

## การทำหน้าที่ต่างกันของข้อสอบวัดความสามารถด้านเหตุผลระหว่างนักเรียนชาย และนักเรียนหญิง: การวิเคราะห์จากผลการทดสอบระดับชาติ

### Differential Item Functioning in Reasoning Abilities between Males and Females : An analysis of the National Test Result

นิตยา ดวงจันทร์ทิพย์\*

นลินี ฌ นคร\*\*

สัจจวรรณ ังดกระโทก\*\*

#### บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์เพื่อวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบวัดความสามารถด้านเหตุผลของแบบทดสอบระดับชาติระหว่างนักเรียนชายและนักเรียนหญิง กลุ่มประชากรคือ นักเรียนชั้นประถมศึกษาปีที่ 3 ปีการศึกษา 2557 จำนวน 464,532 คน เครื่องมือที่ใช้ในการศึกษา คือ แบบทดสอบวัดความสามารถด้านเหตุผล ชั้นประถมศึกษาปีที่ 3 ของสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน จำนวน 27 ข้อ วิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ (DIF) ด้วยการใช้การวิเคราะห์ถดถอยโลจิสติก และวัดขนาดการทำหน้าที่ต่างกันของข้อสอบ โดยใช้เกณฑ์Jodoin and Gierl (2001) ผลการวิจัยพบว่า ข้อสอบทำหน้าที่ต่างกันทุกข้อ โดยมีข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูป จำนวน 2 ข้อ และข้อสอบที่ทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป จำนวน 25 ข้อ แต่เมื่อพิจารณาขนาดของการทำหน้าที่ต่างกันของข้อสอบ พบว่าทุกข้อมีขนาดของการทำหน้าที่ต่างกันเพียงขนาดเดียวคือ ขนาดเล็ก ดังนั้นจึงสรุปได้ว่าข้อสอบวัดความสามารถด้านเหตุผลของการทดสอบระดับชาติ(NT) มีความยุติธรรมทั้งนักเรียนชายและนักเรียนหญิง

**คำสำคัญ :** การทำหน้าที่ต่างกันของข้อสอบ, ความสามารถด้านเหตุผล

\* นักศึกษาระดับปริญญาโท แขนงวิชาวัดและประเมินผลการศึกษา สาขาวิชาศึกษาศาสตร์  
มหาวิทยาลัยสุโขทัยธรรมาธิราช

\*\* ผู้ช่วยศาสตราจารย์ ดร. แขนงวิชาการวัดและประเมินผลการศึกษา สาขาวิชาศึกษาศาสตร์  
มหาวิทยาลัยสุโขทัยธรรมาธิราช

## Abstract

The purposes of this research was to analyze the differential item functioning (DIF) in reasoning abilities of the national test between male and female students. The sample were 464,532 grade 3 students in the academic year 2014. The research instruments was the reasoning ability test for third grade students. The logistic regression was used to analyze DIF. The magnitude of Differential Item Functioning was measured by criteria proposed by Jodoin and Gierl (2001). The result showed that all items were identified as DIF. Specifically, 2 items were flagged as uniform DIF, while 25 items were non-uniform DIF. When consider the magnitude of differential item functioning, it was found that all test items were categorized into the little or non DIF category. Therefore, it can be summarized that NT test items were fair both the male and female.

Keywords : Differential item function , Reasoning

## บทนำ

การทดสอบวัดความสามารถพื้นฐานของผู้เรียนระดับชาติ เป็นการประเมินผู้เรียนเพื่อสะท้อนคุณภาพในการจัดการศึกษาของสถานศึกษา และสำนักงานเขตพื้นที่การศึกษา อีกทั้งยังเป็นตัวบ่งชี้คุณภาพการศึกษาขั้นพื้นฐานในภาพรวม ใช้ประกอบการตัดสินใจในการกำหนดนโยบายในระดับต่าง ๆ ในระดับสถานศึกษาเป็นการประเมิน เพื่อปรับปรุงและพัฒนาการเรียนการสอนของครู สร้างความมั่นใจกับสถานศึกษาและหน่วยงานต้นสังกัดจัดการศึกษาได้อย่างมีประสิทธิภาพตามมาตรฐานที่กำหนด การวัดและประเมินผลการเรียนรู้ของสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน ก็เป็นหลักประกันการเรียนรู้ (accountability) อย่างหนึ่ง ในการประเมินคุณภาพของนักเรียนว่ามีความรู้หรือคุณลักษณะที่ต้องการวัดอยู่ในระดับใด ซึ่งผลที่ได้จากการวัดมีความสำคัญต่อการพัฒนาคุณภาพของการศึกษา การทดสอบเป็นวิธีการวัดผลการศึกษาวิธีหนึ่งที่นิยมใช้มาก

การทดสอบเพื่อวัดความสามารถของผู้เรียนนั้นต้องมีการกำหนดจุดมุ่งหมายของการวัดที่ชัดเจน เครื่องมือวัดที่มีคุณภาพ ปัจจุบันเครื่องมือวัดทางการศึกษาและจิตวิทยา ส่วนใหญ่ใช้วัดพัฒนาการของผู้เรียน เช่น แบบสอบ แบบสัมภาษณ์ แบบสังเกตและแบบสอบถาม แบบประเมินการเขียน เป็นต้น ผลของการวัดถูกกำหนดค่าให้เป็นคะแนนซึ่งถือว่าเป็นเพียงตัวอย่างของพฤติกรรมที่แสดงออกมา แต่ไม่ใช่พฤติกรรมทั้งหมดของบุคคล เมื่อแปลผลของคะแนนแล้วต้องมีความหมายที่สื่อถึงคุณลักษณะภายในที่ต้องการจะวัดได้ถูกต้องตามความเป็นจริง คะแนนที่ได้เป็นคะแนนดิบไม่มีความสมบูรณ์มีคุณลักษณะเชิงสัมพัทธ์ (Relative) ต้อง

เปรียบเทียบความหมายกับเกณฑ์มาตรฐานอื่นแล้วจึงแปลความหมายต่อไปได้ (ศิริชัย กาญจนวาสี, 2555 : 10) ดังนั้นในการสร้างเครื่องมือต้องคำนึงถึงคุณภาพเครื่องมือ

การวิเคราะห์ข้อสอบเป็นเทคนิคของการตรวจสอบคุณภาพของข้อสอบเป็นรายข้อ ตัวบ่งชี้คุณภาพของข้อสอบที่สำคัญได้แก่ ค่าความยาก ค่าอำนาจจำแนกของข้อสอบ ส่วนคุณภาพของข้อสอบทั้งฉบับ ประกอบด้วยหลักฐานแสดงความเที่ยง (Reliability) และความตรง (Validity) ที่ได้มาตรฐาน โดยเฉพาะประเด็นด้านความตรง เป็นหัวใจสำคัญของคุณภาพของแบบทดสอบ ทั้งนี้เพราะความตรงเป็นคุณสมบัติของแบบทดสอบที่แสดงถึงความสามารถในการวัดได้ถูกต้องแม่นยำ และการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ (item and test unfairness) ก็เป็นอีกลักษณะหนึ่งของการตรวจสอบคุณภาพด้านความตรง ในประเด็นของความยุติธรรม (ศิริชัย กาญจนวาสี, 2555 : 115)

การตรวจสอบในประเด็นของความยุติธรรมของข้อสอบและแบบสอบ เป็นการจัดข้อสอบที่ทำให้เกิดปัญหาความยุติธรรมระหว่างข้อสอบกับกลุ่มผู้สอบต่างๆ ที่มีลักษณะบางอย่างแตกต่างกัน การทำหน้าที่ต่างกันของข้อสอบมักเกิดจากการปฏิสัมพันธ์ (Interaction) ระหว่างลักษณะของข้อสอบกับลักษณะของกลุ่มผู้สอบ ที่มีผลต่อคะแนน และลักษณะข้อสอบเช่น การใช้ภาษา ความเฉพาะของเนื้อหา โครงสร้างของการวัด วิธีการที่เลือกใช้ เป็นลักษณะของกลุ่มผู้สอบ เช่น เชื้อชาติ ศาสนา วัฒนธรรม ภูมิลำเนา สังคม เพศ ภาษา อายุ ประสบการณ์ เป็นต้น เมื่อทั้งสองปัจจัยมีปฏิสัมพันธ์ต่อกันจะทำให้ข้อสอบข้อนั้นเกิดการทำหน้าที่แตกต่างกัน (Differential Item Functioning : DIF) (ศิริชัย กาญจนวาสี, 2555 : 116) จึงมีผลต่อคะแนนของข้อสอบข้อนั้น อันอาจนำไปสู่ความลำเอียงของข้อสอบด้านต่าง ๆ เช่น ความลำเอียงด้านเพศ ความลำเอียงด้านเชื้อชาติ เป็นต้น เพื่อพัฒนาแบบสอบให้มีคุณภาพเหมาะสมสำหรับนำไปใช้ทดสอบต่อไป วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ที่นิยมใช้ตั้งแต่เริ่มศึกษาการทำหน้าที่ต่างกันของข้อสอบ จนถึงปัจจุบันก็ยังได้รับความนิยมอยู่ คือ วิธีแมนเทิล-แฮนส์เชล วิธี SIBTEST และวิธีถดถอย โลจิสติก ทุกวิธีตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ แต่จะมีข้อดี ข้อเสีย ข้อจำกัด ขนาดของกลุ่มตัวอย่าง ความยาวของข้อสอบ และลักษณะของเครื่องมือ ของแต่ละวิธีแตกต่างกันออกไป ดังนั้นผู้วิจัยควรศึกษาให้เข้าใจและเลือกวิธีที่เหมาะสม เพื่อความถูกต้องและประสิทธิภาพของการศึกษา ในการวิจัยครั้งนี้ผู้วิจัยเลือกวิธีถดถอยโลจิสติก เพราะเหมาะกับการสอบที่มีผู้สอบจำนวนมาก ๆ เป็นวิธีที่สอดคล้องกับธรรมชาติของการทำหน้าที่ต่างกันของข้อสอบได้ดี สามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูปและอเนกรูปได้ดี

การวัดความสามารถพื้นฐานของนักเรียนชั้นประถมศึกษาปีที่ 3 ได้ดำเนินการจัดทำแบบทดสอบวัดความสามารถพื้นฐานที่สำคัญ 3 ด้าน คือ ด้านภาษา (Literacy) ด้านคำนวณ (Numeracy) และด้านเหตุผล (Reasoning abilities) ซึ่งถือเป็นความสามารถพื้นฐานเบื้องต้นสำคัญที่ใช้ในการเรียนรู้ในระดับที่สูงขึ้น สำหรับการประเมินความสามารถด้านเหตุผลของผู้เรียนนั้น มีผลการประเมินหลายแหล่งและกระแสวิชาชีพวิจารณ์ว่า เด็กไทยควรได้รับการพัฒนาทักษะด้านการคิดวิเคราะห์ สังเคราะห์ และการให้เหตุผล

ดังนั้นจึงมีการประเมินความสามารถด้านเหตุผลของผู้เรียนด้วย โดยใช้บริบทหรือสถานการณ์เกี่ยวกับวิทยาศาสตร์และสิ่งแวดล้อม สังคมศาสตร์และ เศรษฐศาสตร์รวมถึงการดำรงชีวิต (สำนักทดสอบทางการศึกษา, 2556 : 3)

สำหรับการทดสอบวัดความสามารถพื้นฐานของผู้เรียนระดับชาติเพื่อการประเมินคุณภาพการศึกษาในระดับประเทศ เป็นการประเมินเพื่อกำกับ ติดตาม และควบคุม คุณภาพการจัดการศึกษาขั้นพื้นฐานของประเทศ ซึ่งผลการประเมินยังสะท้อนสู่คุณภาพการดำเนินการจัดการศึกษาของสถานศึกษา และเขตพื้นที่การศึกษา ในการหาคุณภาพของข้อสอบ ด้านความตรง ความเที่ยงและค่าอำนาจจำแนกของข้อสอบนั้นผู้วิจัยหลายท่านได้ศึกษาไปมากแล้ว แต่ในด้านของความยุติธรรมของข้อสอบ ยังมีการศึกษาน้อย ผู้วิจัยจึงสนใจศึกษาการทำหน้าที่ต่างกันของข้อสอบ ในความแตกต่างกันทางด้านเพศ ซึ่งอาจจะส่งผลให้มีการได้เปรียบเสียเปรียบกันเกิดขึ้นในการทำข้อสอบหรือแบบทดสอบ สำหรับแบบทดสอบที่ใช้มีลักษณะเป็นข้อสอบแบบเลือกตอบ 4 ตัวเลือก ทดสอบกับนักเรียนชั้นประถมศึกษาปีที่ 3 โดยสำนักทดสอบทางการศึกษาของสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน ด้วยเหตุนี้ทำให้ผู้วิจัยสนใจที่จะศึกษาการทำหน้าที่ต่างกันของข้อสอบด้านเหตุผลระหว่างนักเรียนชายกับนักเรียนหญิง โดยใช้การวิเคราะห์จากผลการทดสอบระดับชาติ ระดับชั้นประถมศึกษาปีที่ 3 ปีการศึกษา 2557 ด้วยวิธีของโลจิสติก โดยจำแนกตามตัวแปรเพศ เพื่อเป็นแนวทางในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบต่อไป

### วัตถุประสงค์การวิจัย

เพื่อวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบวัดความสามารถด้านเหตุผลของแบบทดสอบระดับชาติระหว่างนักเรียนชายและนักเรียนหญิง

### วิธีการดำเนินการวิจัย

การวิจัยครั้งนี้เป็นวิจัยเชิงสำรวจ ประชากรที่ใช้ในการศึกษา คือ นักเรียนชั้นประถมศึกษาปีที่ 3 ปีการศึกษา 2557 จำนวน 464,532 คน เป็นนักเรียนชาย 224,615 คน และเป็นนักเรียนหญิง 239,917 คน

### เครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้ในการวิจัย คือ แบบทดสอบวัดความสามารถด้านเหตุผล ชั้นประถมศึกษาปีที่ 3 ของสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน ปีการศึกษา 2557 จำนวน 27 ข้อ เครื่องมือที่ใช้ในการวิจัยครั้งนี้ได้ผ่านกระบวนการให้ผู้เชี่ยวชาญจากคณะกรรมการในการออกข้อสอบเพื่อหาคุณภาพของข้อสอบเป็นที่เรียบร้อยแล้ว ทั้งนี้ผู้วิจัยได้ให้ผู้เชี่ยวชาญที่มีความสามารถด้านเนื้อหา มีประสบการณ์ในการสอนวิทยาศาสตร์ คณิตศาสตร์ ประสบการณ์ในการออกข้อสอบด้านเหตุผล อีกทั้งมีความเชี่ยวชาญในการวัดและประเมินผลการศึกษา จำนวน 3 ท่าน ในการวิเคราะห์สาเหตุการทำหน้าที่ต่างกันของข้อสอบต่อไป

## การวิเคราะห์ข้อมูล

ผู้วิจัยดำเนินการวิเคราะห์ข้อมูลตามขั้นตอน ดังนี้

1. การกำหนดกลุ่มผู้เข้าสอบโดยกำหนดกลุ่มเพศชายเป็นกลุ่มอ้างอิง (Reference group) และกลุ่มเพศหญิงเป็นกลุ่มเปรียบเทียบ (Focal group)
2. การกำหนดค่าของตัวแปรที่ใช้ในการศึกษา เพศชายเป็น 1 และเพศหญิงเป็น 0 เพื่อคำนวณหาค่าสถิติต่าง ๆ
3. การวิเคราะห์การแจกแจงของคะแนนและคุณภาพของแบบทดสอบรายข้อ โดยนำคะแนนที่ได้จากการทดสอบหาค่าสถิติต่างๆ ได้แก่ ค่าเฉลี่ย (mean) ส่วนเบี่ยงเบนมาตรฐาน (standard deviation) ค่าความเบ้ (skewness) ค่าความโด่ง (kurtosis) ค่าความยาก (difficulty) และ ค่าอำนาจจำแนก (discrimination)
4. การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบวัดความสามารถด้านเหตุผลระหว่างนักเรียนชายและนักเรียนหญิง

การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธีถดถอยโลจิสติก ของ Bruno and Zumbo (1999 : 26) เป็นการทดสอบค่าทำนายสำคัญทางสถิติ ด้วยสถิติไค - สแควร์ และมีการกำหนดวิธีการศึกษาเป็นรูปแบบโมเดลที่เป็นขั้น ๆ (step) โดยมีขั้นสำหรับการทดสอบ 3 ขั้น ดังนี้

1) ขั้นที่ 1 การกำหนดตัวแปรควบคุมหรือเลือกตัวแปรควบคุม (คะแนนรวมของผู้สอบ) เพื่อหาค่าความสามารถในการตอบข้อสอบถูกแต่ละข้อ สามารถแสดงเป็นสมการการวิเคราะห์ดังนี้

$$Y = a + b_1TOT$$

โดยที่  $Y$  = คะแนนรวมทำนายการตอบถูกแต่ละข้อ (ตัวแปรตาม)

$a$  = ค่าพารามิเตอร์จุดตัด

$b_1$  = อิทธิพลของความสามารถของผู้สอบ(TOT) ต่อการตอบข้อสอบถูก

TOT = ความสามารถของผู้สอบ

2) ขั้นที่ 2 การนำตัวแปรเพศเข้าสู่สมการโลจิสติก เพื่อศึกษาว่าเพศมีผลต่อการตอบถูกของแต่ละข้อ ผู้วิจัยได้กำหนดตัวแปรเพศชาย(1) เป็นกลุ่มอ้างอิง และเพศหญิง(0) เป็นกลุ่มเปรียบเทียบเพื่อเข้าสู่สมการการวิเคราะห์ดังนี้

$$Y = a + b_1TOT + b_2 GENDER$$

โดยที่  $Y$  = คะแนนรวมทำนายการตอบถูกแต่ละข้อ(ตัวแปรตาม)

$a$  = ค่าพารามิเตอร์จุดตัด

TOT = ความสามารถของผู้สอบ

GENDER = เพศชาย

- $b_1$  = อิทธิพลของความสามารถของผู้สอบ(TOT) ต่อการตอบข้อสอบถูก  
 $b_2$  = ความแตกต่างของการตอบถูกระหว่างนักเรียนชายและนักเรียนหญิง

จากสมการ ข้อสอบที่ทำหน้าที่ต่างกันของข้อสอบหรือลำเอียง เมื่อควบคุมคะแนนรวม มีหลักการพิจารณาดังนี้

2.1) กรณีที่ค่าสัมประสิทธิ์ของ  $b_2$  มีค่าเป็นบวกแสดงว่าข้อสอบข้อนั้นผู้ชายมีโอกาสตอบถูกมากกว่าผู้หญิงเมื่อควบคุมคะแนนรวม

2.2) กรณีที่ค่าสัมประสิทธิ์ของ  $b_2$  มีค่าเป็นลบแสดงว่าข้อสอบข้อนั้นผู้หญิงมีโอกาสตอบถูกมากกว่าผู้ชายเมื่อควบคุมคะแนนรวม

2.3) ถ้า  $b_2$  ไม่ sig แสดงว่าผู้หญิงกับผู้ชายมีโอกาสตอบถูกไม่แตกต่างกันนั้น หมายถึงข้อสอบข้อนั้นมีความยุติธรรมกับผู้สอบ

3) ขั้นที่ 3 เป็นการศึกษาการทำหน้าที่ต่างกันของข้อสอบ ว่าเป็นการทำหน้าที่ต่างกันแบบใด ระหว่างแบบเอกรูปกับอนเอกรูป มีสมการการวิเคราะห์ดังนี้

$$Y = a_1 + b_1TOT + b_2GENDER + b_3(GENDER*TOT)$$

โดยที่  $Y$  = คะแนนรวมทำนายการตอบถูกแต่ละข้อ(รายข้อ)หรือคะแนนรายข้อเป็นตัวแปรตาม

$a$  = ค่าพารามิเตอร์จุดตัด

TOT = ความสามารถของผู้สอบ

GENDER = เพศชาย

GENDER\*TOT = ปฏิสัมพันธ์ระหว่างความสามารถของผู้สอบกับเพศชาย

$b_1$  = อิทธิพลของความสามารถของผู้สอบ(TOT) ต่อการตอบข้อสอบถูก

$b_2$  = ความแตกต่างของการตอบถูกระหว่างนักเรียนชาย  
และนักเรียนหญิง

$b_3$  = ปฏิสัมพันธ์ระหว่างความสามารถของผู้สอบ(คะแนนรวม)กับเพศชายต่อการตอบข้อสอบถูก

4.1 หลักการจำแนกรูปแบบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูป (Uniform) และแบบอนเอกรูป (Non-uniform) มีหลักการพิจารณา Bruno and Zumbo (1999 : 27) ดังต่อไปนี้

1) การทำหน้าที่ต่างกันแบบเอกรูป พิจารณาจากการเปรียบเทียบค่า  $\chi^2$  ของขั้นที่ 1 กับขั้นที่ 2 ถ้า  $\Delta \chi^2_{\text{คำนวณ}} > \Delta \chi^2_{\text{ตาราง}}$  ยอมรับ ขั้นที่ 2 แสดงว่าข้อสอบข้อนั้นเป็นเอกรูป แต่ถ้า  $\Delta \chi^2_{\text{คำนวณ}} < \Delta \chi^2_{\text{ตาราง}}$  ยอมรับ ขั้นที่ 1 แสดงว่าข้อสอบข้อนั้นไม่ทำหน้าที่ต่างกัน คือผู้หญิงและผู้ชายมีโอกาสตอบถูกไม่ต่างกัน

2) การทำหน้าที่ต่างกันแบบอเนกรูป พิจารณาจากการเปรียบเทียบค่า  $\chi^2$  ของชั้นที่ 2 กับชั้นที่ 3 ถ้า  $\Delta \chi^2_{\text{คำนวณ}} > \Delta \chi^2_{\text{ตาราง}}$  ยอมรับ ชั้นที่ 3 แสดงว่าข้อสอบข้อนั้นเป็นอเนกรูป แต่ถ้า  $\Delta \chi^2_{\text{คำนวณ}} < \Delta \chi^2_{\text{ตาราง}}$  ยอมรับ ชั้นที่ 2 แสดงว่าข้อสอบข้อนั้นเป็นเอกรูป

#### 4.2 การพิจารณาขนาดของการทำหน้าที่ต่างกันของข้อสอบ (effect size)

การหาขนาดของการทำหน้าที่ต่างกันของข้อสอบโดยใช้ค่าสถิติ  $R^2$  ซึ่งจะตัดสินข้อสอบที่ทำหน้าที่ต่างกันเมื่อขนาดของการทำหน้าที่ต่างกันของข้อสอบที่มีขนาดปานกลางและขนาดใหญ่ ใช้เกณฑ์ของ Jodoin and Gierl (2001 : 335) เพื่อตัดสินขนาดของ DIF 3 ระดับ ได้แก่

DIF ขนาดเล็กน้อย เมื่อมีค่าความแตกต่าง  $\Delta R^2 < .035$

DIF ขนาดปานกลาง เมื่อมีค่าความแตกต่าง  $.035 \leq \Delta R^2 \leq .07$

DIF ขนาดใหญ่ เมื่อมีค่าความแตกต่าง  $\Delta R^2 > .26$

### ผลการวิจัย

#### ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้านเหตุผลของแบบทดสอบระดับชาติระหว่างนักเรียนชายและนักเรียนหญิง

การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีถดถอยโลจิสติก แบบทดสอบด้านเหตุผล จำนวน 27 ข้อ ผลปรากฏว่า ข้อสอบทุกข้อทำหน้าที่ต่างกันของข้อสอบ และยังจำแนกการทำหน้าที่ต่างกันได้ 2 แบบ คือ แบบเอกรูปและแบบอเนกรูป ข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูป จำนวน 2 ข้อ คือ ข้อ 2 และ 10 คือนักเรียนชายมีโอกาสตอบถูกมากกว่านักเรียนหญิงทุกระดับความสามารถและข้อสอบที่ทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป จำนวน 25 ข้อ คือ ข้อ 1 3 4 5 6 7 8 9 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 และ 27 พบว่า โอกาสในการตอบข้อสอบถูกระหว่างนักเรียนชายและนักเรียนหญิงไม่สม่ำเสมอ ข้อสอบข้อนี้นักเรียนชายบางระดับความสามารถมีโอกาสตอบถูกได้ดีกว่านักเรียนหญิงและนักเรียนหญิงบางระดับความสามารถมีโอกาสตอบถูกได้ดีกว่านักเรียนชาย และเมื่อพิจารณาขนาดการทำหน้าที่ต่างกันของข้อสอบพบว่า ทุกข้อมีขนาดของการทำหน้าที่ต่างกันเพียงขนาดเดียว คือ ขนาดเล็ก ดังนั้นเมื่อพิจารณาขนาดของอิทธิพลแล้วสรุปว่าข้อสอบทุกข้อทำหน้าที่ไม่แตกต่างกัน แสดงดังตารางที่ 1

ตารางที่ 1 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวัดความสามารถด้านเหตุผลของแบบทดสอบระดับชาติระหว่างนักเรียนชายและนักเรียนหญิง

ข้อ	$b_2$	กลุ่มที่ได้เปรียบ	$\chi^2$		รูปแบบของ DIF	$R^2$		ขนาดของ DIF
			$\chi_2^2 - \chi_1^2$	$\chi_3^2 - \chi_2^2$		$R_2^2 - R_1^2$	$R_3^2 - R_2^2$	
1	0.35	ชาย	3090.43*	190.67*	อเนกรูป	0.01	0.00	เล็ก
2	0.08	ชาย	143.82*	0.04	เอกรูป	0.00	0.00	เล็ก



3	0.02	ชาย	8.36*	48.47*	อเนกรูป	0.00	0.00	เล็ก
4	-0.02	หญิง	7.63*	25.65*	อเนกรูป	0.00	0.00	เล็ก
5	0.13	ชาย	357.88*	10.93*	อเนกรูป	0.00	0.00	เล็ก
6	0.04	ชาย	42.65*	9.34*	อเนกรูป	0.00	0.00	เล็ก
7	-0.11	หญิง	310.66*	639.23*	อเนกรูป	0.00	0.00	เล็ก
8	0.33	ชาย	2483.64*	15.29*	อเนกรูป	0.01	0.00	เล็ก
9	-0.06	หญิง	71.38*	78.82*	อเนกรูป	0.00	0.00	เล็ก
10	0.31	ชาย	2215.85*	2.43	เอกรูป	0.01	0.00	เล็ก
11	0.22	ชาย	35.73*	178.62*	อเนกรูป	0.00	0.00	เล็ก
12	0.20	ชาย	1191.07*	95.04*	อเนกรูป	0.00	0.00	เล็ก
13	0.18	ชาย	3180.36*	272.37*	อเนกรูป	0.01	0.00	เล็ก
14	-0.04	หญิง	29.74*	10.86*	อเนกรูป	0.00	0.00	เล็ก
15	0.06	ชาย	74.65*	136.17*	อเนกรูป	0.00	0.00	เล็ก
16	0.04	ชาย	39.20*	330.09*	อเนกรูป	0.00	0.00	เล็ก
17	0.13	ชาย	392.32*	62.54*	อเนกรูป	0.00	0.00	เล็ก
18	0.02	ชาย	5.53*	71.66*	อเนกรูป	0.00	0.00	เล็ก
19	-0.02	หญิง	10.02*	11.40*	อเนกรูป	0.00	0.00	เล็ก
20	-0.24	หญิง	1120.81*	460.30*	อเนกรูป	0.00	0.00	เล็ก
21	0.10	หญิง	236.98*	510.46*	อเนกรูป	0.00	0.00	เล็ก
22	-0.01	หญิง	2.86	136.72	อเนกรูป	0.00	0.00	เล็ก
23	-0.25	หญิง	1238.60*	332.69*	อเนกรูป	0.00	0.00	เล็ก
24	0.30	ชาย	1864.67	311.57*	อเนกรูป	0.01	0.00	เล็ก
25	-0.28	หญิง	1528.67*	199.28*	อเนกรูป	0.00	0.00	เล็ก
26	-0.24	หญิง	1193.25*	1087.66*	อเนกรูป	0.00	0.01	เล็ก
27	-0.10	หญิง	223.38*	586.37*	อเนกรูป	0.00	0.00	เล็ก

\* $\chi^2_{ตาราง} = 3.84$ , ทุกข้อมีค่า  $p < .05$

### สรุปและอภิปรายผลการวิจัย

#### สรุปผลการวิจัย

การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้านเหตุผลระหว่างนักเรียนชายและนักเรียนหญิงจากผลการทดสอบระดับชาติ โดยใช้วิธีถดถอยโลจิสติก ได้ผลดังนี้

1. ผลการวิเคราะห์การทำหน้าที่ต่างกันของนักเรียนชายและนักเรียนหญิงโดยพิจารณาควบคู่กับขนาดของการทำหน้าที่ต่างกันของข้อสอบ



1.1 ข้อสอบที่มีการทำหน้าที่ต่างกันทางนักเรียนชายหรือนักเรียนหญิงเมื่อพิจารณาจากตารางได้แก่

1.1.1 ข้อสอบที่มีการทำหน้าที่ต่างกันไปทางนักเรียนชาย คือ 1 2 5 6 8 10 11 15 16 17 18 21 และ 24

1.1.2 ข้อสอบที่มีการทำหน้าที่ต่างกันไปทางนักเรียนหญิง คือ 3 4 7 9 12 13 14 19 20 22 23 25 26 และ 27

1.2 ขนาดของการทำหน้าที่ต่างกันของข้อสอบ พบว่า ทุกข้อมีขนาดของการทำหน้าที่ต่างกันเพียงขนาดเล็ก

1.3 การทำหน้าที่ต่างกันของข้อสอบมี 2 แบบ คือ

1.3.1 ข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูป จำนวน 2 ข้อ คือ 2 และ 10 คือทุกระดับความสามารถนักเรียนชายมีโอกาสตอบถูกมากกว่านักเรียนหญิง

1.3.2 แบบอนเอกรูป จำนวน 25 ข้อ 1 3 4 5 6 7 8 9 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 พบว่าโอกาสในการตอบข้อสอบถูกระหว่างนักเรียนชายและนักเรียนหญิงไม่สม่ำเสมอเช่นนักเรียนชายบางระดับความสามารถมีโอกาสตอบถูกได้มากกว่านักเรียนหญิงและนักเรียนหญิงบางระดับความสามารถมีโอกาสตอบถูกได้มากกว่านักเรียนชาย

ดังนั้นจึงสรุปได้ว่า ข้อสอบทุกข้อทำหน้าที่ไม่แตกต่างกัน และมีขนาดของการทำหน้าที่ต่างกันของข้อสอบมีขนาดเล็กซึ่งถือว่ายังอยู่ในเกณฑ์ดีและข้อสอบวัดความสามารถด้านเหตุผลของการทดสอบระดับชาติมีความยุติธรรมทั้งนักเรียนชายและนักเรียนหญิง

### อภิปรายผล

ผู้วิจัยขออภิปรายในประเด็นการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบในการตรวจสอบผลการทำหน้าที่ต่างกันของข้อสอบวัดความสามารถด้านเหตุผลของแบบทดสอบระดับชาติ พบว่าเมื่อคำนวณด้วยวิธีถดถอยโลจิสติกพิจารณาค่า  $\chi^2$  ของ Bruno and Zumbo (1999 : 26) แล้วพบข้อสอบที่ทำหน้าที่ต่างกันของข้อสอบเป็นจำนวนมาก แต่ถ้าพิจารณาร่วมกับขนาดของการทำหน้าที่ต่างกันของข้อสอบของ Jodoin and Gierl พบว่าขนาดของการทำหน้าที่ต่างกัน มีขนาดเล็ก ซึ่งไม่มีอิทธิพลหรือส่งผลต่อการทำหน้าที่ต่างกันของข้อสอบ ดังนั้นจึงสรุปได้ว่า ไม่พบข้อสอบที่ทำหน้าที่ต่างกันหรือทำหน้าที่ไม่แตกต่างกัน อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 มีความยุติธรรมทั้งนักเรียนชายและนักเรียนหญิง ซึ่งสอดคล้องผลการวิจัยของ Jodoin and Gierl (2001 : 329 - 349) ได้ศึกษาประสิทธิภาพของการวัดขนาดอิทธิพลในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ พบว่า การวัดขนาดอิทธิพลโดยสถิติ R ร่วมกับการทดสอบนัยสำคัญจะได้ค่าที่ลดลงจนเกือบจะเป็นศูนย์ ทั้งที่ความเป็นจริงข้อสอบไม่ได้ทำหน้าที่ต่างกัน ( NO DIF ) ( False Positive:FP) โดยเฉพาะอย่างยิ่งข้อสอบยังมีความยาวเพิ่มขึ้น FP ยิ่งใกล้ศูนย์และในทางกลับกันการทดสอบนัยสำคัญของสถิติเพียง

อย่างเดียวกันทำให้ได้ FP สูงกว่าเล็กน้อยหรือใกล้เคียงจากค่าปกติทั่วไป และสอดคล้องผลการวิจัยของเกียรติกมล ทองงอก ( 2554 : 203) ที่พบว่าผลการวัดระดับนัยสำคัญพบข้อสอบที่ทำหน้าที่ต่างกัน 36 ข้อ ถ้ายึดสาระสนเทศนี้เป็นหลัก ทำให้ต้องตัดข้อสอบออกจากแบบสอบเป็นจำนวนมาก สาระสนเทศดังกล่าวบอกให้ทราบเพียงว่าข้อสอบเกิดการทำหน้าที่ต่างกันแต่ไม่รู้สาระสนเทศอื่นเลย นั่นก็ถือว่ายังไม่เพียงพอสำหรับการตัดสินใจว่าจะตัดข้อสอบออกจากแบบสอบหรือคงข้อสอบข้อนั้นไว้ในแบบสอบ ส่วนการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl พบข้อสอบที่ทำหน้าที่ต่างกันที่มีขนาดอิทธิพลของการทำหน้าที่ต่างกันขนาดกลาง จำนวน 2 ข้อ พบข้อสอบที่ทำหน้าที่ต่างกันที่มีขนาดอิทธิพลของการทำหน้าที่ต่างกันขนาดใหญ่ จำนวน 2 ข้อ ซึ่งทั้ง 4 ข้อ ให้ผลตรงกับผลการวัดระดับนัยสำคัญว่าข้อสอบทำหน้าที่ต่างกัน หากพิจารณาผลการตรวจสอบด้วยวิธีถดถอยโลจิสติก โดยการวัดระดับนัยสำคัญ ร่วมกับผลการตรวจสอบด้วยขนาดอิทธิพลที่มีอิทธิพลระดับขนาดปานกลางขึ้นไป พบว่าต้องพิจารณาดำเนินการปรับปรุงข้อสอบแล้วตรวจสอบคุณภาพใหม่จากแบบสอบเพียง 4 ข้อ จากผลการวิจัยการทำหน้าที่ต่างกันของข้อสอบของแบบทดสอบวัดความสามารถด้านเหตุผล ชั้นประถมศึกษาปีที่ 3 พบว่านักเรียนชายและนักเรียนหญิงมีโอกาสตอบข้อสอบได้ไม่แตกต่างกันทุกระดับความสามารถ หมายความว่านักเรียนหญิงและนักเรียนชายมีโอกาสตอบข้อสอบถูกเท่ากัน ซึ่งแสดงให้เห็นว่าแบบทดสอบวัดความสามารถด้านเหตุผล ชั้นประถมศึกษาปีที่ 3 มีความยุติธรรมทุกคน สามารถนำผลการทดสอบไปใช้อ้างอิงพัฒนาการของผู้เรียน ได้และถือว่าแบบทดสอบมีคุณภาพสามารถนำไปเก็บในคลังข้อสอบเพื่อลดค่าใช้จ่ายในการออกข้อสอบ

ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการถดถอยโลจิสติก ของ Bruno and Zumbo (1999 : 27) นี้ถือว่าเป็นวิธีการที่ถูกออกแบบมาสำหรับตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันทั้งแบบเอกรูปและอนเอกรูป และการวัดขนาดการทำหน้าที่ต่างกันของข้อสอบได้ถูกพัฒนาขึ้นจากวิธีถดถอยโลจิสติก การตัดสินใจขนาดของการทำหน้าที่ต่างกันของข้อสอบของ Jodoin and Gierl (2001 : 335) ถือว่าเป็นวิธีที่ดีเข้าใจง่ายและสะดวกในการวิเคราะห์ข้อสอบ การวิเคราะห์ข้อสอบสามารถวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและขนาดอิทธิพลของข้อสอบไปพร้อมกันได้ ซึ่งถือว่าลดขั้นตอนของการวิเคราะห์ข้อสอบได้

### ข้อเสนอแนะในการนำผลการวิจัยไปใช้

การวิจัยครั้งนี้เป็นการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบวัดความสามารถด้านเหตุผลของแบบทดสอบระดับชาติระหว่างนักเรียนชายและนักเรียนหญิง ด้วยวิธีการถดถอยโลจิสติก (Logistic Regression) เป็นลำดับขั้นในการวิเคราะห์ข้อสอบด้วยพารามิเตอร์ของข้อสอบ ซึ่งผลการวิจัยพบว่า ข้อสอบวัดความสามารถด้านเหตุผลของแบบทดสอบระดับชาติ ทำหน้าที่ไม่ต่างกัน แสดงว่านักเรียนชายและนักเรียนหญิงมีโอกาสตอบข้อสอบข้อนั้นถูกไม่ต่างกัน ดังนั้นจากผลการวิเคราะห์ข้อสอบทำให้ได้ข้อมูลสารสนเทศเกี่ยวกับข้อสอบที่ไม่ทำหน้าที่ต่างกันของข้อสอบ ซึ่งเป็นประโยชน์ต่อครูผู้สอนในการพัฒนาข้อสอบและแบบสอบเพื่อยกระดับคุณภาพของสถานศึกษา

การวิจัยครั้งต่อไปควรศึกษาผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบลักษณะทำนองเดียวกันในตัวแปรอื่น เช่น ภาษา เชื้อชาติ ศาสนา ภูมิภาค เป็นเกณฑ์ในการแบ่งกลุ่มเพราะในแต่ละภูมิภาคจะมีภาษาถิ่นและศาสนา ภูมิภาคที่แตกต่างกัน วิเคราะห์พร้อมกันหลาย ๆ ด้าน เพื่อให้ได้ข้อมูลสารสนเทศมากยิ่งขึ้น และยืนยันข้อสอบอีกครั้ง

### เอกสารอ้างอิง

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

สำนักทดสอบทางการศึกษา สำนักงานคณะกรรมการ การศึกษาขั้นพื้นฐาน (2556). **นิยามความสามารถของผู้เรียนด้านภาษา ด้านคำนวณและด้านเหตุผล**. กรุงเทพมหานคร: ชุมนุมสหกรณ์การเกษตรแห่งประเทศไทย.

Gomez-Benito, J.Hidalgo,M.D. and Padilla, J.L.,(2009). **Efficacy of Effect Size Measures in Logistic Regression An Application for Detecting DIF**. Methodology :European Journal of Research Method for the Behavioral and Social Sciences.Vol 5(1), 2009,18-25.

Jodoin, M.G.,and Gierl, M.J.(2001). Evaluating Type I Error and Power Rates Using an Effect Size Measure With the Logistic Regression Procedure for DIF Detection. Applied Measurement in Education,14(4),329 – 349.

Zumbo, B.D. (1999). A handbook on the theory and methods of differential item functioning (DIF) : Logistic regression modeling as a unitary framework for binary and Likert-type item scores. Ottawa, Canada : Directorate of Human Resources Research and Evaluation, Department of National Defense. Retrieved From.