดำเนินการ
Mesur	หมายถึง	องค์ประกอบด้านการวัด
Alger	หมายถึง	องค์ประกอบด้านพีชคณิต
Proba	หมายถึง	องค์ประกอบด้านการวิเคราะห์ข้อมูลและความน่าจะเป็น
Science	หมายถึง	แบบทดสอบด้านวิทยาศาสตร์ (ตัวแปรแฝง)
Living	หมายถึง	องค์ประกอบด้านสิ่งมีชีวิตกับกระบวนการดำรงชีวิต
Life	หมายถึง	องค์ประกอบด้านชีวิตกับสิ่งแวดล้อม
Substan	หมายถึง	องค์ประกอบด้านสารและสมบัติของสาร
Force	หมายถึง	องค์ประกอบด้านแรงและการเคลื่อนที่
Energy	หมายถึง	องค์ประกอบด้านพลังงาน
Change	หมายถึง	องค์ประกอบด้านกระบวนการเปลี่ยนแปลงของโลก
Astro	หมายถึง	องค์ประกอบด้านดาราศาสตร์และอวกาศ
Nature	หมายถึง	องค์ประกอบด้านธรรมชาติของวิทยาศาสตร์และเทคโนโลยี
Physical	หมายถึง	แบบทดสอบวิชาสุขศึกษาและพละศึกษา (ตัวแปรแฝง)
Human	หมายถึง	องค์ประกอบด้านการเจริญเติบโตและพัฒนาการของมนุษย์
Life	หมายถึง	องค์ประกอบด้านชีวิตและครอบครัว
Movement	หมายถึง	องค์ประกอบด้านการเคลื่อนไหวการออกกำลังกาย การเล่นเกม กีฬาไทยและกีฬาสากล
Strong	หมายถึง	องค์ประกอบด้านการสร้างเสริมสุขภาพ สมรรถภาพและการป้องกันโรค
Safety	หมายถึง	องค์ประกอบด้านความปลอดภัยในชีวิต
Art	หมายถึง	แบบทดสอบวิชาศิลปะ
Visual	หมายถึง	องค์ประกอบด้านทัศนศิลป์
Music	หมายถึง	องค์ประกอบด้านดนตรี
Drama	หมายถึง	องค์ประกอบด้านนาฏศิลป์
Techno	หมายถึง	แบบทดสอบวิชาการงานอาชีพและเทคโนโลยี (ตัวแปรแฝง)
Life	หมายถึง	องค์ประกอบด้านการดำรงชีวิตและครอบครัว
Design	หมายถึง	องค์ประกอบด้านการออกแบบและเทคโนโลยี

Inform	หมายถึง	องค์ประกอบด้านเทคโนโลยีสารสนเทศและการสื่อสาร
Occupa	หมายถึง	องค์ประกอบด้านการอาชีพ
p - Value	หมายถึง	ระดับนัยสำคัญทางสถิติ
DIF	หมายถึง	ข้อสอบที่ทำหน้าที่ต่างกัน
NO DIF	หมายถึง	ข้อสอบที่ทำหน้าที่ไม่ต่างกัน

ตอนที่ 1 ผลการวิเคราะห์คุณภาพของข้อสอบ (a, b และ c) และค่าความเที่ยงของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 จำนวน 8 กลุ่มสาระการเรียนรู้ ทั้งฉบับก่อนและฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกัน ตามหลักทฤษฎีการตอบสนองข้อสอบ แบบ 3 พารามิเตอร์

การวิเคราะห์คุณภาพของข้อสอบ O-NET เป็นการวิเคราะห์ค่าพารามิเตอร์ของข้อสอบตามหลักทฤษฎีการตอบสนองข้อสอบ (IRT) ประกอบด้วย ค่าอำนาจจำแนกของข้อสอบ (a) ค่าความยากของข้อสอบ (b) และค่าการเดาของข้อสอบ (c) โดยใช้โปรแกรม Xcalibre Version 4.2.2 สำหรับการประมาณค่าพารามิเตอร์ของข้อสอบ ซึ่งมีเกณฑ์การคัดเลือกข้อสอบ ดังนี้

เกณฑ์การคัดเลือกข้อสอบ (Urry, 1977)

1) ค่าอำนาจจำแนกของข้อสอบ (a) มีค่าตั้งแต่ 0.50 ถึง 2.50

2) ค่าความยากของข้อสอบ (b) มีค่าตั้งแต่ -2.50 ถึง 2.50

3) ค่าการเดาของข้อสอบ (c) มีค่าไม่เกิน 0.30

โดยเกณฑ์การประเมินค่าความยากของข้อสอบ (b) ของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 จำนวน 8 กลุ่มสาระการเรียนรู้ ดังนี้

ค่าความยากเฉลี่ยมากกว่า	2.0000		หมายถึง	ข้อสอบยากมาก
ค่าความยากเฉลี่ยตั้งแต่	1.0001	ถึง 2.0000	หมายถึง	ข้อสอบยาก
ค่าความยากเฉลี่ยตั้งแต่	0.5001	ถึง 1.0000	หมายถึง	ข้อสอบค่อนข้างยาก
ค่าความยากเฉลี่ยตั้งแต่	-0.4999	ถึง 0.5000	หมายถึง	ข้อสอบปานกลาง
ค่าความยากเฉลี่ยตั้งแต่	-0.9999	ถึง -0.5000	หมายถึง	ข้อสอบค่อนข้างง่าย
ค่าความยากเฉลี่ยตั้งแต่	-1.9999	ถึง -1.0000	หมายถึง	ข้อสอบง่าย
ค่าความยากเฉลี่ยน้อยกว่า	-2.0000		หมายถึง	ข้อสอบง่ายมาก

(จารุจิตร สิทธิปรุ ปิยะทิพย์ ตินวร และโสพส สุขานนท์สวัสดิ์, 2559)

สรุปผลค่าพารามิเตอร์ของข้อสอบรายข้อฉบับก่อนและหลังตัดการทำหน้าที่ต่างกันของข้อสอบ ชั้นมัธยมศึกษาปีที่ 6 จำนวน 8 กลุ่มสาระการเรียนรู้ ได้แก่ 1) ภาษาไทย 2) สังคมศึกษา ศาสนา และวัฒนาธรรม 3) ภาษาต่างประเทศ 4) คณิตศาสตร์ 5) วิทยาศาสตร์ 6) สุขศึกษาและพลศึกษา 7) ศิลปะ และ 8) การงานอาชีพ และเทคโนโลยี และแสดงผลการทดสอบสมมติฐานได้ดังตารางที่ 4-1

ตารางที่ 4-1 ผลการวิเคราะห์คุณภาพของข้อสอบ (a, b และ c) ฉบับก่อนและฉบับหลังตัดการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบภาษาไทย ชั้นมัธยมศึกษาปีที่ 6 จำนวน 70 ข้อ

แบบทดสอบ	ข้อที่	ค่าพารามิเตอร์ก่อนตัด DIF			ค่าพารามิเตอร์หลังตัด DIF		
		a	b	c	a	b	c
ภาษาไทย	1	0.311	0.846	0.178	Cut DIF		
	2	0.389	0.296	0.178	0.152	1.244	0.220
	3	0.584	0.192	0.176	0.564	0.377	0.212
	4	0.467	1.306	0.163	0.955	1.296	0.255
	5	0.794	-1.189	0.172	0.765	-1.105	0.246
	6	1.021	1.249	0.147	0.590	2.305	0.198
	7	0.449	0.968	0.178	0.782	1.118	0.264
	8	0.539	1.826	0.160	0.940	1.510	0.191
	9	0.513	-0.468	0.173	Cut DIF		
	10	0.511	-0.923	0.174	0.587	-0.618	0.254
	11	0.449	0.607	0.188	0.322	1.368	0.262
	12	0.270	0.678	0.175	0.615	0.856	0.279
	13	0.649	0.006	0.171	0.610	0.262	0.237
	14	0.910	0.942	0.167	0.620	1.435	0.200
	15	1.074	-1.113	0.172	Cut DIF		
	16	0.659	1.790	0.165	1.469	1.338	0.182
	17	0.517	0.930	0.172	0.473	1.492	0.242
	18	0.769	1.262	0.162	0.602	2.027	0.222
	19	0.582	-0.227	0.172	0.443	-0.004	0.239
	20	0.730	-0.684	0.167	0.611	-0.616	0.229
	21	0.368	2.662	0.174	0.757	1.992	0.223
	22	0.842	0.697	0.182	Cut DIF		
	23	0.957	-0.146	0.188	0.691	-0.024	0.238
	24	0.618	1.241	0.189	0.858	1.348	0.240
	25	0.825	1.194	0.173	1.225	1.216	0.206
	26	0.585	1.009	0.178	1.014	1.051	0.238
	27	1.203	1.970	0.200	0.899	3.011	0.232
	28	0.591	0.380	0.160	0.757	0.587	0.221

ตารางที่ 4-1 (ต่อ)

แบบทดสอบ	ข้อที่	ค่าพารามิเตอร์ก่อนตัด DIF			ค่าพารามิเตอร์หลังตัด DIF		
		a	b	c	a	b	c
	29	0.320	0.758	0.177	0.942	0.890	0.299
	30	1.309	0.703	0.163	0.953	1.021	0.198
	31	0.775	1.289	0.175	0.836	1.736	0.237
	32	0.901	0.994	0.161	0.766	1.389	0.204
	33	0.729	0.872	0.187	0.553	1.454	0.245
	34	Low Point-Biserial			0.975	3.254	0.244
	35	0.478	1.134	0.170	0.341	2.338	0.254
	36	0.546	-0.224	0.173	Cut DIF		
	37	1.056	0.507	0.161	1.462	0.643	0.200
	38	0.404	0.374	0.169	0.259	1.212	0.249
	39	0.870	-0.374	0.165	0.869	-0.209	0.229
	40	0.735	0.271	0.176	0.740	0.574	0.247
	41	1.021	-0.816	0.171	0.755	-0.883	0.226
	42	0.479	0.160	0.172	0.691	0.441	0.250
	43	0.853	-0.630	0.177	0.578	-0.669	0.236
	44	0.931	1.232	0.174	Cut DIF		
	45	0.434	0.820	0.175	0.785	0.958	0.259
	46	0.185	0.232	0.179	0.387	0.616	0.270
	47	0.750	1.759	0.134	1.290	1.545	0.158
	48	0.411	-0.140	0.173	0.678	0.228	0.261
	49	0.887	0.137	0.180	0.989	0.341	0.237
	50	0.241	1.951	0.175	Cut DIF		
	51	0.875	-0.830	0.177	1.001	-0.643	0.252
	52	0.850	0.829	0.193	0.802	1.110	0.235
	53	0.581	-1.121	0.176	Cut DIF		
	54	0.916	-0.009	0.165	0.816	0.176	0.220
	55	0.990	0.634	0.183	0.755	0.960	0.224
	56	1.139	2.392	0.217	1.150	2.589	0.230
	57	0.706	1.551	0.161	0.581	2.273	0.208
	58	0.253	1.885	0.176	0.810	1.363	0.279
	59	0.873	1.295	0.179	0.964	1.460	0.208

ตารางที่ 4-1 (ต่อ)

แบบทดสอบ	ข้อที่	ค่าพารามิเตอร์ก่อนตัด DIF			ค่าพารามิเตอร์หลังตัด DIF		
		a	b	c	a	b	c
	60	0.519	0.481	0.163	0.882	0.634	0.231
	61	0.531	1.057	0.183	0.826	1.186	0.252
	62	0.849	0.400	0.178	0.651	0.729	0.232
	63	0.880	1.063	0.175	0.531	1.950	0.228
	64	0.521	1.856	0.183		Cut DIF	
	65	0.734	3.080	0.208		Cut DIF	
	66	0.776	0.593	0.175		Cut DIF	
	67	0.559	1.508	0.184	0.698	1.854	0.252
	68	0.718	0.346	0.161		Cut DIF	
	69	0.580	0.227	0.174	0.672	0.534	0.251
	70	0.838	0.731	0.171	1.187	0.835	0.212
ค่าเฉลี่ย (Mean)		0.684	0.671	0.174	0.767	0.989	0.234

จากตารางที่ 4-1 ข้อสอบฉบับก่อนตัดข้อสอบที่ทำหน้าที่ต่างกันในแบบทดสอบภาษาไทย จำนวน 70 ข้อ มีข้อสอบที่ผ่านเกณฑ์ จำนวน 69 ข้อ เป็นข้อสอบที่มีอำนาจจำแนกของข้อสอบ (a) ตั้งแต่ 0.185 ถึง 1.309 ค่าความยากของข้อสอบ (b) ตั้งแต่ -1.189 ถึง 3.080 ค่าการเดาของข้อสอบ (c) ตั้งแต่ 0.134 ถึง 0.217

ข้อสอบฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกันในแบบทดสอบภาษาไทย มีการตัดข้อสอบออก จำนวน 12 ข้อ จากจำนวนข้อสอบทั้งหมด 70 ข้อ มีข้อสอบที่ผ่านเกณฑ์ จำนวน 58 ข้อ เป็นข้อสอบที่มีอำนาจจำแนกของข้อสอบ (a) ตั้งแต่ 0.322 ถึง 1.469 ค่าความยากของข้อสอบ (b) ตั้งแต่ -0.883 ถึง 3.254 ค่าการเดาของข้อสอบ (c) ตั้งแต่ 0.158 ถึง 0.279 ซึ่งเป็นไปตามสมมติฐานของการวิจัยข้อที่ 1

ตารางที่ 4-2 ผลการวิเคราะห์คุณภาพของข้อสอบ (a, b และ c) ฉบับก่อนและฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบสังคมศึกษา ศาสนา และวัฒนธรรม ชั้นมัธยมศึกษาปีที่ 6 จำนวน 65 ข้อ

แบบทดสอบ	ข้อที่	ค่าพารามิเตอร์ก่อนตัด DIF			ค่าพารามิเตอร์หลังตัด DIF		
		a	b	c	a	b	c
สังคมศึกษา ศาสนา	1	0.489	2.069	0.180	0.176	4.000	0.199
และวัฒนธรรม	2	0.443	0.501	0.177	0.152	2.417	0.243
	3	0.567	1.814	0.188	Cut DIF		
	4	1.824	1.673	0.213	0.638	4.000	0.258
	5	Low Point-Biserial			0.932	4.000	0.316
	6	0.809	2.518	0.158	Cut DIF		
	7	0.450	2.016	0.178	Cut DIF		
	8	1.055	1.142	0.155	0.858	2.431	0.232
	9	0.520	1.185	0.174	0.833	1.350	0.222
	10	0.739	1.323	0.166	0.528	3.051	0.252
	11	0.879	1.580	0.195	1.125	3.058	0.275
	12	0.455	1.318	0.185	1.146	2.919	0.373
	13	0.533	1.147	0.172	0.839	1.388	0.229
	14	0.773	1.632	0.173	Cut DIF		
	15	0.493	1.628	0.181	1.114	3.029	0.329
	16	0.581	1.863	0.178	Cut DIF		
	17	0.897	1.778	0.152	1.055	2.175	0.165
	18	1.027	2.404	0.122	1.049	2.908	0.127
	19	0.941	2.371	0.133	1.037	4.000	0.169
	20	0.447	1.753	0.182	Cut DIF		
	21	1.166	1.562	0.135	Cut DIF		
	22	0.341	1.525	0.180	0.891	4.000	0.402
	23	0.678	1.608	0.168	0.576	3.386	0.254
	24	1.676	1.308	0.126	1.326	2.528	0.171
	25	1.108	1.477	0.173	Cut DIF		
	26	0.590	0.568	0.178	0.289	2.137	0.278
	27	1.476	1.825	0.142	Cut DIF		
	28	1.170	1.181	0.187	0.759	2.165	0.219
	29	0.712	2.132	0.196	0.905	4.000	0.266

ตารางที่ 4-2 (ต่อ)

แบบทดสอบ	ข้อที่	ค่าพารามิเตอร์ก่อนตัด DIF			ค่าพารามิเตอร์หลังตัด DIF		
		a	b	c	a	b	c
	30	0.856	1.702	0.164	0.784	4.000	0.248
	31	1.067	1.739	0.148	0.799	3.411	0.193
	32	0.968	2.866	0.156	0.797	3.899	0.166
	33	0.770	2.893	0.210	0.840	3.991	0.235
	34	0.834	2.208	0.175	0.731	2.887	0.184
	35	0.729	2.145	0.172	1.082	2.963	0.223
	36	1.109	2.258	0.224	0.899	4.000	0.254
	37	1.473	1.412	0.138	0.949	2.419	0.156
	38	0.939	1.412	0.188	1.073	2.146	0.232
	39	0.569	1.841	0.180	Cut DIF		
	40	1.015	1.717	0.133	1.160	1.990	0.128
	41	0.714	2.198	0.191	0.917	4.000	0.258
	42	0.994	2.827	0.182	0.907	3.836	0.196
	43	0.371	3.266	0.192	0.483	3.638	0.236
	44	1.180	1.874	0.181	0.771	3.313	0.203
	45	0.475	2.268	0.186	Cut DIF		
	46	0.687	2.392	0.180	0.956	3.030	0.219
	47	0.620	1.776	0.182	0.716	3.509	0.279
	48	1.073	2.353	0.164	0.885	4.000	0.194
	49	Low Point-Biserial			0.985	4.000	0.188
	50	1.369	1.859	0.170	0.973	4.000	0.210
	51	0.589	1.259	0.173	0.402	2.614	0.239
	52	0.472	0.913	0.174	1.038	0.955	0.226
	53	0.774	1.105	0.171	0.604	2.433	0.259
	54	0.502	1.526	0.171	0.868	1.425	0.194
	55	0.636	1.895	0.170	0.753	2.222	0.192
	56	0.302	2.423	0.188	1.177	2.802	0.336
	57	0.283	2.118	0.179	0.433	2.056	0.233
	58	0.580	2.208	0.180	Cut DIF		
	59	0.989	2.030	0.158	0.866	4.000	0.207
	60	0.749	1.646	0.163	0.762	2.282	0.195

ตารางที่ 4-2 (ต่อ)

แบบทดสอบ	ข้อที่	ค่าพารามิเตอร์ก่อนตัด DIF			ค่าพารามิเตอร์หลังตัด DIF		
		a	b	c	a	b	c
	61	0.901	3.022	0.172	0.992	3.492	0.182
	62	0.634	2.204	0.151	0.801	2.283	0.159
	63	0.599	2.525	0.196	0.720	2.418	0.191
	64	0.353	0.944	0.178	Cut DIF		
	65	0.380	0.726	0.180	0.161	3.792	0.306
ค่าเฉลี่ย (Mean)		0.784	1.817	0.173	0.818	3.014	0.229

จากตารางที่ 4-2 ข้อสอบฉบับก่อนตัดข้อสอบที่ทำหน้าที่ต่างกันในแบบทดสอบสังคมศึกษา ศาสนา และวัฒนธรรม จำนวน 65 ข้อ มีข้อสอบที่ผ่านเกณฑ์ จำนวน 63 ข้อ เป็นข้อสอบที่มีอำนาจจำแนกของข้อสอบ (a) 0.283 ถึง 1.824 ค่าความยากของข้อสอบ (b) 0.501 ถึง 3.266 ค่าการเดาของข้อสอบ (c) ตั้งแต่ 0.122 ถึง 0.224

ข้อสอบฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกันในแบบทดสอบสังคมศึกษา ศาสนา และวัฒนธรรม มีการตัดข้อสอบออก จำนวน 13 ข้อ จากจำนวนข้อสอบทั้งหมด 52 ข้อ มีข้อสอบที่ผ่านเกณฑ์ จำนวน 52 ข้อ เป็นข้อสอบที่มีอำนาจจำแนกของข้อสอบ (a) ตั้งแต่ 0.152 ถึง 1.326 ค่าความยากของข้อสอบ (b) ตั้งแต่ -1.350 ถึง 4.000 ค่าการเดาของข้อสอบ (c) ตั้งแต่ 0.127 ถึง 0.402 ซึ่งเป็นไปตามสมมติฐานของการวิจัยข้อที่ 1

ตารางที่ 4-3 ผลการวิเคราะห์คุณภาพของข้อสอบ (a, b และ c) ฉบับก่อนและฉบับหลังตัดข้อสอบทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบภาษาอังกฤษ ชั้นมัธยมศึกษาปีที่ 6 จำนวน 80 ข้อ

แบบทดสอบ	ข้อที่	ค่าพารามิเตอร์ก่อนตัด DIF			ค่าพารามิเตอร์หลังตัด DIF		
		a	b	c	a	b	c
ภาษาอังกฤษ	1	1.046	0.620	0.181	0.887	0.886	0.138
	2	1.071	1.567	0.217	0.808	2.000	0.195
	3	1.986	1.122	0.317	1.811	1.534	0.337
	4	2.313	1.210	0.191	2.061	1.645	0.218
	5	1.894	1.255	0.211	1.615	1.683	0.219
	6	0.934	0.685	0.175	0.930	1.110	0.194
	7	1.029	1.479	0.197	0.997	1.934	0.210
	8	1.759	0.930	0.288	1.506	1.311	0.294
	9	1.631	1.134	0.184	1.468	1.536	0.191
	10	1.989	1.305	0.207	1.627	1.727	0.212
	11	1.212	0.870	0.175	1.191	1.274	0.189
	12	0.983	1.631	0.188	0.909	2.113	0.198
	13	1.072	2.042	0.163	0.996	2.563	0.173
	14	1.482	0.691	0.204	1.386	1.074	0.215
	15	0.899	2.856	0.181	0.878	3.416	0.191
	16	1.072	1.957	0.154	1.003	2.462	0.164
	17	1.400	1.741	0.184	1.246	2.219	0.192
	18	1.806	1.962	0.173	1.415	2.520	0.180
	19	1.254	1.560	0.213	1.097	2.039	0.221
	20	1.562	2.215	0.214	1.211	2.827	0.221
	21	0.970	1.781	0.184	0.892	2.287	0.195
	22	1.400	1.414	0.182	1.243	1.867	0.190
	23	1.856	1.959	0.195	1.449	2.510	0.202
	24	1.380	1.721	0.195	1.208	2.201	0.202
	25	Low Point-Biserial			1.497	3.281	0.241
	26	1.726	2.522	0.208	Cut DIF		
	27	1.108	2.087	0.179	1.008	2.612	0.188
	28	2.058	1.907	0.157	1.585	2.437	0.163
	29	1.446	1.579	0.210	1.231	2.048	0.216

ตารางที่ 4-3 (ต่อ)

แบบทดสอบ	ข้อที่	ค่าพารามิเตอร์ก่อนตัด DIF			ค่าพารามิเตอร์หลังตัด DIF		
		a	b	c	a	b	c
	30	1.395	2.071	0.231	1.149	2.629	0.238
	31	1.259	1.947	0.203	1.075	2.485	0.210
	32	2.002	1.314	0.166	1.682	1.736	0.173
	33	1.700	1.842	0.191	1.387	2.364	0.198
	34	1.577	1.828	0.207	1.263	2.343	0.213
	35	1.665	2.005	0.162	1.383	2.549	0.169
	36	1.929	2.449	0.163	1.428	3.115	0.169
	37	1.645	2.361	0.214	1.231	2.991	0.220
	38	1.828	1.278	0.191	1.558	1.707	0.199
	39	1.586	1.233	0.194	1.390	1.663	0.202
	40	1.733	1.157	0.174	1.511	1.568	0.182
	41	2.114	1.983	0.124	1.622	2.554	0.131
	42	1.681	1.821	0.207	1.316	2.341	0.212
	43	1.912	1.364	0.161	1.613	1.810	0.169
	44	2.469	1.166	0.170	2.039	1.571	0.177
	45	1.750	1.460	0.289	1.353	1.895	0.292
	46	Low Point-Biserial			1.335	4.000	0.187
	47	0.954	1.234	0.196	0.922	1.682	0.211
	48	0.826	1.692	0.186	0.788	2.189	0.200
	49	2.244	1.803	0.196	1.637	2.319	0.202
	50	2.591	2.075	0.222	1.858	2.641	0.228
	51	0.797	-0.634	0.206	0.870	-0.126	0.255
	52	1.095	0.175	0.170	1.129	0.573	0.190
	53	1.212	0.288	0.140	1.234	0.671	0.156
	54	1.125	-0.422	0.233	1.191	0.013	0.265
	55	0.899	-0.292	0.177	0.945	0.162	0.212
	56	1.187	0.137	0.143	1.217	0.532	0.163
	57	0.990	1.561	0.200	0.971	2.018	0.214
	58	1.061	1.551	0.217	0.982	2.019	0.227
	59	1.316	0.904	0.206	Cut DIF		
	60	1.603	1.430	0.189	1.371	1.881	0.196

ตารางที่ 4-3 (ต่อ)

แบบทดสอบ	ข้อที่	ค่าพารามิเตอร์ก่อนตัด DIF			ค่าพารามิเตอร์หลังตัด DIF		
		a	b	c	a	b	c
	61	1.388	1.417	0.210	1.237	1.871	0.219
	62	3.193	2.040	0.282	1.751	2.660	0.288
	63	2.647	1.715	0.175	1.762	2.220	0.180
	64	2.428	1.954	0.168	1.716	2.504	0.174
	65	1.905	2.151	0.206	1.451	2.717	0.212
	66	2.020	2.468	0.188	1.337	3.182	0.194
	67	2.224	2.205	0.276	1.461	2.866	0.282
	68	1.543	1.459	0.164	1.338	1.925	0.172
	69	0.772	1.953	0.174	0.758	2.456	0.188
	70	1.158	1.390	0.199	1.063	1.841	0.209
	71	1.433	1.984	0.181	1.194	2.521	0.188
	72	1.065	1.891	0.278	0.924	2.409	0.285
	73	1.482	1.666	0.214	1.247	2.161	0.221
	74	1.199	1.319	0.240	Cut DIF		
	75	1.803	1.402	0.175	1.516	1.855	0.183
	76	0.569	2.376	0.200	0.658	2.836	0.226
	77	0.882	2.229	0.183	0.846	2.755	0.194
	78	1.579	1.456	0.240	1.331	1.902	0.247
	79	1.780	1.962	0.232	1.358	2.525	0.239
	80	1.121	1.756	0.243	1.021	2.246	0.253
ค่าเฉลี่ย (Mean)		1.521	1.530	0.199	1.280	2.058	0.207

จากตารางที่ 4-3 ข้อสอบฉบับก่อนตัดข้อสอบที่ทำหน้าที่ต่างกันในแบบทดสอบภาษาอังกฤษ จากจำนวนข้อสอบ 80 ข้อ มีข้อสอบที่ผ่านเกณฑ์ จำนวน 78 ข้อ เป็นข้อสอบที่มีอำนาจจำแนกของข้อสอบ (a) ตั้งแต่ 0.569 ถึง 3.193 ค่าความยากของข้อสอบ (b) ตั้งแต่ -0.634 ถึง 2.856 ค่าการเดาของข้อสอบ (c) ตั้งแต่ 0.124 ถึง 0.224

ข้อสอบฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกันในแบบทดสอบภาษาอังกฤษ มีการตัดข้อสอบออก จำนวน 3 ข้อ จากจำนวนข้อสอบทั้งหมด 77 ข้อ มีข้อสอบที่ผ่านเกณฑ์ จำนวน 77 ข้อ เป็นข้อสอบที่มีอำนาจจำแนกของข้อสอบ (a) ตั้งแต่ 0.658 ถึง 2.061 ค่าความยากของข้อสอบ (b) ตั้งแต่ -0.126 ถึง 4.000 ค่าการเดาของข้อสอบ (c) ตั้งแต่ 0.131 ถึง 0.337 ซึ่งเป็นไปตามสมมติฐานของการวิจัยข้อที่ 1

ตารางที่ 4-4 ผลการวิเคราะห์คุณภาพของข้อสอบ (a, b และ c) ฉบับก่อนและฉบับหลังตัดข้อสอบทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 6 จำนวน 32 ข้อ

แบบทดสอบ	ข้อที่	ค่าพารามิเตอร์ก่อนตัด DIF			ค่าพารามิเตอร์หลังตัด DIF		
		a	b	c	a	b	c
คณิตศาสตร์	1	1.600	1.538	0.125	1.844	2.728	0.141
	2	1.415	2.008	0.162	1.599	3.232	0.177
	3	0.876	1.652	0.158	0.961	2.907	0.184
	4	1.374	1.446	0.175	1.375	2.688	0.206
	5	0.852	1.737	0.179	0.827	3.028	0.199
	6	1.582	1.624	0.131	Cut DIF		
	7	1.462	0.939	0.181	1.272	2.075	0.194
	8	1.605	1.106	0.200	1.372	2.268	0.214
	9	1.471	1.315	0.202	1.240	2.501	0.215
	10	1.020	1.181	0.175	0.979	2.396	0.196
	11	0.647	1.773	0.206	0.782	3.038	0.242
	12	1.513	1.746	0.145	1.156	3.029	0.156
	13	1.348	1.943	0.126	1.098	3.241	0.137
	14	1.691	1.990	0.209	1.166	3.316	0.219
	15	2.182	1.669	0.184	1.385	2.934	0.194
	16	1.466	1.760	0.176	1.196	3.017	0.188
	17	1.731	1.648	0.194	1.278	2.874	0.204
	18	1.819	1.529	0.178	1.297	2.757	0.188
	19	1.419	0.909	0.111	Cut DIF		
	20	1.135	1.079	0.174	1.077	2.258	0.192
	21	1.109	1.558	0.184	0.990	2.786	0.198
	22	1.005	1.106	0.160	0.945	2.304	0.179
	23	1.000	1.232	0.155	0.975	2.434	0.175
	24	1.742	0.897	0.171	1.454	2.014	0.183
	25	1.091	1.438	0.167	1.027	2.639	0.183
	26	0.773	2.036	0.185	0.868	3.361	0.211
	27	1.163	2.102	0.156	0.957	3.421	0.167
	28	0.449	1.382	0.173	0.520	2.770	0.228
	29	0.692	1.293	0.165	0.741	2.563	0.198

ตารางที่ 4-4 (ต่อ)

แบบทดสอบ	ข้อที่	ค่าพารามิเตอร์ก่อนตัด DIF			ค่าพารามิเตอร์หลังตัด DIF		
		a	b	c	a	b	c
	30	0.811	0.217	0.161	0.810	1.505	0.211
	31	0.955	0.956	0.183	0.968	2.150	0.209
	32	0.854	1.329	0.181	0.865	2.518	0.203
ค่าเฉลี่ย (Mean)		1.245	1.442	0.170	1.101	2.692	0.193

จากตารางที่ 4-4 ข้อสอบฉบับก่อนตัดข้อสอบที่ทำหน้าที่ต่างกันในแบบทดสอบคณิตศาสตร์ จำนวน 32 ข้อ มีข้อสอบที่ผ่านเกณฑ์ จำนวน 32 ข้อ เป็นข้อสอบที่มีอำนาจจำแนกของข้อสอบ (a) ตั้งแต่ 0.449 ถึง 2.182 ค่าความยากของข้อสอบ (b) ตั้งแต่ 0.217 ถึง 2.102 ค่าการเดาของข้อสอบ (c) ตั้งแต่ 0.111 ถึง 0.209

ข้อสอบฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกันในแบบทดสอบคณิตศาสตร์ มีการตัดข้อสอบออก จำนวน 2 ข้อ จากจำนวนข้อสอบทั้งหมด 30 ข้อ มีข้อสอบที่ผ่านเกณฑ์ จำนวน 30 ข้อ เป็นข้อสอบที่มีอำนาจจำแนกของข้อสอบ (a) ตั้งแต่ 0.520 ถึง 1.844 ค่าความยากของข้อสอบ (b) ตั้งแต่ 1.505 ถึง 3.421 ค่าการเดาของข้อสอบ (c) ตั้งแต่ 0.137 ถึง 0.242 ซึ่งเป็นไปตามสมมติฐานของการวิจัยข้อที่ 1

ตารางที่ 4-5 ผลการวิเคราะห์คุณภาพของข้อสอบ (a, b และ c) ฉบับก่อนและฉบับหลังตัดข้อสอบทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบวิทยาศาสตร์ ชั้นมัธยมศึกษาปีที่ 6 จำนวน 80 ข้อ

แบบทดสอบ	ข้อที่	ค่าพารามิเตอร์ก่อนตัด DIF			ค่าพารามิเตอร์หลังตัด DIF		
		a	b	c	a	b	c
วิทยาศาสตร์	1	1.017	1.464	0.205	1.028	2.155	0.227
	2	1.020	2.056	0.161	0.842	2.875	0.162
	3	1.301	1.328	0.157	1.189	2.078	0.183
	4	0.447	1.679	0.195	0.482	2.657	0.267
	5	1.316	1.219	0.166	1.121	1.918	0.184
	6	1.066	0.972	0.204	0.881	1.709	0.239
	7	0.556	2.099	0.217	0.702	2.954	0.273
	8	1.067	1.583	0.196	0.913	2.443	0.218
	9	1.191	0.837	0.199	0.990	1.450	0.225
	10	0.762	1.117	0.198	Cut DIF		
	11	0.980	1.452	0.168	Cut DIF		
	12	0.704	1.213	0.194	0.638	2.007	0.238
	13	0.853	1.676	0.197	0.838	2.416	0.224
	14	0.676	1.061	0.193	0.656	1.762	0.242
	15	0.685	1.098	0.186	0.594	1.935	0.236
	16	0.857	1.519	0.188	Cut DIF		
	17	0.869	1.101	0.194	0.728	1.918	0.234
	18	1.385	1.597	0.226	1.188	2.484	0.244
	19	1.658	0.873	0.190	1.284	1.492	0.206
	20	1.427	1.851	0.190	Cut DIF		
	21	1.027	2.185	0.170	0.892	3.191	0.187
	22	Low Point-Biserial			1.016	4.000	0.235
	23	1.596	1.626	0.197	1.162	2.529	0.209
	24	1.768	1.491	0.223	1.287	2.313	0.233
	25	1.968	1.375	0.160	1.457	2.149	0.170
	26	1.123	2.237	0.176	0.989	3.246	0.191
	27	0.588	1.166	0.211	0.631	1.935	0.276
	28	1.160	1.634	0.181	0.950	2.536	0.200
	29	0.825	1.984	0.160	0.751	2.916	0.185

ตารางที่ 4-5 (ต่อ)

แบบทดสอบ	ข้อที่	ค่าพารามิเตอร์ก่อนตัด DIF			ค่าพารามิเตอร์หลังตัด DIF		
		a	b	c	a	b	c
	30	0.736	1.847	0.178	0.678	2.703	0.206
	31	0.882	1.376	0.206	Cut DIF		
	32	0.863	1.408	0.202	0.748	2.280	0.235
	33	1.004	2.412	0.199	0.851	3.420	0.212
	34	1.363	2.180	0.191	1.087	3.100	0.200
	35	2.002	1.859	0.184	1.394	2.788	0.192
	36	1.197	1.764	0.188	1.022	2.582	0.201
	37	1.427	1.399	0.178	1.100	2.215	0.193
	38	1.190	2.781	0.165	0.984	3.815	0.174
	39	0.965	2.288	0.152	Cut DIF		
	40	0.452	2.571	0.187	0.531	3.448	0.238
	41	1.149	2.051	0.191	0.969	2.898	0.201
	42	0.815	1.333	0.203	0.847	2.034	0.242
	43	0.611	2.402	0.207	0.753	3.205	0.248
	44	1.306	2.104	0.182	1.058	2.977	0.190
	45	0.883	0.964	0.219	0.860	1.687	0.268
	46	0.463	2.408	0.195	Cut DIF		
	47	0.987	2.187	0.160	0.890	3.051	0.174
	48	1.017	1.173	0.184	0.916	1.829	0.208
	49	1.685	1.862	0.205	1.297	2.739	0.214
	50	1.579	1.676	0.183	1.282	2.493	0.193
	51	0.397	2.268	0.193	0.479	3.099	0.258
	52	0.595	1.469	0.193	0.620	2.208	0.242
	53	0.665	1.301	0.181	0.656	2.033	0.227
	54	0.657	2.018	0.204	0.674	2.773	0.234
	55	Low Point-Biserial			1.000	4.000	0.264
	56	0.555	0.966	0.191	0.480	1.916	0.260
	57	0.713	1.351	0.183	0.633	2.208	0.224
	58	0.741	2.452	0.177	0.713	3.405	0.200
	59	Low Point-Biserial			Cut DIF		
	60	0.802	2.665	0.215	0.814	3.592	0.234

ตารางที่ 4-5 (ต่อ)

แบบทดสอบ	ข้อที่	ค่าพารามิเตอร์ก่อนตัด DIF			ค่าพารามิเตอร์หลังตัด DIF		
		a	b	c	a	b	c
	61	1.469	2.418	0.148	1.124	3.496	0.157
	62	0.932	2.917	0.208	0.866	3.908	0.220
	63	0.757	0.781	0.179	0.630	1.472	0.224
	64	Low Point-Biserial			0.930	4.000	0.299
	65	0.463	1.191	0.188	0.500	1.929	0.256
	66	0.489	1.807	0.192	0.534	2.592	0.246
	67	0.727	2.502	0.189	0.723	3.348	0.210
	68	Low Point-Biserial			0.966	4.000	0.252
	69	0.756	2.666	0.209	0.807	3.532	0.231
	70	1.840	2.416	0.149	1.246	3.506	0.157
	71	1.033	2.659	0.207	0.930	3.647	0.219
	72	0.602	1.066	0.191	0.574	1.817	0.246
	73	0.798	2.294	0.191	0.822	3.224	0.216
	74	0.606	0.737	0.197	Cut DIF		
	75	1.055	2.161	0.230	0.969	3.005	0.244
	76	1.425	2.006	0.170	1.105	2.961	0.180
	77	0.645	2.036	0.194	0.708	2.896	0.236
	78	0.344	0.779	0.189	Cut DIF		
	79	0.616	0.364	0.183	0.469	1.075	0.248
	80	0.447	0.359	0.190	Cut DIF		
ค่าเฉลี่ย (Mean)		0.968	1.696	0.189	0.876	2.666	0.221

จากตารางที่ 4-5 ข้อสอบฉบับก่อนตัดข้อสอบที่ทำหน้าที่ต่างกันในแบบทดสอบวิทยาศาสตร์ จำนวน 80 ข้อ มีข้อสอบที่ผ่านเกณฑ์ จำนวน 75 ข้อ เป็นข้อสอบที่มีอำนาจจำแนกของข้อสอบ (a) ตั้งแต่ 0.344 ถึง 2.002 ค่าความยากของข้อสอบ (b) ตั้งแต่ 0.359 ถึง 2.917 ค่าการเดาของข้อสอบ (c) ตั้งแต่ 0.148 ถึง 0.230

ข้อสอบฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกันในแบบทดสอบวิทยาศาสตร์ มีการตัดข้อสอบออก จำนวน 11 ข้อ จากจำนวนข้อสอบทั้งหมด 69 ข้อ มีข้อสอบที่ผ่านเกณฑ์ จำนวน 69 ข้อ เป็นข้อสอบที่มีอำนาจจำแนกของข้อสอบ (a) ตั้งแต่ 0.469 ถึง 1.457 ค่าความยากของข้อสอบ (b) ตั้งแต่ 1.075 ถึง 4.000 ค่าการเดาของข้อสอบ (c) ตั้งแต่ 0.157 ถึง 0.299 ซึ่งเป็นไปตามสมมติฐานของการวิจัยข้อที่ 1

ตารางที่ 4-6 ผลการวิเคราะห์คุณภาพของข้อสอบ (a, b และ c) ฉบับก่อนและฉบับหลังตัดข้อสอบทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบสุขศึกษา และพละศึกษา ชั้นมัธยมศึกษาปีที่ 6 จำนวน 32 ข้อ

แบบทดสอบ	ข้อที่	ค่าพารามิเตอร์ก่อนตัด DIF			ค่าพารามิเตอร์หลังตัด DIF		
		a	b	c	a	b	c
สุขศึกษาและ	1	2.025	0.234	0.195	Cut DIF		
พละศึกษา	2	1.152	-0.559	0.224	1.076	-0.404	0.314
	3	0.104	-2.768	0.254	Cut DIF		
	4	0.217	-0.881	0.226	0.265	-0.261	0.300
	5	0.526	-0.212	0.226	0.565	0.600	0.401
	6	0.090	0.113	0.226	Cut DIF		
	7	0.759	-2.219	0.222	0.604	-2.739	0.249
	8	Low Point-Biserial			0.159	1.452	0.270
	9	0.263	-0.126	0.221	0.298	0.159	0.255
	10	0.075	2.431	0.203	0.125	3.333	0.285
	11	0.482	0.792	0.203	Cut DIF		
	12	0.065	-1.795	0.261	0.101	-1.069	0.278
	13	0.551	-0.477	0.233	Cut DIF		
	14	0.929	-0.769	0.220	0.723	-0.851	0.250
	15	0.847	0.412	0.212	0.767	0.803	0.258
	16	0.261	0.204	0.222	0.279	0.590	0.264
	17	1.228	-0.144	0.196	1.110	-0.007	0.227
	18	1.055	-1.386	0.220	0.800	-1.723	0.247
	19	0.144	-3.382	0.247	Cut DIF		
	20	0.089	3.552	0.201	Cut DIF		
	21	0.134	3.544	0.214	Cut DIF		
	22	0.491	-1.515	0.224	0.461	-1.548	0.253
	23	0.376	-3.308	0.228	0.361	-3.469	0.253
	24	0.645	-1.029	0.222	0.469	-1.229	0.252
	25	1.254	0.321	0.237	Cut DIF		
	26	1.056	3.490	0.223	1.032	4.000	0.227
	27	0.109	-2.045	0.245	Cut DIF		
	28	0.062	2.459	0.197	0.115	3.581	0.290
	29	0.140	-1.795	0.237	Cut DIF		

ตารางที่ 4-6 (ต่อ)

แบบทดสอบ	ข้อที่	ค่าพารามิเตอร์ก่อนตัด DIF			ค่าพารามิเตอร์หลังตัด DIF		
		a	b	c	a	b	c
	30	0.671	-1.416	0.220	0.531	-1.673	0.249
	31	2.244	0.384	0.189	1.044	0.790	0.210
	32	0.236	0.980	0.217	0.279	1.336	0.262
ค่าเฉลี่ย (Mean)		0.590	-0.223	0.221	0.532	0.080	0.226

จากตารางที่ 4-6 ข้อสอบฉบับก่อนตัดข้อสอบที่ทำหน้าที่ต่างกันในแบบทดสอบสุขศึกษาและพละศึกษา จำนวน 32 ข้อ มีข้อสอบที่ผ่านเกณฑ์ จำนวน 31 ข้อ เป็นข้อสอบที่มีอำนาจจำแนกของข้อสอบ (a) ตั้งแต่ 0.062 ถึง 2.244 ค่าความยากของข้อสอบ (b) ตั้งแต่ -3.382 ถึง 3.552 ค่าการเดาของข้อสอบ (c) ตั้งแต่ 0.189 ถึง 0.261

ข้อสอบฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกันในแบบทดสอบสุขศึกษา และพละศึกษา มีการตัดข้อสอบออก จำนวน 11 ข้อ จากจำนวนข้อสอบทั้งหมด 21 ข้อ มีข้อสอบที่ผ่านเกณฑ์ จำนวน 21 ข้อ เป็นข้อสอบที่มีอำนาจจำแนกของข้อสอบ (a) ตั้งแต่ 0.101 ถึง 1.110 ค่าความยากของข้อสอบ (b) ตั้งแต่ -3.469 ถึง 4.000 ค่าการเดาของข้อสอบ (c) ตั้งแต่ 0.210 ถึง 0.401 ซึ่งเป็นไปตามสมมติฐานของการวิจัยข้อที่ 1

ตารางที่ 4-7 ผลการวิเคราะห์คุณภาพของข้อสอบ (a, b และ c) ฉบับก่อนและฉบับหลังตัดข้อสอบทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบศิลปะ ชั้นมัธยมศึกษาปีที่ 6 จำนวน 30 ข้อ

แบบทดสอบ	ข้อที่	ค่าพารามิเตอร์ก่อนตัด DIF			ค่าพารามิเตอร์หลังตัด DIF		
		a	b	c	a	b	c
ศิลปะ	1	0.552	-0.006	0.201	Cut DIF		
	2	0.864	2.493	0.234	0.942	4.000	0.260
	3	Low Point-Biserial			0.506	3.354	0.269
	4	0.178	3.034	0.145	0.998	4.000	0.334
	5	1.251	2.912	0.263	0.951	4.000	0.273
	6	Low Point-Biserial			0.946	4.000	0.167
	7	1.396	3.034	0.227	0.995	4.000	0.236
	8	1.248	2.990	0.277	Cut DIF		
	9	1.248	2.975	0.279	0.959	4.000	0.288
	10	1.401	2.875	0.202	1.034	4.000	0.209
	11	1.928	0.316	0.173	1.076	0.876	0.203
	12	Low Point-Biserial			1.006	4.000	0.393
	13	Low Point-Biserial			0.758	3.791	0.161
	14	0.278	-0.875	0.212	Cut DIF		
	15	0.810	1.760	0.176	0.571	3.257	0.203
	16	1.271	3.034	0.191	0.966	4.000	0.197
	17	0.263	2.842	0.219	Cut DIF		
	18	Low Point-Biserial			0.724	2.638	0.147
	19	1.114	1.173	0.195	0.827	2.220	0.221
	20	1.401	1.696	0.158	0.906	3.128	0.170
	21	0.232	3.034	0.194	0.981	4.000	0.355
	22	0.809	0.582	0.202	0.576	1.265	0.240
	23	1.548	3.034	0.225	1.066	4.000	0.233
	24	3.620	0.245	0.162	2.129	0.605	0.162
	25	1.038	2.330	0.189	0.842	3.848	0.203
	26	1.109	2.685	0.185	0.875	4.000	0.196
	27	0.881	1.216	0.180	Cut DIF		
	28	1.597	3.034	0.138	1.064	4.000	0.145
	29	1.041	2.165	0.187	0.809	3.840	0.206
	30	1.423	3.034	0.141	Cut DIF		
ค่าเฉลี่ย (Mean)		1.140	2.064	0.198	0.938	3.368	0.228

จากตารางที่ 4-7 ข้อสอบฉบับก่อนตัดข้อสอบที่ทำหน้าที่ต่างกันในแบบทดสอบศิลปะ จำนวน 30 ข้อ มีข้อสอบที่ผ่านเกณฑ์ จำนวน 25 ข้อ เป็นข้อสอบที่มีอำนาจจำแนกของข้อสอบ (a) ตั้งแต่ 0.176 ถึง 3.620 ค่าความยากของข้อสอบ (b) ตั้งแต่ -0.875 ถึง 3.034 ค่าการเดาของข้อสอบ (c) ตั้งแต่ 0.138 ถึง 0.279

ข้อสอบฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกันในแบบทดสอบศิลปะ มีการตัดข้อสอบออก จำนวน 6 ข้อ จากจำนวนข้อสอบ 24 ข้อ มีข้อสอบที่ผ่านเกณฑ์ จำนวน 24 ข้อ เป็นข้อสอบที่มีอำนาจจำแนกของข้อสอบ (a) ตั้งแต่ 1.076 ถึง 2.129 ค่าความยากของข้อสอบ (b) ตั้งแต่ 0.605 ถึง 4.000 ค่าการเดาของข้อสอบ (c) ตั้งแต่ 0.147 ถึง 0.393 ซึ่งเป็นไปตามสมมติฐานของการวิจัยข้อที่ 1

ตารางที่ 4-8 ผลการวิเคราะห์คุณภาพของข้อสอบ (a, b และ c) ฉบับก่อนและฉบับหลังตัดข้อสอบทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบการงานอาชีพ และเทคโนโลยี ชั้นมัธยมศึกษาปีที่ 6 จำนวน 41 ข้อ

แบบทดสอบ	ข้อที่	ค่าพารามิเตอร์ก่อนตัด DIF			ค่าพารามิเตอร์หลังตัด DIF		
		a	b	c	a	b	c
การงานอาชีพและเทคโนโลยี	1	0.668	1.093	0.197	0.067	4.000	0.061
	2	1.710	0.490	0.254	0.076	2.038	0.086
	3	0.226	-2.283	0.204	0.574	-0.954	0.230
	4	0.072	4.000	0.164	2.979	0.554	0.201
	5	0.901	-0.512	0.229	0.167	-1.335	0.264
	6	0.364	0.842	0.194	0.345	1.542	0.259
	7	3.685	0.174	0.235	Cut DIF		
	8	1.516	0.603	0.197	1.219	4.000	0.397
	9	0.660	-0.372	0.189	Cut DIF		
	10	0.169	2.793	0.195	1.023	0.874	0.218
	11	0.105	4.000	0.170	1.120	0.975	0.224
	12	1.695	0.122	0.292	0.105	0.892	0.274
	13	3.249	0.288	0.239	0.102	3.917	0.280
	14	1.558	0.712	0.190	1.180	4.000	0.367
	15	0.278	0.479	0.196	0.412	0.707	0.245
	16	0.079	2.373	0.195	1.946	0.423	0.230
	17	Low Point-Biserial			1.085	3.648	0.191
	18	0.288	-1.329	0.194	1.005	-0.398	0.238
	19	0.211	4.000	0.162	0.843	1.615	0.178
	20	0.091	4.000	0.186	1.264	0.703	0.214
	21	0.313	-0.291	0.194	0.536	0.013	0.227
	22	1.387	1.159	0.125	1.256	4.000	0.235
	23	0.066	1.624	0.199	2.187	0.369	0.257
	24	0.244	2.057	0.190	0.641	1.233	0.227
	25	0.304	-1.257	0.192	1.056	-0.361	0.242
	26	0.156	1.262	0.199	0.740	0.564	0.239
	27	0.339	-0.649	0.188	0.725	-0.205	0.220
	28	1.384	0.688	0.182	1.193	3.861	0.370
	29	1.081	0.710	0.183	1.221	3.879	0.382

ตารางที่ 4-8 (ต่อ)

แบบทดสอบ	ข้อที่	ค่าพารามิเตอร์ก่อนตัด DIF			ค่าพารามิเตอร์หลังตัด DIF		
		a	b	c	a	b	c
	30	0.176	2.722	0.195	1.033	1.122	0.264
	31	0.897	3.537	0.300	0.916	1.920	0.241
	32	0.863	3.712	0.221	Cut DIF		
	33	0.628	0.575	0.183	0.252	2.448	0.269
	34	0.772	-0.439	0.179	0.307	-0.325	0.243
	35	0.253	4.000	0.169	1.244	1.777	0.210
	36	1.132	-0.462	0.185	0.261	-0.666	0.251
	37	0.302	-0.687	0.192	0.655	-0.182	0.233
	38	0.893	-1.394	0.181	0.749	-1.353	0.232
	39	0.534	-0.206	0.183	Cut DIF		
	40	1.037	-1.371	0.183	Cut DIF		
	41	0.768	-0.526	0.196	Cut DIF		
ค่าเฉลี่ย (Mean)		0.776	0.906	0.197	0.871	1.291	0.243

จากตารางที่ 4-8 ข้อสอบฉบับก่อนตัดข้อสอบที่หน้าที่ต่างกันในแบบทดสอบการงานอาชีพและเทคโนโลยี จำนวน 41 ข้อ มีข้อสอบที่ผ่านเกณฑ์ จำนวน 40 ข้อ ซึ่งเป็นข้อสอบที่มีอำนาจจำแนกของข้อสอบ (a) ตั้งแต่ 0.066 ถึง 3.685 ค่าความยากของข้อสอบ (b) ตั้งแต่ -2.283 ถึง 4.000 ค่าการเดาของข้อสอบ (c) ตั้งแต่ 0.125 ถึง 0.300

ข้อสอบฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกันในแบบทดสอบการงานอาชีพ และเทคโนโลยี มีการตัดข้อสอบออก จำนวน 6 ข้อ จากจำนวนข้อสอบ 35 ข้อ มีข้อสอบที่ผ่านเกณฑ์การ จำนวน 35 ข้อ เป็นข้อสอบที่มีอำนาจจำแนกของข้อสอบ (a) ตั้งแต่ 0.067 ถึง 2.979 ค่าความยากของข้อสอบ (b) ตั้งแต่ -1.353 ถึง 4.000 ค่าการเดาของข้อสอบ (c) ตั้งแต่ 0.067 ถึง 0.397 ซึ่งเป็นไปตามสมมติฐานของการวิจัยข้อที่ 1

ตารางที่ 4-9 ผลการตรวจสอบค่าความเที่ยงของข้อสอบฉบับก่อนและฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกัน ตามหลักทฤษฎีการตอบสนองข้อสอบ (IRT) แบบ 3 พารามิเตอร์

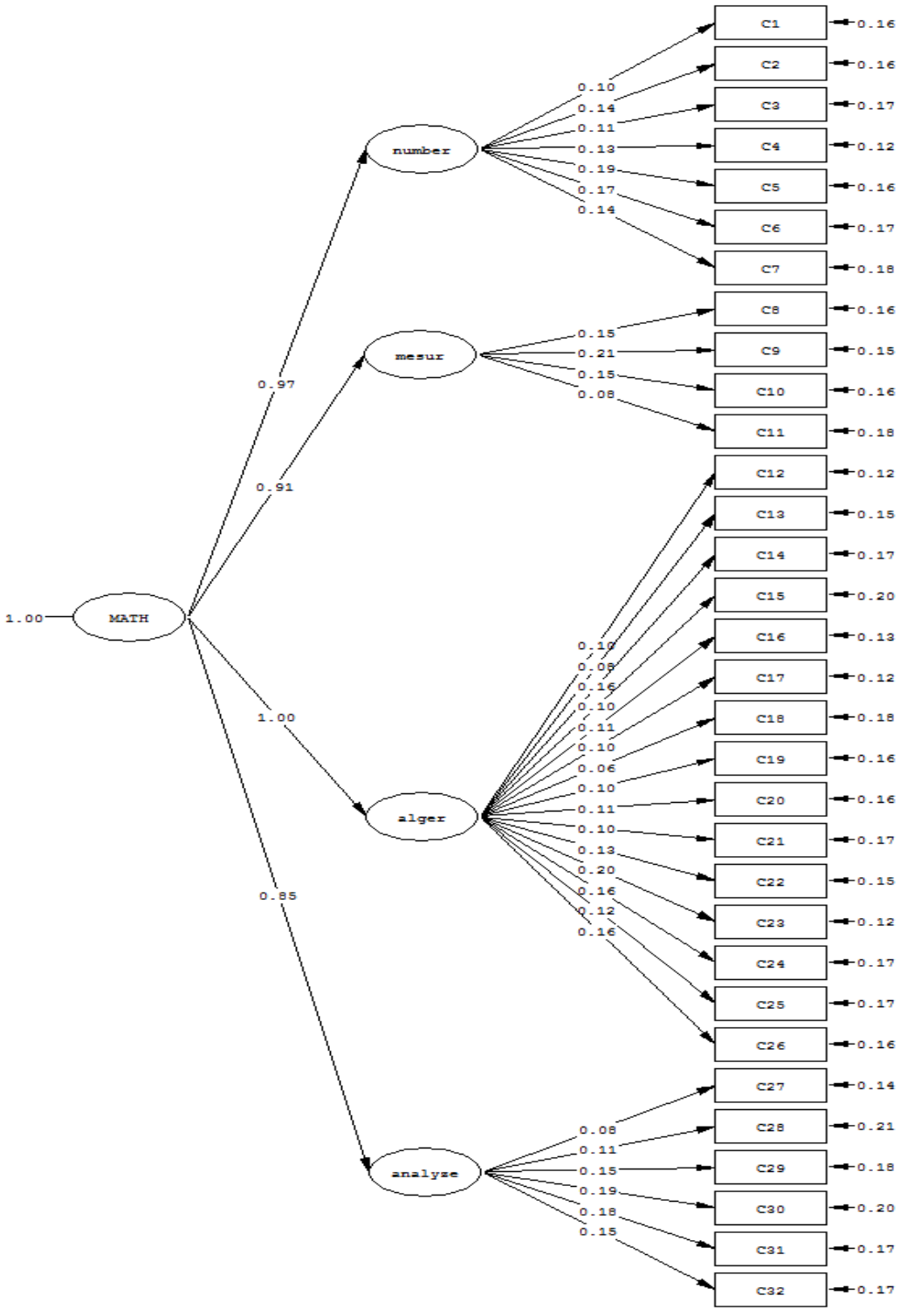
แบบทดสอบ	จำนวน (ข้อ)	ค่าความเที่ยง ก่อนตรวจพบ DIF	จำนวน (ข้อ)	ค่าความเที่ยง หลังตรวจพบ DIF
ภาษาไทย	70	0.867	57	0.847
สังคมศึกษา ศาสนาและวัฒนธรรม	65	0.669	52	0.604
ภาษาอังกฤษ	80	0.866	77	0.857
คณิตศาสตร์	32	0.775	30	0.753
วิทยาศาสตร์	80	0.776	69	0.728
สุขศึกษาและพลศึกษา	32	0.560	21	0.478
ศิลปะ	30	0.309	24	0.137
การงานอาชีพและเทคโนโลยี	41	0.724	35	0.667

จากตารางที่ 4-9 ผลการตรวจสอบค่าความเที่ยงของข้อสอบฉบับก่อนและฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกัน ตามหลักทฤษฎีการตอบสนองข้อสอบ (IRT) แบบ 3 พารามิเตอร์ พบว่า แบบทดสอบฉบับก่อนตัดข้อสอบที่ทำหน้าที่ต่างกัน มีค่าความเที่ยง (Reliability) มากกว่าฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกัน ทั้ง 8 กลุ่มสาระการเรียนรู้ โดยมี 2 กลุ่มสาระการเรียนรู้ ที่มีค่าความเที่ยงน้อยที่สุด คือ แบบทดสอบสุขศึกษาและพลศึกษา ก่อนตัดมีค่าเท่ากับ 0.560 และหลังตัดมีค่าเท่ากับ 0.478 และแบบทดสอบศิลปะก่อนตัดมีค่าเท่ากับ 0.309 และหลังตัดมีค่าเท่ากับ 0.137 ซึ่งเป็นไปตามสมมติฐานของการวิจัยข้อที่ 2

ตอนที่ 2 ผลการตรวจสอบความตรงเชิงโครงสร้างของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 จำนวน 8 กลุ่มสาระการเรียนรู้ ทั้งฉบับก่อนและฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกัน

การตรวจสอบความตรงเชิงโครงสร้าง ของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ฉบับก่อนตัดข้อสอบที่ทำหน้าที่ต่างกันออกจากแบบทดสอบใช้วิธีการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสอง (Second Oder Factor Analysis) มีการกำหนดเกณฑ์การพิจารณาดังนี้ ค่า Chi-Square ต้องมีค่าน้อยกว่าไคสแควร์เกณฑ์หรือค่าไคสแควร์สัมพัทธ์ น้อยกว่า 2.00 หรือตำราบางเล่มกล่าวว่าควรมีค่าน้อยกว่า 5.00 (Diamantopopulos & Siguaw, 2000) ถือเป็นค่าที่ยอมรับได้ ดัชนีตรวจสอบความกลมกลืน ได้แก่ค่า CFI , GFI, AGFI, NFI และ NNFI ต้องมีค่ามากกว่า 0.95 (พูลพงศ์ สุขสว่าง, 2561) ค่า NNFI และ CFI ที่ยอมรับได้ว่าสอดคล้อง ≤ 0.97 (สุวิมล ติรกานันท์, 2555) ในขณะที่ค่าความคลาดเคลื่อนของการประมาณค่า ได้แก่ SMSEA, RMR, SRMR ต้องมีค่าน้อยกว่า 0.05 (Schumaker & Lomax, 2010) เนื่องจากกลุ่มตัวอย่างมีขนาดใหญ่ การพิจารณาความสอดคล้องของโมเดลไม่ควรใช้ดัชนีชุดใดชุดหนึ่ง (สุวิมล ติรกานันท์, 2555) จึงนำค่า NCP มาใช้เป็นเกณฑ์ยอมรับสมมติฐาน ในกรณีที่มีการปรับโมเดล และค่า AIC และ ECVI ในการพิจารณาความเหมาะสมของโมเดล โดยใช้เกณฑ์มีค่าใกล้เคียงกับ Structure Model (สถาพร สุขสมบูรณ์, 2559) ซึ่งแสดงให้เห็นถึงความสอดคล้องของข้อมูลเชิงประจักษ์ จำนวน 8 กลุ่มสาระการเรียนรู้ ได้แก่ 1) ภาษาไทย 2) สังคมศึกษา ศาสนาและวัฒนาธรรม 3) ภาษาต่างประเทศ 4) คณิตศาสตร์ 5) วิทยาศาสตร์ 6) สุขศึกษาและพลศึกษา 7) ศิลปะ และ 8) การงานอาชีพและเทคโนโลยี จากการตรวจสอบความตรงเชิงโครงสร้างของแบบทดสอบ O-NET ฉบับก่อนตัดกับฉบับหลังตัด DIF ทั้ง 8 กลุ่มสาระการเรียนรู้ พบว่า 2 กลุ่มสาระการเรียนรู้ มีความสอดคล้องกับข้อมูลเชิงประจักษ์ ได้แก่ แบบทดสอบคณิตศาสตร์ และแบบทดสอบศิลปะ แสดงผลการทดสอบสมมติฐานทางสถิติได้ดังต่อไปนี้

1. ผลการตรวจสอบความตรงเชิงโครงสร้างของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ฉบับก่อนตัดข้อสอบที่ทำหน้าที่ต่างกันออกจากแบบทดสอบคณิตศาสตร์ จำนวน 32 ข้อ มี 4 องค์ประกอบ คือ องค์ประกอบด้าน 1) จำนวนและการดำเนินการ 2) การวัด 3) พิชคณิต 4) การวิเคราะห์ข้อมูลและความน่าจะเป็น จากกลุ่มตัวอย่าง 2000 คนผลการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับที่สอง ภาพที่ 4-1 และตารางที่ 4-10 ถึง ตารางที่ 4-11



ภาพที่ 4-1 โมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสองของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ฉบับก่อนตัด DIF ของแบบทดสอบคณิตศาสตร์

ตารางที่ 4-10 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสองของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ฉบับก่อนตัด DIF ของแบบทดสอบคณิตศาสตร์

ข้อที่	Number		Mesur		Alger		Analyze		r^2
	b_i	S.E	b_i	S.E	b_i	S.E	b_i	S.E	
1	0.10*	-	-	-	-	-	-	-	0.06
2	0.14*	0.02	-	-	-	-	-	-	0.11
3	0.11*	0.02	-	-	-	-	-	-	0.07
4	0.13*	0.02	-	-	-	-	-	-	0.13
5	0.19*	0.02	-	-	-	-	-	-	0.19
6	0.17*	0.02	-	-	-	-	-	-	0.15
7	0.14*	0.02	-	-	-	-	-	-	0.10
8	-	-	0.15*	-	-	-	-	-	0.13
9	-	-	0.21*	0.02	-	-	-	-	0.23
10	-	-	0.15*	0.02	-	-	-	-	0.12
11	-	-	0.08*	0.01	-	-	-	-	0.03
12	-	-	-	-	0.10*	-	-	-	0.08
13	-	-	-	-	0.08*	0.01	-	-	0.04
14	-	-	-	-	0.16*	0.02	-	-	0.13
15	-	-	-	-	0.10*	0.02	-	-	0.05
16	-	-	-	-	0.11*	0.02	-	-	0.08
17	-	-	-	-	0.10*	0.02	-	-	0.08
18	-	-	-	-	0.06*	0.01	-	-	0.02
19	-	-	-	-	0.10*	0.02	-	-	0.06
20	-	-	-	-	0.11*	0.02	-	-	0.07
21	-	-	-	-	0.10*	0.02	-	-	0.06
22	-	-	-	-	0.13*	0.02	-	-	0.10
23	-	-	-	-	0.20*	0.03	-	-	0.26
24	-	-	-	-	0.16*	0.02	-	-	0.14
25	-	-	-	-	0.12*	0.02	-	-	0.08
26	-	-	-	-	0.16*	0.02	-	-	0.14
27	-	-	-	-	-	-	0.08*	-	0.05
28	-	-	-	-	-	-	0.11*	0.02	0.06
29	-	-	-	-	-	-	0.15*	0.02	0.11

ตารางที่ 4-10 (ต่อ)

ข้อที่	Number		Mesur		Alger		Analyze		r^2
	b_i	S.E	b_i	S.E	b_i	S.E	b_i	S.E	
30	-	-	-	-	-	-	0.19*	0.03	0.15
31	-	-	-	-	-	-	0.18*	0.03	0.16
32	-	-	-	-	-	-	0.15*	0.02	0.11

ตัวแปรแฝง	Math		r^2
	b_i	S.E	
Number	0.97*	0.11	0.94
Mesur	0.91*	0.08	0.84
Alger	1.00*	0.14	1.08
Analyze	0.85*	0.11	0.72

ตารางที่ 4-11 ค่าสถิติความสอดคล้องของโมเดลความสัมพันธ์ตามสมมติฐานกับข้อมูลเชิงประจักษ์องค์ประกอบเชิงยืนยันอันดับสองของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ฉบับก่อนตัด DIF ของแบบทดสอบคณิตศาสตร์

สถิติที่ใช้วัดความสอดคล้อง	เกณฑ์	ค่าที่ได้	ผลการตรวจสอบ
χ^2		442.30	-
df		396	-
χ^2/df	$(\chi^2/df) < 2$	1.12	ผ่าน
p	$p > 0.05$	0.05	ผ่าน
RMSEA	ค่า <0.05 (เข้าใกล้ 0)	0.01	ผ่าน
CFI	ค่า *CFI* > 0.95 (เข้าใกล้ 1)	1.00	ผ่าน
NNFI	ค่า *NNFI* > 0.95 (เข้าใกล้ 1)	1.00	ผ่าน
GFI	ค่า *GFI* > 0.95 (เข้าใกล้ 1)	0.99	ผ่าน
AGFI	ค่า *AGFI* > 0.95 (เข้าใกล้ 1)	0.98	ผ่าน
RMR	ค่า RMR < 0.05 (เข้าใกล้ 0)	0.01	ผ่าน
SRMR	ค่า SRMR < 0.05 (เข้าใกล้ 0)	0.02	ผ่าน
NCP	ยอมรับสมมุติฐาน	46.30	ผ่าน

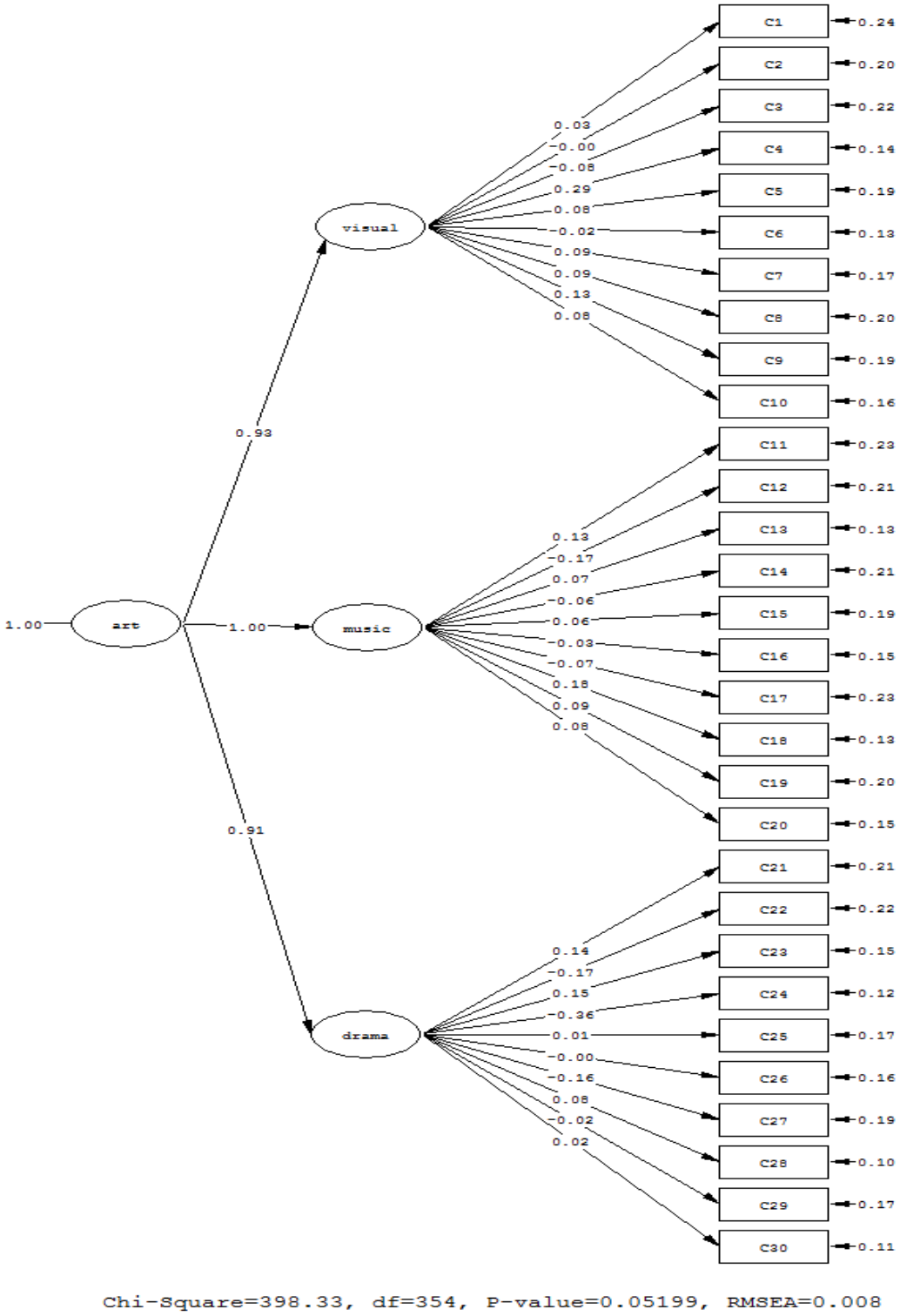
จากภาพที่ 4-1 และตารางที่ 4-10 ถึง ตารางที่ 4-11 พบว่า ผลการวิเคราะห์ความตรงเชิงโครงสร้างของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ฉบับก่อนตัด DIF ของแบบทดสอบ

คณิตศาสตร์ จำนวน 32 ข้อ มีองค์ประกอบย่อย 4 องค์ประกอบ ได้แก่ 1) จำนวนและการดำเนินการ มี 7 ตัวแปรสังเกตได้ 2) การวัด มี 4 ตัวแปรสังเกตได้ 3) พีชคณิต มี 15 ตัวแปรสังเกตได้ 4) การวิเคราะห์ข้อมูลและความน่าจะเป็น มี 6 ตัวแปรสังเกตได้ ดังตารางที่ 4-11 และตารางที่ 4-12 พบว่าผลการวิเคราะห์ทำให้ได้ค่าดัชนีต่างๆ ได้แก่

ค่าสัมประสิทธิ์พยากรณ์ (Coefficient of Determination: r^2) มีพิสัยของค่าสัมประสิทธิ์พยากรณ์ เท่ากับ 0.02 - 0.26 โดยข้อคำถามที่ 23 มีค่าน้ำหนักองค์ประกอบสูงที่สุด และข้อคำถามข้อที่ 18 มีค่าน้ำหนักองค์ประกอบต่ำที่สุด

ผลการวิเคราะห์ความสอดคล้องของโมเดลความสัมพันธ์ตามสมมติฐานกับข้อมูลเชิงประจักษ์องค์ประกอบเชิงยืนยันอันดับสองของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ฉบับก่อนตัด DIF ของแบบทดสอบคณิตศาสตร์ พบว่า ค่าไค-สแควร์ ($\boldsymbol{x^2}$) เท่ากับ 442.30 ที่องศาอิสระ (*df*) เท่ากับ 396 เมื่อพิจารณาไค-สแควร์สัมพัทธ์ (Relative $\boldsymbol{x^2}$) ได้เท่ากับ 1.12 ถือว่าโมเดลมีความสอดคล้องระดับดี ค่าความน่าจะเป็น (*p*) เท่ากับ 0.05 เมื่อพิจารณาค่าไค-สแควร์ โดยใช้ค่า NCP ในกรณีที่ผลลัพธ์เป็น 0.12 (NCP = 46.30, *df* = 396) ซึ่งถือว่ามีความสอดคล้องในระดับดีมาก ผลการวิเคราะห์ข้อมูลเชิงประจักษ์พบว่าเมื่อปรับโมเดลแล้ว มีค่า RMSEA เท่ากับ 0.01 ค่า RMR เท่ากับ 0.01 ค่า SRMR เท่ากับ 0.02 ซึ่งมีค่าน้อยกว่า 0.05 และค่าดัชนี GFI เท่ากับ 0.99 ค่า AGFI เท่ากับ 0.98 ค่า NFI เท่ากับ 0.96 และค่า NNFI เท่ากับ 1.00 ซึ่งมีค่ามากกว่า 0.95 แสดงให้เห็นว่าแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ฉบับก่อนตัด DIF ของแบบทดสอบคณิตศาสตร์ มีความสอดคล้องของโมเดลความสัมพันธ์ตามสมมติฐานกับข้อมูลเชิงประจักษ์

2. ผลการตรวจสอบความตรงเชิงโครงสร้างของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ฉบับก่อนตัด DIF ของแบบทดสอบศิลปะ จำนวน 25 ข้อ มี 3 องค์ประกอบ คือ องค์ประกอบด้าน 1) ทัศนศิลป์ 2) ดนตรี 3) นาฏศิลป์ จากกลุ่มตัวอย่าง 2000 คน ผลการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับที่สอง ดังภาพที่ 4-2 และตารางที่ 4-12 ถึงตารางที่ 4-13



ภาพที่ 4-2 โมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสองของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ฉบับก่อนตัด DIF ของแบบทดสอบศิลปะ

ตารางที่ 4-12 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสองของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ฉบับก่อนตัดตัด DIF ของแบบทดสอบศิลปะ

ข้อที่	Visual		Music		Drama		r^2
	b_i	S.E	b_i	S.E	b_i	S.E	
1	0.03*	-	-	-	-	-	0.00
2	0.00*	0.01	-	-	-	-	0.00
3	-0.08*	0.05	-	-	-	-	0.03
4	0.29*	0.18	-	-	-	-	0.36
5	0.08*	0.05	-	-	-	-	0.04
6	-0.02*	0.02	-	-	-	-	0.00
7	0.09*	0.06	-	-	-	-	0.04
8	0.09*	0.06	-	-	-	-	0.04
9	0.13*	0.08	-	-	-	-	0.09
10	0.08*	0.05	-	-	-	-	0.04
11	-	-	0.13*	-	-	-	0.07
12	-	-	-0.17*	0.03	-	-	0.11
13	-	-	0.07*	0.02	-	-	0.04
14	-	-	-0.06*	0.02	-	-	0.02
15	-	-	0.06*	0.02	-	-	0.02
16	-	-	-0.03*	0.01	-	-	0.01
17	-	-	-0.07*	0.02	-	-	0.02
18	-	-	0.18*	0.03	-	-	0.20
19	-	-	0.09*	0.02	-	-	0.04
20	-	-	0.08*	0.02	-	-	0.04
21	-	-	-	-	0.14*	-	0.09
22	-	-	-	-	-0.17*	0.02	0.12
23	-	-	-	-	0.15*	0.02	0.13
24	-	-	-	-	-0.36*	0.04	0.51
25	-	-	-	-	0.01*	0.01	0.00
26	-	-	-	-	0.00*	0.01	0.00
27	-	-	-	-	-0.16*	0.02	0.12
28	-	-	-	-	0.08*	0.01	0.06
29	-	-	-	-	-0.02*	0.01	0.00

ตารางที่ 4-12 (ต่อ)

ข้อที่	Visual		Music		Drama		r^2
	b_i	S.E	b_i	S.E	b_i	S.E	
30	-	-	-	-	0.02*	0.01	0.00

ตัวแปรแฝง	Art		r^2
	b_i	S.E	
Visual	0.14*	0.19	0.86
Music	0.21*	0.11	0.79
Drama	0.18*	0.08	0.82

ตารางที่ 4-13 ค่าสถิติความสอดคล้องของโมเดลความสัมพันธ์ตามสมมติฐานกับข้อมูลเชิงประจักษ์องค์ประกอบเชิงยืนยันอันดับสองของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ฉบับก่อนตัด DIF ของแบบทดสอบศิลปะ

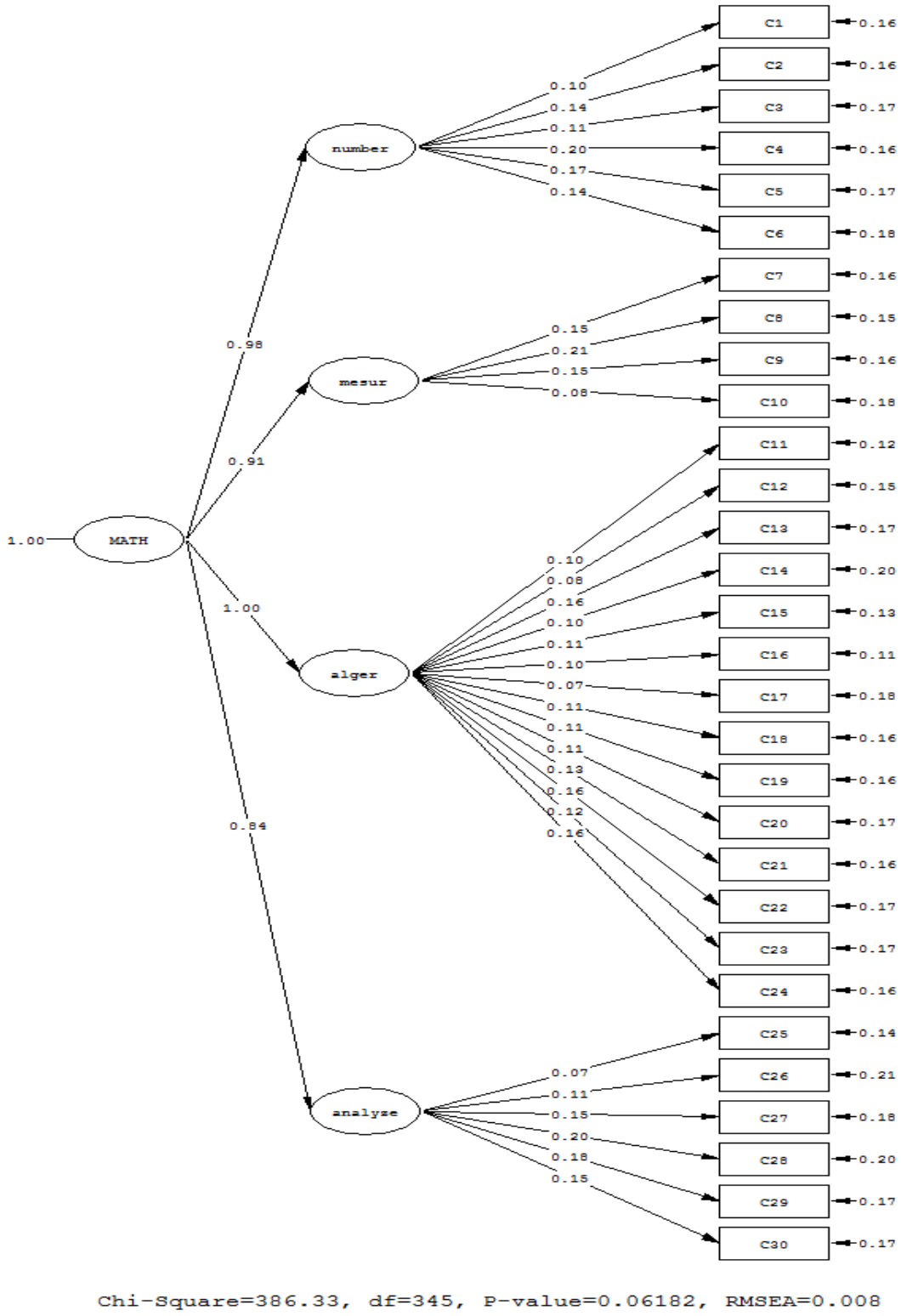
สถิติที่ใช้วัดความสอดคล้อง	เกณฑ์	ค่าที่ได้	ผลการตรวจสอบ
χ^2		398.33	-
df		354	-
χ^2/df	$(\chi^2/df) < 2$	1.12	ผ่าน
p	$p > 0.05$	0.05	ผ่าน
RMSEA	ค่า <0.05 (เข้าใกล้ 0)	0.01	ผ่าน
CFI	ค่า *CFI* > 0.95 (เข้าใกล้ 1)	0.99	ผ่าน
NNFI	ค่า *NNFI* > 0.95 (เข้าใกล้ 1)	0.99	ผ่าน
GFI	ค่า *GFI* > 0.95 (เข้าใกล้ 1)	0.99	ผ่าน
AGFI	ค่า *AGFI* > 0.95 (เข้าใกล้ 1)	0.98	ผ่าน
RMR	ค่า RMR < 0.05 (เข้าใกล้ 0)	0.01	ผ่าน
SRMR	ค่า SRMR < 0.05 (เข้าใกล้ 0)	0.02	ผ่าน
NCP	ยอมรับสมมุติฐาน	44.33	ผ่าน

จากภาพที่ 4-2 และตารางที่ 4-12 ถึง ตารางที่ 4-13 พบว่า ผลการวิเคราะห์ความตรงเชิงโครงสร้างของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ฉบับก่อนตัด DIF ของแบบทดสอบศิลปะ จำนวน 30 ข้อ มีองค์ประกอบย่อย 3 องค์ประกอบ ได้แก่ 1) ทัศนศิลป์ มี 10 ตัวแปรสังเกตได้ 2) ดนตรี มี 10 ตัวแปรสังเกตได้ 3) นาฏศิลป์ มี 10 ตัวแปรสังเกตได้ ดังตารางที่ 4-12 และตารางที่ 4-13 พบว่าผลการวิเคราะห์ทำให้ได้ค่าดัชนีต่างๆ ได้แก่

ค่าสัมประสิทธิ์พยากรณ์ (Coefficient of Determiantion: r^2) มีพิสัยของค่าสัมประสิทธิ์พยากรณ์ เท่ากับ 0.00 - 0.51 โดยข้อคำถามที่ 24 มีค่าน้ำหนักองค์ประกอบสูงที่สุด และข้อคำถามข้อที่ 1, 2, 6, 25, 26, 29 และ 30 มีค่าน้ำหนักองค์ประกอบต่ำที่สุด

ผลการวิเคราะห์ความสอดคล้องของโมเดลความสัมพันธ์ตามสมมติฐานกับข้อมูลเชิงประจักษ์องค์ประกอบเชิงยืนยันอันดับสองของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ฉบับก่อนตัด DIF ของแบบทดสอบศิลปะ พบว่า ค่าไค-สแควร์ ($\boldsymbol{\chi^2}$) เท่ากับ 398.33 ที่องศาอิสระ (*df*) เท่ากับ 354 เมื่อพิจารณาไค-สแควร์สัมพัทธ์ (Relative $\boldsymbol{\chi^2}$) ได้เท่ากับ 1.12 ถือว่าโมเดลมีความสอดคล้องระดับดี ค่าความน่าจะเป็น (*p*) เท่ากับ 0.05 เมื่อพิจารณาค่าไค-สแควร์โดยใช้ค่า NCP ในกรณีที่ผลลัพธ์เป็น 0.12 (NCP = 44.33, *df* = 354) ซึ่งถือว่ามีความสอดคล้องในระดับดีมาก ผลการวิเคราะห์ข้อมูลเชิงประจักษ์พบว่าเมื่อปรับโมเดลแล้ว มีค่า RMSEA เท่ากับ 0.01 ค่า RMR เท่ากับ 0.01 ค่า SRMR เท่ากับ 0.02 ซึ่งมีค่าน้อยกว่า 0.05 และค่าดัชนี GFI เท่ากับ 0.99 ค่า AGFI เท่ากับ 0.98 ค่า NFI เท่ากับ 0.92 และค่า NNFI เท่ากับ 0.99 ซึ่งมีค่ามากกว่า 0.95 แสดงให้เห็นว่า แบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ฉบับก่อนตัด DIF ของแบบทดสอบศิลปะมีความสอดคล้องของโมเดลความสัมพันธ์ตามสมมติฐานกับข้อมูลเชิงประจักษ์

3. ผลการตรวจสอบความตรงเชิงโครงสร้างของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ฉบับหลังตัด DIF ของแบบทดสอบคณิตศาสตร์ จำนวน 30 ข้อ มี 4 องค์ประกอบ คือ องค์ประกอบด้าน 1) จำนวนและการดำเนินการ 2) การวัด 3) พีชคณิต 4) การวิเคราะห์ข้อมูลและความน่าจะเป็น จากกลุ่มตัวอย่าง 2000 คนผลการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับที่สองดังภาพที่ 4-3 และตารางที่ 4-14 ถึง ตารางที่ 4-15



ภาพที่ 4-3 โมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสองของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ฉบับหลังตัด DIF ของแบบทดสอบคณิตศาสตร์

ตารางที่ 4-14 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสองของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ฉบับหลังตัด DIF ของแบบทดสอบคณิตศาสตร์

ข้อที่	Number		Mesur		Alger		Analyze		r^2
	b_i	S.E	b_i	S.E	b_i	S.E	b_i	S.E	
1	0.10*	-	-	-	-	-	-	-	0.06
2	0.14*	0.02	-	-	-	-	-	-	0.11
3	0.11*	0.02	-	-	-	-	-	-	0.07
4	0.20*	0.02	-	-	-	-	-	-	0.19
5	0.17*	0.02	-	-	-	-	-	-	0.14
6	0.14*	0.02	-	-	-	-	-	-	0.10
7	-	-	0.15*	-	-	-	-	-	0.13
8	-	-	0.21*	0.02	-	-	-	-	0.23
9	-	-	0.15*	0.02	-	-	-	-	0.12
10	-	-	0.08*	0.01	-	-	-	-	0.03
11	-	-	-	-	0.10*	-	-	-	0.08
12	-	-	-	-	0.08*	0.01	-	-	0.04
13	-	-	-	-	0.16*	0.02	-	-	0.13
14	-	-	-	-	0.10*	0.02	-	-	0.05
15	-	-	-	-	0.11*	0.02	-	-	0.08
16	-	-	-	-	0.10*	0.02	-	-	0.08
17	-	-	-	-	0.07*	0.01	-	-	0.03
18	-	-	-	-	0.11*	0.02	-	-	0.07
19	-	-	-	-	0.11*	0.02	-	-	0.07
20	-	-	-	-	0.11*	0.02	-	-	0.06
21	-	-	-	-	0.13*	0.02	-	-	0.10
22	-	-	-	-	0.16*	0.02	-	-	0.14
23	-	-	-	-	0.12*	0.02	-	-	0.08
24	-	-	-	-	0.16*	0.02	-	-	0.14
25	-	-	-	-	-	-	0.07*	-	0.04
26	-	-	-	-	-	-	0.11*	0.02	0.06
27	-	-	-	-	-	-	0.15*	0.03	0.11
28	-	-	-	-	-	-	0.20*	0.03	0.16
29	-	-	-	-	-	-	0.18*	0.03	0.16

ตารางที่ 4-14 (ต่อ)

ข้อที่	Number		Mesur		Alger		Analyze		r^2
	b_i	S.E	b_i	S.E	b_i	S.E	b_i	S.E	
30	-	-	-	-	-	-	0.15*	0.03	0.12

ตัวแปรแฝง	Math		r^2
	b_i	S.E	
Number	0.98*	0.11	0.96
Mesur	0.91*	0.07	0.83
Alger	1.00*	0.04	1.04
Analyze	0.84*	0.13	0.71

ตารางที่ 4-15 ค่าสถิติความสอดคล้องของโมเดลความสัมพันธ์ตามสมมติฐานกับข้อมูลเชิงประจักษ์องค์ประกอบเชิงยืนยันอันดับสองของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ฉบับหลังตัด DIF ของแบบทดสอบคณิตศาสตร์

สถิติที่ใช้วัดความสอดคล้อง	เกณฑ์	ค่าที่ได้	ผลการตรวจสอบ
χ^2		386.33	-
df		345	-
χ^2/df	$(\chi^2/df) < 2$	1.12	ผ่าน
p	$p > 0.05$	0.06	ผ่าน
RMSEA	ค่า <0.05 (เข้าใกล้ 0)	0.01	ผ่าน
CFI	ค่า *CFI* > 0.95 (เข้าใกล้ 1)	1.00	ผ่าน
NNFI	ค่า *NNFI* > 0.95 (เข้าใกล้ 1)	0.99	ผ่าน
GFI	ค่า *GFI* > 0.95 (เข้าใกล้ 1)	0.99	ผ่าน
AGFI	ค่า *AGFI* > 0.95 (เข้าใกล้ 1)	0.98	ผ่าน
RMR	ค่า RMR < 0.05 (เข้าใกล้ 0)	0.01	ผ่าน
SRMR	ค่า SRMR < 0.05 (เข้าใกล้ 0)	0.02	ผ่าน
NCP	ยอมรับสมมุติฐาน	41.33	ผ่าน

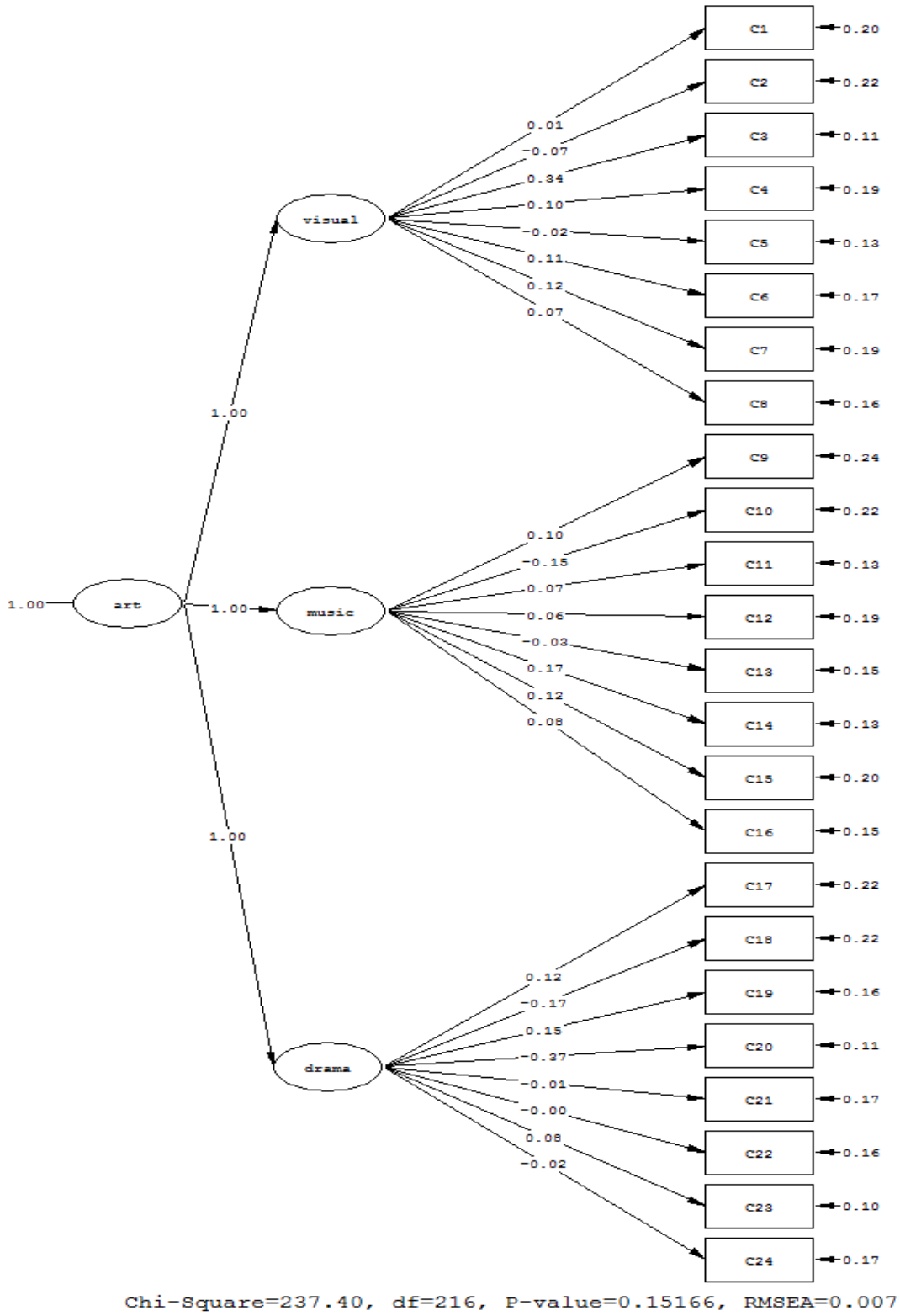
จากภาพที่ 4-3 และตารางที่ 4-14 ถึง ตารางที่ 4-15 พบว่า ผลการวิเคราะห์ความตรงเชิงโครงสร้างของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ฉบับหลังตัด DIF ของแบบทดสอบคณิตศาสตร์ จำนวน 30 ข้อ มีองค์ประกอบย่อย 4 องค์ประกอบ ได้แก่ 1) จำนวนและการดำเนินการ

มี 6 ตัวแปรสังเกตได้ 2) การวัด มี 4 ตัวแปรสังเกตได้ 3) พีชคณิต มี 14 ตัวแปรสังเกตได้ 4) การวิเคราะห์ข้อมูลและความน่าจะเป็น มี 6 ตัวแปรสังเกตได้ ดังตารางที่ 4-14 และตารางที่ 4-15 พบว่าผลการวิเคราะห์ทำให้ได้ค่าต่างๆ ได้แก่

ค่าสัมประสิทธิ์พยากรณ์ (Coefficient of Determination: r^2) มีพิสัยของค่าสัมประสิทธิ์พยากรณ์ เท่ากับ 0.03 - 0.23 โดยข้อคำถามที่ 8 มีค่าน้ำหนักองค์ประกอบสูงที่สุด และข้อคำถามข้อที่ 10 และ 17 มีค่าน้ำหนักองค์ประกอบต่ำที่สุด

ผลการวิเคราะห์ความสอดคล้องของโมเดลความสัมพันธ์ตามสมมติฐานกับข้อมูลเชิงประจักษ์องค์ประกอบเชิงยืนยันอันดับสองของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ฉบับหลังตัด DIF ของแบบทดสอบคณิตศาสตร์ พบว่า ค่าไค-สแควร์ ($\boldsymbol{x^2}$) เท่ากับ 386.33 ที่องศาอิสระ (*df*) เท่ากับ 345 เมื่อพิจารณาไค-สแควร์สัมพัทธ์ (Relative $\boldsymbol{x^2}$) ได้เท่ากับ 1.12 ถือว่าโมเดลมีความสอดคล้องระดับดี ค่าความน่าจะเป็น (*p*) เท่ากับ 0.06 เมื่อพิจารณาค่าไค-สแควร์ โดยใช้ค่า NCP ในกรณีที่ผลลัพธ์เป็น 0.12 (NCP = 41.33, *df* = 345) ซึ่งถือว่ามีความสอดคล้องในระดับดีมาก ผลการวิเคราะห์ข้อมูลเชิงประจักษ์พบว่าเมื่อปรับโมเดลแล้ว มีค่า RMSEA เท่ากับ 0.01 ค่า RMR เท่ากับ 0.01 ค่า SRMR เท่ากับ 0.02 ซึ่งมีค่าน้อยกว่า 0.05 และค่าดัชนี GFI เท่ากับ 0.99 ค่า AGFI เท่ากับ 0.98 ค่า NFI เท่ากับ 0.96 และค่า NNFI เท่ากับ 0.99 ซึ่งมีค่ามากกว่า 0.95 แสดงให้เห็นว่าแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ฉบับก่อนตัด DIF ของแบบทดสอบคณิตศาสตร์ มีความสอดคล้องของโมเดลความสัมพันธ์ตามสมมติฐานกับข้อมูลเชิงประจักษ์

4. ผลการตรวจสอบความตรงเชิงโครงสร้างของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ฉบับหลังตัด DIF ของแบบทดสอบศิลปะ จำนวน 24 ข้อ มี 3 องค์ประกอบ คือ องค์ประกอบด้าน 1) ทัศนศิลป์ 2) ดนตรี 3) นาฏศิลป์ จากกลุ่มตัวอย่าง 2000 คน ผลการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับที่สอง ดังภาพที่ 4-4 และตารางที่ 4-16 ถึง ตารางที่ 4-17



ภาพที่ 4-4 โมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสองของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ฉบับหลังตัด DIF ของแบบทดสอบศิลปะ

ตารางที่ 4-16 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสองของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ฉบับหลังตัด DIF ของแบบทดสอบศิลปะ

ข้อที่	Visual		Music		Drama		r^2
	b_i	S.E	b_i	S.E	b_i	S.E	
1	0.01*	-	-	-	-	-	0.00
2	-0.07*	0.07	-	-	-	-	0.02
3	0.34*	0.32	-	-	-	-	0.51
4	0.10*	0.10	-	-	-	-	0.05
5	-0.02*	0.02	-	-	-	-	0.00
6	0.11*	0.11	-	-	-	-	0.07
7	0.12*	0.11	-	-	-	-	0.07
8	0.07*	0.07	-	-	-	-	0.03
9	-	-	0.10*	-	-	-	0.04
10	-	-	-0.15*	0.04	-	-	0.10
11	-	-	0.07*	0.02	-	-	0.04
12	-	-	0.06*	0.02	-	-	0.02
13	-	-	-0.03*	0.01	-	-	0.01
14	-	-	0.17*	0.04	-	-	0.18
15	-	-	0.12*	0.03	-	-	0.07
16	-	-	0.08*	0.02	-	-	0.04
17	-	-	-	-	0.12*	-	0.06
18	-	-	-	-	-0.17*	0.04	0.12
19	-	-	-	-	0.15*	0.03	0.12
20	-	-	-	-	-0.37*	0.08	0.55
21	-	-	-	-	-0.01*	0.01	0.00
22	-	-	-	-	0.00*	0.01	0.00
23	-	-	-	-	0.08*	0.02	0.06
24	-	-	-	-	-0.02*	0.01	0.00

ตัวแปรแฝง	Art		r^2
	b_i	S.E	
Visual	1.00*	0.73	0.59
Music	1.00*	0.26	1.02
Drama	1.00*	0.19	0.74

ตารางที่ 4- 17 ค่าสถิติความสอดคล้องของโมเดลความสัมพันธ์ตามสมมติฐานกับข้อมูลเชิงประจักษ์องค์ประกอบเชิงยืนยันอันดับสองของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ฉบับหลังตัด DIF ของแบบทดสอบศิลปะ

สถิติที่ใช้วัดความสอดคล้อง	เกณฑ์	ค่าที่ได้	ผลการตรวจสอบ
χ^2		237.15	-
df		216	-
χ^2/df	$(\chi^2/df) < 2$	1.09	ผ่าน
p	$p > 0.05$	0.15	ผ่าน
RMSEA	ค่า <0.05 (เข้าใกล้ 0)	0.01	ผ่าน
CFI	ค่า *CFI* > 0.95 (เข้าใกล้ 1)	0.99	ผ่าน
NNFI	ค่า *NNFI* > 0.95 (เข้าใกล้ 1)	0.99	ผ่าน
GFI	ค่า *GFI* > 0.95 (เข้าใกล้ 1)	0.99	ผ่าน
AGFI	ค่า *AGFI* > 0.95 (เข้าใกล้ 1)	0.99	ผ่าน
RMR	ค่า RMR < 0.05 (เข้าใกล้ 0)	0.01	ผ่าน
SRMR	ค่า SRMR < 0.05 (เข้าใกล้ 0)	0.02	ผ่าน
NCP	ยอมรับสมมุติฐาน	21.40	ผ่าน

จากภาพที่ 4-4 และตารางที่ 4-16 ถึง ตารางที่ 4-17 พบว่า ผลการวิเคราะห์ความตรงเชิงโครงสร้างของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ฉบับหลังตัด DIF ของแบบทดสอบศิลปะ จำนวน 24 ข้อ มีองค์ประกอบย่อย 3 องค์ประกอบ ได้แก่ 1) ทัศนศิลป์ มี 8 ตัวแปรสังเกตได้ 2) ดนตรี มี 8 ตัวแปรสังเกตได้ 3) นาฏศิลป์ มี 8 ตัวแปรสังเกตได้ ดังตารางที่ 4-16 และตารางที่ 4-17 พบว่าผลการวิเคราะห์ทำให้ได้ค่าดัชนีต่างๆ ได้แก่

ค่าสัมประสิทธิ์พยากรณ์ (Coefficient of Determiantion: r^2) มีพิสัยของค่าสัมประสิทธิ์พยากรณ์ เท่ากับ 0.00 - 0.55 โดยข้อคำถามที่ 20 มีค่าน้ำหนักองค์ประกอบสูงที่สุด และข้อคำถามข้อที่ 1, 5, 21, 22 และ 24 มีค่าน้ำหนักองค์ประกอบต่ำที่สุด

ผลการวิเคราะห์ความสอดคล้องของโมเดลความสัมพันธ์ตามสมมติฐานกับข้อมูลเชิงประจักษ์องค์ประกอบเชิงยืนยันอันดับสองของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ฉบับหลังตัด DIF ของแบบทดสอบศิลปะ พบว่า ค่าไค-สแควร์ ($\boldsymbol{\chi^2}$) เท่ากับ 237.15 ที่องศาอิสระ (*df*) เท่ากับ 216 เมื่อพิจารณาไค-สแควร์สัมพัทธ์ (Relative $\boldsymbol{\chi^2}$) ได้เท่ากับ 1.09 ถือว่าโมเดลมีความสอดคล้องระดับดี ค่าความน่าจะเป็น (*p*) เท่ากับ 0.15 เมื่อพิจารณาค่าไค-สแควร์โดยใช้ค่า NCP ในกรณีที่ผลลัพธ์เป็น 0.09 (NCP = 21.40, *df* = 216) ซึ่งถือว่ามีความสอดคล้องในระดับดีมาก ผลการวิเคราะห์ข้อมูลเชิงประจักษ์พบว่าเมื่อปรับโมเดลแล้ว มีค่า RMSEA เท่ากับ 0.01 ค่า RMR เท่ากับ 0.01 ค่า SRMR เท่ากับ 0.02 ซึ่งมีค่าน้อยกว่า 0.05 และค่าดัชนี GFI เท่ากับ

0.99 ค่า AGFI เท่ากับ 0.99 ค่า NFI เท่ากับ 0.94 และค่า NNFI เท่ากับ 0.99 ซึ่งมีค่ามากกว่า 0.95 แสดงให้เห็นว่า แบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ฉบับก่อนตัด DIF ของแบบทดสอบคณิตศาสตร์ มีความสอดคล้องของโมเดลความสัมพันธ์ตามสมมติฐานกับข้อมูลเชิงประจักษ์

ตอนที่ 3 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ทั้ง 8 กลุ่มสาระการเรียนรู้ ด้วยวิธี IRT-LR และ วิธี SIBTEST ฉบับก่อนตัดข้อสอบที่ทำหน้าที่ต่างกัน

ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ได้จากการวิเคราะห์ด้วยวิธี IRT-LR และวิธี SIBTEST ของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 จำนวน 8 กลุ่มสาระการเรียนรู้ ได้แก่ 1) ภาษาไทย 2) สังคมศึกษา ศาสนาและวัฒนาธรรม 3) ภาษาต่างประเทศ 4) คณิตศาสตร์ 5) วิทยาศาสตร์ 6) สุขศึกษาและพลศึกษา 7) ศิลปะ และ 8) การงานอาชีพและเทคโนโลยี แสดงผลการทดสอบสมมติฐานทางสถิติได้ ดังนี้

1. ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบภาษาไทย ด้วยวิธี IRT-LR และวิธี SIBTEST จำนวน 70 ข้อ ดังตารางที่ 4-18

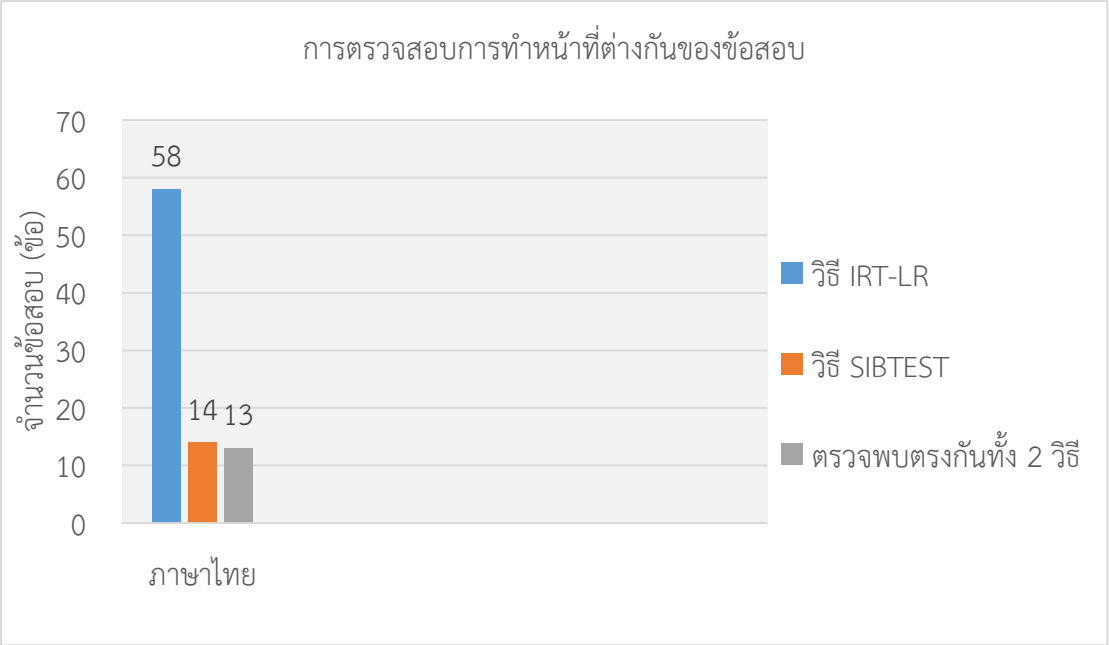
ตารางที่ 4-18 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบภาษาไทย ด้วยวิธี IRT-LR และวิธี SIBTEST จำนวน 70 ข้อ

แบบทดสอบ	จำนวน (ข้อ)	ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ			
		วิธี IRT-LR		วิธี SIBTEST	
		DIF	NO DIF	DIF	NO DIF
ภาษาไทย	70	58 ข้อ (1-13, 15-16, 18-19, 21-29, 33-39, 41-51, 53-54, 58-66, 68-70)	12 ข้อ	14 ข้อ (1, 9, 15-16, 22, 36, 44, 50, 53, 64-68)	56 ข้อ
คิดเป็นร้อยละของข้อสอบที่พบ DIF		82.86		20.00	

หมายเหตุ DIF หมายถึง ข้อสอบที่พบว่าทำหน้าที่ต่างกัน
NO DIF หมายถึง ข้อสอบที่ไม่พบว่าทำหน้าที่ต่างกัน

จากตารางที่ 4-18 พบว่า แบบทดสอบภาษาไทย จำนวน 70 ข้อ วิธี IRT-LR ตรวจพบการทำหน้าที่ต่างกันของข้อสอบ จำนวน 58 ข้อ 1-13, 15-16, 18-19, 21-29, 33-39, 41-51, 53-54, 58-66, 68-70 คิดเป็นร้อยละ 82.86 และวิธี SIBTEST ตรวจพบการทำหน้าที่ต่างกันของข้อสอบ จำนวน 14 ข้อ ได้แก่ข้อที่ 1, 9, 15-16, 22, 36, 44, 50, 53, 64-68 คิดเป็นร้อยละ 20.00 ซึ่งสอดคล้องกับสมมติฐานของการวิจัยข้อที่ 4

จากผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบภาษาไทย ด้วยวิธี IRT-LR และวิธี SIBTEST ดังภาพที่ 4-5



ภาพที่ 4-5 จำนวนข้อสอบที่ตรวจพบว่าทำหน้าที่ต่างกันในแบบทดสอบภาษาไทย ด้วยวิธี IRT-LR และวิธี SIBTEST

2. ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบสังคมศึกษา ศาสนาและวัฒนธรรม ด้วยวิธี IRT-LR และวิธี SIBTEST จำนวน 65 ข้อ ดังตารางที่ 4-19

ตารางที่ 4-19 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบสังคมศึกษา ศาสนา และวัฒนธรรม ด้วยวิธี IRT-LR และวิธี SIBTEST จำนวน 65 ข้อ

แบบทดสอบ	จำนวน (ข้อ)	ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ			
		วิธี IRT-LR		วิธี SIBTEST	
		DIF	NO DIF	DIF	NO DIF
สังคมศึกษา ศาสนา และ วัฒนธรรม	65	59 ข้อ (1-32, 34,36-41, 43-45, 47-52, 54-55, 57-65)	6 ข้อ	13 ข้อ (3, 6-7, 14, 16, 20-21, 25, 27, 39, 45, 58, 64)	52 ข้อ
คิดเป็นร้อยละของข้อสอบที่พบ DIF		90.77		20.00	

หมายเหตุ DIF หมายถึง ข้อสอบที่พบว่าทำหน้าที่ต่างกัน
NO DIF หมายถึง ข้อสอบที่ไม่พบว่าทำหน้าที่ต่างกัน

จากตารางที่ 4-19 พบว่า แบบทดสอบสังคมศึกษา ศาสนาและวัฒนธรรม จำนวน 65 ข้อ ด้วยวิธี IRT-LR ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน จำนวน 59 ข้อ 1-32, 34,36-41, 43-45, 47-52, 54-55, 57-65 คิดเป็นร้อยละ 90.77 และวิธี SIBTEST มีข้อสอบที่ทำหน้าที่ต่างกัน จำนวน 13 ข้อ ได้แก่ข้อที่ 3, 6-7, 14, 16, 20-21, 25, 27, 39, 45, 58, 64 คิดเป็นร้อยละ 20.00 ซึ่งสอดคล้องกับสมมติฐานของการวิจัยข้อที่ 4

จากผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบสังคมศึกษา ศาสนา และวัฒนธรรม ด้วยวิธี IRT-LR และวิธี SIBTEST ดังภาพที่ 4-6

ภาพที่ 4- 6 จำนวนข้อสอบที่ตรวจพบพบว่าทำหน้าที่ต่างกันในแบบทดสอบสังคมศึกษา ศาสนา และวัฒนธรรม ด้วยวิธี IRT-LR และวิธี SIBTEST

3. ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบภาษาอังกฤษ ด้วยวิธี IRT-LR และวิธี SIBTEST จำนวน 80 ข้อ ดังตารางที่ 4-20

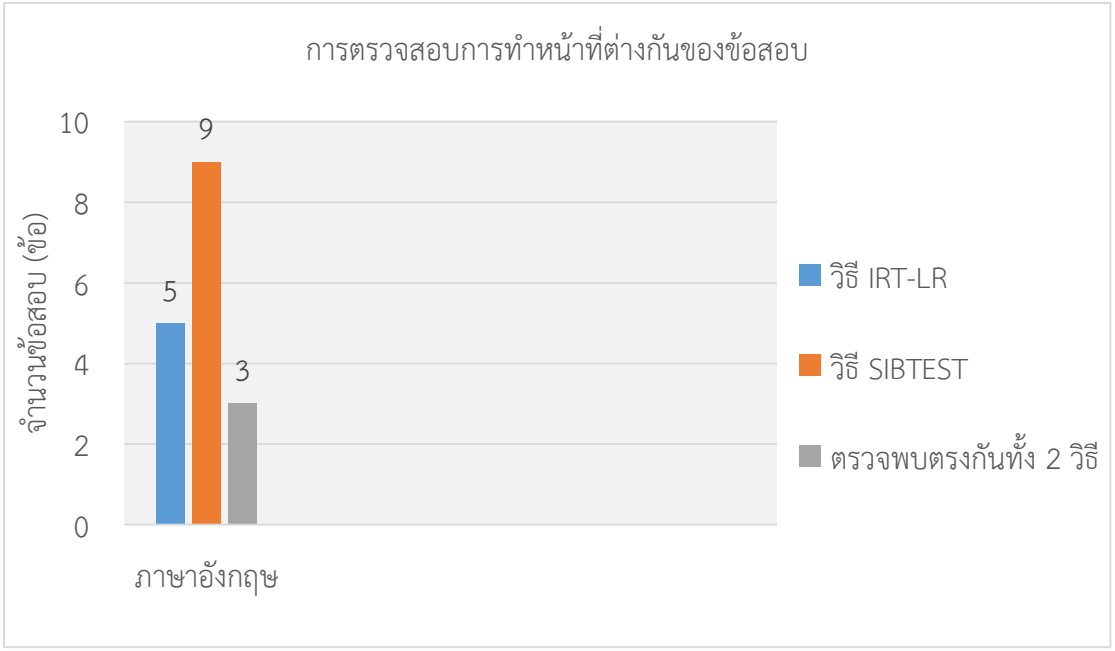
ตารางที่ 4-20 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบภาษาอังกฤษ ด้วยวิธี IRT-LR และวิธี SIBTEST จำนวน 80 ข้อ

แบบทดสอบ	จำนวน (ข้อ)	ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ			
		วิธี IRT-LR		วิธี SIBTEST	
		DIF	NO DIF	DIF	NO DIF
ภาษาอังกฤษ	80	5 ข้อ (26, 44, 59, 68, 74)	75 ข้อ	9 ข้อ (6, 16, 26, 32, 43, 50, 58, 59, 74)	71 ข้อ
คิดเป็นร้อยละของข้อสอบที่พบ DIF		6.25		11.25	

หมายเหตุ DIF หมายถึง ข้อสอบที่พบว่าทำหน้าที่ต่างกัน
NO DIF หมายถึง ข้อสอบที่ไม่พบว่าทำหน้าที่ต่างกัน

จากตารางที่ 4-20 พบว่า แบบทดสอบภาษาอังกฤษ จำนวน 80 ข้อ ด้วยวิธี IRT-LR ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน จำนวน 5 ข้อ ได้แก่ข้อที่ 26, 44, 59, 68, 74 คิดเป็นร้อยละ 6.25 และวิธี SIBTEST มีข้อสอบที่ทำหน้าที่ต่างกัน จำนวน 9 ข้อ ได้แก่ข้อที่ 6, 16, 26, 32, 43, 50, 58, 59, 74 คิดเป็นร้อยละ 11.25 ซึ่งสอดคล้องกับสมมติฐานของการวิจัยข้อที่ 4

จากผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบภาษาอังกฤษ ด้วยวิธี IRT-LR และวิธี SIBTEST ดังภาพที่ 4-7



ภาพที่ 4-7 จำนวนข้อสอบที่ตรวจพบว่าทำหน้าที่ต่างกันในแบบทดสอบภาษาอังกฤษ ด้วยวิธี IRT-LR และวิธี SIBTEST

4. ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบคณิตศาสตร์ ด้วยวิธี IRT-LR และวิธี SIBTEST จำนวน 32 ข้อ ดังตารางที่ 4-21

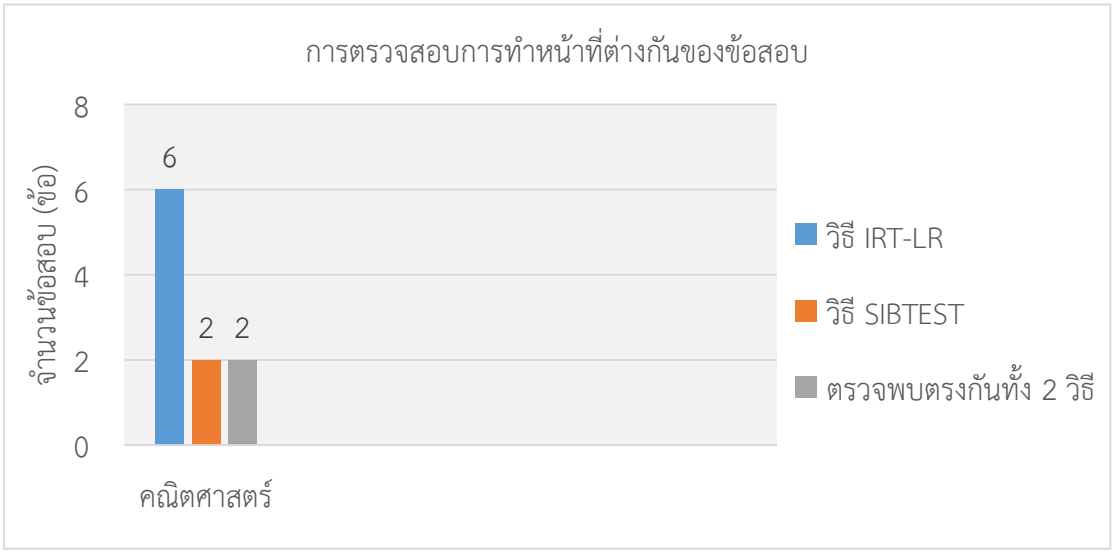
ตารางที่ 4-21 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบคณิตศาสตร์ ด้วยวิธี IRT-LR และวิธี SIBTEST จำนวน 32 ข้อ

แบบทดสอบ	จำนวน (ข้อ)	ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ			
		วิธี IRT-LR		วิธี SIBTEST	
		DIF	NO DIF	DIF	NO DIF
คณิตศาสตร์	32	6 ข้อ (6, 14, 19, 26, 28, 30)	26 ข้อ	2 ข้อ (6,19)	30 ข้อ
คิดเป็นร้อยละของข้อสอบที่พบ DIF		18.75		6.25	

หมายเหตุ DIF หมายถึง ข้อสอบที่พบว่าทำหน้าที่ต่างกัน
NO DIF หมายถึง ข้อสอบที่ไม่พบว่าทำหน้าที่ต่างกัน

จากตารางที่ 4-21 พบว่า แบบทดสอบคณิตศาสตร์ จำนวน 32 ข้อ ด้วยวิธี IRT-LR ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน จำนวน 6 ข้อ ได้แก่ข้อที่ 6, 14, 19, 26, 28, 30 คิดเป็นร้อยละ 18.75 และวิธี SIBTEST ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน จำนวน 2 ข้อ ได้แก่ข้อที่ 6,19 คิดเป็นร้อยละ 6.25 ซึ่งสอดคล้องกับสมมติฐานของการวิจัยข้อที่ 4

จากผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบคณิตศาสตร์ ด้วยวิธี IRT-LR และวิธี SIBTEST ดังภาพที่ 4-8



ภาพที่ 4-8 จำนวนข้อสอบที่ตรวจพบว่าทำหน้าที่ต่างกันในแบบทดสอบคณิตศาสตร์ ด้วยวิธี IRT-LR และวิธี SIBTEST

5. ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบวิทยาศาสตร์ ด้วยวิธี IRT-LR และวิธี SIBTEST จำนวน 80 ข้อ ดังตารางที่ 4-22

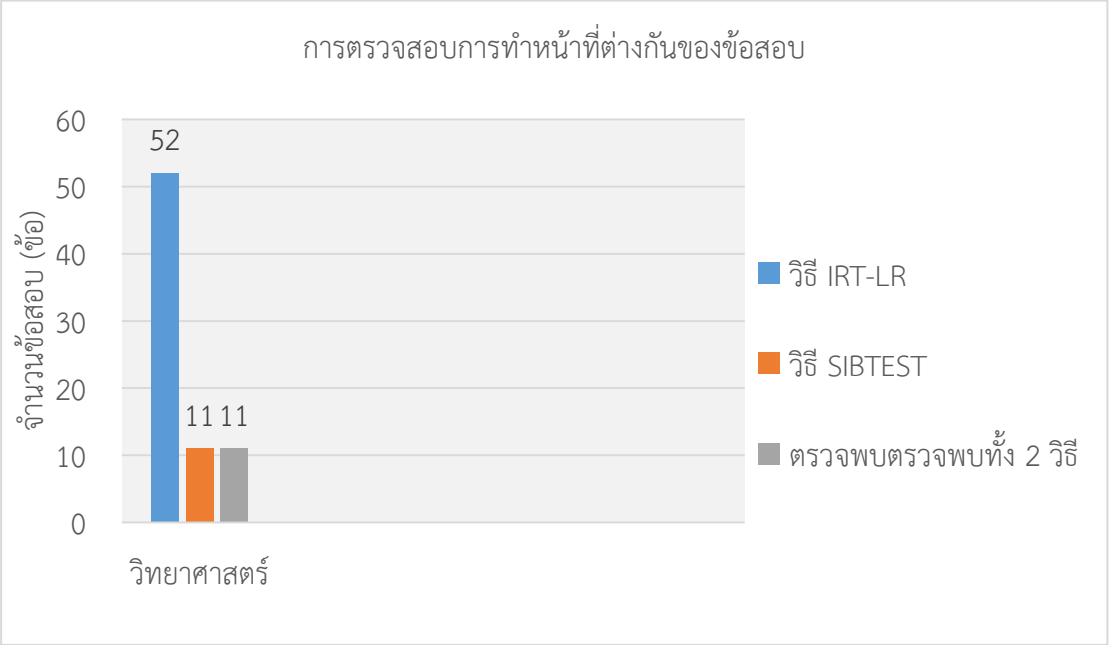
ตารางที่ 4-22 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบวิทยาศาสตร์ ด้วยวิธี IRT-LR และวิธี SIBTEST จำนวน 80 ข้อ

แบบทดสอบ	จำนวน(ข้อ)	ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ			
		วิธี IRT-LR		วิธี SIBTEST	
		DIF	NO DIF	DIF	NO DIF
วิทยาศาสตร์	80	52 ข้อ (1-3, 5, 8-14, 16, 17, 19, 21, 22, 25-27, 30, 31, 34-36, 39-48, 50-56, 59, 63, 65-67, 69-70, 72, 74, 75, 77-80)	28 ข้อ	11 ข้อ (10, 11, 16, 21, 31, 39, 46, 59, 74, 78, 80)	69 ข้อ
คิดเป็นร้อยละของข้อสอบที่พบ DIF		65.00		13.75	

หมายเหตุ DIF หมายถึง ข้อสอบที่พบว่าทำหน้าที่ต่างกัน
NO DIF หมายถึง ข้อสอบที่ไม่พบว่าทำหน้าที่ต่างกัน

จากตารางที่ 4-22 พบว่า แบบทดสอบวิทยาศาสตร์ จำนวน 80 ข้อ ด้วยวิธี IRT-LR ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน จำนวน 52 ข้อ ได้แก่ข้อที่ 1-3, 5, 8-14, 16, 17, 19, 21, 22, 25-27, 30, 31, 34-36, 39-48, 50-56, 59, 63, 65-67, 69-70, 72, 74, 75, 77-80 คิดเป็นร้อยละ 65.00 และวิธี SIBTEST ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน จำนวน 11 ข้อ ได้แก่ข้อที่ 10, 11, 16, 21, 31, 39, 46, 59, 74, 78, 80 คิดเป็นร้อยละ 13.75 ซึ่งสอดคล้องกับสมมติฐานของการวิจัยข้อที่ 4

จากผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบวิทยาศาสตร์ ด้วยวิธี IRT-LR และวิธี SIBTEST ดังภาพที่ 4-9



ภาพที่ 4-9 จำนวนข้อสอบที่ตรวจพบว่าทำหน้าที่ต่างกันในแบบทดสอบวิทยาศาสตร์ ด้วยวิธี IRT-LR และวิธี SIBTEST

6. ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบสุขศึกษาและพลศึกษา ด้วยวิธี IRT-LR และวิธี SIBTEST จำนวน 32 ข้อ ดังตารางที่ 4-23

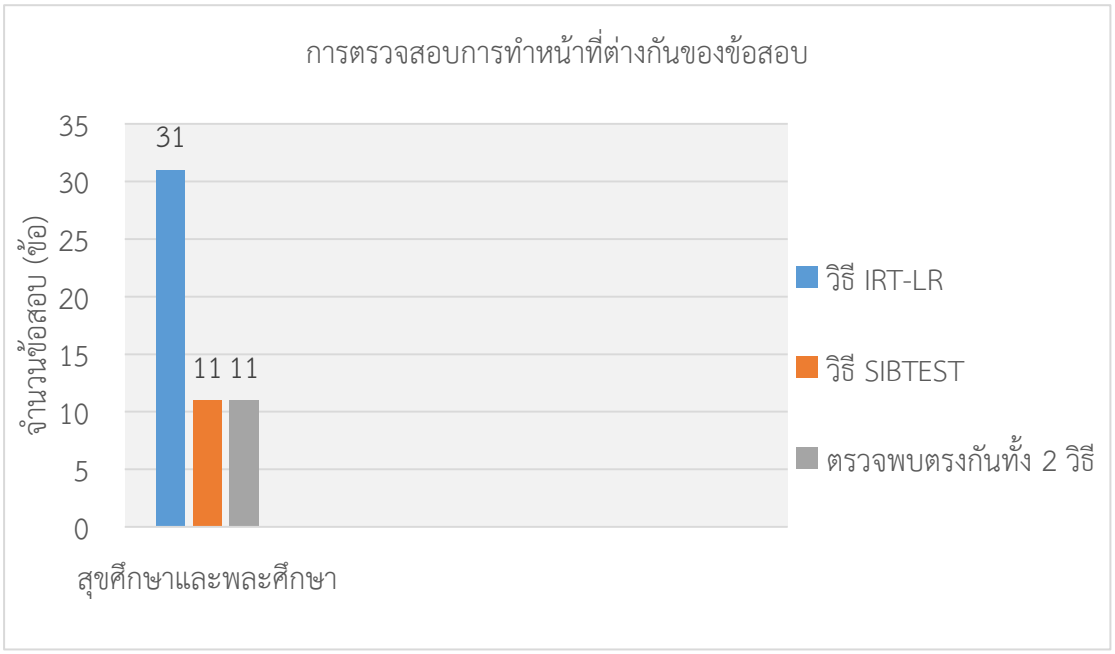
ตารางที่ 4-23 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบสุขศึกษาและพลศึกษา ด้วยวิธี IRT-LR และวิธี SIBTEST จำนวน 32 ข้อ

แบบทดสอบ	จำนวน (ข้อ)	ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ			
		วิธี IRT-LR		วิธี SIBTEST	
		DIF	NO DIF	DIF	NO DIF
สุขศึกษา และพลศึกษา	32	31 ข้อ (1-7, 9-32)	1 ข้อ	11 ข้อ (1, 3, 6, 11, 19-21, 25, 27, 29)	21 ข้อ
คิดเป็นร้อยละของข้อสอบที่พบ DIF		96.88		34.36	

หมายเหตุ DIF หมายถึง ข้อสอบที่พบว่าทำหน้าที่ต่างกัน
NO DIF หมายถึง ข้อสอบที่ไม่พบว่าทำหน้าที่ต่างกัน

จากตารางที่ 4-23 พบว่า แบบทดสอบสุขศึกษา และพละศึกษา จำนวน 32 ข้อ ด้วยวิธี IRT-LR ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน จำนวน 31 ข้อ ได้แก่ข้อที่ 1-7, 9-32 คิดเป็นร้อยละ 96.88 และวิธี SIBTEST ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน จำนวน 11 ข้อ ได้แก่ข้อที่ 1, 3, 6, 11, 19-21, 25, 27, 29 คิดเป็นร้อยละ 34.36 ซึ่งสอดคล้องกับสมมติฐานของการวิจัยข้อที่ 4

จากผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบสุขศึกษา และพละศึกษา ด้วยวิธี IRT-LR และวิธี SIBTEST สรุปได้ดังภาพที่ 4-10



ภาพที่ 4-10 จำนวนข้อสอบที่ตรวจพบว่าทำหน้าที่ต่างกันในแบบทดสอบสุขศึกษา และพละศึกษา ด้วยวิธี IRT-LR และวิธี SIBTEST

7. ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบศิลปะ ด้วยวิธี IRT-LR และวิธี SIBTEST จำนวน 30 ข้อ ดังตารางที่ 4-24

ตารางที่ 4-24 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบศิลปะ ด้วยวิธี IRT-LR และวิธี SIBTEST จำนวน 30 ข้อ

แบบทดสอบ	จำนวน (ข้อ)	ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ			
		วิธี IRT-LR		วิธี SIBTEST	
		DIF	NO DIF	DIF	NO DIF
ศิลปะ	30	26 ข้อ (1, 3-24, 27-28, 30)	4 ข้อ	6 ข้อ (1, 8, 14, 17, 27, 30)	24 ข้อ
คิดเป็นร้อยละของข้อสอบที่พบ DIF		86.67		20.00	

หมายเหตุ DIF หมายถึง ข้อสอบที่พบว่าทำหน้าที่ต่างกัน
NO DIF หมายถึง ข้อสอบที่ไม่พบว่าทำหน้าที่ต่างกัน

จากตารางที่ 4-24 พบว่า แบบทดสอบศิลปะ จำนวน 30 ข้อ ด้วยวิธี IRT-LR ตรวจพบข้อสอบที่ทำหน้าที่ต่างกันอย่าง จำนวน 26 ข้อ ได้แก่ข้อที่ 1, 3-24, 27-28, 30 คิดเป็นร้อยละ 86.67 และวิธี SIBTEST มีข้อสอบที่ทำหน้าที่ต่างกัน จำนวน 6 ข้อ ได้แก่ข้อที่ 1, 8, 14, 17, 27, 30 คิดเป็นร้อยละ 20.00 ซึ่งสอดคล้องกับสมมติฐานของการวิจัยข้อที่ 4

จากผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบศิลปะ ด้วยวิธี IRT-LR และวิธี SIBTEST ดังภาพที่ 4-11

ภาพที่ 4-11 จำนวนข้อสอบที่ตรวจพบว่าทำหน้าที่ต่างกันในแบบทดสอบศิลปะ ด้วยวิธี IRT-LR และวิธี SIBTEST

8. ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบการงานอาชีพและเทคโนโลยี ด้วยวิธี IRT-LR และวิธี SIBTEST จำนวน 41 ข้อ ดังตารางที่ 4-25

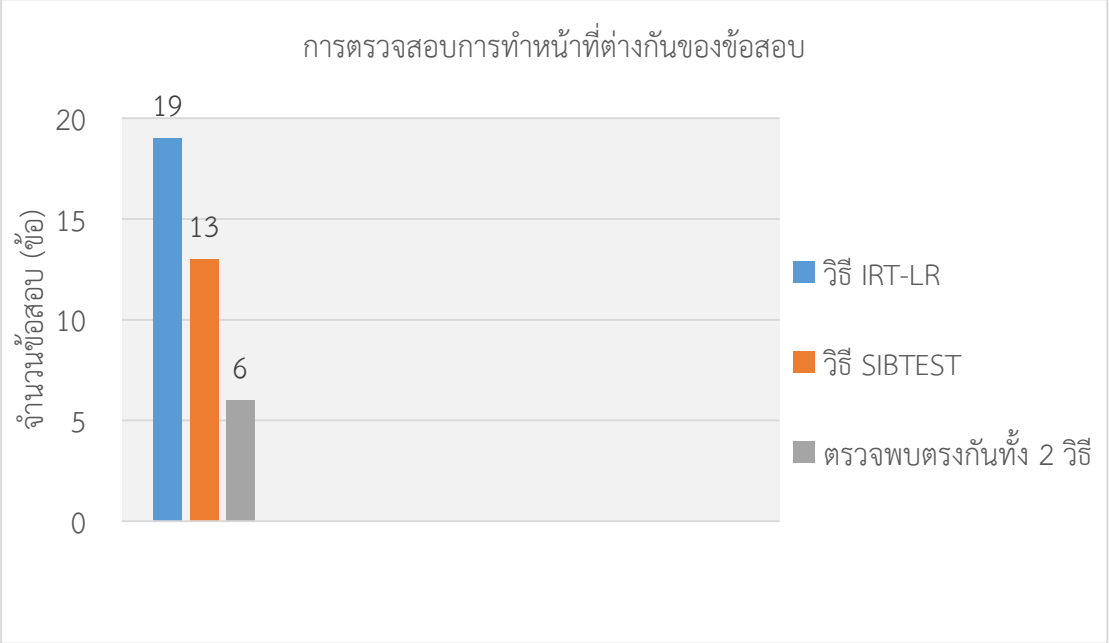
ตารางที่ 4-25 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบการงานอาชีพและเทคโนโลยี ด้วยวิธี IRT-LR และวิธี SIBTEST จำนวน 41 ข้อ

แบบทดสอบ	จำนวน (ข้อ)	ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ			
		วิธี IRT-LR		วิธี SIBTEST	
		DIF	NO DIF	DIF	NO DIF
การงานอาชีพและเทคโนโลยี	41	19 ข้อ (3, 6-7, 9-10, 12, 16, 18, 21, 25, 30-32, 34, 37-41)	22 ข้อ	13 ข้อ (4, 7, 9, 15, 20, 22, 27-28, 32-33, 39-41)	28 ข้อ
คิดเป็นร้อยละของข้อสอบที่พบ DIF		46.34		31.71	

หมายเหตุ DIF หมายถึง ข้อสอบที่พบว่าทำหน้าที่ต่างกันของข้อสอบ
NO DIF หมายถึง ข้อสอบที่ไม่พบว่าทำหน้าที่ต่างกันของข้อสอบ

จากตารางที่ 4-25 พบว่า แบบทดสอบการงานอาชีพและเทคโนโลยี จำนวน 41 ข้อ ด้วยวิธี IRT-LR ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน จำนวน 19 ข้อ ได้แก่ข้อที่ 3, 6-7, 9-10, 12, 16, 18, 21, 25, 30-32, 34, 37-41 คิดเป็นร้อยละ 46.34 และวิธี SIBTEST ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน จำนวน 13 ข้อ ได้แก่ข้อที่ 4, 7, 9, 15, 20, 22, 27-28, 32-33, 39-41 คิดเป็นร้อยละ 31.71 ซึ่งสอดคล้องกับสมมติฐานของการวิจัยข้อที่ 4

จากผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบการงานอาชีพและเทคโนโลยี ด้วยวิธี IRT-LR และวิธี SIBTEST สรุปได้ดังภาพที่ 4-12



ภาพที่ 4-12 จำนวนข้อสอบที่ตรวจพบว่าทำหน้าที่ต่างกันในแบบทดสอบการงานอาชีพและเทคโนโลยี ด้วยวิธี IRT-LR และวิธี SIBTEST

ตอนที่ 4 ผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 และ 2 ของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ระหว่างวิธี IRT-LR กับวิธี SIBTEST

การเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 และ 2 ของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในครั้งนี้เป็นการคำนวณอัตราความคลาดเคลื่อนประเภทที่ 1 และ 2 ของวิธี IRT-LR โดยผู้วิจัยใช้ วิธี SIBTEST เป็นวิธีเกณฑ์ ของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ทั้ง 8 กลุ่มสาระการเรียนรู้ ได้แก่ 1) ภาษาไทย 2) สังคมศึกษา ศาสนา และวัฒนาธรรม 3)ภาษาต่างประเทศ 4) คณิตศาสตร์ 5) วิทยาศาสตร์ 6) สุขศึกษาและพลศึกษา 7) ศิลปะ และ 8) การงานอาชีพและเทคโนโลยี ดังตารางที่ 4-26

ตารางที่ 4-26 ผลการคำนวณอัตราความคลาดเคลื่อนประเภทที่ 1 และ 2 ของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ทั้ง 8 กลุ่มสาระการเรียนรู้

แบบทดสอบ	จำนวนข้อสอบ (ข้อ)	อัตราความคลาดเคลื่อนประเภทที่ 1	อัตราความคลาดเคลื่อนประเภทที่ 2
1. ภาษาไทย	70	0.803	0.071
2. สังคมศึกษา ศาสนา และวัฒนธรรม	65	0.884	0.000
3. ภาษาอังกฤษ	80	0.028	0.667
4. คณิตศาสตร์	32	0.133	0.000
5. วิทยาศาสตร์	80	0.637	0.000
6. สุขศึกษา และพลศึกษา	32	0.952	0.000
7. ศิลปะ	30	0.833	0.000
8. การงานอาชีพ และเทคโนโลยี	41	0.464	0.538

จากตารางที่ 4-26 ผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 และ 2 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธี IRT-LR และวิธี SIBTEST โดยใช้วิธี SIBTEST เป็นเกณฑ์ พบว่า อัตราความคลาดเคลื่อนประเภทที่ 1 มี 2 กลุ่มสาระการเรียนรู้ที่มีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำ คือ แบบทดสอบภาษาอังกฤษ มีอัตราความคลาดเคลื่อนประเภทที่ 1 เท่ากับ0.028 และแบบทดสอบคณิตศาสตร์ มีอัตราความคลาดเคลื่อนประเภทที่ 1 เท่ากับ 0.133 ส่วนอัตราความคลาดเคลื่อนประเภทที่ 2 มี 2 กลุ่มสาระการเรียนรู้ ที่มีอัตราความคลาดเคลื่อนประเภทที่ 2 สูง คือ แบบทดสอบภาษาอังกฤษ มีอัตราความคลาดเคลื่อนประเภทที่ 2 เท่ากับ 0.667 และแบบทดสอบการงานอาชีพ และเทคโนโลยี มีอัตราความคลาดเคลื่อนประเภทที่ 2 เท่ากับ 0.538

ตารางที่ 4-27 ผลการเปรียบเทียบการทดสอบความแตกต่างระหว่างวีธี IRT-LR กับวิธี SIBTEST

แบบทดสอบ	จำนวนข้อสอบ (ข้อ)	จำนวนข้อที่พบ DIF (ข้อ)			x^2	*p*-value	แปลความหมาย
		วีธี IRT-LR	วิธี SIBTEST	พบ DIF ตรงกัน			
1. ภาษาไทย	70	58	14	13	0.729	0.393	ไม่แตกต่าง
2. สังคมศึกษา ศาสนาและวัฒนธรรม	65	59	13	13	1.653	0.199	ไม่แตกต่าง
3. ภาษาอังกฤษ	80	5	9	3	12.695	0.000	แตกต่าง
4. คณิตศาสตร์	32	6	2	2	9.244	0.002	แตกต่าง
5. วิทยาศาสตร์	80	52	11	11	5.797	0.016	แตกต่าง
6. สุขศึกษาและพลศึกษา	32	31	11	11	0.541	0.462	ไม่แตกต่าง
7. ศิลปะ	30	26	6	6	1.154	0.283	ไม่แตกต่าง
8. การงานอาชีพและเทคโนโลยี	41	19	13	6	0.000	0.987	ไม่แตกต่าง

จากตารางที่ 4-27 ผลการเปรียบเทียบการทดสอบความแตกต่างระหว่างวีธี IRT-LR กับวิธี SIBTEST พบว่า แบบทดสอบที่สามารถตรวจพบ DIF ทั้ง 2 วิธี ได้ไม่แตกต่างกัน มี 5 กลุ่มสาระการเรียนรู้ ได้แก่ ภาษาไทย สังคมศึกษา ศาสนาและวัฒนธรรม สุขศึกษาและพละศึกษา ศิลปะ และการงานอาชีพและเทคโนโลยี ซึ่งมีความสอดคล้องกับสมมติฐานของการวิจัยข้อที่ 5 แบบทดสอบที่สามารถตรวจพบ DIF ทั้ง 2 วิธี ได้แตกต่างกัน มี 3 กลุ่มสาระการเรียนรู้ ได้แก่ ภาษาอังกฤษ คณิตศาสตร์ และวิทยาศาสตร์ ซึ่งไม่สอดคล้องกับสมมติฐานของการวิจัยข้อที่ 5 ดังนั้นผลการทดลองมีความสอดคล้องกับสมมติฐานของการวิจัยข้อที่ 5 บางส่วน

บทที่ 5
สรุปและอภิปรายผล

การวิจัยในครั้งนี้มีวัตถุประสงค์เพื่อวิเคราะห์คุณภาพของข้อสอบ ตรวจสอบความเที่ยง ความตรงเชิงโครงสร้าง ฉบับก่อนและฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกัน ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบและเปรียบเทียบอัตราความคลาดเคลื่อน ของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 จำนวน 8 กลุ่มสาระการเรียนรู้ ได้แก่ 1) ภาษาไทย 2) สังคมศึกษา ศาสนา และวัฒนธรรม 3) ภาษาอังกฤษ 4) คณิตศาสตร์ 5) วิทยาศาสตร์ 6) สุขศึกษา และพลศึกษา 7) ศิลปะ และ 8) การงานอาชีพ และเทคโนโลยี ระหว่างวิธี IRT-LR กับวิธี SIBTEST ซึ่งเป็นข้อมูลแบบทุติยภูมิ (Secondary Data) ที่ได้มาจากสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) มีวิธีการดำเนินการวิจัย เป็น 4 ระยะ ดังนี้ ระยะที่ 1 การวิเคราะห์คุณภาพของข้อสอบ (a, b, c) และค่าความเที่ยงของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 จำนวน 8 กลุ่มสาระการเรียนรู้ ฉบับก่อนและฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกัน ตามหลักทฤษฎีการตอบสนองข้อสอบ แบบ 3 พารามิเตอร์ ระยะที่ 2 การตรวจสอบความตรงเชิงโครงสร้างของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 จำนวน 8 กลุ่มสาระการเรียนรู้ ฉบับก่อนและฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกัน ระยะที่ 3 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ระหว่างวิธี IRT-LR กับวิธี SIBTEST ฉบับก่อนตัดข้อสอบที่ทำหน้าที่ต่างกัน และระยะที่ 4 การเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 และ 2 ของการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ระหว่างวิธี IRT-LR กับวิธี SIBTEST

สรุปผลการวิจัย

1. ผลการวิเคราะห์คุณภาพของข้อสอบ (a, b, c) และค่าความเที่ยงของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ฉบับก่อนตัดกับฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกัน ตามหลักทฤษฎีการตอบสนองข้อสอบ แบบ 3 พารามิเตอร์ เป็นการวิเคราะห์ค่าพารามิเตอร์ของข้อสอบตามหลักการของทฤษฎีการตอบสนองข้อสอบซึ่งประกอบด้วยค่าอำนาจจำแนกของข้อสอบ ค่าความยากของข้อสอบ และค่าการเดาของข้อสอบ โดยใช้โปรแกรม Xcaribre Version 4.2.2 สำหรับประมาณค่าพารามิเตอร์ของข้อสอบซึ่งมีเกณฑ์การคัดเลือกข้อสอบ คือค่าอำนาจจำแนกของข้อสอบมีค่าตั้งแต่ 0.50 ถึง 2.50 ค่าความยากของข้อสอบมีค่าตั้งแต่ -2.50 ถึง 2.50 และค่าการเดาของข้อสอบมีค่าไม่เกิน 0.30 (Urry, 1977) ฉบับก่อนตัดกับฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกันออกจากแบบทดสอบ พบว่า

1.1 แบบทดสอบภาษาไทย ฉบับก่อนตัดข้อสอบที่ทำหน้าที่ต่างกัน มีค่าอำนาจจำแนกของข้อสอบ (a) เฉลี่ยเท่ากับ 0.684 ค่าความยากของข้อสอบ (b) เฉลี่ยเท่ากับ 0.671 และค่าโอกาสการเดาของข้อสอบ (c) เฉลี่ยเท่ากับ 0.174 แสดงให้เห็นว่า แบบทดสอบภาษาไทยมีค่าอำนาจจำแนกของข้อสอบ (a) ทั้งฉบับอยู่ในระดับค่อนข้างดีมีค่าความยากของข้อสอบ (b) ทั้งฉบับอยู่ในระดับค่อนข้างยากและมีค่าโอกาสการเดาของข้อสอบ (c) ทั้งฉบับไม่เกิน 0.3 ส่วนแบบทดสอบภาษาไทย

ฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกัน มีค่าอำนาจจำแนกของข้อสอบ (a) เฉลี่ยเท่ากับ 0.767 ค่าความยากของข้อสอบ (b) เฉลี่ยเท่ากับ 0.989 และค่าโอกาสการเดาของข้อสอบ (c) เฉลี่ยเท่ากับ 0.234 แสดงให้เห็นว่า แบบทดสอบภาษาไทยมีค่าอำนาจจำแนกของข้อสอบ (a) ทั้งฉบับอยู่ในระดับค่อนข้างดี มีค่าความยากของข้อสอบ (b) ทั้งฉบับอยู่ในระดับค่อนข้างยาก และมีค่าโอกาสการเดาของข้อสอบ (c) ทั้งฉบับไม่เกิน 0.3 ซึ่งเป็นไปตามสมมติฐานของการวิจัยข้อที่ 1

1.2 แบบทดสอบสังคมศึกษา ศาสนา และวัฒนธรรม ฉบับก่อนตัดข้อสอบที่ทำหน้าที่ต่างกัน มีค่าอำนาจจำแนกของข้อสอบ (a) เฉลี่ยเท่ากับ 0.784 ค่าความยากของข้อสอบ (b) เฉลี่ยเท่ากับ 1.817 และค่าโอกาสการเดาของข้อสอบ (c) เฉลี่ยเท่ากับ 0.173 แสดงให้เห็นว่า แบบทดสอบสังคมศึกษา ศาสนาและวัฒนธรรม มีค่าอำนาจจำแนกของข้อสอบ (a) ทั้งฉบับอยู่ในระดับค่อนข้างดี มีค่าความยากของข้อสอบ (b) ทั้งฉบับอยู่ในระดับยาก และมีค่าโอกาสการเดาของข้อสอบ (c) ทั้งฉบับไม่เกิน 0.3 ส่วนแบบทดสอบสังคมศึกษา ศาสนาและวัฒนธรรม ฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกัน มีค่าอำนาจจำแนกของข้อสอบ (a) เฉลี่ยเท่ากับ 0.818 ค่าความยากของข้อสอบ (b) เฉลี่ยเท่ากับ 3.014 และค่าโอกาสการเดาของข้อสอบ (c) เฉลี่ยเท่ากับ 0.229 แสดงให้เห็นว่า แบบทดสอบสังคมศึกษา ศาสนาและวัฒนธรรม มีค่าอำนาจจำแนกของข้อสอบ (a) ทั้งฉบับอยู่ในระดับค่อนข้างดี มีค่าความยากของข้อสอบ (b) ทั้งฉบับอยู่ในระดับยากมาก และมีค่าโอกาสการเดาของข้อสอบ (c) ทั้งฉบับไม่เกิน 0.3 ซึ่งเป็นไปตามสมมติฐานของการวิจัยข้อที่ 1

1.3 แบบทดสอบภาษาอังกฤษ ฉบับก่อนตัดข้อสอบที่ทำหน้าที่ต่างกัน มีค่าอำนาจจำแนกของข้อสอบ (a) เฉลี่ยเท่ากับ 1.521 ค่าความยากของข้อสอบ (b) เฉลี่ยเท่ากับ 1.530 และค่าโอกาสการเดาของข้อสอบ (c) เฉลี่ยเท่ากับ 0.199 แสดงให้เห็นว่า แบบทดสอบภาษาอังกฤษ มีค่าอำนาจจำแนกของข้อสอบ (a) ทั้งฉบับอยู่ในระดับค่อนข้างดี มีค่าความยากของข้อสอบ (b) ทั้งฉบับอยู่ในระดับยาก และมีค่าโอกาสการเดาของข้อสอบ (c) อยู่ในระดับทั้งฉบับไม่เกิน 0.3 ส่วนแบบทดสอบภาษาอังกฤษ ฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกัน มีค่าอำนาจจำแนกของข้อสอบ (a) เฉลี่ยเท่ากับ 1.280 ค่าความยากของข้อสอบ (b) เฉลี่ยเท่ากับ 2.058 และค่าโอกาสการเดาของข้อสอบ (c) เฉลี่ยเท่ากับ 0.207 แสดงให้เห็นว่า แบบทดสอบภาษาอังกฤษ มีค่าอำนาจจำแนกของข้อสอบ (a) ทั้งฉบับอยู่ในระดับค่อนข้างดี มีค่าความยากของข้อสอบ (b) ทั้งฉบับอยู่ในระดับยากมาก และมีค่าโอกาสการเดาของข้อสอบ (c) ทั้งฉบับไม่เกิน 0.3 ซึ่งเป็นไปตามสมมติฐานของการวิจัยข้อที่ 1

1.4 แบบทดสอบคณิตศาสตร์ ฉบับก่อนตัดข้อสอบที่ทำหน้าที่ต่างกัน มีค่าอำนาจจำแนกของข้อสอบ (a) เฉลี่ยเท่ากับ 1.245 ค่าความยากของข้อสอบ (b) เฉลี่ยเท่ากับ 1.442 และค่าโอกาสการเดาของข้อสอบ (c) เฉลี่ยเท่ากับ 0.170 แสดงให้เห็นว่า แบบทดสอบคณิตศาสตร์ มีค่าอำนาจจำแนกของข้อสอบ (a) ทั้งฉบับอยู่ในระดับค่อนข้างดี มีค่าความยากของข้อสอบ (b) ทั้งฉบับอยู่ในระดับยาก และมีค่าโอกาสการเดาของข้อสอบ (c) ทั้งฉบับไม่เกิน 0.3 ส่วนแบบทดสอบคณิตศาสตร์ ฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกัน มีค่าอำนาจจำแนกของข้อสอบ (a) เฉลี่ยเท่ากับ 1.101 ค่าความยากของข้อสอบ (b) เฉลี่ยเท่ากับ 2.692 และค่าโอกาสการเดาของข้อสอบ (c) เฉลี่ยเท่ากับ 0.193 แสดงให้เห็นว่า แบบทดสอบคณิตศาสตร์ มีค่าอำนาจจำแนกของข้อสอบ (a)

ทั้งฉบับอยู่ในระดับค่อนข้างดี มีค่าความยากของข้อสอบ (b) ทั้งฉบับอยู่ในระดับยากมาก และมีค่าโอกาสการเดาของข้อสอบ (c) ทั้งฉบับไม่เกิน 0.3

1.5 แบบทดสอบวิทยาศาสตร์ ฉบับก่อนตัดข้อสอบที่ทำหน้าที่ต่างกัน มีค่าอำนาจจำแนกของข้อสอบ (a) เฉลี่ยเท่ากับ 0.968 ค่าความยากของข้อสอบ (b) เฉลี่ยเท่ากับ 1.696 และค่าโอกาสการเดาของข้อสอบ (c) เฉลี่ยเท่ากับ 0.189 แสดงให้เห็นว่า แบบทดสอบวิทยาศาสตร์ มีค่าอำนาจจำแนกของข้อสอบ (a) ทั้งฉบับอยู่ในระดับค่อนข้างดี มีค่าความยากของข้อสอบ (b) ทั้งฉบับอยู่ในระดับยาก และมีค่าโอกาสการเดาของข้อสอบ (c) ทั้งฉบับไม่เกิน 0.3 ส่วนแบบทดสอบวิทยาศาสตร์ ฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกัน มีค่าอำนาจจำแนกของข้อสอบ (a) เฉลี่ยเท่ากับ 0.876 ค่าความยากของข้อสอบ (b) เฉลี่ยเท่ากับ 2.666 และค่าโอกาสการเดาของข้อสอบ (c) เฉลี่ยเท่ากับ 0.221 แสดงให้เห็นว่า แบบทดสอบวิทยาศาสตร์ มีค่าอำนาจจำแนกของข้อสอบ (a) ทั้งฉบับอยู่ในระดับค่อนข้างดี มีค่าความยากของข้อสอบ (b) ทั้งฉบับอยู่ในระดับยากมาก และมีค่าโอกาสการเดาของข้อสอบ (c) ทั้งฉบับไม่เกิน 0.3 ซึ่งเป็นไปตามสมมติฐานของการวิจัยข้อที่ 1

1.6 แบบทดสอบสุขศึกษา และพลศึกษา ฉบับก่อนตัดข้อสอบที่ทำหน้าที่ต่างกัน มีค่าอำนาจจำแนกของข้อสอบ (a) เฉลี่ยเท่ากับ 0.590 ค่าความยากของข้อสอบ (b) เฉลี่ยเท่ากับ -0.223 และค่าโอกาสการเดาของข้อสอบ (c) เฉลี่ยเท่ากับ 0.221 แสดงให้เห็นว่า แบบทดสอบสุขศึกษา และพลศึกษา มีค่าอำนาจจำแนกของข้อสอบ (a) ทั้งฉบับอยู่ในระดับค่อนข้างดี มีค่าความยากของข้อสอบ (b) ทั้งฉบับอยู่ในระดับปานกลางและมีค่าโอกาสการเดาของข้อสอบ (c) ทั้งฉบับ ไม่เกิน 0.3 ส่วนแบบทดสอบสุขศึกษา และพลศึกษา ฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกัน มีค่าอำนาจจำแนกของข้อสอบ (a) เฉลี่ยเท่ากับ 0.532 ค่าความยากของข้อสอบ (b) เฉลี่ยเท่ากับ 0.080 และค่าโอกาสการเดาของข้อสอบ (c) เฉลี่ยเท่ากับ 0.266 แสดงให้เห็นว่า แบบทดสอบสุขศึกษา และพลศึกษา มีค่าอำนาจจำแนกของข้อสอบ (a) ทั้งฉบับอยู่ในระดับค่อนข้างดี มีค่าความยากของข้อสอบ (b) ทั้งฉบับอยู่ในระดับปานกลางและมีค่าโอกาสการเดาของข้อสอบ (c) ทั้งฉบับ ไม่เกิน 0.3 ซึ่งเป็นไปตามสมมติฐานของการวิจัยข้อที่ 1

1.7 แบบทดสอบศิลปะ ฉบับก่อนตัดข้อสอบที่ทำหน้าที่ต่างกัน มีค่าอำนาจจำแนกของข้อสอบ (a) เฉลี่ยเท่ากับ 1.140 ค่าความยากของข้อสอบ (b) เฉลี่ยเท่ากับ 2.064 และค่าโอกาสการเดาของข้อสอบ (c) เฉลี่ยเท่ากับ 0.198 แสดงให้เห็นว่า แบบทดสอบศิลปะ มีค่าอำนาจจำแนกของข้อสอบ (a) ทั้งฉบับอยู่ในระดับค่อนข้างดี มีค่าความยากของข้อสอบ (b) ทั้งฉบับอยู่ในระดับยากมากและมีค่าโอกาสการเดาของข้อสอบ (c) ทั้งฉบับไม่เกิน 0.3 ส่วนแบบทดสอบศิลปะ ฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกัน มีค่าอำนาจจำแนกของข้อสอบ (a) เฉลี่ยเท่ากับ 0.938 ค่าความยากของข้อสอบ (b) เฉลี่ยเท่ากับ 3.368 และค่าโอกาสการเดาของข้อสอบ (c) เฉลี่ยเท่ากับ 0.228 แสดงให้เห็นว่า แบบทดสอบศิลปะ มีค่าอำนาจจำแนกของข้อสอบ (a) ทั้งฉบับอยู่ในระดับค่อนข้างดี มีค่าความยากของข้อสอบ (b) ทั้งฉบับอยู่ในระดับยากมากและมีค่าโอกาสการเดาของข้อสอบ (c) ทั้งฉบับไม่เกิน 0.3 ซึ่งเป็นไปตามสมมติฐานของการวิจัยข้อที่ 1

1.8 แบบทดสอบการงานอาชีพ และเทคโนโลยี ฉบับก่อนตัดข้อสอบที่ทำหน้าที่ต่างกัน มีค่าอำนาจจำแนกของข้อสอบ (a) เฉลี่ยเท่ากับ 0.776 ค่าความยากของข้อสอบ (b) เฉลี่ยเท่ากับ 0.906 และค่าโอกาสการเดาของข้อสอบ (c) เฉลี่ยเท่ากับ 0.197 แสดงให้เห็นว่า แบบทดสอบการงาน

อาชีพ และเทคโนโลยี มีค่าอำนาจจำแนกของข้อสอบ (a) ทั้งฉบับอยู่ในระดับค่อนข้างดี มีค่าความยากของข้อสอบ (b) ทั้งฉบับอยู่ในระดับค่อนข้างยาก และมีค่าโอกาสการเดาของข้อสอบ (c) ทั้งฉบับไม่เกิน 0.3 ส่วนแบบทดสอบการงานอาชีพ และเทคโนโลยี ฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกัน มีค่าอำนาจจำแนกของข้อสอบ (a) เฉลี่ยเท่ากับ 0.871 ค่าความยากของข้อสอบ (b) เฉลี่ยเท่ากับ 1.291 และค่าโอกาสการเดาของข้อสอบ (c) เฉลี่ยเท่ากับ 0.243 แสดงให้เห็นว่า แบบทดสอบการงานอาชีพ และเทคโนโลยี มีค่าอำนาจจำแนกของข้อสอบ (a) ทั้งฉบับอยู่ในระดับค่อนข้างดี มีค่าความยากของข้อสอบ (b) ทั้งฉบับอยู่ในระดับยาก และมีค่าโอกาสการเดาของข้อสอบ (c) ทั้งฉบับไม่เกิน 0.3 ซึ่งเป็นไปตามสมมติฐานของการวิจัยข้อที่ 1

ค่าความเที่ยงของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ฉบับก่อนตัดกับฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกัน โดยใช้ทฤษฎีการตอบสนองข้อสอบ (IRT) แบบ 3 พารามิเตอร์ ทั้ง 8 กลุ่มสาระการเรียนรู้ พบว่า แบบทดสอบฉบับก่อนตัดข้อสอบที่ทำหน้าที่ต่างกัน มีค่าความเที่ยง (Reliability) มากกว่าฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกัน ทั้ง 8 กลุ่มสาระการเรียนรู้ โดยแบบทดสอบภาษาอังกฤษมีค่าความเที่ยงมากที่สุด และแบบทดสอบศิลปะมีค่าความเที่ยงน้อยที่สุด ทั้งฉบับก่อนตัดและหลังตัดข้อสอบที่ทำหน้าที่ต่างกัน ซึ่งเป็นไปตามสมมติฐานของการวิจัยข้อที่ 2

2. การตรวจสอบความตรงเชิงโครงสร้างของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ฉบับก่อนและฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกัน ทั้ง 8 กลุ่มสาระการเรียนรู้ พบว่า โมเดลแบบทดสอบฉบับก่อนตัดข้อสอบที่ทำหน้าที่ต่างกันกับโมเดลแบบทดสอบฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกัน ทั้ง 8 กลุ่มสาระการเรียนรู้ มีความสอดคล้องกับข้อมูลเชิงประจักษ์แตกต่างกัน เมื่อพิจารณาจากค่าสถิติที่ได้ มี 6 กลุ่มสาระการเรียนรู้ ได้แก่ แบบทดสอบภาษาไทย แบบทดสอบสังคมศึกษา ศาสนา และวัฒนธรรม แบบทดสอบภาษาอังกฤษ แบบทดสอบวิทยาศาสตร์ แบบทดสอบสุขศึกษา และพลศึกษา และแบบทดสอบการงานอาชีพ และเทคโนโลยี มีค่าไม่มากพอที่จะยอมรับสมมติฐาน ซึ่งทำให้ผลการตรวจสอบไม่สอดคล้องกับสมมติฐานข้อที่ 3 และมี 2 กลุ่มสาระการเรียนรู้ ได้แก่ แบบทดสอบคณิตศาสตร์และศิลปะ มีค่ามากพอที่จะยอมรับสมมติฐาน ซึ่งทำให้ผลการตรวจสอบสอดคล้องกับสมมติฐานข้อที่ 3 จึงสรุปผลได้ว่า เนื่องจากข้อมูลที่ใช้ในการวิเคราะห์เป็นผลการตอบแบบทดสอบ O-NET ทั้ง 8 กลุ่มสาระการเรียนรู้ ชั้นมัธยมศึกษาปีที่ 6 เป็นข้อมูลทุติยภูมิ (Secondary Data) โดยขอความร่วมมือจาก สถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) หรือ สทศ. มีความแตกต่างกันด้วยจำนวนข้อคำถาม จึงทำให้ผลการตรวจสอบดัชนีความตรงเชิงโครงสร้างโมเดลแบบทดสอบฉบับก่อนและฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกันออกจากแบบทดสอบ 8 กลุ่มสาระการเรียนรู้ มีความสอดคล้องกับสมมติฐานข้อที่ 3 บางส่วน

3. การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ทั้ง 8 กลุ่มสาระการเรียนรู้ ระหว่างวิธี IRT-LR กับวิธี SIBTEST พบว่า

3.1 แบบทดสอบภาษาไทย จากข้อสอบทั้งหมด 70 ข้อ วิธี IRT-LR พบจำนวนข้อสอบที่ทำหน้าที่ต่างกันจำนวน 58 ข้อ คิดเป็นร้อยละ 85.71 ของจำนวนข้อสอบทั้งหมด วิธี SIBTEST พบจำนวนข้อสอบที่ทำหน้าที่ต่างกันจำนวน 14 ข้อ คิดเป็นร้อยละ 20.00 ของจำนวนข้อสอบทั้งหมด

จำนวนข้อสอบรายข้อที่ถูกตัดออกเมื่อพบการทำหน้าที่ต่างกันทั้ง 2 วิธี จำนวน 13 ข้อ คิดเป็นร้อยละ 18.57 ข้อจำนวนข้อสอบทั้งหมด

3.2 แบบทดสอบสังคมศึกษา ศาสนาและวัฒนธรรม จากข้อสอบทั้งหมด 65 ข้อ วิธี IRT-LR พบจำนวนข้อสอบที่ทำหน้าที่ต่างกันจำนวน 59 ข้อ คิดเป็นร้อยละ 90.77 ของจำนวนข้อสอบทั้งหมด วิธี SIBTEST พบจำนวนข้อสอบที่ทำหน้าที่ต่างกันจำนวน 13 ข้อ คิดเป็นร้อยละ 20.00 ของจำนวนข้อสอบทั้งหมด จำนวนข้อสอบรายข้อที่ถูกตัดออกเมื่อพบการทำหน้าที่ต่างกันทั้ง 2 วิธี จำนวน 13 ข้อ คิดเป็นร้อยละ 20.00 ข้อจำนวนข้อสอบทั้งหมด

3.3 แบบทดสอบภาษาอังกฤษ จากข้อสอบทั้งหมด 80 ข้อ วิธี IRT-LR พบจำนวนข้อสอบที่ทำหน้าที่ต่างกันจำนวน 5 ข้อ คิดเป็นร้อยละ 6.25 ของจำนวนข้อสอบทั้งหมด วิธี SIBTEST พบจำนวนข้อสอบที่ทำหน้าที่ต่างกันจำนวน 9 ข้อ คิดเป็นร้อยละ 11.25 ของจำนวนข้อสอบทั้งหมด จำนวนข้อสอบรายข้อที่ถูกตัดออกเมื่อพบการทำหน้าที่ต่างกันทั้ง 2 วิธี จำนวน 3 ข้อ คิดเป็นร้อยละ 3.75 ข้อจำนวนข้อสอบทั้งหมด

3.4 แบบทดสอบคณิตศาสตร์ จากข้อสอบทั้งหมด 32 ข้อ วิธี IRT-LR พบจำนวนข้อสอบที่ทำหน้าที่ต่างกันจำนวน 6 ข้อ คิดเป็นร้อยละ 8.75 ของจำนวนข้อสอบทั้งหมด วิธี SIBTEST พบจำนวนข้อสอบที่ทำหน้าที่ต่างกันจำนวน 2 ข้อ คิดเป็นร้อยละ 6.25 ของจำนวนข้อสอบทั้งหมด จำนวนข้อสอบรายข้อที่ถูกตัดออกเมื่อพบการทำหน้าที่ต่างกันทั้ง 2 วิธี จำนวน 2 ข้อ คิดเป็นร้อยละ 6.25 ข้อจำนวนข้อสอบทั้งหมด

3.5 แบบทดสอบวิทยาศาสตร์ จากข้อสอบทั้งหมด 80 ข้อ วิธี IRT-LR พบจำนวนข้อสอบที่ทำหน้าที่ต่างกันจำนวน 52 ข้อ คิดเป็นร้อยละ 65.00 ของจำนวนข้อสอบทั้งหมด วิธี SIBTEST พบจำนวนข้อสอบที่ทำหน้าที่ต่างกันจำนวน 11 ข้อ คิดเป็นร้อยละ 13.75 ของจำนวนข้อสอบทั้งหมด จำนวนข้อสอบรายข้อที่ถูกตัดออกเมื่อพบการทำหน้าที่ต่างกันทั้ง 2 วิธี จำนวน 11 ข้อ คิดเป็นร้อยละ 13.75 ข้อจำนวนข้อสอบทั้งหมด

3.6 แบบทดสอบสุขศึกษา และพลศึกษา จากข้อสอบทั้งหมด 32 ข้อ วิธี IRT-LR พบจำนวนข้อสอบที่ทำหน้าที่ต่างกันจำนวน 31 ข้อ คิดเป็นร้อยละ 96.88 ของจำนวนข้อสอบทั้งหมด วิธี SIBTEST พบจำนวนข้อสอบที่ทำหน้าที่ต่างกันจำนวน 11 ข้อ คิดเป็นร้อยละ 34.36 ของจำนวนข้อสอบทั้งหมด จำนวนข้อสอบรายข้อที่ถูกตัดออกเมื่อพบการทำหน้าที่ต่างกันทั้ง 2 วิธี จำนวน 11 ข้อ คิดเป็นร้อยละ 34.36 ข้อจำนวนข้อสอบทั้งหมด

3.7 แบบทดสอบศิลปะ จากข้อสอบทั้งหมด 30 ข้อ วิธี IRT-LR พบจำนวนข้อสอบที่ทำหน้าที่ต่างกันจำนวน 26 ข้อ คิดเป็นร้อยละ 86.67 ของจำนวนข้อสอบทั้งหมด วิธี SIBTEST พบจำนวนข้อสอบที่ทำหน้าที่ต่างกันจำนวน 6 ข้อ คิดเป็นร้อยละ 20.00 ของจำนวนข้อสอบทั้งหมด จำนวนข้อสอบรายข้อที่ถูกตัดออกเมื่อพบการทำหน้าที่ต่างกันทั้ง 2 วิธี จำนวน 6 ข้อ คิดเป็นร้อยละ 20.00 ข้อจำนวนข้อสอบทั้งหมด

3.8 แบบทดสอบการงานอาชีพและเทคโนโลยี จากข้อสอบทั้งหมด 41 ข้อ วิธี IRT-LR พบจำนวนข้อสอบที่ทำหน้าที่ต่างกันจำนวน 11 ข้อ คิดเป็นร้อยละ 46.34 ของจำนวนข้อสอบทั้งหมด วิธี SIBTEST พบจำนวนข้อสอบที่ทำหน้าที่ต่างกันจำนวน 13 ข้อ คิดเป็นร้อยละ 31.71 ของจำนวน

ข้อสอบทั้งหมด จำนวนข้อสอบรายข้อที่ถูกตัดออกเมื่อพบการทำหน้าที่ต่างกันทั้ง 2 วิธี จำนวน 6 ข้อ คิดเป็นร้อยละ 14.63 ข้อจำนวนข้อสอบทั้งหมด

จากผลการทดลองแสดงให้เห็นว่า แบบทดสอบทั้ง 8 กลุ่มสาระการเรียนรู้ ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน แตกต่างกัน เป็นไปตามสมมติฐานของการวิจัยข้อที่ 4

4. การเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 และ 2 ของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 จำนวน 8 กลุ่มสาระการเรียนรู้ ระหว่างวิธี IRT-LR กับวิธี SIBTEST จากการคำนวณอัตราความคลาดเคลื่อนประเภทที่ 1 และอัตราความคลาดเคลื่อนประเภทที่ 2 พบว่า แบบทดสอบที่ตรวจพบ DIF ทั้ง 2 วิธี มีอัตราความคลาดเคลื่อนประเภทที่ 1 และ 2 ไม่แตกต่างกัน มีจำนวน 5 กลุ่มสาระการเรียนรู้ ได้แก่ ภาษาไทย สังคมศึกษา ศาสนาและวัฒนธรรม สุขศึกษาและพละศึกษา ศิลปะ และการงานอาชีพ และเทคโนโลยี ดังนั้น ผลการเปรียบเทียบทดลองมีความสอดคล้องกับสมมติฐานของการวิจัยข้อที่ 5 บางส่วน

อภิปรายผล

ผลวิเคราะห์คุณภาพของข้อสอบ ตรวจสอบความเที่ยง ความตรงเชิงโครงสร้าง ฉบับก่อนและฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกันออกจากแบบทดสอบ ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบและเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 และ 2 ของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ในแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ระหว่างวิธี IRT-LR กับวิธี SIBTEST มีประเด็นที่ควรอภิปรายดังนี้

1. ผลการวิเคราะห์คุณภาพข้อสอบ O-NET ระดับชั้นมัธยมศึกษาปีที่ 6 จำนวน 8 กลุ่มสาระการเรียนรู้ พบว่า แบบทดสอบฉบับก่อนตัดข้อสอบที่พบ DIF มีค่าความยากของข้อสอบเฉลี่ยปานกลาง เมื่อตัด ข้อสอบที่พบ DIF ออกไปแล้ว ปรากฏว่า มีค่าความยากของข้อสอบเฉลี่ยค่อนข้างยาก แสดงว่า ค่าความยากของข้อสอบเพิ่มมากขึ้น เมื่อเปรียบเทียบกับแบบทดสอบฉบับก่อนตัดข้อสอบที่พบ DIF มีความสอดคล้องกับสมมติฐานข้อที่ 1 ซึ่งสอดคล้องกับผลการวิจัยของ Finch & French (2014) ที่ศึกษาเกี่ยวกับผลกระทบของค่าโอกาสการเดาข้อสอบในค่าพารามิเตอร์ที่มีความแตกต่างกันของการทำหน้าที่ต่างกันของข้อสอบในรูปแบบเอกรูปและอเนกรูป พบว่า เมื่อค่าโอกาสของการเดาของข้อสอบมีความแตกต่าง ค่าความยากของข้อสอบและค่าอำนาจจำแนกของข้อสอบจะมีความแตกต่าง ตามหลักทฤษฎีการตอบสนองข้อสอบ แบบ 3 พารามิเตอร์ เมื่อตรวจสอบค่าความเที่ยงของแบบทดสอบ พบว่า แบบทดสอบทั้งฉบับก่อน และฉบับหลังตัดข้อสอบที่พบ DIF พบว่า มีค่าความเที่ยงของแบบทดสอบไม่แตกต่างกัน สอดคล้องกับผลการวิจัยของ ณรงค์ จันทรมหา (2554) ที่ศึกษาการเปรียบเทียบค่าความเที่ยงของแบบทดสอบผลสัมฤทธิ์ทางการเรียนทีมีจำนวนข้อสอบทำหน้าที่ต่างกันแตกต่างกัน พบว่า ค่าความเที่ยงของแบบทดสอบมีค่าไม่สูงมากนักและมีค่าไม่แตกต่างกัน มีความสอดคล้องกับสมมติฐานข้อที่ 2

2. ผลการตรวจสอบความตรงเชิงโครงสร้างของแบบทดสอบ ทั้ง 8 กลุ่มสาระการเรียนรู้ของแบบทดสอบฉบับก่อนและฉบับหลังตัดข้อสอบที่พบ DIF พบว่า มีแบบทดสอบ O-NET จำนวน 2 กลุ่มสาระการเรียนรู้ คือ สาระการเรียนรู้คณิตศาสตร์ และศิลปะ ที่มีความตรงเชิงโครงสร้างไม่แตกต่างกันทั้งฉบับก่อน และฉบับหลังตัดข้อสอบที่พบ DIF เนื่องจากข้อมูลที่ใช้ในการวิเคราะห์

เป็นผลการตอบแบบทดสอบ O-NET ทั้ง 8 กลุ่มสาระการเรียนรู้ ชั้นมัธยมศึกษาปีที่ 6 เป็นข้อมูลทุติยภูมิ (Secondary Data) เนื้อหาของข้อสอบที่ได้รับแต่ละกลุ่มสาระการเรียนรู้ผู้วิจัยไม่ได้เป็นผู้กำหนด มีความแตกต่างกันด้วยจำนวนข้อคำถาม จึงทำให้ผลการตรวจสอบดัชนีความตรงเชิงโครงสร้างโมเดลแบบทดสอบฉบับก่อนและฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกันออกจากแบบทดสอบ 8 กลุ่มสาระการเรียนรู้ มีความสอดคล้องกับสมมติฐานข้อที่ 3 บางส่วน แสดงว่าเมื่อตัดข้อสอบที่ทำหน้าที่ต่างกันออกออกจากแบบทดสอบไม่ได้ทำให้โครงสร้างของแบบทดสอบเปลี่ยนแปลงไปจากเดิม มีความสอดคล้องกับข้อมูลเชิงประจักษ์ สอดคล้องกับงานวิจัยของพิชชา สุริอาจ และประกฤติยา ทักษิโณ (2559) ได้วิเคราะห์องค์ประกอบเชิงยืนยันอันดับที่สองของแบบวัดความตระหนักต่อโลกในยุค ศตวรรษที่ 21 ของนักเรียนมัธยมศึกษาตอนต้นที่พัฒนาขึ้น ผลการวิจัยพบว่า การวิเคราะห์องค์ประกอบเชิงยืนยันอันดับที่สองหลังการปรับโครงสร้างแบบวัดหลังการตัดข้อสอบที่พบ DIF มีความสอดคล้องกับข้อมูลเชิงประจักษ์ และสอดคล้องกับงานวิจัยของ ชัยยศ ชาวระนอง และคณะ (2553) ได้ศึกษาประสิทธิภาพของโมเดลและการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบพหุมิติ โดยวิธีการวิเคราะห์องค์ประกอบเชิงยืนยัน ผลการวิจัยพบว่า ค่าไค-สแควร์ของโมเดลจะไม่แปรเปลี่ยนไปตามกลุ่มตัวอย่าง แต่แปลเปลี่ยนตามจำนวนข้อสอบต่อองค์ประกอบ

3. ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธี IRT-LR และวิธี SIBTEST ทั้ง 8 กลุ่มสาระการเรียนรู้ พบว่า วิธี IRT-LR สามารถตรวจพบข้อสอบที่ทำหน้าที่ต่างกันได้มากที่สุด ในแบบทดสอบสังคมศึกษา ศาสนาและวัฒนธรรม ซึ่งสอดคล้องกับงานวิจัยของ Paek & Han (2012) พบว่าโปรแกรม IRTPRO 2.1 มีประสิทธิภาพสูง ซึ่งทำให้การประมาณค่าพารามิเตอร์ที่ได้จากการวิเคราะห์ด้วยโปรแกรมนี้เป็นที่ยอมรับในวงกว้างของการวิจัย เมื่อเปรียบเทียบผลตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ทั้งฉบับก่อนและฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกันออกจากแบบทดสอบ พบว่า วิธี IRT-LR ตรวจพบ DIF มากกว่าวิธี SIBTEST และพบข้อสอบที่ทำหน้าที่ต่างกันตรงกัน ทั้ง 2 วิธี จำนวน 65 ข้อ คิดเป็นร้อยละ 15.12 ($p < 0.05$) เมื่อพิจารณาสัดส่วนของร้อยละของการทำหน้าที่ต่างกันของข้อสอบ พบว่า แบบทดสอบ O-NET ทั้ง 8 กลุ่มสาระการเรียนรู้ มีสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันอยู่ในเกณฑ์ปกติ สอดคล้องกับงานวิจัยของ Narayanan & Swaminathan (1994) ที่ศึกษาเกี่ยวกับประสิทธิภาพของการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบมาตรฐาน พบว่า แบบทดสอบที่มีสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน ร้อยละ 10 ถึง 15 อยู่ในระดับเกณฑ์ปกติ แต่ถ้ามีสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน มากกว่าร้อยละ 20 ถือว่าอยู่ในระดับเกณฑ์ผิดปกติ

4. ผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 และ 2 ระหว่างวิธี IRT-LR กับวิธี SIBTEST (โดยใช้วิธี SIBTEST เป็นเกณฑ์) ทั้ง 8 กลุ่มสาระการเรียนรู้ พบว่า มี 2 กลุ่มสาระการเรียนรู้ที่มีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำ คือ กลุ่มสาระการเรียนรู้ภาษาอังกฤษ และคณิตศาสตร์ โดย มี 2 กลุ่มสาระการเรียนรู้ ที่มีอัตราความคลาดเคลื่อนประเภทที่ 2 สูง คือ กลุ่มสาระการเรียนรู้ภาษาอังกฤษ และการงานอาชีพและเทคโนโลยี ซึ่งสอดคล้องกับงานวิจัยของ Demir (2016) ได้ตรวจสอบอัตราความคลาดเคลื่อนประเภทที่ 1 ของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธี IRT-LR สอดคล้องกับงานวิจัยของ Kose & Demirtasli (2012) ได้เปรียบเทียบมีอัตราความคลาดเคลื่อนและมีอัตราความถูกต้องของโมเดลแบบเอกมิติ และแบบ

พหุมิติ ตามหลักทฤษฎีการตอบสนองข้อสอบ ทั้งในแง่ของจำนวนของแบบทดสอบและขนาดของกลุ่มตัวอย่าง พบว่า ค่าพารามิเตอร์ภายใต้ทฤษฎีการตอบสนองข้อสอบ (IRT) มีอัตราความคลาดเคลื่อนน้อยลงและมีอัตราความถูกต้องเพิ่มมากขึ้น และข้อสอบที่มากขึ้นและขนาดกลุ่มตัวอย่างที่ใหญ่ขึ้นสามารถลดการประมาณค่าอัตราความคลาดเคลื่อนและมีอัตราความถูกต้องเพิ่มมาก

ข้อเสนอแนะในการนำผลการวิจัยไปใช้

จากผลวิเคราะห์คุณภาพของข้อสอบ ตรวจสอบความเที่ยง ความตรงเชิงโครงสร้าง ฉบับก่อนตัดกับฉบับหลังตัดข้อสอบที่ทำหน้าที่ต่างกันออกจากแบบทดสอบ ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบและเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 และ 2 ของแบบทดสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 จำนวน 8 กลุ่มสาระการเรียนรู้ ระหว่างวิธี IRT-LR กับวิธี SIBTEST มีข้อเสนอแนะในการนำผลการวิจัยไปใช้ดังนี้

1. สถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) หรือ สทศ. สามารถนำผลการวิเคราะห์คุณภาพของแบบทดสอบ O-NET ในกลุ่มสาระที่ผ่านเกณฑ์การวิเคราะห์คุณภาพโดยใช้ทฤษฎีการตอบสนองข้อสอบ แบบ 3 พารามิเตอร์ ไปใช้เป็นแนวทางในการทำข้อสอบต่อไปเพื่อใช้วัดความสามารถของนักเรียนในระดับชั้นมัธยมศึกษาปีที่ 6 ของสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) หรือ สทศ.

2. นักวิจัยและนักวัดผลทางการศึกษาที่มีความสนใจเกี่ยวกับการตรวจสอบทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธีการตรวจสอบที่อยู่บนพื้นฐานของทฤษฎีการตอบสนองข้อสอบ (IRT) ทั้งวิธี IRT-LR และวิธี SIBTEST เป็นวิธีที่มีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบทั้ง 2 วิธี

ข้อเสนอแนะสำหรับการวิจัยต่อไป

1. วิธี IRT-LR มีประสิทธิภาพในการตรวจพบข้อสอบที่ทำหน้าที่ต่างกันของข้อสอบเมื่อตรวจสอบการทำหน้าที่ต่างกันของข้อสอบกับข้อสอบที่มีการให้คะแนนแบบ 2 ค่าจึงควรมีการเปรียบเทียบเพิ่มเติมกับวิธีการตรวจสอบอื่นๆและศึกษาตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบที่มีการตรวจให้คะแนนแบบมากกว่า 2 ค่า

2. ควรมีการศึกษาวิธีการทำหน้าที่ต่างกันของข้อสอบด้วยทฤษฎีการตอบสนองข้อสอบด้วยวิธี Lord's x^2 วิธี General IRT Likelihood Ratio และ วิธี Log linear IRT Likelihood Ratio ว่าวิธีใดมีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบมากกว่ากัน

3. ควรมีการศึกษาวิธีการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบ NT แบบทดสอบ GAT แบบทดสอบ PAT และแบบทดสอบ A-NET ในการวัดความสามารถของผู้เรียนที่หลากหลายเพิ่มมากขึ้น

4. ควรมีการศึกษาความตรงเชิงโครงสร้างโดยการวิเคราะห์องค์ประกอบ โดยใช้แบบทดสอบทั่วไปในสถานศึกษาซึ่งเป็นข้อมูลปฐมภูมิแทนข้อมูลทุติยภูมิ เพื่อดูว่าคำถามแต่ละข้อที่พัฒนาขึ้นของแต่ละองค์ประกอบเกาะกลุ่มกันเป็นโครงสร้างขององค์ประกอบนั้น ๆ หรือไม่

บรรณานุกรม

กระทรวงศึกษาธิการ. (2551). *หลักสูตรแกนกลางการศึกษาการศึกษาขั้นพื้นฐาน พุทธศักราช 2551.* กรุงเทพฯ: กระทรวงศึกษาธิการ.

กนกกร ศิริสุข, อิศรัฏฐ์ รินไธสง และจรัส อติวิทยาภรณ์ (2555). พัฒนาการคุณภาพการศึกษาไทยโดยผ่านโมเดลโค้งพัฒนาการในการศึกษาการเปลี่ยนแปลงระยะยาวของผลคะแนนการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน (O-NET) ของนักเรียนชั้นมัธยมศึกษาปีที่ 6. *วารสารศึกษาศาสตร์ มหาวิทยาลัยสงขลานครินทร์ วิทยาเขตปัตตานี*, *24*(2), 28-63.

จารุจิตร สิทธิปรุ ปิยะทิพย์ ตินวร และโสพส สุขานนท์สวัสดิ์. (2559). การพัฒนาโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ สำหรับการจัดสอบ O-NET ระดับชั้นมัธยมศึกษาปีที่ 3.*วารสารการวัดผลการศึกษา มหาวิทยาลัยมหาสารคาม, 22*(1), 47-62.

จิระนาฎ ฉวีพัฒน์, บัณฑิต อินสมบัติ และธีรยุทธ ภูเขา. (2554). การเปรียบเทียบคุณภาพของการปรับเทียบคะแนนตามแนวตั้งโดยการใช้แบบทดสอบร่วมภายในระหว่างวิธีเชิงเส้นตรงกับวิธีการตอบสนองข้อสอบแบบสามพารามิเตอร์. *วิทยาการวิจัยและวิทยาการปัญญา*, *9*(1), 78-89.

ชัยยศ ชาวระนอง, ไพรัตน์ วงษ์นาม, สมศักดิ์ ลิลา และสุรีพร อนุศาสนนันท์. (2553). ประสิทธิภาพของโมเดลของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบพหุมิติโดยวิธีการวิเคราะห์องค์ประกอบยืนยันแฝงภายใน. *วารสารศึกษาศาสตร์*, *22*(1), 23-35.

ชัยวัฒน์ หฤทัยพันธ์. (2558). การพัฒนาวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการัดสินของผู้เชี่ยวชาญ. *วารสารครุศาสตร์, 43*(1), 1-18.

ณรงค์ จันทรมหา (2554).การเปรียบเทียบค่าความเที่ยงของแบบทดสอบผลสัมฤทธิ์ทางการเรียนที่มีจำนวนข้อสอบทำหน้าที่ต่างกันแตกต่างกัน. *วิทยาการวิจัยและวิทยาการปัญญา, 8*(2), 58-71.

ธเกียรติกมล ทองงอก, โชติกา ภาษีผล และศิริชัย กาญจนวาสี. (2556). ประสิทธิภาพการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบภายใต้วิธีถดถอยโลจิสติกสำหรับข้อสอบที่ตรวจให้คะแนนแบบทวิภาค:เปรียบเทียบขนาดอิทธิพลสองเกณฑ์. *วารสารวิจัย มสด. สาขามนุษยศาสตร์สังคมศาสตร์, 9*(2), 31-49.

นุภาพรรณ ปลื้มใจ, ปิยะทิพย์ ตินวร และโสพส สุขานนท์สวัสดิ์. (2558). การพัฒนาโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์สำหรับการจัดสอบ O-NET ระดับชั้นมัธยมศึกษาปีที่ 6. *วิทยาการวิจัยและวิทยาการปัญญา, 13*(2), 109-125.

พิชชา สุริอาจ และประกฤติยา ทักษิโณ. (2559). การพัฒนาแบบวัดความตระหนักต่อโลกในยุคศตวรรษที่ 21 ของนักเรียนชั้นมัธยมศึกษาตอนต้นโดยใช้แบบวัดเชิงสถานการณ์: การประยุกต์ใช้การทำหน้าที่ต่างกันของข้อสอบ. *วารสารศึกษาศาสตร์ ฉบับวิจัยบัณฑิตศึกษา, 10*(พิเศษ), 94-100.

พีรญา สูงเนิน, เสรี ชัดแช้ม และสมโภชน์ อเนกสุข. (2552). การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบพหุมิติ: การเปรียบเทียบระหว่างรายข้อกับรายหมวดข้อสอบ โดยใช้วิธีซิปเทสท์. *วิทยาการวิจัยและวิทยาการปัญญา, 6*(2), 49-62.

พูลพงศ์ สุขสว่าง. (2556). *โมเดลสมการโครงสร้าง.* กรุงเทพฯ: โรงพิมพ์วัฒนาพานิชจำกัด.

พูลพงศ์ สุขสว่าง. (2561). *โมเดลสมการโครงสร้าง.* ชลบุรี: เอ.พี.บลูปริ้นท์.

มนัส สายแก้ว และนิตยา ภัสสรศิริ. (2555).ความสัมพันธ์ระหว่างสุขภาพองค์การกับคะแนนการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐานของนักเรียนระดับมัธยมศึกษา ในจังหวัดสุราษฎร์ธานี. *วารสารศึกษาศาสตร์ มสธ, 5*(1), 97-108.

เยาวดี รางชัยกุล วิบูลย์ศรี. (2556). *การวัดและการสร้างแบบสอบผลสัมฤทธิ์* (พิมพ์ครั้งที่ 11). กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

วลัยพร สุขปลั่ง และบรรพต วิรุณราช. (2558). การวิเคราะห์องค์ประกอบเชิงยืนยันอันดับที่สองของโมเดลสมรรถนะบุคลากรสายผู้สอนสาขาวิชาการบริหารทรัพยากรมนุษย์ สังกัดมหาวิทยาลัยราชภัฏเพื่อรองรับการเข้าสู่ประชาคมอาเซียน. *วารสารวิทยาลัยพาณิชยศาสตร์บูรพาปริทัศน์*, *10*(2), 92-103.

วีรวรรณ โฆษิตสุคต. (2557). การพัฒนาแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐานพุทธศักราช 2551 กลุ่มสาระการเรียนรู้ภาษาไทย ชั้นมัธยมศึกษาปีที่ 3, *วารสารวิชาการ Veridian E-Journal, 7*(2), 962-973.

ศิริชัย กาญจนวาสี. (2555). *ทฤษฎีการทดสอบแนวใหม่* (พิมพ์ครั้งที่ 4). กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

ศิริชัย กาญจนวาสี. (2556). *ทฤษฎีการทดสอบแบบดั้งเดิม* (พิมพ์ครั้งที่ 7). กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

ศุภวัฒน์ มะลิเผือก. (2549). อิทธิพลของการทำหน้าที่ต่างกันของข้อสอบที่ส่งผลต่อคุณภาพของแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนระดับชาติ ชั้นประถมศึกษาปีที่ 6. *วารสารวิจัยและวัดผลการศึกษา มหาวิทยาลัยบูรพา*, *14*(1), 46-60.

สถาพร สุขสมบูรณ์. (2559). รูปแบบการจัดการรักษาความปลอดภัยในการแข่งขันฟุตบอลอาชีพไทยพรีเมียร์ลีก. *วารสารบัณฑิตศึกษา: มหาวิทยาลัยราชภัฏสวนสุนันทา, 2*(2), 163-183.

สัมพันธ์ พันธุ์พฤกษ์. (2557). *เอกสารประกอบการบรรยายการพัฒนาสมรรถนะด้านการวัดผลประเมินผลการเรียนรู้.* สถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน).

สุชาดา กรเพชรปาณี, ปิยะทิพย์ ตินวร และโสพส สุขานนท์สวัสดิ์. (2559). การพัฒนาโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์สำหรับการจัดสอบ O-NET. *วิทยาการวิจัยและวิทยาการปัญญา, 14*(1), 14-31.

สุนทร เทียนงาม, ศิริชัย กาญจนวาสี และดิเรก ศรีสุโข. (2553). ผลของความไม่เป็นอิสระของข้อสอบที่มีต่อค่าความเที่ยง ค่าพารามิเตอร์ของข้อสอบ ค่าความสามารถของผู้สอบ และสารสนเทศของแบบสอบเมื่อมีเงื่อนไขการทดสอบที่แตกต่างกัน. *วารสารวิธีวิทยาการวิจัย, 23*(3), 247-271.

สุพัฒนา หอมบุปผา, ไพรัตน์ วงษ์นาม และสมพงษ์ ปั้นหุ่น. (2556). การเปรียบเทียบการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธี HGLM วิธี MIMIC และวิธี BAYESIAN.*วารสารวิจัยราชภัฏพระนคร, 8*(2), 10-24.

สุธาทิพย์ ตรีสิน และปิยทิพย์ ประดุจพรม. (2560). การเปรียบเทียบผลการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบระดับชาติ ด้านภาษา ด้านคำนวณ และด้านเหตุผล ชั้นประถมศึกษาปีที่ 3 ด้วยวิธี HGLM วิธี MIMIC และวิธี IRT- LR. *วิทยาการวิจัยและวิทยาการปัญญา, 15*(2), 109-119.

สุวิมล ติรกานันท์. (2555). *การวิเคราะห์ตัวแปรพหุในงานวิจัยทางสังคมศาสตร์.* กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

สำราญ มีแจ้ง, ประภัสสร วงษ์ดี และยุพิน โกณฑา. (2553). การเปรียบเทียบคะแนน O-NET ระหว่างปีโดยวิธีการเปรียบเทียบเป็นมาตรเดี่ยวกันกับโดยใช้ทฤษฎีการตอบสนองข้อสอบ. *วิทยาการวิจัยและวิทยาการปัญญา*, *7*(2), 83-92.

อิทธิพัทธ์ สุวทันพรกุล. (2558). การวิเคราะห์และสังเคราะห์กลยุทธ์การพัฒนาคุณภาพผู้เรียนโดยใช้ผลการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน: พหุกรณีศึกษาโรงเรียนที่มีผลการทดสอบทางการศึกษาในระดับดี. *วารสารวิธีวิทยาการวิจัย*, *28*(1), 27-47.

เอกสารประกอบหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551. (2552). สำนักวิชาการมาตรฐานการศึกษา.

Tan P., ม.ร.ว. สมพร สุทัศนีย์ และสุพิมพ์ ศรีพันธ์วรสกุล. (2557). การพัฒนามาตรวัดแรงจูงใจในการเรียนภาษาอังกฤษสำหรับนักเรียนชั้นมัธยมศึกษาตอนปลาย ในราชอาณาจักรกัมพูชา. *วิทยาการวิจัยและวิทยาการปัญญา, 11*(2), 16-27.

Acar, T. (2011). Sample size in differential item functioning: An application of hierarchical linear modeling. *Kuram ve Uygulamada Eğitim Bilimleri, 11*(1), 284-288.

Abd-El-Fattah, M. S., AL-Sinani, Y., Shourbagi, E. S. & Fakhroo, A. H. (2014). Using Rasch analysis to examine the dimensionality structure and differential item functioning of the Arabic version of the Perceived Physical Ability Scale for Children. *Australian Journal of Educational & Developmental Psychology, 14*(2014), 29-44.

Angoff, W. H., & Ford, S. F. (1973). Item-race interaction on the test of scholastic apitude. *Journal of Education Measurement***,** *10*, 95-105.

Apinyapibal, S., Lawthong, N., & Kanjanawasee, S. (2015). A Comparative Analysis of the Efficacy of Differential Item Functioning Detection for Dichotomously Scored Items among Logistic Regression, SIBTEST and Raschtree Methods. *Procedia - Social and Behavioral Sciences, 191*(2015), 21–25.

Aryadoust, V. (2018). Using recursive partitioning Rasch trees to investigate differential item functioning in second language reading tests. *Studies in Educational Evaluation, 56*(2018), 197–204.

Caballero, O. A., Casares, V. A. J., & Caballero, O. M. (2017). Visual perception in art education. Gender and intercultural study. *Procedia - Social and Behavioral Sciences, 237*(2017), 588–593.

Camilli, G., & Shapard, L. A. (1994).*Methods for identifying biased test items.* London: Sage Publications.

Chang, H. H., Mazzeo, J., & Roussos, L. (1996). Detecting DIF for polytomously scored items: An adaptation of the SIBTEST procedure. *Journal of Educational Measurment*, *33*(3), 333-353.

Chanpleng, P., Lawthong, N., & Ngudgratoke, S. (2015). A study of Quality Assessment of Science Instructional Management in Thailand: An analysis of Differential Item Functioning and Test Functioning in Mixed format Tests. *Procedia - Social and Behavioral Sciences, 191*(2015), 121–125.

Cleary, T. A., & Hilton, T. L. (1968). An investigation of item bias. *Educational and Psychological Measurement*, *28*(1), 61-75.

Cronbach, L. J. (1951). Coefficient Alpha and the Internal Structure of Tests. *Psychometrika*, *16*(3), 297-334.

Cronbach, L., & Furby, L. (1970). How should we measure change – Or should we?. *Psychological Bulletin*, *74*(1), 68-80.

Cronbach, L. J., & Meehl, P. E. (1955). Construct Validity in Psychological Test. *Psychological Bulletin*, *52*(4), 281-302.

Diamantopoulos, A., & Siguaw, A. D. (2000). *Introducing LISREL: A Guide for the Uninitiated*. London: Sage Publication.

Dimitrov, D. M. (2003). Marginal true score measures and reliability of binary item as a function of their IRT parameters. *Applied Psychological Measurement*, *27*(6), 440-458.

Dogan, N., & Yurtcu, M. (2015). Invariability of Item Parameter in the Sample with Restricted Real Data Results. *Procedia-Social and Behavioral Sc*ience, *197*(2015), 2207-2214.

Dorans, N. J., & Kulick, E. M. (1986). Demonstrating the utility of the standardization approach to assessing unexpected differential item performance on the Scholastic Aptitude Test. *Journal of Educational Measurement*, *23*, 355-358.

Ellis, B. B., & Mead, A. D. (2002). *Item analysis: Theory and practice using classical and modern test theory.* In S. G. Rogelberg (Ed.), Handbook of research methods in industrial and organizational psychology (pp. 324-343). Malden, Massachusetts: Blackwell Publishing, Inc.

Elosua, E., & Wells, C. S. (2013).Detecting DIF in Polytomous Items Using MACS, IRT and Ordinal Logistic Regression International Journal of Methodology and Experimental Psychology*, 34*(2), 327-342.

Finch, H. W. & French, F. B. (2014). The impact of group pseudo-guessing parameter differences on the detection of uniform and nonuniform DIF. *Psychological Test and Assessment Modeling, 56*(1), 25-44.

Freund, J. F., & Walpole, R. E. (1980). *Mathematical statistics*. (3rd ed.). New Jersey: Prentice-Hall. 548.

Gierl, J. M., Gotzmann, A. & Boughton, A. K. (2004). Performance of SIBTEST When the Percentage of DIF Items is Large. *Applied Measurement in Education, 17*(3), 241–264.

Goh, C., & Aryadoust, S. V. (2010).Investigating the Construct Validity of the MELAB Listening Test through the Rasch Analysis and Correlated Uniqueness Modeling. *Spaan Fellow Working Papers in Second or Foreign Language Assessment, 8*, 31–68.

Gulliksen, H. (1950). Theory of mental tests. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc.

Guyer, R., & Thompson, N. A. (2014). User's Manual for Xcalibre item response theory calibration software, version 4.2.2 and later. Woodbury Minnesota: Assessment Systems Corporation.

Hambleton, R. K., & Swaminathan, H. (1985). Item Response Theory: Principles and applications. Norwell, Massachusetts: Kluwer Academic Publishers.

Hambleton, R. K., & Swaminathan, H., & Roger, H. J. (1991). *Fundamentals of Item Response Theory.* California: Sage Publications.

Holland, P. W., & Thayer, D. T. (1988). *Differential item functioning and the Mantel – Haenzel procedure.In H. Wainer & H.I Braun (Eds.).* Test validity. (pp.129-145) Hillsdale, New Jersey: Lawrence Erlbaum.

Holland, P. W., & Wainer, H. (1993). D*ifferential Item Functioning.* Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Hoyt, C. (1941). Test Reliability Obtained by Analysis of Variance. *Psychometric*, *6*(3), 153-160.

Jansen, M., Schroeders, U., & Lüdtke, O. (2014). Academic self-concept in science: Multidimensionality, relations to achievement measures, and gender differences. *Learning and Individual Differences, 30*(2014), 11–21.

Jensen, A. R. (1974). How biased are culture – loaded Test?. *Genetic Psychology Monographs*, *90*, 185-224.

Jehangir, K., Glas, C. A. W., & Berg, S. M. (2015). Correcting for Differential Item Functioning in Multi-level Regression Models in Cross-National Surveys. *Measurement*, *66*, 263-271.

Joreskog, K. G., & Sorbom, D. (1985). *LISREL VI. Analysis of Linear Structural Relationships by Minimum Likelihood, Instrumental Variables, and Least Squares Method*. Mooresville, Indiana: Scientific Software, Inc.

Joreskog, K. G., & Sorbom, D. (1993). *LISREL 8 Structural equation modeling with the simplis command language.* Hillsdale, New Jersey: Erlbaum.

Kabasakal, A. K., Arsan, N., Gök, B. & Kelecioğlu, H. (2014). Comparing Performances (Type I error and Power) of IRT Likelihood Ratio SIBTEST and Mantel-Haenszel Methods in the Determination of Differential Item Functioning. *Educational Sciences: Theory & Practice, 14*(6), 2186-2193.

Kim, S. H., & Cohen, A. S. (1991). A comparison of two area measures of detecting differential Item functioning. *Applied Psychological Measurement*, *15*(3), 269-278.

Kose, A. I., & Demirtasli, C. N. (2012). Comparison of unidimensional and multidimensional models based on item response theory in terms of both variables of test length and sample size. *Procedia - Social and Behavioral Sciences, 46*(2012), 135–140.

Kuder, G. F., & Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika*, *2*(3), 151-160.

Li, H., & Stout, W. (1996). Multidimensional DIF analyses: The effects of matching on unidimensional subtest scores. *Psychological Measurement*, *22*(4), 357-367.

Linn, R. L., & Levine, M. V., Hastings, C.N., & Wardrop, J. L. (1981). Item Bias in a Test of Reading Comprehension. *Applied Psychological Measurement, 5*(2), 159-173.

Load, F. M. (1980). *Applications of item response theory to practical testing problems.* Hillsdale, New Jersey: Erlbaum.

Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports*, *3*, 635-694.

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

Mark, J. G., Andrea, G., & Keith, A. B. (2004). Performance of SIBTEST When the Percentage of DIF Items is Large. *Applied Measurement in Education*, *17*(3), 241-264.

Master, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika***,** *47*(2), 149-174.

Maydeu-Olivares, A. (2001). Multidimensional item response theory modeling of binary data: Large sample properties of noharm estimates. *Journal of Educational and Behavioral Statistics*, *26*(1), 51-71.

Mazor, K. M., Clauser, B. E., & Hambleton, R. K. (1992). The effect of sample size on the functioning of the Mantel-Haenszel statistic. *Educational and Psychological Measurement*, *52*(3), 443-451.

Maier, N. R. F., & Casselman, G. G. (1970). Locating the difficulty in insight problems. *Individual and Sex Differences Psychological Rep*, *26*(1), 77-103.

Mellenbergh, G. J. (1982). Contingency table models for assessing item bias. *Journal of Educational Statistics*, *7*(2), 105-118.

MEI, L. O. (2014). *A Comparison of Differential Item Functioning (DIF) Detection for Dichotomusly Scored Items By 2.1 IRTPRO, BILOG-MG 3, and IRTLRDIF V.2*. Paper Presented 79th Annual Meeting of the Psychometric Society Madison, Wisconsin. Retrieved June, 2, 2014 from: http://conferencing.uwex.edu/ conferences/ps2014/docs/IMPS_Abstracts_Web.pdf.

Messick, S. (1980). Test validity and the ethics of assessment. *American Psychologist*, *35*(11), 1012-1027.

Millsap, R. E., & Everson, H. T. (1993). Methodology review: Statistical approaches for assessing measurement bias. *Applied Psychological Measurement*, *17*(4), 297-334.

Moghadamzadeh, A., Salehi, K., & Khodaie, E. (2011). A Comparison the Information Functions of the Item and Test in One, Two and Three Parametric Model of the Item Response Theory (IRT). *Procedia - Social and Behavioral Sciences, 29*(2011), 1359–1367

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement*, *16*(2), 159-176.

Muthén, B., & Asparouhov, T. (2012). Bayesian SEM: A more flexible representation of substantive theory. *Psychological Methods*, *17*(3), 313-335.

Nandakumar, R. (1993).Assessing Essential Dimensionality of Real Data. *Applied Psychological Measurement, 17*, 29-38.

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

Narayanan, P., & Swaminathan, H. (1996). Identification of items that show nonuniform DIF. *Applied Psychological Measurement*, *20*(3), 257-274.

Na Wichian, S., Wongwanich, S., & Saengsiri, P. (2014). The Development of an O-NET Score Forecasting System. *Procedia - Social and Behavioral Sciences, 116*(2014), 1594 – 1598.

Nenty, H. J. (2015). *Dimensionality Analysis of students' Performance in 2013 BGCSE Agricultural Examination: Implications for Differential Item Functioning*. Degree of Master of Education in Research and Evaluation, University of Botswana.

Nenty, H. J., Kgosidialwa, K., & Batsalelwang, K. J. (2017). Sex, Gender-typing and Performance in 2015 JSC 3 Mathematics by Students in Gaborone. *Advances in Social Sciences Research Journal, 4*(20), 9-23.

Nordin, B. A. R., Ahmad, Z. B. K., & Lee, M. T. (2012). Examining Quality Mathematics Test Item Using Rasch Model: Preminarily Analysis. *Procedia-Social and Behavioral Sc*iences, *69*, 2205-2214.

Nunnally, J. C., & Bernstein, I. C. (1994). Psychometric theory (3rd ed.). New York: McGraw-Hill. Inc.

Öğretmen, T. (2015). DIF analysis across genders for reading comprehension part of English language achievement exam as a foreign language. *Academic Journals, 10*(11), 1505-1513.

Özdemir, B. (2015). A comparison of IRT-based methods for examining differential item functioning in TIMSS 2011 mathematics subtest. *Procedia - Social and Behavioral Sciences, 174*(2015), 2075–2083.

Peak, I., & Han, K. T. (2012). IRTPRO 2.1 for Windows (Item Response Theory for Patient Reported Outcomes). *Applied Psychological Measurment*, *37*(3), 242-252.

Penner, M. A., & Paret, M. (2008). Gender differences in mathematics achievement: Exploring the early grades and the extremes. *Social Science Research. 37*(2008), 239–253.

Potenza, M. T., & Doran, N. J. (1995). DIF assessment for polytomous scored item: A framework for classification and envaluation. *Psicológica, 34*, 327-342.

Raju, N. S. (1990). Determining the significance of estimated signed and unsigned areas between two item response functions. *Applied Psychological Measurement*, *14*(2), 197-207.

Roussos, L. A., & Stout, W. F. (1996). Simulation Studies of the Effects of Small Sample Size and Studied Item Parameters on SIBTEST and Mantel-Haenszel Type I Error Performance. *Journal of Educational Measurement*, *33*(2), 215–230.

Rozowski, M., & Reith, J. (1999). Examining the measurement quality of tests containing differentially functioning items: Do biased item result in poor measurement. *Educational and Psychological Measurement*, *59*(2), 248-269.

Rudner, L. M., (1977). *An Approach to Biased Item Identification Using Latent Trait Measurement Theory.* Paper presented at the annual meeting of the American Educational Research Association.

Rudner, L. M., Getson, P. R., & Knight, D. L. (1980). Biased Item Detection Techniques. *Journal of Educational Statistics*, *5*(3), 213-233.

Schumacker, R. E., & Lomax, R. G. (2010). A beginner's guide to structural equation modeling. (3rd ed.). New Jersey: Lawrence Erlbaum Associates.

Scott, N. W., Fayers, P. M., Aaronson, N. K., Bottomley, A., Graeff, A. D., Groenvold, M., et al. (2009). A simulation study provided sample size guidance for differential item functioning (DIF) studies using short scales. *Journal of Clinical Epidemiology*, *62*(3), 288-295.

Shealy, R., & Stout, W. F. (1993). An item response theory model for test bias. In P.W. Holland & H. Wainer (Eds.). *Differential item functioning: Theory and practice*. (197-239) Hillsdale, New Jersey: Lawrence Erlbaum.

Shepard, L. A., Camilli, G., & Williams, D. M. (1984). Accounting for Statistical Artifacts in Item Bias Research. *Journal of Educational Statistics*, *9*(2), 93-128.

Sokuler, F. N., & Justman, M. (2016). Gender streaming and prior achievement in high school science and mathematics. *Economics of Education Review, 53*(2016), 230–253.

Stout, W. F., Hsin-Hung, L., Nandakumar, R., & Bolt, D. (1997). MULTISIB: A procedure to investigate DIF when a test is intentionally two- dimensional. *Applied Psychologocal Measurement*, *21*, 195-213.

Stout, W. F., & Roussos, L. A. (1999). Dimensionality-based DIF/DBF Package [Computer Program]. Champaign – Urbana, Illinois: William Stout Institute for Measurement, University of Illinois.

Swaminathan, H., & Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement*, *27*(4), 361-370.

Thissen, D. M., Steinberg, L., & Wainer, H. (1993). Detection of differential item functioning using the parameters of item response models .In P.W. Holland & H. Wainer (Eds.). *Differential item functioning: Theory and practice*. (pp.67-113) Hillsdale, New Jersey: Lawrence Erlbaum.

Thissen, D. (2009). The MEDPRO project: An SBIR project for a comprehensive IRT and CAT software system - IRT software. In D. J. Weiss (Ed.), *Proceedings of the 2009 GMAC Conference on Computerized Adaptive Testing.* Retrieved from www.psych.umn.edu/psylabs/CATCentral/

Urry, V. W. (1977). Tailored testing: A Successful application of latent trait theory. *Journal of Educational Measurement*, *14*(2), 181-196.

Wainer, H. (2000). *Computerized Adaptive Testing: A Primer.* New Jersey: Lawrence Erlbaum Associates, Inc., Publishers.

ภาคผนวก

ภาคผนวก ก
หนังสือขอความอนุเคราะห์ขอข้อมูลเพื่อการวิจัย

ที่ ศธ ๖๖๒๘/๐๑๑๖

วิทยาลัยวิทยาการวิจัยและวิทยาการปัญญา
มหาวิทยาลัยบูรพา
ต.แสนสุข อ.เมือง จ.ชลบุรี ๒๐๑๓๑

๒ พฤษภาคม ๒๕๕๙

เรื่อง ขอความอนุเคราะห์ขอข้อมูลเพื่อการวิจัย

เรียน ผู้อำนวยการสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน)

สิ่งที่ส่งมาด้วย ๑. เค้าโครงวิทยานิพนธ์ฉบับย่อ จำนวน ๑ ชุด
๒. ตัวอย่างเครื่องมือประกอบการวิจัย จำนวน ๑ ชุด
๓. แบบรายงานผลการพิจารณาจริยธรรมการวิจัยในคน จำนวน ๑ ชุด

ด้วย นางพิจักษณา กาวี รหัสประจำตัว ๕๖๙๑๐๓๙๒ นิสิตหลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาการวิจัยและสถิติทางวิทยาการปัญญา ได้รับอนุมัติให้ทำวิทยานิพนธ์เรื่อง "การเปรียบเทียบผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบ O-NET ระดับชั้นมัธยมศึกษาปีที่ ๖ ระหว่างวิธี IRT และวิธี SIBTEST" ซึ่งอยู่ในความควบคุมดูแลของ ดร.ปิยะทิพย์ ประดุจพรม อาจารย์ที่ปรึกษาหลัก ในการนี้ ผู้วิจัยมีความประสงค์ขอความอนุเคราะห์ขอข้อสอบ O-NET ระดับชั้นมัธยมศึกษาปีที่ ๖ ปีการศึกษา ๒๕๕๖ ทั้ง ๘ กลุ่มสาระการเรียนรู้ ได้แก่ ๑. ภาษาไทย ๒. คณิตศาสตร์ ๓. วิทยาศาสตร์ ๔. สังคมศึกษา ศาสนาและวัฒนธรรม ๕. ศิลปะ ๖. การงานอาชีพและเทคโนโลยี ๗. สุขศึกษาและพลศึกษา ๘. ภาษาต่างประเทศ และผลการตอบข้อสอบ O-NET ระดับชั้นมัธยมศึกษาปีที่ ๖ ปีการศึกษา ๒๕๕๖ จำแนกเพศชายและเพศหญิง ทั้ง ๘ กลุ่มสาระการเรียนรู้ พร้อมเฉลย

จึงเรียนมาเพื่อโปรดพิจารณา วิทยาลัยวิทยาการวิจัยและวิทยาการปัญญา หวังเป็นอย่างยิ่งว่าคงจะได้รับความอนุเคราะห์จากท่านด้วยดี และขอขอบคุณมา ณ โอกาสนี้

ขอแสดงความนับถือ

(ผู้ช่วยศาสตราจารย์ ดร.สุชาดา กรเพชรปาณี)
คณบดีวิทยาลัยวิทยาการวิจัยและวิทยาการปัญญา

วิทยาลัยวิทยาการวิจัยและวิทยาการปัญญา
โทร. ๐ ๓๘๑๐ ๒๐๗๗-๘
โทร/ โทรสาร ๐ ๓๘๓๙ ๓๔๘๔
http ://www.rmcs.buu.ac.th

ภาคผนวก ข
บันทึกข้อตกลงว่าด้วยการขอใช้ประโยชน์จากข้อสอบ ในการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน (O-NET)

คู่ฉบับ

บันทึกข้อตกลง
ว่าด้วยการขอใช้ประโยชน์จากข้อสอบ
ในการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน (O-NET)
ระหว่าง
สถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) กับ นางพิจักษณา กาวี

บันทึกข้อตกลงฉบับนี้ทำขึ้น ณ สถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) เมื่อวันที่ ๒๒ มิถุนายน ๒๕๕๙ ระหว่าง สถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) โดย ดร.ทรายทอง พวกสันเทียะ ตำแหน่ง รองผู้อำนวยการ ปฏิบัติงานแทนผู้อำนวยการสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) สำนักงานตั้งอยู่ที่ ๑๒๘ อาคารพญาไทพลาซ่า ชั้น ๓๖ แขวงทุ่งพญาไท เขตราชเทวี กรุงเทพมหานคร ซึ่งต่อไปนี้เรียกว่า "ผู้ให้ข้อมูล" ฝ่ายหนึ่ง กับ นางพิจักษณา กาวี อยู่ที่ ๙๘ หมู่ที่ ๕ ตำบลหันทราย อำเภออรัญประเทศ จังหวัดสระแก้ว ซึ่งต่อไปนี้เรียกว่า "ผู้ขอข้อมูล" อีกฝ่ายหนึ่ง ทั้งสองฝ่ายได้ตกลงร่วมกันจัดทำบันทึกข้อตกลงฯ ดังนี้

๑. บันทึกข้อตกลงนี้จัดทำขึ้นระหว่าง ผู้ให้ข้อมูล กับ ผู้ขอข้อมูล โดยมีวัตถุประสงค์ที่จะให้ผู้ขอข้อมูล ใช้ประโยชน์จากข้อมูลข้อสอบการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน (O-NET) ที่ผู้ให้ข้อมูลได้ดำเนินการจัดการทดสอบและรวบรวมข้อมูลขึ้นโดยข้อมูลที่จะนำไปใช้ประโยชน์ ได้แก่ **ข้อมูลข้อสอบการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน (O-NET) ของนักเรียนระดับชั้นมัธยมศึกษา ปีที่ ๖ ปีการศึกษา ๒๕๕๖**

๒. ผู้ขอข้อมูล ตกลงว่าจะนำข้อมูลไปใช้ประกอบภารกิจในการทำวิทยานิพนธ์ เรื่อง **"การเปรียบเทียบผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบ O-NET ระดับชั้นมัธยมศึกษาปีที่ ๖ ระหว่างวิธี IRT และวิธี SIBTEST"**

๓. ผู้ขอข้อมูล จะต้องเป็นผู้รับผิดชอบในการจัดหาวัสดุอุปกรณ์เพื่อใช้ในการบันทึกหรือบรรจุข้อมูลที่ขอใช้ประโยชน์ ตลอดจนรับผิดชอบค่าธรรมเนียมและค่าใช้จ่ายต่างๆ ที่เกิดขึ้นจากการนี้ด้วย

๔. ผู้ขอข้อมูล จะต้องไม่นำข้อมูลไปใช้ผิดจากข้อตกลงหรือนำไปใช้ในลักษณะที่ส่งผลกระทบหรือก่อให้เกิดความเสียหายแก่เจ้าของข้อมูล หากฝ่าฝืนหรือเกิดความรับผิดทางแพ่งหรืออาญาผู้ขอข้อมูลต้องเป็นผู้รับผิดชอบในฐานะผู้ก่อให้เกิดความเสียหายขั้นต้น และเป็นผู้รับผิดชอบทางแพ่งแต่เพียงฝ่ายเดียว

๕. เพื่อให้การปฏิบัติงาน เกิดความคล่องตัวและมีประสิทธิภาพ ผู้ให้ข้อมูล ได้มอบหมายให้นางวิลาวัณย์ บำรุงเวช เป็นผู้ประสานงานในการดำเนินการตามข้อตกลงนี้

๖. ผู้ให้ข้อมูล ขอสงวนสิทธิ์ในการระงับการใช้ประโยชน์จากข้อมูลข้อสอบการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน (O-NET) เป็นการชั่วคราวหรือการเปลี่ยนแปลงวิธีการขอใช้ประโยชน์ข้อมูลได้ตามที่เห็นสมควร

/๗. หากผู้ขอ ...

สทศ NIETS

๗. หากผู้ขอข้อมูลไม่ปฏิบัติตามหลักเกณฑ์หรือวิธีการปฏิบัติที่ได้กำหนดไว้ในบันทึกข้อตกลงฉบับนี้ ผู้ให้ข้อมูล มีสิทธิยกเลิกหรือระงับการใช้ประโยชน์จากข้อมูลข้อสอบการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน (O-NET) ได้ในทันทีที่ตรวจสอบพบ โดยไม่ต้องแจ้งให้ทราบล่วงหน้า

๘. เมื่อการใช้ประโยชน์จากข้อมูลข้อสอบการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน (O-NET) ตามข้อ ๑ เสร็จสิ้น ผู้ขอข้อมูล ตกลงว่าจะส่งผลงานวิทยานิพนธ์ฉบับสมบูรณ์ ตามข้อ ๒ จำนวน ๑ ชุด ให้แก่ผู้ให้ข้อมูลต่อไป

บันทึกข้อตกลงฉบับนี้มีผลบังคับใช้ตั้งแต่วันที่มีการลงนามในบันทึกข้อตกลงเป็นต้นไป และทำขึ้นเป็นสองฉบับ มีข้อความถูกต้องตรงกันทุกประการ ทั้งสองฝ่ายได้อ่านและเข้าใจข้อความในบันทึกข้อตกลงเป็นอย่างดีแล้ว เห็นว่าถูกต้องตรงตามความประสงค์ จึงได้ลงนามไว้เป็นสำคัญต่อหน้าพยานและเก็บรักษาไว้ฝ่ายละหนึ่งฉบับ

ลงชื่อ.....	ลงชื่อ.....
(นางพิจักษณา กาวี)	(ดร.ทรายทอง พวกสันเทียะ)
ผู้ขอข้อมูล	รองผู้อำนวยการ
	ปฏิบัติงานแทนผู้อำนวยการ
	สถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน)
ลงชื่อ.....พยาน	ลงชื่อ.....พยาน
(น.ส. พรทิพย์ โสภาเจริญสกล)	(นายธนพงษ์ แซ่ตั้ง)

ภาคผนวก ค
แบบรายงานผลการพิจารณาจริยธรรมการวิจัยในมนุษย์
วิทยาลัยวิทยาการวิจัยและวิทยาการปัญญา
มหาวิทยาลัยบูรพา

แบบรายงานผลการพิจารณาจริยธรรมการวิจัยในมนุษย์
วิทยาลัยวิทยาการวิจัยและวิทยาการปัญญา
มหาวิทยาลัยบูรพา

๑. ชื่อเรื่องวิทยานิพนธ์

ชื่อเรื่องวิทยานิพนธ์ (ภาษาไทย) การเปรียบเทียบผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบ O-NET ระดับชั้นมัธยมศึกษาปีที่ ๖ ระหว่างวิธี IRT กับวิธี SIBTEST

ชื่อเรื่องวิทยานิพนธ์ (ภาษาอังกฤษ) COMPARISON OF DIFFERENTIAL ITEM-FUNCTIONING ON O-NET FOR GRADE XII ANALYSES BETWEEN IRT AND SIBTEST METHODS

๒. ชื่อนิสิต (นาย, นาง, นางสาว): พิจักษณา กาวี

หลักสูตรวิทยาศาสตรมหาบัณฑิต (M.Sc.) สาขาวิชาการวิจัยและสถิติทางวิทยาการปัญญา

☑ ภาคปกติ ☐ ภาคพิเศษ

รหัสประจำตัว ๕๖๙๑๐๓๙๒ คณะ/วิทยาลัยวิทยาการวิจัยและวิทยาการปัญญา

๓. หน่วยงานที่สังกัด: วิทยาลัยวิทยาการวิจัยและวิทยาการปัญญา

๔. ผลการพิจารณาของคณะกรรมการพิจารณาจริยธรรมการวิจัยในมนุษย์:

คณะกรรมการพิจารณาจริยธรรมการวิจัยในมนุษย์ ได้พิจารณารายละเอียดวิทยานิพนธ์เรื่องดังกล่าวข้างต้นแล้ว ในประเด็นที่เกี่ยวกับ

๑) การเคารพในศักดิ์ศรี และสิทธิของมนุษย์ที่ใช้เป็นตัวอย่างการวิจัย

๒) วิธีการอย่างเหมาะสมในการได้รับความยินยอมจากกลุ่มตัวอย่างก่อนเข้าร่วมโครงการวิจัย (Informed consent) รวมทั้งการป้องกันสิทธิประโยชน์ และรักษาความลับกลุ่มตัวอย่างในการวิจัย

๓) การดำเนินการวิจัยอย่างเหมาะสม เพื่อไม่ก่อความเสียหายต่อสิ่งที่ศึกษาวิจัย ไม่ว่าจะเป็นสิ่งที่มีชีวิตหรือไม่มีชีวิต

คณะกรรมการพิจารณาจริยธรรมการวิจัยในมนุษย์ มีมติเห็นชอบ ดังนี้

(√) รับรองโครงการวิจัย

( ) ไม่รับรอง

๕. วันที่ที่ให้การรับรอง: ๑๙ เดือน สิงหาคม พ.ศ. ๒๕๕๘

ลงนาม.....

(ผู้ช่วยศาสตราจารย์ ดร.สุชาดา กรเพชรปาณี)
ประธานกรรมการพิจารณาจริยธรรมการวิจัยในมนุษย์
คณบดีวิทยาลัยวิทยาการวิจัยและวิทยาการปัญญา
วันที่ ๑๙ สิงหาคม พ.ศ. ๒๕๕๘

ภาคผนวก ง
ตัวอย่าง Print Out ผลการวิเคราะห์คุณภาพของข้อสอบ และค่าความเที่ยง ของแบบทดสอบ O-NET ฉบับก่อนและหลังตัดข้อสอบที่ทำหน้าที่ต่างกัน

IRT Item Parameter Calibration Report

User Test 1

Report created on 17/10/2017

Xcalibre 4.2.2.0: IRT Item Parameter Estimation Software

Copyright © 2014 - Assessment Systems Corporation

ASC

Introduction

This report provides the results of the IRT item parameter calibration by the computer program Xcalibre Version 4.2.2.0 (Assessment Systems Corporation, 2014) for User Test 1. The output is divided into four sections:

1. Specifications
2. E-M Algorithm
3. Summary statistics
4. Item-by-item results.

The statistical output is also recorded in a comma-separated value (CSV) file of the same name.

Specifications

This section records the input/output specifications and settings for historical purposes.

The Windows paths for the input files used in this analysis were:

C:\Users\Admin\Desktop\PJ-Gawee-(17 Oct 2017)\Xcariber cut DIF -2\04_math\LrtpXcal_Data.txt
C:\Users\Admin\Desktop\PJ-Gawee-(17 Oct 2017)\Xcariber cut DIF -2\04_math\LrtpXcal_ICF.txt

The Windows paths for the output files produced by this analysis were:

C:\Users\Admin\Desktop\PJ-Gawee-(17 Oct 2017)\Xcariber cut DIF -2\04_math\math-2.rtf
C:\Users\Admin\Desktop\PJ-Gawee-(17 Oct 2017)\Xcariber cut DIF -2\04_math\math-2.csv
C:\Users\Admin\Desktop\PJ-Gawee-(17 Oct 2017)\Xcariber cut DIF -2\04_math\math-2 Scores.csv

Table 1 presents the file specifications. Table 2 presents the IRT specifications used to perform the IRT item parameter calibration. Table 3 presents the flag specifications.

Table 1: File Specifications

Specification	Value	Specification	Value
Number of examinees	2000	Total Items	30
Calibrated Items	30	Pretest Items	0
Excluded Items	0	Number of domains	1
Classic Data Header	No	Delimited input	Yes
Delimiter for input	Tab	Number of ID columns	N/A
ID begins in column	N/A	Responses begin in column	N/A
Omit character	O	Not Admin character	-
Save item parameters	No	Item parameter format	N/A
Save data matrix	No	Omit codes are	N/A
Not Admin codes are	N/A	Score Not Admin as omits	No
Plot the IRFs	Yes	Save the IRFs and IIFs	No
Produce the fit line	Yes	# Groups for Plot	15
Type of score groups	Equally sized	# Groups for Chi-square	15
Perform classification	No	Classify using	N/A
Two-group cutpoint	N/A	Low group label	N/A
High group label	N/A	Merge empty poly categories	N/A

Table 2: IRT Calibration Specifications

Specification	Value	Specification	Value
IRT Specification	Dichotomous	Model constant	1.7
Polytomous IRT Model	N/A	Dichotomous IRT Model	3-parameter
Center the boundary locations	No	Centered value	N/A
Floating priors	Yes	a parameter prior mean (sd)	1.000 (0.250)
b parameter prior mean (sd)	0.000 (1.000)	c parameter prior mean (sd)	0.250 (0.025)
Theta estimation method	EAP	Bayesian prior mean (sd)	0.000 (1.000)
Maximum E-M loops	60	Convergence criterion	0.001
Quadrature points	40	Center dich item parameters on	theta
Acceptable P range	0.00 to 1.00	Acceptable item-corr range	-1.00 to 1.00
Acceptable item mean range	0.00 to 15.00	Correct for spuriousness	Yes
Fit statistic critical alpha	0.050	Minimum a	0.05
Maximum a	6.00	Minimum b	-4.00
Maximum b	4.00	Minimum c	0.00
Maximum c	0.70	Minimum theta	-7.00
Maximum theta	7.00	Treat scored items as poly	No
Center poly parameters on theta	No	Test for DIF	No
Group status column	N/A	Ability levels for DIF Test	N/A
Group 1 code	N/A	Group 2 code	N/A
Group 1 label	N/A	Group 2 label	N/A
Exclude items with low N	No	Minimum valid N	N/A
Compute scaled scores	No	Mean (SD) of scaled scores	N/A
Minimum scaled score	N/A	Maximum scaled score	N/A
Save statistics output	Yes	Delimiter	Comma
Save scores output	Yes	Delimiter	Comma
Save test information output	Yes	Delimiter	Comma
Save item information output	Yes	Delimiter	Comma

Table 3: Flag Specifications

Specification	Value	Specification	Value
Low a Flag Bound	0.30	High a Flag Bound	4.00
Low b Flag Bound	-3.00	High b Flag Bound	3.00
Low c Flag Bound	0.00	High c Flag Bound	0.40
Key Flag	K	Fit Flag	F
Low a Flag	La	High a Flag	Ha
Low b Flag	Lb	High b Flag	Hb
Low c Flag	Lc	High c Flag	Hc

E-M Algorithm

Xcalibre uses the expectation-maximization approach to calibrate item parameters. The estimation process is iterative, and repeated in loops until the convergence criterion is satisfied. The following list presents the item with the largest parameter change after each loop, and the value of the change.

The number of loops needed is evidence regarding the fit of the data; if many loops are required, or convergence is never reached, it means that the data does not fit well with the selected IRT model.

Item 1 failed to converge on this loop
Item 4 failed to converge on this loop
Item 7 failed to converge on this loop
Item 12 failed to converge on this loop
Item 14 failed to converge on this loop
Item 15 failed to converge on this loop
Item 16 failed to converge on this loop
Item 18 failed to converge on this loop
Item 19 failed to converge on this loop
Item 23 failed to converge on this loop
Item 24 failed to converge on this loop
Item 29 failed to converge on this loop
Item 30 failed to converge on this loop
Maximum change after Loop 1 was 3.7850 for Item 10 for the b parameter
Item 2 failed to converge on this loop
Maximum change after Loop 2 was -1.4330 for Item 10 for the b parameter
Maximum change after Loop 3 was -0.4509 for Item 10 for the a parameter
Maximum change after Loop 4 was 0.3283 for Item 2 for the a parameter
Maximum change after Loop 5 was 0.3331 for Item 1 for the a parameter
Maximum change after Loop 6 was 0.1404 for Item 1 for the a parameter
Maximum change after Loop 7 was 0.0760 for Item 1 for the a parameter
Maximum change after Loop 8 was -0.0526 for Item 25 for the b parameter
Maximum change after Loop 9 was -0.0405 for Item 12 for the b parameter
Maximum change after Loop 10 was -0.0306 for Item 3 for the b parameter
Maximum change after Loop 11 was -0.0241 for Item 4 for the b parameter
Maximum change after Loop 12 was -0.0168 for Item 26 for the b parameter
Maximum change after Loop 13 was -0.0138 for Item 26 for the b parameter
Maximum change after Loop 14 was -0.0116 for Item 26 for the b parameter
Maximum change after Loop 15 was -0.0099 for Item 26 for the b parameter
Maximum change after Loop 16 was -0.0089 for Item 2 for the b parameter
Maximum change after Loop 17 was -0.0080 for Item 2 for the b parameter
Maximum change after Loop 18 was -0.0072 for Item 2 for the b parameter
Maximum change after Loop 19 was -0.0065 for Item 2 for the b parameter
Maximum change after Loop 20 was -0.0059 for Item 2 for the b parameter
Maximum change after Loop 21 was -0.0053 for Item 2 for the b parameter
Maximum change after Loop 22 was -0.0048 for Item 2 for the b parameter
Maximum change after Loop 23 was -0.0043 for Item 2 for the b parameter
Maximum change after Loop 24 was -0.0038 for Item 2 for the b parameter
Maximum change after Loop 25 was -0.0035 for Item 2 for the b parameter
Maximum change after Loop 26 was -0.0031 for Item 2 for the b parameter
Maximum change after Loop 27 was -0.0028 for Item 2 for the b parameter
Maximum change after Loop 28 was -0.0026 for Item 2 for the b parameter
Maximum change after Loop 29 was -0.0023 for Item 2 for the b parameter
Maximum change after Loop 30 was -0.0021 for Item 2 for the b parameter
Maximum change after Loop 31 was -0.0019 for Item 2 for the b parameter
Maximum change after Loop 32 was -0.0017 for Item 2 for the b parameter
Maximum change after Loop 33 was -0.0016 for Item 2 for the b parameter
Maximum change after Loop 34 was -0.0014 for Item 2 for the b parameter
Maximum change after Loop 35 was -0.0013 for Item 2 for the b parameter
Maximum change after Loop 36 was -0.0012 for Item 2 for the b parameter
Maximum change after Loop 37 was -0.0011 for Item 2 for the b parameter
Maximum change after Loop 38 was -0.0010 for Item 2 for the b parameter

Summary statistics

Table 4 presents the summary statistics for the item parameters for all calibrated items. Table 5 summarizes the total scores for the full test for just the calibrated items. Table 6 summarizes the theta estimates for the full test. Table 7 provides the overall model fit chi-square(s) for the full test. Definitions of these statistics are found in the Xcalibre manual.

Table 4: Summary Statistics for All Calibrated Items

Parameter	Items	Mean	SD	Min	Max
a	30	1.101	0.279	0.520	1.844
b	30	2.692	0.450	1.505	3.421
c	30	0.193	0.023	0.137	0.242

Table 5: Summary Statistics for the Total Scores

Test	Items	Alpha	Mean	SD	Skew	Min	Q1	Median	Q3	Max	IQR
Full Test	30	0.753	7.301	4.476	2.163	0	5.00	6.0	8.00	30	3.00

Table 6: Summary Statistics for the Theta Estimates

Test	Examinees	Mean	SD	Skew	Min	Q1	Median	Q3	Max	IQR
Full Test	2000	0.155	0.843	2.320	-0.548	-0.338	-0.144	0.223	3.807	0.561

Table 7: Overall Model Fit

Test	Items	Chi-square	df	p	-2LL
Full Test	30	2511.507	360	0.000	60554

Table 8 presents the item control information and item status for each item

Table 8: Item Control and Item Status for All Items

Seq.	Item ID	Key	Options	Domain	Inclusion	Item Type	Status
1	C1	1	2	1	Y	P	Included
2	C2	1	2	1	Y	P	Included
3	C3	1	2	1	Y	P	Included
4	C4	1	2	1	Y	P	Included
5	C5	1	2	1	Y	P	Included
6	C7	1	2	1	Y	P	Included
7	C8	1	2	1	Y	P	Included
8	C9	1	2	1	Y	P	Included
9	C10	1	2	1	Y	P	Included
10	C11	1	2	1	Y	P	Included
11	C12	1	2	1	Y	P	Included
12	C13	1	2	1	Y	P	Included
13	C14	1	2	1	Y	P	Included
14	C15	1	2	1	Y	P	Included
15	C16	1	2	1	Y	P	Included
16	C17	1	2	1	Y	P	Included
17	C18	1	2	1	Y	P	Included
18	C20	1	2	1	Y	P	Included
19	C21	1	2	1	Y	P	Included
20	C22	1	2	1	Y	P	Included
21	C23	1	2	1	Y	P	Included
22	C24	1	2	1	Y	P	Included
23	C25	1	2	1	Y	P	Included
24	C26	1	2	1	Y	P	Included
25	C27	1	2	1	Y	P	Included
26	C28	1	2	1	Y	P	Included
27	C29	1	2	1	Y	P	Included
28	C30	1	2	1	Y	P	Included
29	C31	1	2	1	Y	P	Included
30	C32	1	2	1	Y	P	Included

Table 9 presents the classical statistics, the item parameters, and any flags for each calibrated item.
The K flag indicates that the keyed alternative did not have the highest correlation with total score. The F flag indicates that the item fit statistic (z Resid for dichotomous / chi-square for polytomous) was significant, and the item did not fit the IRT model.The La, Lb, and Lc flags indicate that the a/b/c parameters were lower than the minimum acceptable value.The Ha, Hb, and Hc flags indicate that the a/b/c parameters were higher than the maximum acceptable value

Table 9: Item Parameters for All Calibrated Items

Seq.	Item ID	P	R	a	b	c	Flag(s)
1	C1	0.162	0.334	1.844	2.728	0.141	
2	C2	0.187	0.185	1.599	3.232	0.177	Hb
3	C3	0.220	0.235	0.961	2.907	0.184	
4	C4	0.230	0.290	1.375	2.688	0.206	
5	C5	0.240	0.217	0.827	3.028	0.199	Hb
6	C7	0.274	0.359	1.272	2.075	0.194	
7	C8	0.277	0.321	1.372	2.268	0.214	
8	C9	0.267	0.278	1.240	2.501	0.215	
9	C10	0.263	0.307	0.979	2.396	0.196	
10	C11	0.290	0.194	0.782	3.038	0.242	Hb
11	C12	0.175	0.249	1.156	3.029	0.156	Hb
12	C13	0.147	0.257	1.098	3.241	0.137	Hb
13	C14	0.238	0.133	1.166	3.316	0.219	Hb
14	C15	0.219	0.232	1.385	2.934	0.194	
15	C16	0.212	0.227	1.196	3.017	0.188	Hb
16	C17	0.234	0.237	1.278	2.874	0.204	
17	C18	0.222	0.261	1.297	2.757	0.188	
18	C20	0.264	0.318	1.077	2.258	0.192	
19	C21	0.241	0.241	0.990	2.786	0.198	
20	C22	0.254	0.318	0.945	2.304	0.179	
21	C23	0.237	0.321	0.975	2.434	0.175	
22	C24	0.262	0.373	1.454	2.014	0.183	
23	C25	0.231	0.290	1.027	2.639	0.183	
24	C26	0.237	0.168	0.868	3.361	0.211	Hb
25	C27	0.181	0.183	0.957	3.421	0.167	Hb
26	C28	0.324	0.181	0.520	2.770	0.228	
27	C29	0.273	0.251	0.741	2.563	0.198	
28	C30	0.375	0.319	0.810	1.505	0.211	
29	C31	0.296	0.312	0.968	2.150	0.209	
30	C32	0.270	0.279	0.865	2.518	0.203	

ภาคผนวก จ
ตัวอย่าง Print Out ผลการตรวจสอบความตรงเชิงโครงสร้าง (CFA) ในแบบทดสอบ O-NET ฉบับก่อนและหลังตัดข้อสอบที่ทำหน้าที่ต่างกัน

DATE: 3/12/2018
TIME: 22:52

L I S R E L 8.54

BY

Karl G. J”reskog & Dag S”rbom

This program is published exclusively by
Scientific Software International, Inc.
7383 N. Lincoln Avenue, Suite 100
Lincolnwood, IL 60712, U.S.A.
Phone: (800)247-6113, (847)675-0720, Fax: (847)675-2140
Copyright by Scientific Software International, Inc., 1981-2002
Use of this program is subject to the terms specified in the
Universal Copyright Convention.
Website: www.ssicentral.com

The following lines were read from file D:\2NDCFA\CFA\04 math\04 math..LS8:

```
TI
DA NI=32 NO=2000 NG=1 MA=CM
RA FI='D:\2NDCFA\CFA\04 math\04 MATH.psf'
MO NY=32 NK=1 NE=4 BE=FU GA=FI TE=FU
LE
number mesur alger analyze
LK
MATH
FR LY(2,1) LY(3,1) LY(4,1) LY(5,1) LY(6,1) LY(7,1)
FR LY(9,2) LY(10,2) LY(11,2)
FR LY(13,3) LY(14,3) LY(15,3) LY(16,3) LY(17,3) LY(18,3)
FR LY(19,3) LY(20,3) LY(21,3) LY(22,3) LY(23,3) LY(24,3) LY(25,3) LY(26,3)
FR LY(28,4) LY(29,4) LY(30,4) LY(31,4) LY(32,4)
FR GA(1,1) GA(2,1) GA(3,1) GA(4,1)
FI LY(1,1) LY(8,2) LY(12,3) LY(27,4)
FR TE(12,8) TE(12,1) TE(8,1) TE(18,17) TE(30,1) TE(27,12) TE(30,23)
FR TE(27,8) TE(30,8) TE(30,26) TE(28,18) TE(29,13) TE(21,19) TE(21,10)
FR TE(23,11) TE(23,1) TE(5,1) TE(30,11) TE(16,12) TE(27,13) TE(23,12)
FR TE(23,8) TE(22,5) TE(19,17) TE(9,8) TE(24,12) TE(8,4) TE(21,13)
FR TE(19,13) TE(12,6) TE(32,27) TE(10,8) TE(31,8) TE(31,1) TE(7,3)
FR TE(31,12) TE(29,11) TE(12,10) TE(12,5) TE(22,12) TE(12,2) TE(8,6)
FR TE(25,8) TE(10,1) TE(21,1) TE(32,8) TE(14,8) TE(4,1) TE(7,1) TE(32,12)
FR TE(29,8) TE(28,11) TE(29,12) TE(12,9) TE(12,7) TE(17,12) TE(19,12)
FR TE(12,4) TE(15,12) TE(25,12) TE(20,12) TE(28,12) TE(26,12) TE(21,12)
PD
OU AD=off mi

TI
```

Number of Input Variables 32
Number of Y - Variables 32
Number of X - Variables 0
Number of ETA - Variables 4
Number of KSI - Variables 1
Number of Observations 2000

TI

Covariance Matrix

	C1	C2	C3	C4	C5	C6
C1	0.17					
C2	0.01	0.18				
C3	0.01	0.01	0.18			
C4	0.02	0.02	0.02	0.14		
C5	0.03	0.02	0.03	0.02	0.20	

C6	0.02	0.03	0.02	0.03	0.03	0.20
C7	0.02	0.02	0.03	0.02	0.03	0.03
C8	0.02	0.02	0.02	0.02	0.02	0.03
C9	0.02	0.03	0.02	0.02	0.04	0.03
C10	0.02	0.02	0.01	0.02	0.02	0.02
C11	0.01	0.01	0.00	0.01	0.01	0.01
C12	0.01	0.02	0.01	0.02	0.03	0.03
C13	0.01	0.01	0.00	0.01	0.02	0.01
C14	0.02	0.02	0.02	0.02	0.03	0.03
C15	0.01	0.02	0.01	0.01	0.02	0.01
C16	0.01	0.01	0.01	0.02	0.02	0.02
C17	0.01	0.02	0.02	0.01	0.02	0.02
C18	0.01	0.02	0.01	0.01	0.01	0.02
C19	0.01	0.02	0.02	0.02	0.03	0.01
C20	0.01	0.02	0.02	0.02	0.02	0.02
C21	0.02	0.01	0.01	0.01	0.02	0.01
C22	0.01	0.02	0.02	0.02	0.02	0.03
C23	0.03	0.03	0.02	0.02	0.04	0.04
C24	0.02	0.02	0.03	0.02	0.03	0.02
C25	0.01	0.02	0.01	0.02	0.03	0.02
C26	0.02	0.02	0.02	0.02	0.03	0.03
C27	0.01	0.01	0.00	0.01	0.01	0.01
C28	0.01	0.01	0.01	0.01	0.02	0.02
C29	0.01	0.02	0.01	0.02	0.02	0.02
C30	0.03	0.03	0.02	0.02	0.04	0.02
C31	0.02	0.02	0.02	0.02	0.03	0.03
C32	0.02	0.02	0.01	0.01	0.03	0.03

Covariance Matrix

	C7	C8	C9	C10	C11	C12
C7	0.20					
C8	0.02	0.18				
C9	0.02	0.03	0.19			
C10	0.02	0.03	0.03	0.18		
C11	0.01	0.02	0.01	0.01	0.18	
C12	0.02	0.02	0.03	0.02	0.01	0.14
C13	0.01	0.01	0.01	0.01	0.01	0.01
C14	0.01	0.03	0.03	0.02	0.01	0.02
C15	0.02	0.02	0.02	0.01	0.01	0.02
C16	0.02	0.02	0.02	0.01	0.01	0.02
C17	0.02	0.01	0.03	0.02	0.01	0.02
C18	0.01	0.01	0.01	0.01	0.00	0.01
C19	0.02	0.01	0.02	0.01	0.00	0.02
C20	0.02	0.02	0.02	0.01	0.01	0.02
C21	0.01	0.02	0.03	0.03	0.01	0.01
C22	0.02	0.02	0.03	0.01	0.01	0.02
C23	0.03	0.03	0.04	0.03	0.03	0.03
C24	0.02	0.02	0.04	0.02	0.01	0.03
C25	0.02	0.02	0.02	0.02	0.01	0.02
C26	0.02	0.02	0.04	0.02	0.01	0.02
C27	0.02	0.01	0.02	0.01	0.01	0.01
C28	0.02	0.02	0.02	0.02	0.02	0.02
C29	0.01	0.02	0.02	0.01	0.02	0.02
C30	0.02	0.04	0.04	0.03	0.03	0.02
C31	0.02	0.03	0.03	0.02	0.02	0.03
C32	0.02	0.02	0.03	0.02	0.01	0.02

Covariance Matrix

	C13	C14	C15	C16	C17	C18
C13	0.15					
C14	0.02	0.19				
C15	0.01	0.02	0.21			
C16	0.01	0.02	0.01	0.14		
C17	0.01	0.02	0.01	0.01	0.13	
C18	0.01	0.01	0.01	0.01	0.02	0.18
C19	0.02	0.02	0.01	0.01	0.02	0.01
C20	0.02	0.02	0.00	0.01	0.01	0.01
C21	0.02	0.02	0.01	0.01	0.01	0.01
C22	0.00	0.02	0.01	0.02	0.01	0.01

C23	0.02	0.03	0.02	0.02	0.02	0.01
C24	0.01	0.03	0.02	0.02	0.01	0.01
C25	0.01	0.02	0.01	0.01	0.01	0.01
C26	0.01	0.02	0.02	0.02	0.01	0.01
C27	0.01	0.01	0.01	0.01	0.01	0.01
C28	0.01	0.02	0.00	0.01	0.00	-0.01
C29	0.00	0.02	0.02	0.01	0.01	0.01
C30	0.01	0.03	0.01	0.01	0.02	0.00
C31	0.01	0.02	0.01	0.01	0.01	0.01
C32	0.01	0.02	0.02	0.01	0.02	0.02

Covariance Matrix

	C19	C20	C21	C22	C23	C24
C19	0.17					
C20	0.01	0.17				
C21	0.02	0.02	0.18			
C22	0.02	0.02	0.01	0.17		
C23	0.02	0.02	0.02	0.03	0.16	
C24	0.02	0.02	0.02	0.02	0.04	0.19
C25	0.01	0.01	0.01	0.02	0.02	0.01
C26	0.01	0.02	0.01	0.02	0.03	0.03
C27	0.01	0.02	0.01	0.01	0.02	0.02
C28	0.00	0.01	0.01	0.01	0.03	0.02
C29	0.01	0.01	0.02	0.02	0.03	0.02
C30	0.01	0.01	0.01	0.02	0.05	0.03
C31	0.02	0.02	0.02	0.02	0.03	0.03
C32	0.02	0.01	0.02	0.01	0.03	0.02

Covariance Matrix

	C25	C26	C27	C28	C29	C30
C25	0.18					
C26	0.02	0.19				
C27	0.01	0.01	0.15			
C28	0.02	0.02	0.00	0.22		
C29	0.02	0.02	0.00	0.02	0.20	
C30	0.02	0.04	0.01	0.02	0.03	0.23
C31	0.02	0.03	0.02	0.02	0.03	0.04
C32	0.01	0.02	0.02	0.02	0.02	0.03

Covariance Matrix

	C31	C32
C31	0.21	
C32	0.03	0.20

TI

Parameter Specifications

LAMBDA-Y

	number	mesur	alger	analyze
C1	0	0	0	0
C2	0	0	0	0
C3	1	0	0	0
C4	2	0	0	0
C5	3	0	0	0
C6	4	0	0	0
C7	5	0	0	0
C8	0	0	0	0
C9	0	0	0	0
C10	0	6	0	0
C11	0	7	0	0
C12	0	0	0	0
C13	0	0	0	0
C14	0	0	8	0

C15	0	0	9	0
C16	0	0	10	0
C17	0	0	11	0
C18	0	0	12	0
C19	0	0	13	0
C20	0	0	14	0
C21	0	0	15	0
C22	0	0	16	0
C23	0	0	17	0
C24	0	0	18	0
C25	0	0	19	0
C26	0	0	20	0
C27	0	0	0	0
C28	0	0	0	0
C29	0	0	0	21
C30	0	0	0	22
C31	0	0	0	23
C32	0	0	0	24

GAMMA

	MATH
number	25
mesur	26
alger	27
analyze	28

PSI

number	mesur	alger	analyze
29	30	31	32

THETA-EPS

	C1	C2	C3	C4	C5	C6
C1	33					
C2	0	34				
C3	0	0	35			
C4	36	0	0	37		
C5	38	0	0	0	39	
C6	0	0	0	0	0	40
C7	41	0	42	0	0	0
C8	44	0	0	45	0	46
C9	0	0	0	0	0	0
C10	50	0	0	0	0	0
C11	0	0	0	0	0	0
C12	54	55	0	56	57	58
C13	0	0	0	0	0	0
C14	0	0	0	0	0	0
C15	0	0	0	0	0	0
C16	0	0	0	0	0	0
C17	0	0	0	0	0	0
C18	0	0	0	0	0	0
C19	0	0	0	0	0	0
C20	0	0	0	0	0	0
C21	81	0	0	0	0	0
C22	0	0	0	0	87	0
C23	90	0	0	0	0	0
C24	0	0	0	0	0	0
C25	0	0	0	0	0	0
C26	0	0	0	0	0	0
C27	0	0	0	0	0	0
C28	0	0	0	0	0	0
C29	0	0	0	0	0	0
C30	115	0	0	0	0	0
C31	121	0	0	0	0	0
C32	0	0	0	0	0	0

THETA-EPS

	C7	C8	C9	C10	C11	C12
C7	43					
C8	0	47				
C9	0	48	49			
C10	0	51	0	52		
C11	0	0	0	0	53	
C12	59	60	61	62	0	63
C13	0	0	0	0	0	0
C14	0	65	0	0	0	0
C15	0	0	0	0	0	67
C16	0	0	0	0	0	69
C17	0	0	0	0	0	71
C18	0	0	0	0	0	0
C19	0	0	0	0	0	75
C20	0	0	0	0	0	79
C21	0	0	0	82	0	83
C22	0	0	0	0	0	88
C23	0	91	0	0	92	93
C24	0	0	0	0	0	95
C25	0	97	0	0	0	98
C26	0	0	0	0	0	100
C27	0	102	0	0	0	103
C28	0	0	0	0	106	107
C29	0	110	0	0	111	112
C30	0	116	0	0	117	0
C31	0	122	0	0	0	123
C32	0	125	0	0	0	126

THETA-EPS

	C13	C14	C15	C16	C17	C18
C13	64					
C14	0	66				
C15	0	0	68			
C16	0	0	0	70		
C17	0	0	0	0	72	
C18	0	0	0	0	73	74
C19	76	0	0	0	77	0
C20	0	0	0	0	0	0
C21	84	0	0	0	0	0
C22	0	0	0	0	0	0
C23	0	0	0	0	0	0
C24	0	0	0	0	0	0
C25	0	0	0	0	0	0
C26	0	0	0	0	0	0
C27	104	0	0	0	0	0
C28	0	0	0	0	0	108
C29	113	0	0	0	0	0
C30	0	0	0	0	0	0
C31	0	0	0	0	0	0
C32	0	0	0	0	0	0

THETA-EPS

	C19	C20	C21	C22	C23	C24
C19	78					
C20	0	80				
C21	85	0	86			
C22	0	0	0	89		
C23	0	0	0	0	94	
C24	0	0	0	0	0	96
C25	0	0	0	0	0	0
C26	0	0	0	0	0	0
C27	0	0	0	0	0	0
C28	0	0	0	0	0	0
C29	0	0	0	0	0	0
C30	0	0	0	0	118	0
C31	0	0	0	0	0	0
C32	0	0	0	0	0	0

THETA-EPS

	C25	C26	C27	C28	C29	C30
C25	99					
C26	0	101				
C27	0	0	105			
C28	0	0	0	109		
C29	0	0	0	0	114	
C30	0	119	0	0	0	120
C31	0	0	0	0	0	0
C32	0	0	127	0	0	0

THETA-EPS

	C31	C32
C31	124	
C32	0	128

TI

Number of Iterations = 55

LISREL Estimates (Maximum Likelihood)

LAMBDA-Y

	number	mesur	alger	analyze
C1	- -	- -	- -	- -
C2	0.14	- -	- -	- -
C3	0.11 (0.01) 8.08	- -	- -	- -
C4	0.12 (0.01) 9.24	- -	- -	- -
C5	0.19 (0.02) 10.36	- -	- -	- -
C6	0.16 (0.02) 9.77	- -	- -	- -
C7	0.13 (0.02) 8.71	- -	- -	- -
C8	- -	- -	- -	- -
C9	- -	0.20	- -	- -
C10	- -	0.13 (0.01) 9.56	- -	- -
C11	- -	0.07 (0.01) 5.43	- -	- -
C12	- -	- -	- -	- -
C13	- -	- -	0.07	- -
C14	- -	- -	0.15	- -

			(0.02)	
			6.22	
C15	- -	- -	0.09	- -
			(0.02)	
			5.24	
C16	- -	- -	0.10	- -
			(0.02)	
			5.83	
C17	- -	- -	0.10	- -
			(0.02)	
			5.88	
C18	- -	- -	0.06	- -
			(0.01)	
			4.40	
C19	- -	- -	0.10	- -
			(0.02)	
			5.79	
C20	- -	- -	0.11	- -
			(0.02)	
			5.76	
C21	- -	- -	0.09	- -
			(0.02)	
			5.61	
C22	- -	- -	0.13	- -
			(0.02)	
			6.08	
C23	- -	- -	0.18	- -
			(0.03)	
			6.55	
C24	- -	- -	0.16	- -
			(0.02)	
			6.25	
C25	- -	- -	0.11	- -
			(0.02)	
			5.75	
C26	- -	- -	0.16	- -
			(0.03)	
			6.31	
C27	- -	- -	- -	- -
C28	- -	- -	- -	0.11
C29	- -	- -	- -	0.15
				(0.02)
				6.79
C30	- -	- -	- -	0.17
				(0.02)
				6.99
C31	- -	- -	- -	0.17
				(0.02)
				7.10
C32	- -	- -	- -	0.14
				(0.02)
				6.67

GAMMA

	MATH
number	0.97
	(0.08)
	12.41
mesur	0.90
	(0.05)
	16.65
alger	1.04
	(0.15)
	6.85
analyze	0.78
	(0.10)
	8.07

Covariance Matrix of ETA and KSI

	number	mesur	alger	analyze	MATH
number	1.00				
mesur	0.87	1.00			
alger	1.01	0.94	1.00		
analyze	0.75	0.70	0.81	1.00	
MATH	0.97	0.90	1.04	0.78	1.00

PHI

MATH
1.00

PSI
Note: This matrix is diagonal.

number	mesur	alger	analyze
0.07	0.19	-0.09	0.40
(0.07)	(0.13)	(0.06)	(0.12)
1.01	1.44	-1.47	3.42

Squared Multiple Correlations for Structural Equations

number	mesur	alger	analyze
0.93	0.81	1.09	0.60

Squared Multiple Correlations for Reduced Form

number	mesur	alger	analyze
0.93	0.81	1.09	0.60

THETA-EPS

	C1	C2	C3	C4	C5	C6
C1	0.17					
	(0.01)					
	31.63					
C2	- -	0.16				
		(0.01)				
		29.97				
C3	- -	- -	0.17			
			(0.01)			

			30.81			
C4	0.01 (0.00) 2.81	- -	- -	0.12 (0.00) 29.88		
C5	0.01 (0.00) 3.93	- -	- -	- -	0.16 (0.01) 28.23	
C6	- -	- -	- -	- -	- -	0.17 (0.01) 29.35
C7	0.01 (0.00) 2.73	- -	0.01 (0.00) 2.62	- -	- -	- -
C8	0.01 (0.00) 3.01	- -	- -	0.01 (0.00) 3.87	- -	0.01 (0.00) 2.90
C9	- -	- -	- -	- -	- -	- -
C10	0.01 (0.00) 2.83	- -	- -	- -	- -	- -
C11	- -	- -	- -	- -	- -	- -
C12	0.01 (0.00) 2.19	0.02 (0.00) 4.53	- -	0.01 (0.00) 3.83	0.02 (0.00) 5.44	0.02 (0.00) 5.67
C13	- -	- -	- -	- -	- -	- -
C14	- -	- -	- -	- -	- -	- -
C15	- -	- -	- -	- -	- -	- -
C16	- -	- -	- -	- -	- -	- -
C17	- -	- -	- -	- -	- -	- -
C18	- -	- -	- -	- -	- -	- -
C19	- -	- -	- -	- -	- -	- -
C20	- -	- -	- -	- -	- -	- -
C21	0.01 (0.00) 2.93	- -	- -	- -	- -	- -
C22	- -	- -	- -	- -	-0.01 (0.00) -2.78	- -
C23	0.02 (0.00) 4.83	- -	- -	- -	- -	- -
C24	- -	- -	- -	- -	- -	- -
C25	- -	- -	- -	- -	- -	- -
C26	- -	- -	- -	- -	- -	- -
C27	- -	- -	- -	- -	- -	- -
C28	- -	- -	- -	- -	- -	- -
C29	- -	- -	- -	- -	- -	- -

C30	0.02 (0.00) 4.80	- -	- -	- -	- -	- -
C31	0.01 (0.00) 3.17	- -	- -	- -	- -	- -
C32	- -	- -	- -	- -	- -	- -

THETA-EPS

	C7	C8	C9	C10	C11	C12
C7	0.18 (0.01) 30.26					
C8	- -	0.18 (0.01) 31.64				
C9	- -	0.02 (0.00) 4.32	0.15 (0.01) 20.97			
C10	- -	0.01 (0.00) 3.64	- -	0.16 (0.01) 28.61		
C11	- -	- -	- -	- -	0.18 (0.01) 31.15	
C12	0.01 (0.00) 3.88	0.01 (0.00) 3.34	0.02 (0.00) 5.14	0.02 (0.00) 5.15	- -	0.13 (0.00) 31.66
C13	- -	- -	- -	- -	- -	- -
C14	- -	0.01 (0.00) 2.80	- -	- -	- -	- -
C15	- -	- -	- -	- -	- -	0.01 (0.00) 3.29
C16	- -	- -	- -	- -	- -	0.02 (0.00) 5.59
C17	- -	- -	- -	- -	- -	0.01 (0.00) 4.61
C18	- -	- -	- -	- -	- -	- -
C19	- -	- -	- -	- -	- -	0.01 (0.00) 3.83
C20	- -	- -	- -	- -	- -	0.01 (0.00) 3.49
C21	- -	- -	- -	0.01 (0.00) 3.06	- -	0.01 (0.00) 2.94
C22	- -	- -	- -	- -	- -	0.02

						(0.00) 4.64
C23	- -	0.01 (0.00) 4.24	- -	- -	0.01 (0.00) 3.39	0.02 (0.00) 6.97
C24	- -	- -	- -	- -	- -	0.02 (0.00) 6.15
C25	- -	0.01 (0.00) 3.12	- -	- -	- -	0.01 (0.00) 3.55
C26	- -	- -	- -	- -	- -	0.01 (0.00) 3.44
C27	- -	0.01 (0.00) 2.70	- -	- -	- -	0.01 (0.00) 1.90
C28	- -	- -	- -	- -	0.01 (0.00) 2.55	0.01 (0.00) 3.72
C29	- -	0.01 (0.00) 3.04	- -	- -	0.01 (0.00) 2.39	0.02 (0.00) 4.48
C30	- -	0.02 (0.00) 4.59	- -	- -	0.01 (0.00) 3.34	- -
C31	- -	0.02 (0.00) 3.86	- -	- -	- -	0.02 (0.00) 5.22
C32	- -	0.01 (0.00) 3.30	- -	- -	- -	0.01 (0.00) 3.99

THETA-EPS

	C13	C14	C15	C16	C17	C18
C13	0.15 (0.00) 31.34					
C14	- -	0.17 (0.01) 30.47				
C15	- -	- -	0.20 (0.01) 31.16			
C16	- -	- -	- -	0.13 (0.00) 30.77		
C17	- -	- -	- -	- -	0.12 (0.00) 30.70	
C18	- -	- -	- -	- -	0.01 (0.00) 3.20	0.18 (0.01) 31.43
C19	0.01	- -	- -	- -	0.01	- -

	(0.00)				(0.00)	
	2.78				3.06	
C20	- -	- -	- -	- -	- -	- -
C21	0.01	- -	- -	- -	- -	- -
	(0.00)					
	2.76					
C22	- -	- -	- -	- -	- -	- -
C23	- -	- -	- -	- -	- -	- -
C24	- -	- -	- -	- -	- -	- -
C25	- -	- -	- -	- -	- -	- -
C26	- -	- -	- -	- -	- -	- -
C27	0.01	- -	- -	- -	- -	- -
	(0.00)					
	2.76					
C28	- -	- -	- -	- -	- -	-0.01
						(0.00)
						-3.00
C29	-0.01	- -	- -	- -	- -	- -
	(0.00)					
	-2.87					
C30	- -	- -	- -	- -	- -	- -
C31	- -	- -	- -	- -	- -	- -
C32	- -	- -	- -	- -	- -	- -

THETA-EPS

	C19	C20	C21	C22	C23	C24
C19	0.16					
	(0.01)					
	30.92					
C20	- -	0.16				
		(0.01)				
		30.84				
C21	0.01	- -	0.17			
	(0.00)		(0.01)			
	3.59		31.02			
C22	- -	- -	- -	0.15		
				(0.01)		
				30.28		
C23	- -	- -	- -	- -	0.12	
					(0.00)	
					28.25	
C24	- -	- -	- -	- -	- -	0.17
						(0.01)
						30.04
C25	- -	- -	- -	- -	- -	- -
C26	- -	- -	- -	- -	- -	- -
C27	- -	- -	- -	- -	- -	- -
C28	- -	- -	- -	- -	- -	- -

C29	- -	- -	- -	- -	- -	- -
C30	- -	- -	- -	- -	0.02 (0.00) 4.25	- -
C31	- -	- -	- -	- -	- -	- -
C32	- -	- -	- -	- -	- -	- -

THETA-EPS

	C25	C26	C27	C28	C29	C30
C25	0.17 (0.01) 30.81					
C26	- -	0.16 (0.01) 29.82				
C27	- -	- -	0.15 (0.00) 31.62			
C28	- -	- -	- -	0.21 (0.01) 30.18		
C29	- -	- -	- -	- -	0.17 (0.01) 28.57	
C30	- -	0.01 (0.00) 3.13	- -	- -	- -	0.20 (0.01) 28.14
C31	- -	- -	- -	- -	- -	- -
C32	- -	- -	0.01 (0.00) 3.11	- -	- -	- -

THETA-EPS

	C31	C32
C31	0.17 (0.01) 27.19	
C32	- -	0.17 (0.01) 28.93

Squared Multiple Correlations for Y - Variables

C1	C2	C3	C4	C5	C6
- -	0.11	0.07	0.11	0.18	0.14

Squared Multiple Correlations for Y - Variables

C7	C8	C9	C10	C11	C12
0.09	- -	0.21	0.10	0.02	- -

Squared Multiple Correlations for Y - Variables

C13	C14	C15	C16	C17	C18
0.03	0.11	0.04	0.07	0.08	0.02

Squared Multiple Correlations for Y - Variables

C19	C20	C21	C22	C23	C24
0.06	0.07	0.05	0.10	0.22	0.13

Squared Multiple Correlations for Y - Variables

C25	C26	C27	C28	C29	C30
0.07	0.14	- -	0.06	0.11	0.13

Squared Multiple Correlations for Y - Variables

C31	C32
0.15	0.10

Goodness of Fit Statistics

Degrees of Freedom = 400
Minimum Fit Function Chi-Square = 735.36 (P = 0.0)
Normal Theory Weighted Least Squares Chi-Square = 742.61 (P = 0.0)
Estimated Non-centrality Parameter (NCP) = 342.61
90 Percent Confidence Interval for NCP = (269.86 ; 423.18)

Minimum Fit Function Value = 0.37
Population Discrepancy Function Value (F0) = 0.17
90 Percent Confidence Interval for F0 = (0.13 ; 0.21)
Root Mean Square Error of Approximation (RMSEA) = 0.021
90 Percent Confidence Interval for RMSEA = (0.018 ; 0.023)
P-Value for Test of Close Fit (RMSEA < 0.05) = 1.00

Expected Cross-Validation Index (ECVI) = 0.50
90 Percent Confidence Interval for ECVI = (0.46 ; 0.54)
ECVI for Saturated Model = 0.53
ECVI for Independence Model = 5.59

Chi-Square for Independence Model with 496 Degrees of Freedom = 11103.98
Independence AIC = 11167.98
Model AIC = 998.61
Saturated AIC = 1056.00
Independence CAIC = 11379.21
Model CAIC = 1843.52
Saturated CAIC = 4541.28

Normed Fit Index (NFI) = 0.93
Non-Normed Fit Index (NNFI) = 0.96
Parsimony Normed Fit Index (PNFI) = 0.75
Comparative Fit Index (CFI) = 0.97
Incremental Fit Index (IFI) = 0.97
Relative Fit Index (RFI) = 0.92

Critical N (CN) = 1275.18

Root Mean Square Residual (RMR) = 0.0069
Standardized RMR = 0.040
Goodness of Fit Index (GFI) = 0.98
Adjusted Goodness of Fit Index (AGFI) = 0.97
Parsimony Goodness of Fit Index (PGFI) = 0.74

TI

Modification Indices and Expected Change

Modification Indices for LAMBDA-Y

	number	mesur	alger	analyze
C1	64.51	64.01	64.61	62.80
C2	- -	0.84	1.10	0.34
C3	- -	3.64	9.89	3.38
C4	- -	0.26	0.31	0.05
C5	- -	0.16	0.01	1.03
C6	- -	1.68	0.67	3.32
C7	- -	1.18	0.45	0.01
C8	113.63	115.33	113.98	113.37
C9	2.34	- -	1.24	0.10
C10	2.09	- -	0.24	1.61
C11	0.22	- -	0.28	4.03
C12	37.68	38.70	38.13	35.21
C13	2.49	0.06	- -	0.01
C14	0.01	0.68	- -	0.45
C15	0.84	0.03	- -	0.09
C16	4.61	0.32	- -	4.75
C17	1.91	3.73	- -	3.68
C18	3.70	7.62	- -	0.20
C19	1.52	4.34	- -	0.84
C20	10.62	2.92	- -	4.50
C21	1.35	1.62	- -	0.00
C22	0.99	0.67	- -	4.44
C23	4.71	0.61	- -	2.11
C24	0.67	0.88	- -	0.52
C25	2.05	0.49	- -	0.13
C26	0.11	2.23	- -	0.02
C27	56.92	60.33	56.87	46.12
C28	0.41	0.36	0.31	- -
C29	0.70	1.96	1.03	- -
C30	0.10	1.44	0.59	- -
C31	0.41	0.10	0.18	- -
C32	1.19	1.56	1.74	- -

Expected Change for LAMBDA-Y

	number	mesur	alger	analyze
C1	0.09	0.10	0.09	0.11
C2	- -	0.07	0.18	0.02
C3	- -	-0.15	-0.53	-0.06
C4	- -	-0.03	-0.08	-0.01
C5	- -	-0.03	-0.02	0.04
C6	- -	-0.11	0.15	0.07
C7	- -	-0.09	-0.12	0.00
C8	0.13	0.14	0.13	0.16
C9	-0.29	- -	-0.18	0.01
C10	-0.20	- -	-0.06	0.04
C11	0.04	- -	0.05	0.09
C12	0.10	0.11	0.10	0.10
C13	0.25	-0.02	- -	0.00
C14	0.02	0.08	- -	0.03
C15	0.17	0.02	- -	-0.01
C16	0.33	0.05	- -	-0.07
C17	-0.20	0.15	- -	-0.06
C18	0.33	-0.25	- -	-0.02
C19	0.21	-0.18	- -	-0.03
C20	0.54	-0.15	- -	-0.08
C21	-0.21	0.13	- -	0.00
C22	0.17	-0.08	- -	-0.08
C23	-0.38	0.07	- -	0.05
C24	-0.15	0.09	- -	0.03
C25	-0.25	-0.06	- -	0.01
C26	-0.06	0.14	- -	0.01
C27	0.08	0.08	0.08	0.08
C28	0.02	0.02	0.02	- -
C29	-0.03	-0.05	-0.04	- -
C30	0.01	0.05	0.04	- -
C31	-0.03	-0.01	-0.02	- -
C32	0.04	0.05	0.05	- -

Modification Indices for BETA

	number	mesur	alger	analyze
number	- -	6.54	2.29	1.53
mesur	6.54	- -	1.53	2.29
alger	2.29	1.53	- -	6.54
analyze	1.53	2.29	6.54	- -

Expected Change for BETA

	number	mesur	alger	analyze
number	- -	-0.90	-1.53	0.18
mesur	-2.50	- -	-1.00	0.23
alger	1.98	0.47	- -	-0.39
analyze	1.03	0.50	1.80	- -

No Non-Zero Modification Indices for GAMMA

No Non-Zero Modification Indices for PHI

Modification Indices for PSI

	number	mesur	alger	analyze
number	- -			
mesur	6.54	- -		
alger	2.29	1.53	- -	
analyze	1.53	2.29	6.54	- -

Expected Change for PSI

	number	mesur	alger	analyze
number	- -			
mesur	-0.17	- -		
alger	0.13	0.09	- -	
analyze	0.07	0.09	-0.16	- -

Modification Indices for THETA-EPS

	C1	C2	C3	C4	C5	C6
C1	- -					
C2	0.74	- -				
C3	0.06	0.80	- -			
C4	- -	0.78	0.54	- -		
C5	- -	1.11	1.79	1.51	- -	
C6	1.90	1.81	1.81	2.01	0.13	- -
C7	- -	0.51	- -	0.35	0.11	0.13
C8	- -	0.88	4.31	- -	2.40	- -
C9	2.49	0.03	3.31	0.56	0.27	0.12
C10	- -	3.03	0.59	0.06	2.36	2.15
C11	0.08	0.01	0.36	0.00	1.72	0.00
C12	- -	- -	6.06	- -	- -	- -
C13	3.47	0.69	2.38	1.47	1.06	0.25
C14	3.14	1.06	0.40	0.16	0.97	0.08
C15	3.30	2.10	0.44	1.13	0.74	1.79
C16	3.68	1.58	0.21	2.44	0.28	0.66
C17	0.90	0.97	2.16	0.00	0.09	0.03
C18	1.12	4.92	0.16	0.32	0.45	3.01
C19	0.00	0.91	1.76	1.15	3.68	1.75
C20	0.00	0.04	2.10	1.06	0.76	0.09
C21	- -	0.15	0.37	0.30	0.87	0.93
C22	0.00	0.01	0.60	4.21	- -	2.63
C23	- -	0.27	0.10	4.50	0.02	0.11
C24	5.99	1.84	3.06	0.02	0.21	4.26
C25	0.77	0.04	0.73	4.83	1.02	0.00
C26	2.19	0.65	0.25	1.54	0.15	0.01
C27	0.78	2.75	2.97	4.91	2.32	0.00
C28	0.64	1.11	0.02	0.36	0.53	0.46
C29	1.15	1.34	1.72	3.60	0.74	0.26

C30	- -	1.25	0.08	0.83	1.55	0.80
C31	- -	0.32	0.11	0.00	0.17	2.84
C32	2.18	0.01	5.51	0.83	1.78	3.64

Modification Indices for THETA-EPS

	C7	C8	C9	C10	C11	C12
C7	- -					
C8	6.03	- -				
C9	0.31	- -	- -			
C10	0.20	- -	1.62	- -		
C11	0.74	3.58	0.68	0.62	- -	
C12	- -	- -	- -	- -	6.81	- -
C13	0.73	3.62	3.88	0.25	0.13	1.63
C14	5.95	- -	0.15	0.55	0.24	8.09
C15	0.37	3.23	1.48	0.00	3.75	- -
C16	0.90	4.46	0.18	0.67	2.66	- -
C17	0.08	0.98	4.68	3.45	0.04	- -
C18	0.80	0.97	3.08	1.46	1.99	1.97
C19	3.01	2.68	0.21	2.97	0.14	- -
C20	0.43	4.78	2.66	3.36	0.15	- -
C21	0.82	3.53	1.90	- -	0.84	- -
C22	1.35	7.56	0.14	1.87	1.60	- -
C23	0.09	- -	1.15	0.29	- -	- -
C24	0.03	4.47	0.76	0.12	0.72	- -
C25	0.00	- -	1.14	4.35	1.78	- -
C26	1.99	1.05	2.18	0.08	0.81	- -
C27	4.43	- -	6.31	1.30	0.07	- -
C28	1.80	2.56	0.54	0.74	- -	- -
C29	0.67	- -	1.02	0.72	- -	- -
C30	0.07	- -	1.54	3.88	- -	6.94
C31	1.19	- -	0.33	0.35	3.76	- -
C32	2.07	- -	0.06	2.39	0.06	- -

Modification Indices for THETA-EPS

	C13	C14	C15	C16	C17	C18
C13	- -					
C14	1.65	- -				
C15	4.35	0.18	- -			
C16	1.16	0.00	0.78	- -		
C17	1.05	0.22	0.05	1.44	- -	
C18	1.85	0.28	0.21	0.15	- -	- -
C19	- -	0.05	0.01	0.86	- -	0.86
C20	3.68	0.48	5.31	0.01	0.54	0.01
C21	- -	0.00	0.82	0.63	0.10	0.20
C22	2.26	0.08	0.68	0.24	0.07	0.25
C23	0.92	0.25	0.01	0.03	0.01	2.01
C24	0.01	1.51	0.74	0.01	3.80	0.13
C25	0.04	0.05	0.01	0.53	3.77	0.13
C26	0.20	0.15	1.19	0.36	1.20	0.09
C27	- -	1.93	0.08	1.12	3.91	0.21
C28	0.03	0.37	1.75	0.14	2.41	- -
C29	- -	0.25	2.00	2.06	2.57	0.00
C30	0.03	2.04	0.65	1.67	0.39	4.61
C31	1.88	0.25	2.77	0.73	2.26	0.07
C32	1.10	1.80	2.02	0.26	1.08	4.13

Modification Indices for THETA-EPS

	C19	C20	C21	C22	C23	C24
C19	- -					
C20	0.00	- -				
C21	- -	3.75	- -			
C22	0.65	0.10	1.69	- -		
C23	2.36	0.01	0.43	0.19	- -	
C24	0.01	0.10	0.07	1.77	0.35	- -
C25	2.18	0.05	0.52	0.40	2.48	1.66
C26	0.35	0.68	0.41	0.00	0.12	3.54
C27	0.06	6.30	0.86	0.00	3.63	3.39

C28	1.03	0.13	0.39	0.54	5.19	2.20
C29	2.99	0.26	1.02	0.25	0.04	0.01
C30	0.74	3.94	2.29	2.18	- -	0.83
C31	0.00	0.23	0.07	0.31	1.15	0.46
C32	4.79	3.82	1.08	1.24	0.00	2.82

Modification Indices for THETA-EPS

	C25	C26	C27	C28	C29	C30
C25	- -					
C26	0.44	- -				
C27	0.21	0.01	- -			
C28	0.57	0.45	0.26	- -		
C29	1.09	0.00	1.07	0.80	- -	
C30	0.78	- -	0.85	0.08	1.17	- -
C31	0.06	0.04	3.84	0.02	1.51	0.16
C32	1.86	0.22	- -	0.03	0.26	0.62

Modification Indices for THETA-EPS

	C31	C32
C31	- -	
C32	0.01	- -

Expected Change for THETA-EPS

	C1	C2	C3	C4	C5	C6
C1	- -					
C2	0.00	- -				
C3	0.00	0.00	- -			
C4	- -	0.00	0.00	- -		
C5	- -	0.00	0.01	0.00	- -	
C6	0.01	0.01	-0.01	0.00	0.00	- -
C7	- -	0.00	- -	0.00	0.00	0.00
C8	- -	0.00	0.01	- -	0.01	- -
C9	0.01	0.00	-0.01	0.00	0.00	0.00
C10	- -	0.01	0.00	0.00	-0.01	-0.01
C11	0.00	0.00	0.00	0.00	-0.01	0.00
C12	- -	- -	0.01	- -	- -	- -
C13	0.01	0.00	-0.01	0.00	0.00	0.00
C14	0.01	0.00	0.00	0.00	0.00	0.00
C15	0.01	0.01	0.00	0.00	0.00	-0.01
C16	0.01	0.00	0.00	0.00	0.00	0.00
C17	0.00	0.00	0.00	0.00	0.00	0.00
C18	0.00	0.01	0.00	0.00	0.00	0.01
C19	0.00	0.00	0.00	0.00	0.01	-0.01
C20	0.00	0.00	0.01	0.00	0.00	0.00
C21	- -	0.00	0.00	0.00	0.00	0.00
C22	0.00	0.00	0.00	0.01	- -	0.01
C23	- -	0.00	0.00	-0.01	0.00	0.00
C24	0.01	-0.01	0.01	0.00	0.00	-0.01
C25	0.00	0.00	0.00	0.01	0.00	0.00
C26	0.01	0.00	0.00	0.00	0.00	0.00
C27	0.00	0.01	-0.01	0.01	-0.01	0.00
C28	0.00	0.00	0.00	0.00	0.00	0.00
C29	0.00	0.00	-0.01	0.01	0.00	0.00
C30	- -	0.00	0.00	0.00	0.01	0.00
C31	- -	0.00	0.00	0.00	0.00	0.01
C32	0.01	0.00	-0.01	0.00	0.01	0.01

Expected Change for THETA-EPS

	C7	C8	C9	C10	C11	C12
C7	- -					
C8	0.01	- -				
C9	0.00	- -	- -			
C10	0.00	- -	0.01	- -		
C11	0.00	0.01	0.00	0.00	- -	
C12	- -	- -	- -	- -	0.01	- -

C13	0.00	0.01	-0.01	0.00	0.00	0.00
C14	-0.01	- -	0.00	0.00	0.00	0.01
C15	0.00	0.01	-0.01	0.00	0.01	- -
C16	0.00	0.01	0.00	0.00	0.01	- -
C17	0.00	0.00	0.01	0.01	0.00	- -
C18	0.00	0.00	-0.01	0.00	-0.01	0.00
C19	0.01	0.01	0.00	-0.01	0.00	- -
C20	0.00	0.01	-0.01	-0.01	0.00	- -
C21	0.00	0.01	0.01	- -	0.00	- -
C22	0.00	0.01	0.00	-0.01	0.00	- -
C23	0.00	- -	0.00	0.00	- -	- -
C24	0.00	0.01	0.00	0.00	0.00	- -
C25	0.00	- -	0.00	0.01	0.01	- -
C26	-0.01	0.00	0.01	0.00	0.00	- -
C27	0.01	- -	0.01	0.00	0.00	- -
C28	0.01	0.01	0.00	0.00	- -	- -
C29	0.00	- -	0.00	0.00	- -	- -
C30	0.00	- -	0.01	0.01	- -	0.01
C31	0.00	- -	0.00	0.00	0.01	- -
C32	0.01	- -	0.00	0.01	0.00	- -

Expected Change for THETA-EPS

	C13	C14	C15	C16	C17	C18
C13	- -					
C14	0.00	- -				
C15	0.01	0.00	- -			
C16	0.00	0.00	0.00	- -		
C17	0.00	0.00	0.00	0.00	- -	
C18	0.00	0.00	0.00	0.00	- -	- -
C19	- -	0.00	0.00	0.00	- -	0.00
C20	0.01	0.00	-0.01	0.00	0.00	0.00
C21	- -	0.00	0.00	0.00	0.00	0.00
C22	-0.01	0.00	0.00	0.00	0.00	0.00
C23	0.00	0.00	0.00	0.00	0.00	0.00
C24	0.00	0.00	0.00	0.00	-0.01	0.00
C25	0.00	0.00	0.00	0.00	-0.01	0.00
C26	0.00	0.00	0.00	0.00	0.00	0.00
C27	- -	0.01	0.00	0.00	0.01	0.00
C28	0.00	0.00	-0.01	0.00	-0.01	- -
C29	- -	0.00	0.01	-0.01	-0.01	0.00
C30	0.00	0.01	0.00	0.00	0.00	-0.01
C31	-0.01	0.00	-0.01	0.00	-0.01	0.00
C32	0.00	-0.01	0.01	0.00	0.00	0.01

Expected Change for THETA-EPS

	C19	C20	C21	C22	C23	C24
C19	- -					
C20	0.00	- -				
C21	- -	0.01	- -			
C22	0.00	0.00	0.00	- -		
C23	0.00	0.00	0.00	0.00	- -	
C24	0.00	0.00	0.00	-0.01	0.00	- -
C25	-0.01	0.00	0.00	0.00	-0.01	-0.01
C26	0.00	0.00	0.00	0.00	0.00	0.01
C27	0.00	0.01	0.00	0.00	0.01	0.01
C28	0.00	0.00	0.00	0.00	0.01	0.01
C29	-0.01	0.00	0.00	0.00	0.00	0.00
C30	0.00	-0.01	-0.01	-0.01	- -	0.00
C31	0.00	0.00	0.00	0.00	0.00	0.00
C32	0.01	-0.01	0.00	0.00	0.00	-0.01

Expected Change for THETA-EPS

	C25	C26	C27	C28	C29	C30
C25	- -					
C26	0.00	- -				
C27	0.00	0.00	- -			
C28	0.00	0.00	0.00	- -		

C29	0.00	0.00	0.00	0.00	- -	
C30	0.00	- -	0.00	0.00	0.01	- -
C31	0.00	0.00	0.01	0.00	0.01	0.00
C32	-0.01	0.00	- -	0.00	0.00	0.00

Expected Change for THETA-EPS

	C31	C32
C31	- -	
C32	0.00	- -

Maximum Modification Index is 115.33 for Element (8, 2) of LAMBDA-Y

Time used: 0.375 Seconds

ภาคผนวก ฉ
ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบ O-NET ด้วยวิธี IRT-LR และวิธี SIBTEST

ตารางที่ ฉ-1 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ด้วยวิธี IRT-LR และวิธี SIBTEST ของแบบทดสอบภาษาไทย

ข้อที่	ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF)			
	วิธี IRT-LR		วิธี SIBTEST	
	p - value		*p* - value	
1	0.000	(DIF)	0.010	(DIF)
2	0.000	(DIF)	0.083	(NO DIF)
3	0.000	(DIF)	0.239	(NO DIF)
4	0.000	(DIF)	0.209	(NO DIF)
5	0.000	(DIF)	0.154	(NO DIF)
6	0.000	(DIF)	0.377	(NO DIF)
7	0.000	(DIF)	0.300	(NO DIF)
8	0.000	(DIF)	0.516	(NO DIF)
9	0.000	(DIF)	0.030	(DIF)
10	0.000	(DIF)	0.997	(NO DIF)
11	0.000	(DIF)	0.882	(NO DIF)
12	0.000	(DIF)	0.598	(NO DIF)
13	0.015	(DIF)	0.711	(NO DIF)
14	0.269	(NO DIF)	0.570	(NO DIF)
15	0.018	(DIF)	0.027	(DIF)
16	0.000	(DIF)	0.011	(DIF)
17	0.171	(NO DIF)	0.207	(NO DIF)
18	0.000	(DIF)	0.161	(NO DIF)
19	0.002	(DIF)	0.552	(NO DIF)
20	0.483	(NO DIF)	0.907	(NO DIF)
21	0.000	(DIF)	0.298	(NO DIF)
22	0.000	(DIF)	0.038	(DIF)
23	0.004	(DIF)	0.927	(NO DIF)
24	0.000	(DIF)	0.511	(NO DIF)
25	0.000	(DIF)	0.501	(NO DIF)
26	0.000	(DIF)	0.870	(NO DIF)
27	0.000	(DIF)	0.281	(NO DIF)
28	0.000	(DIF)	0.490	(NO DIF)
29	0.000	(DIF)	0.598	(NO DIF)
30	0.813	(NO DIF)	0.443	(NO DIF)
31	0.710	(NO DIF)	0.887	(NO DIF)
32	0.296	(NO DIF)	0.055	(NO DIF)
33	0.019	(DIF)	0.166	(NO DIF)
34	0.000	(DIF)	0.060	(NO DIF)
35	0.000	(DIF)	0.270	(NO DIF)

ตารางที่ ฉ-1 (ต่อ)

ข้อที่	ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF)	
	วิธี IRT-LR	วิธี SIBTEST
	p - value	*p* - value
36	0.000 (DIF)	0.024 (DIF)
37	0.000 (DIF)	0.072 (NO DIF)
38	0.000 (DIF)	0.076 (NO DIF)
39	0.000 (DIF)	0.594 (NO DIF)
40	0.649 (NO DIF)	0.817 (NO DIF)
41	0.002 (DIF)	0.124 (NO DIF)
42	0.000 (DIF)	0.778 (NO DIF)
43	0.000 (DIF)	0.186 (NO DIF)
44	0.000 (DIF)	0.033 (DIF)
45	0.000 (DIF)	0.929 (NO DIF)
46	0.000 (DIF)	0.519 (NO DIF)
47	0.000 (DIF)	0.632 (NO DIF)
48	0.000 (DIF)	0.070 (NO DIF)
49	0.000 (DIF)	0.935 (NO DIF)
50	0.000 (DIF)	0.042 (DIF)
51	0.000 (DIF)	0.087 (NO DIF)
52	0.224 (NO DIF)	0.985 (NO DIF)
53	0.000 (DIF)	0.019 (DIF)
54	0.002 (DIF)	0.069 (NO DIF)
55	0.818 (NO DIF)	0.265 (NO DIF)
56	0.511 (NO DIF)	0.236 (NO DIF)
57	0.365 (NO DIF)	0.854 (NO DIF)
58	0.000 (DIF)	0.582 (NO DIF)
59	0.000 (DIF)	0.056 (NO DIF)
60	0.000 (DIF)	0.515 (NO DIF)
61	0.001 (DIF)	0.372 (NO DIF)
62	0.002 (DIF)	0.258 (NO DIF)
63	0.000 (DIF)	0.473 (NO DIF)
64	0.000 (DIF)	0.001 (DIF)
65	0.008 (DIF)	0.006 (DIF)
66	0.000 (DIF)	0.045 (DIF)
67	0.052 (NO DIF)	0.027 (DIF)
68	0.000 (DIF)	0.008 (DIF)
69	0.000 (DIF)	0.130 (NO DIF)
70	0.000 (DIF)	0.107 (NO DIF)
รวม	58 ข้อ	14 ข้อ

ตารางที่ ฉ-2 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ด้วยวิธี IRT-LR และวิธี SIBTEST ของแบบทดสอบสังคมศึกษา ศาสนาและวัฒนธรรม

ข้อที่	ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF)			
	วิธี IRT-LR		วิธี SIBTEST	
	p - value		*p* - value	
1	0.000	(DIF)	0.153	(NO DIF)
2	0.002	(DIF)	0.278	(NO DIF)
3	0.000	(DIF)	0.000	(DIF)
4	0.000	(DIF)	0.538	(NO DIF)
5	0.000	(DIF)	0.547	(NO DIF)
6	0.000	(DIF)	0.036	(DIF)
7	0.000	(DIF)	0.017	(DIF)
8	0.001	(DIF)	0.541	(NO DIF)
9	0.000	(DIF)	0.165	(NO DIF)
10	0.035	(DIF)	0.738	(NO DIF)
11	0.000	(DIF)	0.151	(NO DIF)
12	0.000	(DIF)	0.085	(NO DIF)
13	0.000	(DIF)	0.557	(NO DIF)
14	0.001	(DIF)	0.012	(DIF)
15	0.000	(DIF)	0.943	(NO DIF)
16	0.000	(DIF)	0.019	(DIF)
17	0.000	(DIF)	0.470	(NO DIF)
18	0.000	(DIF)	0.362	(NO DIF)
19	0.000	(DIF)	0.119	(NO DIF)
20	0.000	(DIF)	0.012	(DIF)
21	0.000	(DIF)	0.010	(DIF)
22	0.000	(DIF)	0.906	(NO DIF)
23	0.005	(DIF)	0.186	(NO DIF)
24	0.039	(DIF)	0.930	(NO DIF)
25	0.004	(DIF)	0.000	(DIF)
26	0.000	(DIF)	0.583	(NO DIF)
27	0.000	(DIF)	0.004	(DIF)
28	0.000	(DIF)	0.130	(NO DIF)
29	0.000	(DIF)	0.169	(NO DIF)
30	0.000	(DIF)	0.375	(NO DIF)
31	0.011	(DIF)	0.066	(NO DIF)
32	0.000	(DIF)	0.202	(NO DIF)
33	0.683	(NO DIF)	0.359	(NO DIF)
34	0.000	(DIF)	0.115	(NO DIF)
35	0.708	(NO DIF)	0.644	(NO DIF)

ตารางที่ ฉ-2 (ต่อ)

ข้อที่	ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF)	
	วิธี IRT-LR	วิธี SIBTEST
	p - value	*p* - value
36	0.001 (DIF)	0.609 (NO DIF)
37	0.000 (DIF)	0.850 (NO DIF)
38	0.000 (DIF)	0.434 (NO DIF)
39	0.000 (DIF)	0.008 (DIF)
40	0.000 (DIF)	0.632 (NO DIF)
41	0.000 (DIF)	0.153 (NO DIF)
42	0.899 (NO DIF)	0.867 (NO DIF)
43	0.000 (DIF)	0.883 (NO DIF)
44	0.021 (DIF)	0.146 (NO DIF)
45	0.000 (DIF)	0.001 (DIF)
46	0.591 (NO DIF)	0.894 (NO DIF)
47	0.000 (DIF)	0.804 (NO DIF)
48	0.014 (DIF)	0.879 (NO DIF)
49	0.000 (DIF)	0.452 (NO DIF)
50	0.000 (DIF)	0.056 (NO DIF)
51	0.042 (DIF)	0.057 (NO DIF)
52	0.000 (DIF)	0.056 (NO DIF)
53	0.817 (NO DIF)	0.686 (NO DIF)
54	0.000 (DIF)	0.508 (NO DIF)
55	0.000 (DIF)	0.711 (NO DIF)
56	0.291 (NO DIF)	0.144 (NO DIF)
57	0.000 (DIF)	0.428 (NO DIF)
58	0.000 (DIF)	0.000 (DIF)
59	0.000 (DIF)	0.528 (NO DIF)
60	0.000 (DIF)	0.682 (NO DIF)
61	0.010 (DIF)	0.121 (NO DIF)
62	0.000 (DIF)	0.602 (NO DIF)
63	0.000 (DIF)	0.778 (NO DIF)
64	0.000 (DIF)	0.000 (DIF)
65	0.000 (DIF)	0.396 (NO DIF)
รวม	59 ข้อ	13 ข้อ

ตารางที่ ฉ-3 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ด้วยวิธี IRT-LR และวิธี SIBTEST ของแบบทดสอบภาษาอังกฤษ

ข้อที่	ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF)	
	วิธี IRT-LR	วิธี SIBTEST
	p - value	*p* - value
1	0.317 (NO DIF)	0.074 (NO DIF)
2	0.466 (NO DIF)	0.452 (NO DIF)
3	0.646 (NO DIF)	0.760 (NO DIF)
4	0.192 (NO DIF)	0.467 (NO DIF)
5	0.081 (NO DIF)	0.395 (NO DIF)
6	0.089 (NO DIF)	0.025 (DIF)
7	0.833 (NO DIF)	0.736 (NO DIF)
8	0.348 (NO DIF)	0.064 (NO DIF)
9	0.127 (NO DIF)	0.266 (NO DIF)
10	0.774 (NO DIF)	0.544 (NO DIF)
11	0.404 (NO DIF)	0.779 (NO DIF)
12	0.776 (NO DIF)	0.253 (NO DIF)
13	0.484 (NO DIF)	0.762 (NO DIF)
14	0.234 (NO DIF)	0.464 (NO DIF)
15	0.612 (NO DIF)	0.289 (NO DIF)
16	0.212 (NO DIF)	0.046 (DIF)
17	0.272 (NO DIF)	0.655 (NO DIF)
18	0.903 (NO DIF)	0.518 (NO DIF)
19	0.339 (NO DIF)	0.611 (NO DIF)
20	0.275 (NO DIF)	0.055 (NO DIF)
21	0.254 (NO DIF)	0.877 (NO DIF)
22	0.655 (NO DIF)	0.845 (NO DIF)
23	0.681 (NO DIF)	0.957 (NO DIF)
24	0.716 (NO DIF)	0.230 (NO DIF)
25	0.107 (NO DIF)	0.658 (NO DIF)
26	0.012 (DIF)	0.007 (DIF)
27	0.182 (NO DIF)	0.769 (NO DIF)
28	0.850 (NO DIF)	0.864 (NO DIF)
29	0.305 (NO DIF)	0.470 (NO DIF)
30	0.442 (NO DIF)	0.403 (NO DIF)
31	0.061 (NO DIF)	0.106 (NO DIF)
32	0.425 (NO DIF)	0.029 (DIF)
33	0.989 (NO DIF)	0.748 (NO DIF)
34	0.613 (NO DIF)	0.069 (NO DIF)
35	0.135 (NO DIF)	0.608 (NO DIF)
36	0.287 (NO DIF)	0.183 (NO DIF)
37	0.884 (NO DIF)	0.647 (NO DIF)
38	0.455 (NO DIF)	0.637 (NO DIF)
39	0.130 (NO DIF)	0.055 (NO DIF)
40	0.116 (NO DIF)	0.372 (NO DIF)

ตารางที่ ฉ-3 (ต่อ)

ข้อที่	ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF)	
	วิธี IRT-LR	วิธี SIBTEST
	p - value	*p* - value
41	0.698 (NO DIF)	0.566 (NO DIF)
42	0.997 (NO DIF)	0.124 (NO DIF)
43	0.255 (NO DIF)	0.020 (DIF)
44	0.006 (NO DIF)	0.086 (NO DIF)
45	0.251 (NO DIF)	0.543 (NO DIF)
46	0.646 (NO DIF)	0.198 (NO DIF)
47	0.959 (NO DIF)	0.976 (NO DIF)
48	0.418 (NO DIF)	0.523 (NO DIF)
49	0.358 (NO DIF)	0.578 (NO DIF)
50	0.327 (NO DIF)	0.028 (DIF)
51	0.357 (NO DIF)	0.226 (NO DIF)
52	0.695 (NO DIF)	0.498 (NO DIF)
53	0.753 (NO DIF)	0.406 (NO DIF)
54	0.085 (NO DIF)	0.418 (NO DIF)
55	0.398 (NO DIF)	0.109 (NO DIF)
56	0.376 (NO DIF)	0.906 (NO DIF)
57	0.383 (NO DIF)	0.193 (NO DIF)
58	0.222 (NO DIF)	0.037 (DIF)
59	0.020 (DIF)	0.002 (DIF)
60	0.572 (NO DIF)	0.192 (NO DIF)
61	0.468 (NO DIF)	0.504 (NO DIF)
62	0.482 (NO DIF)	0.922 (NO DIF)
63	0.491 (NO DIF)	0.503 (NO DIF)
64	0.336 (NO DIF)	0.238 (NO DIF)
65	0.289 (NO DIF)	0.769 (NO DIF)
66	0.439 (NO DIF)	0.999 (NO DIF)
67	0.572 (NO DIF)	0.663 (NO DIF)
68	0.026 (DIF)	0.283 (NO DIF)
69	0.415 (NO DIF)	0.452 (NO DIF)
70	0.345 (NO DIF)	0.483 (NO DIF)
71	0.276 (NO DIF)	0.085 (NO DIF)
72	0.596 (NO DIF)	0.352 (NO DIF)
73	0.582 (NO DIF)	0.993 (NO DIF)
74	0.005 (DIF)	0.046 (DIF)
75	0.703 (NO DIF)	0.568 (NO DIF)
76	0.629 (NO DIF)	0.193 (NO DIF)
77	0.238 (NO DIF)	0.192 (NO DIF)
78	0.652 (NO DIF)	0.114 (NO DIF)
79	0.321 (NO DIF)	0.106 (NO DIF)
80	0.402 (NO DIF)	0.692 (NO DIF)
รวม	4 ข้อ	9 ข้อ

ตารางที่ ฉ-4 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ด้วยวิธี IRT-LR และวิธี SIBTEST ของแบบทดสอบคณิตศาสตร์

ข้อที่	ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF)	
	วิธี IRT-LR	วิธี SIBTEST
	p - value	*p* - value
1	0.928 NO DIF	0.446 NO DIF
2	0.607 NO DIF	0.640 NO DIF
3	0.940 NO DIF	0.905 NO DIF
4	0.926 NO DIF	0.885 NO DIF
5	0.152 NO DIF	0.715 NO DIF
6	0.033 DIF	0.018 DIF
7	0.464 NO DIF	0.684 NO DIF
8	0.114 NO DIF	0.934 NO DIF
9	0.243 NO DIF	0.332 NO DIF
10	0.121 NO DIF	0.115 NO DIF
11	0.538 NO DIF	0.400 NO DIF
12	0.714 NO DIF	0.330 NO DIF
13	0.146 NO DIF	0.659 NO DIF
14	0.004 DIF	0.233 NO DIF
15	0.221 NO DIF	0.753 NO DIF
16	0.613 NO DIF	0.778 NO DIF
17	0.982 NO DIF	0.381 NO DIF
18	0.423 NO DIF	0.642 NO DIF
19	0.008 DIF	0.007 DIF
20	0.615 NO DIF	0.631 NO DIF
21	0.020 NO DIF	0.968 NO DIF
22	0.655 NO DIF	0.940 NO DIF
23	0.796 NO DIF	0.715 NO DIF
24	0.353 NO DIF	0.633 NO DIF
25	0.206 NO DIF	0.121 NO DIF
26	0.021 DIF	0.058 NO DIF
27	0.376 NO DIF	0.103 NO DIF
28	0.030 DIF	0.133 NO DIF
29	0.656 NO DIF	0.830 NO DIF
30	0.026 DIF	0.070 NO DIF
31	0.273 NO DIF	0.321 NO DIF
32	0.166 NO DIF	0.451 NO DIF
รวม	6 ข้อ	2 ข้อ

ตารางที่ ฉ-5 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ด้วยวิธี IRT-LR และวิธี SIBTEST ของแบบทดสอบวิทยาศาสตร์

ข้อที่	ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF)			
	วิธี IRT-LR		วิธี SIBTEST	
	p - value		*p* - value	
1	0.000	(DIF)	0.595	(NO DIF)
2	0.000	(DIF)	0.695	(NO DIF)
3	0.000	(DIF)	0.060	(NO DIF)
4	0.166	(NO DIF)	0.363	(NO DIF)
5	0.000	(DIF)	0.177	(NO DIF)
6	0.339	(NO DIF)	0.304	(NO DIF)
7	0.064	(NO DIF)	0.110	(NO DIF)
8	0.001	(DIF)	0.260	(NO DIF)
9	0.000	(DIF)	0.894	(NO DIF)
10	0.011	(DIF)	0.002	(DIF)
11	0.000	(DIF)	0.008	(DIF)
12	0.002	(DIF)	0.451	(NO DIF)
13	0.000	(DIF)	0.419	(NO DIF)
14	0.000	(DIF)	0.334	(NO DIF)
15	0.118	(NO DIF)	0.100	(NO DIF)
16	0.000	(DIF)	0.031	(DIF)
17	0.011	(DIF)	0.328	(NO DIF)
18	0.778	(NO DIF)	0.874	(NO DIF)
19	0.000	(DIF)	0.471	(NO DIF)
20	0.790	(NO DIF)	0.477	(NO DIF)
21	0.007	(DIF)	0.044	(DIF)
22	0.000	(DIF)	0.331	(NO DIF)
23	0.132	(NO DIF)	0.844	(NO DIF)
24	0.354	(NO DIF)	0.400	(NO DIF)
25	0.000	(DIF)	0.095	(NO DIF)
26	0.008	(DIF)	0.969	(NO DIF)
27	0.000	(DIF)	0.673	(NO DIF)
28	0.608	(NO DIF)	0.318	(NO DIF)
29	0.063	(NO DIF)	0.161	(NO DIF)
30	0.006	(DIF)	0.084	(NO DIF)
31	0.012	(DIF)	0.006	(DIF)
32	0.860	(NO DIF)	0.106	(NO DIF)
33	0.643	(NO DIF)	0.241	(NO DIF)
34	0.012	(DIF)	0.807	(NO DIF)
35	0.028	(DIF)	0.435	(NO DIF)
36	0.000	(DIF)	0.443	(NO DIF)
37	0.057	(NO DIF)	0.550	(NO DIF)
38	0.207	(NO DIF)	0.856	(NO DIF)
39	0.000	(DIF)	0.040	(DIF)
40	0.000	(DIF)	0.253	(NO DIF)

ตารางที่ ฉ-5 (ต่อ)

ข้อที่	ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF)	
	วิธี IRT-LR	วิธี SIBTEST
	p - value	*p* - value
41	0.000 (DIF)	0.109 (NO DIF)
42	0.000 (DIF)	0.331 (NO DIF)
43	0.000 (DIF)	0.142 (NO DIF)
44	0.000 (DIF)	0.717 (NO DIF)
45	0.000 (DIF)	0.106 (NO DIF)
46	0.000 (DIF)	0.017 (DIF)
47	0.000 (DIF)	0.435 (NO DIF)
48	0.000 (DIF)	0.505 (NO DIF)
49	0.081 (NO DIF)	0.399 (NO DIF)
50	0.000 (DIF)	0.798 (NO DIF)
51	0.000 (DIF)	0.061 (NO DIF)
52	0.000 (DIF)	0.180 (NO DIF)
53	0.000 (DIF)	0.528 (NO DIF)
54	0.000 (DIF)	0.861 (NO DIF)
55	0.000 (DIF)	0.734 (NO DIF)
56	0.000 (DIF)	0.106 (NO DIF)
57	0.554 (NO DIF)	0.231 (NO DIF)
58	0.135 (NO DIF)	0.253 (NO DIF)
59	0.003 (DIF)	0.028 (DIF)
60	0.697 (NO DIF)	0.295 (NO DIF)
61	0.126 (NO DIF)	0.111 (NO DIF)
62	0.044 (NO DIF)	0.910 (NO DIF)
63	0.000 (DIF)	0.530 (NO DIF)
64	0.086 (NO DIF)	0.568 (NO DIF)
65	0.000 (DIF)	0.210 (NO DIF)
66	0.000 (DIF)	0.647 (NO DIF)
67	0.000 (DIF)	0.198 (NO DIF)
68	0.406 (NO DIF)	0.348 (NO DIF)
69	0.002 (DIF)	0.352 (NO DIF)
70	0.001 (DIF)	0.094 (NO DIF)
71	0.956 (NO DIF)	0.975 (NO DIF)
72	0.000 (DIF)	0.456 (NO DIF)
73	0.441 (NO DIF)	0.422 (NO DIF)
74	0.000 (DIF)	0.000 (DIF)
75	0.000 (DIF)	0.977 (NO DIF)
76	0.058 (NO DIF)	0.717 (NO DIF)
77	0.000 (DIF)	0.069 (NO DIF)
78	0.000 (DIF)	0.002 (DIF)
79	0.003 (DIF)	0.095 (NO DIF)
80	0.000 (DIF)	0.005 (DIF)
รวม	56 ข้อ	11 ข้อ

ตารางที่ ฉ-6 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ด้วยวิธี IRT-LR และวิธี SIBTEST ของแบบทดสอบสุขศึกษา และพลศึกษา

ข้อที่	ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF)	
	วิธี IRT-LR	วิธี SIBTEST
	p - value	*p* - value
1	0.000 (DIF)	0.047 (DIF)
2	0.000 (DIF)	0.260 (NO DIF)
3	0.000 (DIF)	0.008 (DIF)
4	0.000 (DIF)	0.334 (NO DIF)
5	0.000 (DIF)	0.618 (NO DIF)
6	0.000 (DIF)	0.000 (DIF)
7	0.000 (DIF)	0.081 (NO DIF)
8	0.377 (NO DIF)	0.806 (NO DIF)
9	0.000 (DIF)	0.443 (NO DIF)
10	0.000 (DIF)	0.535 (NO DIF)
11	0.000 (DIF)	0.000 (DIF)
12	0.000 (DIF)	0.682 (NO DIF)
13	0.000 (DIF)	0.045 (DIF)
14	0.000 (DIF)	0.606 (NO DIF)
15	0.000 (DIF)	0.110 (NO DIF)
16	0.042 (DIF)	0.997 (NO DIF)
17	0.000 (DIF)	0.439 (NO DIF)
18	0.000 (DIF)	0.976 (NO DIF)
19	0.000 (DIF)	0.004 (DIF)
20	0.000 (DIF)	0.000 (DIF)
21	0.000 (DIF)	0.008 (DIF)
22	0.000 (DIF)	0.807 (NO DIF)
23	0.000 (DIF)	0.183 (NO DIF)
24	0.000 (DIF)	0.089 (NO DIF)
25	0.000 (DIF)	0.003 (DIF)
26	0.000 (DIF)	0.624 (NO DIF)
27	0.000 (DIF)	0.004 (DIF)
28	0.000 (DIF)	0.500 (NO DIF)
29	0.000 (DIF)	0.000 (DIF)
30	0.000 (DIF)	0.139 (NO DIF)
31	0.000 (DIF)	0.813 (NO DIF)
32	0.003 (DIF)	0.326 (NO DIF)
รวม	31 ข้อ	11 ข้อ

ตารางที่ ฉ-7 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ด้วยวิธี IRT-LR และวิธี SIBTEST ของแบบทดสอบศิลปะ

ข้อที่	ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF)	
	วิธี IRT-LR	วิธี SIBTEST
	p - value	*p* - value
1	0.003 (DIF)	0.019 (DIF)
2	0.440 (NO DIF)	0.522 (NO DIF)
3	0.000 (DIF)	0.895 (NO DIF)
4	0.000 (DIF)	0.800 (NO DIF)
5	0.000 (DIF)	0.294 (NO DIF)
6	0.004 (DIF)	0.394 (NO DIF)
7	0.000 (DIF)	0.963 (NO DIF)
8	0.000 (DIF)	0.020 (DIF)
9	0.000 (DIF)	0.966 (NO DIF)
10	0.000 (DIF)	0.377 (NO DIF)
11	0.000 (DIF)	0.260 (NO DIF)
12	0.000 (DIF)	0.168 (NO DIF)
13	0.000 (DIF)	0.195 (NO DIF)
14	0.000 (DIF)	0.000 (DIF)
15	0.000 (DIF)	0.744 (NO DIF)
16	0.026 (DIF)	0.990 (NO DIF)
17	0.000 (DIF)	0.000 (DIF)
18	0.000 (DIF)	0.886 (NO DIF)
19	0.000 (DIF)	0.283 (NO DIF)
20	0.000 (DIF)	0.059 (NO DIF)
21	0.000 (DIF)	0.239 (NO DIF)
22	0.000 (DIF)	0.065 (NO DIF)
23	0.000 (DIF)	0.336 (NO DIF)
24	0.000 (DIF)	0.051 (NO DIF)
25	0.491 (NO DIF)	0.753 (NO DIF)
26	0.241 (NO DIF)	0.308 (NO DIF)
27	0.000 (DIF)	0.010 (DIF)
28	0.000 (DIF)	0.811 (NO DIF)
29	0.066 (NO DIF)	0.153 (NO DIF)
30	0.000 (DIF)	0.000 (DIF)
รวม	26 ข้อ	6 ข้อ

ตารางที่ ฉ-8 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ด้วยวิธี IRT-LR และวิธี SIBTEST ของแบบทดสอบการงานอาชีพ และเทคโนโลยี

ข้อที่	ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF)	
	วิธี IRT-LR	วิธี SIBTEST
	p - value	*p* - value
1	0.889 (NO DIF)	0.891 (NO DIF)
2	0.056 (NO DIF)	0.943 (NO DIF)
3	0.007 (DIF)	0.504 (NO DIF)
4	0.759 (NO DIF)	0.001 (DIF)
5	0.110 (NO DIF)	0.658 (NO DIF)
6	0.011 (DIF)	0.539 (NO DIF)
7	0.002 (DIF)	0.004 (DIF)
8	0.532 (NO DIF)	0.155 (NO DIF)
9	0.000 (DIF)	0.003 (DIF)
10	0.000 (DIF)	0.137 (NO DIF)
11	0.994 (NO DIF)	0.159 (NO DIF)
12	0.030 (DIF)	0.051 (NO DIF)
13	0.525 (NO DIF)	0.975 (NO DIF)
14	0.055 (NO DIF)	0.068 (NO DIF)
15	0.098 (NO DIF)	0.040 (DIF)
16	0.000 (DIF)	0.345 (NO DIF)
17	0.562 (NO DIF)	0.525 (NO DIF)
18	0.000 (DIF)	0.313 (NO DIF)
19	0.187 (NO DIF)	0.409 (NO DIF)
20	0.531 (NO DIF)	0.024 (DIF)
21	0.022 (DIF)	0.195 (NO DIF)
22	0.773 (NO DIF)	0.044 (DIF)
23	0.199 (NO DIF)	0.060 (NO DIF)
24	0.505 (NO DIF)	0.146 (NO DIF)
25	0.000 (DIF)	0.983 (NO DIF)
26	0.056 (NO DIF)	0.364 (NO DIF)
27	0.324 (NO DIF)	0.032 (DIF)
28	0.211 (NO DIF)	0.025 (DIF)
29	0.144 (NO DIF)	0.374 (NO DIF)
30	0.013 (DIF)	0.585 (NO DIF)
31	0.008 (DIF)	0.416 (NO DIF)
32	0.002 (DIF)	0.000 (DIF)
33	0.163 (NO DIF)	0.027 (DIF)
34	0.005 (DIF)	0.571 (NO DIF)
35	0.712 (NO DIF)	0.904 (NO DIF)
36	0.102 (NO DIF)	0.546 (NO DIF)
37	0.000 (DIF)	0.074 (NO DIF)
38	0.000 (DIF)	0.410 (NO DIF)
39	0.000 (DIF)	0.002 (DIF)
40	0.000 (DIF)	0.039 (DIF)
41	0.000 (DIF)	0.011 (DIF)
รวม	19 ข้อ	13 ข้อ

ภาคผนวก ช
ตัวอย่าง Print Out ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ในแบบทดสอบ O-NET ด้วยวิธี IRT-LR โดยใช้โปรแกรม IRTPRO

IRTPRO Version 2.1

Output generated by IRTPRO estimation engine Version 4.54 (32-bit)

Project:	04 MATH
Description:	
Date:	22 June 2017
Time:	02:10 PM

Table of Contents

2PL Model Item Parameter Estimates for Group 1, logit: $a\theta + c$ or $a(\theta - b)$

Summed-Score Based Item Diagnostic Tables and X^2s for Group 1

2PL Model Item Parameter Estimates for Group 2, logit: $a\theta + c$ or $a(\theta - b)$

Summed-Score Based Item Diagnostic Tables and X^2s for Group 2

Group Parameter Estimates

DIF Statistics for Graded Items

Marginal fit (X^2) and Standardized LD X^2 Statistics for Group 1

Marginal fit (X^2) and Standardized LD X^2 Statistics for Group 2

Item Information Function Values for Group 1 at 15 Values of θ from -2.8 to 2.8

Item Information Function Values for Group 2 at 15 Values of θ from -2.8 to 2.8

Likelihood-based Values and Goodness of Fit Statistics

Summary of the Data and Control Parameters

2PL Model Item Parameter Estimates for Group 1, logit: $a\theta + c$ or $a(\theta - b)$ (Back to TOC)

Item	Label	a	s.e.	c	s.e.	b	s.e.
1	C1	[2] 1.00	0.11	[1] -1.80	0.10	1.80	0.18
2	C2	[4] 0.44	0.09	[3] -1.45	0.08	3.28	0.61
3	C3	[6] 0.68	0.09	[5] -1.33	0.08	1.98	0.26
4	C4	[8] 0.81	0.10	[7] -1.26	0.08	1.57	0.18
5	C5	[10] 0.66	0.09	[9] -1.11	0.08	1.67	0.23
6	C6	[12] 0.89	0.10	[11] -1.61	0.09	1.81	0.19
7	C7	[14] 1.20	0.13	[13] -1.03	0.08	0.86	0.10
8	C8	[16] 1.12	0.12	[15] -1.09	0.08	0.97	0.11
9	C9	[18] 0.79	0.10	[17] -0.98	0.08	1.24	0.16
10	C10	[20] 0.86	0.10	[19] -1.19	0.08	1.38	0.16
11	C11	[22] 0.51	0.08	[21] -0.85	0.07	1.66	0.28
12	C12	[24] 0.66	0.09	[23] -1.63	0.09	2.49	0.33
13	C13	[26] 0.63	0.09	[25] -1.77	0.10	2.80	0.38
14	C14	[28] 0.34	0.08	[27] -1.01	0.07	2.96	0.69

15	C15	30	0.58	0.09	29	−1.21	0.08	2.10	0.31
16	C16	32	0.69	0.09	31	−1.38	0.08	2.01	0.26
17	C17	34	0.61	0.09	33	−1.21	0.08	1.98	0.28
18	C18	36	0.82	0.10	35	−1.36	0.09	1.65	0.19
19	C19	38	1.50	0.15	37	−1.86	0.12	1.24	0.10
20	C20	40	1.03	0.11	39	−1.14	0.08	1.11	0.12
21	C21	42	0.85	0.10	41	−1.23	0.08	1.45	0.17
22	C22	44	0.89	0.10	43	−1.13	0.08	1.27	0.15
23	C23	46	0.93	0.10	45	−1.22	0.08	1.31	0.15
24	C24	48	1.17	0.12	47	−1.06	0.08	0.91	0.10
25	C25	50	0.82	0.10	49	−1.17	0.08	1.44	0.17
26	C26	52	0.58	0.09	51	−1.34	0.08	2.29	0.33
27	C27	54	0.48	0.09	53	−1.46	0.08	3.02	0.52
28	C28	56	0.58	0.09	55	−0.87	0.07	1.51	0.24
29	C29	58	0.66	0.09	57	−1.02	0.08	1.54	0.21
30	C30	60	1.09	0.12	59	−0.64	0.08	0.59	0.09
31	C31	62	0.82	0.10	61	−0.89	0.08	1.09	0.14
32	C32	64	0.65	0.09	63	−1.00	0.08	1.54	0.22

Summed-Score Based Item Diagnostic Tables and X^2s for Group 1 (Back to TOC)

S-X^2 Item Level Diagnostic Statistics

Item	Label	X^2	*d.f.*	Probability
1	C1	40.02	24	0.0212
2	C2	34.13	26	0.1314
3	C3	25.89	26	0.4708
4	C4	40.72	23	0.0127
5	C5	51.72	25	0.0013
6	C6	35.47	24	0.0616
7	C7	41.22	19	0.0022
8	C8	29.44	21	0.1036
9	C9	50.76	24	0.0011
10	C10	20.14	21	0.5137
11	C11	28.91	25	0.2668
12	C12	40.14	26	0.0377
13	C13	23.10	24	0.5150
14	C14	53.13	26	0.0013
15	C15	40.77	25	0.0242
16	C16	41.15	25	0.0221
17	C17	44.93	24	0.0059

18	C18	39.19	23	0.0189
19	C19	30.14	21	0.0889
20	C20	37.61	21	0.0143
21	C21	53.74	24	0.0005
22	C22	31.21	22	0.0917
23	C23	34.53	23	0.0577
24	C24	43.49	20	0.0018
25	C25	28.12	23	0.2105
26	C26	41.48	27	0.0369
27	C27	29.80	25	0.2313
28	C28	37.96	25	0.0466
29	C29	22.09	25	0.6315
30	C30	26.31	20	0.1554
31	C31	29.65	23	0.1593
32	C32	33.09	25	0.1284

2PL Model Item Parameter Estimates for Group 2, logit: $a\boldsymbol{\theta} + c$ or $a(\boldsymbol{\theta} - b)$ (Back to TOC)

Item	Label	*a*	*s.e.*	*c*	*s.e.*	*b*	*s.e.*
1	C1	[66] 0.97	0.14	[65] −1.84	0.10	1.90	0.25
2	C2	[68] 0.56	0.12	[67] −1.54	0.09	2.74	0.57
3	C3	[70] 0.69	0.12	[69] −1.30	0.08	1.88	0.32
4	C4	[72] 0.77	0.12	[71] −1.29	0.08	1.67	0.26
5	C5	[74] 0.51	0.11	[73] −1.27	0.08	2.51	0.56
6	C6	[76] 0.84	0.13	[75] −1.94	0.10	2.31	0.35
7	C7	[78] 1.02	0.13	[77] −1.10	0.08	1.07	0.15
8	C8	[80] 0.78	0.12	[79] −0.97	0.07	1.24	0.20
9	C9	[82] 0.68	0.12	[81] −1.13	0.08	1.67	0.29
10	C10	[84] 0.99	0.13	[83] −0.99	0.08	1.01	0.15
11	C11	[86] 0.49	0.11	[85] −0.96	0.07	1.98	0.46
12	C12	[88] 0.77	0.12	[87] −1.61	0.09	2.09	0.33
13	C13	[90] 0.93	0.14	[89] −1.97	0.10	2.13	0.29
14	C14	[92] 0.26	0.11	[91] −1.34	0.08	5.16	2.26
15	C15	[94] 0.66	0.12	[93] −1.41	0.08	2.14	0.38
16	C16	[96] 0.55	0.11	[95] −1.34	0.08	2.45	0.52
17	C17	[98] 0.61	0.11	[97] −1.23	0.08	2.01	0.38
18	C18	[100] 0.63	0.12	[99] −1.29	0.08	2.03	0.38
19	C19	[102] 1.83	0.21	[101] −1.56	0.11	0.85	0.09
20	C20	[104] 0.88	0.12	[103] −1.06	0.08	1.20	0.18
21	C21	[106] 0.44	0.11	[105] −1.17	0.08	2.69	0.69
22	C22	[108] 1.04	0.13	[107] −1.18	0.08	1.13	0.15

23	C23	[110]	1.00	0.13	[109]	−1.29	0.08	1.30	0.18
24	C24	[112]	1.25	0.15	[111]	−1.24	0.09	0.99	0.12
25	C25	[114]	0.84	0.12	[113]	−1.38	0.08	1.64	0.24
26	C26	[116]	0.38	0.11	[115]	−1.06	0.07	2.79	0.81
27	C27	[118]	0.54	0.12	[117]	−1.63	0.09	3.02	0.66
28	C28	[120]	0.50	0.11	[119]	−0.61	0.07	1.22	0.30
29	C29	[122]	0.79	0.12	[121]	−1.00	0.08	1.26	0.20
30	C30	[124]	1.20	0.15	[123]	−0.36	0.07	0.30	0.08
31	C31	[126]	1.09	0.14	[125]	−0.91	0.08	0.84	0.12
32	C32	[128]	0.94	0.13	[127]	−1.06	0.08	1.13	0.16

Summed-Score Based Item Diagnostic Tables and X^2s for Group 2 (Back to TOC)

S-X^2 Item Level Diagnostic Statistics

Item	Label	X^2	*d.f.*	Probability
1	C1	35.18	22	0.0370
2	C2	46.18	21	0.0012
3	C3	41.64	21	0.0047
4	C4	42.15	22	0.0060
5	C5	34.90	23	0.0531
6	C6	23.89	20	0.2465
7	C7	33.96	21	0.0365
8	C8	30.33	21	0.0854
9	C9	34.41	22	0.0445
10	C10	27.33	21	0.1599
11	C11	34.73	22	0.0412
12	C12	30.00	20	0.0697
13	C13	43.63	22	0.0039
14	C14	39.16	20	0.0063
15	C15	35.67	23	0.0445
16	C16	35.97	22	0.0305
17	C17	32.65	22	0.0668
18	C18	38.43	21	0.0115
19	C19	35.01	18	0.0094
20	C20	43.86	20	0.0016
21	C21	35.33	23	0.0480
22	C22	42.52	20	0.0024
23	C23	24.00	19	0.1955
24	C24	44.92	20	0.0011
25	C25	29.19	22	0.1392

26	C26	33.64	21	0.0395
27	C27	40.97	22	0.0083
28	C28	25.42	21	0.2288
29	C29	35.29	21	0.0262
30	C30	29.66	18	0.0408
31	C31	30.40	20	0.0634
32	C32	19.83	20	0.4698

Group Parameter Estimates (Back to TOC)

Group	Label	μ	*s.e.*	σ^2	*s.e.*	σ	*s.e.*
1	G1	0.00	-----	1.00	-----	1.00	-----
2	G2	-0.12	-----	0.65	-----	0.80	-----

DIF Statistics for Graded Items (Back to TOC)

Item numbers in:										
Group 1	Group 2	Total X^2	*d.f.*	*p*	X^2_a	*d.f.*	*p*	X^2_{cla}	*d.f.*	*p*
1	1	0.1	2	0.9282	0.0	1	0.8620	0.1	1	0.7304
2	2	1.0	2	0.6067	0.7	1	0.4157	0.3	1	0.5619
3	3	0.1	2	0.9396	0.0	1	0.9192	0.1	1	0.7355
4	4	0.2	2	0.9261	0.0	1	0.8311	0.1	1	0.7426
5	5	3.8	2	0.1516	1.2	1	0.2687	2.6	1	0.1104
6	6	6.8	2	0.0325	0.1	1	0.7635	6.8	1	0.0093
7	7	1.5	2	0.4636	0.9	1	0.3468	0.7	1	0.4197
8	8	4.3	2	0.1140	3.9	1	0.0477	0.4	1	0.5160
9	9	2.8	2	0.2426	0.6	1	0.4584	2.3	1	0.1311
10	10	4.2	2	0.1207	0.6	1	0.4553	3.7	1	0.0554
11	11	1.2	2	0.5383	0.0	1	0.8413	1.2	1	0.2739
12	12	0.7	2	0.7139	0.5	1	0.4674	0.1	1	0.7031
13	13	3.9	2	0.1457	3.2	1	0.0741	0.7	1	0.4097
14	14	10.3	2	0.0058	0.3	1	0.5624	10.0	1	0.0016
15	15	3.0	2	0.2210	0.3	1	0.5791	2.7	1	0.1000
16	16	1.0	2	0.6134	0.9	1	0.3326	0.0	1	0.8467
17	17	0.0	2	0.9818	0.0	1	0.9841	0.0	1	0.8489
18	18	1.7	2	0.4225	1.6	1	0.2079	0.1	1	0.7140
19	19	9.8	2	0.0076	1.7	1	0.1917	8.0	1	0.0046
20	20	1.0	2	0.6151	0.7	1	0.3900	0.2	1	0.6299
21	21	7.9	2	0.0196	7.9	1	0.0050	0.0	1	0.9600
22	22	0.8	2	0.6547	0.8	1	0.3702	0.0	1	0.8355
23	23	0.5	2	0.7957	0.2	1	0.6965	0.3	1	0.5810
24	24	2.1	2	0.3530	0.2	1	0.6615	1.9	1	0.1695

25	25	3.2	2	0.2058	0.0	1	0.8694	3.1	1	0.0761
26	26	7.7	2	0.0213	2.2	1	0.1397	5.5	1	0.0189
27	27	2.0	2	0.3761	0.2	1	0.6969	1.8	1	0.1796
28	28	7.0	2	0.0296	0.3	1	0.5727	6.7	1	0.0095
29	29	0.8	2	0.6556	0.8	1	0.3859	0.1	1	0.7622
30	30	7.3	2	0.0261	0.3	1	0.5552	6.9	1	0.0084
31	31	2.6	2	0.2727	2.6	1	0.1075	0.0	1	0.9697
32	32	3.6	2	0.1657	3.5	1	0.0597	0.1	1	0.7972

Marginal fit (X^2) and Standardized LD X^2 Statistics for Group 1 (Back to TOC)

Item	Label	Marginal X^2	1	2	3	4	5	6	7	8	9	10
1	C1	0.1										
2	C2	0.0	-0.6									
3	C3	0.0	-0.0	-0.6								
4	C4	0.1	1.6	-0.5	-0.6							
5	C5	0.0	-0.6	0.5	-0.2	0.1						
6	C6	0.1	-0.2	-0.7	0.5	-0.4	1.4					
7	C7	0.9	0.2	1.0	1.4	0.2	2.9	1.8				
8	C8	0.6	3.5	-0.2	0.7	0.6	-0.2	0.7	0.5			
9	C9	0.1	0.8	-0.4	0.7	0.5	2.9	1.0	0.9	0.1		
10	C10	0.2	-0.5	-0.3	-0.5	-0.4	0.6	3.1	0.1	0.1	-0.5	
11	C11	0.0	1.5	1.6	-0.1	0.7	-0.3	-0.2	-0.1	0.1	-0.6	1.9
12	C12	0.0	6.6	1.8	0.7	0.2	0.4	2.1	0.3	0.4	0.5	-0.5
13	C13	0.0	0.1	-0.7	-0.6	3.6	0.9	-0.5	0.7	0.1	0.1	2.7
14	C14	0.0	0.6	-0.4	-0.5	5.9	-0.6	-0.5	-0.1	2.3	-0.2	0.1
15	C15	0.0	-0.6	-0.1	-0.5	-0.6	0.2	1.0	1.5	0.2	1.7	-0.5
16	C16	0.0	1.1	0.8	-0.6	-0.6	4.8	4.3	-0.1	0.3	1.1	-0.5
17	C17	0.0	0.4	1.2	2.1	0.3	-0.4	-0.4	0.8	-0.3	-0.6	-0.1
18	C18	0.1	3.8	-0.4	0.0	-0.5	2.4	0.8	0.1	2.2	0.5	2.2
19	C19	1.0	1.4	2.2	4.6	1.2	0.5	0.0	3.0	1.1	1.4	0.5
20	C20	0.4	3.4	-0.4	1.3	-0.1	1.6	0.9	0.1	-0.0	1.1	0.0
21	C21	0.1	1.0	-0.6	-0.5	-0.2	-0.2	3.1	2.1	1.0	0.0	0.1
22	C22	0.2	-0.0	-0.6	-0.4	-0.5	-0.0	0.0	-0.0	2.1	0.3	-0.1
23	C23	0.2	0.0	0.4	0.0	-0.5	2.3	1.4	0.2	0.5	-0.4	-0.2
24	C24	0.8	-0.0	0.3	-0.1	0.8	-0.1	0.3	0.4	0.1	0.1	0.0
25	C25	0.1	1.0	-0.6	2.4	1.5	-0.4	-0.6	0.2	-0.0	-0.5	-0.4
26	C26	0.0	1.3	0.2	0.1	0.0	0.3	-0.6	0.2	-0.2	2.5	-0.5
27	C27	0.0	0.3	4.7	-0.2	0.5	0.5	-0.2	1.2	-0.3	1.8	-0.2
28	C28	0.0	0.7	0.1	-0.7	-0.5	0.9	-0.0	0.1	0.2	-0.2	-0.6

29	C29	0.0	0.4	−0.3	−0.7	1.8	−0.6	−0.5	−0.1	−0.2	0.2	−0.1
30	C30	0.7	0.4	−0.1	0.9	−0.2	−0.1	0.1	0.3	0.4	0.4	0.3
31	C31	0.2	0.4	−0.6	0.2	−0.5	−0.6	−0.5	1.8	1.0	−0.0	−0.3
32	C32	0.0	0.7	1.1	−0.3	0.1	−0.2	−0.6	0.1	0.2	1.8	−0.6

Item	Label	Marginal X^2	11	12	13	14	15	16	17	18	19	20
11	C11	0.0										
12	C12	0.0	−0.6									
13	C13	0.0	−0.7	−0.0								
14	C14	0.0	−0.1	−0.6	4.5							
15	C15	0.0	0.7	0.7	5.2	−0.6						
16	C16	0.0	0.8	−0.5	−0.0	0.0	−0.3					
17	C17	0.0	0.3	1.1	−0.5	−0.3	4.3	−0.1				
18	C18	0.1	−0.6	2.2	−0.5	0.3	0.1	−0.1	−0.3			
19	C19	1.0	0.3	1.1	0.2	0.1	0.0	0.9	0.1	1.7		
20	C20	0.4	1.8	−0.2	1.0	−0.4	0.1	0.2	−0.3	−0.4	0.6	
21	C21	0.1	−0.5	−0.1	−0.3	0.1	−0.6	−0.3	−0.6	1.9	0.1	0.1
22	C22	0.2	−0.5	0.7	−0.1	−0.5	−0.4	1.6	0.1	1.2	0.6	2.1
23	C23	0.2	−0.5	−0.2	−0.5	−0.3	−0.5	−0.4	−0.1	2.0	0.4	−0.3
24	C24	0.8	−0.2	0.6	2.8	−0.1	−0.1	0.4	0.3	1.1	6.3	2.3
25	C25	0.1	−0.4	−0.6	0.3	−0.4	0.0	−0.6	4.4	0.1	2.7	−0.0
26	C26	0.0	1.7	−0.2	−0.6	0.2	−0.7	−0.6	0.1	0.8	3.4	−0.4
27	C27	0.0	−0.5	−0.1	−0.1	−0.6	−0.3	1.2	−0.1	−0.6	0.9	0.2
28	C28	0.0	0.5	−0.6	−0.7	1.5	−0.7	−0.3	0.7	−0.2	0.5	−0.1
29	C29	0.0	0.8	−0.5	−0.1	−0.6	−0.1	−0.4	−0.5	0.4	0.1	−0.3
30	C30	0.7	0.4	0.4	−0.2	0.9	−0.1	−0.1	0.4	0.6	3.5	0.7
31	C31	0.2	−0.5	−0.6	−0.6	−0.3	−0.1	−0.5	−0.5	−0.5	0.7	−0.1
32	C32	0.0	−0.6	−0.6	−0.3	−0.1	0.8	−0.7	1.0	−0.6	0.2	1.1

Item	Label	Marginal X^2	21	22	23	24	25	26	27	28	29	30
21	C21	0.1										
22	C22	0.2	0.4									
23	C23	0.2	1.0	−0.4								
24	C24	0.8	−0.1	2.7	0.5							
25	C25	0.1	1.5	−0.5	1.8	0.6						
26	C26	0.0	−0.6	−0.6	−0.5	0.1	−0.6					
27	C27	0.0	−0.0	−0.5	0.9	0.4	0.7	−0.7				
28	C28	0.0	0.2	0.5	−0.4	−0.1	−0.1	1.2	−0.5			

29	C29	0.0	0.2	0.6	0.5	0.4	−0.6	2.8	−0.1	−0.6		
30	C30	0.7	−0.2	1.9	1.5	0.3	−0.1	1.7	1.4	2.9	1.9	
31	C31	0.2	−0.1	0.3	−0.5	−0.0	−0.5	2.1	0.1	0.1	3.9	−0.1
32	C32	0.0	−0.6	−0.1	0.9	−0.1	0.9	0.3	1.5	−0.6	2.2	−0.2

		Marginal	
Item	Label	X^2	31
31	C31	0.2	
32	C32	0.0	0.5

Marginal fit (X^2) and Standardized LD X^2 Statistics for Group 2 (Back to TOC)

		Marginal										
Item	Label	X^2	1	2	3	4	5	6	7	8	9	10
1	C1	0.0										
2	C2	0.0	−0.7									
3	C3	0.0	−0.7	1.2								
4	C4	0.0	−0.5	1.2	−0.7							
5	C5	0.0	−0.6	−0.7	−0.7	1.3						
6	C6	0.0	−0.1	−0.7	−0.5	−0.3	−0.2					
7	C7	0.1	0.4	0.6	1.4	−0.6	−0.2	−0.1				
8	C8	0.0	−0.5	−0.7	0.4	2.4	−0.6	4.0	−0.2			
9	C9	0.0	0.1	−0.3	−0.3	−0.2	2.9	−0.1	−0.6	0.2		
10	C10	0.1	−0.5	0.9	−0.5	−0.6	−0.6	−0.6	−0.4	−0.5	0.5	
11	C11	0.0	−0.7	0.2	−0.6	0.2	3.2	−0.0	−0.6	−0.7	0.2	−0.4
12	C12	0.0	0.1	−0.5	−0.6	0.5	−0.6	2.1	1.7	−0.7	0.4	0.5
13	C13	0.0	2.1	−0.7	−0.7	−0.6	1.8	1.4	−0.4	0.5	0.0	−0.7
14	C14	0.0	−0.1	1.1	−0.7	−0.4	0.7	0.7	−0.6	0.3	−0.0	−0.5
15	C15	0.0	4.1	7.7	0.3	3.6	1.3	0.0	1.9	0.6	0.0	−0.2
16	C16	0.0	−0.7	2.2	−0.7	1.4	−0.7	−0.7	4.1	0.0	−0.6	2.0
17	C17	0.0	−0.7	3.8	−0.4	−0.5	−0.4	−0.7	−0.0	−0.6	−0.5	−0.6
18	C18	0.0	−0.6	−0.6	−0.7	1.6	−0.6	8.6	−0.1	2.1	1.4	−0.6
19	C19	0.7	2.0	−0.0	0.1	−0.1	0.3	−0.1	−0.1	0.4	−0.0	0.8
20	C20	0.0	−0.2	−0.4	−0.7	−0.2	0.9	−0.5	0.1	0.8	−0.7	1.9
21	C21	0.0	−0.6	−0.5	−0.2	−0.3	−0.2	1.0	−0.4	−0.6	−0.7	−0.5
22	C22	0.1	−0.6	−0.7	−0.3	−0.1	0.0	−0.6	1.1	−0.6	−0.0	−0.6
23	C23	0.0	−0.1	−0.6	−0.5	−0.7	0.9	0.7	−0.2	−0.6	−0.3	0.2
24	C24	0.2	0.8	−0.6	−0.3	−0.6	−0.3	−0.6	1.0	−0.4	−0.6	−0.1
25	C25	0.0	0.3	1.6	−0.7	0.2	−0.2	0.0	−0.6	−0.4	−0.6	1.8
26	C26	0.0	−0.7	−0.3	0.8	−0.3	−0.7	−0.7	0.2	−0.6	−0.2	0.3
27	C27	0.0	−0.0	0.2	0.5	−0.1	−0.1	1.6	−0.6	−0.7	−0.4	−0.2

28	C28	0.0	−0.0	−0.2	−0.7	−0.3	0.9	0.0	0.3	−0.7	−0.2	0.2
29	C29	0.0	0.2	4.1	−0.7	−0.7	−0.4	6.2	−0.5	0.2	−0.3	−0.5
30	C30	0.2	0.1	0.2	2.9	−0.1	−0.5	−0.1	−0.5	0.5	−0.5	−0.4
31	C31	0.1	1.0	0.0	0.2	−0.6	−0.6	−0.2	−0.1	0.0	−0.6	−0.5
32	C32	0.1	−0.2	−0.6	−0.7	−0.6	0.9	−0.7	1.2	1.7	−0.6	−0.1

Item	Label	Marginal X^2	11	12	13	14	15	16	17	18	19	20
11	C11	0.0										
12	C12	0.0	2.4									
13	C13	0.0	−0.0	1.8								
14	C14	0.0	−0.7	0.5	3.9							
15	C15	0.0	−0.3	−0.1	1.6	1.1						
16	C16	0.0	−0.1	−0.5	−0.4	−0.4	2.3					
17	C17	0.0	−0.6	−0.6	−0.7	−0.6	2.9	2.6				
18	C18	0.0	−0.7	−0.7	1.2	−0.2	−0.3	−0.2	−0.7			
19	C19	0.7	−0.0	−0.0	0.7	0.4	0.0	0.0	−0.1	−0.1		
20	C20	0.0	−0.5	−0.5	−0.5	−0.7	−0.6	−0.6	−0.7	−0.7	1.0	
21	C21	0.0	−0.6	−0.5	−0.6	−0.6	1.1	−0.4	−0.7	−0.7	−0.1	1.1
22	C22	0.1	2.2	−0.4	−0.3	−0.6	−0.5	−0.6	−0.2	−0.6	0.7	0.5
23	C23	0.0	−0.4	−0.6	−0.2	0.1	−0.7	−0.4	−0.7	−0.7	0.3	−0.6
24	C24	0.2	−0.5	−0.4	2.0	−0.2	−0.4	0.1	1.5	0.8	−0.1	−0.4
25	C25	0.0	−0.7	−0.4	2.6	−0.7	−0.7	0.0	1.4	0.1	−0.2	−0.3
26	C26	0.0	−0.4	0.6	−0.7	0.3	−0.1	−0.7	−0.7	−0.6	3.0	−0.2
27	C27	0.0	−0.3	−0.4	2.1	−0.1	−0.7	1.7	−0.6	−0.6	0.1	0.1
28	C28	0.0	−0.6	−0.5	1.2	2.5	0.6	−0.5	−0.4	−0.6	2.9	0.1
29	C29	0.0	−0.4	−0.6	−0.7	−0.4	−0.4	−0.4	0.2	−0.2	0.1	−0.0
30	C30	0.2	−0.5	−0.5	−0.4	1.2	0.3	1.9	0.9	−0.4	1.8	−0.5
31	C31	0.1	1.9	−0.6	0.0	0.0	−0.3	−0.6	−0.6	−0.0	0.1	0.0
32	C32	0.1	1.5	−0.7	1.5	3.6	2.2	0.5	−0.3	−0.6	−0.1	−0.6

Item	Label	Marginal X^2	21	22	23	24	25	26	27	28	29	30
21	C21	0.0										
22	C22	0.1	0.0									
23	C23	0.0	0.1	0.2								
24	C24	0.2	−0.6	0.8	1.4							
25	C25	0.0	0.8	0.2	0.0	2.1						
26	C26	0.0	1.1	0.1	0.1	−0.3	−0.7					
27	C27	0.0	−0.3	−0.6	−0.6	1.1	−0.7	−0.6				

28	C28	0.0	−0.7	−0.6	−0.6	−0.3	−0.5	1.3	−0.6			
29	C29	0.0	−0.0	−0.6	−0.1	−0.5	−0.4	−0.0	−0.2	−0.4		
30	C30	0.2	−0.1	1.4	−0.3	−0.4	0.2	0.6	−0.2	0.2	0.3	
31	C31	0.1	−0.6	−0.4	3.5	0.3	−0.2	−0.4	−0.4	0.5	1.6	−0.1
32	C32	0.1	0.3	−0.0	−0.6	−0.5	−0.3	−0.1	1.3	−0.2	−0.5	−0.5

Item	Label	Marginal X^2	31
31	C31	0.1	
32	C32	0.1	0.6

Item Information Function Values for Group 1 at 15 Values of θ from −2.8 to 2.8 (Back to TOC)

Item	Label	−2.8	−2.4	−2.0	−1.6	−1.2	−0.8	−0.4	−0.0	0.4	0.8	1.2	1.6	2.0	2.4	2.8
1	C1	0.01	0.01	0.02	0.03	0.05	0.06	0.09	0.12	0.16	0.20	0.23	0.25	0.25	0.23	0.20
2	C2	0.01	0.01	0.02	0.02	0.02	0.02	0.03	0.03	0.03	0.04	0.04	0.04	0.05	0.05	0.05
3	C3	0.02	0.02	0.03	0.03	0.04	0.05	0.06	0.08	0.09	0.10	0.11	0.11	0.11	0.11	0.11
4	C4	0.02	0.02	0.03	0.04	0.06	0.07	0.09	0.11	0.13	0.15	0.16	0.16	0.16	0.15	0.13
5	C5	0.02	0.03	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10	0.11	0.11	0.11	0.10	0.10
6	C6	0.01	0.02	0.02	0.03	0.05	0.06	0.08	0.11	0.14	0.16	0.18	0.20	0.20	0.18	0.16
7	C7	0.02	0.03	0.04	0.07	0.10	0.15	0.21	0.28	0.33	0.36	0.34	0.30	0.23	0.17	0.12
8	C8	0.02	0.03	0.04	0.06	0.09	0.13	0.18	0.24	0.28	0.31	0.31	0.28	0.23	0.18	0.13
9	C9	0.02	0.03	0.04	0.05	0.07	0.09	0.10	0.12	0.14	0.15	0.16	0.15	0.14	0.13	0.11
10	C10	0.02	0.03	0.04	0.05	0.07	0.09	0.11	0.13	0.16	0.18	0.19	0.18	0.17	0.15	0.13
11	C11	0.02	0.03	0.03	0.04	0.04	0.05	0.05	0.06	0.06	0.06	0.06	0.07	0.07	0.06	0.06
12	C12	0.01	0.02	0.02	0.03	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10	0.11	0.11	0.11
13	C13	0.01	0.01	0.02	0.02	0.03	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.09	0.10	0.10
14	C14	0.01	0.01	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.03	0.03	0.03	0.03	0.03	0.03
15	C15	0.02	0.02	0.03	0.03	0.04	0.04	0.05	0.06	0.07	0.07	0.08	0.08	0.08	0.08	0.08
16	C16	0.02	0.02	0.03	0.03	0.04	0.05	0.06	0.08	0.09	0.10	0.11	0.12	0.12	0.12	0.11
17	C17	0.02	0.02	0.03	0.03	0.04	0.05	0.06	0.07	0.07	0.08	0.09	0.09	0.09	0.09	0.09
18	C18	0.02	0.02	0.03	0.04	0.05	0.07	0.09	0.11	0.13	0.15	0.16	0.17	0.17	0.15	0.14
19	C19	0.01	0.01	0.02	0.03	0.06	0.10	0.16	0.26	0.39	0.51	0.56	0.52	0.41	0.29	0.18
20	C20	0.02	0.03	0.04	0.06	0.08	0.11	0.15	0.19	0.23	0.26	0.26	0.25	0.22	0.17	0.13
21	C21	0.02	0.03	0.03	0.05	0.06	0.08	0.10	0.13	0.15	0.17	0.18	0.18	0.17	0.15	0.13
22	C22	0.02	0.03	0.04	0.05	0.07	0.09	0.12	0.15	0.17	0.19	0.20	0.19	0.18	0.16	0.13
23	C23	0.02	0.03	0.04	0.05	0.07	0.09	0.12	0.15	0.18	0.20	0.22	0.21	0.20	0.17	0.14
24	C24	0.02	0.03	0.04	0.07	0.10	0.14	0.20	0.26	0.31	0.34	0.33	0.29	0.23	0.17	0.12
25	C25	0.02	0.03	0.04	0.05	0.06	0.08	0.10	0.12	0.14	0.16	0.17	0.17	0.16	0.14	0.12
26	C26	0.02	0.02	0.02	0.03	0.03	0.04	0.05	0.06	0.06	0.07	0.08	0.08	0.08	0.09	0.08

27	C27	0.01	0.01	0.02	0.02	0.02	0.03	0.03	0.04	0.04	0.04	0.05	0.05	0.05	0.06	0.06
28	C28	0.02	0.03	0.03	0.04	0.05	0.06	0.06	0.07	0.08	0.08	0.08	0.08	0.08	0.08	0.07
29	C29	0.02	0.03	0.04	0.04	0.05	0.06	0.07	0.09	0.10	0.10	0.11	0.11	0.11	0.10	0.09
30	C30	0.03	0.04	0.06	0.09	0.13	0.18	0.22	0.27	0.29	0.29	0.27	0.22	0.17	0.13	0.09
31	C31	0.03	0.03	0.05	0.06	0.08	0.10	0.12	0.14	0.15	0.17	0.17	0.16	0.15	0.13	0.11
32	C32	0.02	0.03	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10	0.10	0.10	0.10	0.10	0.09
	Test Information:	1.56	1.75	2.01	2.36	2.81	3.37	4.05	4.79	5.51	6.05	6.28	6.15	5.71	5.12	4.48
	Expected s.e.:	0.80	0.76	0.70	0.65	0.60	0.54	0.50	0.46	0.43	0.41	0.40	0.40	0.42	0.44	0.47

Marginal Reliability for Response Pattern Scores: 0.76

Item Information Function Values for Group 2 at 15 Values of θ from −2.8 to 2.8 (Back to TOC)

Item	Label	−2.8	−2.4	−2.0	−1.6	−1.2	−0.8	−0.4	−0.0	0.4	0.8	1.2	1.6	2.0	2.4	2.8
1	C1	0.01	0.01	0.02	0.03	0.04	0.06	0.08	0.11	0.14	0.18	0.21	0.23	0.23	0.22	0.20
2	C2	0.01	0.02	0.02	0.02	0.03	0.03	0.04	0.05	0.05	0.06	0.07	0.07	0.08	0.08	0.08
3	C3	0.02	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10	0.11	0.12	0.12	0.12	0.11
4	C4	0.02	0.02	0.03	0.04	0.05	0.07	0.08	0.10	0.12	0.13	0.14	0.15	0.15	0.14	0.12
5	C5	0.02	0.02	0.02	0.03	0.03	0.03	0.04	0.04	0.05	0.05	0.06	0.06	0.06	0.06	0.06
6	C6	0.01	0.01	0.02	0.02	0.03	0.04	0.06	0.08	0.10	0.12	0.14	0.16	0.17	0.18	0.17
7	C7	0.02	0.03	0.04	0.06	0.09	0.12	0.16	0.20	0.23	0.26	0.26	0.24	0.21	0.17	0.13
8	C8	0.02	0.03	0.04	0.05	0.07	0.09	0.10	0.12	0.14	0.15	0.15	0.15	0.14	0.13	0.11
9	C9	0.02	0.03	0.03	0.04	0.05	0.06	0.07	0.08	0.10	0.11	0.11	0.11	0.11	0.11	0.10
10	C10	0.02	0.03	0.05	0.06	0.09	0.12	0.16	0.19	0.22	0.24	0.24	0.22	0.19	0.16	0.12
11	C11	0.02	0.02	0.03	0.03	0.03	0.04	0.04	0.05	0.05	0.05	0.06	0.06	0.06	0.06	0.06
12	C12	0.01	0.02	0.02	0.03	0.04	0.05	0.07	0.08	0.10	0.12	0.13	0.14	0.15	0.15	0.14
13	C13	0.01	0.01	0.02	0.03	0.04	0.05	0.07	0.09	0.12	0.15	0.18	0.20	0.21	0.21	0.19
14	C14	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.02
15	C15	0.02	0.02	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10	0.11	0.11	0.11	0.10
16	C16	0.02	0.02	0.02	0.03	0.03	0.04	0.04	0.05	0.06	0.06	0.07	0.07	0.07	0.07	0.07
17	C17	0.02	0.02	0.03	0.03	0.04	0.05	0.06	0.07	0.07	0.08	0.09	0.09	0.09	0.09	0.09
18	C18	0.02	0.02	0.03	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.09	0.10	0.10	0.10	0.09
19	C19	0.00	0.01	0.02	0.04	0.07	0.15	0.28	0.48	0.71	0.84	0.76	0.54	0.32	0.17	0.09
20	C20	0.02	0.03	0.04	0.06	0.07	0.10	0.12	0.15	0.17	0.19	0.20	0.19	0.17	0.15	0.12
21	C21	0.01	0.02	0.02	0.02	0.02	0.03	0.03	0.03	0.04	0.04	0.04	0.04	0.05	0.05	0.05
22	C22	0.02	0.03	0.04	0.06	0.08	0.11	0.15	0.20	0.24	0.26	0.27	0.26	0.22	0.18	0.14
23	C23	0.02	0.02	0.03	0.05	0.07	0.10	0.13	0.17	0.20	0.23	0.25	0.24	0.22	0.19	0.15
24	C24	0.01	0.02	0.04	0.06	0.09	0.14	0.20	0.27	0.34	0.39	0.38	0.34	0.27	0.20	0.13
25	C25	0.02	0.02	0.03	0.04	0.05	0.07	0.09	0.11	0.14	0.16	0.17	0.18	0.17	0.16	0.14

26	C26	0.01	0.02	0.02	0.02	0.02	0.02	0.03	0.03	0.03	0.03	0.03	0.03	0.04	0.04	0.04
27	C27	0.01	0.01	0.02	0.02	0.02	0.03	0.03	0.04	0.05	0.05	0.06	0.06	0.07	0.07	0.07
28	C28	0.03	0.03	0.03	0.04	0.04	0.05	0.05	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.05
29	C29	0.02	0.03	0.04	0.05	0.07	0.09	0.10	0.12	0.14	0.15	0.16	0.15	0.14	0.13	0.11
30	C30	0.03	0.05	0.08	0.12	0.18	0.24	0.30	0.35	0.36	0.33	0.27	0.21	0.15	0.10	0.07
31	C31	0.02	0.03	0.05	0.07	0.11	0.15	0.20	0.24	0.28	0.30	0.29	0.25	0.20	0.16	0.11
32	C32	0.02	0.03	0.04	0.06	0.08	0.11	0.14	0.17	0.20	0.22	0.22	0.21	0.19	0.16	0.13
	Test Information:	2.08	2.27	2.52	2.87	3.33	3.93	4.67	5.51	6.32	6.85	6.94	6.63	6.10	5.50	4.91
	Expected s.e.:	0.69	0.66	0.63	0.59	0.55	0.50	0.46	0.43	0.40	0.38	0.38	0.39	0.40	0.43	0.45

Marginal Reliability for Response Pattern Scores: 0.68

Likelihood-based Values and Goodness of Fit Statistics (Back to TOC)

Statistics based on the loglikelihood	
-2loglikelihood:	66262.86
Akaike Information Criterion (AIC):	66518.86
Bayesian Information Criterion (BIC):	67235.77

Statistics based on the full item x item x ... classification
The table is too sparse to compute the general multinomial goodness of fit statistics.

Statistics based on one- and two-way marginal tables
The M_2 statistics were not requested.

Summary of the Data and Control Parameters (Back to TOC)

Group:	Group 1	Group 2
Sample Size	1000	1000
Number of Items	32	32
Number of Dimensions	1	1

Group 1

Item	Label	Categories	Model
1	C1	2	2PL
2	C2	2	2PL
3	C3	2	2PL
4	C4	2	2PL
5	C5	2	2PL
6	C6	2	2PL

7	C7	2	2PL
8	C8	2	2PL
9	C9	2	2PL
10	C10	2	2PL
11	C11	2	2PL
12	C12	2	2PL
13	C13	2	2PL
14	C14	2	2PL
15	C15	2	2PL
16	C16	2	2PL
17	C17	2	2PL
18	C18	2	2PL
19	C19	2	2PL
20	C20	2	2PL
21	C21	2	2PL
22	C22	2	2PL
23	C23	2	2PL
24	C24	2	2PL
25	C25	2	2PL
26	C26	2	2PL
27	C27	2	2PL
28	C28	2	2PL
29	C29	2	2PL
30	C30	2	2PL
31	C31	2	2PL
32	C32	2	2PL

Group 2

Item	Label	Categories	Model
1	C1	2	2PL
2	C2	2	2PL
3	C3	2	2PL
4	C4	2	2PL
5	C5	2	2PL
6	C6	2	2PL
7	C7	2	2PL
8	C8	2	2PL
9	C9	2	2PL
10	C10	2	2PL
11	C11	2	2PL

12	C12	2	2PL
13	C13	2	2PL
14	C14	2	2PL
15	C15	2	2PL
16	C16	2	2PL
17	C17	2	2PL
18	C18	2	2PL
19	C19	2	2PL
20	C20	2	2PL
21	C21	2	2PL
22	C22	2	2PL
23	C23	2	2PL
24	C24	2	2PL
25	C25	2	2PL
26	C26	2	2PL
27	C27	2	2PL
28	C28	2	2PL
29	C29	2	2PL
30	C30	2	2PL
31	C31	2	2PL
32	C32	2	2PL

Parameter Estimation Control Values

Bock–Aitkin EM Algorithm		
Maximum number of cycles:	500	
Convergence criterion:	1.00e–005	
Maximum number of M–step iterations:	50	
Convergence criterion for iterative M–steps:	1.00e–006	
Number of rectangular quadrature points:	49	
Minimum, Maximum quadrature points:	–6.00	6.00
SEM algorithm tolerance:	1.00e–003	
Standard error computation algorithm:	Supplemented EM	

DIF Analysis

All items are evaluated for DIF

(Conditional on population distribution estimates obtained with all items constrained equal)

Contrasts among groups:

Group:	1	2
Contrast 1	1.000	–1.000

Miscellaneous Control Values

Print parameter numbers?	Yes

Z tolerance, max. abs. logit value:	50.00
Number of processor cores used:	8
Number of cycles completed:	40
Maximum parameter change:	0.00e+000
Number of free parameters:	128

Processing times (in seconds)

E-step computations:	0.11
M-step computations:	0.03
Standard error computations:	1.78
Goodness-of-fit statistics:	0.13
Total:	2.04

Output Files

HTML results and control parameters:	F:\เตรียมข้อมูล IRT PRO\04 math\04 math.Test1-irt.htm

Convergence and Numerical Stability

Engine status:	Normal termination
SEM algorithm status:	Normal
First-order test:	Convergence criteria satisfied
Condition number of information matrix:	9.92e+000
Second-order test:	Solution is a possible local maximum

ภาคผนวก ซ
ตัวอย่าง Print Out ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ในแบบทดสอบ O-NET ด้วยวิธี SIBTEST โดยใช้โปรแกรม SIBTEST

```
name of input parameter file = sib.in
 number of items on test =  32
 name of file for Ref. grp. scores = f:\àµÃÕÂÁ¢éÍÁÙÅ SIBTEST\04 math\Mmath.dat
 name of file for Focal grp. scores = f:\àµÃÕÂÁ¢éÍÁÙÅ SIBTEST\04 math\FMmath.dat
 minimum no. of examinees per matching score cell =   2
 number of runs for this data set =  32
 number of examinees in Reference group =   1000
 number of examinees in Focal group     =   1000

 Examinee Test Score Summary Statistics

     Reference Group:               Mean =   8.03
                     Standard deviation =   5.24
     Focal Group:                   Mean =   7.30
                     Standard deviation =   4.29

     Standardized Score Difference =  0.15

 Item Statistics
    # = item number
    p = proportion right on item (Classical Test Theory p-value)
    r = point biserial (item score-test score correlation)

    #:     1     2     3     4     5     6     7     8     9    10
    p: 0.162 0.187 0.220 0.229 0.240 0.164 0.273 0.277 0.266 0.263
    r: 0.406 0.264 0.322 0.371 0.306 0.383 0.443 0.410 0.366 0.394

    #:    11    12    13    14    15    16    17    18    19    20
    p: 0.290 0.176 0.147 0.238 0.219 0.213 0.234 0.221 0.199 0.264
    r: 0.284 0.328 0.328 0.219 0.314 0.312 0.318 0.348 0.523 0.405

    #:    21    22    23    24    25    26    27    28    29    30
    p: 0.242 0.254 0.237 0.262 0.231 0.237 0.182 0.324 0.273 0.375
    r: 0.328 0.402 0.405 0.456 0.372 0.260 0.264 0.279 0.341 0.415

    #:    31    32
    p: 0.296 0.270
    r: 0.400 0.363

$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$
 OUTPUT FOR RUN NUMBER   1          OUTPUT FOR RUN NUMBER   1

 Suspect subtest items:
   1

 Matching subtest items:
    2   3   4   5   6   7   8   9  10  11
   12  13  14  15  16  17  18  19  20  21
   22  23  24  25  26  27  28  29  30  31
   32
 estimate of guessing on this matching subtest = 0.20

 proportion of Ref. grp. examinees eliminated =  0.504
 proportion of Focal grp. examinees eliminated =  0.537
 KR-20 for Ref. grp. =   0.956
 KR-20 for Foc. grp. =   0.955

 Matching Subtest Summary Statistics
     Reference Group:               Mean =   7.85
                     Standard deviation =  5.08
     Focal Group:                   Mean =   7.16
                     Standard deviation =  4.18

     Standardized Score Difference =  0.15


                SIBTEST-Focal weighting Results
                                              p-value
  Beta estimate   standard error   for DIF against either grp.
      0.018           0.023                 0.445935
```

```
                    SIBTEST error flag = 0
   No errors.  This SIBTEST run had a normal successful completion.

$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$
 OUTPUT FOR RUN NUMBER   2          OUTPUT FOR RUN NUMBER   2

 Suspect subtest items:
   2

 Matching subtest items:
   1   3   4   5   6   7   8   9  10  11
  12  13  14  15  16  17  18  19  20  21
  22  23  24  25  26  27  28  29  30  31
  32
 estimate of guessing on this matching subtest = 0.20

 proportion of Ref. grp. examinees eliminated =  0.522
 proportion of Focal grp. examinees eliminated =  0.531
 KR-20 for Ref. grp. =   0.958
 KR-20 for Foc. grp. =   0.956

 Matching Subtest Summary Statistics
     Reference Group:              Mean =   7.83
                     Standard deviation =  5.15
     Focal Group:                  Mean =   7.13
                     Standard deviation =  4.21

     Standardized Score Difference =  0.15


                SIBTEST-Focal weighting Results
                                          p-value
  Beta estimate   standard error   for DIF against either grp.
     -0.012           0.025               0.640424

                    SIBTEST error flag = 0
   No errors.  This SIBTEST run had a normal successful completion.

$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$
 OUTPUT FOR RUN NUMBER   3          OUTPUT FOR RUN NUMBER   3

 Suspect subtest items:
   3

 Matching subtest items:
   1   2   4   5   6   7   8   9  10  11
  12  13  14  15  16  17  18  19  20  21
  22  23  24  25  26  27  28  29  30  31
  32
 estimate of guessing on this matching subtest = 0.20

 proportion of Ref. grp. examinees eliminated =  0.527
 proportion of Focal grp. examinees eliminated =  0.542
 KR-20 for Ref. grp. =   0.958
 KR-20 for Foc. grp. =   0.957

 Matching Subtest Summary Statistics
     Reference Group:              Mean =   7.80
                     Standard deviation =  5.11
     Focal Group:                  Mean =   7.09
                     Standard deviation =  4.19

     Standardized Score Difference =  0.15


                SIBTEST-Focal weighting Results
                                          p-value
  Beta estimate   standard error   for DIF against either grp.
      0.003           0.028               0.904868

                    SIBTEST error flag = 0
   No errors.  This SIBTEST run had a normal successful completion.
```

```
$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$
 OUTPUT FOR RUN NUMBER   4          OUTPUT FOR RUN NUMBER   4

 Suspect subtest items:
   4

 Matching subtest items:
   1   2   3   5   6   7   8   9  10  11
  12  13  14  15  16  17  18  19  20  21
  22  23  24  25  26  27  28  29  30  31
  32
 estimate of guessing on this matching subtest = 0.20

 proportion of Ref. grp. examinees eliminated =  0.516
 proportion of Focal grp. examinees eliminated =  0.540
 KR-20 for Ref. grp. =   0.958
 KR-20 for Foc. grp. =   0.956

 Matching Subtest Summary Statistics
     Reference Group:              Mean =   7.78
                       Standard deviation =  5.08
     Focal Group:                  Mean =   7.09
                       Standard deviation =  4.17

     Standardized Score Difference =  0.15


                SIBTEST-Focal weighting Results
                                              p-value
  Beta estimate   standard error   for DIF against either grp.
      0.004           0.027                 0.884942

                     SIBTEST error flag = 0
    No errors.  This SIBTEST run had a normal successful completion.

$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$
 OUTPUT FOR RUN NUMBER   5          OUTPUT FOR RUN NUMBER   5

 Suspect subtest items:
   5

 Matching subtest items:
   1   2   3   4   6   7   8   9  10  11
  12  13  14  15  16  17  18  19  20  21
  22  23  24  25  26  27  28  29  30  31
  32
 estimate of guessing on this matching subtest = 0.20

 proportion of Ref. grp. examinees eliminated =  0.519
 proportion of Focal grp. examinees eliminated =  0.548
 KR-20 for Ref. grp. =   0.960
 KR-20 for Foc. grp. =   0.958

 Matching Subtest Summary Statistics
     Reference Group:              Mean =   7.76
                       Standard deviation =  5.10
     Focal Group:                  Mean =   7.09
                       Standard deviation =  4.21

     Standardized Score Difference =  0.14


                SIBTEST-Focal weighting Results
                                              p-value
  Beta estimate   standard error   for DIF against either grp.
     -0.010           0.028                 0.715317

                     SIBTEST error flag = 0
    No errors.  This SIBTEST run had a normal successful completion.

$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$
 OUTPUT FOR RUN NUMBER   6          OUTPUT FOR RUN NUMBER   6
```

```
 Suspect subtest items:
   6

 Matching subtest items:
   1   2   3   4   5   7   8   9  10  11
  12  13  14  15  16  17  18  19  20  21
  22  23  24  25  26  27  28  29  30  31
  32
 estimate of guessing on this matching subtest = 0.20

 proportion of Ref. grp. examinees eliminated =  0.515
 proportion of Focal grp. examinees eliminated =  0.538
 KR-20 for Ref. grp. =   0.956
 KR-20 for Foc. grp. =   0.956

 Matching Subtest Summary Statistics
     Reference Group:              Mean =   7.83
                       Standard deviation =  5.08
     Focal Group:                  Mean =   7.17
                       Standard deviation =  4.20

     Standardized Score Difference =  0.14


                SIBTEST-Focal weighting Results
                                          p-value
  Beta estimate   standard error   for DIF against either grp.
      0.057           0.024                0.018186

                     SIBTEST error flag = 0
    No errors.  This SIBTEST run had a normal successful completion.

$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$
 OUTPUT FOR RUN NUMBER   7          OUTPUT FOR RUN NUMBER   7

 Suspect subtest items:
   7

 Matching subtest items:
   1   2   3   4   5   6   8   9  10  11
  12  13  14  15  16  17  18  19  20  21
  22  23  24  25  26  27  28  29  30  31
  32
 estimate of guessing on this matching subtest = 0.20

 proportion of Ref. grp. examinees eliminated =  0.525
 proportion of Focal grp. examinees eliminated =  0.557
 KR-20 for Ref. grp. =   0.959
 KR-20 for Foc. grp. =   0.957

 Matching Subtest Summary Statistics
     Reference Group:              Mean =   7.73
                       Standard deviation =  5.03
     Focal Group:                  Mean =   7.05
                       Standard deviation =  4.14

     Standardized Score Difference =  0.15


                SIBTEST-Focal weighting Results
                                          p-value
  Beta estimate   standard error   for DIF against either grp.
      0.012           0.029                0.683634

                     SIBTEST error flag = 0
    No errors.  This SIBTEST run had a normal successful completion.

$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$
 OUTPUT FOR RUN NUMBER   8          OUTPUT FOR RUN NUMBER   8

 Suspect subtest items:
   8
```

```
 Matching subtest items:
   1   2   3   4   5   6   7   9  10  11
  12  13  14  15  16  17  18  19  20  21
  22  23  24  25  26  27  28  29  30  31
  32
 estimate of guessing on this matching subtest = 0.20

 proportion of Ref. grp. examinees eliminated =  0.521
 proportion of Focal grp. examinees eliminated =  0.548
 KR-20 for Ref. grp. =   0.959
 KR-20 for Foc. grp. =   0.959

 Matching Subtest Summary Statistics
     Reference Group:               Mean =   7.74
                     Standard deviation =  5.04
     Focal Group:                   Mean =   7.03
                     Standard deviation =  4.16

     Standardized Score Difference =  0.15


                SIBTEST-Focal weighting Results
                                              p-value
  Beta estimate   standard error   for DIF against either grp.
     -0.002           0.029                 0.933892

                     SIBTEST error flag = 0
    No errors.  This SIBTEST run had a normal successful completion.

$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$
 OUTPUT FOR RUN NUMBER   9          OUTPUT FOR RUN NUMBER   9

 Suspect subtest items:
   9

 Matching subtest items:
   1   2   3   4   5   6   7   8  10  11
  12  13  14  15  16  17  18  19  20  21
  22  23  24  25  26  27  28  29  30  31
  32
 estimate of guessing on this matching subtest = 0.20

 proportion of Ref. grp. examinees eliminated =  0.526
 proportion of Focal grp. examinees eliminated =  0.549
 KR-20 for Ref. grp. =   0.960
 KR-20 for Foc. grp. =   0.958

 Matching Subtest Summary Statistics
     Reference Group:               Mean =   7.74
                     Standard deviation =  5.07
     Focal Group:                   Mean =   7.06
                     Standard deviation =  4.18

     Standardized Score Difference =  0.15


                SIBTEST-Focal weighting Results
                                              p-value
  Beta estimate   standard error   for DIF against either grp.
      0.027           0.028                 0.331682

                     SIBTEST error flag = 0
    No errors.  This SIBTEST run had a normal successful completion.

$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$
 OUTPUT FOR RUN NUMBER  10          OUTPUT FOR RUN NUMBER  10

 Suspect subtest items:
  10

 Matching subtest items:
   1   2   3   4   5   6   7   8   9  11
  12  13  14  15  16  17  18  19  20  21
```

```
 22  23  24  25  26  27  28  29  30  31
 32
estimate of guessing on this matching subtest = 0.20

proportion of Ref. grp. examinees eliminated =  0.523
proportion of Focal grp. examinees eliminated =  0.553
KR-20 for Ref. grp. =   0.959
KR-20 for Foc. grp. =   0.959

Matching Subtest Summary Statistics
    Reference Group:               Mean =   7.77
                     Standard deviation =  5.07
    Focal Group:                   Mean =   7.03
                     Standard deviation =  4.15

    Standardized Score Difference =  0.16


               SIBTEST-Focal weighting Results
                                          p-value
 Beta estimate   standard error   for DIF against either grp.
    -0.046           0.029                0.115008

                    SIBTEST error flag = 0
   No errors.  This SIBTEST run had a normal successful completion.

$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$
OUTPUT FOR RUN NUMBER  11          OUTPUT FOR RUN NUMBER  11

Suspect subtest items:
 11

Matching subtest items:
  1   2   3   4   5   6   7   8   9  10
 12  13  14  15  16  17  18  19  20  21
 22  23  24  25  26  27  28  29  30  31
 32
estimate of guessing on this matching subtest = 0.20

proportion of Ref. grp. examinees eliminated =  0.523
proportion of Focal grp. examinees eliminated =  0.555
KR-20 for Ref. grp. =   0.962
KR-20 for Foc. grp. =   0.961

Matching Subtest Summary Statistics
    Reference Group:               Mean =   7.72
                     Standard deviation =  5.11
    Focal Group:                   Mean =   7.03
                     Standard deviation =  4.20

    Standardized Score Difference =  0.15


               SIBTEST-Focal weighting Results
                                          p-value
 Beta estimate   standard error   for DIF against either grp.
     0.026           0.031                0.399648

                    SIBTEST error flag = 0
   No errors.  This SIBTEST run had a normal successful completion.

$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$
OUTPUT FOR RUN NUMBER  12          OUTPUT FOR RUN NUMBER  12

Suspect subtest items:
 12

Matching subtest items:
  1   2   3   4   5   6   7   8   9  10
 11  13  14  15  16  17  18  19  20  21
 22  23  24  25  26  27  28  29  30  31
 32
estimate of guessing on this matching subtest = 0.20
```

```
 proportion of Ref. grp. examinees eliminated =  0.513
 proportion of Focal grp. examinees eliminated =  0.535
 KR-20 for Ref. grp. =   0.957
 KR-20 for Foc. grp. =   0.955

 Matching Subtest Summary Statistics
     Reference Group:              Mean =   7.85
                       Standard deviation =  5.12
     Focal Group:                  Mean =   7.13
                       Standard deviation =  4.19

     Standardized Score Difference =  0.15


                SIBTEST-Focal weighting Results
                                           p-value
  Beta estimate   standard error   for DIF against either grp.
     -0.024           0.025                0.330065

                     SIBTEST error flag = 0
    No errors.  This SIBTEST run had a normal successful completion.

$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$
 OUTPUT FOR RUN NUMBER  13          OUTPUT FOR RUN NUMBER  13

 Suspect subtest items:
  13

 Matching subtest items:
   1   2   3   4   5   6   7   8   9  10
  11  12  14  15  16  17  18  19  20  21
  22  23  24  25  26  27  28  29  30  31
  32
 estimate of guessing on this matching subtest = 0.20

 proportion of Ref. grp. examinees eliminated =  0.509
 proportion of Focal grp. examinees eliminated =  0.526
 KR-20 for Ref. grp. =   0.958
 KR-20 for Foc. grp. =   0.956

 Matching Subtest Summary Statistics
     Reference Group:              Mean =   7.86
                       Standard deviation =  5.13
     Focal Group:                  Mean =   7.17
                       Standard deviation =  4.19

     Standardized Score Difference =  0.15


                SIBTEST-Focal weighting Results
                                           p-value
  Beta estimate   standard error   for DIF against either grp.
      0.010           0.024                0.658992

                     SIBTEST error flag = 0
    No errors.  This SIBTEST run had a normal successful completion.

$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$
 OUTPUT FOR RUN NUMBER  14          OUTPUT FOR RUN NUMBER  14

 Suspect subtest items:
  14

 Matching subtest items:
   1   2   3   4   5   6   7   8   9  10
  11  12  13  15  16  17  18  19  20  21
  22  23  24  25  26  27  28  29  30  31
  32
 estimate of guessing on this matching subtest = 0.20

 proportion of Ref. grp. examinees eliminated =  0.527
 proportion of Focal grp. examinees eliminated =  0.542
```

```
 KR-20 for Ref. grp. =   0.961
 KR-20 for Foc. grp. =   0.958

 Matching Subtest Summary Statistics
     Reference Group:               Mean =   7.76
                      Standard deviation =  5.15
     Focal Group:                   Mean =   7.10
                      Standard deviation =  4.24

     Standardized Score Difference =  0.14


                SIBTEST-Focal weighting Results
                                          p-value
  Beta estimate   standard error   for DIF against either grp.
      0.033           0.027                0.233014

                     SIBTEST error flag = 0
    No errors.  This SIBTEST run had a normal successful completion.

$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$
 OUTPUT FOR RUN NUMBER  15          OUTPUT FOR RUN NUMBER  15

 Suspect subtest items:
  15

 Matching subtest items:
    1   2   3   4   5   6   7   8   9  10
  11  12  13  14  16  17  18  19  20  21
  22  23  24  25  26  27  28  29  30  31
  32
 estimate of guessing on this matching subtest = 0.20

 proportion of Ref. grp. examinees eliminated =  0.523
 proportion of Focal grp. examinees eliminated =  0.539
 KR-20 for Ref. grp. =   0.959
 KR-20 for Foc. grp. =   0.956

 Matching Subtest Summary Statistics
     Reference Group:               Mean =   7.78
                      Standard deviation =  5.11
     Focal Group:                   Mean =   7.11
                      Standard deviation =  4.19

     Standardized Score Difference =  0.14


                SIBTEST-Focal weighting Results
                                          p-value
  Beta estimate   standard error   for DIF against either grp.
      0.008           0.027                0.753179

                     SIBTEST error flag = 0
    No errors.  This SIBTEST run had a normal successful completion.

$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$
 OUTPUT FOR RUN NUMBER  16          OUTPUT FOR RUN NUMBER  16

 Suspect subtest items:
  16

 Matching subtest items:
    1   2   3   4   5   6   7   8   9  10
  11  12  13  14  15  17  18  19  20  21
  22  23  24  25  26  27  28  29  30  31
  32
 estimate of guessing on this matching subtest = 0.20

 proportion of Ref. grp. examinees eliminated =  0.534
 proportion of Focal grp. examinees eliminated =  0.546
 KR-20 for Ref. grp. =   0.958
 KR-20 for Foc. grp. =   0.956
```

```
 Matching Subtest Summary Statistics
     Reference Group:              Mean =   7.81
                       Standard deviation =  5.11
     Focal Group:                  Mean =   7.10
                       Standard deviation =  4.20

     Standardized Score Difference =  0.15


                SIBTEST-Focal weighting Results
                                          p-value
  Beta estimate   standard error   for DIF against either grp.
      0.007          0.026               0.778134

                     SIBTEST error flag = 0
    No errors.  This SIBTEST run had a normal successful completion.

$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$
 OUTPUT FOR RUN NUMBER  17          OUTPUT FOR RUN NUMBER  17

 Suspect subtest items:
  17

 Matching subtest items:
    1   2   3   4   5   6   7   8   9  10
  11  12  13  14  15  16  18  19  20  21
  22  23  24  25  26  27  28  29  30  31
  32
 estimate of guessing on this matching subtest = 0.20

 proportion of Ref. grp. examinees eliminated =  0.519
 proportion of Focal grp. examinees eliminated =  0.541
 KR-20 for Ref. grp. =   0.959
 KR-20 for Foc. grp. =   0.957

 Matching Subtest Summary Statistics
     Reference Group:              Mean =   7.78
                       Standard deviation =  5.11
     Focal Group:                  Mean =   7.08
                       Standard deviation =  4.19

     Standardized Score Difference =  0.15


                SIBTEST-Focal weighting Results
                                          p-value
  Beta estimate   standard error   for DIF against either grp.
     -0.024          0.027               0.380712

                     SIBTEST error flag = 0
    No errors.  This SIBTEST run had a normal successful completion.

$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$
 OUTPUT FOR RUN NUMBER  18          OUTPUT FOR RUN NUMBER  18

 Suspect subtest items:
  18

 Matching subtest items:
    1   2   3   4   5   6   7   8   9  10
  11  12  13  14  15  16  17  19  20  21
  22  23  24  25  26  27  28  29  30  31
  32
 estimate of guessing on this matching subtest = 0.20

 proportion of Ref. grp. examinees eliminated =  0.524
 proportion of Focal grp. examinees eliminated =  0.544
 KR-20 for Ref. grp. =   0.958
 KR-20 for Foc. grp. =   0.957

 Matching Subtest Summary Statistics
     Reference Group:              Mean =   7.80
                       Standard deviation =  5.08
```

```
    Focal Group:                    Mean =   7.09
                        Standard deviation =  4.19

    Standardized Score Difference =  0.15


               SIBTEST-Focal weighting Results
                                              p-value
 Beta estimate   standard error   for DIF against either grp.
    -0.012           0.026                 0.641971

                     SIBTEST error flag = 0
   No errors.  This SIBTEST run had a normal successful completion.

$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$
 OUTPUT FOR RUN NUMBER  19          OUTPUT FOR RUN NUMBER  19

 Suspect subtest items:
  19

 Matching subtest items:
   1   2   3   4   5   6   7   8   9  10
  11  12  13  14  15  16  17  18  20  21
  22  23  24  25  26  27  28  29  30  31
  32
 estimate of guessing on this matching subtest = 0.20

 proportion of Ref. grp. examinees eliminated =  0.504
 proportion of Focal grp. examinees eliminated =  0.535
 KR-20 for Ref. grp. =   0.955
 KR-20 for Foc. grp. =   0.953

 Matching Subtest Summary Statistics
     Reference Group:               Mean =   7.83
                        Standard deviation =  5.03
     Focal Group:                   Mean =   7.10
                        Standard deviation =  4.10

     Standardized Score Difference =  0.16


               SIBTEST-Focal weighting Results
                                              p-value
 Beta estimate   standard error   for DIF against either grp.
    -0.067           0.025                 0.007367

                     SIBTEST error flag = 0
   No errors.  This SIBTEST run had a normal successful completion.

$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$
 OUTPUT FOR RUN NUMBER  20          OUTPUT FOR RUN NUMBER  20

 Suspect subtest items:
  20

 Matching subtest items:
   1   2   3   4   5   6   7   8   9  10
  11  12  13  14  15  16  17  18  19  21
  22  23  24  25  26  27  28  29  30  31
  32
 estimate of guessing on this matching subtest = 0.20

 proportion of Ref. grp. examinees eliminated =  0.517
 proportion of Focal grp. examinees eliminated =  0.544
 KR-20 for Ref. grp. =   0.959
 KR-20 for Foc. grp. =   0.958

 Matching Subtest Summary Statistics
     Reference Group:               Mean =   7.76
                        Standard deviation =  5.05
     Focal Group:                   Mean =   7.05
                        Standard deviation =  4.16
```

```
     Standardized Score Difference =  0.15


                SIBTEST-Focal weighting Results
                                              p-value
  Beta estimate   standard error   for DIF against either grp.
     -0.014           0.029                 0.630893

                     SIBTEST error flag = 0
    No errors.  This SIBTEST run had a normal successful completion.

$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$
 OUTPUT FOR RUN NUMBER  21          OUTPUT FOR RUN NUMBER  21

 Suspect subtest items:
  21

 Matching subtest items:
    1   2   3   4   5   6   7   8   9  10
   11  12  13  14  15  16  17  18  19  20
   22  23  24  25  26  27  28  29  30  31
   32
 estimate of guessing on this matching subtest = 0.20

 proportion of Ref. grp. examinees eliminated =  0.520
 proportion of Focal grp. examinees eliminated =  0.551
 KR-20 for Ref. grp. =   0.958
 KR-20 for Foc. grp. =   0.959

 Matching Subtest Summary Statistics
     Reference Group:              Mean =   7.78
                       Standard deviation =  5.07
     Focal Group:                  Mean =   7.07
                       Standard deviation =  4.22

     Standardized Score Difference =  0.15


                SIBTEST-Focal weighting Results
                                              p-value
  Beta estimate   standard error   for DIF against either grp.
      0.001           0.027                 0.968109

                     SIBTEST error flag = 0
    No errors.  This SIBTEST run had a normal successful completion.

$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$
 OUTPUT FOR RUN NUMBER  22          OUTPUT FOR RUN NUMBER  22

 Suspect subtest items:
  22

 Matching subtest items:
    1   2   3   4   5   6   7   8   9  10
   11  12  13  14  15  16  17  18  19  20
   21  23  24  25  26  27  28  29  30  31
   32
 estimate of guessing on this matching subtest = 0.20

 proportion of Ref. grp. examinees eliminated =  0.530
 proportion of Focal grp. examinees eliminated =  0.552
 KR-20 for Ref. grp. =   0.959
 KR-20 for Foc. grp. =   0.957

 Matching Subtest Summary Statistics
     Reference Group:              Mean =   7.76
                       Standard deviation =  5.07
     Focal Group:                  Mean =   7.06
                       Standard deviation =  4.15

     Standardized Score Difference =  0.15
```

```
               SIBTEST-Focal weighting Results
                                              p-value
 Beta estimate   standard error   for DIF against either grp.
    -0.002           0.028                 0.939813

                   SIBTEST error flag = 0
   No errors.  This SIBTEST run had a normal successful completion.

$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$
 OUTPUT FOR RUN NUMBER  23          OUTPUT FOR RUN NUMBER  23

 Suspect subtest items:
  23

 Matching subtest items:
   1   2   3   4   5   6   7   8   9  10
  11  12  13  14  15  16  17  18  19  20
  21  22  24  25  26  27  28  29  30  31
  32
 estimate of guessing on this matching subtest = 0.20

 proportion of Ref. grp. examinees eliminated =  0.526
 proportion of Focal grp. examinees eliminated =  0.542
 KR-20 for Ref. grp. =   0.958
 KR-20 for Foc. grp. =   0.956

 Matching Subtest Summary Statistics
     Reference Group:              Mean =   7.77
                      Standard deviation =  5.06
     Focal Group:                  Mean =   7.08
                      Standard deviation =  4.16

     Standardized Score Difference =  0.15


               SIBTEST-Focal weighting Results
                                              p-value
 Beta estimate   standard error   for DIF against either grp.
     0.010           0.029                 0.715493

                   SIBTEST error flag = 0
   No errors.  This SIBTEST run had a normal successful completion.

$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$
 OUTPUT FOR RUN NUMBER  24          OUTPUT FOR RUN NUMBER  24

 Suspect subtest items:
  24

 Matching subtest items:
   1   2   3   4   5   6   7   8   9  10
  11  12  13  14  15  16  17  18  19  20
  21  22  23  25  26  27  28  29  30  31
  32
 estimate of guessing on this matching subtest = 0.20

 proportion of Ref. grp. examinees eliminated =  0.519
 proportion of Focal grp. examinees eliminated =  0.542
 KR-20 for Ref. grp. =   0.959
 KR-20 for Foc. grp. =   0.956

 Matching Subtest Summary Statistics
     Reference Group:              Mean =   7.74
                      Standard deviation =  5.04
     Focal Group:                  Mean =   7.07
                      Standard deviation =  4.13

     Standardized Score Difference =  0.15


               SIBTEST-Focal weighting Results
                                              p-value
 Beta estimate   standard error   for DIF against either grp.
```

```
     0.013           0.027                0.632513

                   SIBTEST error flag = 0
   No errors.  This SIBTEST run had a normal successful completion.

$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$
 OUTPUT FOR RUN NUMBER  25          OUTPUT FOR RUN NUMBER  25

 Suspect subtest items:
  25

 Matching subtest items:
   1   2   3   4   5   6   7   8   9  10
  11  12  13  14  15  16  17  18  19  20
  21  22  23  24  26  27  28  29  30  31
  32
 estimate of guessing on this matching subtest = 0.20

 proportion of Ref. grp. examinees eliminated =  0.517
 proportion of Focal grp. examinees eliminated =  0.545
 KR-20 for Ref. grp. =   0.959
 KR-20 for Foc. grp. =   0.955

 Matching Subtest Summary Statistics
     Reference Group:               Mean =   7.77
                    Standard deviation =  5.08
     Focal Group:                   Mean =   7.10
                    Standard deviation =  4.18

     Standardized Score Difference =  0.14


               SIBTEST-Focal weighting Results
                                          p-value
  Beta estimate   standard error   for DIF against either grp.
      0.043           0.028                0.120651

                   SIBTEST error flag = 0
   No errors.  This SIBTEST run had a normal successful completion.

$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$
 OUTPUT FOR RUN NUMBER  26          OUTPUT FOR RUN NUMBER  26

 Suspect subtest items:
  26

 Matching subtest items:
   1   2   3   4   5   6   7   8   9  10
  11  12  13  14  15  16  17  18  19  20
  21  22  23  24  25  27  28  29  30  31
  32
 estimate of guessing on this matching subtest = 0.20

 proportion of Ref. grp. examinees eliminated =  0.512
 proportion of Focal grp. examinees eliminated =  0.542
 KR-20 for Ref. grp. =   0.958
 KR-20 for Foc. grp. =   0.960

 Matching Subtest Summary Statistics
     Reference Group:               Mean =   7.80
                    Standard deviation =  5.12
     Focal Group:                   Mean =   7.05
                    Standard deviation =  4.22

     Standardized Score Difference =  0.16


               SIBTEST-Focal weighting Results
                                          p-value
  Beta estimate   standard error   for DIF against either grp.
     -0.053           0.028                0.058280

                   SIBTEST error flag = 0
```

```
   No errors.  This SIBTEST run had a normal successful completion.

$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$
 OUTPUT FOR RUN NUMBER  27          OUTPUT FOR RUN NUMBER  27

 Suspect subtest items:
  27

 Matching subtest items:
   1   2   3   4   5   6   7   8   9  10
  11  12  13  14  15  16  17  18  19  20
  21  22  23  24  25  26  28  29  30  31
  32
 estimate of guessing on this matching subtest = 0.20

 proportion of Ref. grp. examinees eliminated =  0.513
 proportion of Focal grp. examinees eliminated =  0.538
 KR-20 for Ref. grp. =   0.958
 KR-20 for Foc. grp. =   0.957

 Matching Subtest Summary Statistics
     Reference Group:               Mean =   7.83
                      Standard deviation =  5.14
     Focal Group:                   Mean =   7.14
                      Standard deviation =  4.22

     Standardized Score Difference =  0.15


               SIBTEST-Focal weighting Results
                                          p-value
  Beta estimate   standard error   for DIF against either grp.
      0.040           0.025                0.103430

                    SIBTEST error flag = 0
    No errors.  This SIBTEST run had a normal successful completion.

$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$
 OUTPUT FOR RUN NUMBER  28          OUTPUT FOR RUN NUMBER  28

 Suspect subtest items:
  28

 Matching subtest items:
   1   2   3   4   5   6   7   8   9  10
  11  12  13  14  15  16  17  18  19  20
  21  22  23  24  25  26  27  29  30  31
  32
 estimate of guessing on this matching subtest = 0.20

 proportion of Ref. grp. examinees eliminated =  0.527
 proportion of Focal grp. examinees eliminated =  0.567
 KR-20 for Ref. grp. =   0.962
 KR-20 for Foc. grp. =   0.964

 Matching Subtest Summary Statistics
     Reference Group:               Mean =   7.72
                      Standard deviation =  5.11
     Focal Group:                   Mean =   6.96
                      Standard deviation =  4.20

     Standardized Score Difference =  0.16


               SIBTEST-Focal weighting Results
                                          p-value
  Beta estimate   standard error   for DIF against either grp.
     -0.047           0.031                0.133133

                    SIBTEST error flag = 0
    No errors.  This SIBTEST run had a normal successful completion.

$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$
```

```
 OUTPUT FOR RUN NUMBER  29          OUTPUT FOR RUN NUMBER  29

 Suspect subtest items:
  29

 Matching subtest items:
   1   2   3   4   5   6   7   8   9  10
  11  12  13  14  15  16  17  18  19  20
  21  22  23  24  25  26  27  28  30  31
  32
 estimate of guessing on this matching subtest = 0.20

 proportion of Ref. grp. examinees eliminated =  0.535
 proportion of Focal grp. examinees eliminated =  0.550
 KR-20 for Ref. grp. =   0.960
 KR-20 for Foc. grp. =   0.959

 Matching Subtest Summary Statistics
     Reference Group:               Mean =   7.75
                       Standard deviation =  5.10
     Focal Group:                   Mean =   7.04
                       Standard deviation =  4.17

     Standardized Score Difference =  0.15


                SIBTEST-Focal weighting Results
                                          p-value
  Beta estimate   standard error   for DIF against either grp.
     -0.006           0.030                0.829678

                     SIBTEST error flag = 0
    No errors.  This SIBTEST run had a normal successful completion.

$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$
 OUTPUT FOR RUN NUMBER  30          OUTPUT FOR RUN NUMBER  30

 Suspect subtest items:
  30

 Matching subtest items:
   1   2   3   4   5   6   7   8   9  10
  11  12  13  14  15  16  17  18  19  20
  21  22  23  24  25  26  27  28  29  31
  32
 estimate of guessing on this matching subtest = 0.20

 proportion of Ref. grp. examinees eliminated =  0.531
 proportion of Focal grp. examinees eliminated =  0.565
 KR-20 for Ref. grp. =   0.962
 KR-20 for Foc. grp. =   0.964

 Matching Subtest Summary Statistics
     Reference Group:               Mean =   7.67
                       Standard deviation =  5.04
     Focal Group:                   Mean =   6.91
                       Standard deviation =  4.12

     Standardized Score Difference =  0.16


                SIBTEST-Focal weighting Results
                                          p-value
  Beta estimate   standard error   for DIF against either grp.
     -0.057           0.032                0.069697

                     SIBTEST error flag = 0
    No errors.  This SIBTEST run had a normal successful completion.

$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$
 OUTPUT FOR RUN NUMBER  31          OUTPUT FOR RUN NUMBER  31

 Suspect subtest items:
```

```
 31

 Matching subtest items:
   1   2   3   4   5   6   7   8   9  10
  11  12  13  14  15  16  17  18  19  20
  21  22  23  24  25  26  27  28  29  30
  32
 estimate of guessing on this matching subtest = 0.20

 proportion of Ref. grp. examinees eliminated =  0.525
 proportion of Focal grp. examinees eliminated =  0.554
 KR-20 for Ref. grp. =   0.961
 KR-20 for Foc. grp. =   0.959

 Matching Subtest Summary Statistics
     Reference Group:              Mean =   7.72
                       Standard deviation =  5.07
     Focal Group:                  Mean =   7.02
                       Standard deviation =  4.13

     Standardized Score Difference =  0.15


                SIBTEST-Focal weighting Results
                                          p-value
  Beta estimate   standard error   for DIF against either grp.
     -0.030           0.030                0.320933

                     SIBTEST error flag = 0
    No errors.  This SIBTEST run had a normal successful completion.

$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$
 OUTPUT FOR RUN NUMBER  32          OUTPUT FOR RUN NUMBER  32

 Suspect subtest items:
  32

 Matching subtest items:
   1   2   3   4   5   6   7   8   9  10
  11  12  13  14  15  16  17  18  19  20
  21  22  23  24  25  26  27  28  29  30
  31
 estimate of guessing on this matching subtest = 0.20

 proportion of Ref. grp. examinees eliminated =  0.527
 proportion of Focal grp. examinees eliminated =  0.552
 KR-20 for Ref. grp. =   0.960
 KR-20 for Foc. grp. =   0.958

 Matching Subtest Summary Statistics
     Reference Group:              Mean =   7.74
                       Standard deviation =  5.09
     Focal Group:                  Mean =   7.05
                       Standard deviation =  4.15

     Standardized Score Difference =  0.15


                SIBTEST-Focal weighting Results
                                          p-value
  Beta estimate   standard error   for DIF against either grp.
     -0.023           0.030                0.450687

                     SIBTEST error flag = 0
    No errors.  This SIBTEST run had a normal successful completion.
$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$

                          SUMMARY OF THE RUNS

    p-value notation:
 R  denotes p-value for test of DIF/DBF against Ref. group
 F  denotes p-value for test of DIF/DBF against Foc. group
 E  denotes p-value for test of DIF/DBF against either the
```

Ref. or Foc. group.

NOTES:

MS/SSD = Matching Subtest Standardized Score Difference.
Standardized difference in mean observed scores
between Reference group and Focal group on the
matching subtest.

p-elim = proportion of Reference (R) and Focal (F) groups
eliminated (not used) in SIBTEST calculations.

Positive Beta estimate indicates DIF/DBF favoring Ref. grp.
Negative Beta estimate indicates DIF/DBF favoring Foc. grp.

FLAG = error flag indicator. FLAG=0 indicates a normal
successful completion of a SIBTEST run. All other values
of FLAG come with short error messages.

SIBTEST-Focal weighting

Run no.	Suspect Subtest Item Numbers	Beta estimate	standard error	p-value	p-elim R	p-elim F	MS SSD	FLAG
1	1	0.018	0.023	0.446 E	.50	.54	0.15	0
2	2	-0.012	0.025	0.640 E	.52	.53	0.15	0
3	3	0.003	0.028	0.905 E	.53	.54	0.15	0
4	4	0.004	0.027	0.885 E	.52	.54	0.15	0
5	5	-0.010	0.028	0.715 E	.52	.55	0.14	0
6	6	0.057	0.024	0.018 E	.51	.54	0.14	0
7	7	0.012	0.029	0.684 E	.52	.56	0.15	0
8	8	-0.002	0.029	0.934 E	.52	.55	0.15	0
9	9	0.027	0.028	0.332 E	.53	.55	0.15	0
10	10	-0.046	0.029	0.115 E	.52	.55	0.16	0
11	11	0.026	0.031	0.400 E	.52	.56	0.15	0
12	12	-0.024	0.025	0.330 E	.51	.54	0.15	0
13	13	0.010	0.024	0.659 E	.51	.53	0.15	0
14	14	0.033	0.027	0.233 E	.53	.54	0.14	0
15	15	0.008	0.027	0.753 E	.52	.54	0.14	0
16	16	0.007	0.026	0.778 E	.53	.55	0.15	0
17	17	-0.024	0.027	0.381 E	.52	.54	0.15	0
18	18	-0.012	0.026	0.642 E	.52	.54	0.15	0
19	19	-0.067	0.025	0.007 E	.50	.54	0.16	0
20	20	-0.014	0.029	0.631 E	.52	.54	0.15	0
21	21	0.001	0.027	0.968 E	.52	.55	0.15	0
22	22	-0.002	0.028	0.940 E	.53	.55	0.15	0
23	23	0.010	0.029	0.715 E	.53	.54	0.15	0
24	24	0.013	0.027	0.633 E	.52	.54	0.15	0
25	25	0.043	0.028	0.121 E	.52	.55	0.14	0
26	26	-0.053	0.028	0.058 E	.51	.54	0.16	0
27	27	0.040	0.025	0.103 E	.51	.54	0.15	0
28	28	-0.047	0.031	0.133 E	.53	.57	0.16	0
29	29	-0.006	0.030	0.830 E	.54	.55	0.15	0
30	30	-0.057	0.032	0.070 E	.53	.56	0.16	0
31	31	-0.030	0.030	0.321 E	.52	.55	0.15	0
32	32	-0.023	0.030	0.451 E	.53	.55	0.15	0

ภาคผนวก ฌ
ตัวอย่าง Print Out ผลการคำนวณอัตราความคลาดเคลื่อนของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบ O-NET ระหว่างวิธี IRT-LR กับวิธี SIBTEST

ตารางที่ ฌ-1 ผลการคำนวณอัตราความคลาดเคลื่อนของการทำหน้าที่ต่างกันของข้อสอบ
ในแบบทดสอบ O-NET ระหว่างวิธี IRT-LR กับวิธี SIBTEST แบบทดสอบภาษาไทย

IRT-LR * SIBTEST Crosstabulation

			SIBTEST		
			DIF	NO DIF	Total
IRT-LR	DIF	Count	13[a]	45[a]	58
		% within IRT-LR	22.4%	77.6%	100.0%
		% within SIBTEST	92.9%	80.4%	82.9%
		% of Total	18.6%	64.3%	82.9%
		Residual	1.4	-1.4	
	NO DIF	Count	1[a]	11[a]	12
		% within IRT-LR	8.3%	91.7%	100.0%
		% within SIBTEST	7.1%	19.6%	17.1%
		% of Total	1.4%	15.7%	17.1%
		Residual	-1.4	1.4	
Total		Count	14	56	70
		% within IRT-LR	20.0%	80.0%	100.0%
		% within SIBTEST	100.0%	100.0%	100.0%
		% of Total	20.0%	80.0%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	1.232[a]	1	.267		
Continuity Correction[b]	.509	1	.476		
Likelihood Ratio	1.449	1	.229		
Fisher's Exact Test				.436	.248
N of Valid Cases	70				

a. 1 cells (25.0%) have expected count less than 5. The minimum expected count is 2.40.

b. Computed only for a 2x2 table

ตารางที่ ฌ-2 ผลการคำนวณอัตราความคลาดเคลื่อนของการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบ O-NET ระหว่างวิธี IRT-LR กับวิธี SIBTEST แบบทดสอบสังคมศึกษา ศาสนา และวัฒนธรรม

IRT-LR * SIBTEST Crosstabulation

			SIBTEST		
			DIF	NO DIF	Total
IRT-LR	DIF	Count	13	46	59
		% within IRT-LR	22.0%	78.0%	100.0%
		% within SIBTEST	100.0%	88.5%	90.8%
		% of Total	20.0%	70.8%	90.8%
		Residual	1.2	-1.2	
	NO	Count	0	6	6
		% within IRT-LR	0.0%	100.0%	100.0%
		% within SIBTEST	0.0%	11.5%	9.2%
		% of Total	0.0%	9.2%	9.2%
		Residual	-1.2	1.2	
Total		Count	13	52	65
		% within IRT-LR	20.0%	80.0%	100.0%
		% within SIBTEST	100.0%	100.0%	100.0%
		% of Total	20.0%	80.0%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	1.653[a]	1	.199		
Continuity Correction[b]	.562	1	.453		
Likelihood Ratio	2.827	1	.093		
Fisher's Exact Test				.335	.246
N of Valid Cases	65				

a. 2 cells (50.0%) have expected count less than 5. The minimum expected count is 1.20.

b. Computed only for a 2x2 table

ตารางที่ ฌ-3 ผลการคำนวณอัตราความคลาดเคลื่อนของการทำหน้าที่ต่างกันของข้อสอบ
ในแบบทดสอบ O-NET ระหว่างวิธี IRT-LR กับวิธี SIBTEST แบบทดสอบภาษาอังกฤษ

IRT-LR * SIBTEST Crosstabulation

			SIBTEST		
			DIF	NO DIF	Total
IRT-LR	DIF	Count	3	2	5
		% within IRT-LR	60.0%	40.0%	100.0%
		% within SIBTEST	33.3%	2.8%	6.3%
		% of Total	3.8%	2.5%	6.3%
		Residual	2.4	-2.4	
	NO DIF	Count	6	69	75
		% within IRT-LR	8.0%	92.0%	100.0%
		% within SIBTEST	66.7%	97.2%	93.8%
		% of Total	7.5%	86.3%	93.8%
		Residual	-2.4	2.4	
Total		Count	9	71	80
		% within IRT-LR	11.3%	88.8%	100.0%
		% within SIBTEST	100.0%	100.0%	100.0%
		% of Total	11.3%	88.8%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	12.695[a]	1	.000		
Continuity Correction[b]	8.021	1	.005		
Likelihood Ratio	7.728	1	.005		
Fisher's Exact Test				.009	.009
N of Valid Cases	80				

a. 2 cells (50.0%) have expected count less than 5. The minimum expected count is .56.

b. Computed only for a 2x2 table

ตารางที่ ฌ-4 ผลการคำนวณอัตราความคลาดเคลื่อนของการทำหน้าที่ต่างกันของข้อสอบ
ในแบบทดสอบ O-NET ระหว่างวิธี IRT-LR กับวิธี SIBTEST แบบทดสอบคณิตศาสตร์

IRT-LR * SIBTEST Crosstabulation

			SIBTEST		
			DIF	NO DIF	Total
IRT-LR	DIF	Count	2	4	6
		% within IRT-LR	33.3%	66.7%	100.0%
		% within SIBTEST	100.0%	13.3%	18.8%
		% of Total	6.3%	12.5%	18.8%
		Residual	1.6	-1.6	
	NO DIF	Count	0	26	26
		% within IRT-LR	0.0%	100.0%	100.0%
		% within SIBTEST	0.0%	86.7%	81.3%
		% of Total	0.0%	81.3%	81.3%
		Residual	-1.6	1.6	
Total		Count	2	30	32
		% within IRT-LR	6.3%	93.8%	100.0%
		% within SIBTEST	100.0%	100.0%	100.0%
		% of Total	6.3%	93.8%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	9.244[a]	1	.002		
Continuity Correction[b]	4.431	1	.035		
Likelihood Ratio	7.324	1	.007		
Fisher's Exact Test				.030	.030
N of Valid Cases	32				

a. 2 cells (50.0%) have expected count less than 5. The minimum expected count is .38.

b. Computed only for a 2x2 table

ตารางที่ ฌ-5 ผลการคำนวณอัตราความคลาดเคลื่อนของการทำหน้าที่ต่างกันของข้อสอบ
ในแบบทดสอบ O-NET ระหว่างวิธี IRT-LR กับวิธี SIBTEST แบบทดสอบวิทยาศาสตร์

IRT-LR * SIBTEST Crosstabulation

			SIBTEST		
			DIF	NO DIF	Total
IRT-LR	DIF	Count	11	44	55
		% within IRT-LR	20.0%	80.0%	100.0%
		% within SIBTEST	100.0%	63.8%	68.8%
		% of Total	13.8%	55.0%	68.8%
		Residual	3.4	-3.4	
	NO DIF	Count	0	25	25
		% within IRT-LR	0.0%	100.0%	100.0%
		% within SIBTEST	0.0%	36.2%	31.3%
		% of Total	0.0%	31.3%	31.3%
		Residual	-3.4	3.4	
Total		Count	11	69	80
		% within IRT-LR	13.8%	86.3%	100.0%
		% within SIBTEST	100.0%	100.0%	100.0%
		% of Total	13.8%	86.3%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	5.797[a]	1	.016		
Continuity Correction[b]	4.233	1	.040		
Likelihood Ratio	9.020	1	.003		
Fisher's Exact Test				.014	.011
N of Valid Cases	80				

a. 1 cells (25.0%) have expected count less than 5. The minimum expected count is 3.44.

b. Computed only for a 2x2 table

ตารางที่ ฌ-6 ผลการคำนวณอัตราความคลาดเคลื่อนของการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบ O-NET ระหว่างวิธี IRT-LR กับวิธี SIBTEST แบบทดสอบสุขศึกษาและพลศึกษา

IRT-LR * SIBTEST Crosstabulation

			SIBTEST		
			DIF	NO DIF	Total
IRT-LR	DIF	Count	11	20	31
		% within IRT-LR	35.5%	64.5%	100.0%
		% within SIBTEST	100.0%	95.2%	96.9%
		% of Total	34.4%	62.5%	96.9%
		Residual	.3	-.3	
	NO DIF	Count	0	1	1
		% within IRT-LR	0.0%	100.0%	100.0%
		% within SIBTEST	0.0%	4.8%	3.1%
		% of Total	0.0%	3.1%	3.1%
		Residual	-.3	.3	
Total		Count	11	21	32
		% within IRT-LR	34.4%	65.6%	100.0%
		% within SIBTEST	100.0%	100.0%	100.0%
		% of Total	34.4%	65.6%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	.541[a]	1	.462		
Continuity Correction[b]	0.000	1	1.000		
Likelihood Ratio	.859	1	.354		
Fisher's Exact Test				1.000	.656
N of Valid Cases	32				

a. 2 cells (50.0%) have expected count less than 5. The minimum expected count is .34.

b. Computed only for a 2x2 table

ตารางที่ ฌ-7 ผลการคำนวณอัตราความคลาดเคลื่อนของการทำหน้าที่ต่างกันของข้อสอบ
ในแบบทดสอบ O-NET ระหว่างวิธี IRT-LR กับวิธี SIBTEST แบบทดสอบศิลปะ

IRT-LR * SIBTEST Crosstabulation

			SIBTEST		
			DIF	NO DIF	Total
IRT-LR	DIF	Count	6	20	26
		% within IRT-LR	23.1%	76.9%	100.0%
		% within SIBTEST	100.0%	83.3%	86.7%
		% of Total	20.0%	66.7%	86.7%
		Residual	.8	-.8	
	NO DIF	Count	0	4	4
		% within IRT-LR	0.0%	100.0%	100.0%
		% within SIBTEST	0.0%	16.7%	13.3%
		% of Total	0.0%	13.3%	13.3%
		Residual	-.8	.8	
Total		Count	6	24	30
		% within IRT-LR	20.0%	80.0%	100.0%
		% within SIBTEST	100.0%	100.0%	100.0%
		% of Total	20.0%	80.0%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	1.154[a]	1	.283		
Continuity Correction[b]	.162	1	.687		
Likelihood Ratio	1.934	1	.164		
Fisher's Exact Test				.557	.388
N of Valid Cases	30				

a. 2 cells (50.0%) have expected count less than 5. The minimum expected count is .80.

b. Computed only for a 2x2 table

ตารางที่ ฌ-8 ผลการคำนวณอัตราความคลาดเคลื่อนของการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบ O-NET ระหว่างวิธี IRT-LR กับวิธี SIBTEST แบบทดสอบการงานอาชีพและเทคโนโลยี

IRT-LR * SIBTEST Crosstabulation

			SIBTEST		
			DIF	NO DIF	Total
IRT-LR	DIF	Count	6	13	19
		% within IRT-LR	31.6%	68.4%	100.0%
		% within SIBTEST	46.2%	46.4%	46.3%
		% of Total	14.6%	31.7%	46.3%
		Residual	.0	.0	
	NO DIF	Count	7	15	22
		% within IRT-LR	31.8%	68.2%	100.0%
		% within SIBTEST	53.8%	53.6%	53.7%
		% of Total	17.1%	36.6%	53.7%
		Residual	.0	.0	
Total		Count	13	28	41
		% within IRT-LR	31.7%	68.3%	100.0%
		% within SIBTEST	100.0%	100.0%	100.0%
		% of Total	31.7%	68.3%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	.000[a]	1	.987		
Continuity Correction[b]	0.000	1	1.000		
Likelihood Ratio	.000	1	.987		
Fisher's Exact Test				1.000	.626
N of Valid Cases	41				

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 6.02.

b. Computed only for a 2x2 table

ภาคผนวก ญ
ผลการคำนวณอัตราความคลาดเคลื่อนประเภทที่ 1 และ 2
ของผลการตรวจสอบการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

การคำนวณอัตราความคลาดเคลื่อนประเภทที่ 1 และ 2 ของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ แบบทดสอบภาษาไทย ระหว่างวิธี IRT-LR กับวิธี SIBTEST

กลุ่มตัวอย่าง 2,000 คน จำนวน 70 ข้อ

ตรวจสอบ DIF ด้วยวิธี IRT-LR โดยใช้วิธี SIBTEST เป็นเกณฑ์

		วิธี SIBTEST		
		NO DIF	DIF	
วิธี IRT-LR	NO DIF	11	1	12
	DIF	45	13	58
		56	14	70

อัตราความถูกต้องประเภทที่ 1 $=\frac{11}{56}=0.196$

อัตราความถูกต้องประเภทที่ 2 $=\frac{13}{14}=0.928$

อัตราความคลาดเคลื่อนประเภทที่ 1 $=\frac{45}{54}=0.803$

อัตราความคลาดเคลื่อนประเภทที่ 2 $=\frac{1}{14}=0.071$

การคำนวณอัตราความคลาดเคลื่อนประเภทที่ 1 และ 2 ของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ แบบทดสอบสังคมศึกษา ศาสนา และวัฒนธรรม ระหว่างวิธี IRT-LR กับวิธี SIBTEST

กลุ่มตัวอย่าง 2,000 คน จำนวน 65 ข้อ

ตรวจสอบ DIF ด้วยวิธี IRT-LR โดยใช้วิธี SIBTEST เป็นเกณฑ์

		วิธี SIBTEST		
		NO DIF	DIF	
วิธี IRT-LR	NO DIF	6	0	6
	DIF	46	13	59
		52	13	65

อัตราความถูกต้องประเภทที่ 1 $=\frac{6}{52}=0.115$

อัตราความถูกต้องประเภทที่ 2 $=\frac{13}{13}=1.000$

อัตราความคลาดเคลื่อนประเภทที่ 1 $=\frac{46}{52}=0.884$

อัตราความคลาดเคลื่อนประเภทที่ 2 $=\frac{0}{13}=0.000$

การคำนวณอัตราความคลาดเคลื่อนประเภทที่ 1 และ 2 ของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ แบบทดสอบภาษาอังกฤษ ระหว่างวิธี IRT-LR กับวิธี SIBTEST

กลุ่มตัวอย่าง 2,000 คน จำนวน 80 ข้อ

ตรวจสอบ DIF ด้วยวิธี IRT-LR โดยใช้วิธี SIBTEST เป็นเกณฑ์

		วิธี SIBTEST		
		NO DIF	DIF	
วิธี IRT-LR	NO DIF	69	6	75
	DIF	2	3	5
		71	9	80

อัตราความถูกต้องประเภทที่ 1 $=\frac{69}{71}=0.971$

อัตราความถูกต้องประเภทที่ 2 $=\frac{3}{9}=0.333$

อัตราความคลาดเคลื่อนประเภทที่ 1 $=\frac{2}{71}=0.028$

อัตราความคลาดเคลื่อนประเภทที่ 2 $=\frac{6}{9}=0.667$

การคำนวณอัตราความคลาดเคลื่อนประเภทที่ 1 และ 2 ของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ แบบทดสอบคณิตศาสตร์ ระหว่างวิธี IRT-LR กับวิธี SIBTEST

กลุ่มตัวอย่าง 2,000 คน จำนวน 32 ข้อ

ตรวจสอบ DIF ด้วยวิธี IRT-LR โดยใช้วิธี SIBTEST เป็นเกณฑ์

		วิธี SIBTEST		
		NO DIF	DIF	
วิธี IRT-LR	NO DIF	26	0	26
	DIF	4	2	6
		30	2	32

อัตราความถูกต้องประเภทที่ 1 $=\frac{26}{30}=0.867$

อัตราความถูกต้องประเภทที่ 2 $=\frac{2}{2}=1.000$

อัตราความคลาดเคลื่อนประเภทที่ 1 $=\frac{4}{30}=0.133$

อัตราความคลาดเคลื่อนประเภทที่ 2 $=\frac{0}{2}=0.000$

การคำนวณอัตราความคลาดเคลื่อนประเภทที่ 1 และ 2 ของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ แบบทดสอบวิทยาศาสตร์ ระหว่างวิธี IRT-LR กับวิธี SIBTEST

กลุ่มตัวอย่าง 2,000 คน จำนวน 80 ข้อ

ตรวจสอบ DIF ด้วยวิธี IRT-LR โดยใช้วิธี SIBTEST เป็นเกณฑ์

		วิธี SIBTEST		
		NO DIF	DIF	
วิธี IRT-LR	NO DIF	25	0	25
	DIF	44	11	55
		69	11	80

อัตราความถูกต้องประเภทที่ 1 $=\frac{25}{69}=0.362$

อัตราความถูกต้องประเภทที่ 2 $=\frac{11}{11}=1.000$

อัตราความคลาดเคลื่อนประเภทที่ 1 $=\frac{44}{69}=0.637$

อัตราความคลาดเคลื่อนประเภทที่ 2 $=\frac{0}{11}=0.000$

การคำนวณอัตราความคลาดเคลื่อนประเภทที่ 1 และ 2 ของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ แบบทดสอบสุขศึกษา และพลศึกษา ระหว่างวิธี IRT-LR กับวิธี SIBTEST

กลุ่มตัวอย่าง 2,000 คน จำนวน 32 ข้อ

ตรวจสอบ DIF ด้วยวิธี IRT-LR โดยใช้วิธี SIBTEST เป็นเกณฑ์

		วิธี SIBTEST		
		NO DIF	DIF	
วิธี IRT-LR	NO DIF	1	0	1
	DIF	20	11	31
		21	11	32

อัตราความถูกต้องประเภทที่ 1 $=\frac{1}{21}=0.047$

อัตราความถูกต้องประเภทที่ 2 $=\frac{11}{11}=1.000$

อัตราความคลาดเคลื่อนประเภทที่ 1 $=\frac{20}{21}=0.952$

อัตราความคลาดเคลื่อนประเภทที่ 2 $=\frac{0}{11}=0.000$

การคำนวณอัตราความคลาดเคลื่อนประเภทที่ 1 และ 2 ของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ แบบทดสอบศิลปะ ระหว่างวิธี IRT-LR กับวิธี SIBTEST

กลุ่มตัวอย่าง 2,000 คน จำนวน 30 ข้อ

ตรวจสอบ DIF ด้วยวิธี IRT-LR โดยใช้วิธี SIBTEST เป็นเกณฑ์

		วิธี SIBTEST: NO DIF	วิธี SIBTEST: DIF	
วิธี IRT-LR	NO DIF	4	0	4
	DIF	20	6	26
		24	6	30

อัตราความถูกต้องประเภทที่ 1 $=\frac{4}{24}=0.167$

อัตราความถูกต้องประเภทที่ 2 $=\frac{6}{6}=1.000$

อัตราความคลาดเคลื่อนประเภทที่ 1 $=\frac{20}{24}=0.833$

อัตราความคลาดเคลื่อนประเภทที่ 2 $=\frac{0}{6}=0.000$

การคำนวณอัตราความคลาดเคลื่อนประเภทที่ 1 และ 2 ของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ แบบทดสอบการงานอาชีพ และเทคโนโลยี ระหว่างวิธี IRT-LR กับวิธี SIBTEST

กลุ่มตัวอย่าง 2000 คน จำนวน 41 ข้อ

ตรวจสอบ DIF ด้วยวิธี IRT-LR โดยใช้วิธี SIBTEST เป็นเกณฑ์

		วิธี SIBTEST: NO DIF	วิธี SIBTEST: DIF	
วิธี IRT-LR	NO DIF	15	7	22
	DIF	13	6	19
		28	13	41

อัตราความถูกต้องประเภทที่ 1 $=\frac{15}{28}=0.535$

อัตราความถูกต้องประเภทที่ 2 $=\frac{6}{13}=0.461$

อัตราความคลาดเคลื่อนประเภทที่ 1 $=\frac{13}{28}=0.464$

อัตราความคลาดเคลื่อนประเภทที่ 2 $=\frac{7}{13}=0.538$