

หัวข้อวิทยานิพนธ์ การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบวัดความสามารถด้านการคำนวณ
ระหว่างนักเรียนชายกับนักเรียนหญิงของการทดสอบระดับชาติ

ชื่อและนามสกุล นางสาวภาภรณ์ อินทิ

แขนงวิชา การวัดและประเมินผลการศึกษา

สาขาวิชา ศึกษาศาสตร์ มหาวิทยาลัยสุโขทัยธรรมาธิราช

อาจารย์ที่ปรึกษา 1. รองศาสตราจารย์ ดร.สังวรณี ใจกระโทก
2. รองศาสตราจารย์ ดร.นลินี ฦ นคร

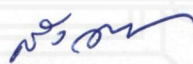
วิทยานิพนธ์นี้ ได้รับความเห็นชอบให้เป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรระดับปริญญาโท เมื่อวันที่ 6 กันยายน 2560

คณะกรรมการสอบวิทยานิพนธ์



ประธานกรรมการ

(รองศาสตราจารย์ ดร.กาญจนา วันสุนทร)



กรรมการ

(รองศาสตราจารย์ ดร.สังวรณี ใจกระโทก)



กรรมการ

(รองศาสตราจารย์ ดร.นลินี ฦ นคร)



ประธานกรรมการบัณฑิตศึกษา

(รองศาสตราจารย์ ดร.วรางคณา จันทร์คง)

ชื่อวิทยานิพนธ์ การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบวัดความสามารถด้านการคำนวณ
ระหว่างนักเรียนชาย กับนักเรียนหญิงของการทดสอบระดับชาติ

ผู้วิจัย นางสุภาภรณ์ อันทิ รหัสนักศึกษา 2572500250

ปริญญา ศึกษาศาสตรมหาบัณฑิต (การประเมินการศึกษา)

อาจารย์ที่ปรึกษา (1) ผู้ช่วยศาสตราจารย์ ดร.นลินี ฌ นคร

(2) ผู้ช่วยศาสตราจารย์ ดร.สังวรณ์ ังคระโทก **ปีการศึกษา** 2562

บทคัดย่อ

การวิจัยในครั้งนี้มีวัตถุประสงค์เพื่อ (1) ศึกษาการทำหน้าที่ต่างกันของข้อสอบวัดความสามารถด้านการคำนวณระหว่างนักเรียนชายและนักเรียนหญิงของการทดสอบระดับชาติ (2) เพื่อวิเคราะห์สาเหตุของข้อสอบวัดความสามารถด้านการคำนวณของการทดสอบระดับชาติที่ทำหน้าที่ต่างกัน กลุ่มประชากรคือ นักเรียนระดับชั้นประถมศึกษาปีที่ 3 ปีการศึกษา 2557 จำนวน 464,532 คน เป็นเพศชาย 224,615 คน เป็นเพศหญิง 239,917 คน เครื่องมือและข้อมูลที่ใช้ในการวิจัย คือ ข้อสอบ และผลการสอบวัดความสามารถด้านการคำนวณ ชั้นประถมศึกษาปีที่ 3 ของสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน ปีการศึกษา 2557 จำนวน 27 ข้อ วิเคราะห์ค่าสถิติพื้นฐาน หาค่าดัชนีการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีดอดอยโลจิตติก ตรวจสอบขนาดอิทธิพลการทำหน้าที่ต่างกัน และ วิเคราะห์สาเหตุของข้อสอบ ผลการวิจัยสรุปได้ดังนี้

การทำหน้าที่ต่างกันของข้อสอบวัดความสามารถด้านการคำนวณระหว่างนักเรียนชายและนักเรียนหญิงของการทดสอบระดับชาติ จำนวน 27 ข้อ ด้วยวิธีโลจิตติก พบข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูป จำนวน 5 ข้อ และข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป จำนวน 22 ข้อ โดยขนาดอิทธิพลการทำหน้าที่ต่างกันของข้อสอบ พบว่า ขนาดของการทำหน้าที่ต่างกันขนาดเล็กมาก ทั้ง 27 ข้อ ซึ่งแสดงว่าข้อสอบวัดความสามารถด้านการคำนวณฉบับนี้มีคุณภาพด้านความตรงในประเด็นด้านความยุติธรรมต่อผู้สอบที่มีเพศแตกต่างกันจากผลการตรวจสอบการทำหน้าที่ต่างกันของแบบทดสอบวัดความสามารถด้านคำนวณของการทดสอบระดับชาติ

สาเหตุของข้อสอบวัดความสามารถด้านการคำนวณของการทดสอบระดับชาติที่ทำหน้าที่ต่างกันมีขนาดเล็กมาก และนักเรียนชายจะมีโอกาสในการตอบข้อสอบถูกในฉบับนี้มากกว่านักเรียนหญิง เนื่องจากนักเรียนชายจะมีความสนใจเนื้อเรื่องที่เกี่ยวข้องกับ คณิตศาสตร์ วิทยาศาสตร์ การทดลอง เครื่องยนต์ การผจญภัย กีฬา การต่อสู้ ดังนั้นการออกข้อสอบที่อยู่ในความสนใจของเพศชาย เพศชายจะมีโอกาสในการตอบ ข้อสอบเหล่านั้นมากกว่าเพศหญิง

คำสำคัญ การทำหน้าที่ต่างกันของข้อสอบ ความสามารถด้านคำนวณ การทดสอบระดับชาติ

Thesis title: Differential Item Functioning Analysis of National Numeracy Ability
Test between Male and Female Students

Researcher: Mrs. Supaporn Aontee; **ID:** 2572500250

Degree: Master of Education (Educational Evaluation);

Thesis advisors: (1) Dr. Nalinee Na Nakorn, Assistant Professor;

(2) Dr. Sungworn Ngudgratoke, Assistant Professor; **Academic year:** 2019

Abstract

The purposes of this research were to study the differential item functioning between male and female students for numeracy subsection of the national test and to analyze the causal of differential item functioning to a numeracy subsection of the national test. The research sample consisted of 464,532 of third grade students in the academic year 2014, which included 224,615 male students and 239,917 female students. The research instrument was a 27 items of numeracy test developed by the Office of the Basic Education Commission in 2014, which were analyzed through basic statistics and logistic regressions. The research findings were as follows: there were 5 uniform differential items functioning and 22 nonuniform differential items functioning, The sizes of the different functional influences of the national test found that the sizes of the different functions were very small, all 27 items, which showed that the test of this numeracy subsection was the quality of straightness in the issue in the aspect of fairness of the examiners with different genders, which was different from the results of the test of the different functions of the numeracy subsection of the national test.

The causes of the test measures the numeracy subsection of the national test that performed different functions were very small in terms of sizes. The male students had more opportunities to get more numbers of the answers in this test than female students due to the reason that male students seem to be interested in the subjects of mathematics, science, experiment, engine, adventure, sports, fighting. Therefore, the test with the scopes of the subjects that interest male students would give more chances for male students to answer more questions in the test correctly than female students

Keywords: Differential item functioning, Numeracy, National test (NT)

กิตติกรรมประกาศ

วิทยานิพนธ์ เรื่อง การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบวัดความสามารถด้านการคำนวณระหว่างนักเรียนชายกับนักเรียนหญิงของการทดสอบระดับชาติ สำเร็จได้เป็นอย่างดีด้วยความร่วมมือ และช่วยเหลือของบุคคลหลายท่านรวมถึงหน่วยงานต่างๆที่เกี่ยวข้องที่สำคัญคือ รองศาสตราจารย์ ดร. นลินี ฦ นคร และ รองศาสตราจารย์ ดร. สัจวรรณ ังคระโทก ผู้เป็นที่ปรึกษาทางวิชาการเป็นอย่างดี ในการสละเวลาอันมีค่าในการให้ความรู้ คำปรึกษา พร้อมทั้งการตรวจสอบแก้ไข ผู้วิจัยขอขอบพระคุณเป็นอย่างสูง

ขอขอบคุณพระคุณรองศาสตราจารย์ ดร.กาญจนา วัฒนสุนทร ประธานกรรมการสอบวิทยานิพนธ์ ในการให้คำแนะนำที่เป็นประโยชน์ต่อการวิจัยครั้งนี้

นอกจากนี้ผู้วิจัยขอขอบพระคุณคณาจารย์ผู้สอนในแขนงวิชาการวัดและประเมินผล การศึกษาทุกท่านในการประสาทความรู้อันสามารถประยุกต์กับการทำวิจัยครั้งนี้ รวมถึงอาจารย์ สุริยันต์ ยางศรี ซึ่งเป็นผู้เชี่ยวชาญในการให้ข้อมูลเพื่อการวิจัย

ขอขอบพระคุณพี่ๆ ทุกคนในการให้การช่วยเหลือ ให้กำลังใจ และแลกเปลี่ยนความรู้เพื่อการทำวิจัย

สุดท้ายขอกราบขอบพระคุณบิดา มารดา และบุคคลในครอบครัวที่คอยห่วงใย ให้กำลังใจ และสนับสนุนผู้วิจัยในการศึกษาครั้งนี้

สุภาภรณ์ อันทิ
เมษายน 2563

การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบวัดความสามารถด้านการคำนวณ
ระหว่างนักเรียนชาย กับนักเรียนหญิงของการทดสอบระดับชาติ

นางสุภาภรณ์ อันทิ



วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาศึกษาศาสตรมหาบัณฑิต
แขนงวิชาการวัดและประเมินผลการศึกษา สาขาวิชาศึกษาศาสตร์ มหาวิทยาลัยสุโขทัยธรรมมาธิราช

พ.ศ. 2562

Differential Item Functioning Analysis of National Numeracy Ability

Test between Male and Female Students

Mrs. Supaporn Aontee



A Thesis Submitted in Partial Fulfillment of the Requirements for
the Degree of Master of Education in Educational Evaluation

School of Educational Studies

Sukhothai Thammathirat Open University

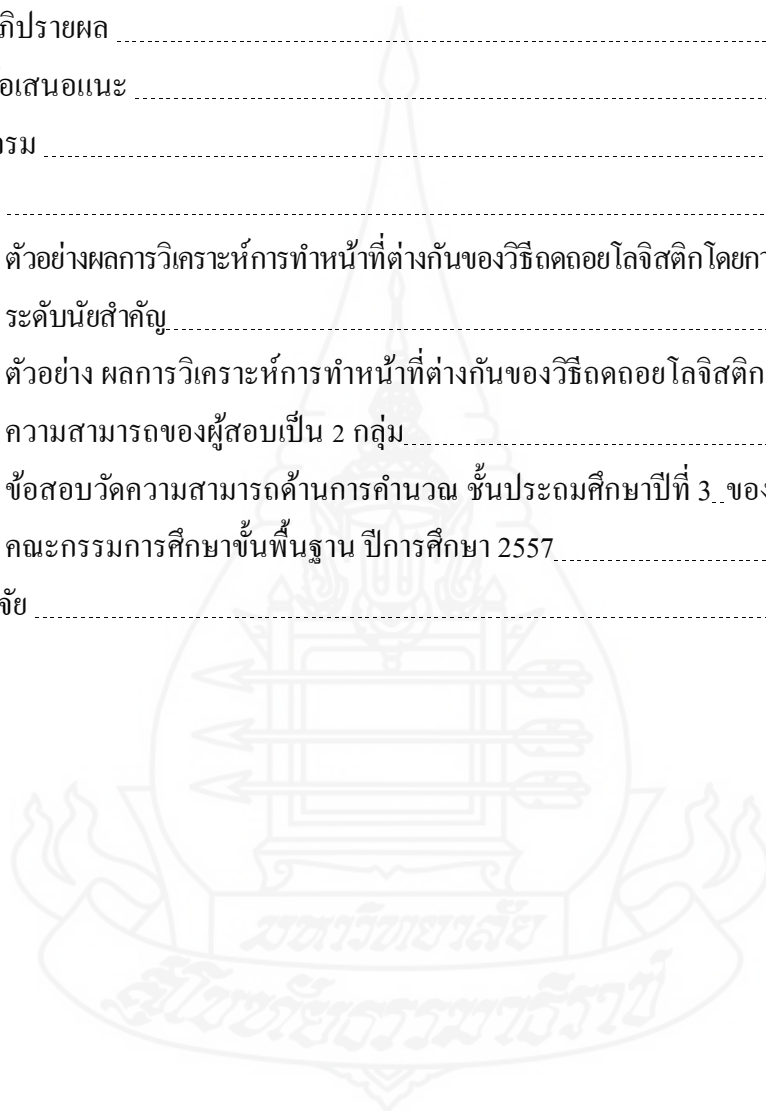
2019

สารบัญ

	หน้า
บทคัดย่อภาษาไทย	ง
บทคัดย่อภาษาอังกฤษ	จ
กิตติกรรมประกาศ	ฉ
สารบัญตาราง	ฉ
สารบัญภาพ	ญ
บทที่ 1 บทนำ	1
ความเป็นมาและความสำคัญของปัญหา	1
วัตถุประสงค์ของการวิจัย	4
สมมติฐานของการวิจัย	4
ขอบเขตของการวิจัย	5
นิยามศัพท์เฉพาะ	5
ประโยชน์ที่คาดว่าจะได้รับ	6
บทที่ 2 วรรณกรรมที่เกี่ยวข้อง	7
คุณภาพของแบบทดสอบ	7
การทำหน้าที่ต่างกันของข้อสอบ	12
โครงการประเมินคุณภาพการศึกษาขั้นพื้นฐาน เพื่อการประกันคุณภาพ งานวิจัยที่เกี่ยวข้อง	24 35
บทที่ 3 วิธีดำเนินการวิจัย	41
ประชากรและกลุ่มตัวอย่าง	41
เครื่องมือที่ใช้ในการวิจัย	41
การวิเคราะห์ข้อมูล	41
บทที่ 4 ผลการวิเคราะห์ข้อมูล	44
ตอนที่ 1 ผลการวิเคราะห์ค่าสถิติพื้นฐานของผลการตอบข้อสอบและคุณภาพของข้อสอบ ..	44
ตอนที่ 2 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีถดถอยโลจิสติก	48
ตอนที่ 3 วิเคราะห์สาเหตุของข้อสอบวัดความสามารถด้านการคำนวณของการทดสอบ ระดับชาติที่ทำหน้าที่ต่างกัน	65

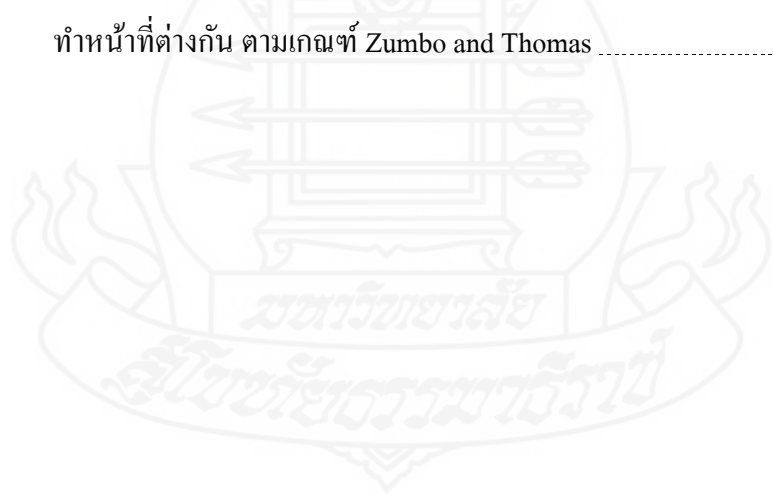
สารบัญ (ต่อ)

	หน้า
บทที่ 5 สรุปการวิจัย อภิปรายผล และข้อเสนอแนะ	66
สรุปการวิจัย	67
อภิปรายผล	68
ข้อเสนอแนะ	69
บรรณานุกรม	71
ภาคผนวก	76
ก ตัวอย่างผลการวิเคราะห์การทำหน้าที่ต่างกันของวิธีถดถอยโลจิสติกโดยการทดสอบ ระดับนัยสำคัญ	77
ข ตัวอย่าง ผลการวิเคราะห์การทำหน้าที่ต่างกันของวิธีถดถอยโลจิสติกโดยการจัดกลุ่ม ความสามารถของผู้สอบเป็น 2 กลุ่ม	87
ค ข้อสอบวัดความสามารถด้านการคำนวณ ชั้นประถมศึกษาปีที่ 3 ของสำนักงาน คณะกรรมการศึกษาขั้นพื้นฐาน ปีการศึกษา 2557	90
ประวัติผู้วิจัย	103



สารบัญตาราง

	หน้า
ตารางที่ 2.1	รายละเอียดเครื่องมือที่ใช้ในการประเมินความสามารถพื้นฐานที่สำคัญ จำเป็นของนักเรียนชั้นประถมศึกษาปีที่ 3 26
ตารางที่ 2.2	รายละเอียดกรอบโครงสร้างข้อสอบวัดความสามารถด้านภาษา 27
ตารางที่ 2.3	รายละเอียดกรอบโครงสร้างข้อสอบวัดความสามารถด้านคำนวณ 27
ตารางที่ 2.4	รายละเอียดกรอบโครงสร้างข้อสอบวัดความสามารถด้านเหตุผล 28
ตารางที่ 4.1	ค่าสถิติพื้นฐานของแบบทดสอบ 45
ตารางที่ 4.2	ค่าเฉลี่ยการตอบข้อสอบถูกของนักเรียนชาย และนักเรียนหญิง ค่าความยากค่า อำนาจจำแนก และค่าความเที่ยงของข้อสอบวัดความสามารถด้านการคำนวณ ... 46
ตารางที่ 4.3	ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีถดถอยโลจิสติก โดยการทดสอบระดับนัยสำคัญของข้อสอบวัดความสามารถด้านการคำนวณ ... 48
ตารางที่ 4.4	กราฟแสดงการตอบข้อสอบถูกระหว่างนักเรียนชายและนักเรียนหญิง ในกลุ่ม นักเรียนที่มีระดับความสามารถสูง และระดับความสามารถต่ำ 50
ตารางที่ 4.5	ผลการการวัดขนาดอิทธิพลในข้อสอบวัดความสามารถด้านการคำนวณที่ ทำหน้าที่ต่างกัน ตามเกณฑ์ Zumbo and Thomas 64



ญ

สารบัญภาพ

	หน้า
ภาพที่ 2.1	ข้อสอบการทำหน้าที่ต่างกันแบบเอกรูป (Uniform DIF) 15
ภาพที่ 2.2	ข้อสอบทำหน้าที่ต่างกันแบบไม่มีทิศทาง (Non-unidirectional DIF) 16
ภาพที่ 2.3	ข้อสอบทำหน้าที่ต่างกันแบบมีทิศทางเดียวกัน (Unidirectional DIF) 16



บทที่ 1

บทนำ

1. ความเป็นมาและความสำคัญของปัญหา

พระราชบัญญัติการศึกษาแห่งชาติ พ.ศ. 2542 และที่แก้ไขเพิ่มเติม (ฉบับที่ 3) พ.ศ. 2553 มาตรา 47 กำหนดให้มีระบบการประกันคุณภาพการศึกษา เพื่อพัฒนาคุณภาพและมาตรฐานการศึกษาในทุกระดับ และมาตรา 48 กำหนดให้หน่วยงานต้นสังกัด และสถานศึกษาจัดให้มีระบบการประกันคุณภาพการศึกษาภายในสถานศึกษา และให้ถือว่าการประกันคุณภาพภายในเป็นส่วนหนึ่งของกระบวนการบริหารการศึกษาที่ต้องดำเนินการอย่างต่อเนื่อง (ราชกิจจานุเบกษา, 2553, น.22-35) หลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 จึงกำหนดให้มีการวัด และประเมินผลการเรียนรู้ใน 4 ระดับ ได้แก่ ระดับชั้นเรียน ระดับสถานศึกษา ระดับเขตพื้นที่การศึกษา และระดับชาติ เพื่อตรวจสอบพัฒนาการ ความก้าวหน้า ความสำเร็จในการเรียนรู้ของผู้เรียน และนำผลการประเมินมาใช้เป็นข้อมูลในการพัฒนาอย่างต่อเนื่องต่อไป ทั้งนี้เพื่อเป็นการสร้างความมั่นใจให้กับผู้เกี่ยวข้องทั้งภายในและภายนอกสถานศึกษาเกี่ยวกับคุณภาพของผู้เรียน (หลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช, 2551, น. 23)

สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐานจึงจัดให้มีการประเมินคุณภาพการศึกษาขั้นพื้นฐานเพื่อการประกันคุณภาพผู้เรียนในระดับชั้นประถมศึกษาปีที่ 3 โดยวัดความสามารถพื้นฐานสำคัญ ๓ ด้าน คือ ด้านภาษา (literacy) ด้านคำนวณ (numeracy) และด้านเหตุผล (reasoning abilities) เพื่อพัฒนาการปรับปรุงการเรียนการสอนของครู ซึ่งถือเป็นกระบวนการตรวจสอบการขับเคลื่อนนโยบายการศึกษาและการเพิ่มประสิทธิภาพสถานศึกษา สร้างความมั่นใจว่าสถานศึกษา และหน่วยงานต้นสังกัดจัดการศึกษาอย่างมีคุณภาพ ได้มาตรฐานตามที่กำหนด นักเรียนมีความรู้ความสามารถ และคุณลักษณะอันพึงประสงค์ตามที่หลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 การทดสอบระดับชาติ (National Test : NT) เพื่อประกันคุณภาพผู้เรียน จึงมีความสำคัญทั้งระดับผู้เรียน ระดับครูผู้สอน ระดับสถานศึกษา ระดับเขตพื้นที่การศึกษา และสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน (คู่มือการจัดสอบการประเมินคุณภาพการศึกษาขั้นพื้นฐานเพื่อประกันคุณภาพผู้เรียน, 2557, น. 1-2) ในการทดสอบจึงต้องคำนึงถึงคุณภาพของแบบทดสอบ ซึ่งดัชนีบ่งชี้คุณภาพของเครื่องมือวัดที่สำคัญแบ่งเป็นสองกลุ่มคือ 1) ดัชนีบ่งชี้คุณภาพ

รายชื่อ ซึ่งประกอบด้วย ความยากง่าย (difficulty) และอำนาจจำแนก (discrimination) 2) ดัชนีคุณภาพ ทั้งฉบับประกอบด้วย ความตรง (validity) และความเที่ยง (reliability) และหัวใจสำคัญของคุณภาพ แบบทดสอบที่ต้องคำนึงถึงคือ “ความตรง” เพราะความตรงแสดงถึงความสามารถในการวัดได้ ถูกต้องแม่นยำ และความยุติธรรมของข้อสอบและแบบทดสอบ (Item and test unfairness) ข้อสอบ หรือแบบทดสอบที่ขาดคุณภาพด้านความตรงในประเด็นของความยุติธรรมของข้อสอบ หรือแบบทดสอบเกิดขึ้นเมื่อผู้สอบกลุ่มย่อยต่างกลุ่มกัน และมีลักษณะเฉพาะบางอย่างแตกต่างกัน เช่น เพศ ภาษา อายุ เชื้อชาติ ศาสนา วัฒนธรรม ภูมิฐานะ สังคม ประสบการณ์ เป็นต้น ทำให้เกิดการ ได้เปรียบเสียเปรียบกันทั้งที่ความสามารถจริงเท่ากัน เดิมใช้คำว่า ความลำเอียงของข้อสอบ (Item bias) หรือ ความลำเอียงของแบบทดสอบ (Test bias) ซึ่งต่อมามีการเปลี่ยนมาใช้เป็น การทำหน้าที่ ต่างกันของข้อสอบ (Differential item Functioning ; DIF) หรือการทำหน้าที่ต่างกันของแบบทดสอบ (Differential Test Functioning ; DIF) (ศิริชัย กาญจนวาสี, 2555, น. 115)

การทดสอบในแต่ละครั้งผู้สอบอาจจะมีลักษณะเฉพาะบางอย่างที่แตกต่างกัน ซึ่งในการทดสอบอาจมีข้อสอบบางข้อที่ทำให้กลุ่มย่อยบางกลุ่มที่มีความสามารถระดับเดียวกัน มีโอกาสของการตอบข้อสอบข้อนั้น ได้ถูกต้อง ไม่เท่ากัน ซึ่งอาจมีสาเหตุจากการใช้ภาษา หรือเนื้อหา บางอย่างที่ไม่เอื้อประโยชน์กับผู้สอบกลุ่มใดกลุ่มหนึ่ง จึงแสดงให้เห็นว่าข้อสอบข้อนั้นขาดความตรง โดยเฉพาะแบบทดสอบที่มีภาษาเข้ามาเกี่ยวข้อง เช่น แบบทดสอบวิชาคณิตศาสตร์เป็นวิชาหนึ่งที่มี ลักษณะบางข้อเป็นข้อสอบ โจทย์ปัญหาที่ประกอบไปด้วยภาษาและตัวเลขที่ต้องการคำตอบ ซึ่งต้องอาศัยการอ่าน โจทย์เพื่อนำไปสู่การแปลงประโยคภาษาเป็นประโยคสัญลักษณ์ทางคณิตศาสตร์เพื่อ หาคำตอบ อาจมีข้อความของโจทย์ปัญหาบางข้อความทำให้ผู้สอบเข้าใจไม่ตรงกัน ส่งผลให้เกิด การได้ประโยชน์ และเสียประโยชน์ระหว่างผู้สอบกลุ่มย่อย ซึ่งจากการศึกษางานวิจัยที่ผ่านมาพบว่า เพศเป็นตัวแปรที่มีผลต่อการทำหน้าที่ต่างกันของข้อสอบ อาทิ อมลฉัฐ อุบลรัตน์ (2552) ได้ศึกษา การทำหน้าที่ต่างกันของข้อสอบวัดผลสัมฤทธิ์ทางการเรียนชั้นมัธยมศึกษาปีที่ 2 สำนักงานเขตพื้นที่ การศึกษาปัตตานี เขต 2 พบว่าเพศเป็นตัวแปรที่มีผลต่อการทำหน้าที่ต่างกันของข้อสอบในวิชา คณิตศาสตร์ ข้อสอบที่ทำหน้าที่ต่างกันส่วนใหญ่จะลำเอียงเข้าข้างกลุ่มเพศชายมากกว่ากลุ่มเพศ หญิง Gierl, Bisanz, Bisanz and Boughton (2003) ได้ศึกษาการระบุเนื้อหาและทักษะทางพุทธิ ปัญญาที่ทำให้เกิดความแตกต่างทางเพศที่มีต่อวิชาคณิตศาสตร์ โดยใช้กระบวนการ วิเคราะห์การทำหน้าที่ต่างกันของข้อสอบแบบพหุมิติพบว่า นักเรียนชายทำคะแนนในส่วนของมิติ สัมพันธ์ได้ดีกว่านักเรียนหญิง ในขณะที่นักเรียนหญิงทำคะแนนในส่วนทักษะความจำได้ดีกว่า นักเรียนชาย และสุกัญญา ทองนาค (2549) ได้ศึกษาการทำหน้าที่ต่างกันของข้อสอบเข้าศึกษาต่อ ประเภทโควตา มหาวิทยาลัยเชียงใหม่ พบว่าข้อสอบที่ใช้สอบส่วนใหญ่เข้าข้างเพศชายมากกว่าเพศ

หญิง ในวิชาภาษาไทยข้อสอบที่ทำหน้าที่ต่างกันเข้าข้างเพศหญิงมากกว่าเพศชาย ดังนั้น การตรวจสอบการทำหน้าที่ต่างกันจึงเป็นวิธีหนึ่งที่น่ามาใช้ตรวจสอบเพื่อช่วยพัฒนาแบบทดสอบให้มีคุณภาพ

สำหรับการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ มีวิธีหลายวิธีซึ่งวิธีที่นิยมใช้กันทั่วไปได้แก่ วิธีแมนเทล-แฮนเซล (Mantel – Haenszel; MH) โดยใช้คะแนนของแบบสอบเป็นตัวทำนายโอกาสการตอบข้อสอบได้ถูกต้อง วิธีถดถอยโลจิสติก (Logistic Regression ; LR) ใช้คะแนนรวมของแบบสอบเป็นตัวทำนายโอกาสการตอบข้อสอบได้ถูกต้อง วิธีชิปเทสต์ (SIBTEST) ใช้คะแนนรวมของแบบสอบเป็นตัวแทนความสามารถมีมิติการวัด 2 มิติ คือ มิติลักษณะแฝงเป้าหมาย (θ) และมิติแฝงแทรกซ้อน (η) และวิธีตรวจสอบด้วยทฤษฎี IRT ซึ่งประกอบด้วยวิธีวัดความแตกต่างของพื้นที่ วิธีความแตกต่างของค่าพารามิเตอร์ (b parameter difference) และวิธีทดสอบไค-สแควร์ของลอร์ด (Lord's χ^2 test) (ศิริชัย กาญจนวาสี, 2555, น. 115)

การศึกษาเกี่ยวกับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบมีนักวิจัยหลายท่านได้ศึกษาทั้งแบบทดสอบที่สร้างขึ้นเองหรือแบบทดสอบมาตรฐานที่ได้สร้างขึ้นโดยหน่วยงานอื่นๆ ของประเทศ ในสภาพข้อมูลจริงหรือการจำลองข้อมูลขึ้นมา อาทิ รัชรินทร์ มุกดา (2540) ได้เปรียบเทียบประสิทธิภาพระหว่างวิธีแมนเทล-แฮนเซลกับวิธีถดถอยโลจิสติก ในการตรวจสอบการทำหน้าที่ต่างกันของแบบสอบแบบอเนกรูป ผลการวิจัยพบว่า โดยภาพรวมวิธีแมนเทล-แฮนเซลกับวิธีถดถอยโลจิสติกมีประสิทธิภาพการตรวจสอบข้อสอบทำหน้าที่ต่างกันแบบอเนกรูปเท่าเทียมกันในทุกระดับความสามารถของผู้สอบ แต่วิธีถดถอยโลจิสติกมีอัตราความถูกต้องสูงกว่าวิธีแมนเทล-แฮนเซล และมีอัตราความคลาดเคลื่อนต่ำกว่าวิธีแมนเทล-แฮนเซล ซึ่งวิธีที่มีความเหมาะสมและควรนำไปใช้ตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูปคือ วิธีถดถอยโลจิสติก Narayanan and Swaminatan (1996) ได้ศึกษาเปรียบเทียบวิธีการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปซึ่งผลการวิจัยพบว่าวิธีแมนเทล-แฮนเซล มีอัตราการตรวจสอบต่ำกว่าวิธีถดถอยโลจิสติก และวิธี CRO-SIB และ วลีมาศ แซ่อึ้ง (2543) ทองอยู่ สาระ (2543) และอรินทร์ น่วมถนอม (2549) ได้ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่วัดความสามารถมิติเดียวและให้คะแนนสองค่า ผลการศึกษาสรุปได้ว่า วิธีชิปเทสต์ และวิธีแมนเทล-แฮนเซล สามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่เป็นรูปแบบเดียวกันแต่ไม่สามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ไม่เป็นรูปแบบเดียวกัน ส่วนวิธีการถดถอยโลจิสติกสามารถตรวจสอบได้ทั้งรูปแบบเดียวกันและไม่เป็นรูปแบบเดียวกัน ซึ่งจากการศึกษาเอกสารงานวิจัยพบว่าวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยวิธีการถดถอยโลจิสติกมีความยืดหยุ่น และมีประสิทธิภาพในการตรวจสอบดีที่สุด และ ธเกียรติกมล ทองงอก (2554) ได้เสนอแนะการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีถดถอยโลจิสติก ควรใช้ผลการทดสอบนี้สำคัญในการตัดสินข้อสอบที่

ทำหน้าที่ต่างกันร่วมกับผลการวัดขนาดอิทธิพลเนื่องจากการตรวจสอบการทำหน้าที่ต่างกันทราบเพียงว่าข้อสอบเกิดการทำหน้าที่ต่างกัน แต่ไม่ทราบสารสนเทศอื่น ซึ่งถือว่าข้อมูลยังไม่เพียงพอสำหรับการตัดสินใจข้อสอบว่าควรปรับปรุงแก้ไข หรือคงข้อสอบข้อนั้นไว้ในแบบสอบเพื่อเก็บเข้าคลังข้อสอบต่อไป

จากที่กล่าวมาข้างต้นและจากการศึกษางานวิจัย ผู้วิจัยมีความสนใจในการศึกษาการทำหน้าที่ต่างกันของข้อสอบวัดความสามารถด้านการคำนวณระหว่างนักเรียนชาย และนักเรียนหญิงของการทดสอบระดับชาติ เนื่องจากเพศ น่าจะเป็นตัวแปรที่มีผลต่อการทำหน้าที่ต่างกันของข้อสอบวัดความสามารถด้านการคำนวณ โดยทำการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการถดถอยโลจิสติกส์ร่วมกับการวัดขนาดอิทธิพล ซึ่งเป็นวิธีหนึ่งที่มีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกัน และผลที่ได้จากการศึกษาจะมีประโยชน์และความสำคัญในการเป็นแนวทางในการพัฒนาแบบทดสอบให้มีมาตรฐาน มีคุณภาพและมีความยุติธรรม ทำให้การประเมินผลความรู้ ความสามารถอย่างแท้จริง เป็นประโยชน์แก่ผู้บริหารระดับเขตพื้นที่การศึกษา และโรงเรียนเพื่อจะได้มีนโยบายส่งเสริมและแก้ปัญหาทางการเรียนให้ความช่วยเหลือแก่ผู้เรียนต่อไป

2. วัตถุประสงค์ของการวิจัย

ในการศึกษาค้นคว้าครั้งนี้ผู้วิจัยได้ตั้งวัตถุประสงค์ดังต่อไปนี้

- 2.1 เพื่อศึกษาการทำหน้าที่ต่างกันของข้อสอบวัดความสามารถด้านการคำนวณระหว่างนักเรียนชายและนักเรียนหญิงของการทดสอบระดับชาติ
- 2.2 เพื่อวิเคราะห์สาเหตุของข้อสอบวัดความสามารถด้านการคำนวณของการทดสอบระดับชาติ ที่ทำหน้าที่ต่างกัน

3. สมมติฐานของการวิจัย

ผู้เรียนชายมีโอกาสของการตอบข้อสอบวัดความสามารถด้านการคำนวณของการทดสอบระดับชาติถูกมากกว่านักเรียนหญิง

4. ขอบเขตของการวิจัย

4.1 ประชากรที่ใช้ในการศึกษา เป็นนักเรียนชั้นประถมศึกษาปีที่ 3 ปีการศึกษา 2557 จำนวน 464,532 คน เป็นเพศชาย 224,615 คน เป็นเพศหญิง 239,917 คน

4.2 ตัวแปรที่ใช้ในการวิจัย

4.2.1 *ตัวแปรอิสระ* คือ เพศ จำแนกได้ดังนี้

- 1) *กลุ่มอ้างอิง (Reference Group)* คือ นักเรียนเพศชาย
- 2) *กลุ่มเปรียบเทียบ (Focal Group)* คือ นักเรียนเพศหญิง

4.2.2 *ตัวแปรตาม* คือ ดัชนีการทำหน้าที่ต่างกันของข้อสอบ

4.3 *เครื่องมือและข้อมูลที่ใช้ในการวิจัย* คือ ข้อสอบ และ ผลการสอบวัดความสามารถด้านการคำนวณ ชั้นประถมศึกษาปีที่ 3 ของสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน ปีการศึกษา 2557

5. นิยามศัพท์เฉพาะ

5.1 *การทำหน้าที่ต่างกันของข้อสอบ (Differential Item Functioning)* หมายถึง ข้อสอบที่ทำให้ผู้สอบต่างกลุ่มกันที่มีความสามารถในระดับเดียวกัน มีโอกาสของการตอบข้อสอบข้อนั้นได้ถูกต้อง ไม่เท่ากัน จึงส่งผลให้ผู้สอบกลุ่มหนึ่งได้ประโยชน์ แต่ผู้สอบอีกกลุ่มหนึ่งนั้นเสียประโยชน์

5.2 *การไม่ทำหน้าที่ต่างกันของข้อสอบ* หมายถึง ข้อสอบที่ทำให้ผู้สอบต่างกลุ่มกันที่มีความสามารถในระดับเดียวกัน มีโอกาสของการตอบข้อสอบข้อนั้นได้ถูกต้องเท่ากัน

5.3 *กลุ่มอ้างอิง (Reference Group - R)* หมายถึง กลุ่มผู้สอบที่คาดว่าจะจะเป็นกลุ่มที่ได้ประโยชน์จาก ข้อสอบข้อสอบวัดความสามารถด้านการคำนวณ ซึ่งเป็นกลุ่มที่คาดว่าจะสามารถตอบข้อสอบถูกต้องมากกว่าผู้สอบอีกกลุ่มหนึ่ง ซึ่งการวิจัยครั้งนี้กลุ่มอ้างอิงคือนักเรียนเพศชาย

5.4 *กลุ่มเปรียบเทียบ (Focal Group - F)* หมายถึง กลุ่มผู้สอบที่คาดว่าจะจะเป็นกลุ่มที่เสียประโยชน์ข้อสอบวัดความสามารถด้านการคำนวณ คือ เป็นกลุ่มที่มีคาดว่าจะสามารถตอบข้อสอบถูกต้องน้อยกว่าผู้สอบกลุ่มอ้างอิง ที่มีคะแนนจากแบบทดสอบ

5.5 *ขนาดอิทธิพล (Effect size)* หมายถึง ขนาดของการทำหน้าที่ต่างกันของข้อสอบ ซึ่งจะตัดสินข้อสอบที่ทำหน้าที่ต่าง โดยค่าสถิติ R^2 เมื่อขนาดอิทธิพลมีขนาดปานกลาง และขนาดใหญ่ โดยเทียบกับเกณฑ์ Zumbo and Thomas

6. ประโยชน์ที่คาดว่าจะได้รับ

6.1 ทราบว่าข้อสอบข้อใดบ้างที่ทำหน้าที่ต่างกันระหว่างเพศ และสาเหตุการทำหน้าที่ต่างกันของข้อสอบซึ่งเป็นประโยชน์ต่อคณะกรรมการออกข้อสอบวัดผลสัมฤทธิ์ทางการเรียนของนักเรียนเพื่อเป็นข้อมูลในการสร้างแบบทดสอบเพื่อให้เกิดความยุติธรรมในครั้งต่อไป

6.2 เพื่อเป็นแนวทางในการศึกษาค้นคว้าเกี่ยวกับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบให้กับผู้ที่เกี่ยวข้อง และผู้ที่สนใจต่อไป



บทที่ 2

วรรณกรรมที่เกี่ยวข้อง

ผู้วิจัยได้ศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องกับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบดังนี้

1. คุณภาพของแบบทดสอบ
2. การทำหน้าที่ต่างกันของข้อสอบ
3. โครงการประเมินคุณภาพการศึกษาขั้นพื้นฐาน เพื่อการประกันคุณภาพ
4. งานวิจัยที่เกี่ยวข้อง

1. คุณภาพของแบบทดสอบ

แบบทดสอบที่มีคุณภาพต้องมีการตรวจสอบคุณภาพแบบทดสอบก่อนการวัดและประเมินผล เพื่อให้แน่ใจว่าจะสามารถให้ข้อมูลที่ถูกต้องเหมาะสม ยุติธรรมสำหรับทุกคน เครื่องมือวัดที่มีคุณภาพยอมให้ข้อมูลหรือคะแนนที่น่าเชื่อถือได้ และเมื่อนำคะแนนไปใช้ในการประเมินผลย่อมมีความถูกต้องและเชื่อถือได้ การตรวจสอบคุณภาพแบบทดสอบควรดำเนินการในประเด็นต่อไปนี้

1.1 ตรวจสอบคุณภาพโดยการพิจารณาความเหมาะสม เป็นการตรวจสอบคุณภาพโดยใช้คุณลักษณะพิจารณาตามความเหมาะสมถูกต้องตามหลักการ โดยผู้สอนหรือให้ผู้อื่นที่มีความเชี่ยวชาญเป็นผู้พิจารณา ซึ่งพิจารณาในประเด็นต่อไปนี้

1.1.1 ตรวจสอบข้อคำถามตรงตามพฤติกรรมหรือเนื้อหาที่ต้องการวัดหรือไม่ โดยการพิจารณาข้อคำถามในแต่ละข้อ โดยตรวจสอบความตรงตามโครงสร้าง เพื่อพิจารณาว่าข้อคำถามออกได้ครอบคลุมพฤติกรรมการเรียนรู้ที่กำหนดไว้หรือไม่

1.1.2 ตรวจสอบความถูกต้องตามหลักวิชา ข้อคำถามทุกข้อที่ถามควรมีความถูกต้องตามหลักวิชาแต่ละวิชาซึ่งควรที่จะตรวจสอบก่อนนำแบบทดสอบไปใช้

1.1.3 ตรวจสอบความเป็นปรนัยของข้อคำถาม คือตรวจสอบความชัดเจนหรือความรัดกุมของภาษาที่ใช้ ว่าอ่านตรงกันหรือไม่ หรือตรวจสอบความซ้ำซ้อน หรือการขึ้นาคำตอบในแต่ละข้อซึ่งข้อคำถามแต่ละข้อต้องมีความชัดเจน เพื่อความเข้าใจที่ตรงกันของทุกคน

1.1.4 ตรวจสอบความสำคัญหรือคุณค่าของข้อคำถาม ข้อคำถามที่ถามควรเป็นประเด็นสำคัญ หรือประเด็นหลักที่มีคุณค่า ไม่ควรถามประเด็นปลีกย่อยมากๆ เช่น ถามชื่อบุคคล ถามเหตุการณ์ที่เกิดในพ.ศ.ใด ซึ่งเป็นลักษณะถามความจำ จึงควรพิจารณาว่าต้องการถามความจำหรือระดับพฤติกรรมที่สูงกว่านั้น จึงควรเขียนข้อคำถามให้สอดคล้องกับพฤติกรรมที่ต้องการวัด

1.1.5 ตรวจสอบความเหมาะสมของเวลา และระดับของผู้เรียน โดยพิจารณาว่าจำนวนข้อสอบมีความเหมาะสมกับเวลาหรือไม่

1.2 ตรวจสอบคุณภาพโดยการทดลองใช้ (try-out) เป็นการนำแบบทดสอบที่จะดำเนินการวัดและประเมินผลไปใช้กับกลุ่มเป้าหมายที่มีลักษณะใกล้เคียงกับกลุ่มตัวอย่างที่จะเก็บข้อมูลจริงเพื่อนำผลมาวิเคราะห์คุณภาพของเครื่องมือก่อนนำไปใช้จริง สามารถตรวจได้ดังนี้

1.2.1 วิเคราะห์คุณภาพรายข้อของแบบทดสอบ คือการหาค่าความยาก และค่าอำนาจจำแนก ในการทดสอบ โดยทั่วไปจะแบ่งการทดสอบเป็นสองลักษณะ คือการทดสอบแบบอิงกลุ่ม และอิงเกณฑ์ โดยดังรายละเอียดต่อไปนี้

1) ค่าความยากง่าย

(1) ค่าความยากง่ายของแบบทดสอบอิงกลุ่ม หมายถึง สัดส่วนหรือเปอร์เซ็นต์ของผู้สอบที่ตอบแต่ละข้อคำถามถูก สัญลักษณ์ที่ใช้แทนค่าความยาก คือ “p”

สูตรในการหาค่าความยากง่าย $P = \frac{R}{N}$

เมื่อ	p	แทน	ค่าความยาก
	R	แทน	จำนวนผู้สอบที่ตอบแต่ละข้อคำถามถูก
	N	แทน	จำนวนผู้สอบทั้งหมด

ในกรณีที่มีผู้สอบไม่มากเป็นจำนวนหลักร้อย การวิเคราะห์ค่าความยากจะนำจำนวนผู้เข้าสอบทั้งหมดมาคำนวณ แต่ถ้ากรณีที่มีผู้สอบมีจำนวนมากอาจใช้วิธีแบ่งกลุ่มผู้สอบตามคะแนนรวมโดยแบ่งกลุ่มผู้สอบที่ได้คะแนนสูง ร้อยละ 27 หรือ 50 และกลุ่มที่ได้คะแนนรวมต่ำ ร้อยละ 27 หรือ 50 โดยใช้การคำนวณค่าความยากรายข้อดังนี้

สูตรในการหาค่าความยากง่าย $P = \frac{H + L}{N_H + N_L}$

เมื่อ	p	แทน	ค่าความยาก
	H	แทน	จำนวนผู้ตอบถูกในกลุ่มสูง
	L	แทน	จำนวนผู้ตอบถูกในกลุ่มต่ำ
	N _H	แทน	จำนวนผู้สอบทั้งหมดในกลุ่มสูง
	N _L	แทน	จำนวนผู้สอบทั้งหมดในกลุ่มต่ำ

(2) ค่าความยากของแบบทดสอบอิงเกณฑ์ จะใช้สูตรการหาเช่นเดียวกับการหาค่าความยากของแบบทดสอบอิงกลุ่ม

การแปลความหมายค่าความยาก ค่าความยากมีค่าตั้งแต่ 0 - 1 ถ้ามีค่า 0 แสดงว่าไม่มีผู้สอบคนใดตอบข้อสอบข้อนั้นถูกเลย หมายความว่าข้อสอบข้อนั้นยากที่สุด ถ้ามีค่า 1 แสดงว่าผู้สอบทำข้อสอบข้อนั้นถูกทุกคน หมายความว่าข้อสอบข้อนั้นง่ายที่สุด ตามอุดมคติข้อที่มีความยากง่ายพอเหมาะคือ ข้อที่มีความยากง่ายเท่ากับ 0.5 ซึ่งการที่จะสร้างข้อสอบที่มีความยาก 0.5 เป็นเรื่องยาก ฉะนั้นในการพิจารณาค่าความยากที่เหมาะสมจึงกำหนดว่าค่าควรมีค่าระหว่าง 0.2 - 0.8

2) ค่าอำนาจจำแนก แยกเป็นค่าอำนาจจำแนก ของแบบทดสอบอิงกลุ่มและอิงเกณฑ์

(1) ค่าอำนาจจำแนกของแบบทดสอบอิงกลุ่ม หมายถึง ความสามารถของข้อสอบที่จะจำแนกให้เห็นถึงความแตกต่างระหว่างบุคคล โดยยึดหลักการว่า คนเก่งจะต้องทำข้อสอบข้อนั้นถูก คนไม่เก่งต้องทำผิด สัญลักษณ์ที่ใช้แทนค่าอำนาจจำแนก คือ “r”

สูตรในการหาค่าอำนาจจำแนก

ค่าอำนาจจำแนกของตัวเลือกที่เป็นตัวถูก

$$r = \frac{H-L}{N_H} \text{ หรือ } r = \frac{H-L}{N_L}$$

ค่าอำนาจจำแนกของตัวเลือกที่เป็นตัวลวง

$$r = \frac{L-H}{N_H} \text{ หรือ } r = \frac{L-H}{N_L}$$

(ในกรณีที่จำนวนกลุ่มสูงและกลุ่มต่ำมีจำนวนเท่ากัน)

เมื่อ r แทน ค่าอำนาจจำแนก

H แทน จำนวนผู้สอบในกลุ่มสูงที่เลือกตัวเลือกนั้น

L แทน จำนวนผู้สอบในกลุ่มต่ำที่เลือกตัวเลือกนั้น

N_H แทน จำนวนผู้สอบในกลุ่มสูงทั้งหมด

N_L แทน จำนวนผู้สอบในกลุ่มต่ำทั้งหมด

การแปลความหมายค่าอำนาจจำแนก มีค่าตั้งแต่ -1.0 ถึง +1.0 ค่าอำนาจจำแนกติดลบหมายความว่าคนที่ไม่เก่งทำข้อสอบถูกมากกว่าคนเก่งค่าอำนาจจำแนกเป็น 0 หมายความว่าคนเก่งและคนไม่เก่งทำข้อสอบถูกเท่ากันถ้าอำนาจจำแนกเป็นบวก หมายความว่าคนเก่งทำข้อสอบถูกมากกว่าคนเก่งและยิ่งมีค่าตัวเลขมากก็ยิ่งแยกคนเก่งออกจากคนไม่เก่งได้มากเท่านั้น สำหรับค่าอำนาจจำแนกที่เหมาะสมควรเท่ากับ 0.20 ขึ้นไป

(2) การหาค่าอำนาจจำแนกของแบบทดสอบอิงเกณฑ์ คือ ประสิทธิภาพของข้อสอบแต่ละข้อที่สามารถจำแนกกลุ่มผู้เรียนออกเป็นกลุ่มรอบรู้และไม่รอบ วิธีการคำนวณหาอำนาจจำแนกสามารถดำเนินการได้หลายวิธี ดังต่อไปนี้

ก. ดัชนีความไว (sensitivity index) เป็นการวิเคราะห์โดยใช้ความแตกต่างระหว่างการสอบก่อนสอนและการสอบหลังสอนมีสูตรการคำนวณดังนี้

สูตรในการหาค่าดัชนีความไว

$$S = \frac{R_{post} - R_{pre}}{N} \quad (\text{กรณีตัวเลือกเป็นตัวเลือก})$$

$$S = \frac{R_{post} - R_{pre}}{N} \quad (\text{กรณีตัวเลือกเป็นตัวเลือก})$$

เมื่อ S แทน ดัชนีความไว

R_{post} แทน จำนวนผู้ตอบถูกหลังสอน

R_{pre} แทน จำนวนผู้ตอบถูกก่อนสอน

N แทน จำนวนนักเรียนทั้งหมด

ในการพิจารณาค่าอำนาจจำแนก (S) ถ้ามีค่าเป็นลบ หมายถึง ก่อนเรียนผู้เรียนมีความรู้ แต่เมื่อเรียนจบเนื้อหาปรากฏว่าผู้เรียนไม่มีความรู้เลย ค่า S มีค่าเป็นศูนย์ หมายถึง เมื่อเรียนจบเนื้อหาปรากฏว่าผู้เรียนมีความรู้เท่ากับความรู้ก่อนเรียน และถ้าค่า S เป็นบวก หมายถึง ก่อนเรียนผู้เรียนไม่มีความรู้ หลังเรียนแล้วปรากฏว่ามีความรู้ตามจุดประสงค์ที่ตั้งไว้

การแปลความหมายค่าดัชนีความไว มีเกณฑ์การพิจารณา ดังนี้

ค่าดัชนีความไว (S)	ความหมาย
1.0	เป็นข้อสอบที่ดี เป็นไปตามทฤษฎี
0.80 - 0.99	เป็นข้อสอบที่ดี หาได้ในเชิงปฏิบัติ
0.30 - 0.79	เป็นข้อสอบที่พอใช้ได้
0.00 - 0.29	เป็นข้อสอบที่ไม่ดี ควรตัดทิ้ง
-1.00 - 0.00	เป็นข้อสอบที่ใช้ไม่ได้ ควรตัดทิ้ง

ข. ดัชนีบี (B-Index) เปรนแนน (Brannan) ได้พัฒนาสูตรการหาค่าอำนาจจำแนกที่เรียกว่า โดยมีสูตรการคำนวณ ดังนี้

$$\text{สูตรในการหาค่าดัชนีบี} \quad B = \frac{H}{N_H} - \frac{L}{N_L}$$

เมื่อ B แทน ดัชนีค่าอำนาจจำแนก

H แทน จำนวนนักเรียนในกลุ่มรอบรู้ที่ตอบ ข้อสอบข้อนั้นถูก

L แทน จำนวนนักเรียนในกลุ่มไม่รอบรู้ที่ตอบ ข้อสอบข้อนั้นถูก

N_H แทน จำนวนนักเรียนในกลุ่มรอบรู้ทั้งหมด

N_L แทน จำนวนนักเรียนในกลุ่มไม่รอบรู้ทั้งหมด

การแปลความหมายค่าดัชนีบี มีเกณฑ์ในการพิจารณา ดังนี้

ค่าดัชนีบี ความหมาย

1.00 บ่งชี้ผู้รอบรู้-ไม่รอบรู้ได้ถูกต้องทุกคน

0.50-0.99 บ่งชี้ผู้รอบรู้-ไม่รอบรู้ได้ถูกต้องเป็นบางส่วน

0.20-0.49 บ่งชี้ผู้รอบรู้-ไม่รอบรู้ได้เป็นบางส่วน

0.00-0.19 บ่งชี้ผู้รอบรู้-ไม่รอบรู้ได้ถูกต้องน้อยมาก หรือไม่ถูกต้องเลย

ต่ำกว่า 0.00 บ่งชี้ผู้รอบรู้-ไม่รอบรู้ผิดพลาด หรือตรงกันข้ามกับความเป็นจริง

3) การตรวจสอบความเที่ยง ความเที่ยง (reliability) หมายถึง ความคงเส้นคงวาของผลการวัด ไม่ว่าจะทำการวัดเมื่อใดก็ตาม ซึ่งวิธีที่นิยมใช้ในการตรวจสอบความเที่ยงของแบบทดสอบคือ วิธีของคูเดอร์-ริชาร์ด (Kuder- Richardson) เป็นการหาความเที่ยงที่คูเดอร์และริชาร์ดได้เสนอไว้และนิยมใช้ คือ KR-20 ใช้กับแบบทดสอบที่มีคะแนนแบบ 0 กับ 1

$$\text{สูตร } KR-20 = \left[\frac{K}{K-1} \right] \left[1 - \frac{\sum p(1-p)}{s_x^2} \right]$$

เมื่อ KR-20 แทน ค่าความเที่ยง

K แทน จำนวนข้อสอบ

P แทน สัดส่วนของผู้ตอบข้อสอบถูก

s_x^2 แทน ค่าแปรปรวนของคะแนนรวม

4) การตรวจสอบความตรง (validity) หมายถึง ความสามารถในการวัดสิ่งที่ต้องการจะวัด การตรวจสอบความตรงสามารถตรวจสอบได้หลายวิธี ได้แก่ ความตรงตามโครงสร้าง ความตรงตามเนื้อหา ความตรงตามเกณฑ์สัมพัทธ์ในที่นี้จะนำเสนอการตรวจสอบความตรงเชิงเนื้อหา ซึ่งเป็นวิธีที่นิยมใช้

ความตรงเชิงเนื้อหา (content validity) เป็นการพิจารณาว่าข้อคำถามเป็นตัวแทนของเนื้อหาทั้งหมดที่ต้องการจะวัดได้หรือไม่เพียงใด หรือเป็นการพิจารณาว่าข้อคำถามแต่ละข้อถามตรงตามเนื้อหาและพฤติกรรมที่ต้องการให้วัดหรือไม่ โดยอาศัยทักษะและประสบการณ์ของผู้เชี่ยวชาญด้านเนื้อหาพิจารณาว่าข้อคำถามแต่ละข้อสอดคล้องกับวัตถุประสงค์เชิงพฤติกรรมที่ต้องการ

วัดหรือไม่ โดยคำนวณจากดัชนีความสอดคล้องระหว่างข้อสอบแต่ละข้อ (Index of Item Objective Congruence : IOC)

การให้ผู้เชี่ยวชาญพิจารณาความตรงเชิงเนื้อหา ควรให้ผู้เชี่ยวชาญแสดงความคิดเห็นแล้วนำความคิดเห็นของผู้เชี่ยวชาญมาประมวลเพื่อสรุปผล โดยพิจารณาจากความคิดเห็นของผู้เชี่ยวชาญ ดังนี้

ให้ 1 เมื่อแน่ใจว่าข้อคำถามสอดคล้องกับวัตถุประสงค์เชิงพฤติกรรมที่ต้องการวัด

ให้ 0 เมื่อไม่แน่ใจว่าข้อคำถามสอดคล้องกับวัตถุประสงค์เชิงพฤติกรรมที่ต้องการวัด

ให้ -1 เมื่อแน่ใจว่าข้อคำถามไม่สอดคล้องกับวัตถุประสงค์เชิงพฤติกรรมที่ต้องการวัด

การคำนวณผลสรุปความคิดเห็นของผู้เชี่ยวชาญ สามารถคำนวณจากสูตร IOC

$$\text{สูตรการคำนวณ} \quad IOC = \frac{\sum R}{N}$$

เมื่อ IOC แทน ดัชนีความสอดคล้องระหว่างข้อสอบระหว่างข้อคำถามกับจุดประสงค์การเรียนรู้

R แทน ความคิดเห็นของผู้เชี่ยวชาญ

N แทน จำนวนผู้เชี่ยวชาญ

ถ้าค่า IOC มีค่าเท่ากับหรือมากกว่า 0.5 แสดงว่าข้อคำถามสอดคล้องกับวัตถุประสงค์เชิงพฤติกรรมที่ต้องการวัด (มหาวิทยาลัยสุโขทัยธรรมมาธิราช, 2554, น.37-50)

2. การทำหน้าที่ต่างกันของข้อสอบ

การทำหน้าที่ต่างกันของข้อสอบ (DIF) เป็นกระบวนการส่วนหนึ่งในการพิจารณาความลำเอียงของข้อสอบซึ่งการทำหน้าที่ต่างกันของข้อสอบ เป็นกระบวนการที่ใช้วิธีการทางสถิติสำหรับตรวจสอบเพื่อให้ได้สารสนเทศเกี่ยวกับการทำหน้าที่ของข้อสอบสำหรับกลุ่มผู้สอบย่อยที่มีลักษณะเฉพาะบางอย่างที่แตกต่างกัน ส่วนการพิจารณาความลำเอียงของข้อสอบนั้นเป็นกระบวนการในการตัดสินความยุติธรรมของข้อสอบ โดยการนำสารสนเทศการทำหน้าที่ต่างกันของข้อสอบมาวิเคราะห์เชิงตรรกะ แล้วให้ผู้เชี่ยวชาญพิจารณาถึงการเขียนข้อสอบ เนื้อหาสาระของข้อสอบ และจุดมุ่งหมายของการวัด เพื่อสรุปว่าข้อสอบข้อนั้นลำเอียงเข้าข้างกลุ่มใดกลุ่มหนึ่งหรือไม่ (Camilli & Shapard, 1994; ศิริชัย กาญจนวาสิ, 2555, น.116)

2.1 ความหมายของการทำหน้าที่ต่างกันของข้อสอบ

นักวิจัยทางการวัดผลการศึกษาลายท่าน ได้ให้ความหมายของการทำหน้าที่ต่างกันของข้อสอบไว้ดังนี้

Shealy and Stout (1993) กล่าวว่า การทำหน้าที่ต่างกันของข้อสอบ หมายถึง ข้อสอบเข้าข้างผู้สอบกลุ่มหนึ่งมากกว่าผู้สอบอีกกลุ่มหนึ่งที่นำมาจับคู่เปรียบเทียบ ซึ่งส่งผลให้ผู้สอบกลุ่มหนึ่งได้ประโยชน์ แต่ผู้สอบอีกกลุ่มหนึ่งนั้นเสียประโยชน์

Holland and Wainer (1993) กล่าวว่า การทำหน้าที่ต่างกันของข้อสอบ หมายถึง สารสนเทศทางสถิติของข้อสอบ ที่ได้จากการทำข้อสอบของผู้สอบต่างกลุ่มกันที่มีความสามารถเท่ากัน แต่มีโอกาสตอบข้อสอบถูกต้องต่างกัน

Camilli and Shapard (1994) กล่าวว่า การทำหน้าที่ต่างกันของข้อสอบ หมายถึง ความเป็นพหุมิติในการวัดของข้อสอบ ซึ่งได้จากการแจกแจงความสามารถหลัก (Primary ability) ของกลุ่มผู้สอบตั้งแต่ 2 กลุ่มขึ้นไปที่มีความเท่ากัน แต่มีการแจกแจงความสามารถรอง (Secondary ability) ที่แตกต่างกัน

Potanza and Dorans (1995) กล่าวว่า การทำหน้าที่ต่างกันของข้อสอบ หมายถึง ข้อสอบที่ทำหน้าที่ต่างกัน ผลการตอบข้อสอบระหว่างกลุ่มผู้สอบสองกลุ่มย่อยที่นำมาเปรียบเทียบมีความแตกต่างกัน

Narayanan and Swaminathan (1996) กล่าวว่า การทำหน้าที่ต่างกันของข้อสอบ หมายถึง ฟังก์ชันการตอบสนองข้อสอบมีค่าไม่เท่ากัน ซึ่งคำนวณจากกลุ่มของผู้สอบมีความสามารถระดับเดียวกันแต่มาจากกลุ่มย่อยต่างกัน

McCallon and Schumacker (2002) กล่าวว่า การทำหน้าที่ต่างกันของข้อสอบ หมายถึง ลักษณะที่แสดงถึงความไม่ยุติธรรมของข้อสอบ โดยเกิดขึ้นจากผู้สอบที่มีความรู้ความสามารถเท่ากัน แต่มาจากกลุ่มที่ต่างกัน และได้รับการตอบสนองจากข้อสอบเกี่ยวกับโอกาสในการตอบข้อสอบได้ถูกต้องแตกต่างกัน

ศิริชัย กาญจนวาสิ (2550) กล่าวว่า การทำหน้าที่ต่างกันของข้อสอบ หมายถึง ข้อสอบที่ทำให้ผู้สอบจากต่างกลุ่มกันที่มีความสามารถหรือคุณลักษณะที่มุ่งวัดเท่ากัน มีโอกาสในการตอบข้อสอบได้ถูกต้องแตกต่างกัน หรือมีฟังก์ชันการตอบสนองข้อสอบแตกต่างกัน การทำหน้าที่ต่างกันของข้อสอบเกิดขึ้นเมื่อนำข้อสอบไปทดสอบกับผู้สอบกลุ่มย่อยต่างกัน ที่มีความสามารถหลัก (Primary ability) ระดับเดียวกันหรือมีคุณลักษณะแฝง (Latent trait) ที่ต้องการวัดเท่ากัน แต่มีความสามารถรอง (Secondary ability) แตกต่างกัน ทำให้ผู้สอบต่างกลุ่มที่นำมาจับคู่เปรียบเทียบมีโอกาสตอบข้อสอบถูกต้องแตกต่างกัน

ชเกียรติกมล ทองงอก (2554) กล่าวว่า การทำหน้าที่ต่างกันของข้อสอบ หมายถึง ผู้สอบกลุ่มย่อยต่างกันที่มีความสามารถหลักระดับเดียวกัน หรือมีคุณลักษณะแฝง (Latent Trait) ที่ต้องการวัดเท่ากัน แต่มีความสามารถรองแตกต่างกันทำให้ผู้สอบต่างกลุ่มที่นำมาจับคู่เปรียบเทียบมีโอกาสตอบถูกแตกต่างกัน

สุชาติ สิริมินันท์ (2554) กล่าวว่า การทำหน้าที่ต่างกันของข้อสอบ หมายถึง ความแตกต่างในการทำหน้าที่ของข้อสอบ หลังจากกลุ่มผู้สอบที่แตกต่างกันมีความสามารถ หรือคุณลักษณะที่ข้อสอบนั้นวัดเท่ากัน มีความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องไม่เท่ากัน

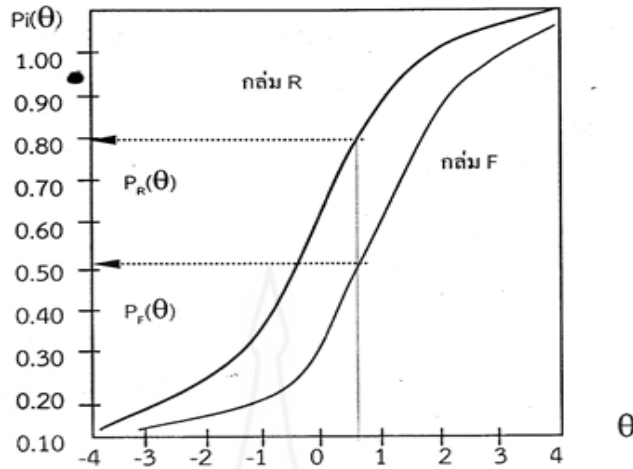
ดังนั้นจึงสรุปได้ว่าการทำหน้าที่ต่างกันของข้อสอบหมายถึง ข้อสอบที่ทำให้ผู้สอบต่างกลุ่มกันที่มีความสามารถในระดับเดียวกัน มีโอกาสของการตอบข้อสอบข้อนั้น ได้ถูกต้องไม่เท่ากัน จึงส่งผลให้ผู้สอบกลุ่มหนึ่งได้ประโยชน์ แต่ผู้สอบอีกกลุ่มหนึ่งนั้นเสียประโยชน์ และการไม่ทำหน้าที่ต่างกันของข้อสอบ หมายถึง ข้อสอบที่ทำให้ผู้สอบต่างกลุ่มกันที่มีความสามารถในระดับเดียวกัน มีโอกาสของการตอบข้อสอบข้อนั้น ได้ถูกต้องเท่ากันซึ่งไม่ส่งผลให้กลุ่มใดกลุ่มหนึ่งเสียประโยชน์

2.2 ประเภทของการทำหน้าที่ต่างกันของข้อสอบ

การทำหน้าที่ต่างกันของข้อสอบ เป็นการเปรียบเทียบผลการตอบข้อสอบระหว่างกลุ่มผู้สอบอย่างน้อย 2 กลุ่มขึ้นไป ซึ่งโดยปกตินิยมทำการเปรียบเทียบ 2 กลุ่ม คือกลุ่มเปรียบเทียบ (Focal group หรือกลุ่ม F) เป็นกลุ่มที่สนใจศึกษาและคาดว่าจะจะเป็นกลุ่มที่เสียเปรียบในการตอบข้อสอบ และกลุ่มอ้างอิง (Reference group หรือกลุ่ม R) เป็นกลุ่มที่คาดว่าจะได้เปรียบในการตอบข้อสอบได้ถูกต้อง

ในการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบจะพบว่า ข้อสอบสามารถทำหน้าที่แตกต่างกันได้ 2 ประเภท (Mellenbergh, 1982 ; สิริชัย กาญจนวาสิ ,2555, น.118) ได้แก่ การทำหน้าที่ต่างกันของข้อสอบแบบเอกรูป (Uniform) และแบบอนกรูป (Nonuniform)

2.2.1 ข้อสอบทำหน้าที่ต่างกันแบบเอกรูป (Uniform DIF) หมายถึง การที่ข้อสอบที่ทำให้ผู้สอบกลุ่มหนึ่งมีโอกาสตอบข้อสอบถูกมากกว่าผู้สอบอีกกลุ่มหนึ่งอย่างสม่ำเสมอ ในทุกระดับความสามารถเมื่อพิจารณาโครงสร้างคุณลักษณะข้อสอบของผู้สอบสองกลุ่มจะพบว่าไม่มีปฏิสัมพันธ์ระหว่างระดับความสามารถของผู้สอบกับการเป็นสมาชิกของกลุ่ม (Group membership) ดังภาพที่ 2.1

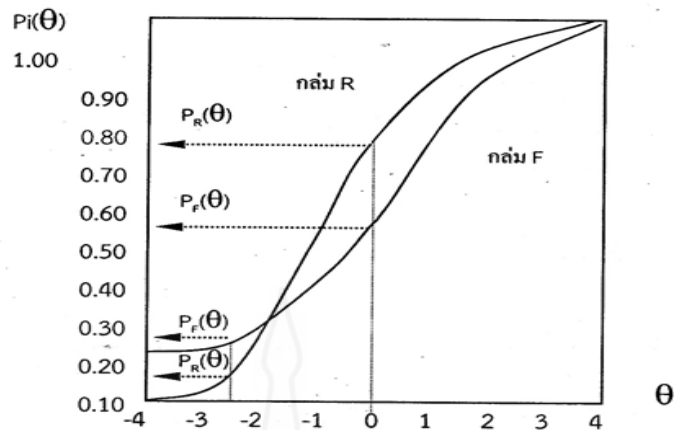


ภาพที่ 2.1 ข้อสอบการทำหน้าที่ต่างกันแบบเอกรูป (Uniform DIF) ที่มา: ศิริชัย กาญจนวาสี (2555:118)

2.2.2 ข้อสอบทำหน้าที่ต่างกันแบบอนกรุป (Nonuniform DIF) หมายถึง ข้อสอบที่ทำให้ผู้สอบระหว่างกลุ่มแตกต่างกันมีโอกาสในการตอบข้อสอบถูกต้องอย่างไม่สม่ำเสมอในทุกระดับความสามารถเมื่อพิจารณาไค์คุณลักษณะข้อสอบของผู้สอบ 2 กลุ่ม พบว่ามีปฏิสัมพันธ์ร่วมกันระหว่างระดับความสามารถของผู้สอบกับการเป็นสมาชิกของกลุ่ม เช่น ที่ระดับความสามารถหนึ่ง กลุ่มอ้างอิง (กลุ่ม R) มีโอกาสในการตอบข้อสอบข้อนั้นถูกต้องมากกว่ากลุ่มผู้สอบกลุ่มเปรียบเทียบ (กลุ่ม F) แต่ระดับความสามารถอีกระดับหนึ่งกลุ่มผู้สอบกลุ่มเปรียบเทียบมีโอกาสในการตอบข้อสอบถูกต้องมากกว่ากลุ่มผู้สอบกลุ่มอ้างอิง

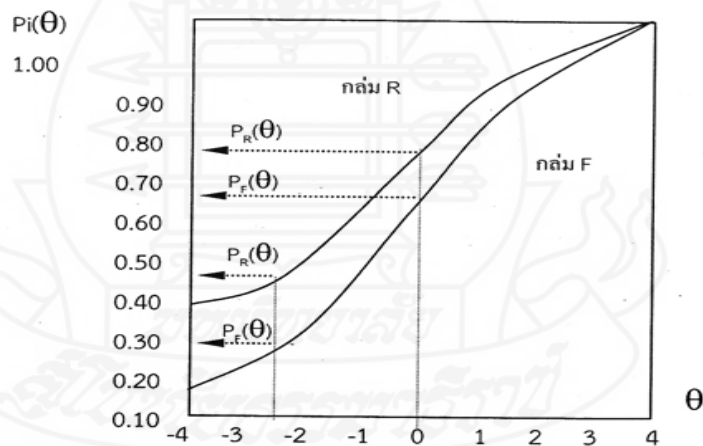
ข้อสอบที่ทำหน้าที่ต่างกันแบบอนกรุปสามารถจำแนกได้เป็น 2 ลักษณะ (Swaminathan and Rogers, 1990; ศิริชัย กาญจนวาสี , 2555, น.119-120) ดังนี้

1) ข้อสอบทำหน้าที่ต่างกับแบบอนกรุปโดยมีปฏิสัมพันธ์ไม่เป็นลำดับ (Disordinal interaction) เป็นการทำหน้าที่ต่างกันของข้อสอบสำหรับกลุ่มผู้สอบซึ่งเกิดขึ้น เมื่อไค์คุณลักษณะข้อสอบตัดกันระหว่างช่วงความสามารถของผู้สอบหรือเรียกว่าข้อสอบทำหน้าที่ต่างกันแบบไม่มีทิศทาง (Non-Unidirectional DIF) ดังภาพที่ 2.2



ภาพที่ 2.2 ข้อสอบทำหน้าที่ต่างกันแบบไม่มีทิศทาง (Non-unidirectional DIF) ที่มา: ศิริชัย กาญจนวาสี (2555:119)

2) ข้อสอบทำหน้าที่ต่างกับแบบอนุกรม โดยมีปฏิสัมพันธ์เป็นลำดับ (Ordinal interaction) เป็นการทำหน้าที่ต่างกันของกลุ่มผู้สอบซึ่งเกิดขึ้น เมื่อโค้งลักษณะข้อสอบต่างกันอย่างไม่สม่ำเสมอแต่ไม่ตัดกัน หรืออาจตัดกันนอกช่วงความสามารถของผู้สอบตรงปลายสุดของช่วงความสามารถต่ำหรือสูง อาจเรียกข้อสอบลักษณะนี้ว่า ข้อสอบทำหน้าที่ต่างกันแบบมีทิศทางเดียวกัน (Unidirectional DIF) ดังภาพที่ 2.3



ภาพที่ 2.3 ข้อสอบทำหน้าที่ต่างกันแบบมีทิศทางเดียวกัน (Unidirectional DIF) ที่มา: ศิริชัย กาญจนวาสี (2555:120)

2.3 หลักการตรวจการทำหน้าที่ต่างกันของข้อสอบ

การตรวจการทำหน้าที่ต่างกันของข้อสอบ (DIF detection) เป็นการเปรียบเทียบผลการตอบข้อสอบเป็นรายข้อระหว่างกลุ่มของผู้สอบอย่างน้อย 2 กลุ่ม ที่มีความสามารถหลัก (primary ability) ที่มุ่งวัดเท่ากันแต่คาดว่าจะมีความได้เปรียบหรือเสียเปรียบกันโดยกลุ่มหนึ่งถือเป็นกลุ่มอ้างอิง (Reference group) ซึ่งคาดว่าจะได้เปรียบในการตอบข้อสอบข้อนั้น หรือมีโอกาสตอบข้อสอบได้

ถูกต้องมากกว่า ส่วนอีกกลุ่มคือ กลุ่มเปรียบเทียบ (Focal group) ซึ่งเป็นกลุ่มที่สนใจศึกษาและคาดว่าน่าจะเป็นกลุ่มที่เสียเปรียบ

ในการเปรียบเทียบผลการตอบข้อสอบระหว่างกลุ่มอ้างอิง และกลุ่มเปรียบเทียบ จำเป็นต้องจับคู่ (matching) ผู้สอบตามความสามารถ ซึ่งเป็นเงื่อนไขสำคัญของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ เกณฑ์การจับคู่ (matching criteria) ที่นิยมใช้กันมี 2 วิธีดังนี้

2.3.1 เกณฑ์ภายนอก (External Criterion)

การวิเคราะห์การทำหน้าที่ต่างกัน โดยใช้เกณฑ์ภายนอกนี้ สามารถนำไปใช้ได้ทั้งข้อสอบรายข้อ และแบบสอบทั้งฉบับ โดยการใส่คะแนนจากแบบทดสอบอื่นเป็นเกณฑ์ภายนอกแล้ว ใช้เทคนิคการวิเคราะห์การถดถอย (Regression analysis) เพื่อทำการเปรียบเทียบเส้นกราฟความสัมพันธ์ระหว่างตัวแปรเกณฑ์กับตัวแปรทำนายระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ

หลักการนี้มีจุดมุ่งหมายเพื่อสร้างสมการทำหน้าที่ตัวแปรเกณฑ์ ซึ่งเป็นคะแนนของแบบสอบอื่นจากตัวแปรทำนายซึ่งเป็นคะแนนรายข้อ หรือคะแนนแบบสอบระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบในการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ จะใช้คะแนนรายข้อเป็นตัวแปรทำนาย แต่ถ้าเป็นการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบ จะใช้คะแนนรวมของแบบสอบทั้งฉบับเป็นตัวแปรทำนาย สำหรับตัวแปรเกณฑ์ที่ใช้เป็นเกณฑ์ภายนอก อาจใช้คะแนนรวมทั้งฉบับ หรือเกรดเฉลี่ย หรือผลสัมฤทธิ์ในงานที่เกี่ยวข้องของผู้สอบ (Cronbach, 1970 ; ศิริชัย กาญจนวาสี, 2555, น.121-122) ข้อดีการใช้เกณฑ์ภายนอก คือ เกณฑ์ที่ใช้มีความเป็นอิสระจากข้อสอบ แต่ถ้าเกณฑ์ภายนอกที่นำมาในการหาความสัมพันธ์นั้น ไม่มีมาตรฐานจะส่งผลให้การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบหรือแบบสอบขาดความถูกต้อง และแม่นยำ เพราะในทางปฏิบัติเป็นการยากที่จะหาตัวแปรเกณฑ์ภายนอกจากแบบสอบฉบับอื่นที่มีความตรงเชิงทำนายและมีความยุติธรรมมาใช้ในการหาความสัมพันธ์

สมการทำนายสำหรับกลุ่มอ้างอิงและกลุ่มเปรียบเทียบแสดงได้ดังนี้

$$\text{กลุ่มอ้างอิง (R) } Y_i = AR + BR X_i$$

$$\text{กลุ่มเปรียบเทียบ (F) } Y_i = AF + BF X_i$$

เมื่อ Y_i แทน คะแนนของตัวแปรเกณฑ์ภายนอก

X_i แทน คะแนนของตัวแปรทำนาย

A แทน ค่าคงที่หรือค่าตัดแกน(intercept)

B แทน ค่าความชัน (slope)

จากฟังก์ชันการทำนายทั้ง 2 สมการ สามารถเปรียบเทียบค่าตัดแกน และค่าความชัน ของเส้นกราฟระหว่างกลุ่มอ้างอิง และกลุ่มเปรียบเทียบ ได้ถ้าเส้นกราฟดังกล่าวมีค่าความชัน

หรือค่าตัดแกน แตกต่างกัน สำหรับข้อสอบใด หรือแบบสอบใดแสดงว่าข้อสอบหรือแบบสอบนั้นมีการทำหน้าที่ต่างกัน โดยเข้าข้างกลุ่มผู้สอบที่มีค่าตัดเกณฑ์หรือค่าความชันที่สูงกว่า

2.3.2 เกณฑ์ภายใน (Internal Criterion) การวิเคราะห์การทำหน้าที่ต่างกัน โดยใช้เกณฑ์ภายในเป็นการนำวิธีการทางสถิติมา ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบหรือแบบทดสอบ โดยพิจารณาโครงสร้างภายใน ของแบบทดสอบเป็นหลัก ด้วยการวิเคราะห์ผลจากการตอบข้อสอบและความสามารถหรือคะแนนจริง ของผู้สอบที่ได้จากแบบทดสอบฉบับนั้น เพื่อนำมาเปรียบเทียบระหว่างผู้สอบจากกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ ที่มีความสามารถหรือคะแนนจริงเท่ากันว่าจะมีผลการตอบหรือโอกาสในการตอบ ข้อสอบได้ถูกต้องแตกต่างกันหรือไม่ เพื่อบ่งชี้ถึงการทำหน้าที่ต่างกันของข้อสอบ การวิเคราะห์ใน ลักษณะนี้นิยมใช้ค่าสถิติต่างๆเป็นตัวบ่งชี้ถึงการทำหน้าที่ต่างกันของข้อสอบ ค่าสถิติที่นำมาใช้สรุปได้ดังนี้

1) **การทดสอบปฏิสัมพันธ์ (Interaction)** ในระยะเริ่มแรกของการศึกษาความลำเอียงของข้อสอบ มีการใช้สถิติทดสอบ เอฟ (F-test) จากการวิเคราะห์ความแปรปรวน (Analysis of variance :ANOVA) เพื่อทดสอบปฏิสัมพันธ์ระหว่างกลุ่มผู้สอบกับข้อสอบ ถ้าการทดสอบมีนัยสำคัญ แสดงว่าข้อสอบทำหน้าที่ต่างกันจากนั้นจึงทำการวิเคราะห์ต่อด้วยวิธีการ Post Hoc เพื่อระบุข้อสอบที่มีต่อการเกิดปฏิสัมพันธ์ ซึ่งเป็นข้อที่ทำหน้าที่ต่างกัน ข้อดีของการทดสอบปฏิสัมพันธ์คือสามารถศึกษาผู้สอบหลายๆ กลุ่มได้สะดวก แต่ยังมีจุดอ่อนในเรื่องการควบคุมกลุ่มต่างๆ ให้มีความสามารถที่ทัดเทียมกัน ขนาดกลุ่มตัวอย่างของกลุ่มต่างๆ และอัตราความคลาดเคลื่อนประเภทที่ 1 จะสูงขึ้น ถ้าจำนวนข้อสอบเพิ่มมากขึ้น

2) **การวัดความเบี่ยงเบนสัมพัทธ์ (Relative Deviation)** การวัดความเบี่ยงเบนสัมพัทธ์เป็นการคำนวณค่าความยากของข้อสอบ เมื่อคำนวณแยกระหว่างกลุ่ม และแปลงให้เป็นค่าความยากมาตรฐาน (Δ) สามารถนำมาพล็อตเปรียบเทียบเป็นรายข้อ ถ้าข้อใดเบี่ยงเบนไปจากแกนหลักที่คาดหมาย หรือเบี่ยงเบนเกิดจากความคลาดเคลื่อนมาตรฐานของค่าความยากที่กำหนด แสดงถึงการทำหน้าที่ต่างกันของข้อสอบ รวมทั้งสามารถคำนวณค่าสหสัมพันธ์ระหว่างค่าความยากรายข้อระหว่างกลุ่มได้ เพื่อแสดงถึงการทำหน้าที่ต่างกันของแบบสอบ ถ้าสหสัมพันธ์เข้าใกล้ 1.00 แสดงว่าค่าความยากสัมพัทธ์ของข้อสอบมีค่าใกล้เคียงกันระหว่างกลุ่ม ดังนั้นแบบสอบวัดคุณลักษณะคล้ายกันระหว่างกลุ่ม ข้อดีและข้อเสียการวัดความเบี่ยงเบนสัมพัทธ์ คล้ายการทดสอบปฏิสัมพันธ์ นอกจากนี้ค่าความยากของข้อสอบ มิใช่ตัวแทนของค่าความยากจริงของข้อสอบ และอาจได้รับอิทธิพลจากค่าแทรกซ้อนอื่น ได้แก่ ค่าอำนาจจำแนก และความสามารถของผู้สอบ

3) **การเปรียบเทียบน้ำหนักตัวประกอบ (Factor Loading)** การวิเคราะห์ตัวประกอบ (Factor Analysis) เป็นเทคนิคทางสถิติที่นิยมใช้ตรวจสอบความตรงเชิงหรือ โครงสร้าง

(Construct Validity) เมื่อนำการวิเคราะห์ตัวประกอบมาใช้ในการวิเคราะห์โครงสร้างของแบบสอบแยกตามกลุ่มผู้สอบ ความไม่สอดคล้องกันระหว่างน้ำหนักตัวประกอบบนคุณลักษณะสำคัญที่มุ่งวัด หรือ ความแตกต่างของค่าเฉลี่ยคะแนน ตัวประกอบ (Factor scores) ระหว่างกลุ่มผู้สอบจะสะท้อนถึงการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ

การใช้เทคนิคการวิเคราะห์ตัวประกอบเชิงสำรวจ (Exploratory Factor Analysis; EFA) สำหรับศึกษาการทำหน้าที่ต่างกัน นั้นมีจุดอ่อนในเรื่องความไม่สอดคล้องกันระหว่างน้ำหนักตัวประกอบอาจเกิดจากความแตกต่างของความสามารถระหว่างกลุ่มก็ได้ แนวทางที่เหมาะสมจึงควรใช้เทคนิคการวิเคราะห์ตัวประกอบเชิงยืนยัน (Confirmatory Factor Analysis; CFA) นอกจากนี้ยังสามารถใช้ CFA สำหรับตรวจสอบความแตกต่างระหว่างกลุ่ม ในด้านคุณลักษณะหรือความสามารถหลัก และความสามารถรองได้อีกด้วย

4) การเปรียบเทียบโอกาสตอบข้อสอบถูก การวิเคราะห์โอกาสตอบข้อสอบถูกของผู้สอบจากกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่มีความสามารถเท่ากัน เป็นแนวทางสำคัญที่นิยมใช้กัน และเป็นที่ยอมรับในปัจจุบัน สำหรับบ่งชี้การทำหน้าที่ต่างกันของข้อสอบ มีการคำนวณค่าสถิติ 2 แนวทางดังนี้

(1) เปรียบเทียบค่าสัดส่วน หรือความน่าจะเป็นในการตอบข้อสอบถูกของผู้สอบต่างกลุ่มที่มีความสามารถเท่ากันเช่น วิธیمانเทส-แฮนส์เชล เป็นต้น

(2) เปรียบเทียบค่าฟังก์ชันการตอบสนองข้อสอบ หรือโค้งลักษณะข้อสอบระหว่างกลุ่มที่มีระดับความสามารถเท่ากัน เป็นวิธีที่อยู่บนพื้นฐานของทฤษฎี IRT เช่น วิธีวัดความแตกต่างของพื้นที่ วิธีวัดความแตกต่างของค่าพารามิเตอร์ความยาก วิธีการทดสอบไค-สแควของ ลอร์ด (Lord's X^2 -test) เป็นต้น

การเปรียบเทียบโอกาสตอบข้อสอบถูกมีข้อดีที่สำคัญได้แก่ การคำนวณค่าสถิติของข้อสอบมีความน่าเชื่อถือ มีกลไกควบคุมความสามารถของผู้สอบโดยการจับคู่กลุ่มความสามารถ เพื่อใช้ทำการเปรียบเทียบ ณ ตำแหน่งต่างๆ ที่มีความสามารถเท่ากัน จึงเป็นวิธีการที่ยอมรับกันทั่วไป แต่มีข้อจำกัดในด้านความสลับซับซ้อนของแนวคิดพื้นฐาน และการวิเคราะห์มีความจำเป็นต้องใช้โปรแกรมคอมพิวเตอร์ โดยเฉพาะ (ศิริชัย กาญจนวาสี, 2555, น.120-123)

2.4 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีถดถอยโลจิสติก

Swaminathan & Rogers (1990) ได้พัฒนาวิธีถดถอยโลจิสติก เพื่อใช้ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบสองค่า (Dichotomous) เป็นวิธีที่พัฒนามาจากวิธีล็อกลิเนียร์ (Loglinear) ของเมลเลนเบอร์เกอร์ (Mellenber, 1982) ซึ่งวิธีนี้อยู่บนพื้นฐานของโมเดลการวิเคราะห์สมการถดถอยโลจิสติก ซึ่งเป็นโมเดลที่มีพื้นฐานเป็นแบบจำลองที่สามารถเพิ่มตัวแปรความสามารถ

และปฏิสัมพันธ์เข้าไปในสมการได้ และสอดคล้องกับธรรมชาติของการทำหน้าที่ต่างกันของแบบสอบ
ได้ดี จึงสามารถตรวจสอบการทำหน้าที่ต่างกันทั้งแบบเอกรูปและแบบอนเอกรูปได้หลักการตรวจสอบ
ด้วยวิธีถดถอย

ในการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ โดยแบ่งสมการสำหรับผู้สอบ
ออกเป็น 2 กลุ่ม สมการมาตรฐานของโมเดลการถดถอยโลจิสติก สำหรับคำนวณโอกาสการตอบ
ข้อสอบถูก เป็นดังนี้

$$P(U_{ij} = 1/\theta_{ij}) = \frac{e^{(\beta_{0j} + \beta_{1j}\theta_{ij})}}{1 + e^{(\beta_{0j} + \beta_{1j}\theta_{ij})}} \quad , i = 1, 2, \dots, n_j : j = 1, 2$$

เมื่อ U_{ij} แทน โอกาสของการตอบข้อสอบถูกของบุคคล ที่ i ในกลุ่ม j
 θ_{ij} แทน ความสามารถของผู้สอบ ที่ i ในกลุ่ม j
 β_{0j} แทน พารามิเตอร์จุดตัด (intercept parameter)
 β_{1j} แทน ค่าพารามิเตอร์ความชัน (slope parameter)

ถ้า $\beta_{01} = \beta_{02}$ และ $\beta_{11} = \beta_{12}$ เป็นลักษณะโค้งถดถอยโลจิสติก สำหรับผู้สอบสองกลุ่ม
เหมือนกันแสดงว่าข้อสอบทำหน้าที่ไม่ต่างกัน (No DIF)

$\beta_{01} \neq \beta_{02}$ และ $\beta_{11} = \beta_{12}$ เป็นลักษณะโค้งถดถอยโลจิสติก สำหรับผู้สอบสองกลุ่ม
เท่าเทียมกันแสดงว่าข้อสอบทำหน้าที่ต่างกันแบบเอกรูป (Uniform DIF)

$\beta_{01} = \beta_{02}$ และ $\beta_{11} \neq \beta_{12}$ เป็นลักษณะโค้งถดถอยโลจิสติก สำหรับผู้สอบสองกลุ่มไม่
เท่าเทียมกันแสดงว่าข้อสอบทำหน้าที่ต่างกันแบบอนเอกรูป (Nonuniform DIF)

จากสมการถดถอยโลจิสติกข้างต้นสามารถเขียนใหม่ในรูปการวิเคราะห์การทำหน้าที่
ต่างกันของข้อสอบได้เป็น

$$P(U_{ij} = 1/\theta_{ij}) = \frac{e^{Z_{ij}}}{1 + e^{Z_{ij}}}$$

$$\text{โดยที่ } \log \left[\frac{p}{1-p} \right] = Z_{ij} = \tau_0 + \tau_1\theta_{ij} + \tau_2G_j + \tau_3\theta_{ij}G_j$$

เมื่อ $P(U_{ij} = 1/\theta_{ij})$ แทน โอกาสในการตอบข้อสอบถูกของผู้สอบที่ i ในกลุ่ม j

ซึ่งมีความสามารถ θ

θ_{ij} แทน ความสามารถของผู้สอบคนที่ i ในกลุ่มที่ j

G_j แทน กลุ่มผู้สอบ ($G=1$ สมาชิกกลุ่มที่ 1 หรือกลุ่มเปรียบเทียบ $G=2$ สมาชิก กลุ่ม
2 หรือกลุ่มอ้างอิง)

$\theta_{ij}G_j$ แทน ปฏิสัมพันธ์ของตัวแปรอิสระ 2 ตัว คือ θ_{ij} กับ G_j

τ_0 แทน พารามิเตอร์จุดตัด

τ_1 แทน ผลของความสามารถของผู้สอบ
 τ_2 แทน ความแตกต่างระหว่างกลุ่มในการตอบข้อสอบถูกโดย $\tau_2 = \beta_{01} - \beta_{02}$
 τ_3 แทน ปฏิสัมพันธ์ระหว่างกลุ่มกับความสามารถผู้สอบโดย $\tau_3 = \beta_{11} - \beta_{12}$
 ในการตัดสินใจว่าข้อสอบทำหน้าที่ต่างกันแบบเอกรูป หรืออนเอกรูป พิจารณาดังนี้
 ถ้า $\tau_2 \neq 0$ และ $\tau_3 = 0$ หมายความว่า ข้อสอบทำหน้าที่ต่างกันแบบเอกรูป แต่ถ้า $\tau_3 \neq 0$

โดยที่ $\tau_2 = 0$ หรือ $\neq 0$ หมายความว่า ข้อสอบทำหน้าที่ต่างกันแบบอนเอกรูป

นอกจากนี้ คามิลลี และเชฟเพอร์ค (Camilli and Shepard, 1994) ได้เสนอวิธีทดสอบสมมติฐานเกี่ยวกับการทำหน้าที่ต่างกันของข้อสอบ โดยเปรียบเทียบระหว่างโมเดลการถดถอย โลจิสติกแบบ 2 ขั้นตอน ดังนี้

ขั้นตอนที่ 1 วิเคราะห์นัยสำคัญของปฏิสัมพันธ์

$$\text{Model 1.} \quad \psi_{ij} = \tau_0 + \tau_1 X_{ij} + \tau_2 G_j + \tau_3 (G_j X_{ij})$$

เมื่อ ψ_{ij} แทน ค่าโลจิท หรือค่าลออถอัตราส่วนโอกาสตอบถูกต้องโอกาส

ตอบผิดของสอบ i ในกลุ่มที่ j

X_{ij} แทน คะแนนรวมของผู้สอบ i ในกลุ่มที่ j

G_j แทน กลุ่มผู้สอบ

τ_0 แทน พารามิเตอร์จุดตัด

τ_1 แทน ผลของความสามารถของผู้สอบ

τ_2 แทน ความแตกต่างระหว่างกลุ่มในการตอบข้อสอบถูก

$$\text{โดย } \tau_2 = \beta_{01} - \beta_{02}$$

τ_3 แทน ปฏิสัมพันธ์ระหว่างกลุ่มกับความสามารถผู้สอบ

$$\text{โดย } \tau_3 = \beta_{11} - \beta_{12}$$

β_{0j} แทน พารามิเตอร์จุดตัด (intercept parameter)

β_{1j} แทน ค่าพารามิเตอร์ความชัน (slope parameter)

$$\text{Model 2.} \quad \psi_{ij} = \tau_0 + \tau_1 X_{ij} + \tau_2 G_j$$

การทดสอบความแตกต่างระหว่างโมเดล 1 และ โมเดล 2 ด้วยการทดสอบไค-สแควร์ที่ระดับชั้นความเป็นอิสระเท่ากับ 1 ($df = 1$) ถ้าพบความแตกต่างอย่างมีนัยสำคัญ หรือ τ_3 มีนัยสำคัญทางสถิติแสดงว่าข้อสอบข้อนั้นทำหน้าที่ต่างกันแบบอนเอกรูป (Nonuniform DIF) แต่ถ้า τ_3 ไม่มีนัยสำคัญ ให้ดำเนินการขั้นที่ 2

ขั้นตอนที่ 2 วิเคราะห์นัยสำคัญของกลุ่มผู้สอบ

$$\text{Model 2.} \quad \psi_{ij} = \tau_0 + \tau_1 X_{ij} + \tau_2 G_j$$

$$\text{Model 3.} \quad \psi_{ij} = \tau_0 + \tau_1 X_{ij}$$

ทดสอบความแตกต่างระหว่างโมเดล 2 และ โมเดล 3 ด้วยการทดสอบไค-สแควร์ ที่ระดับชั้นความเป็นอิสระเท่ากับ 1 ($df=1$) ถ้าพบความแตกต่างอย่างมีนัยสำคัญหรือ τ_2 มีนัยสำคัญทางสถิติแสดงว่า ข้อสอบข้อนั้นทำหน้าที่แตกต่างกันแบบเอกรูป (uniform DIF)

2.5 ขนาดอิทธิพลกับการทำหน้าที่ต่างกันของข้อสอบ

การนัยทางสถิติ (Statistical Significance) มาใช้ในการตัดสินยอมรับหรือปฏิเสธสมมติฐานทางการวิจัยซึ่งมีข้อพิจารณาอยู่ 2 ประเด็น คือ 1) ความแตกต่างอย่างมีนัยสำคัญทางสถิติจะเปลี่ยนแปลงไปตามขนาดของกลุ่มตัวอย่าง 2) ความแตกต่างอย่างมีนัยสำคัญทางสถิติไม่ได้บ่งบอกถึงปริมาณความแตกต่างของค่าเฉลี่ยของประชากร 2 กลุ่ม จึงได้มีการนัยสำคัญทางปฏิบัติ (Practical Significance) ที่รู้จักกันทั่วไป คือ ขนาดอิทธิพล (Effect size) ซึ่งเป็นสัดส่วนระหว่างความแตกต่างของค่าเฉลี่ย 2 กลุ่ม ต่อส่วนเบี่ยงเบนมาตรฐาน นำมารวมใช้ในการพิจารณาเพื่อใช้ในการตัดสินปริมาณความแตกต่างของค่าเฉลี่ยที่พบมากพอต่อการนำไปใช้ในการปฏิบัติหรือไม่ นัยสำคัญทางปฏิบัติมีความแน่นอนไม่ว่ากลุ่มตัวอย่างจะมีขนาดเล็กหรือขนาดใหญ่ (ชเกียรติกรมล ทองงอก, 2554)

สวามินาธาน และ โรเจอร์ส (Swaminathan and Rogers, 1990) เสนอวิธีการทดสอบนัยสำคัญใช้สถิติไค-สแควร์ที่ระดับ $df=2$ ซึ่งยอมให้มีการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูปและอเนกรูป โดยสามารถตรวจสอบพร้อมกันได้ ดังนั้นขนาดของผล (Effect size) ที่ได้จะเป็น ΔR^2 (มาจากผลต่างของ R^2 โมเดล 1 กับ โมเดล 3) ซึ่งเป็นผลมาจากทั้งอิทธิพลของการเป็นสมาชิกของกลุ่มและ/หรือ กลุ่มอิทธิพลของปฏิสัมพันธ์ระหว่างการเป็นสมาชิกของกลุ่ม กับระดับความสามารถของผู้สอบพร้อมกัน

ซัม โบและ โทมัส (Hidalho and Pina, 2004, p.907) ได้เสนอดัชนี ΔR^2 เพื่อใช้ในการจำแนกขนาดการทำหน้าที่ต่างกันของข้อสอบ โดยมพื้นฐานจากค่าประมาณกำลังสองน้อยที่สุดแบบถ่วงน้ำหนัก (Weighted least squares) เกณฑ์สำหรับดัชนี ΔR^2 มีดังนี้

$$\Delta R^2 < 0.13 \quad \text{แสดงว่า DIF ระดับ A มีขนาดเล็ก}$$

$$0.13 \leq \Delta R^2 \leq 0.26 \quad \text{แสดงว่า DIF ระดับ B มีขนาดปานกลาง}$$

$$\Delta R^2 > 0.26 \quad \text{แสดงว่า DIF ระดับ B มีขนาดปานใหญ่}$$

2.6 สาเหตุการทำหน้าที่ต่างกันของข้อสอบ

ชูนเนแมน (Scheuneman, 1982; จันทนา เปรมฤดีปริษาชาญ, 2551, น.12) ได้ศึกษาข้อสอบหลังจากการวิเคราะห์ด้วยค่าสถิติเพื่อตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแล้ว และจากการศึกษาผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบมาตรฐานหลายฉบับ

พบว่าสามารถวิเคราะห์ข้อสอบที่ทำหน้าที่ต่างกันเหล่านั้นได้ว่า น่าจะเกิดจากแหล่งสาเหตุใดมากที่สุด ซึ่งสามารถสรุปสาเหตุที่ก่อให้เกิดการทำหน้าที่ต่างกันของข้อสอบมากที่สุดได้ดังนี้

1. การเดา (guessing) ซึ่งอาจจะเกิดจากข้อสอบยากเกินไปหรือเวลาไม่เพียงพอ ก่อให้เกิดความไม่เท่าเทียมกันในโอกาสการตอบข้อสอบข้อนั้นถูก
2. ความเร็ว (Speed) หรือเวลาในการตอบ ซึ่งจะเป็นผลให้เกิดการเดาหรือทำไม่ทันในกรณีเวลาน้อยจะมีผลกับข้อสอบข้อหลัง ๆ โดยเฉพาะในการศึกษาความลำเอียงของข้อสอบวัดความถนัด
3. ความกำกวมหรือความไม่ชัดเจน (Unclear) ของข้อคำถามนั้นคือข้อคำถามขาดความเป็นปรนัย หรือใช้ภาษาหรือคำที่ไม่สื่อเป็นภาษากลาง ซึ่งจะก่อให้เกิดการทำหน้าที่ต่างกับกลุ่มภาษาใดภาษาหนึ่งได้
4. ลำดับชั้นของคำถาม (Series) จะเป็นสิ่งที่จะก่อให้เกิดความสับสน หรือชี้แนะคำตอบบางข้อได้
5. สภาพทางสังคมหรือความเกี่ยวข้องทางสังคม (Social implication) ก็จะเป็นสิ่งที่จะก่อให้เกิดความแตกต่างระหว่างกลุ่มได้
6. ประสบการณ์หรือการฝึกของแต่ละกลุ่มย่อย เป็นสิ่งที่ก่อให้เกิดการได้เปรียบเสียเปรียบค่อนข้างจะชัดเจน
7. องค์ประกอบทางวัฒนธรรม ความเป็นอยู่ ประเพณี เชื้อชาติก็จะเอื้อให้กับบางกลุ่มย่อย

นอกจากนั้นการทำหน้าที่ต่างกันของข้อสอบอาจจะเกิดจากสาเหตุหรือแหล่งที่สำคัญ 2 แหล่ง คือ

1. การเลือกเนื้อหา คือ ผู้สร้างข้อสอบเลือกเนื้อหาในการสร้างข้อสอบเฉพาะส่วนใดส่วนหนึ่งมาสร้าง ทำให้ได้ข้อสอบมีเนื้อหาไม่ครอบคลุม และไม่เป็นสัดส่วนที่สมดุล
2. การสร้างข้อสอบ คือ การใช้ภาษา หรือข้อความบางอย่างที่เอื้อประโยชน์กับผู้สอบกลุ่มใดกลุ่มหนึ่ง

Hambleton และ Rogers (1995) เชื่อว่าสาเหตุของการทำหน้าที่ต่างกันของข้อสอบมาจากแบบทดสอบ จำแนกสาเหตุของการทำหน้าที่ต่างกันและวิธีการสังเกตดังนี้

1. การทำหน้าที่ต่างกันที่เกิดจากเนื้อหาของแบบทดสอบ (content bias) สังเกตได้จาก
 - 1.1 ข้อสอบมีเนื้อหาที่แตกต่างกันหรือมีความหมายที่ไม่คุ้นเคยกับกลุ่มผู้สอบบางกลุ่ม
 - 1.2 ผู้สอบสามารถตอบข้อสอบถูกหรือผิด โดยให้เหตุผลการตอบข้อสอบต่างกัน
 - 1.3 เนื้อหาของข้อสอบแสดงให้เห็นทักษะอื่นที่ไม่ต้องการที่จะศึกษา
2. ความลำเอียงทางด้านภาษา (Language bias) สังเกตได้จาก

- 2.1 ข้อสอบมีค่าที่มีความหมายแตกต่างกันหรือมีความหมายไม่เป็นที่คุ้นเคยกับกลุ่มผู้สอบบางกลุ่ม
- 2.2 ข้อสอบมีคำศัพท์ยาก
- 2.3 ข้อสอบมีภาษา คำศัพท์ หรือคำอ้างอิง เฉพาะกลุ่มใดกลุ่มหนึ่ง
3. ความลำเอียงจากโครงสร้างและรูปแบบของข้อสอบ (item structure and format bias)
 - 3.1 ข้อสอบมีค่าที่อำนวยความสะดวกให้ผู้สอบบางกลุ่ม
 - 3.2 มีคำชี้แจง ข้อคำถาม ตัวลวง ของข้อสอบ ที่ไม่เพียงพอหรือกำกวม
 - 3.3 มีคำอธิบายในข้อสอบสร้างความสับสนให้กับกลุ่มผู้สอบ

3. โครงการประเมินคุณภาพการศึกษาขั้นพื้นฐานเพื่อการประกันคุณภาพผู้เรียน ปีการศึกษา 2557

3.1 เหตุผลและความสำคัญ

โครงการประเมินคุณภาพการศึกษาขั้นพื้นฐาน เพื่อการประกันคุณภาพผู้เรียน ปีการศึกษา 2557 เป็นการประเมินภาพรวมของการจัดการศึกษาของประเทศไทยและเป็นภาพสะท้อนคุณภาพของเด็กไทย เพื่อนำผลการประเมินไปพัฒนาและปรับปรุงการเรียนการสอนของครู ซึ่งการประเมินคุณภาพการศึกษาขั้นพื้นฐานถือว่าเป็นกระบวนการตรวจสอบการขับเคลื่อนนโยบายการศึกษา และการเพิ่มประสิทธิภาพสถานศึกษา สร้างความมั่นใจว่าสถานศึกษา และหน่วยงานต้นสังกัดจัดการศึกษาอย่างมีคุณภาพ ได้มาตรฐานตามที่กำหนด นักเรียนมีความรู้ความสามารถ และคุณลักษณะอันพึงประสงค์ตามที่หลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 กำหนด และยังเป็นองค์ประกอบ/ตัวชี้วัด ที่สำคัญตามมาตรฐานการประกันคุณภาพการศึกษา ซึ่งความสำคัญของการประกันคุณภาพจะเห็นได้จากพระราชบัญญัติการศึกษาแห่งชาติ พุทธศักราช 2542 มาตรา 47 กำหนดให้มีระบบการประกันคุณภาพการศึกษา เพื่อพัฒนาคุณภาพและมาตรฐานการศึกษาในทุกระดับ และมาตรา 48 ให้หน่วยงานต้นสังกัด และสถานศึกษา จัดให้มีระบบการประกันคุณภาพการศึกษาภายในสถานศึกษา และให้ถือว่าการประกันคุณภาพภายในเป็นส่วนหนึ่งของกระบวนการบริหารการศึกษาที่ต้องดำเนินการอย่างต่อเนื่อง โดยมีการจัดทำรายงานประจำปีเสนอต่อหน่วยงาน ต้นสังกัด หน่วยงานที่เกี่ยวข้อง และเปิดเผยต่อสาธารณชน เพื่อนำไปสู่การพัฒนาคุณภาพและมาตรฐานการศึกษา และเพื่อรองรับการประกันคุณภาพภายนอก

หลักการสำคัญประการหนึ่งของการประกันคุณภาพการศึกษาก็คือ การแสดงถึงภาระรับผิดชอบ (Accountability) ซึ่งตรวจสอบได้ ที่สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน เขตพื้นที่การศึกษา สถานศึกษา และหน่วยงานที่เกี่ยวข้อง ได้ร่วมกันกำหนดคำรับรองในการปฏิบัติราชการของ

ส่วนราชการประจำปีงบประมาณ โดยกำหนดเป้าหมาย และจุดเน้น ที่ต้องการพัฒนาผลสัมฤทธิ์ทางการเรียน รวมทั้งความสามารถในการคิดของผู้เรียนให้สูงขึ้น โดยหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 ได้กำหนดเกณฑ์สำหรับการจบการศึกษาว่า นอกจากผู้เรียนได้เรียนรายวิชาพื้นฐาน รายวิชากิจกรรมเพิ่มเติมตาม โครงสร้างเวลาเรียน และผ่านเกณฑ์การประเมินตามที่สถานศึกษากำหนดแล้ว ผู้เรียนต้องมีผลการประเมิน การอ่าน คิววิเคราะห์ เขียน และคุณลักษณะอันพึงประสงค์ ในระดับผ่านเกณฑ์การประเมินที่สถานศึกษากำหนด ประกอบกับสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน ได้จัดทำประกาศเรื่อง แนวปฏิบัติในการยกระดับคุณภาพนักเรียนระดับการศึกษาขั้นพื้นฐาน ประกาศ ณ วันที่ 18 เมษายน 2557 ลงนามโดยเลขาธิการคณะกรรมการการศึกษาขั้นพื้นฐาน (นายอภิชาติ จีระวุฒิ) โดยเฉพาะข้อ 7 ให้เพิ่มข้อสอบแบบเขียนตอบทั้งการเขียนตอบแบบสั้นและแบบยาว ในการวัดและประเมินผลเมื่อจบหน่วยการเรียนรู้ในการสอบระหว่างภาคเรียนและการสอบปลายภาคเรียน อย่างน้อยร้อยละ 30 ของการสอบแต่ละครั้ง จากนโยบาย ยุทธศาสตร์ กลยุทธ์ และประกาศของ สพฐ. ดังกล่าว สำนักทดสอบทางการศึกษา ได้ดำเนินการจัดทำแบบทดสอบวัดความสามารถพื้นฐานสำคัญ 3 ด้าน คือ ด้านภาษา (Literacy) ด้านคำนวณ (Numeracy) และด้านเหตุผล (Reasoning Abilities) ชั้นประถมศึกษาปีที่ 3 ซึ่งเป็นความสามารถที่ตกผลึก (Crystallization) จากการเรียนรู้ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 และเป็นความสามารถพื้นฐานที่จำเป็นสำหรับใช้เป็นเครื่องมือในการเรียนรู้ของนักเรียนในระดับสูงขึ้นต่อไป

การประเมินคุณภาพการศึกษาขั้นพื้นฐานในปีการศึกษา 2557 นี้ สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐานดำเนินการประเมินคุณภาพการศึกษาขั้นพื้นฐานกับนักเรียน ชั้นประถมศึกษาปีที่ 3 โดยวัดความสามารถพื้นฐานสำคัญ 3 ด้าน คือ ด้านภาษา (Literacy) ด้านคำนวณ (Numeracy) และด้านเหตุผล (Reasoning Abilities) ซึ่งถือเป็นความสามารถพื้นฐานเบื้องต้นสำคัญที่ใช้ในการเรียนรู้ในระดับที่สูงขึ้นและยังสะท้อนไปสู่การยกระดับ ผลการประเมินระดับชาติ (O-NET) และนานาชาติ (PISA) ผลการประเมินที่ได้จะเป็นข้อมูลสำคัญที่สะท้อนคุณภาพการดำเนินงานการจัดการศึกษาของสถานศึกษา เขตพื้นที่การศึกษา สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐานจำเป็นต้องมีข้อมูลผลการเรียนรู้ ไปเตรียมความพร้อมของผู้เรียน และเป็นตัวบ่งชี้คุณภาพการศึกษาขั้นพื้นฐานในภาพรวม เพื่อใช้เป็นข้อมูลประกอบการตัดสินใจในการกำหนดนโยบาย กำหนดยุทธศาสตร์ แผนการศึกษาของชาติ ระดับสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน ระดับเขตพื้นที่การศึกษา ระดับสถานศึกษา มีคุณภาพมาตรฐานระดับสากล บนพื้นฐานของความเป็นไทย ให้นักเรียนได้รับการพัฒนาศักยภาพสูงสุดในตัว มีความรู้และทักษะที่แข็งแกร่งและเหมาะสม เป็นพื้นฐานสำคัญในการเรียนรู้ระดับสูงขึ้นไป และการดำรงชีวิตในอนาคต โดยเฉพาะข้อ 2 เร่งพัฒนาความแข็งแกร่งทางการศึกษาให้ผู้เรียนทุกระดับ ทุกประเภท รวมถึงเด็กพิการ และด้อยโอกาส ให้มีความรู้และทักษะแห่งโลกยุคใหม่

ควบคู่กันไป โดยเฉพาะทักษะการอ่าน เขียน และการคิด เพื่อให้มีความพร้อมเข้าสู่การศึกษาระดับสูง และโลกของการทำงาน โดยได้กำหนดเป็นยุทธศาสตร์ พัฒนาคุณภาพผู้เรียนทุกระดับ ทุกประเภท เพื่อให้นักเรียนระดับการศึกษาขั้นพื้นฐานทุกคนมีพัฒนาการเหมาะสมตามช่วงวัยและมีคุณภาพเพื่อพัฒนานักเรียนให้มีคุณภาพและมาตรฐานใกล้เคียงกัน อีกทั้งเป็นการส่งเสริมการประกันคุณภาพภายในของสถานศึกษาให้มีความเข้มแข็งเพื่อรองรับการประเมินภายนอก ซึ่งได้กำหนดเป็นกลยุทธ์ในการดำเนินงาน เช่น ส่งเสริมสนับสนุนการนำผลการทดสอบ O-NET ประเมินผล PISA และเพื่อเป็นการประกันคุณภาพภายในสถานศึกษา และเพื่อเตรียมการให้ผู้เรียน มีความพร้อมสำหรับรองรับการประเมินภายนอกของสถานศึกษา ทั้งการทดสอบระดับชาติหรือระดับนานาชาติ โดยจะมุ่งประเมินให้ทัดเทียมกับนานาชาติ เช่น การประเมินระดับนานาชาติ (PISA, TIMSS) ที่มีรูปแบบการประเมินที่หลากหลาย มุ่งเน้นคุณภาพของนักเรียนโดยพิจารณาจากความสามารถพื้นฐานที่สำคัญจำเป็น ผลการประเมินของโรงเรียนและเขตพื้นที่การศึกษา จะเป็นตัวบ่งชี้คุณภาพการศึกษาขั้นพื้นฐานในภาพรวม และใช้เป็นข้อมูลประกอบการตัดสินใจในการกำหนดนโยบาย วางแผน ในการพัฒนาคุณภาพการศึกษา ระดับสถานศึกษา ระดับเขตพื้นที่การศึกษา และสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน ส่วนผลการประเมินนักเรียนทุกคน จะเป็นข้อมูลสำคัญในการปรับปรุงเพื่อการพัฒนา

3.2 วัตถุประสงค์

เพื่อประเมินความสามารถพื้นฐานที่สำคัญจำเป็นของนักเรียนชั้นประถมศึกษาปีที่ 3 ในด้านภาษา (Literacy) ด้านคำนวณ (Numeracy) และด้านเหตุผล (Reasoning Abilities)

3.3 กลุ่มเป้าหมาย

นักเรียนที่ศึกษาในระดับชั้นประถมศึกษาปีที่ 3 ปีการศึกษา 2557 ทุกคน ทุกโรงเรียน ซึ่งอยู่ในสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน และตำรวจตระเวนชายแดน

3.4 เครื่องมือที่ใช้ในการประเมิน

เครื่องมือที่ใช้ในการประเมิน ชั้นประถมศึกษาปีที่ 3 แสดงรายละเอียด ดังตาราง 2.1

ตารางที่ 2.1 รายละเอียดเครื่องมือที่ใช้ในการประเมินความสามารถพื้นฐานที่สำคัญจำเป็นของนักเรียนชั้นประถมศึกษาปีที่ 3

แบบทดสอบความสามารถ	จำนวนข้อ	คะแนนเต็ม	เวลา (นาที)
ด้านภาษา (Literacy)	30	35	60
ด้านคำนวณ (Numeracy)	30	35	90
ด้านเหตุผล (Reasoning Abilities)	30	35	60

โครงสร้างข้อสอบ

1. ความสามารถด้านภาษาชั้นประถมศึกษาปีที่ 3 มีกรอบโครงสร้างในการประเมิน ดังตาราง 2.2

ตารางที่ 2.2 รายละเอียดกรอบโครงสร้างข้อสอบวัดความสามารถด้านภาษา

ตัวชี้วัด	เลือกตอบ	เติมคำตอบ
1. บอกความหมายของคำและประโยคจากเรื่องที่ ฟัง ดู และอ่าน	4	
2. บอกความหมายของเครื่องหมายสัญลักษณ์	3	
3. ตอบคำถามจากเรื่องที่ ฟัง ดู และอ่าน	4	2
4. บอก เล่าเรื่องราวที่ได้จากการฟัง ดู และอ่านอย่างง่ายๆ	5	
5. คาดคะเนเหตุการณ์ที่จะเกิดขึ้นจากเรื่องที่ ฟัง ดู และอ่าน	6	
6. สื่อสารความรู้ ความเข้าใจข้อคิดเห็นจากเรื่องที่ ฟังดู และอ่านอย่างเหมาะสม	5	1
รวมทั้งหมด	27	3

2. ความสามารถด้านคำนวณชั้นประถมศึกษาปีที่ 3 มีกรอบ โครงสร้างในการประเมิน ดังตารางที่ 2.3

ตารางที่ 2.3 รายละเอียดกรอบ โครงสร้างข้อสอบวัดความสามารถด้านคำนวณ

ตัวชี้วัด	เลือกตอบ	เติมคำตอบ	แสดงวิธีทำ
1. ใช้ทักษะกระบวนการทางคณิตศาสตร์ หรือทักษะการคิดคำนวณเพื่อตัดสินใจเลือกแนวทางปฏิบัติหรือหาคำตอบจากสถานการณ์ต่างๆ ในชีวิตประจำวันเกี่ยวกับความคิดรวบยอดทางคณิตศาสตร์ เรื่อง จำนวนและการดำเนินการตามขอบข่ายสิ่งเร้า	9	1	
2. ใช้ทักษะกระบวนการทางคณิตศาสตร์ หรือทักษะการคิดคำนวณเพื่อตัดสินใจเลือกแนวทางปฏิบัติหรือหาคำตอบจากสถานการณ์ต่างๆ ในชีวิต ประจำวันเกี่ยวกับความคิดรวบยอดทางคณิตศาสตร์ เรื่อง การวัด ตามขอบข่ายสิ่งเร้า	7		1

ตารางที่ 2.3 (ต่อ)

ตัวชี้วัด	เลือกตอบ	เติมคำตอบ	แสดงวิธีทำ
3. ใช้ทักษะกระบวนการทางคณิตศาสตร์ หรือทักษะการคิดคำนวณเพื่อตัดสินใจเลือกแนวทางปฏิบัติหรือหาคำตอบจากสถานการณ์ต่างๆ ในชีวิตประจำวันเกี่ยวกับความคิดรวบยอดทางคณิตศาสตร์ เรื่อง เรขาคณิต ตามขอบข่ายสิ่งเร้า	4		
4. ใช้ทักษะกระบวนการทางคณิตศาสตร์ หรือทักษะการคิดคำนวณเพื่อตัดสินใจเลือกแนวทางปฏิบัติหรือหาคำตอบจากสถานการณ์ต่างๆ ในชีวิตประจำวันเกี่ยวกับความคิดรวบยอดทางคณิตศาสตร์ เรื่อง พีชคณิต ตามขอบข่ายสิ่งเร้า	4		
5. ใช้ทักษะกระบวนการทางคณิตศาสตร์ หรือทักษะการคิดคำนวณเพื่อตัดสินใจเลือกแนวทางปฏิบัติหรือหาคำตอบจากสถานการณ์ต่างๆ ในชีวิตประจำวันเกี่ยวกับความคิดรวบยอดทางคณิตศาสตร์ เรื่อง การวิเคราะห์ข้อมูลและความน่าจะเป็นตามขอบข่ายสิ่งเร้า	3	1	
รวมทั้งหมด	27	2	1

3. ความสามารถด้านเหตุผลชั้นประถมศึกษาปีที่ 3 มีกรอบโครงสร้างในการประเมิน ดัง
ตารางที่ 2.4

ตารางที่ 2.4 รายละเอียดกรอบโครงสร้างข้อสอบวัดความสามารถด้านเหตุผล

ตัวชี้วัด	เลือกตอบ	เขียนตอบ
1. มีความเข้าใจในข้อมูล สถานการณ์ หรือสารสนเทศทางด้านวิทยาศาสตร์ สิ่งแวดล้อม ด้านสังคมศาสตร์ เศรษฐศาสตร์ และด้านการดำเนินชีวิตอย่างมีเหตุผล	3	1

ตารางที่ 2.4 (ต่อ)

ตัวชี้วัด	เลือกตอบ	เขียนตอบ
2. วิเคราะห์ข้อมูล สถานการณ์ หรือสารสนเทศ โดยใช้องค์ความรู้ด้านวิทยาศาสตร์ สิ่งแวดล้อม ด้านสังคมศาสตร์ เศรษฐศาสตร์ และด้านการดำเนินชีวิตอย่างมีเหตุผล	7	1
3. สามารถสร้างข้อสรุปใหม่ ออกแบบ วางแผน บนพื้นฐานของข้อมูล สถานการณ์ หรือสารสนเทศที่ผ่านการวิเคราะห์ โดยใช้องค์ความรู้ด้านวิทยาศาสตร์ สิ่งแวดล้อม ด้านสังคมศาสตร์ เศรษฐศาสตร์ และด้านการดำเนินชีวิตอย่างมีเหตุผล	8	
4. สามารถตัดสินใจและแก้ปัญหาอย่างมีหลักการและเหตุผล หรือให้ข้อสนับสนุนข้อโต้แย้งที่สมเหตุสมผล โดยคำนึงถึงคุณธรรมและจริยธรรม ค่านิยม ความเชื่อ ในกรณีที่มีสถานการณ์ที่ต้องการตัดสินใจ หรือมีปัญหา	9	1
รวมทั้งหมด	27	3

3.5 บทบาทและหน้าที่ในการดำเนินการประเมิน

การประเมินคุณภาพการศึกษาขั้นพื้นฐาน เพื่อการประกันคุณภาพผู้เรียน ปีการศึกษา 2557 เป็นความร่วมมือทั้งระบบตั้งแต่สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน โดยสำนักทดสอบทางการศึกษา สำนักงานเขตพื้นที่การศึกษา และ โรงเรียน/สนามสอบ โดยแต่ละส่วนมีบทบาท หน้าที่ในการดำเนินการประเมิน ดังนี้

3.5.1 บทบาท หน้าที่ของ สพฐ. โดย สทศ.

เพื่อให้การดำเนินงานประเมินคุณภาพการศึกษาขั้นพื้นฐาน เพื่อการประกันคุณภาพผู้เรียน ปีการศึกษา 2557 ดำเนินการไปอย่างมีมาตรฐาน รูปแบบ และแนวทางการประเมิน สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน ได้แต่งตั้งคณะกรรมการ 2 คณะ ดังนี้

1) คณะกรรมการประเมินคุณภาพการศึกษาขั้นพื้นฐานเพื่อการประกันคุณภาพผู้เรียน ปีการศึกษา 2557 ตามคำสั่ง สพฐ. ที่ 980/2557 เลขานุการคณะกรรมการการศึกษาขั้นพื้นฐาน เป็นที่ปรึกษา รองเลขานุการคณะกรรมการการศึกษาขั้นพื้นฐาน (ที่ดูแลสำนักทดสอบทางการศึกษา) เป็นประธานกรรมการ ที่ปรึกษา สพฐ.ทุกท่าน ผู้เชี่ยวชาญ สพฐ.ทุกท่าน ผู้อำนวยการทุกสำนัก หรือหน่วยงานเทียบเท่ากองใน สพฐ. เป็นกรรมการ ผู้แทนสำนักงานคณะกรรมการส่งเสริม

การศึกษาเอกชน ผู้แทนสำนักงานคณะกรรมการอุดมศึกษา ผู้อำนวยการสำนักงานการศึกษากรุงเทพมหานคร รองผู้บัญชาการกองบัญชาการตำรวจตระเวนชายแดน (กิจการพลเรือน) สำนักงานตำรวจแห่งชาติ ผู้อำนวยการส่วนประสานและพัฒนากิจการการศึกษาท้องถิ่น กรมส่งเสริมการปกครองท้องถิ่น เป็นกรรมการ ผู้อำนวยการสำนักทดสอบทางการศึกษา เป็นกรรมการและเลขานุการ และนักวิชาการสำนักทดสอบ ทางการศึกษา เป็นกรรมการและผู้ช่วยเลขานุการ มีอำนาจหน้าที่

(1) พิจารณากำหนดรูปแบบ หลักการ และแนวทางการประเมินคุณภาพ การศึกษาด้วยข้อสอบมาตรฐานกลาง เพื่อตรวจสอบคุณภาพการจัดการศึกษาขั้นพื้นฐาน และรายงาน ความก้าวหน้า ด้านผลสัมฤทธิ์ของนักเรียน รวมทั้งการส่งเสริม สนับสนุน การนำผลไปใช้ และ พัฒนาการประกันคุณภาพภายในระดับการศึกษาขั้นพื้นฐานอย่างต่อเนื่อง

(2) เสนอแนะแนวทางการปฏิบัติงาน การพัฒนาคุณภาพและมาตรฐานการ ประเมิน ของสำนักงานเขตพื้นที่การศึกษา และสถานศึกษา

(3) ให้คำแนะนำ กำกับ ดูแล เร่งรัด ติดตาม ประเมินผล การปฏิบัติงาน การประเมินคุณภาพการศึกษาระดับการศึกษาขั้นพื้นฐานทั้งระบบ

(4) แต่งตั้งคณะกรรมการ (เพิ่มเติม) คณะอนุกรรมการ หรือคณะทำงาน เฉพาะเรื่อง ตามความจำเป็นและเห็นสมควร เพื่อปฏิบัติงานตามหน้าที่ที่คณะกรรมการมอบหมาย

2) คณะกรรมการดำเนินงานประเมินคุณภาพการศึกษาระดับเขตพื้นที่การศึกษา ปีการศึกษา 2557 ตามคำสั่ง สพฐ. 1115/2557 โดยผู้อำนวยการสำนักงานเขตพื้นที่การศึกษาทุกเขต เป็น ประธานกรรมการ รองผู้อำนวยการสำนักงานเขตพื้นที่การศึกษา ที่รับผิดชอบการประเมิน เป็นรอง ประธานกรรมการ รองผู้อำนวยการสำนักงานเขตพื้นที่การศึกษาทุกท่าน ผู้อำนวยการกลุ่มในสำนักงาน เขตพื้นที่การศึกษาทุกกลุ่ม ผู้ทรงคุณวุฒิจากคณะกรรมการติดตาม ตรวจสอบ ประเมินผลและนิเทศการ การศึกษาของเขตพื้นที่การศึกษา ผู้บริหารสถานศึกษาระดับประถมศึกษาและมัธยมศึกษา ผู้แทน สถานศึกษาระดับประถมศึกษาและมัธยมศึกษา ผู้แทนสถานศึกษาสังกัดอื่นที่เกี่ยวข้อง ศึกษานิเทศก์ทุก ท่าน เป็นกรรมการผู้อำนวยการกลุ่มนิเทศ ติดตาม และประเมินผลการศึกษา เป็นกรรมการ และ เลขานุการ ศึกษานิเทศก์กลุ่มงานวัดและประเมินผลฯ และศึกษานิเทศก์กลุ่มงานส่งเสริมพัฒนาระบบ การประกันคุณภาพการศึกษา เป็นกรรมการและผู้ช่วยเลขานุการ มีหน้าที่

(1) กำหนดแผนดำเนินงานประเมินคุณภาพการศึกษาขั้นพื้นฐาน เพื่อการ ประกันคุณภาพผู้เรียน ระดับเขตพื้นที่การศึกษาด้วยข้อสอบมาตรฐานกลาง ในเขตพื้นที่การศึกษา ตาม นโยบาย และวิธีการประเมินคุณภาพการศึกษา ของคณะกรรมการการประเมินคุณภาพการศึกษาขั้น พื้นฐานฯ (ส่วนกลาง) กำหนด

- (2) ให้คำแนะนำ กำกับ ดูแล เร่งรัด ติดตาม ประเมินผลการปฏิบัติงานการประเมินคุณภาพการศึกษาในเขตพื้นที่การศึกษาที่รับผิดชอบ
- (3) แต่งตั้งคณะกรรมการ (เพิ่มเติม) คณะอนุกรรมการ หรือคณะทำงานเฉพาะเรื่อง ตามความจำเป็นและเห็นสมควร เพื่อปฏิบัติงานตามหน้าที่ที่คณะกรรมการมอบหมาย
- (4) ประชุมคณะกรรมการการประเมินคุณภาพการศึกษาขั้นพื้นฐาน เพื่อการประเมินคุณภาพผู้เรียน ปีการศึกษา 2557 เพื่อพิจารณากำหนดรูปแบบ หลักการ และแนวทางการประเมิน
- (5) ประชุมปฏิบัติการคณะกรรมการดำเนินงานระดับเขตพื้นที่การศึกษา เพื่อชี้แจงรูปแบบ และแนวทางการประเมิน พร้อมทั้งจัดทำแผนปฏิบัติงานการประเมินคุณภาพการศึกษาฯ ปีการศึกษา 2557 ในระดับเขตพื้นที่การศึกษาให้สอดคล้องกับแผนการประสานจากส่วนกลาง
- (6) จัดสรรงบประมาณการดำเนินงานประเมินให้เขตพื้นที่การศึกษา โดยจัดสรรงบประมาณตามสัดส่วน จำนวนโรงเรียน และจำนวนนักเรียน ของแต่ละเขตพื้นที่การศึกษา
- (7) จัดทำต้นฉบับแบบทดสอบและเอกสารประกอบการประเมินคุณภาพการศึกษา ปีการศึกษา 2557
- (8) จัดพิมพ์แบบทดสอบและกระดาษคำตอบพร้อมบรรจุตามจำนวนนักเรียนของแต่ละ โรงเรียน
- (9) จัดส่งแบบทดสอบและเอกสารประกอบให้เขตพื้นที่การศึกษา สำหรับข้อสอบและกระดาษคำตอบ สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน โดยสำนักทดสอบทางการศึกษา (ส่วนกลาง) จัดพิมพ์ตามจำนวนนักเรียนและดำเนินการจัดส่งไปทุกเขตพื้นที่การศึกษา สำหรับคู่มือการจัดสอบฯ จะจัดส่งต้นฉบับทางระบบ EPCC ทั้งนี้ให้เขตนำไปจัดทำสำเนาเอง
- (10) คณะติดตามและประเมินผลการดำเนินงานการประเมินจากส่วนกลางติดต่อ ประสานการดำเนินการสอบกับเขตพื้นที่การศึกษา และปฏิบัติงานก่อนวันสอบ วันสอบ หลังสอบ เพื่อให้คำแนะนำ แก้ไขปัญหาอุปสรรคที่เกิดจากการปฏิบัติงาน และรายงานผลการปฏิบัติงานการประเมินให้กับ สพฐ.
- (11) รับกระดาษคำตอบเลือกตอบจากเขตพื้นที่การศึกษาประถมศึกษา สพฐ. รับกระดาษคำตอบจากเขตพื้นที่การศึกษาตามวันที่กำหนด (จัดส่งกระดาษคำตอบ และกำหนด วันส่งกระดาษคำตอบของแต่ละเขตพื้นที่การศึกษาจะประกาศให้ทราบทางระบบ EPCC)
- (12) สพฐ. ดำเนินการตรวจกระดาษคำตอบ วิเคราะห์ ประมวลผลข้อมูล

(13) รายงานผลการประเมินคุณภาพการศึกษาขั้นพื้นฐานเพื่อการประกันคุณภาพผู้เรียน ปีการศึกษา 2557

(14) ติดตาม สนับสนุนเขตพื้นที่การศึกษา/สถานศึกษาทุกแห่ง นำผลการประเมินไปใช้วางแผนและดำเนินการปรับปรุงการเรียนการสอนของโรงเรียน และดำเนินการของเขตพื้นที่การศึกษา เพื่อแก้ไขจุดอ่อนด้อยของนักเรียนเป็นรายบุคคล และรายโรงเรียน

3.5.2 บทบาทหน้าที่ของสำนักงานเขตพื้นที่การศึกษา

เพื่อให้การดำเนินงานประเมินคุณภาพการศึกษาขั้นพื้นฐาน เพื่อการประกันคุณภาพผู้เรียน ปีการศึกษา 2557 ของสำนักงานเขตพื้นที่การศึกษา มีประสิทธิภาพ และมาตรฐานเดียวกันในทุกเขตพื้นที่การศึกษา ควรมีแนวดำเนินการในแต่ละขั้นตอน ดังนี้

1) สำรวจข้อมูลนักเรียนชั้นประถมศึกษาปีที่ 3 ทุกคนสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน และสำรวจตระเวนชายแดน ทั้งนี้ให้แยกข้อมูลเด็กพิเศษด้านการมองเห็น คือ ตาบอด ตาเลือนราง และสำรวจนักเรียนในโรงเรียนสังกัดการศึกษาพิเศษ เช่น ศึกษาสงเคราะห์ ราชประชานุเคราะห์ ฯลฯ

2) จัดทำแผนการประเมิน ฯ ของเขตพื้นที่การศึกษาให้สอดคล้อง กับแนวทางการประเมินของส่วนกลาง และสามารถปฏิบัติได้จริงตามบริบทของแต่ละเขตพื้นที่การศึกษา

3) ปรับปรุงข้อมูลโรงเรียนและจำนวนนักเรียนทางระบบ EPCC ตามข้อมูลจำนวนโรงเรียน จำนวนนักเรียนชั้นประถมศึกษาปีที่ 3

4) แต่งตั้งคณะกรรมการ หรือคณะทำงานชุดต่างๆ ตามความจำเป็น เพื่อรองรับและปฏิบัติงานประเมิน

5) จัดประชุมคณะกรรมการชุดต่างๆ ที่แต่งตั้งขึ้น เพื่อกำหนดมาตรการ แนวปฏิบัติการจัดสอบ รวมทั้งการประชาสัมพันธ์ แจ้งให้โรงเรียนทุกโรงทราบ

6) แต่งตั้งสนามสอบและศูนย์ต่างๆ เพื่อรองรับการดำเนินงานประเมิน

7) รับแบบทดสอบและกระดาษคำตอบ

8) ประชาสัมพันธ์การจัดสอบให้กับ ครูหรืออาจารย์ นักเรียน ผู้ปกครอง ชุมชน และผู้เกี่ยวข้อง ให้ทราบถึงวัตถุประสงค์และความสำคัญของการสอบ

9) จัดสอบในวันที่ 26 กุมภาพันธ์ 2558 ตามตารางสอบที่กำหนด

10) กำกับ ติดตาม ตรวจสอบ และประเมินผลการจัดสอบของสนามสอบ พร้อมให้คำแนะนำปรึกษา แก้ไขปัญหาต่างๆ เพื่อให้การดำเนินการสอบมีความโปร่งใส และเป็นธรรม

11) รับเกณฑ์การตรวจให้คะแนนข้อสอบเขียนตอบจากระบบ EPCC วันที่ 26 กุมภาพันธ์ 2558 และจัดส่งเกณฑ์การตรวจให้คะแนนให้ประธานคณะกรรมการตรวจให้คะแนน

หลังจากสอบเสร็จเพื่อดำเนินการตรวจให้คะแนน พร้อมระบายคะแนนเขียนตอบที่ได้ ลงในกระดาษคำตอบ ให้ดำเนินการให้แล้วเสร็จภายในวันที่ 26 – 27 กุมภาพันธ์ 2558

12) เขตพื้นที่การศึกษา รวบรวมกระดาษคำตอบเลือกตอบนำส่งส่วนกลางตามกำหนด (สำหรับกระดาษคำตอบเขียนตอบให้เขตพื้นที่การศึกษาเก็บรักษาไว้เพื่อการตรวจสอบ)

13) วางแผน และรับผลการสอบจากส่วนกลาง ในภาพรวมระดับประเทศ เขตพื้นที่การศึกษา โรงเรียน รายบุคคล

14) นำผลการประเมินมาใช้ในการพัฒนา ตามแผนการพัฒนายที่วางไว้ สรุปผล และจัดทำรายงานผลการดำเนินงานประเมินของสำนักงานเขตพื้นที่การศึกษาต่อสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน

3.5.3 บทบาท หน้าที่ของโรงเรียน/สนามสอบ

1) สำรวจจำนวนนักเรียนชั้น ป.3 ทุกคน และนักเรียนที่มีปัญหาด้านการมองเห็น คือ ตาบอดเฉพาะที่อ่านอักษรเบรลล์ได้ และสายตาเลือนราง ส่งข้อมูลจำนวนนักเรียนดังกล่าวให้เขตพื้นที่การศึกษา

2) กรอกข้อมูลโรงเรียน และข้อมูลนักเรียนชั้นประถมศึกษาปีที่ 3 ทุกคน ทางเว็บไซต์ <http://juno.tks.co.th/OBECWebRegis/> เพื่อนำข้อมูลนักเรียนพิมพ์ลงกระดาษคำตอบ

3) วางแผนเตรียมความพร้อมในการประเมิน ในด้านบุคลากรที่จะร่วมเป็นคณะกรรมการต่างๆ ของเขตพื้นที่การศึกษา และการเตรียมความพร้อมนักเรียน ฯลฯ

4) ส่งรายชื่อครู และบุคลากรภายในโรงเรียนให้เขตแต่งตั้งเป็นคณะกรรมการต่างๆ

5) เข้าร่วมประชุมกับเขตพื้นที่การศึกษา เพื่อรับทราบนโยบายการประเมินรูปแบบ แนวทางการดำเนินการประเมิน

6) รับคู่มือการจัดสอบ เพื่อเตรียมความพร้อมของครู และบุคลากรภายในโรงเรียนในการดำเนินสอบ

7) รับบัญชีรายชื่อนักเรียนในการจัดห้องสอบแต่ละห้องจากเขตพื้นที่การศึกษาจัดเตรียมห้องสอบช่วงเย็นวันที่ 25 กุมภาพันธ์ 2558 โดยจัดที่นั่งสอบ 35 คน/1ห้องสอบ บัญชีรายชื่อนักเรียนติดที่หน้าห้องสอบทุกห้อง จัดเตรียมดินสอ 2B และยางลบให้นักเรียน หรือกำซับนักเรียนให้เตรียมติดตัวมาในวันสอบ

8) รับข้อสอบกระดาษคำตอบจากศูนย์ประสานการสอบ เข้าวันที่ 26 กุมภาพันธ์ 2558 และกรรมการสนามสอบ ประกอบด้วยผู้อำนวยการ โรงเรียน เป็นประธาน สนามสอบ ครู/อาจารย์ และบุคลากรทางการศึกษา ในจำนวนที่เหมาะสม เป็นกรรมการกลาง โดยกรรมการกลางทำ

หน้าที่รับ – ส่งข้อสอบ และเอกสารประกอบ จากศูนย์ประสานการสอบ ตรวจสอบความถูกต้องของจำนวนเครื่องมือที่ได้รับ ความเรียบร้อยการปิดผนึกของข้อสอบต้องไม่มีร่องรอยการเปิด ส่งต่อไปให้กับกรรมการกำกับการสอบ

9) ปฏิบัติการจัดสอบพร้อมกันวันที่ 26 กุมภาพันธ์ 2558

3.6 การนำผลการประเมินไปใช้

การนำผลการประเมินคุณภาพการศึกษาไปพัฒนาและยกระดับคุณภาพการศึกษาได้อย่างมีประสิทธิภาพ ส่งผลถึงตัวผู้เรียน และปรับปรุงการจัดการเรียนการสอนของโรงเรียน และการกำกับดูแลแนวทางการจัดการศึกษาของเขตพื้นที่การศึกษา ให้เป็นไปตามนโยบายสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน ยุทธศาสตร์ที่ 1 พัฒนาคุณภาพผู้เรียนทุกระดับ ทุกประเภท เป้าหมายที่ 1 นักเรียนระดับก่อนประถมศึกษาที่มีพัฒนาการที่เหมาะสมตามช่วงวัยและได้สมดุล และนักเรียนระดับการศึกษาขั้นพื้นฐานทุกคนมีพัฒนาการเหมาะสมตามช่วงวัยและมีคุณภาพ ทุกส่วนในระบบการศึกษาตั้งแต่ส่วนกลาง สำนักงานเขตพื้นที่การศึกษา โรงเรียน จำเป็นต้องนำผลจากการประเมินมาวิเคราะห์และจัดทำเป็นแผนการยกระดับคุณภาพการศึกษา ในทุกระดับ อย่างเป็นรูปธรรม ตั้งแต่แผนการพัฒนาตัวผู้เรียนรายบุคคล การพัฒนาสภาพการจัดการเรียน การสอนของโรงเรียน สภาพปัญหาและความต้องการ บริบทในการบริหารจัดการ การกำกับดูแลการจัดการเรียนการสอนของเขตพื้นที่การศึกษา ดังนี้

3.6.1 การนำผลการประเมิน ไปยกระดับผลสัมฤทธิ์ทางการเรียน ต้องดำเนินการ ทั้งระบบ โดยกำหนดบทบาทของส่วนกลางว่าแต่ละสำนักก็มีกิจกรรมอะไรบ้างจะสามารถส่งเสริมและสนับสนุนการยกระดับผลสัมฤทธิ์ของเขตพื้นที่การศึกษา โดยศึกษานิเทศก์ร่วมกับศูนย์เครือข่ายเป็นผู้นำในการขับเคลื่อนการยกระดับ และ โรงเรียน ครูผู้สอนต้องมีความรู้ความสามารถด้านการวัดผล โดยทั้งหมดนี้จะต้องมีโครงการ กิจกรรม และงบประมาณในการดำเนินงาน

3.6.2 คณะกรรมการการประเมินคุณภาพการศึกษาขั้นพื้นฐานเพื่อการประกันคุณภาพผู้เรียน ปีการศึกษา 2557 ร่วมกันวางแผนการพัฒนา โดยกำหนดให้ผู้รับผิดชอบในแต่ละสำนักที่เกี่ยวข้องกับการยกระดับคุณภาพการศึกษา เขียนเป็นแผนยกระดับที่เป็นรูปธรรม

3.6.3 มีระบบการติดตามและประเมินผลกระบวนการเรียนการสอนที่สอดคล้องกับจุดเน้นในการยกระดับผลสัมฤทธิ์ทางการเรียน

3.6.4 เขตพื้นที่การศึกษา อาจวิเคราะห์ผลภาพรวมของแต่ละโรงเรียน และภาพรวมของเขตพื้นที่การศึกษา รวมทั้งวิเคราะห์ปัจจัย กระบวนการ แนวทางการส่งเสริม ปรับปรุง การจัดการเรียนการสอนเพื่อยกระดับคุณภาพการศึกษา กำหนดเป็น โครงการ/กิจกรรมในการพัฒนาการจัดการศึกษาของโรงเรียนในเขตพื้นที่การศึกษา

3.6.5 โรงเรียนวิเคราะห์ผลภาพรวม เปรียบเทียบผลการประเมินกับผลภาพรวมระดับเขตพื้นที่การศึกษา และเปรียบเทียบผลระหว่างปีการศึกษา 2556 และปีการศึกษา 2557 เพื่อดูพัฒนาการของโรงเรียน และทราบถึงจุดที่ต้องเร่งดำเนินการพัฒนา ปรับปรุง การจัดการเรียน การสอนของตน จัดทำเป็นแผนยกระดับ โครงการ/กิจกรรม และกำหนดเป้าหมายในการพัฒนา

3.6.6 โรงเรียนวิเคราะห์ผลรายบุคคลของนักเรียน ที่สะท้อนถึงความสามารถของผู้เรียนในแต่ละด้าน ส่งต่อข้อมูลให้ครูผู้สอนในระดับประถมศึกษาปีที่ 4 เพื่อให้ทราบถึงจุดเด่น จุดด้อยที่ต้องเร่งพัฒนา และปรับปรุงเป็นรายบุคคล เพื่อให้ครูผู้สอนสามารถปรับปรุงและพัฒนาผู้เรียนได้อย่างถูกต้อง

4. งานวิจัยที่เกี่ยวข้อง

4.1 งานวิจัยในประเทศ

จากการศึกษารายงานวิจัยภายในประเทศที่ผ่านมา มีผู้ทำวิจัยเกี่ยวกับการทำหน้าที่ต่างกันของ ข้อสอบไว้หลายท่าน โดยศึกษาในรูปแบบการเปรียบเทียบผลการวิเคราะห์การทำหน้าที่ต่างกัน ด้วยวิธี ต่างๆ กับ กลุ่มย่อยที่มีลักษณะแตกต่างกันในด้านภาษา เพศ และสภาพภูมิศาสตร์รวมทั้งการเปรียบเทียบคุณภาพของแบบทดสอบหลังคัดเลือกข้อสอบที่ทำหน้าที่ต่างกันออกแล้ว นอกจากนี้ยังมีการพัฒนาเกณฑ์ที่ใช้ในการตัดสินข้อสอบที่ทำหน้าที่ต่างกันทางเพศ ซึ่งได้เสนอ ดังต่อไปนี้

อมลฉัฐ อุบลรัตน์ (2553) ได้ศึกษาการทำหน้าที่ต่างกัน ของข้อสอบวัดผลสัมฤทธิ์ทางการเรียน ระดับเขตพื้นที่การศึกษา สำนักงานเขตพื้นที่การศึกษาปัตตานี เขต 2 วิชาภาษาไทย และคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 2 ปีการศึกษา 2550 ด้วยวิธีถดถอยโลจิสติก และวิธีแมนเทิล-แฮนส์เซล เมื่อจำแนกตามกลุ่มเพศ และภาษาที่ใช้ในชีวิตประจำวัน ผลการศึกษา จำแนกตามเพศพบว่า

1. ข้อสอบในวิชาคณิตศาสตร์ ข้อสอบที่ทำหน้าที่ต่างกันจะลำเอียงเข้าข้างเพศชายมากกว่าเพศหญิง และในกลุ่มภาษาที่ใช้ในชีวิตประจำวันจะพบว่าข้อสอบที่ทำหน้าที่ต่างกันทั้งในวิชาคณิตศาสตร์ ในวิชาภาษาไทย ส่วนใหญ่จะเข้าข้างกลุ่มผู้สอบที่ใช้ภาษาในชีวิตประจำวันมากกว่ากลุ่มผู้สอบที่ใช้ภาษามลายูถิ่นในชีวิตประจำ

2. การเปรียบเทียบจำนวนข้อสอบที่ทำหน้าที่ต่างกัน ในวิชาคณิตศาสตร์ ในวิชาคณิตศาสตร์ และภาษาไทย เมื่อจำแนกตามกลุ่มเพศ และกลุ่มภาษาที่ใช้ในชีวิตประจำวัน พบว่าวิแมนเทิล-แฮนส์เซล พบข้อที่ทำหน้าที่กันกัน มากกว่าวิธีถดถอยโลจิสติก ยกเว้นวิชาคณิตศาสตร์ ในกลุ่มภาษาวิธีถดถอยโลจิสติกพบมากกว่าวิธีแมนเทิล-แฮนส์เซล 3 ข้อ

3. ค่าสัมประสิทธิ์ของดัชนีการทำหน้าที่ต่างกันของข้อสอบพบว่าวิธีถดถอยคลัสติคกับวิธีแมนเทล-แฮนส์เซลในวิชาคณิตศาสตร์มีความสัมพันธ์ในทางบวก ในวิชาภาษาไทยไม่มีความสัมพันธ์กันทั้งนี้อาจอาจเป็นเพราะข้อสอบวิชาภาษาไทยมีค่าความยาก และอำนาจจำแนกสูง

ญาณภัทร สีหะมงคล (2540) ได้ศึกษาเปรียบเทียบความสอดคล้องของผลการตรวจสอบ ข้อสอบที่ทำหน้าที่ต่างกันระหว่างวิธี Lord's x^2 วิธี Raju's Area Measures และวิธี Closed Interval Area เมื่อขนาดกลุ่มตัวอย่าง ความยาวแบบทดสอบ และสัดส่วนจำนวนข้อสอบที่ทำหน้าที่ต่างกัน ในแบบทดสอบแตกต่างกัน ข้อมูลที่ใช้ในการศึกษาเป็นผลการสอบประเมินคุณภาพและความก้าวหน้าทางการศึกษาวิชาคณิตศาสตร์ของนักเรียนชั้นประถมศึกษาปีที่ 4 ปีการศึกษา 2536 สังกัดสำนักงานการประถมศึกษาแห่งชาติจำนวน 11,404 คน เครื่องมือที่ใช้ในการศึกษาเป็น แบบทดสอบแบบเลือกตอบ จำนวน 80 ข้อ แบ่งเป็น 3 ตอน คือ การคิดคำนวณ ความรู้ความเข้าใจ และโจทย์ปัญหา การศึกษาแบ่งออกเป็น 2 ส่วน ได้แก่ ส่วนที่ 1 เป็นการตรวจสอบจำนวนข้อสอบที่ทำหน้าที่ต่างกัน ความสอดคล้องในการตรวจสอบ ความสัมพันธ์ระหว่างวิธีตรวจสอบ และส่วนที่ 2 เป็น การตรวจสอบอัตราความไม่สอดคล้องในการตรวจสอบ และอัตราความสอดคล้องในการตรวจสอบ ผลการศึกษาพบว่า

1. จำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ได้จากการตรวจสอบ ระหว่าง Lord's x^2 วิธี วิธี Raju's Area Measures และวิธี Closed Interval Area แตกต่างกันเมื่อขนาดกลุ่มตัวอย่างและความยาวของแบบทดสอบต่างกัน ยกเว้นกรณีกลุ่มตัวอย่าง 501 คน แบบทดสอบความยาว 20 , 30 และ 40 ข้อ และกรณีกลุ่มตัวอย่าง 700 คน แบบทดสอบยาว 40 ข้อ

2. ความสอดคล้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ส่วนมากมีค่าปานกลาง ถึงต่ำเกือบทุกเงื่อนไขของการศึกษา ยกเว้นกรณีกลุ่มตัวอย่าง 500 คน แบบทดสอบความยาว 30 , 40 และ 60 ข้อ และกลุ่มตัวอย่าง 700 คน แบบทดสอบยาว 40 ข้อ ที่มีความสอดคล้องกันในระดับปานกลางถึงสูงมาก

3. ความสัมพันธ์ระหว่างวิธีการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันทั้ง 3 วิธี โดยภาพรวมมี สัมประสิทธิ์สหสัมพันธ์ค่อนข้างสูง และมีนัยสำคัญทางสถิติเกือบทุกเงื่อนไขของการศึกษา

4. อัตราความไม่สอดคล้องในการตรวจสอบ เมื่อแบบทดสอบยาวมากขึ้นและมีสัดส่วนจำนวนข้อที่ ต่างกันมากขึ้น อัตราความไม่สอดคล้องชนิดที่ 1 ส่วนใหญ่เพิ่มขึ้น แต่อัตราความไม่สอดคล้อง ชนิดที่ 2 ส่วนใหญ่ลดลง โดยที่อัตราความไม่สอดคล้องชนิดที่ 1 และอัตราความไม่สอดคล้องชนิดที่ 2 ส่วนมากมีค่าใกล้เคียงกัน เมื่อกลุ่มตัวอย่างขนาดใหญ่ Lord's x^2 วิธี Raju's Area Measures มีอัตราความไม่ สอดคล้องสูงกว่าวิธี Closed Interval Area และเมื่อกลุ่มตัวอย่างขนาดเล็ก วิธี Lord's x^2 วิธี วิธี Raju's Area Measures มีอัตราความไม่สอดคล้องต่ำกว่าวิธี Closed Interval Area

5. อัตราความสอดคล้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ เมื่อแบบทดสอบยาวมากขึ้น และสัดส่วนจำนวนข้อสอบที่ทำหน้าที่ต่างกันมากขึ้น อัตราความสอดคล้องชนิดที่1 ส่วนใหญ่ ลดลงแต่อัตราความสอดคล้องชนิดที่2 ส่วนใหญ่เพิ่มขึ้น

พรรณี จินตมาศ (2540) ได้ศึกษาเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันต่อเพศของ แบบทดสอบวิชาคณิตศาสตร์โดยวิธีวิเคราะห์ต่างกัน 3 วิธีคือ วิธีแมนเทล-แฮนส์เซล, วิธี SIBTEST และวิธีแปลงค่าความยากของข้อสอบ โดยใช้ขนาดกลุ่มตัวอย่างต่าง ๆ กัน ผลการศึกษาพบว่า ขนาด กลุ่มตัวอย่าง 500 คน วิธีแปลงค่าความยากของข้อสอบ พบจำนวนข้อสอบที่การทำหน้าที่ต่างกัน น้อย ที่สุด วิธีแปลงค่าความยากของข้อสอบจะตรวจพบจำนวนข้อสอบที่มีความการทำหน้าที่ต่างกันซ้ำ กับ วิธีของแมนเทล – แฮนส์เซล และวิธีSIBTEST ทุกข้อ วิธีSIBTEST พบจำนวนข้อสอบที่มีการทำ หน้าที่ ต่างกันมากที่สุดและเมื่อวิเคราะห์จากขนาดกลุ่มตัวอย่าง 1,000 คน วิธีแมนเทล – แฮนส์เซล พบ จำนวนข้อสอบที่มีการทำหน้าที่ต่างกันมากที่สุด วิธีแปลงค่าความยากของข้อสอบ ไม่พบข้อสอบที่การ ทำหน้าที่ต่างกัน โดยจำนวนข้อสอบที่วิเคราะห์ด้วยวิธีแปลงค่าความยากกับวิธีแมนเทล – แฮนส์เซล และวิธีแปลงค่าความยากกับวิธีSIBTEST แตกต่างกันอย่างมีนัยสำคัญ ที่ระดับ .05 นอกนั้น มีค่า แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติสำหรับจำนวนข้อสอบที่วิเคราะห์ด้วยวิธีแปลงค่าความยาก ระหว่างกลุ่มผู้สอบขนาด 500 คน และขนาด 1,000 คน จะมีจำนวนข้อสอบที่มีการทำหน้าที่ต่างกัน แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 นอกนั้น มีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทาง สถิติ ค่าความเชื่อมั่นของแบบทดสอบหลังการคัดเลือกข้อสอบที่มีการทำหน้าที่ต่างกันออกแล้ว เมื่อ วิเคราะห์จากกลุ่มตัวอย่างขนาด 500 คน จากการวิเคราะห์การทำหน้าที่ต่างกันด้วยวิธีแปลงค่าความ ยาก และวิธีSIBTEST แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติที่ระดับ .05 นอกนั้น มีค่าแตกต่างกัน อย่างไม่มีนัยสำคัญทางสถิติและเมื่อวิเคราะห์จากกลุ่มตัวอย่างขนาด 1,000 คน ค่าความเชื่อมั่นของ แบบทดสอบ หลังการคัดเลือกข้อสอบที่มีการทำหน้าที่ต่างกันออกแล้วจากการตรวจสอบการทำหน้าที่ต่างกัน 3 วิธี นั้น มีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติสำหรับค่าความเชื่อมั่นของแบบทดสอบ หลังการ คัดเลือกข้อสอบที่มีการทำหน้าที่ต่างกันออกแล้วจากการตรวจสอบการทำหน้าที่ต่างกันด้วยวิธี SIBTEST ระหว่างกลุ่มผู้สอบขนาด 500 คน และ 1,000 คน มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติ ที่ ระดับ .05 นอกนั้นมีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

ทอง อยู่สาร (2543) ได้ศึกษาเปรียบเทียบอำนาจการตรวจสอบและจำแนกผิดพลาด ในการ ตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันแบบสมำเสมอและแบบไม่สมำเสมอ ระหว่าง วิธีแมนเทล - แฮนส์เซล และวิธีถดถอยโลจิสติก โดยใช้ความยาวแบบทดสอบและขนาดกลุ่มตัวอย่างต่างกัน พบว่า

1. อำนวยการตรวจสอบและการจำแนกผิดพลาดในการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันทั้ง แบบสมำเสมอและแบบไม่สมำเสมอ ระหว่างวิธีแมนเทิล-แฮนส์เซล และวิธีถดถอยโลจิสติก ภายใต้ ความยาวแบบทดสอบและขนาดกลุ่มตัวอย่างที่ศึกษา เกือบทุกเงื่อนไขไม่แตกต่างกัน
2. ความยาวของแบบทดสอบ ไม่มีผลต่ออำนวยการตรวจสอบและการจำแนกผิดพลาดในการ ตรวจสอบด้วยวิธีแมนเทิล-แฮนส์เซล และวิธีถดถอยโลจิสติก
3. ขนาดกลุ่มตัวอย่างมีผลต่ออำนวยการตรวจสอบด้วย วิธีแมนเทิล-แฮนส์เซล และวิธีถดถอยโลจิสติก เกือบทุกเงื่อนไขของการศึกษา กล่าวคือ เมื่อขนาดกลุ่มตัวอย่างเพิ่มขึ้น อำนวยการ ตรวจสอบจะมีค่าเพิ่มขึ้น แต่พบว่าขนาดกลุ่มตัวอย่างจะไม่มีผลต่อการจำแนกผิดพลาดในเกือบทุกเงื่อนไขของการศึกษา

นวลอนงค์ นันทวงศ์ (2543) ได้ศึกษาเปรียบเทียบการทำหน้าที่ต่างกันของข้อสอบของ แบบทดสอบความเข้าใจในการอ่าน ตามลักษณะการใช้ภาษาพูดในครอบครัว โดยกลุ่มอ้างอิงเป็นกลุ่มใช้ภาษาพูด เป็นภาษาไทยและกลุ่มเปรียบเทียบเป็นกลุ่มที่ใช้ภาษาพูดเป็นภาษาลาวพวนกับ ภาษาลาว ซึ่งทำการวิเคราะห์การทำหน้าที่ต่างกันด้วยวิธี3 วิธีคือ วิธีไคสแควร์วิธีแมนเทิล-แฮนส์เซล และวิธีวิเคราะห์ความแปรปรวน โดยการเปรียบเทียบจำนวนข้อสอบที่มีการทำหน้าที่ต่างกันและ เปรียบเทียบค่าความเชื่อมั่นที่คำนวณด้วยสูตรแบบแบ่งครั้งฉบับของแบบทดสอบหลังการคัดเลือก ข้อสอบที่มีการทำหน้าที่ต่างกันออกแล้ว กลุ่มตัวอย่างที่ใช้ในการศึกษาเป็นนักเรียนชายและนักเรียน หญิง ในระดับชั้นมัธยมศึกษาปีที่3 ภาคเรียนที่2 ปีการศึกษา 2542 ของโรงเรียนสังกัดกรมสามัญ ศึกษา จังหวัดลพบุรีจำนวน 900 คน ซึ่งได้มาโดยการสุ่มแบบสองขั้นตอน ผลการวิจัยพบว่า ในกลุ่ม ผู้สอบที่ใช้ภาษาพูดเป็นภาษาลาวพวน จำนวนข้อสอบที่มีการทำหน้าที่ต่างกันจากการตรวจสอบการทำหน้าที่ต่างกันระหว่างวิธีไคสแควร์กับวิธีแมนเทิล-แฮนส์เซล และวิธีไคสแควร์กับวิธีวิเคราะห์ความแปรปรวนแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนวิธีแมนเทิล-แฮนส์เซล กับวิธี วิเคราะห์ความแปรปรวนรวมแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติส่วนในกลุ่มผู้สอบที่ใช้ภาษาพูด เป็นภาษาลาวพบว่า จำนวนข้อสอบที่มีการทำหน้าที่ต่างกันระหว่างวิธีไคสแควร์กับวิธีวิเคราะห์ความแปรปรวนแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนวิธีไคสแควร์กับวิธีแมนเทิล-แฮนส์ เซล และวิธีแมนเทิล-แฮนส์เซล กับวิธีวิเคราะห์ความแปรปรวนรวม แตกต่างกันอย่างไม่มีนัยสำคัญ ทางสถิติ จำนวนข้อสอบที่มีการทำหน้าที่ต่างกัน เมื่อมีวิธีวิเคราะห์ทั้งวิธีไคสแควร์วิธีแมนเทิล-แฮนส์เซล และวิธีวิเคราะห์ความแปรปรวน ระหว่างกลุ่มผู้สอบที่ใช้ภาษาพูดเป็นภาษาลาวและภาษาลาว พวน พบว่า จำนวนข้อสอบที่มีการทำหน้าที่ต่างกันแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติค่าความ เชื่อมั่นของแบบทดสอบความเข้าใจในการอ่านหลังจากคัดเลือกข้อสอบที่มีการทำหน้าที่ต่างกันออก แล้ว เมื่อกลุ่มผู้สอบที่ใช้ภาษาพูดเป็นภาษาลาวและภาษาลาวพวน พบว่า ต่างก็มีค่าความเชื่อมั่น ระหว่างวิธีตรวจสอบ

การทำหน้าที่ต่างกันทั้งวิธีไคสแควร์วิธีแมนเทล – แชนส์เชล และวิธีวิเคราะห์ ความแปรปรวน แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติส่วนค่าความเชื่อมั่นของแบบทดสอบความ เข้าใจในการอ่านหลังจากคัดเลือกรหัสสอบที่มีการทำหน้าที่ต่างกันออกแล้ว เมื่อมีวิธีตรวจสอบการทำหน้าที่ต่างกันวิธีเดียวกัน เมื่อกลุ่มผู้สอบที่ใช้ภาษาพูดเป็นภาษาลาวและภาษาลาวพวน มีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

4.2 งานวิจัยในต่างประเทศ

Swaminatha และ Rogers(1993) ได้เปรียบเทียบวิธีการตรวจสอบการทำหน้าที่ต่างกันระหว่างวิธีถดถอยโลจิสติกกับวิธีแมนเทล-แชนส์เชล โดยใช้การจำลองข้อมูล ศึกษาเกี่ยวกับการกระจายของสถิติทดสอบและประสิทธิภาพการตรวจสอบการทำหน้าที่ต่างกันของแบบเอกรูป และแบบอนุรูป การศึกษาปัจจัยที่แปรเปลี่ยนคือ ขนาดกลุ่มตัวอย่าง 2 กลุ่ม คือ 250 คน และ 500 คน ความยาวแบบสอบ 2 ขนาด คือ จำนวน 40 ข้อ และ 80 ข้อ

ผลการศึกษาด้านการกระจายของสถิติทดสอบพบว่า การกระจายของสถิติทดสอบของทั้ง 2 วิธี อยู่ในระดับที่น่าพอใจ ยกเว้นการกระจายของสถิติของวิธีถดถอยโลจิสติกไม่เป็นไปตามที่คาดไว้ในกรณีที่ข้อสอบยากมากและค่าอำนาจจำแนกสูง ประสิทธิภาพในการตรวจสอบพบว่า 2 วิธี มีประสิทธิภาพในการตรวจสอบเท่ากันในการตรวจสอบการทำหน้าที่ต่างกันแบบเอกรูป แต่วิธีถดถอยโลจิสติกมีประสิทธิภาพมากกว่าในการตรวจสอบการทำหน้าที่ต่างกันแบบอนุรูป ขนาดกลุ่มตัวอย่างเป็นปัจจัยที่มีผลกระทบต่อตรวจสอบการทำหน้าที่ต่างกันแบบอนุรูป คือเมื่อเพิ่มขนาดกลุ่มตัวอย่างอัตราการตรวจสอบจะเพิ่มขึ้น ส่วนการกระจายคะแนนและแบบสอบไม่มีผลต่ออัตราการตรวจสอบ วิธีถดถอยโลจิสติกตรวจพบข้อสอบที่ทำหน้าที่ต่างกันแบบอนุรูปได้ดีในกรณีข้อสอบยากง่ายปานกลาง และค่าอำนาจจำแนกสูง ส่วนวิธีแมนเทล-แชนส์เชล ตรวจสอบข้อสอบยากง่ายปานกลางได้น้อยมาก แต่วิธีแมนเทล – แชนส์เชลสามารถตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันแบบอนุรูปได้ดีในกรณีข้อสอบยากง่าย และข้อสอบยาก

Mazor และ Clauser (1995) ศึกษาเปรียบเทียบวิธีตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธีถดถอยโลจิสติกกับวิธีแมนเทล-แชนส์เชลกรณีที่ใช้เกณฑ์ภายนอกเป็นความสามารถหลากหลาย (multiple ability) ศึกษาจากข้อมูลจริง จากผลการตอบแบบสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาประวัติศาสตร์ และวิชาเคมีของนักเรียนระดับชั้นมัธยมศึกษา โดยกลุ่มผู้สอบจำแนกตามเพศความสามารถทางภาษา เกณฑ์ที่ใช้ในการตรวจสอบแปรเปลี่ยนคือ 1. ใช้คะแนนรวมของแบบสอบ 2. ใช้คะแนนรวมของแบบสอบและตัวแปรความสามารถภายนอก ซึ่งในที่นี้ใช้ความถนัดทางภาษา และความถนัดทางคณิตศาสตร์เป็นตัวแปรที่เพิ่มเข้ามา ส่วนความยาวแบบสอบ ใช้ 75 ข้อในทุกเงื่อนไข

ผลการศึกษาพบว่าวิธีถดถอยโลจิสติกตรวจพบข้อสอบที่ทำหน้าที่ต่างกันได้มากกว่าวิธีแมนเทล-แฮนส์เซลในทุกเงื่อนไข แต่ผลที่ได้จากการตรวจสอบสอดคล้องกันกล่าวคือ เมื่อนำคะแนนความถนัดทางภาษาและคะแนนความถนัดทางคณิตศาสตร์ เข้ามาเป็นเกณฑ์ร่วมกับคะแนนรวมของแบบสอบจะทำให้ตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันได้ลดลง โดยคะแนนความถนัดทางภาษาเมื่อบรวมกับคะแนนรวมของแบบสอบทำให้การตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันได้ต่ำกว่าเมื่อใช้คะแนนความถนัดทางคณิตศาสตร์ร่วมกับคะแนนรวมของแบบสอบ Clauser ได้อธิบายเพิ่มเติมว่าความสามารถที่เพิ่มเข้ามา หากมีความตรง และมีความสัมพันธ์กับแบบสอบ และทำให้ข้อสอบทำหน้าที่ต่างกันควรที่จะตัดข้อสอบข้อนั้นทิ้ง ซึ่งข้อสอบลักษณะนี้มีลักษณะเป็นข้อสอบพหุมิติที่เกิดจากความแตกต่างของความสามารถที่สองที่มีความสัมพันธ์กับคุณลักษณะแฝงหลัก ดังนั้นหากตรวจสอบข้อสอบพหุมิติที่ทำหน้าที่ต่างกัน ควรนำเกณฑ์ความสามารถมาคิดคำนวณด้วยเพราะทำให้ผลการตอบมีความถูกต้องมากขึ้น และวิธีที่มีความเหมาะสมในการตรวจสอบข้อสอบลักษณะเช่นนี้คือวิธีถดถอยโลจิสติก

Narayanan และ Swaminathan (1996) เปรียบเทียบประสิทธิภาพในการตรวจสอบข้อสอบทำหน้าที่ต่างกันแบบอนุกรมระหว่างวิธีแมนเทล-แฮนส์เซล วิธีถดถอยโลจิสติก และวิธีโครซิบท (CRO-SIB) ศึกษาจากข้อมูลจำลอง ภายใต้ปัจจัยที่แปรเปลี่ยนดังนี้ 1. กลุ่มตัวอย่าง ใช้กลุ่มตัวอย่าง 500 คน และ 1,000 คนในกลุ่มอ้างอิง และ 200 คนและ 500 คนในกลุ่มเปรียบเทียบ 2. การกระจายความสามารถระหว่างกลุ่มอ้างอิง และกลุ่มเปรียบเทียบเท่ากัน และ ไม่เท่ากัน 3. สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันแบบสอบโดยจำลองข้อมูลใน 3 ระดับ คือ 0% ,10 %และ 20% 4. ขนาดของข้อสอบที่ทำหน้าที่ต่างกันหรือพื้นที่ความแตกต่างระหว่างโค้งคุณลักษณะข้อสอบสองกลุ่ม กำหนด 4 ระดับ คือ 0.4 , 0.6 , 0.8 และ 1.0 5. ค่าความยากและค่าอำนาจจำแนกของข้อสอบส่วนค่าการเดา กำหนดให้เท่ากันคือ 0.2 และศึกษาความยาว 40 ข้อในทุกเงื่อนไข

ผลการศึกษาพบว่าวิธีถดถอยโลจิสติกและวิธีโครซิบทให้ผลใกล้เคียงกันในการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันแบบอนุกรม และทั้งสองวิธีตรวจสอบได้ดีกว่าวิธีแมนเทล-แฮนส์เซล ปัจจัยที่มีผลต่อการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันแบบอนุกรมมีดังนี้ ขนาดกลุ่มตัวอย่างเมื่อเพิ่มขนาดกลุ่มตัวอย่างทั้งสามวิธีสามารถตรวจสอบได้เพิ่มขึ้น การกระจายความสามารถระหว่างกลุ่มอ้างอิงกับกลุ่มเปรียบเทียบเพิ่มขึ้นจาก 0.4 เป็น 1.0 ทั้งสามวิธีสามารถตรวจสอบข้อสอบทำหน้าที่ต่างกันแบบอนุกรมได้เพิ่มขึ้น ข้อสอบที่ตรวจพบทำหน้าที่ต่างกันด้วยวิธีโลจิสติก และ วิธีโครซิบท ส่วนใหญ่เป็นข้อสอบที่ค่าความยากต่ำ ค่าอำนาจจำแนกสูง ส่วนวิธีแมนเทล-แฮนส์เซล ตรวจพบข้อสอบที่ทำหน้าที่ต่างกันแบบอนุกรมได้ดีเฉพาะกรณีข้อสอบยาก และข้อสอบง่ายซึ่งเป็นลักษณะที่โค้งคุณลักษณะข้อสอบของผู้สอบสองกลุ่มตัดกันที่ความสามารถสูงหรือความสามารถต่ำเท่านั้น

บทที่ 3

วิธีดำเนินการวิจัย

การวิจัยครั้งนี้เป็นการวิจัยเชิงสำรวจมีวัตถุประสงค์เพื่อศึกษาการทำหน้าที่ต่างกันของข้อสอบวัดความสามารถด้านการคำนวณของการทดสอบระดับชาติชั้นประถมศึกษาปีที่ 3 สังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐานปีการศึกษา 2557 โดยใช้วิธีถดถอยโลจิสติก (LR) ในการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ โดยมีรายละเอียดขั้นตอนการดำเนินการวิจัยดังนี้

1. ประชากรที่ใช้ในการศึกษา

ประชากรที่ใช้ในการศึกษาค้นคว้าครั้งนี้ คือ นักเรียนชั้นประถมศึกษาปีที่ 3 ปีการศึกษา 2557 จำนวน 464,532 คน เป็นเพศชาย 224,615 คน เป็นเพศหญิง 239,917 คน

2. เครื่องมือและข้อมูลที่ใช้ในการวิจัย

2.1 เครื่องมือที่ใช้ในการวิจัย ได้แก่ แบบทดสอบวัดความสามารถด้านการคำนวณชั้นประถมศึกษาปีที่ 3 ของสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน ปีการศึกษา 2557 จำนวน 27 ข้อ ซึ่งเป็นข้อสอบปรนัยแบบเลือกตอบ 4 ตัวเลือก

2.2 ข้อมูลที่ใช้ในการวิจัย ได้แก่ ข้อมูลทุติยภูมิ (Secondary data) จากสำนักทดสอบทางการศึกษาของสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน (สพฐ.) เป็นผลการตอบแบบทดสอบระดับชาติ (NT) ด้านการคำนวณ ของนักเรียนชั้นประถมศึกษาปีที่ 3 ปีการศึกษา 2557 จำนวน 27 ข้อ จำนวน 464,532 ชุด

3. การวิเคราะห์ข้อมูล

3.1 การเตรียมข้อมูล

3.1.1 จัดเตรียมเพิ่มข้อมูลสำหรับใช้ในการวิเคราะห์ โดยแยกผลการตอบข้อสอบเป็นเพศชายและเพศหญิง

3.1.2 ตรวจสอบข้อมูลการกระจายของคะแนน และคุณภาพแบบทดสอบโดยนำคะแนนผลการสอบวัดความสามารถด้านการคำนวณ ชั้นประถมศึกษาปีที่ 3 ของสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน ปีการศึกษา 2557 จำนวน 464,532 คน มาหาค่าสถิติพื้นฐานคุณภาพแบบทดสอบทั้งรายข้อและทั้งฉบับ โดยแยกตามเพศเพื่อศึกษาสารสนเทศของผลการสอบ

3.2 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการนำคะแนนผลการสอบวัดความสามารถด้านการคำนวณ ชั้นประถมศึกษาปีที่ 3 ของสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน ปีการศึกษา 2557 จำนวน 464,532 คน จำนวน 27 ข้อ มาวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบเป็นรายข้อด้วยวิธีถดถอยโลจิสติก (LR) ของคามิลลี และเชฟเพอร์ด (Camilli & Shepard, 1994) โดยใช้โปรแกรม SPSS เพื่อทำนายโอกาสในการตอบข้อสอบถูกหรือผิดของผู้สอบ โดยใช้กลุ่มผู้สอบ (เพศชายเป็นกลุ่มอ้างอิง และเพศหญิงเป็นกลุ่มเปรียบเทียบ) ความสามารถผู้สอบ และปฏิสัมพันธ์กลุ่มผู้สอบเป็นตัวทำนาย ตามลำดับดังนี้

ขั้นตอนที่ 1 วิเคราะห์นัยสำคัญของปฏิสัมพันธ์

$$\text{Model 1} \quad \psi_{ij} = \tau_0 + \tau_1 X_{ij} + \tau_2 G_j + \tau_3 (G_j X_{ij})$$

พิจารณาค่านัยสำคัญทางสถิติของ τ_3 หากมีนัยสำคัญทางสถิติ แสดงว่าข้อสอบข้อนั้นทำหน้าที่ต่างกันแบบอเนกรูป (Nonuniform DIF) แต่ถ้า τ_3 ไม่มีนัยสำคัญทางสถิติ แสดงว่าข้อสอบข้อนั้น ไม่ทำหน้าที่ต่างกันแบบอเนกรูป ให้ดำเนินขั้นที่ 2

ขั้นตอนที่ 2 วิเคราะห์นัยสำคัญของกลุ่มผู้สอบ

$$\text{Model 2} \quad \psi_{ij} = \tau_0 + \tau_1 X_{ij} + \tau_2 G_j$$

ถ้าพบ τ_2 หากมีนัยสำคัญทางสถิติ แสดงว่าข้อสอบข้อนั้นทำหน้าที่ต่างกันแบบเอกรูป (Uniform DIF) แต่ถ้า τ_2 ไม่มีนัยสำคัญทางสถิติ แสดงว่าข้อสอบไม่ทำหน้าที่ต่างกันซึ่งแสดงว่า τ_1 มีนัยสำคัญทางสถิติ เมื่อพิจารณาจาก **Model 3** $\psi_{ij} = \tau_0 + \tau_1 X_{ij}$

3.3 ตรวจสอบโอกาสตอบข้อสอบถูก

ขั้นที่ 1 แบ่งระดับความสามารถผู้สอบ

ระดับความสามารถต่ำ นักเรียนที่มีคะแนนตั้งแต่ 0 ถึง (ค่าเฉลี่ยของคะแนนสอบ – ส่วนเบี่ยงเบนมาตรฐาน)

ระดับความสามารถสูง นักเรียนที่มีคะแนน ตั้งแต่ (ค่าเฉลี่ยของคะแนนสอบ + ส่วนเบี่ยงเบนมาตรฐาน) ถึง 27

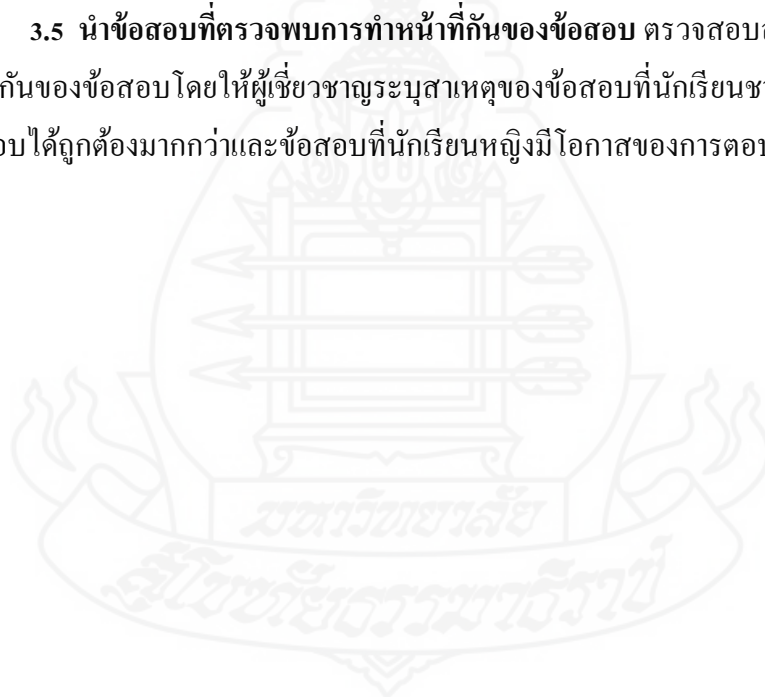
ขั้นที่ 2 เปรียบเทียบคะแนนเฉลี่ยแต่ละข้อระหว่างนักเรียนชายและนักเรียนหญิง ในกรณีกลุ่มใดกลุ่มหนึ่งมีคะแนนเฉลี่ยในข้อนั้นสูงกว่า แสดงว่านักเรียนกลุ่มนั้นมีโอกาสตอบข้อสอบข้อนั้นถูกมากกว่าอีกกลุ่มหนึ่ง

3.4 ตรวจสอบขนาดอิทธิพลการทำหน้าที่ต่างกันของข้อสอบ

นำข้อสอบที่ทำหน้าที่ต่างกันมาตรวจสอบขนาดของอิทธิพลหรือขนาดการทำหน้าที่ต่างกัน (Effect size) ที่ได้จะเป็น ΔR^2 (มาจากผลต่างของ R^2 โมเดล 1 กับ โมเดล 3) โดยใช้เกณฑ์ซัมโบและโทมัส เพื่อใช้ในการจำแนกขนาดการทำหน้าที่ต่างกันของข้อสอบ เกณฑ์สำหรับดัชนี ΔR^2 มีดังนี้

$\Delta R^2 < 0.13$	แสดงว่า DIF มีขนาดเล็ก
$0.13 \leq \Delta R^2 \leq 0.26$	แสดงว่า DIF มีขนาดปานกลาง
$\Delta R^2 > 0.26$	แสดงว่า DIF มีขนาดปานใหญ่

3.5 นำข้อสอบที่ตรวจพบการทำหน้าที่กันของข้อสอบ ตรวจสอบสาเหตุข้อการทำหน้าที่ต่างกันของข้อสอบโดยให้ผู้เชี่ยวชาญระบุสาเหตุของข้อสอบที่นักเรียนชายมีโอกาสของการตอบข้อสอบได้ถูกต้องมากกว่าและข้อสอบที่นักเรียนหญิงมีโอกาสของการตอบข้อสอบได้ถูกต้องมากกว่า



บทที่ 4

ผลการวิเคราะห์ข้อมูล

การวิจัยครั้งนี้มีจุดมุ่งหมาย เพื่อศึกษาการทำหน้าที่ต่างกันของข้อสอบวัดความสามารถด้านคำนวณระหว่างนักเรียนชาย และนักเรียนหญิงของการทดสอบระดับชาติของนักเรียนชั้นประถมศึกษาปีที่ 3 สังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน ปีการศึกษา 2557

ผู้วิจัยนำเสนอผลการวิเคราะห์ข้อมูลตามลำดับ ดังนี้

ตอนที่ 1 ผลการวิเคราะห์ค่าสถิติพื้นฐานของผลการตอบข้อสอบและคุณภาพของข้อสอบ

ตอนที่ 2 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีถดถอยโลจิสติก

ตอนที่ 3 วิเคราะห์สาเหตุของข้อสอบวัดความสามารถด้านการคำนวณของการทดสอบระดับชาติที่ทำหน้าที่ต่างกัน

สัญลักษณ์ที่ใช้ในการนำเสนอผลการวิเคราะห์ข้อมูล

T_1	แทน ผลของความสามารถผู้สอบ
T_2	แทน ความแตกต่างระหว่างกลุ่มในการตอบข้อสอบถูก
T_3	แทน ปฏิสัมพันธ์ระหว่างกลุ่มกับความสามารถผู้สอบ
UDIF	แทน ข้อสอบทำหน้าที่ต่างกันแบบเอกรูป
NDIF	แทน ข้อสอบทำหน้าที่ต่างกันแบบอนเอกรูป
NODIF	แทน ข้อสอบทำหน้าที่ไม่ต่างกัน
R^2	แทน ค่าสถิติอาร์-สแควร์ (R-square)

ตอนที่ 1 ค่าสถิติพื้นฐานของผลการตอบและคุณภาพของข้อสอบ

1.1 ค่าสถิติพื้นฐานของผลการตอบข้อสอบ ผู้วิจัยนำผลการตอบข้อสอบวัดความสามารถด้านคำนวณของนักเรียนชั้นประถมศึกษาปีที่ 3 ในการทดสอบระดับชาติ ปีการศึกษา 2557 จำนวน 27 ข้อ ซึ่งมีผู้เข้าสอบจำนวน 464,532 คน เป็นเพศชาย 224,615 คน เป็นเพศหญิง 239,917 คน มาวิเคราะห์หาค่าเฉลี่ย (Mean) ส่วนเบี่ยงเบนมาตรฐาน (S.D.) ค่าต่ำสุด (Min) ค่าสูงสุด (Max) พิสัย (Rang) มัชยฐาน (Median) ฐานนิยม (Mode) ความเบ้ (Sk) และความโค้ง (Ku) โดยจำแนกเป็นนักเรียนเพศชาย และนักเรียนเพศหญิงได้ผลการวิเคราะห์ดังตารางที่ 4.1

ตารางที่ 4.1 ค่าสถิติพื้นฐานของแบบทดสอบ

ค่าสถิติ	นักเรียนเพศชาย	นักเรียนเพศหญิง
ผู้เข้าสอบ	224,615	239,917
จำนวนข้อสอบ	27	27
ค่าเฉลี่ย (Mean)	11.04	11.94
ส่วนเบี่ยงเบนมาตรฐาน (S.D.)	4.84	4.82
ต่ำสุด (Min)	0	0
สูงสุด (Max)	27	27
พิสัย (Range)	27	27
มัธยฐาน (Median)	10	11
ฐานนิยม (Mode)	8	10
ความเบ้ (Sk)	.62	.43
ความโด่ง (Ku)	-.20	-.41

จากผลการวิเคราะห์ค่าสถิติพื้นฐานของแบบทดสอบวัดความรู้ด้านคำนวณ 27 ข้อ เมื่อจำแนกกลุ่มสอบตามตัวแปรเพศ พบว่า เพศชายมีค่าเฉลี่ยของคะแนนสอบเท่ากับ 11.04 คะแนน มีการกระจายออกจากค่าเฉลี่ย 4.84 คะแนน คะแนนต่ำสุดเท่ากับ 0 คะแนน คะแนนสูงสุดเท่ากับ 27 คะแนน มีฐานนิยมเท่ากับ 8 คะแนน ค่าความเบ้เท่ากับ .62 ซึ่งเมื่อเทียบกับเส้นโค้งปกติ ที่มีค่าเท่ากับ 0 แสดงว่าผลการสอบมีลักษณะเบ้ทางขวา นั่นคือผลสอบส่วนใหญ่ผู้เรียนได้ต่ำกว่าค่าเฉลี่ยเล็กน้อย และมีค่าความโด่งเท่ากับ -.20 เมื่อเทียบกับเส้นโค้งปกติที่มีค่าเท่ากับ 3 แสดงว่าการแจกแจงความถี่ในลักษณะโค้งลาดที่แบนกว่าปกติ

ผู้สอบเพศหญิงมีค่าเฉลี่ยของคะแนนสอบเท่ากับ 11.94 คะแนน มีการกระจายออกจากค่าเฉลี่ย 4.82 คะแนน คะแนนต่ำสุดเท่ากับ 0 คะแนน คะแนนสูงสุดเท่ากับ 27 คะแนน มีฐานนิยมเท่ากับ 10 คะแนน ค่าความเบ้เท่ากับ .43 ซึ่งเมื่อเทียบกับเส้นโค้งปกติ ที่มีค่าเท่ากับ 0 แสดงว่าผลการสอบมีลักษณะเบ้ทางขวา นั่นคือผลสอบส่วนใหญ่ผู้เรียนได้ต่ำกว่าค่าเฉลี่ยเล็กน้อย และมีค่า

ความโค้งเท่ากับ -.41 เมื่อเทียบกับเส้นโค้งปกติที่มีค่าเท่ากับ 3 แสดงว่ามีการแจกแจงความถี่ในลักษณะโค้งลาดที่แบนกว่าปกติ

1.2 คุณภาพรายข้อของข้อสอบวัดผลสัมฤทธิ์ข้อสอบวัดความสามารถด้านการคำนวณ แบบทดสอบที่มีความเหมาะสม ต้องมีค่าความยากง่ายอยู่ในระหว่าง .20 - .80 มีค่าอำนาจจำแนก ตั้งแต่ .20 ขึ้นไป และมีค่าเที่ยง (Reliability) ประมาณ .70 - .80

ตารางที่ 4.2 ค่าเฉลี่ยการตอบข้อสอบถูกของนักเรียนชาย และนักเรียนหญิง ค่าความยาก ค่าอำนาจจำแนก และค่าความเที่ยงของข้อสอบวัดความสามารถด้านการคำนวณ

ข้อ	ค่าความยาก		ค่าอำนาจจำแนก		ค่าความยาก	ค่าอำนาจจำแนก
	นักเรียนชาย	นักเรียนหญิง	นักเรียนชาย	นักเรียนหญิง		
1	.47	.59	.361	.345	.54	.36
2	.26	.30	-.012	.000	.26	-.01
3	.43	.46	.334	.351	.45	.34
4	.35	.41	.299	.284	.38	.29
5	.40	.45	.377	.390	.42	.39
6	.43	.50	.412	.412	.47	.42
7	.51	.53	.368	.370	.52	.37
8	.27	.25	.224	.278	.26	.25
9	.52	.60	.370	.373	.56	.38
10	.40	.45	.344	.334	.43	.34
11	.30	.31	.295	.257	.31	.28
12	.39	.38	.327	.320	.38	.32
13	.36	.39	.294	.265	.38	.28
14	.32	.34	.115	.088	.33	.10
15	.39	.41	.421	.376	.40	.40

ตารางที่ 4.2 (ต่อ)

ข้อ	ค่าความยาก		ค่าอำนาจจำแนก		ค่าความยาก	ค่าอำนาจจำแนก
	นักเรียน	นักเรียน	นักเรียน	นักเรียน		
	ชาย	หญิง	ชาย	หญิง		
16	.39	.38	.295	.305	.38	.30
17	.65	.72	.335	.308	.69	.33
18	.41	.42	.158	.160	.42	.16
19	.39	.40	.221	.237	.39	.23
20	.23	.21	.095	.100	.22	.10
21	.53	.57	.366	.351	.55	.36
22	.39	.43	.295	.319	.41	.31
23	.33	.34	.111	.117	.33	.12
24	.58	.63	.115	.079	.61	.10
25	.45	.52	.361	.358	.49	.36
26	.49	.55	.381	.400	.52	.39
27	.37	.43	.306	.330	.40	.32

*Reliability Kuder – Richardson 20= .76

จากตารางที่ 4.2 คุณภาพรายข้อของข้อสอบวัดความสามารถด้านการคำนวณ จำนวน 27 ข้อ พบว่าค่าเฉลี่ยการตอบข้อสอบถูกนักเรียนชายสูงกว่านักเรียนหญิงจำนวน 5 ข้อ คือข้อ 4, 8, 12, 16 และ 20 ค่าเฉลี่ยการตอบข้อสอบถูกของนักหญิงสูงกว่านักเรียนชาย 22 ข้อ คือ 1, 2, 3, 5, 6, 7, 9, 10, 11, 13, 14, 15, 17, 18, 19, 21, 22, 23, 24, 25, 26 และ 27 ค่าความยากมีค่าระหว่าง .22 ถึง .69 ค่าอำนาจจำแนกมีค่าระหว่าง .07 ถึง .42 และค่าความเที่ยง เท่ากับ .76

ตอนที่ 2 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีถดถอยโลจิสติก

การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบภายใต้วิธีถดถอยโลจิสติก โดยการทดสอบระดับนัยสำคัญของข้อสอบวัดความสามารถด้านการคำนวณ 27 ข้อ รายละเอียดดังตารางที่ 4.3

ตารางที่ 4.3 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีถดถอยโลจิสติก โดยการทดสอบระดับนัยสำคัญของข้อสอบวัดความสามารถด้านการคำนวณ

ข้อ	Model 1		Model 2		Model 3		ผลการ ตรวจสอบ
	ปฏิสัมพันธ์ระหว่าง กลุ่มผู้สอบกับความสามารถ		กลุ่มผู้สอบ		ความสามารถ		
	τ_3	Sig	τ_2	Sig	τ_1	Sig	
1	.000	.988	-.241	.000*	.233	.000*	UDIF
2	-.006	.000*	.010	.123	.039	.000*	NDIF
3	-.014	.000*	.047	.000*	.211	.000*	NDIF
4	.005	.000*	-.101	.000*	.179	.000*	NDIF
5	-.013	.000*	.003	.679	.238	.000*	NDIF
6	-.008	.002*	-.132	.000*	.264	.000*	NDIF
7	-.002	.309	.095	.000*	.237	.000*	UDIF
8	-.036	.002*	.249	.000*	.160	.000*	NDIF
9	-.014	.000*	-.182	.000*	.251	.000*	NDIF
10	.002	.130	-.033	.000*	.208	.000*	UDIF
11	.016	.000*	.104	.000*	.171	.000*	NDIF
12	.002	.317	.268	.000*	.194	.000*	UDIF
13	.013	.000*	.017	.011*	.172	.000*	NDIF
14	.010	.000*	.000	.995	.087	.000*	NDIF
15	.029	.000*	.119	.000*	.245	.000*	NDIF
16	-.008	.000*	.201	.000*	.182	.000*	NDIF

ตารางที่ 4.3 (ต่อ)

ข้อ	Model 1 ปฏิสัมพันธ์ระหว่างกลุ่ม ผู้สอบ กับความสามารถ (g by x)		Model 2 กลุ่มผู้สอบ (g)		Model 3 ความสามารถ (x)		ผลการ ตรวจสอบ
	τ_3	Sig	τ_2	Sig	τ_1	Sig	
	17	.010	.000*	-.200	.000*	.247	
18	-.002	.142	.044	.000*	.112	.000*	UDIF
19	-.010	.000*	.074	.000*	.145	.000*	NDIF
20	-.006	.000*	.192	.000*	.087	.000*	NDIF
21	.009	.000*	.018	.006*	.237	.000*	NDIF
22	-.017	.000*	-.037	.000*	.189	.000*	NDIF
23	-.004	.001*	.034	.000*	.092	.000*	NDIF
24	.016	.000*	-.138	.000*	.090	.000*	NDIF
25	-.004	.011*	-.126	.000*	.229	.000*	NDIF
26	-.020	.000*	-.016	.019*	.256	.000*	NDIF
27	-.018	.000*	-.110	.000*	.195	.000*	NDIF

จากตารางที่ 4.3 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีลดข้อยกเว้นโลจิสติก เมื่อใช้คะแนนรวมของข้อสอบวัดความสามารถด้านการคำนวณเป็นเกณฑ์จับคู่เปรียบเทียบ ซึ่งพิจารณาค่านัยสำคัญ

ค่านัยสำคัญของปฏิสัมพันธ์ระหว่างกลุ่มผู้เข้าสอบกับความสามารถ (g by x) ถ้าข้อสอบข้อใดมีนัยสำคัญที่ระดับ .05 หมายความว่าข้อสอบข้อนั้นทำหน้าที่ต่างกันแบบอนุรูป ผลจากการตรวจสอบ พบว่ามีข้อสอบที่ทำหน้าที่ต่างกันแบบอนุรูป จำนวน 22 ข้อ คือ ข้อ 2, 3, 4, 5, 6, 8, 9, 11, 13, 14, 15, 16, 17, 19, 20, 21, 22, 23, 24, 25, 26 และข้อ 27

ค่านัยสำคัญของกลุ่มผู้สอบ (g) ถ้าข้อสอบข้อใดมีนัยสำคัญที่ระดับ .05 หมายความว่าข้อสอบข้อนั้นทำหน้าที่ต่างกันแบบเอกรูป ผลการตรวจสอบพบข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูป จำนวน 5 ข้อ คือข้อ 1, 7, 10, 12 และ 18

ความแตกต่างระหว่างกลุ่มในการตอบข้อสอบถูก (τ_2) พบว่า ข้อสอบวัดความสามารถด้านการคำนวณฉบับนี้นักเรียนชายจะมีโอกาสในการตอบข้อสอบถูกมากกว่านักเรียนหญิงจำนวน

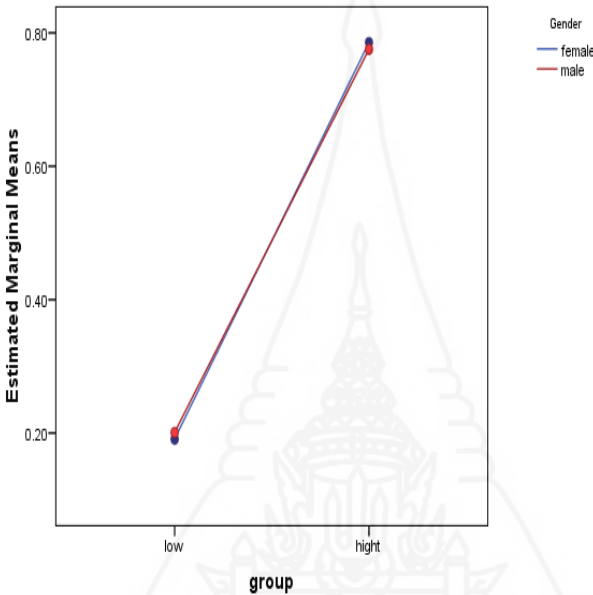
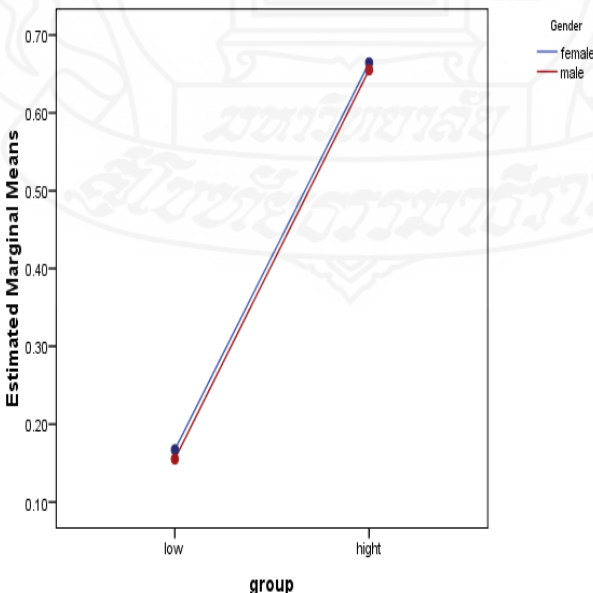
16 ข้อ คือ 2, 3, 5, 7, 8, 11, 12, 13, 14, 15, 16, 18, 19, 20, 21 และ 23 และนักเรียนหญิงมีโอกาสตอบข้อสอบถูกมากกว่านักเรียนชายจำนวน 11 ข้อ คือ 1, 4, 6, 9, 10, 17, 22, 24, 25, 26 และ 27

การตรวจสอบโอกาสในการตอบข้อสอบถูกระหว่างนักเรียนชายและนักเรียนหญิงในระดับความสามารถ 2 ระดับ ของข้อสอบวัดความสามารถด้านการคำนวณ 27 ข้อ รายละเอียดดังตารางที่ 4.4

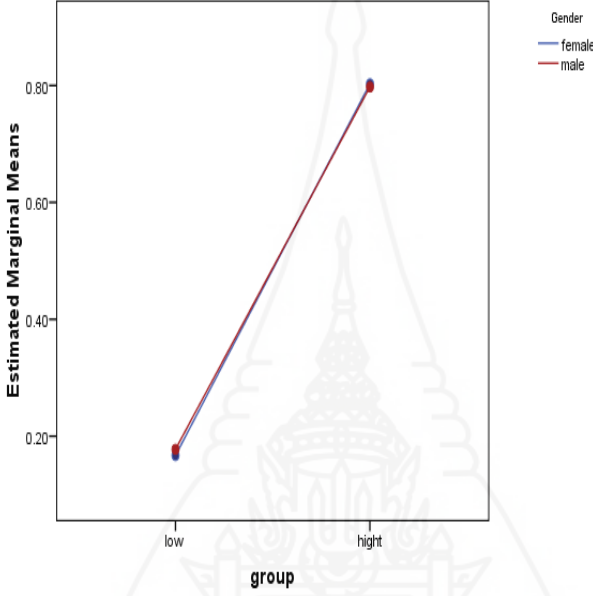
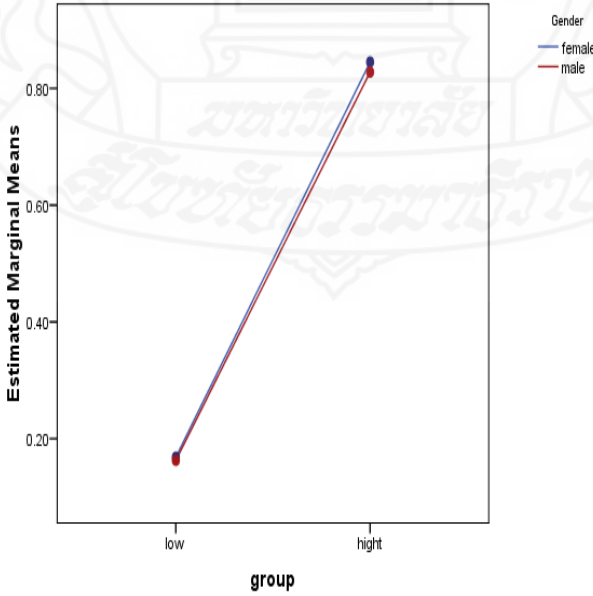
ตารางที่ 4.4 กราฟแสดงการตอบข้อสอบถูกระหว่างนักเรียนชายและนักเรียนหญิง ในกลุ่มนักเรียนที่มีระดับความสามารถสูง และระดับความสามารถต่ำ

ข้อ	กราฟแสดงโอกาสในการตอบข้อสอบถูก	แปลผล
Estimated Marginal Means of score item N1		
1		<ul style="list-style-type: none"> - ระดับความสามารถต่ำ <li style="padding-left: 20px;">นักเรียนหญิงมีโอกาสตอบข้อสอบถูกมากกว่านักเรียนชาย - ระดับความสามารถสูง <li style="padding-left: 20px;">นักเรียนหญิงมีโอกาสตอบข้อสอบถูกมากกว่านักเรียนชาย
Estimated Marginal Means of score item N2		
2		<ul style="list-style-type: none"> - ระดับความสามารถต่ำ <li style="padding-left: 20px;">นักเรียนชายมีโอกาสตอบข้อสอบถูกมากกว่านักเรียนหญิง - ระดับความสามารถสูง <li style="padding-left: 20px;">นักเรียนหญิงมีโอกาสตอบข้อสอบถูกมากกว่านักเรียนชาย

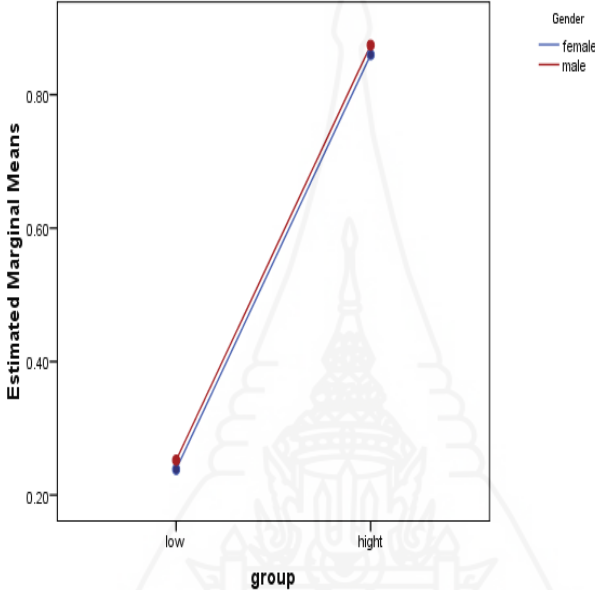
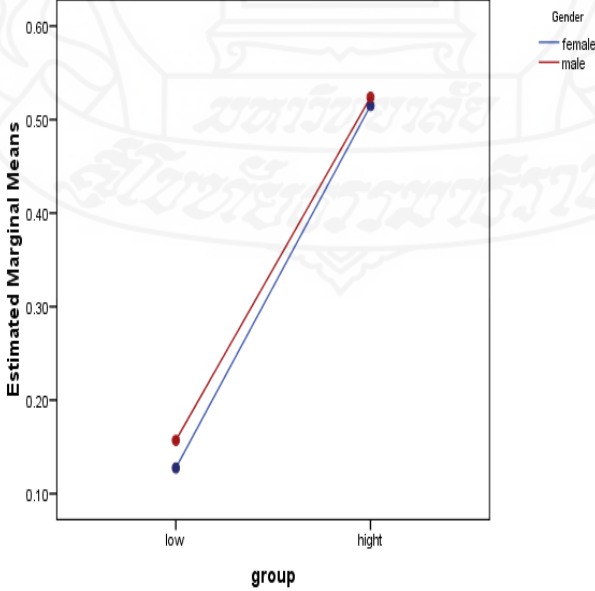
ตารางที่ 4.4 (ต่อ)

ข้อ	กราฟแสดงโอกาสในการตอบข้อสอบถูก	แปลผล									
3	<p data-bbox="584 544 935 573">Estimated Marginal Means of score item N3</p>  <table border="1" data-bbox="443 613 1038 1205"> <caption>Data for Estimated Marginal Means of score item N3</caption> <thead> <tr> <th>Group</th> <th>Female</th> <th>Male</th> </tr> </thead> <tbody> <tr> <td>low</td> <td>~0.20</td> <td>~0.20</td> </tr> <tr> <td>high</td> <td>~0.78</td> <td>~0.79</td> </tr> </tbody> </table>	Group	Female	Male	low	~0.20	~0.20	high	~0.78	~0.79	<ul style="list-style-type: none"> - ระดับความสามารถต่ำ นักเรียนชายมีโอกาสตอบข้อสอบถูกมากกว่านักเรียนหญิง - ระดับความสามารถสูง นักเรียนหญิงมีโอกาสตอบข้อสอบถูกมากกว่านักเรียนชาย
Group	Female	Male									
low	~0.20	~0.20									
high	~0.78	~0.79									
4	<p data-bbox="584 1305 935 1335">Estimated Marginal Means of score item N4</p>  <table border="1" data-bbox="443 1375 1038 1966"> <caption>Data for Estimated Marginal Means of score item N4</caption> <thead> <tr> <th>Group</th> <th>Female</th> <th>Male</th> </tr> </thead> <tbody> <tr> <td>low</td> <td>~0.17</td> <td>~0.16</td> </tr> <tr> <td>high</td> <td>~0.67</td> <td>~0.66</td> </tr> </tbody> </table>	Group	Female	Male	low	~0.17	~0.16	high	~0.67	~0.66	<ul style="list-style-type: none"> - ระดับความสามารถต่ำ นักเรียนหญิงมีโอกาสตอบข้อสอบถูกมากกว่านักเรียนชาย - ระดับความสามารถสูง นักเรียนหญิงมีโอกาสตอบข้อสอบถูกมากกว่านักเรียนชาย
Group	Female	Male									
low	~0.17	~0.16									
high	~0.67	~0.66									

ตารางที่ 4.4 (ต่อ)

ข้อ	กราฟแสดงโอกาสในการตอบข้อสอบถูก	แปลผล									
5	<p data-bbox="600 546 954 573">Estimated Marginal Means of score item N5</p>  <table border="1" data-bbox="459 618 1054 1211"> <caption>Estimated Marginal Means of score item N5</caption> <thead> <tr> <th>Group</th> <th>Female</th> <th>Male</th> </tr> </thead> <tbody> <tr> <td>low</td> <td>~0.18</td> <td>~0.19</td> </tr> <tr> <td>high</td> <td>~0.80</td> <td>~0.81</td> </tr> </tbody> </table>	Group	Female	Male	low	~0.18	~0.19	high	~0.80	~0.81	<ul style="list-style-type: none"> - ระดับความสามารถต่ำ นักเรียนชายมีโอกาสตอบข้อสอบถูกมากกว่านักเรียนหญิง - ระดับความสามารถสูง นักเรียนหญิงและนักเรียนชายมีโอกาสตอบข้อสอบถูกเท่ากัน
Group	Female	Male									
low	~0.18	~0.19									
high	~0.80	~0.81									
6	<p data-bbox="600 1308 954 1335">Estimated Marginal Means of score item N6</p>  <table border="1" data-bbox="459 1379 1054 1973"> <caption>Estimated Marginal Means of score item N6</caption> <thead> <tr> <th>Group</th> <th>Female</th> <th>Male</th> </tr> </thead> <tbody> <tr> <td>low</td> <td>~0.18</td> <td>~0.17</td> </tr> <tr> <td>high</td> <td>~0.85</td> <td>~0.84</td> </tr> </tbody> </table>	Group	Female	Male	low	~0.18	~0.17	high	~0.85	~0.84	<ul style="list-style-type: none"> - ระดับความสามารถต่ำ นักเรียนหญิงมีโอกาสตอบข้อสอบถูกมากกว่านักเรียนชาย - ระดับความสามารถสูง นักเรียนหญิงมีโอกาสตอบข้อสอบถูกมากกว่านักเรียนชาย
Group	Female	Male									
low	~0.18	~0.17									
high	~0.85	~0.84									

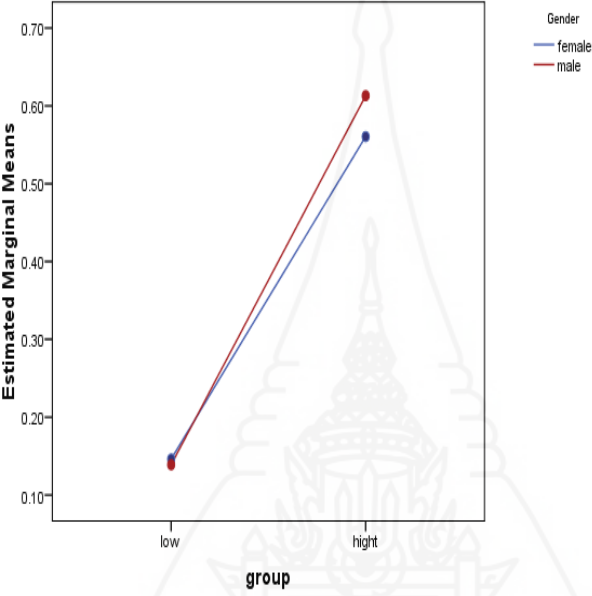
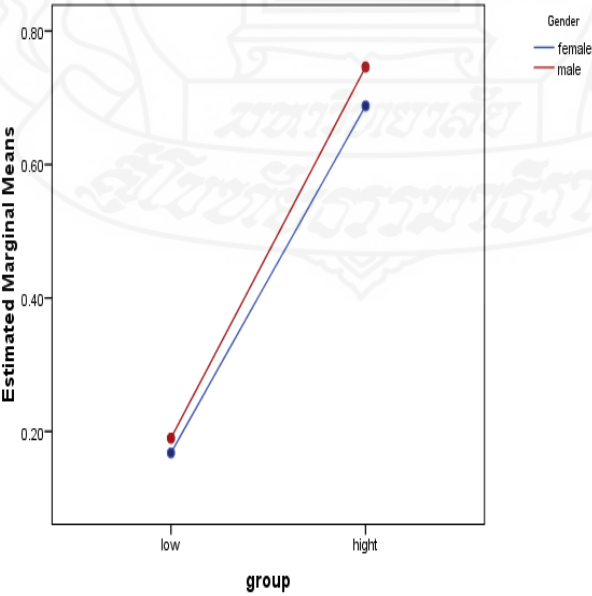
ตารางที่ 4.4 (ต่อ)

ข้อที่	กราฟแสดงโอกาสในการตอบข้อสอบถูก	แปลผล									
7	<p data-bbox="592 544 951 573">Estimated Marginal Means of score item N7</p>  <table border="1" data-bbox="453 618 1050 1205"> <caption>Data for Estimated Marginal Means of score item N7</caption> <thead> <tr> <th>Group</th> <th>Female</th> <th>Male</th> </tr> </thead> <tbody> <tr> <td>low</td> <td>~0.25</td> <td>~0.26</td> </tr> <tr> <td>high</td> <td>~0.85</td> <td>~0.87</td> </tr> </tbody> </table>	Group	Female	Male	low	~0.25	~0.26	high	~0.85	~0.87	<ul style="list-style-type: none"> - ระดับความสามารถต่ำ นักเรียนชายมีโอกาสตอบข้อสอบถูกมากกว่านักเรียนหญิง - ระดับความสามารถสูง นักเรียนชายมีโอกาสตอบข้อสอบถูกมากกว่านักเรียนหญิง
Group	Female	Male									
low	~0.25	~0.26									
high	~0.85	~0.87									
8	<p data-bbox="592 1312 951 1341">Estimated Marginal Means of score item N8</p>  <table border="1" data-bbox="453 1384 1050 1973"> <caption>Data for Estimated Marginal Means of score item N8</caption> <thead> <tr> <th>Group</th> <th>Female</th> <th>Male</th> </tr> </thead> <tbody> <tr> <td>low</td> <td>~0.13</td> <td>~0.16</td> </tr> <tr> <td>high</td> <td>~0.52</td> <td>~0.54</td> </tr> </tbody> </table>	Group	Female	Male	low	~0.13	~0.16	high	~0.52	~0.54	<ul style="list-style-type: none"> - ระดับความสามารถต่ำ นักเรียนชายมีโอกาสตอบข้อสอบถูกมากกว่านักเรียนหญิง - ระดับความสามารถสูง นักเรียนชายมีโอกาสตอบข้อสอบถูกมากกว่านักเรียนหญิง
Group	Female	Male									
low	~0.13	~0.16									
high	~0.52	~0.54									

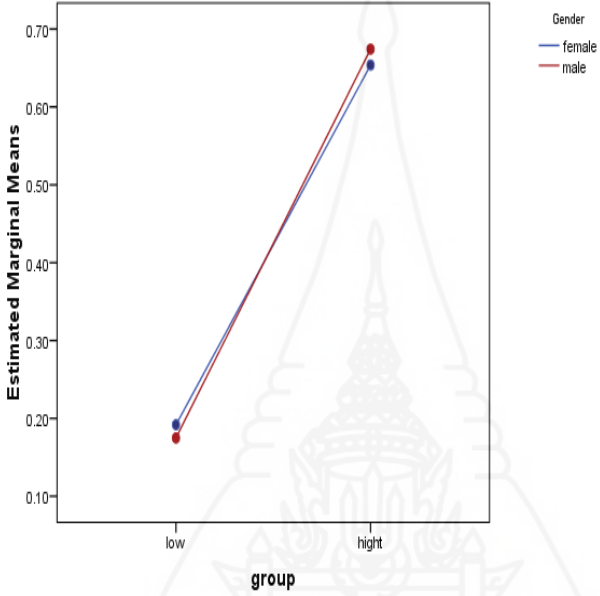
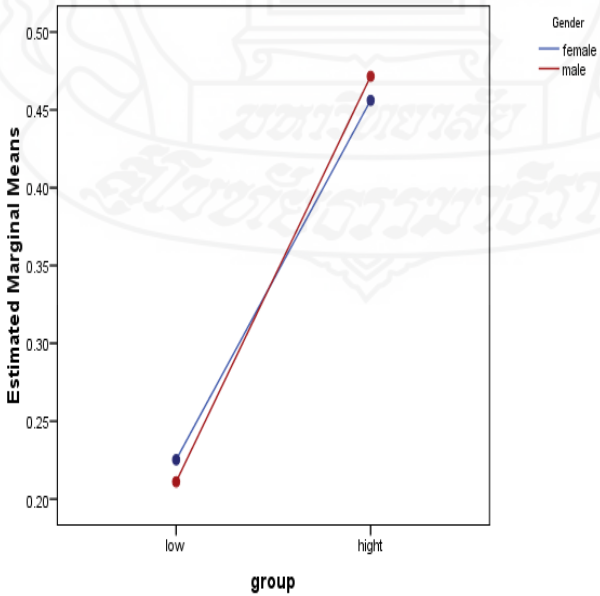
ตารางที่ 4.4 (ต่อ)

ข้อ	กราฟแสดงโอกาสในการตอบข้อสอบถูก	แปลผล									
9	<p style="text-align: center;">Estimated Marginal Means of score item N9</p> <table border="1"> <caption>Data for Item N9</caption> <thead> <tr> <th>Group</th> <th>Female</th> <th>Male</th> </tr> </thead> <tbody> <tr> <td>low</td> <td>~0.25</td> <td>~0.25</td> </tr> <tr> <td>high</td> <td>~0.85</td> <td>~0.85</td> </tr> </tbody> </table>	Group	Female	Male	low	~0.25	~0.25	high	~0.85	~0.85	<ul style="list-style-type: none"> - ระดับความสามารถต่ำ นักเรียนหญิงมีโอกาสตอบข้อสอบถูกมากกว่านักเรียนชาย - ระดับความสามารถสูง นักเรียนหญิงมีโอกาสตอบข้อสอบถูกมากกว่านักเรียนชาย
Group	Female	Male									
low	~0.25	~0.25									
high	~0.85	~0.85									
10	<p style="text-align: center;">Estimated Marginal Means of score item N10</p> <table border="1"> <caption>Data for Item N10</caption> <thead> <tr> <th>Group</th> <th>Female</th> <th>Male</th> </tr> </thead> <tbody> <tr> <td>low</td> <td>~0.18</td> <td>~0.18</td> </tr> <tr> <td>high</td> <td>~0.75</td> <td>~0.75</td> </tr> </tbody> </table>	Group	Female	Male	low	~0.18	~0.18	high	~0.75	~0.75	<ul style="list-style-type: none"> - ระดับความสามารถต่ำ นักเรียนหญิงมีโอกาสตอบข้อสอบถูกมากกว่านักเรียนชาย - ระดับความสามารถสูง นักเรียนหญิงมีโอกาสตอบข้อสอบถูกมากกว่านักเรียนชาย
Group	Female	Male									
low	~0.18	~0.18									
high	~0.75	~0.75									

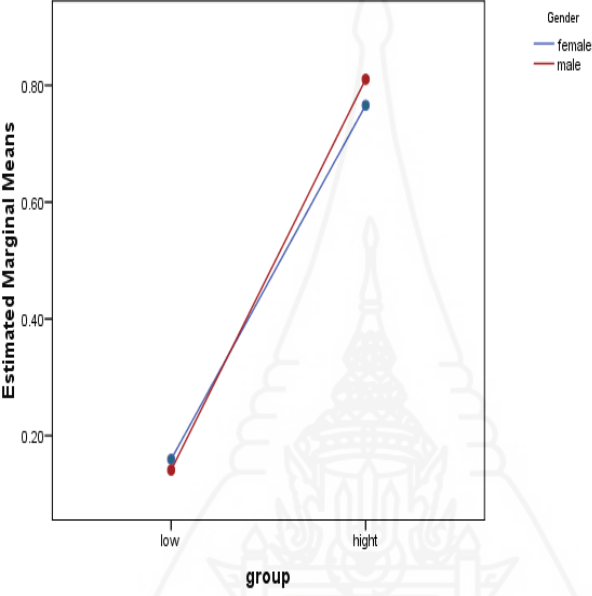
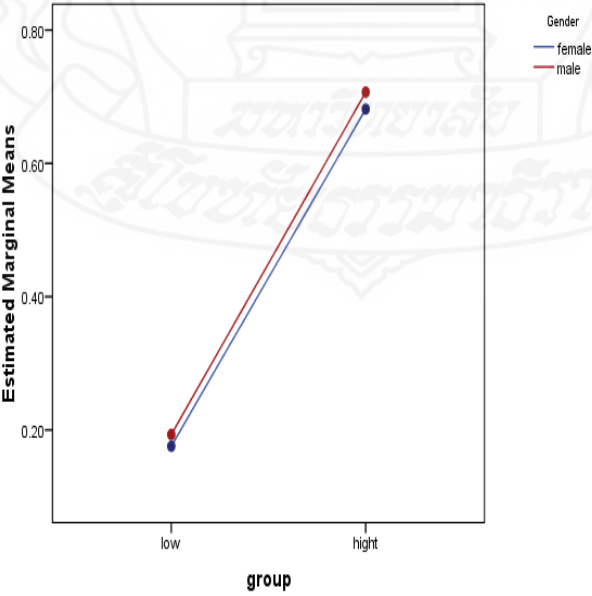
ตารางที่ 4.4 (ต่อ)

ข้อ	กราฟแสดงโอกาสในการตอบข้อสอบถูก	แปลผล									
11	<p data-bbox="563 546 927 573">Estimated Marginal Means of score item N11</p>  <table border="1" data-bbox="432 618 1026 1211"> <caption>Estimated Marginal Means of score item N11</caption> <thead> <tr> <th>Group</th> <th>Female</th> <th>Male</th> </tr> </thead> <tbody> <tr> <td>low</td> <td>0.15</td> <td>0.15</td> </tr> <tr> <td>high</td> <td>0.56</td> <td>0.62</td> </tr> </tbody> </table>	Group	Female	Male	low	0.15	0.15	high	0.56	0.62	<ul style="list-style-type: none"> - ระดับความสามารถต่ำ นักเรียนหญิงมีโอกาสตอบข้อสอบถูกมากกว่านักเรียนชาย - ระดับความสามารถสูง นักเรียนชายมีโอกาสตอบข้อสอบถูกมากกว่านักเรียนหญิง
Group	Female	Male									
low	0.15	0.15									
high	0.56	0.62									
12	<p data-bbox="563 1308 927 1335">Estimated Marginal Means of score item N12</p>  <table border="1" data-bbox="432 1379 1026 1973"> <caption>Estimated Marginal Means of score item N12</caption> <thead> <tr> <th>Group</th> <th>Female</th> <th>Male</th> </tr> </thead> <tbody> <tr> <td>low</td> <td>0.18</td> <td>0.20</td> </tr> <tr> <td>high</td> <td>0.70</td> <td>0.75</td> </tr> </tbody> </table>	Group	Female	Male	low	0.18	0.20	high	0.70	0.75	<ul style="list-style-type: none"> - ระดับความสามารถต่ำ นักเรียนชายมีโอกาสตอบข้อสอบถูกมากกว่านักเรียนหญิง - ระดับความสามารถสูง นักเรียนชายมีโอกาสตอบข้อสอบถูกมากกว่านักเรียนหญิง
Group	Female	Male									
low	0.18	0.20									
high	0.70	0.75									

ตารางที่ 4.4 (ต่อ)

ข้อ	กราฟแสดงโอกาสในการตอบข้อสอบถูก	แปลผล									
13	<p style="text-align: center;">Estimated Marginal Means of score item N13</p>  <table border="1" data-bbox="427 616 1029 1209"> <caption>Data for Estimated Marginal Means of score item N13</caption> <thead> <tr> <th>Group</th> <th>Female</th> <th>Male</th> </tr> </thead> <tbody> <tr> <td>low</td> <td>0.19</td> <td>0.18</td> </tr> <tr> <td>high</td> <td>0.66</td> <td>0.68</td> </tr> </tbody> </table>	Group	Female	Male	low	0.19	0.18	high	0.66	0.68	<ul style="list-style-type: none"> - ระดับความสามารถต่ำ นักเรียนหญิงมีโอกาสตอบข้อสอบถูกมากกว่านักเรียนชาย - ระดับความสามารถสูง นักเรียนชายมีโอกาสตอบข้อสอบถูกมากกว่านักเรียนหญิง
Group	Female	Male									
low	0.19	0.18									
high	0.66	0.68									
14	<p style="text-align: center;">Estimated Marginal Means of score item N14</p>  <table border="1" data-bbox="427 1377 1029 1971"> <caption>Data for Estimated Marginal Means of score item N14</caption> <thead> <tr> <th>Group</th> <th>Female</th> <th>Male</th> </tr> </thead> <tbody> <tr> <td>low</td> <td>0.23</td> <td>0.21</td> </tr> <tr> <td>high</td> <td>0.46</td> <td>0.47</td> </tr> </tbody> </table>	Group	Female	Male	low	0.23	0.21	high	0.46	0.47	<ul style="list-style-type: none"> - ระดับความสามารถต่ำ นักเรียนหญิงมีโอกาสตอบข้อสอบถูกมากกว่านักเรียนชาย - ระดับความสามารถสูง นักเรียนชายมีโอกาสตอบข้อสอบถูกมากกว่านักเรียนหญิง
Group	Female	Male									
low	0.23	0.21									
high	0.46	0.47									

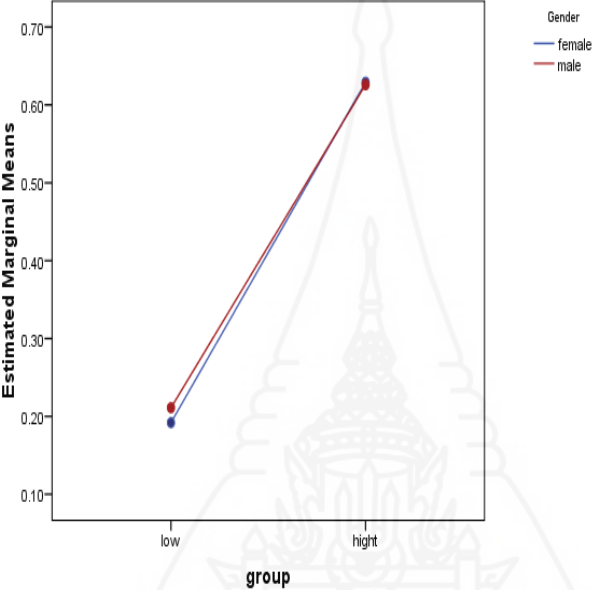
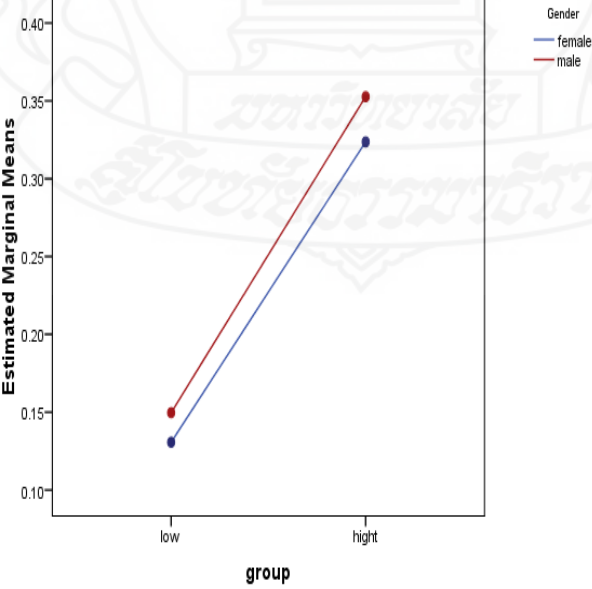
ตารางที่ 4.4 (ต่อ)

ข้อ	กราฟแสดงโอกาสในการตอบข้อสอบถูก	แปลผล									
15	<p data-bbox="563 546 927 573">Estimated Marginal Means of score item N15</p>  <table border="1" data-bbox="432 618 1026 1211"> <caption>Estimated Marginal Means of score item N15</caption> <thead> <tr> <th>Group</th> <th>Female</th> <th>Male</th> </tr> </thead> <tbody> <tr> <td>low</td> <td>0.15</td> <td>0.15</td> </tr> <tr> <td>high</td> <td>0.75</td> <td>0.80</td> </tr> </tbody> </table>	Group	Female	Male	low	0.15	0.15	high	0.75	0.80	<ul style="list-style-type: none"> - ระดับความสามารถต่ำ นักเรียนหญิงมีโอกาสตอบข้อสอบถูกมากกว่านักเรียนชาย - ระดับความสามารถสูง นักเรียนชายมีโอกาสตอบข้อสอบถูกมากกว่านักเรียนหญิง
Group	Female	Male									
low	0.15	0.15									
high	0.75	0.80									
16	<p data-bbox="563 1308 927 1335">Estimated Marginal Means of score item N16</p>  <table border="1" data-bbox="432 1379 1026 1973"> <caption>Estimated Marginal Means of score item N16</caption> <thead> <tr> <th>Group</th> <th>Female</th> <th>Male</th> </tr> </thead> <tbody> <tr> <td>low</td> <td>0.18</td> <td>0.18</td> </tr> <tr> <td>high</td> <td>0.65</td> <td>0.70</td> </tr> </tbody> </table>	Group	Female	Male	low	0.18	0.18	high	0.65	0.70	<ul style="list-style-type: none"> - ระดับความสามารถต่ำ นักเรียนชายมีโอกาสตอบข้อสอบถูกมากกว่านักเรียนหญิง - ระดับความสามารถสูง นักเรียนชายมีโอกาสตอบข้อสอบถูกมากกว่านักเรียนหญิง
Group	Female	Male									
low	0.18	0.18									
high	0.65	0.70									

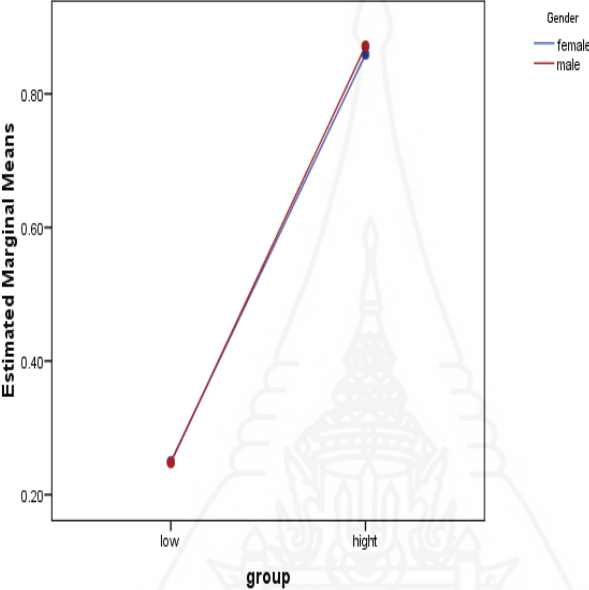
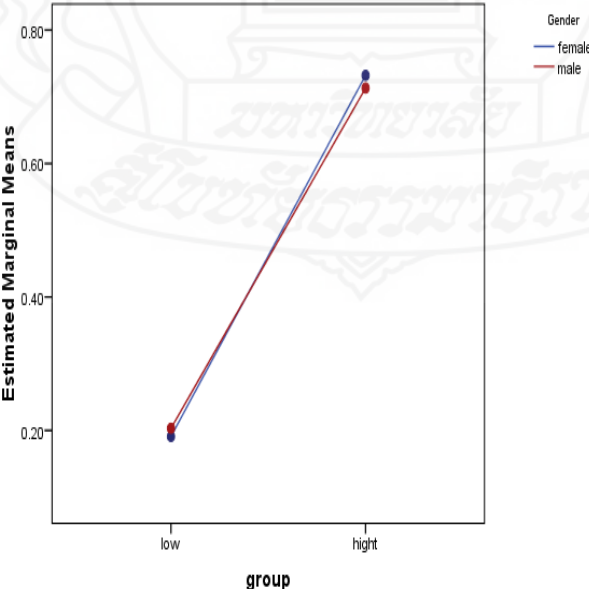
ตารางที่ 4.4 (ต่อ)

ข้อ	กราฟแสดงโอกาสในการตอบข้อสอบถูก	แปลผล									
17	<p style="text-align: center;">Estimated Marginal Means of score item N17</p> <table border="1"> <caption>Data for Estimated Marginal Means of score item N17</caption> <thead> <tr> <th>Group</th> <th>Female</th> <th>Male</th> </tr> </thead> <tbody> <tr> <td>low</td> <td>0.42</td> <td>0.38</td> </tr> <tr> <td>high</td> <td>0.92</td> <td>0.92</td> </tr> </tbody> </table>	Group	Female	Male	low	0.42	0.38	high	0.92	0.92	<ul style="list-style-type: none"> - ระดับความสามารถต่ำ นักเรียนหญิงมีโอกาสตอบข้อสอบถูกมากกว่านักเรียนชาย - ระดับความสามารถสูง นักเรียนหญิงมีโอกาสตอบข้อสอบถูกมากกว่านักเรียนชาย
Group	Female	Male									
low	0.42	0.38									
high	0.92	0.92									
18	<p style="text-align: center;">Estimated Marginal Means of score item N18</p> <table border="1"> <caption>Data for Estimated Marginal Means of score item N18</caption> <thead> <tr> <th>Group</th> <th>Female</th> <th>Male</th> </tr> </thead> <tbody> <tr> <td>low</td> <td>0.28</td> <td>0.28</td> </tr> <tr> <td>high</td> <td>0.62</td> <td>0.62</td> </tr> </tbody> </table>	Group	Female	Male	low	0.28	0.28	high	0.62	0.62	<ul style="list-style-type: none"> - ระดับความสามารถต่ำ นักเรียนชายและนักเรียนหญิงมีโอกาสตอบข้อสอบถูกเท่ากัน - ระดับความสามารถสูง นักเรียนชายมีโอกาสตอบข้อสอบถูกมากกว่านักเรียนหญิง
Group	Female	Male									
low	0.28	0.28									
high	0.62	0.62									

ตารางที่ 4.4 (ต่อ)

ข้อ	กราฟแสดงโอกาสในการตอบข้อสอบถูก	แปลผล									
19	<p data-bbox="563 544 927 573">Estimated Marginal Means of score item N19</p>  <table border="1" data-bbox="432 618 1026 1205"> <caption>Estimated Marginal Means of score item N19</caption> <thead> <tr> <th>Group</th> <th>Female</th> <th>Male</th> </tr> </thead> <tbody> <tr> <td>low</td> <td>~0.19</td> <td>~0.21</td> </tr> <tr> <td>high</td> <td>~0.63</td> <td>~0.63</td> </tr> </tbody> </table>	Group	Female	Male	low	~0.19	~0.21	high	~0.63	~0.63	<ul style="list-style-type: none"> - ระดับความสามารถต่ำ นักเรียนชายมีโอกาสตอบข้อสอบถูกมากกว่านักเรียนหญิง - ระดับความสามารถสูง นักเรียนหญิงมีโอกาสตอบข้อสอบถูกมากกว่านักเรียนชาย
Group	Female	Male									
low	~0.19	~0.21									
high	~0.63	~0.63									
20	<p data-bbox="563 1312 927 1341">Estimated Marginal Means of score item N20</p>  <table border="1" data-bbox="432 1386 1026 1973"> <caption>Estimated Marginal Means of score item N20</caption> <thead> <tr> <th>Group</th> <th>Female</th> <th>Male</th> </tr> </thead> <tbody> <tr> <td>low</td> <td>~0.13</td> <td>~0.15</td> </tr> <tr> <td>high</td> <td>~0.33</td> <td>~0.36</td> </tr> </tbody> </table>	Group	Female	Male	low	~0.13	~0.15	high	~0.33	~0.36	<ul style="list-style-type: none"> - ระดับความสามารถต่ำ นักเรียนชายมีโอกาสตอบข้อสอบถูกมากกว่านักเรียนหญิง - ระดับความสามารถสูง นักเรียนชายมีโอกาสตอบข้อสอบถูกมากกว่านักเรียนหญิง
Group	Female	Male									
low	~0.13	~0.15									
high	~0.33	~0.36									

ตารางที่ 4.4 (ต่อ)

ข้อ	กราฟแสดงโอกาสในการตอบข้อสอบถูก	แปลผล									
21	<p data-bbox="563 544 922 577">Estimated Marginal Means of score item N21</p>  <table border="1" data-bbox="432 618 1023 1205"> <caption>Estimated Marginal Means of score item N21</caption> <thead> <tr> <th>Group</th> <th>Female</th> <th>Male</th> </tr> </thead> <tbody> <tr> <td>low</td> <td>~0.25</td> <td>~0.25</td> </tr> <tr> <td>high</td> <td>~0.85</td> <td>~0.85</td> </tr> </tbody> </table>	Group	Female	Male	low	~0.25	~0.25	high	~0.85	~0.85	<ul style="list-style-type: none"> - ระดับความสามารถต่ำ นักเรียนหญิงมีโอกาสตอบข้อสอบถูกมากกว่านักเรียนชาย - ระดับความสามารถสูง นักเรียนชายมีโอกาสตอบข้อสอบถูกมากกว่านักเรียนหญิง
Group	Female	Male									
low	~0.25	~0.25									
high	~0.85	~0.85									
22	<p data-bbox="563 1305 922 1339">Estimated Marginal Means of score item N22</p>  <table border="1" data-bbox="432 1379 1023 1966"> <caption>Estimated Marginal Means of score item N22</caption> <thead> <tr> <th>Group</th> <th>Female</th> <th>Male</th> </tr> </thead> <tbody> <tr> <td>low</td> <td>~0.20</td> <td>~0.20</td> </tr> <tr> <td>high</td> <td>~0.75</td> <td>~0.75</td> </tr> </tbody> </table>	Group	Female	Male	low	~0.20	~0.20	high	~0.75	~0.75	<ul style="list-style-type: none"> - ระดับความสามารถต่ำ นักเรียนชายมีโอกาสตอบข้อสอบถูกมากกว่านักเรียนหญิง - ระดับความสามารถสูง นักเรียนหญิงมีโอกาสตอบข้อสอบถูกมากกว่านักเรียนชาย
Group	Female	Male									
low	~0.20	~0.20									
high	~0.75	~0.75									

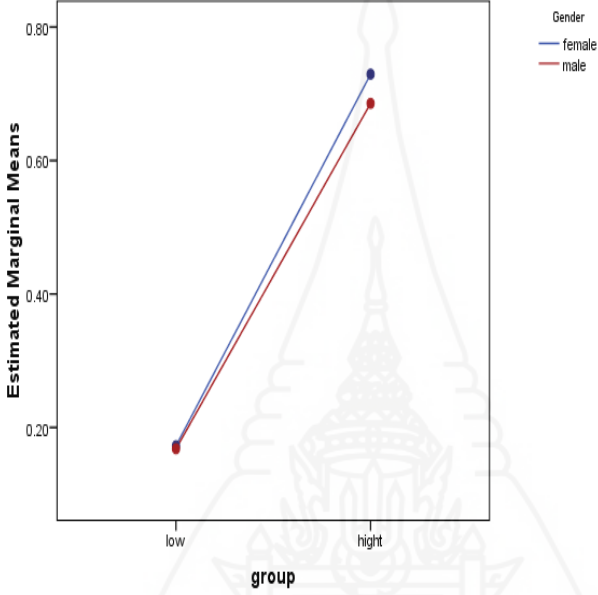
ตารางที่ 4.4 (ต่อ)

ข้อ	กราฟแสดงโอกาสในการตอบข้อสอบถูก	แปลผล									
23	<p style="text-align: center;">Estimated Marginal Means of score item N23</p> <table border="1"> <caption>Data for Item N23</caption> <thead> <tr> <th>Group</th> <th>Female</th> <th>Male</th> </tr> </thead> <tbody> <tr> <td>low</td> <td>0.21</td> <td>0.22</td> </tr> <tr> <td>high</td> <td>0.48</td> <td>0.47</td> </tr> </tbody> </table>	Group	Female	Male	low	0.21	0.22	high	0.48	0.47	<ul style="list-style-type: none"> - ระดับความสามารถต่ำ นักเรียนชายมีโอกาสตอบข้อสอบถูกมากกว่านักเรียนหญิง - ระดับความสามารถสูง นักเรียนหญิงมีโอกาสตอบข้อสอบถูกมากกว่านักเรียนชาย
Group	Female	Male									
low	0.21	0.22									
high	0.48	0.47									
24	<p style="text-align: center;">Estimated Marginal Means of score item N24</p> <table border="1"> <caption>Data for Item N24</caption> <thead> <tr> <th>Group</th> <th>Female</th> <th>Male</th> </tr> </thead> <tbody> <tr> <td>low</td> <td>0.46</td> <td>0.42</td> </tr> <tr> <td>high</td> <td>0.72</td> <td>0.71</td> </tr> </tbody> </table>	Group	Female	Male	low	0.46	0.42	high	0.72	0.71	<ul style="list-style-type: none"> - ระดับความสามารถต่ำ นักเรียนหญิงมีโอกาสตอบข้อสอบถูกมากกว่านักเรียนชาย - ระดับความสามารถสูง นักเรียนหญิงมีโอกาสตอบข้อสอบถูกมากกว่านักเรียนชาย
Group	Female	Male									
low	0.46	0.42									
high	0.72	0.71									

ตารางที่ 4.4 (ต่อ)

ข้อ	กราฟแสดงโอกาสในการตอบข้อสอบถูก	แปลผล
25	<p style="text-align: center;">Estimated Marginal Means of score item N25</p>	<ul style="list-style-type: none"> - ระดับความสามารถต่ำ นักเรียนหญิงมีโอกาสตอบข้อสอบถูกมากกว่านักเรียนชาย - ระดับความสามารถสูง นักเรียนหญิงมีโอกาสตอบข้อสอบถูกมากกว่านักเรียนชาย
26	<p style="text-align: center;">Estimated Marginal Means of score item N26</p>	<ul style="list-style-type: none"> - ระดับความสามารถต่ำ นักเรียนชายมีโอกาสตอบข้อสอบถูกมากกว่านักเรียนหญิง - ระดับความสามารถสูง นักเรียนหญิงมีโอกาสตอบข้อสอบถูกมากกว่านักเรียนชาย

ตารางที่ 4.4 (ต่อ)

ข้อ	กราฟแสดงโอกาสในการตอบข้อสอบถูก	แปลผล									
27	<p style="text-align: center;">Estimated Marginal Means of score item N27</p>  <table border="1" data-bbox="427 618 1027 1211"> <caption>Data for Estimated Marginal Means of score item N27</caption> <thead> <tr> <th>Group</th> <th>Female (Mean)</th> <th>Male (Mean)</th> </tr> </thead> <tbody> <tr> <td>low</td> <td>0.18</td> <td>0.18</td> </tr> <tr> <td>high</td> <td>0.75</td> <td>0.70</td> </tr> </tbody> </table>	Group	Female (Mean)	Male (Mean)	low	0.18	0.18	high	0.75	0.70	<p>- ระดับความสามารถต่ำ</p> <p>นักเรียนหญิงมีโอกาสตอบข้อสอบถูกมากกว่านักเรียนชาย</p> <p>- ระดับความสามารถสูง</p> <p>นักเรียนหญิงมีโอกาสตอบข้อสอบถูกมากกว่านักเรียนชาย</p>
Group	Female (Mean)	Male (Mean)									
low	0.18	0.18									
high	0.75	0.70									

จากตารางที่ 4.4 ผลการตอบข้อสอบทั้ง 27 ข้อพบว่าในระดับความสามารถต่ำนักเรียนหญิงมีโอกาสตอบข้อสอบถูก 14 ข้อ คิดเป็นร้อยละ 51.85 และนักเรียนชายมีโอกาสตอบถูก 12 ข้อ คิดเป็นร้อยละ 44.44 ในระดับความสามารถสูง นักเรียนหญิงมีโอกาสตอบข้อสอบถูกจำนวน 16 ข้อคิดเป็นร้อยละ 59.26 และนักเรียนชายมีโอกาสตอบถูก 11 ข้อคิดเป็นร้อยละ 40.74

ผลจากการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบพบว่าข้อสอบทั้ง 27 ข้อ ทำหน้าที่ต่างกัน ซึ่งทราบเพียงว่าข้อสอบเกิดการทำหน้าที่ต่างกัน แต่ไม่ทราบสารสนเทศอื่น ซึ่งถือว่าข้อมูลยังไม่เพียงพอสำหรับการตัดสินใจข้อสอบว่าควรปรับปรุงแก้ไข หรือคงข้อสอบข้อนั้นไว้ในแบบสอบเพื่อเก็บเข้าคลังข้อสอบต่อไป ผู้วิจัยจึงวัดขนาดอิทธิพลในข้อสอบที่ทำหน้าที่ต่างกัน ตามเกณฑ์ Zumbo and Thomas เพื่อใช้ในการพิจารณา ร่วมในการตัดสินใจ ซึ่งจะทำให้เกิดความรอบคอบต่อการตัดสินใจข้อสอบ

การวัดขนาดอิทธิพล ตามเกณฑ์ Zombo และ Thomas (1997)

การตรวจสอบข้อสอบการที่ทำหน้าต่างกันโดยการวัดขนาดอิทธิพล ตามเกณฑ์ Zombo and Thomas ข้อสอบวัดความสามารถด้านการคำนวณ 27 ข้อ ดัง ตารางที่ 4.5

ตารางที่ 4.5 ผลการการวัดขนาดอิทธิพลในข้อสอบวัดความสามารถด้านการคำนวณที่ทำหน้าที่ต่างกัน ตามเกณฑ์ Zumbo and Thomas

ข้อ	ขนาดอิทธิพล		ข้อ	ขนาดอิทธิพล	
	ΔR^2	ผลขนาดอิทธิพล		ΔR^2	ผลขนาดอิทธิพล
1	.003*	ขนาดเล็ก	15	.001*	ขนาดเล็ก
2	.000*	ขนาดเล็ก	16	.003*	ขนาดเล็ก
3	.000*	ขนาดเล็ก	17	.002*	ขนาดเล็ก
4	.000*	ขนาดเล็ก	18	.000*	ขนาดเล็ก
5	.000*	ขนาดเล็ก	19	.000*	ขนาดเล็ก
6	.001*	ขนาดเล็ก	20	.003*	ขนาดเล็ก
7	.000*	ขนาดเล็ก	21	.001*	ขนาดเล็ก
8	.005*	ขนาดเล็ก	22	.000*	ขนาดเล็ก
9	.002*	ขนาดเล็ก	23	.000*	ขนาดเล็ก
10	.000*	ขนาดเล็ก	24	.002*	ขนาดเล็ก
11	.001*	ขนาดเล็ก	25	.001*	ขนาดเล็ก
12	.004*	ขนาดเล็ก	26	.001*	ขนาดเล็ก
13	.000*	ขนาดเล็ก	27	.001*	ขนาดเล็ก
14	.000*	ขนาดเล็ก			

Effect size (Nagelkerke's R^2) : Zombo and Thomas (ZT) : * $\Delta R^2 < .13$ negligible effect , ** $.13 \leq \Delta R^2 \leq .26$ moderate effect, *** $\Delta R^2 > .26$ large effect

จากตารางที่ 4.5 ผลการวัดขนาดอิทธิพลตามเกณฑ์ Zumbo and tomas (1997) ของข้อสอบวัดความสามารถด้านการคำนวณที่ทำหน้าที่ต่างกัน จำนวน 27 ข้อ พบว่า ขนาดของการทำหน้าที่ต่างกันขนาดเล็กมากๆ ($\Delta R^2 < 0.13$) ทั้ง 27 ข้อ

ตอนที่ 3 วิเคราะห์สาเหตุของข้อสอบวัดความสามารถด้านการคำนวณของการทดสอบ ระดับชาติที่ทำหน้าที่ต่างกัน

จากผลการตรวจสอบการทำหน้าที่ต่างกันของแบบทดสอบวัดความสามารถด้าน
คำนวณของการทดสอบระดับชาติพบว่าข้อสอบทั้ง 27 ข้อ ทำหน้าที่ต่างกันซึ่งมีขนาดการทำหน้าที่
ต่างกัน 0.000-0.005 ซึ่งมีขนาดเล็กมาก จึงไม่สามารถมองเห็นสาเหตุการทำหน้าที่ต่างกันทั้ง 27 ข้อ
ได้ชัดเจน และเมื่อพิจารณาผลการวิเคราะห์การทำหน้าที่ต่างกันทั้ง 27 ข้อ พบว่านักเรียนชายมี
โอกาสตอบข้อสอบถูกมากกว่านักเรียนหญิง 16 ข้อ คิดเป็นร้อยละ 59.26 และนักเรียนหญิงมีโอกา
สตอบข้อสอบถูกมากกว่านักเรียนชาย 11 ข้อ คิดเป็นร้อยละ 40.74 ซึ่งข้อสอบวัดความสามารถด้าน
การคำนวณฉบับนี้นักเรียนชายจะมีโอกาสในการตอบข้อสอบถูกมากกว่านักเรียนหญิง

เนื่องจากเพศชายจะ มีความสนใจในเรื่องที่เกี่ยวกับ คณิตศาสตร์ วิทยาศาสตร์ การ
ทดลอง เครื่องยนต์ การผจญภัย กีฬา การต่อสู้ ดังนั้นการออกข้อสอบที่อยู่ในความสนใจของเพศ
ชาย เพศชายจะมีโอกาสในการตอบข้อสอบเหล่านั้นถูกมากกว่าเพศหญิง



บทที่ 5

สรุปผล อภิปรายผล และข้อเสนอแนะ

การวิจัยครั้งนี้เป็นการวิจัยเชิงสำรวจมีวัตถุประสงค์เพื่อศึกษาการทำหน้าที่ต่างกันของข้อสอบวัดความสามารถด้านการคำนวณของการทดสอบระดับชาติชั้นประถมศึกษาปีที่ 3 สังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน ปีการศึกษา 2557 และเพื่อวิเคราะห์สาเหตุของข้อสอบวัดความสามารถด้านการคำนวณของการทดสอบระดับชาติ ที่ทำหน้าที่ต่างกัน

1. เครื่องมือและข้อมูลที่ใช้ในการวิจัย

ผลการสอบวัดความสามารถด้านการคำนวณ ชั้นประถมศึกษาปีที่ 3 ของสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน ปีการศึกษา 2557 จำนวน 27 ข้อ ซึ่งมีผู้เข้าสอบ จำนวน 464,532 คน เป็นเพศชาย 224,615 คน เป็นเพศหญิง 239,917 คน

2. การวิเคราะห์ข้อมูล

2.1 หาค่าสถิติพื้นฐาน คือค่าเฉลี่ย (Mean) ส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation) ค่าความเบ้ (Skewness) ค่าความโด่ง (Kurtosis) เพื่อบรรยายการแจกแจงของคะแนนของแบบทดสอบ และหาคุณภาพรายข้อของข้อสอบ คือ ค่าความยาก ค่าอำนาจจำแนก ค่าความเที่ยง (Reliability)

2.2 นำแบบทดสอบวัดความสามารถด้านคำนวณ 27 ข้อ วิเคราะห์การทำหน้าที่ต่างกันของข้อสอบเป็นรายข้อด้วยวิธีถดถอยโลจิสติก (LR) วิเคราะห์ค่าสถิติพื้นฐาน หาค่าดัชนีการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีถดถอยโลจิสติกโดยใช้โปรแกรมคอมพิวเตอร์สำเร็จรูป

2.3 ตรวจสอบโอกาสตอบข้อสอบถูก

ขั้นที่ 1 แบ่งระดับความสามารถผู้สอบ

ระดับความสามารถต่ำ นักเรียนที่มีคะแนนตั้งแต่ 0 ถึง (ค่าเฉลี่ยของคะแนนสอบ - ส่วนเบี่ยงเบนมาตรฐาน) ระดับความสามารถสูง นักเรียนที่มีคะแนนตั้งแต่ (ค่าเฉลี่ยของคะแนนสอบ + ส่วนเบี่ยงเบนมาตรฐาน) ถึง 27

ขั้นที่ 2 เปรียบเทียบคะแนนเฉลี่ยแต่ละข้อระหว่างนักเรียนชายและนักเรียนหญิง ในกรณีกลุ่มใดกลุ่มหนึ่งมีคะแนนเฉลี่ยในข้อนั้นสูงกว่า แสดงว่านักเรียนกลุ่มนั้นมีโอกาสตอบข้อสอบข้อนั้นถูกมากกว่าอีกกลุ่มหนึ่ง

2.4 ตรวจสอบขนาดอิทธิพลการทำหน้าที่ต่างกันของข้อสอบ

นำข้อสอบที่ทำหน้าที่ต่างกันมาตรวจสอบขนาดของอิทธิพล (Effect size) ที่ได้จะเป็น ΔR^2 (มาจากผลต่างของ R^2 โมเดล 1 กับ โมเดล 1 กับ โมเดล 3) โดยใช้เกณฑ์ซัมโบและโทมัส เพื่อใช้ในการจำแนกขนาดการทำหน้าที่ต่างกันของข้อสอบ เกณฑ์สำหรับดัชนี ΔR^2 มีดังนี้

$\Delta R^2 < 0.13$	แสดงว่า DIF มีขนาดเล็ก
$0.13 \leq \Delta R^2 \leq 0.26$	แสดงว่า DIF มีขนาดปานกลาง
$\Delta R^2 > 0.26$	แสดงว่า DIF มีขนาดปานใหญ่

1. สรุปผลการวิจัย

1.1 ผลการตรวจสอบการทำหน้าที่ต่างกันของแบบทดสอบวัดความรู้ด้านคำนวณ 27 ข้อ

1.1.1 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวัดความสามารถด้านการคำนวณของการทดสอบระดับชาติ จำนวน 27 ข้อ จำแนกตามกลุ่มเพศ มีข้อสอบที่ทำหน้าที่ต่างกัน 27 ข้อ คิดเป็นร้อยละ 100 มีค่าดัชนีการทำหน้าที่ต่างกันระหว่าง -0.241 ถึง 0.268 ซึ่งพบข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูป จำนวน 5 ข้อ คิดเป็นร้อยละ 18.52 และพบข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป จำนวน 22 ข้อ คิดเป็นร้อยละ 81.48

1.1.2 โอกาสตอบข้อสอบถูกของแบบทดสอบวัดความรู้ด้านคำนวณ 27 ข้อ พบว่าในระดับความสามารถต่ำนักเรียนหญิงมีโอกาสตอบข้อสอบถูก 14 ข้อ และนักเรียนชายมีโอกาสตอบถูก 12 ข้อ ในระดับความสามารถสูง นักเรียนหญิงมีโอกาสตอบข้อสอบถูกจำนวน 16 ข้อ และนักเรียนชายมีโอกาสตอบถูก 11 ข้อ

1.1.3 ผลการตรวจสอบโดยการวัดขนาดอิทธิพลของข้อสอบวัดความสามารถด้านการคำนวณของการทดสอบระดับชาติที่ทำหน้าที่ต่างกัน ตามเกณฑ์ Zumbo and Thomas พบว่าข้อสอบที่ทำหน้าที่ต่างกันมีขนาดอิทธิพลขนาด 0.000-0.005 ซึ่งมีขนาดเล็กมากๆ ($\Delta R^2 < 0.13$) ทั้ง 27 ข้อ คิดเป็นร้อยละ 100

1.2 สาเหตุของข้อสอบวัดความสามารถด้านการคำนวณของการทดสอบระดับชาติ ที่ทำหน้าที่ต่างกัน

จากการตรวจสอบการทำหน้าที่ต่างกันของแบบทดสอบวัดความสามารถด้านการคำนวณของการทดสอบระดับชาติ พบว่านักเรียนชายมีโอกาสตอบข้อสอบถูกมากกว่านักเรียนหญิง 16 ข้อ คิดเป็นร้อยละ 59.26 และนักเรียนหญิงมีโอกาสตอบข้อสอบถูกมากกว่านักเรียนชาย 11 ข้อ

คิดเป็นร้อยละ 40.74 ซึ่งข้อสอบวัดความสามารถด้านการคำนวณฉบับนี้นักเรียนชายจะมีโอกาสในการตอบข้อสอบถูกมากกว่านักเรียนหญิง

เนื่องจากเพศชายจะ มีความสนใจเนื้อเรื่องที่เกี่ยวข้องกับ คณิตศาสตร์ วิทยาศาสตร์ การทดลอง เครื่องยนต์ การผจญภัย กีฬา การต่อสู้ ดังนั้นการออกข้อสอบที่อยู่ในความสนใจของเพศชาย เพศชายจะมีโอกาสในการตอบข้อสอบเหล่านั้นมากกว่าเพศหญิง แต่เมื่อพิจารณาขนาดการทำหน้าที่ต่างกันพบว่า มีขนาดเล็กมากแทบมองไม่เห็นจึงแสดงให้เห็นว่าข้อสอบวัดความสามารถด้านคำนวณของการทดสอบระดับชาติฉบับนี้มีความยุติธรรมของข้อสอบและแบบทดสอบ

2. อภิปรายผล

จากการตรวจสอบการทำหน้าที่ต่างของข้อสอบวัดความสามารถด้านคำนวณของการทดสอบระดับชาติด้วยวิธีถดถอยโลจิสติก พบว่า เพศเป็นตัวแปรที่มีผลต่อการทำหน้าที่ต่างกันของข้อสอบ ซึ่งข้อสอบที่ทำหน้าที่ต่างกันส่วนใหญ่ นักเรียนชายจะได้เปรียบในการตอบข้อสอบสอดคล้องกับงานวิจัยของอมลณัฐ อุบลรัตน์ (2552) ที่ศึกษาการทำหน้าที่ต่างกันของข้อสอบวัดผลสัมฤทธิ์ทางการเรียนชั้นมัธยมศึกษาปีที่ 2 สำนักงานเขตพื้นที่การศึกษาปัตตานีเขต 2 ด้วยวิธีโลจิสติกและวิธีแมนเทล – ฮานส์เซล เมื่อจำแนกตามกลุ่มเพศ พบว่า ข้อสอบวิชาคณิตศาสตร์ที่ทำหน้าที่ต่างกันส่วนใหญ่เพศชายจะได้เปรียบ เนื่องจากเพศชายมีความสนใจวิชาด้านทักษะคำนวณมากกว่าเพศหญิง และยังสอดคล้องกับงานวิจัยของ สุรชาติพิทย์ ตรีสิน และปิยะทิพย์ ประจุพรหม (2560) ที่ศึกษาการเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบระดับชาติชั้นประถมศึกษาปีที่ 3 ด้วยวิธี HGLM วิธี MIMIC และวิธี IRT –LR พบว่าเพศส่งผลให้เกิดการทำหน้าที่ต่างกันของข้อสอบ โดยเพศชายจะได้เปรียบในการตอบข้อสอบด้านคำนวณ

การตรวจสอบการทำหน้าที่ต่างของข้อสอบวัดความสามารถด้านคำนวณของการทดสอบระดับชาติด้วยวิธีถดถอยโลจิสติก โดยพิจารณาค่านัยสำคัญทางสถิติพบข้อสอบทำหน้าที่ต่างกันทั้ง 27 ข้อ ๓ เกียรติกรม ทองงอก (2554) ที่ศึกษาประสิทธิภาพการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในวิธีถดถอยโลจิสติกโดยใช้เกณฑ์ขนาดอิทธิพล 2 วิธีสำหรับข้อสอบที่มีรูปแบบการตรวจให้คะแนนแบบทวิภาค ผลการวิจัยสนับสนุนให้พิจารณาขนาดอิทธิพลของการทำหน้าที่ต่างกันของข้อสอบร่วมกับการตรวจสอบโดยวิธีการหลักเลือกศึกษา ซึ่งข้อสอบที่ต้องดำเนินการปรับปรุงหรือแก้ไขพิจารณาจากขนาดอิทธิพลที่มีขนาดปานกลาง และขนาดใหญ่ เมื่อนำผลการตรวจสอบการทำหน้าที่ต่างของข้อสอบวัดความสามารถด้านคำนวณของการทดสอบระดับชาติด้วยวิธีถดถอยโลจิสติกมาพิจารณาร่วมกับการตรวจสอบขนาดอิทธิพลหรือขนาดการทำ

หน้าที่ต่างกันของข้อสอบวัดความสามารถด้านการคำนวณของการทดสอบระดับชาติพบว่าการทำหน้าที่ต่างกันของข้อสอบมีขนาดตั้งแต่ 0.00-0.01 ซึ่งมีขนาดเล็กมากทั้ง 27 ข้อ แสดงให้เห็นว่าข้อสอบวัดความสามารถด้านการคำนวณฉบับนี้มีคุณภาพ และมีความเหมาะสมสำหรับการนำไปใช้ในการทดสอบ สอดคล้องกับ Gomez-Benito, Hidalgo and Padilla, (2009) ที่ศึกษาประสิทธิภาพของขนาดอิทธิพลในการพัฒนาวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีลดรอยโลจิสติก พบว่า การวัดขนาดอิทธิพลโดยสถิติ R^2 ร่วมกับการทดสอบนัยสำคัญจะได้ค่าลดลงจนเกือบจะเป็นศูนย์ของการสรุปผิดว่าข้อสอบทำหน้าที่ต่างกันทั้งที่ความเป็นจริงข้อสอบไม่ได้ทำหน้าที่ต่างกัน

ผลจากการตรวจสอบการทำหน้าที่ต่างของข้อสอบวัดความสามารถด้านคำนวณของการทดสอบระดับชาติแสดงให้เห็นว่าข้อสอบวัดความสามารถด้านการคำนวณฉบับนี้มีคุณภาพด้านความตรงในประเด็นด้านความยุติธรรมต่อผู้สอบที่มีเพศแตกต่างกันข้อสอบวัดความสามารถด้านการคำนวณของการทดสอบระดับชาติ ชั้นประถมศึกษาปีที่ 3 ปีการศึกษา 2557 มีความเหมาะสมสำหรับการนำไปใช้ในการทดสอบ ซึ่งแบบทดสอบสามารถเก็บเข้าคลังข้อสอบหรือเป็นแนวทางในการพัฒนาข้อสอบหรือการสร้างข้อสอบในการทดสอบครั้งต่อไป และเป็นการสร้างความมั่นใจต่อผลการวัดความสามารถด้านคำนวณของทดสอบครั้งนี้ที่จะนำไปใช้ประกันคุณภาพผู้เรียนในความสามารถด้านคำนวณ (Numeracy) เพื่อพัฒนาการเรียนการสอนของครู

3. ข้อเสนอแนะ

5.1 ข้อเสนอแนะในการนำผลวิจัยไปใช้

5.1.1 การสร้างแบบทดสอบในแต่ละครั้งควรตรวจสอบคุณภาพของข้อสอบซึ่งนอกจากหาค่าความยากง่าย (Difficulty) ค่าอำนาจจำแนก (Discrimination) ค่าความเที่ยง (Reliability) และค่าความตรงเชิงเนื้อหาแล้ว ควรวิเคราะห์ข้อสอบที่ทำหน้าที่ต่างกันเพื่อลดการได้เปรียบเสียเปรียบของผู้สอบ

5.1.2 ในการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบควรวิเคราะห์ควบคู่กับการวัดขนาดอิทธิพลการทำหน้าที่ต่างกันของข้อสอบ เพื่อใช้ในการตัดสินใจข้อสอบที่ทำหน้าที่ต่างกันสำหรับการตัดสินใจข้อสอบว่าควรปรับปรุงแก้ไข หรือคงข้อสอบข้อนั้นไว้ในแบบสอบ เพื่อนำไปใช้หรือเก็บเข้าคลังข้อสอบ

5.2 ข้อเสนอแนะในการวิจัยครั้งต่อไป

5.2.1 ควรศึกษาการทำหน้าที่ต่างกันของข้อสอบตัวแปรกลุ่มอื่น เช่น อายุ ภาษา ศาสนา ภูมิภาค วิถีวัฒนธรรม สังคม และสภาพทางเศรษฐกิจของครอบครัว เป็นต้น

5.2.2 ควรศึกษาการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีวิเคราะห์วิธีอื่นเช่นวิธีแมนเทล-แฮนเซล(Mantel – Haenszel; MH) ,วิธีชิปเทสต์ (SIBTEST) , วิธีตรวจสอบด้วยทฤษฎีIRT, วิธีความแตกต่างของค่าพารามิเตอร์ (b parameter difference) และวิธีทดสอบไค-สแควร์ของลอร์ด (Lord's test) เป็นต้น





บรรณานุกรม

บรรณานุกรม

- กฎกระทรวงว่าด้วยระบบ หลักเกณฑ์ และวิธีการประกันคุณภาพการศึกษา พ.ศ. 2553. (2553).
ราชกิจจานุเบกษา เล่มที่ 127 ตอนที่ 23 ก, 22–35.
- กระทรวงศึกษาธิการ. (2552). *หลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551*
สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน. กรุงเทพฯ: โรงพิมพ์ชุมนุมสหกรณ์
การเกษตรแห่งประเทศไทย จำกัด.
- จันทนา เปรมฤดีปรีชาชาญ. (2551). การเปรียบเทียบความสอดคล้องของผลการตรวจสอบข้อสอบ
ที่ทำหน้าที่ต่างกันตามเชื้อชาติของผู้สอบระหว่างวิธีการของ Mantel – Haenszel วิธี
Lord' x^2 และ วิธี Closed Interval Area. (วิทยานิพนธ์ปริญญาการศึกษามหาบัณฑิต).
มหาวิทยาลัยศรีนครินทรวิโรฒ, กรุงเทพฯ.
- ญาณภัทร สีหะมงคล. (2540). การเปรียบเทียบผลการตรวจสอบที่ทำหน้าที่ต่างกัน ระหว่างวิธี
Lord' x^2 วิธีการของ Mantel- Haenszel และวิธีSIBTEST. (วิทยานิพนธ์ปริญญา
การศึกษาศุภบัณฑิต). มหาวิทยาลัยศรีนครินทรวิโรฒ, กรุงเทพฯ.
- ทองอยู่ สารระ. (2543). *การเปรียบเทียบอำนาจการทดสอบและการจำแนกผิดพลาดในการ
ตรวจสอบ ข้อสอบที่ทำหน้าที่ต่างกันแบบสมมาตรระหว่างวิธีแมนเทล-แฮนส์เซล
และวิธีถดถอยโลจิสติก*. (วิทยานิพนธ์ปริญญาการศึกษามหาบัณฑิต).
มหาวิทยาลัยศรีนครินทรวิโรฒ, กรุงเทพฯ.
- ชเกียรติกมล ทองอก. (2554). *ประสิทธิภาพการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในวิธี
ถดถอยโลจิสติก โดยใช้เกณฑ์ขนาดอิทธิพล 2 วิธี สำหรับข้อสอบที่มีรูปแบบการตรวจ
ในคะแนนแบบทวิภาค: ข้อมูลจำลองและข้อมูลเชิงประจักษ์*. (วิทยานิพนธ์ปริญญา
ครุศาสตร์คุษบัณฑิต). จุฬาลงกรณ์มหาวิทยาลัย, กรุงเทพฯ.
- นวลอนงค์ นันทวงศ์. (2543). *การเปรียบเทียบความลำเอียงของแบบทดสอบความเข้าใจการอ่าน
ตามลักษณะการใช้ภาษาพูดในครอบครัว โดยวิธีวิเคราะห์ต่างกัน*. (วิทยานิพนธ์
ปริญญาการศึกษามหาบัณฑิต). มหาวิทยาลัยศรีนครินทรวิโรฒ, กรุงเทพฯ.
- พรณี จินตมาศ. (2540). *การเปรียบเทียบผลการวิเคราะห์ความลำเอียงของข้อสอบ โดยใช้ขนาด
กลุ่มผู้สอบและวิธีวิเคราะห์ต่างกัน*. (วิทยานิพนธ์ปริญญาการศึกษามหาบัณฑิต).
มหาวิทยาลัยศรีนครินทรวิโรฒ, กรุงเทพฯ.

- นพมาศ พิพัฒน์สุข. (2541). *การเปรียบเทียบประสิทธิภาพระหว่างวิธีแมนเทิล-แฮนส์เชลกับ วิธีถดถอยโลจิสติกในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบเมื่อใช้เกณฑ์จับคู่เปรียบเทียบ แตกต่างกันในแบบทดสอบชนิดพหุมิติ.* (วิทยานิพนธ์ปริญญาครุศาสตรมหาบัณฑิต). จุฬาลงกรณ์มหาวิทยาลัย, กรุงเทพฯ.
- ปิยะทิพย์ ดินวร. (2549). *การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบพหุมิติ: การเปรียบเทียบประสิทธิภาพระหว่างวิธีวิเคราะห์องค์ประกอบจำกัดกับวิธีถดถอยโลจิสติก.* (วิทยานิพนธ์ปริญญาวิทยาศาสตรมหาบัณฑิต). มหาวิทยาลัยบูรพา, ชลบุรี.
- มหาวิทยาลัยสุโขทัยธรรมมาธิราช. (2554). *การวิจัยและสถิติทางการศึกษา. ใน การสร้าง และตรวจสอบคุณภาพแบบทดสอบ.* หน่วยที่ 6 นนทบุรี: สำนักพิมพ์มหาวิทยาลัยสุโขทัยธรรมมาธิราช.
- รัชนีทร์ มุคคา. (2540). *การเปรียบเทียบประสิทธิภาพระหว่างแมนเทิล-แฮนเซนส์เชลกับวิธีถดถอยโลจิสติกในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอนุกรม ในกรณีการ จัดกลุ่มความสามารถ ค่าความยากง่ายของข้อสอบ และค่าอำนาจจำแนกของข้อสอบต่างกัน.* (วิทยานิพนธ์ปริญญาครุศาสตรมหาบัณฑิต). จุฬาลงกรณ์มหาวิทยาลัย, กรุงเทพฯ.
- เรวดี อินทสระ. (2539). *ผลการตรวจสอบความลำเอียงของข้อสอบต่อการศึกษาความเที่ยงตรงเชิง พยากรณ์ของแบบทดสอบคัดเลือกที่คิดคะแนนต่างกัน.* (วิทยานิพนธ์ปริญญาการศึกษาคุุณบัณฑิต). มหาวิทยาลัยศรีนครินทรวิโรฒ, กรุงเทพฯ.
- วลีมาศ แซ่อึ้ง. (2543). *การเปรียบเทียบอำนาจการทดสอบ และอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบหลายรูปแบบระหว่างวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทิล-แฮนเชล และวิธี โลจิสติก.* (วิทยานิพนธ์ปริญญาคุุณบัณฑิต). จุฬาลงกรณ์มหาวิทยาลัย, กรุงเทพฯ.
- ศิริชัย กาญจนวาสี. (2555). *ทฤษฎีการทดสอบแนวใหม่.* (พิมพ์ครั้งที่ 4). กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- สำนักทดสอบทางการศึกษา,สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน. (2557). *คู่มือการตรวจสอบประเมินคุณภาพการศึกษาขั้นพื้นฐานเพื่อการประกันคุณภาพผู้เรียน ปีการศึกษา 2557.*
- สุกัญญา ทองนาค. (2549). *การวิเคราะห์ความลำเอียงของข้อสอบเข้าศึกษาต่อประเภท โควตา มหาวิทยาลัยเชียงใหม่.* (วิทยานิพนธ์ปริญญาวิทยาศาสตรมหาบัณฑิต). มหาวิทยาลัยเชียงใหม่, เชียงใหม่.

- สุชาติพิทย์ ตรีสิน และปิยะทิพย์ ประจักษ์พรหม. (2560). การเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบระดับชาติชั้นประถมศึกษาปีที่ 3 ด้วยวิธี HGLM วิธี MIMIC และวิธี IRT – LR. วิทยาการวิจัยและวิทยาการปัญญา. มหาวิทยาลัยบูรพา.
- สุมาลี แก้วทงกต์. (2547). สาเหตุการทำหน้าที่ต่างกันของข้อสอบสาระการเรียนรู้ภาษาไทย และสาระการเรียนรู้สังคมศึกษา ศาสนาและวัฒนธรรม. (วิทยานิพนธ์ปริญญาครุศาสตรมหาบัณฑิต). จุฬาลงกรณ์มหาวิทยาลัย, กรุงเทพฯ.
- สุชาติ สิริมินนนท์. (2554). การเปรียบเทียบวิธี โพลี โทมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบในแบบทดสอบที่มีการให้คะแนนแบบหลายค่า. (วิทยานิพนธ์ปริญญาการศึกษาคุุณชีพัฒนชาติ). มหาวิทยาลัยศรีนครินทรวิโรฒ, กรุงเทพฯ.
- อมลณัฐ อุบลรัตน์. (2553). การทำหน้าที่ต่างกันด้วยวิธีโลจิสติกและวิธีแมนเทิล-แฮนเซลสำหรับข้อสอบวัดผลสัมฤทธิ์ทางการเรียนชั้นมัธยมศึกษาปีที่ 2. (วิทยานิพนธ์ปริญญาศึกษาศาสตรมหาบัณฑิต). มหาวิทยาลัยสงขลานครินทร์, สงขลา.
- อรินทร์ อ่วมถนอม. (2549). การเปรียบเทียบวิธี โพลี-ซิบเทสท์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า. (วิทยานิพนธ์ปริญญาการศึกษาคุุณชีพัฒนชาติ). มหาวิทยาลัยศรีนครินทรวิโรฒ, กรุงเทพฯ.
- Camilli, G., & Shepard, L.A. (1994). Methods for Identifying biased test items. *Journal of Educational Statistics*, 12,87-99.
- Gierl, M.J.; Gotzmann, A.; & Boughton, K.A. (2004). Performance of SIBTEST when the percentage of DIF Item is large. *Applied Measurement in Education*. 17(3):241-264.
- Gomez-Benito, J.Hidalgo, M.D. and Padilla, J.L. (2009). Efficacy of Effect Size Measurement in Logistic Regression An Application for Detecting DIF. *Methodology : European Journal of Research Methods for the Behavioral and Social Sciences*.Vol 5(1),18-25
- Mazor, K. M., Kanjee, A., & Clauser, B. E. (1995). Using Logistic Regression and the Mantel-Haenszel with Multiple Ability Estimates to Detect Differential Item Functioning. *Journal of Educational Measurement*, 32(2), 131-144.

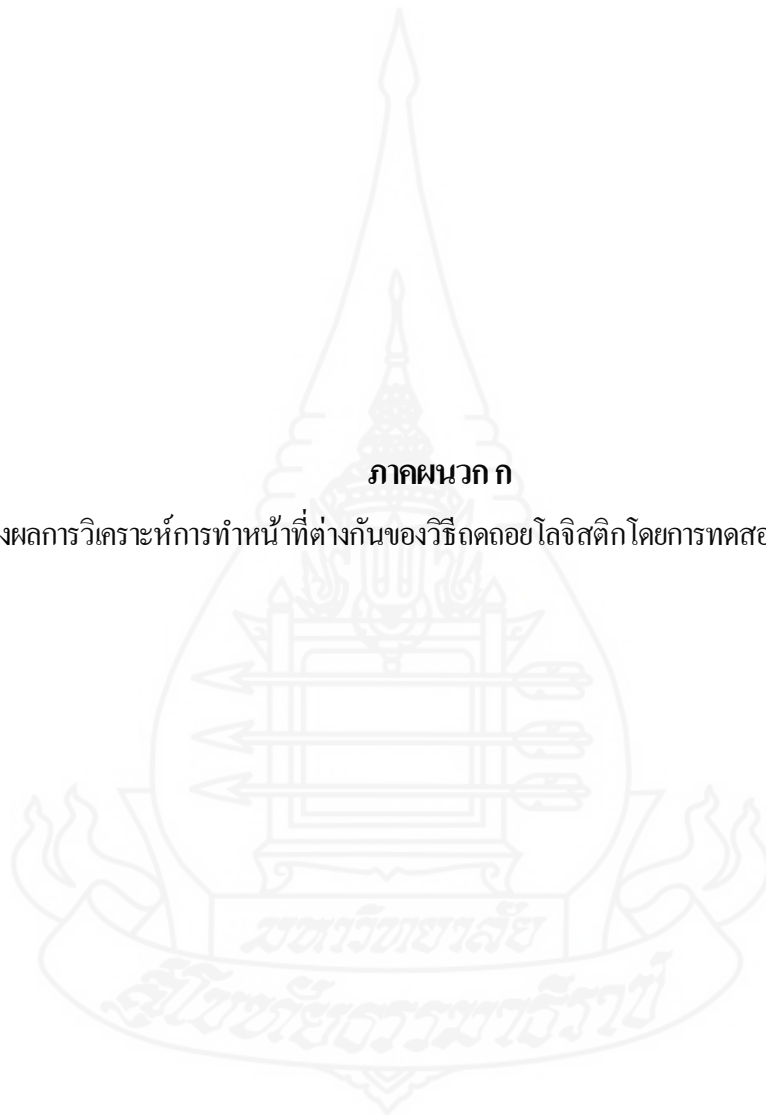
- Narayanan, P.; & Swaminathan, H. (1996). *Identification of items that show nonuniform DIF*. Applied Psychological Measurement, 20(3):257-274.
- Potenza, M.T.; & Dorans, N. J. (1995). *DIF assessment for polytomously scored items: A framework for classification and evaluation*. Applied Psychological Measurement, 19(2), 211-237.
- Rogers, H. J., & Swaminathan, H. (1993). *A comparison of logistic regression and Mantel-Haenszel procedures for detecting differential item functioning*. Applied Psychological Measurement, 17(2), 105-116
- Shealy, R., & Stout, W. (1993). A model-based standardization approach that separates true bias/DIF from group ability differences and detects test bias/DIF as well as item bias/DIF. Psychometrika, 58(2), 159-194.
- Holland & H. Wainer. (Eds.), *Differential item functioning: Theory and practice*. (197- 239) .Hillsdele, N.J. : Lawrence Erlbaum
- Shealy, R., & Stout, W.F. (1993). A model-based standardization approach that separates true bias/DIF from group ability differences and detects test bias/DIF as well as item bias/DIF. Psychometrika, 58(2): 159-194.
- Swaminathan, H., & Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. Journal of Educational Measurement, 27(4):361-370.
- Zumbo, B. D. (1999). A handbook of the theory and methods of differential item functioning (DIF) :Logistic regression modeling as a unitary framework for binary and Likert-type (ordinal) item score. Ottawa, Ontario, Canada: Directorate of Human resources Research and Evaluation, Department National Defense.
- Zumbo, B.D., & Thomas, D.R. (1997). A measure of effect size for a model-based approach for studying DIF. Prince George, Canada: University of Northem British Columbia, Edgeworht Laboratory for Quantitative Behavioral Science.



ภาคผนวก

ภาคผนวก ก

ตัวอย่างผลการวิเคราะห์การทำหน้าที่ต่างกันของวิธีถดถอยโลจิสติกโดยการทดสอบระดับนัยสำคัญ



ตัวอย่างผลการวิเคราะห์การทำหน้าที่ต่างกันของวิธีจดรอยโลจิสติก
โดยการทดสอบระดับนัยสำคัญ

ข้อที่ 1

Model 1.
$$\psi_{ij} = \tau_0 + \tau_1 X_{ij} + \tau_2 G_j + \tau_3 (G_j X_{ij})$$

Logistic Regression

Case Processing Summary

Unweighted Cases ^a		N	Percent
Selected Cases	Included in Analysis	464532	100.0
	Missing Cases	0	.0
	Total	464532	100.0
Unselected Cases		0	.0
Total		464532	100.0

a. If weight is in effect, see classification table for the total number of cases.

Dependent Variable Encoding

Original Value	Internal Value
ผิด	0
ถูก	1

Variables in the Equation

	B	S.E.	Wald	Df	Sig.	Exp(B)
Step 0 Constant	.171	.003	3.368E3	1	.000	1.186

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	group	3.943E3	1	.000
		score	9.226E4	1	.000
		Group by score	3.327E3	1	.000
	Overall Statistics		9.351E4	3	.000

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	Df	Sig.
Step 1	Step	1.049E5	3	.000
	Block	1.049E5	3	.000
	Model	1.049E5	3	.000

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	535720.525 ^a	.202	.270

a. Estimation terminated at iteration number 4 because parameter estimates changed by less than .001.

Classification Table^a

Observed			Predicted		
			score item N1		Percentage
			ผิด	ถูก	Correct
Step 1	score item N1	ผิด	137943	74522	64.9
		ถูก	65848	186219	73.9
	Overall Percentage				69.8

a. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a	group	-.240	.019	154.974	1	.000	.786
	score	.231	.001	3.748E4	1	.000	1.259
	Group by score	.000	.002	.000	1	.988	1.000
	Constant	-2.279	.014	2.730E4	1	.000	.102

a. Variable(s) entered on step 1: group, score, groupbyscore.

Model 2.

$$\psi_{ij} = \tau_0 + \tau_1 X_{ij} + \tau_2 G_j$$

Case Processing Summary

Unweighted Cases ^a		N	Percent
Selected Cases	Included in Analysis	464532	100.0
	Missing Cases	0	.0
	Total	464532	100.0
Unselected Cases		0	.0
Total		464532	100.0

a. If weight is in effect, see classification table for the total number of cases.

Dependent Variable Encoding

Original Value	Internal Value
ผิด	0
ถูก	1

Block 0: Beginning Block**Classification Table^{a,b}**

Observed			Predicted		
			score item N1		Percentage Correct
			ผิด	ถูก	
Step 0	score item N1	ผิด	0	212465	.0
		ถูก	0	252067	100.0
Overall Percentage					54.3

a. Constant is included in the model.

b. The cut value is .500

Classification Table^{a,b}

Observed		Predicted			
		score item N1		Percentage Correct	
		ผิด	ถูก		
Step 0	score item N1	ผิด	0	212465	.0
		ถูก	0	252067	100.0
Overall Percentage					54.3

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	.171	.003	3.368E3	1	.000	1.186

Variables not in the Equation

		Score	df	Sig.
Step 0	Variables			
	group	3.943E3	1	.000
	score	9.226E4	1	.000
Overall Statistics		9.347E4	2	.000

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	Df	Sig.
Step 1	Step	1.049E5	2	.000
	Block	1.049E5	2	.000
	Model	1.049E5	2	.000

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	535720.525 ^a	.202	.270

a. Estimation terminated at iteration number 4 because parameter estimates changed by less than .001.

Variables in the Equation

	B	S.E.	Wald	Df	Sig.	Exp(B)
Step 1 ^a group	-.241	.007	1.322E3	1	.000	.786
score	.231	.001	7.439E4	1	.000	1.259
Constant	-2.279	.010	4.880E4	1	.000	.102

a. Variable(s) entered on step 1: group, score.

LOGISTIC REGRESSION VARIABLES Nu1

/METHOD=ENTER score

/CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).

Model 3. $\psi_{ij} = \tau_0 + \tau_1 X_{ij}$

Case Processing Summary

Unweighted Cases ^a		N	Percent
Selected Cases	Included in Analysis	464532	100.0
	Missing Cases	0	.0
	Total	464532	100.0
Unselected Cases		0	.0
Total		464532	100.0

a. If weight is in effect, see classification table for the total number of cases.

Dependent Variable Encoding

Original Value	Internal Value
ผิด	0
ถูก	1

Block 0: Beginning Block

Classification Table^{a,b}

Observed		Predicted			
		score item N1		Percentage Correct	
		ผิด	ถูก		
Step 0	score item N1	ผิด	0	212465	.0
		ถูก	0	252067	100.0
Overall Percentage					54.3

a. Constant is included in the model.

b. The cut value is .500

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 0 Constant	.171	.003	3.368E3	1	.000	1.186

Variables not in the Equation

		Score	df	Sig.
Step 0	Variables score	9.226E4	1	.000
Overall Statistics		9.226E4	1	.000

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	Df	Sig.
Step 1	Step	1.036E5	1	.000
	Block	1.036E5	1	.000
	Model	1.036E5	1	.000

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	537041.784 ^a	.200	.267

a. Estimation terminated at iteration number 4 because parameter estimates changed by less than .001.

Classification Table^a

Observed			Predicted		
			score item N1		Percentage Correct
			ผิด	ถูก	
Step 1	score item N1	ผิด	147496	64969	69.4
		ถูก	76455	175612	69.7
Overall Percentage					69.6

a. The cut value is .500

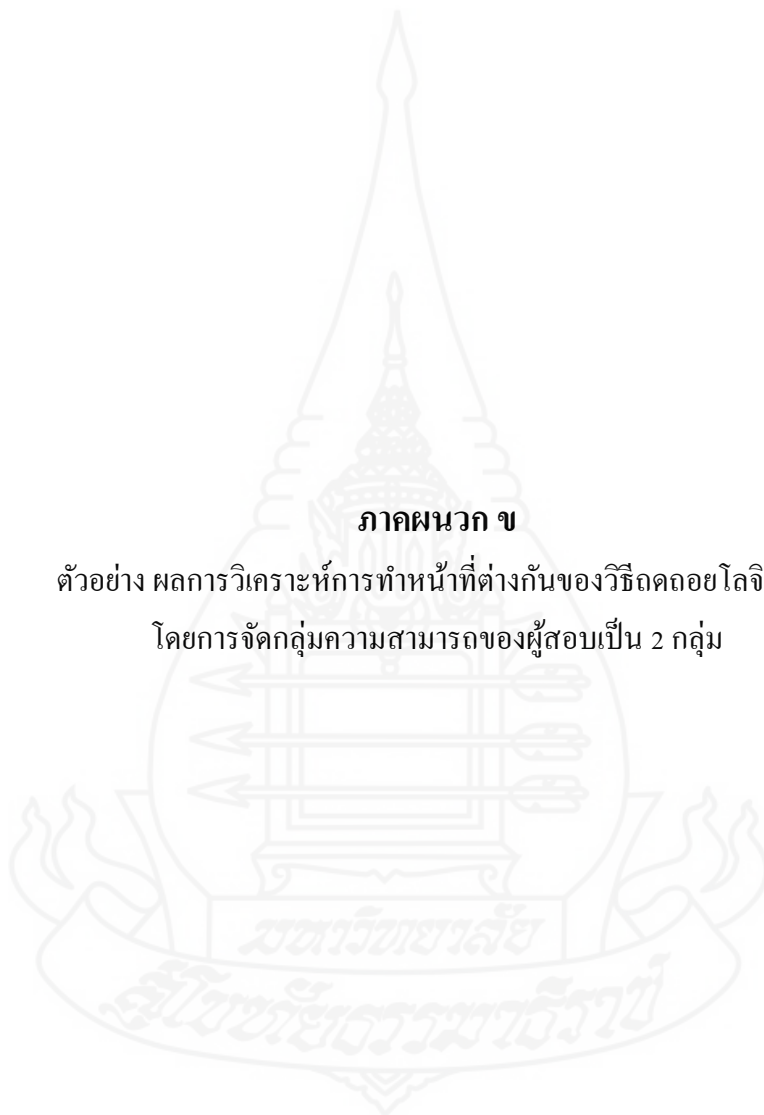
Variables in the Equation

	B	S.E.	Wald	Df	Sig.	Exp(B)
Step 1 ^a score	.233	.001	7.588E4	1	.000	1.262
Constant	-2.419	.010	6.291E4	1	.000	.089

a. Variable(s) entered on step 1: score.

ภาคผนวก ข

ตัวอย่าง ผลการวิเคราะห์การทำหน้าที่ต่างกันของวิธีดออยโลจิตติก
โดยการจัดกลุ่มความสามารถของผู้สอบเป็น 2 กลุ่ม



Statistics

score

N	Valid	464532
	Missing	0
Mean		11.5021
Std. Deviation		4.85236

Between-Subjects Factors

		Value Label	N
Gender	0	female	101893
	1	male	101253
ability	-1	Low	103718
	1	hight	99428

Tests of Between-Subjects Effects

Dependent Variable:score item N1

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	19179.029 ^a	3	6393.010	4.141E4	.000
Intercept	58498.797	1	58498.797	3.789E5	.000
Gender	78.632	1	78.632	509.323	.000
ability	18445.031	1	18445.031	1.195E5	.000
Gender * ability	1.874	1	1.874	12.137	.000
Error	31362.112	203142	.154		
Total	108633.000	203146			
Corrected Total	50541.141	203145			

Tests of Between-Subjects Effects

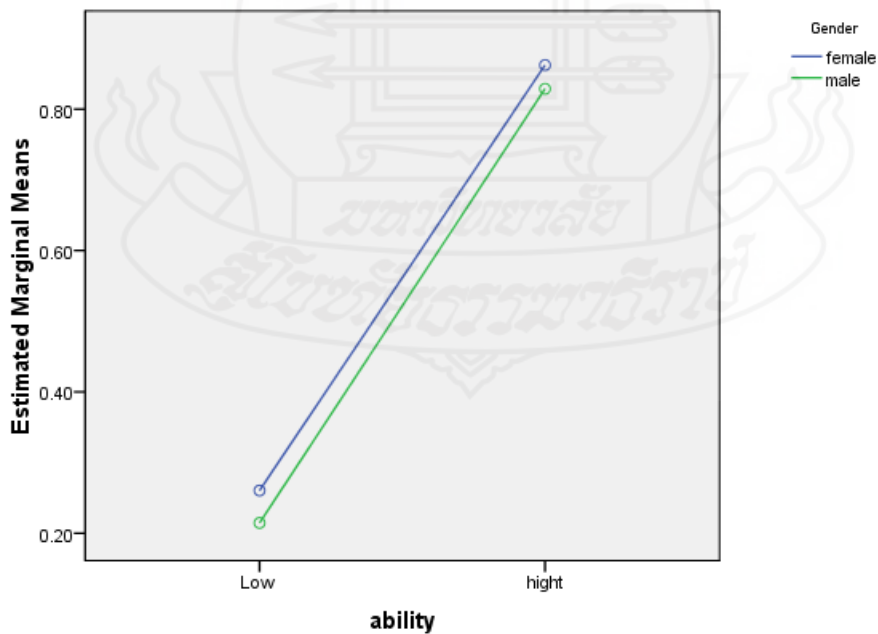
Dependent Variable:score item N1

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	19179.029 ^a	3	6393.010	4.141E4	.000
Intercept	58498.797	1	58498.797	3.789E5	.000
Gender	78.632	1	78.632	509.323	.000
ability	18445.031	1	18445.031	1.195E5	.000
Gender * ability	1.874	1	1.874	12.137	.000
Error	31362.112	203142	.154		
Total	108633.000	203146			

a. R Squared = .379 (Adjusted R Squared = .379)

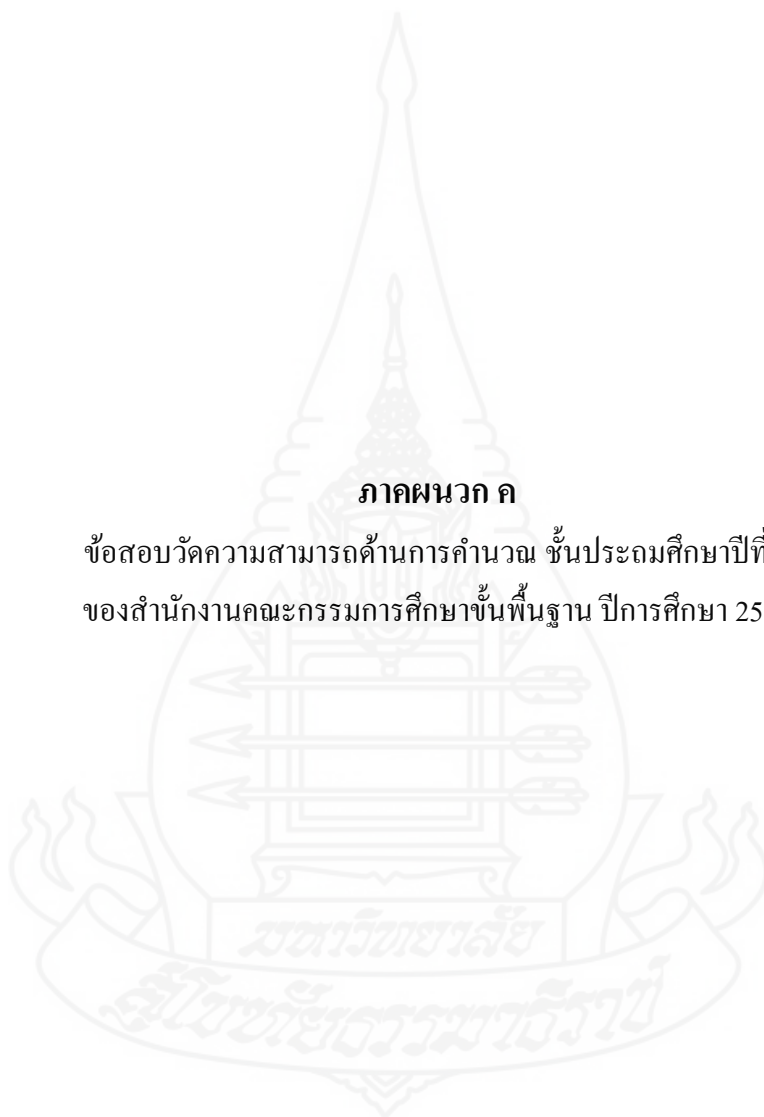
Profile Plots

Estimated Marginal Means of score item N1



ภาคผนวก ก

ข้อสอบวัดความสามารถด้านการคำนวณ ชั้นประถมศึกษาปีที่ 3
ของสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน ปีการศึกษา 2557



ข้อสอบวัดความสามารถด้านการคำนวณ ชั้นประถมศึกษาปีที่ 3
ของสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน ปีการศึกษา 2557

ให้นักเรียนเลือกคำตอบที่ถูกต้องที่สุดเพียงคำตอบเดียว

1. การออมเงินของนักเรียน 3 คน ตั้งแต่วันจันทร์-ศุกร์ในสัปดาห์หนึ่ง

	จันทร์	อังคาร	พุธ	พฤหัสบดี	ศุกร์
ค.ญ.มาลี	5	7	12	4	6
ค.ญ.ซูใจ	3	9	11	2	14
ค.ญ.วันดี	10	5	2	9	8

จากตาราง ข้อใดเรียงลำดับวัน ที่มีเงินออมรวมทั้งสามคนจากมากไปน้อย

- 1) วันพฤหัสบดี จันทร์ อังคาร พุธ ศุกร์
 - 2) วันจันทร์ อังคาร ศุกร์ พฤหัสบดี พุธ
 - 3) วันพุธ พฤหัสบดี อังคาร ศุกร์ จันทร์
 - 4) วันศุกร์ พุธ อังคาร จันทร์ พฤหัสบดี
2. ร้านอาหารแห่งหนึ่งคิดราคาเหมาจ่าย คนละ 140 บาท โดยจัดรายการพิเศษ ดังนี้
เงินจำนวน 1,400 บาท สามารถจ่ายค่าอาหารของคน 10 คน ได้หรือไม่

มา 5 คน	จ่ายเพียง 4 คน
มา 12 คน	จ่ายเพียง 9 คน

- 1) ได้ และยังเหลือเงินอีก 140 บาท
 - 2) ได้ และยังเหลือเงินอีก 280 บาท
 - 3) เงินที่มีอยู่จ่ายค่าอาหารได้พอดี
 - 4) เงินที่มีอยู่ไม่พอจ่ายค่าอาหาร
3. คุณแม่มีผลไม้ 3 ชนิดดังนี้ มังคุด 25 ผล ส้มเขียวหวาน 15 ผล และชมพู 30 ผล
นำมาจัดใส่ถาด 5 ใบ โดยให้ถาดแต่ละใบมีผลไม้แต่ละชนิดจำนวนเท่า ๆ กัน ได้ดังข้อใด
- 1) มังคุด 3 ผล ส้มเขียวหวาน 5 ผล ชมพู 6 ผล
 - 2) มังคุด 5 ผล ส้มเขียวหวาน 6 ผล ชมพู 3 ผล
 - 3) มังคุด 5 ผล ส้มเขียวหวาน 3 ผล ชมพู 6 ผล
 - 4) มังคุด 6 ผล ส้มเขียวหวาน 5 ผล ชมพู 3 ผล

4.

แม่ซื้อจานจำนวนครึ่งโหล ให้ธนบัตรห้าร้อยบาท 1 ฉบับ
ได้รับเงินทอน 260 บาท

ติ่มกับต๋อยช่วยกันบรรจุส้มใส่ถุง ติ่มบรรจุได้ 25 ถุง
ต๋อยบรรจุได้ 30 ถุง โดยได้รับค่าจ้างถุงละ 8 บาท

ข้อใดถูกต้อง

- 1) จานราคาใบละ 40 บาท
2) จานราคาโหลละ 400 บาท
3) แม่ซื้อจานทั้งหมดเป็นเงิน 260 บาท
4) ถ้าซื้อ 1 โหล ได้รับเงินทอน 40 บาท

5.

ข้อใดถูกต้อง

- 1) ติ่มได้รับค่าจ้างน้อยกว่าต๋อย 40 บาท
2) ต๋อยได้รับค่าจ้างมากกว่าติ่ม 50 บาท
3) ติ่มได้รับค่าจ้างมากกว่าต๋อย 55 บาท
4) ต๋อยได้รับค่าจ้างน้อยกว่าติ่ม 45 บาท

6.

จำนวนดอกกุหลาบที่นึ่งทำได้เวลา 5 วัน

วัน	จำนวนดอกกุหลาบที่ทำได้ (ดอก)
วันแรก	4
วันที่สอง	6
วันที่สาม	8
วันที่สี่	10
วันที่ห้า	12

จากตาราง ข้อใดถูกต้อง

- 1) นึ่งทำดอกกุหลาบได้ทั้งหมด 30 ดอก
2) ในแต่ละวันนึ่งทำดอกกุหลาบเพิ่มขึ้น 3 ดอก
3) นึ่งทำดอกกุหลาบวันที่ห้ามากกว่าวันที่สองจำนวน 8 ดอก
4) นึ่งทำดอกกุหลาบวันแรกและวันที่สองรวมกัน ได้น้อยกว่าวันที่ห้าอยู่ 2 ดอก

7. บริษัทผลิตน้ำผลไม้ได้วันละ 6 ชั่วโมง ชั่วโมงละ 250 กระป๋อง แต่ต้องส่งให้ลูกค้าจำนวน 1,600 กระป๋อง ยังขาดน้ำผลไม้ก็กระป๋อง

วิธีคิดหาคำตอบตรงตามข้อใด

- 1) $1,600 - (6 + 250)$
- 2) $(250 + 6) - 1,600$
- 3) $1,600 - (6 \times 250)$
- 4) $(250 - 6) - 1,600$

8.



จิวรรณต้องการซื้อตุ๊กตา 1 ตัว ที่มีราคาถูกที่สุดจะต้องเลือกร้านใด เพราะเหตุใด

- 1) ร้านหนึ่ง เพราะราคาถูกกว่า 182 บาท
- 2) ร้านหนึ่ง เพราะราคาถูกกว่า 4 บาท
- 3) ร้านสอง เพราะราคาถูกกว่า 4 บาท
- 4) ร้านสอง เพราะราคาถูกกว่า 182 บาท

9.

ราคาสินค้า 6 ชนิด ของร้านสหกรณ์โรงเรียนแห่งหนึ่ง

รายการสินค้า	ราคาต่อหน่วย (บาท)
สมุด	20
ดินสอ	7
ปากกา	12
ไม้บรรทัด	5
ยางลบ	5
สีไม้	35

จากตาราง ข้อใดถูกต้อง

- 1) ปิติซื้อสมุด 2 เล่ม ดินสอ 1 แท่ง จ่ายเงิน 57 บาท
- 2) นุติซื้อไม้บรรทัด 4 อัน สีไม้ 1 กล่อง จ่ายเงิน 55 บาท
- 3) ปรีชา ซื้อดินสอ 3 แท่ง ยางลบ 1 แท่ง จ่ายเงิน 30 บาท
- 4) จารุณีซื้อปากกา 2 ค้าม ดินสอ 1 แท่ง จ่ายเงิน 30 บาท

10. ขาวสูง 160 เซนติเมตร และสูงกว่าแดง 20 เซนติเมตร ดำเตี้ยกว่าแดง 15 เซนติเมตร ดำมีความสูงกี่เซนติเมตร

- 1) 125
- 2) 135
- 3) 140
- 4) 145

ใช้ข้อมูลต่อไปนี้ตอบคำถาม ข้อ 11 - 12



11. ต้องการตักน้ำใส่ตู้ปลาให้ได้ปริมาตร 4 ลิตร 500 มิลลิลิตร ข้อใดถูกต้อง

- 1) ประชา ใช้แก้วตักน้ำเต็ม 20 แก้ว
- 2) เสรี ใช้ช้อนตักน้ำเต็ม 8 ช้อน และใช้แก้วตักน้ำเต็ม 1 แก้ว
- 3) แก้วตา ใช้เหยือกตักน้ำเต็ม 4 เหยือก และใช้ช้อนตักน้ำเต็ม 2 ช้อน
- 4) มนตรี ใช้แก้วตักน้ำเต็ม 6 แก้ว และใช้เหยือกตักน้ำเต็ม 3 เหยือก

12. ข้อใดถูกต้อง

- 1) เข็ช็อก 1 ใบ มีความจุเท่ากับ ชั้น 3 ใบ
- 2) ชั้น 2 ใบ มีความจุเท่ากับ แก้ว 4 ใบ
- 3) แก้ว 5 ใบ มีความจุมากกว่า ชั้น 3 ใบ
- 4) เข็ช็อก 2 ใบ มีความจุน้อยกว่า แก้ว 8 ใบ



แก้วตึ้นนอนตามเวลา ดังรูป แล้วทำกิจกรรมต่างๆ
ต้อเนื่องกันตามลำดับ ดังตาราง

ลำดับที่	กิจกรรมที่ทำ	ใช้เวลา
1	อาบน้ำ	10 นาที
2	แต่งตัว	20 นาที
3	รับประทานอาหารเช้า	30 นาที
4	เดินทางจากบ้านถึงโรงเรียน	30 นาที

13.

ถ้าโรงเรียนเลิกเวลา 15.30 น. แก้วใช้เวลาอยู่ในโรงเรียนกี่ชั่วโมง

- 1) 8
- 2) 7
- 3) 6
- 4) 5

14.

รายรับ-รายจ่ายของเด็กหญิงดาวรุ่ง ตั้งแต่วันที่ 10 ถึง วันที่ 12 กันยายน พ.ศ. 2557

วันเดือนปี	รายการ	รายรับ		รายจ่าย		คงเหลือ	
		บาท	สต.	บาท	สต.	บาท	สต.
10 กันยายน 2557	พ่อให้เงิน	50	-	-	-	50	-
	แม่ให้เงิน	60	-	-	-	110	-
11 กันยายน 2557	ซื้อหนังสือ	-	-	34	50	75	50
	ซื้อปากกา	-	-	15	50	60	-
12 กันยายน 2557	ขายขนม	100	-	-	-	160	-
	ซื้อสมุด	-	-	20	-	140	-
	ซื้อถุงเท้า	-	-	50	-	90	-

ข้อใดถูกต้อง

- 1) วันที่ 11 กันยายน มีรายจ่ายทั้งหมด 49 บาท
- 2) รายรับทั้งหมดมากกว่ารายจ่ายทั้งหมด อยู่ 30 บาท
- 3) รายจ่ายวันที่ 12 กันยายน มากกว่า วันที่ 11 กันยายน อยู่ 21 บาท
- 4) รายจ่ายวันที่ 11 กันยายน น้อยกว่า รายรับวันที่ 12 กันยายน อยู่ 50 บาท

15. แม่ซื้อสิ่งของดังภาพต่อไปนี้



ข้อใดถูกต้อง

- 1) มะระหนักกว่าฟักทอง
- 2) ผักกาดขาวหนักกว่ามะระ
- 3) มะระและผักกาดขาว หนักรวมกันเป็น 2 กิโลกรัม 2 ซีด
- 4) ผักกาดขาวและฟักทอง หนักรวมกันเป็น 2 กิโลกรัม 2 ซีด

16.



มานะมีเงินจำนวนดังภาพข้างบน คิดเป็นเงินเท่าใด

- 1) 9,784.75 บาท
- 2) 9,788.25 บาท
- 3) 9,813.25 บาท
- 4) 9,884.75 บาท

17.



จากรูปที่กำหนดให้ รูปในข้อใดที่มีรูปวงกลมเป็นส่วนประกอบทุกรูป

- 1) รูป ①, ② และ ③
- 2) รูป ①, ② และ ④
- 3) รูป ①, ③ และ ④
- 4) รูป ②, ③ และ ④

18. รูปภาพตัวอักษรในข้อใดมีแกนสมมาตรทั้งสองตัว

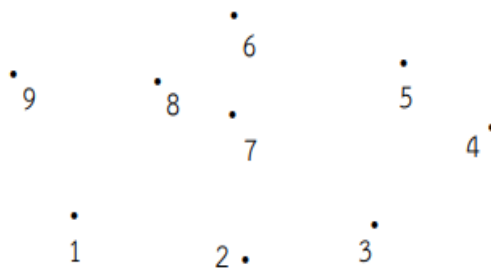
1) A B

2) M N

3) S T

4) Y Z

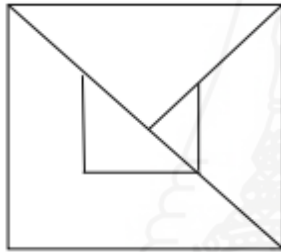
19.



จากภาพ ลากเส้นเชื่อมจุดตามข้อใดจะเกิดเป็นรูปห้าเหลี่ยม

- 1) จุด 1, 2, 7, 5, 4 และ 1 ตามลำดับ
- 2) จุด 1, 8, 5, 4, 2 และ 1 ตามลำดับ
- 3) จุด 3, 2, 8, 9, 1 และ 3 ตามลำดับ
- 4) จุด 3, 4, 5, 8, 6 และ 3 ตามลำดับ

20.



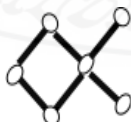
จากรูป ถ้าตัดกระดาษตามเส้นทุกเส้น กระดาษที่ตัดแต่ละชิ้นจะไม่เกิดรูปในข้อใด

- 1) รูปสามเหลี่ยม
- 2) รูปสี่เหลี่ยม
- 3) รูปห้าเหลี่ยม
- 4) รูปหกเหลี่ยม

21. นำก้อนไม้ขีดไฟมาวาง ดังรูป



รูปที่ 1



รูปที่ 2



รูปที่ 3



รูปที่ 4

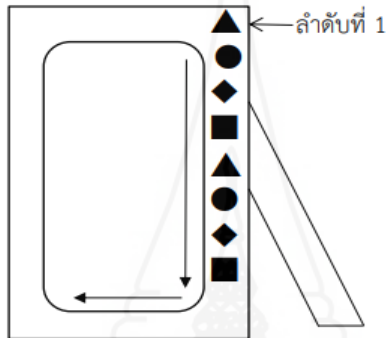


รูปที่ 5

รูปที่ 5 ต้องใช้ไม้ขีดไฟกี่ก้าน

- 1) 12
- 2) 13
- 3) 14
- 4) 15

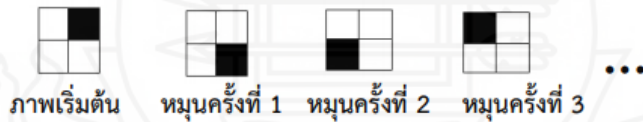
22. เอ ตกแต่งกรอบรูป ด้วยรูปเรขาคณิต โดยวนตามเข็มนาฬิกา ดังรูป



ถ้าเอ ต้องการติดกรอบรูปตามแบบข้างต้น รูปเรขาคณิตลำดับที่ 18 คือข้อใด

- 1) ▲
- 2) ●
- 3) ◆
- 4) ■

23. วัฒนา หมุนบล็อกไม้ตามเข็มนาฬิกา ดังภาพ



ภาพที่ได้จากการหมุนครั้งที่ 5 คือภาพใด

- 1)
- 2)
- 3)
- 4)

24. กำหนดแบบรูป ดังนี้

1, 3, 4, 5, 1, 3, 4, 5, 1, 3, ...


จำนวนที่อยู่ในลำดับที่ 19 คือจำนวนใด

- 1) 1
- 2) 3
- 3) 4
- 4) 5

ใช้ข้อมูลต่อไปนี้ตอบคำถามข้อ 25 - 26

จำนวนนักเรียนชั้น ป.3 ของโรงเรียนแห่งหนึ่ง ที่ชอบเครื่องประดับคนละ 1 ชนิด

นาฬิกา	
สร้อยคอ	
แหวน	
กำไลมือ	

 แทนนักเรียน 3 คน

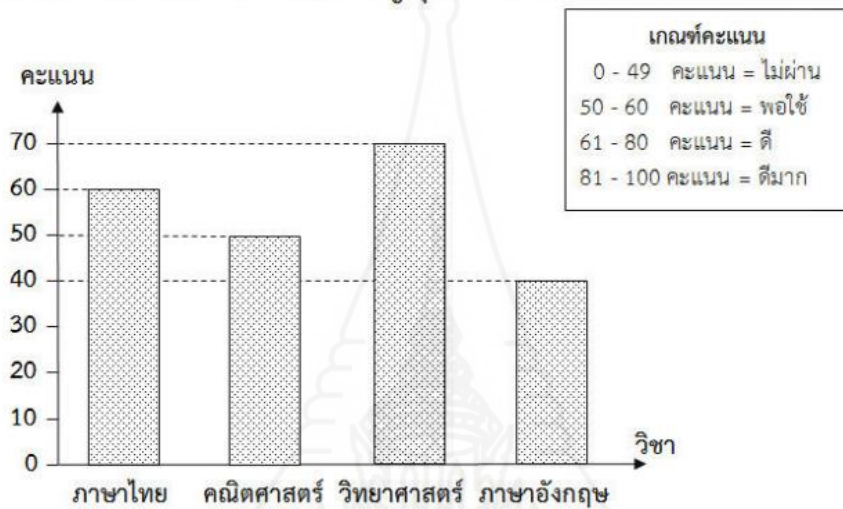
25. จำนวนนักเรียนที่ชอบแหวนมากกว่าชอบสร้อยค้อมีกี่คน

- 1) 2
- 2) 4
- 3) 6
- 4) 8

26. โรงเรียนนี้มีนักเรียนชั้น ป.3 ทั้งหมดกี่คน

- 1) 60
- 2) 66
- 3) 70
- 4) 76

คะแนนผลการสอบใน 5 วิชา ของเด็กหญิงรุจิรา รักการเรียน ปีการศึกษา 2556



27. วิชาใดบ้างที่เด็กหญิงรุจิรา รักการเรียน ท คะแนนอยู่ในระดับพอใช้

- 1) ภาษาไทย เท่านั้น
- 2) วิทยาศาสตร์ และ ภาษาไทย
- 3) คณิตศาสตร์ และ ภาษาอังกฤษ
- 4) ภาษาไทย และ คณิตศาสตร์

ประวัติผู้วิจัย

ชื่อ	นางสุภาภรณ์ อันทิ
วัน เดือน ปีเกิด	19 มีนาคม 2531
สถานที่เกิด	อำเภอเมืองตรัง จังหวัดตรัง
ประวัติการศึกษา	วิทยาศาสตรบัณฑิต (คณิตศาสตร์) มหาวิทยาลัยราชภัฏภูเก็ต พ.ศ. 2553 ประกาศนียบัตรบัณฑิต วิชาชีพครู มหาวิทยาลัยราชภัฏภูเก็ต พ.ศ.2554
สถานที่ทำงาน	โรงเรียนตรังคริสเตียนศึกษา
ตำแหน่ง	ครู

