

Detecting Differential Item Functioning of Ordinary National Education Test Mathematics Subject in 2015 Academic year by Test Using MIMIC

Pichitchai Sukaek

M.Ed. (Educational Research and Evaluation), Candidate
Department of Educational Research and Evaluation, Faculty of Education
Prince of Songkla University, Pattani Campus

Jirawat Tansakul

Ph.D. (Educational Research and Evaluation), Lecturer
Department of Educational Research and Evaluation, Faculty of Education
Prince of Songkla University, Pattani Campus

Abstract

The objective of this research was to check Differential Item Functioning (DIF) by using Multiple Indicators and Multiple Causes (MIMIC) when distinguish genders, academic year 2015. Parameters that used in this research were 424,430 grade 12 students and samples were 500 grade 12 students by using multistage stage sampling. The research instrument was test of Ordinary National Education Test (O-NET) grade 12 students, academic year 2015, and Mathematics subject amount 32 items those were minor information.

This research found that 1)When distinguish students by genders revealed that male students had higher average score than female students. Male students had 9.89 average score, 4.75 standard diversion and female students had 9.37 average score, 4.75 standard diversion. 2) Parameter difficulty of Ability Parameters of Ordinary National Education Test (O-NET), Mathematics subject in 2015 academic year had the fewest difficulty .03 and the highest 5.95. Parameter of Discrimination the fewest was .13 and the highest was .73 3) Parameter of Ability Parameters of Ordinary National Education Test (O-NET), Mathematics subject in 2015 academic year had 500 students. The lowest Parameter of Ability Parameters was -1.45 and the highest was 3.60. 4) Analysis of Differential Item Functioning (DIF) of Ordinary National Education Test (O-NET), Mathematics subject by using Multiple Indicators and Multiple Causes (MIMIC) when distinguish by genders, academic year 2015. There were Differential Item Functioning (DIF) amount 8 items; items numbers 3, 6, 10, 15, 17, 18, 23, and 25. The test was partial male amount 8 items and there was not partial female.

Keywords: Differential Item Functioning (DIF), Multiple Indicators and Multiple Causes (MIMIC), Parameter of Ability Parameters

การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบทางการศึกษาระดับชาติ ขั้นพื้นฐาน (O-NET) รายวิชาคณิตศาสตร์ ปีการศึกษา 2558 โดยใช้วิธี MIMIC

พิชิตชัย สุขเอก

ศษ.ม. (การวิจัยและประเมินผลการศึกษา), นักศึกษาปริญญาโท
ภาควิชาประเมินผลและวิจัยทางการศึกษา คณะศึกษาศาสตร์
มหาวิทยาลัยสงขลานครินทร์ วิทยาเขตปัตตานี

จิระวัฒน์ ต้นสกุล

ค.ด. (การวัดและประเมินผลการศึกษา) , อาจารย์
ภาควิชาประเมินผลและวิจัยทางการศึกษา คณะศึกษาศาสตร์
มหาวิทยาลัยสงขลานครินทร์ วิทยาเขตปัตตานี

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยใช้วิธีมิมิค (MIMIC : Multiple Indicators and Multiple Causes) เมื่อจำแนกตามเพศ ปีการศึกษา 2558 ประชากรที่ใช้ในการวิจัยครั้งนี้เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2558 จำนวน 424,430 คน และกลุ่มตัวอย่าง ได้แก่ นักเรียนชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2558 จำนวน 500 คน โดยใช้การสุ่มตัวอย่างแบบหลายขั้นตอน (Multistage Stage sampling) โดยใช้เครื่องมือที่ใช้ในการวิจัยครั้งนี้เป็นแบบทดสอบการทดสอบทางการศึกษาระดับชาติด้านขั้นพื้นฐาน (O-NET) ชั้นมัธยมศึกษาปีที่ 6 ประจำปีการศึกษา 2558 รายวิชาคณิตศาสตร์ จำนวน 32 ข้อ ซึ่งเป็นข้อมูลทุติยภูมิ

ผลการวิจัยพบว่า 1) เมื่อจำแนกนักเรียนผู้เข้าสอบตามเพศ พบว่า เพศชายมีค่าเฉลี่ยคะแนนสูงกว่าเพศหญิง โดยที่เพศชายมีคะแนนเฉลี่ย 9.89 คะแนน มีส่วนเบี่ยงเบนมาตรฐาน 4.75 และเพศหญิงมีคะแนนเฉลี่ย 9.37 คะแนน มีส่วนเบี่ยงเบนมาตรฐาน 4.75 2) ค่าพารามิเตอร์ความยากของแบบทดสอบการทดสอบทางการศึกษาระดับชาติด้านขั้นพื้นฐาน (O-NET) รายวิชาคณิตศาสตร์ ในปีการศึกษา 2558 มีค่าความยากน้อยสุดเท่ากับ .03 และสูงสุดเท่ากับ 5.95 ค่าพารามิเตอร์อำนาจจำแนกมีค่าอำนาจจำแนกน้อยสุดเท่ากับ .13 และสูงสุดเท่ากับ .73 3) ค่าระดับความสามารถของผู้สอบที่ได้ทำแบบทดสอบการทดสอบทางการศึกษาระดับชาติด้านขั้นพื้นฐาน (O-NET) รายวิชาคณิตศาสตร์ ปีการศึกษา 2558 มีจำนวนผู้สอบจำนวน 500 คน มีค่าความสามารถของผู้สอบน้อยสุด -1.45 และสูงสุดเท่ากับ 3.60 4) ผลการตรวจสอบการทำหน้าที่ต่างกันของแบบทดสอบทางการศึกษาระดับชาติด้านขั้นพื้นฐาน (O-NET) รายวิชาคณิตศาสตร์ โดยใช้วิธี MIMIC จำแนกตามเพศ ปีการศึกษา 2558 มีข้อสอบที่ทำหน้าที่ต่างกันของข้อสอบจำนวน 8 ข้อ ได้แก่ 3, 6, 10, 15, 17, 18, 23, 25 ซึ่งข้อสอบลำเอียงเข้าข้างเพศชาย จำนวน 8 ข้อ และไม่มีข้อสอบลำเอียงเข้าข้างเพศหญิง

คำสำคัญ: การทำหน้าที่ต่างกันของข้อสอบ วิธี MIMIC พารามิเตอร์ความสามารถของผู้สอบ

บทนำ

การวัดผลการศึกษาเป็นการตรวจสอบผู้เรียนว่ามีความรู้หรือคุณลักษณะที่ต้องการวัดอยู่ในระดับใด ซึ่งผลที่ได้จากการวัดมีความสำคัญต่อพัฒนาการศึกษา การทดสอบเป็นวิธีการวัดผลการศึกษาวิธีหนึ่งที่นิยมใช้กันมาก เครื่องมือที่ใช้ในการทดสอบที่สำคัญก็คือ ข้อสอบ ชนิดต่างๆ ข้อสอบเป็นเครื่องมือหลักที่ใช้ในการวัดและประเมินผลทางการศึกษา เพื่อตรวจสอบว่าผู้เข้าสอบมีคุณลักษณะแฝงหรือความสามารถอยู่ในระดับใด ในการสร้างและการตรวจสอบคุณภาพของข้อสอบ ซึ่งการสร้างข้อสอบที่ดีต้องคำนึงถึงความตรง ความเชื่อมั่น ความยาก-ง่ายอย่างเหมาะสม ความเป็นปรนัย ด้านความตรงก็เป็นสิ่งสำคัญ ทั้งนี้เพราะความตรงเป็นคุณสมบัติของแบบทดสอบที่แสดงถึงความสามารถในการวัดได้ถูกต้องแม่นยำ (ศิริชัย กาญจนวาสี, 2555) ในการทดสอบแต่ละครั้ง ผู้สอบอาจมีลักษณะแตกต่างกัน เช่น เชื้อชาติ ศาสนา วัฒนธรรม ภูมิฐานะ สังคม เพศ ภาษา อายุ และ ประสบการณ์ เป็นต้น ผู้สอบดังกล่าวอาจไม่ได้ได้รับความยุติธรรมในการทำข้อสอบ โดยข้อสอบบางข้ออาจมีความลำเอียงเข้าข้างผู้สอบกลุ่มย่อยบางกลุ่มของผู้เข้าสอบทั้งหมด ซึ่งทำให้เกิดการได้เปรียบเสียเปรียบระหว่างผู้สอบกลุ่มย่อยด้วยกัน ทั้ง ๆ ที่สอบด้วยข้อสอบข้อเดียวกัน แสดงว่าข้อสอบขาดความตรง สาเหตุดังกล่าวอาจเนื่องมาจากแบบทดสอบไม่ได้วัดความสามารถเป้าหมายที่ต้องการวัดเพียงอย่างเดียว แต่ยังวัดความสามารถแทรกซ้อนที่ไม่ต้องการวัดอีกด้วย ส่วนข้อสอบที่วัดทั้งความสามารถเป้าหมายและความสามารถแทรกซ้อน เรียกว่า ข้อสอบทำหน้าที่ต่างกัน ซึ่งถ้ามีผู้สอบกลุ่มย่อยใดมีความสามารถแทรกซ้อนสูงกว่าก็มีโอกาสในการตอบได้ถูกต้องมากกว่า ทั้ง ๆ ที่ระดับความสามารถเป้าหมายที่ต้องการวัดเท่ากัน จึงส่งผลทำให้ข้อสอบทำหน้าที่ต่างกัน (ศิริชัย กาญจนวาสี, 2555)

การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบเป็นการเปรียบเทียบผลการตอบข้อสอบระหว่างกลุ่มผู้สอบอย่างน้อย 2 กลุ่มขึ้นไป โดยทั่วไปทำการเปรียบเทียบผู้สอบ 2 กลุ่ม ได้แก่ กลุ่มเปรียบเทียบ (Focal Group) เป็นกลุ่มที่คาดว่าจะเป็นกลุ่มที่เสียประโยชน์ในการตอบข้อสอบ และ กลุ่มอ้างอิง (Reference Group) เป็นกลุ่มที่คาดว่าจะได้ประโยชน์จากการตอบข้อสอบได้ถูกต้อง ในการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ จะพบว่า

ข้อสอบสามารถทำหน้าที่ต่างกันได้ 2 ประเภท ได้แก่ การทำหน้าที่ต่างกันของข้อสอบแบบเอกรูป (Uniform) และแบบอนเอกรูป (Nonuniform) 1) ข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูป (Uniform DIF) หมายถึง ข้อสอบที่ทำให้ผู้สอบกลุ่มหนึ่งมีโอกาสในการตอบข้อสอบถูกมากกว่าผู้สอบอีกหนึ่งกลุ่มอย่างสม่ำเสมอ ในทุกระดับความสามารถเมื่อพิจารณาไค่งคุณลักษณะข้อสอบของผู้สอบ 2 กลุ่ม จะพบว่าไม่มีปฏิสัมพันธ์ระหว่างระดับความสามารถของผู้สอบกับการเป็นสมาชิกของกลุ่ม (Group membership) 2) ข้อสอบทำหน้าที่ต่างกันแบบอนเอกรูป (Nonuniform DIF) หมายถึง ข้อสอบที่ทำให้โอกาสในการตอบข้อสอบถูกของผู้สอบระหว่างกลุ่มแตกต่างกันอย่างไม่สม่ำเสมอ ในทุกระดับความสามารถ เมื่อพิจารณาไค่งคุณลักษณะข้อสอบของผู้สอบ 2 กลุ่ม พบว่ามีปฏิสัมพันธ์ร่วมกันระหว่างระดับความสามารถของผู้สอบกับการเป็นสมาชิกของกลุ่ม เช่น ที่ระดับความสามารถหนึ่ง กลุ่มผู้สอบกลุ่ม R มีโอกาสในการตอบข้อสอบถูกมากกว่ากลุ่มผู้สอบกลุ่ม F แต่ที่ระดับความสามารถอีกระดับหนึ่ง กลุ่มผู้สอบกลุ่ม F มีโอกาสในการตอบข้อสอบถูกมากกว่ากลุ่มผู้สอบกลุ่ม R (ศิริชัย กาญจนวาสี, 2555)

โมเดลมิมิค (Multiple Indicators and Multiple Causes : MIMIC) โมเดลที่มีตัวแปรแฝงเพียงตัวแปรเดียว โดยที่ตัวแปรแฝงนั้นได้รับอิทธิพลจากตัวแปรภายนอกสังเกตได้หลายตัวแปร และส่งอิทธิพลไปยังตัวแปรภายในสังเกตได้หลายตัวแปร กล่าวอีกอย่างหนึ่งคือ เป็นโมเดลที่มีคุณลักษณะแฝงที่มีหลายสาเหตุและวัดได้จากตัวบ่งชี้หลายตัว ซึ่งในการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ (DIF) ด้วยวิธี MIMIC พบว่า ข้อสอบสามารถทำหน้าที่ต่างกันได้ 2 ประเภท ได้แก่ การทำหน้าที่ต่างกันของข้อสอบแบบเอกรูป และ แบบอนเอกรูป ถ้าข้อสอบหรือข้อคำถามได้รับอิทธิพลอย่างมีนัยสำคัญจากตัวแปรตัวแปรภายนอก แสดงให้เห็นว่าข้อสอบหรือข้อคำถามขึ้นอยู่กับตัวแปรภายนอกไม่ได้อธิบายตัวแปรแฝงแสดงว่าข้อสอบหรือข้อคำถามข้อนั้นทำหน้าที่ต่างกันของข้อสอบ ซึ่งผู้วิจัยได้สนใจการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบโดย MIMIC เพราะ 1) โมเดล MIMIC เป็นวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบค่อนข้างใหม่ 2) มีการลดจำนวนพารามิเตอร์ในการทดสอบความแตกต่างระหว่างกลุ่ม (Ian A. Carroll, 2014)

3) มีการลดลงของขนาดตัวอย่างให้น้อยลง ในขณะที่รูปแบบอื่น ๆ เช่น การวิเคราะห์องค์ประกอบเชิงยืนยันหลายกลุ่ม (multiple - group CFA models) จำเป็นต้องใช้ขนาดตัวอย่างจำนวนมาก เพื่อให้ได้ผลลัพธ์ที่มีความน่าเชื่อถือ (Woods, 2009) 4) โมเดล MIMIC มีกระบวนการในการวิเคราะห์ที่สะดวกมากยิ่งขึ้น เมื่อมีจำนวนกลุ่มตัวแปรหลายกลุ่ม เมื่อเปรียบเทียบกับรูปแบบอื่น ๆ เช่น การวิเคราะห์องค์ประกอบเชิงยืนยันหลายกลุ่ม (multiple - group CFA models) (Brown, 2006) 5) มีข้อดกของเบื้องต้นเกี่ยวกับขนาดตัวอย่างที่น้อยกว่าเมื่อเปรียบเทียบกับรูปแบบอื่นๆ เช่น การวิเคราะห์องค์ประกอบเชิงยืนยันหลายกลุ่ม (multiple - group CFA models) ที่ข้อดกของต้องมีขนาดตัวอย่างใหญ่ (Brown, 2006) ปัจจุบันได้มีการพัฒนาโปรแกรม Mplus เพื่อให้ให้นักวิจัยมีเครื่องมือสำหรับการวิเคราะห์ข้อมูลด้วยสถิติวิเคราะห์ขั้นสูง ที่ให้ผลการวิเคราะห์ข้อมูลที่มีความถูกต้องมากกว่าสถิติแบบเดิมซึ่งในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยใช้วิธี MIMIC ในครั้งนี้ผู้วิจัยได้ใช้โปรแกรม Mplus เพราะ 1) มีการพัฒนาอย่างต่อเนื่องของโปรแกรม Mplus สามารถออกแบบวิธีการวิเคราะห์ข้อมูลเป็นรูปแบบเดียวกันเกือบทุกโมเดล เพื่อลดความยุ่งยากในขั้นตอนของการวิเคราะห์สถิติขั้นสูงของโปรแกรมดั้งเดิมที่ใช้วิเคราะห์กันอยู่โดยทั่วไป 2) โปรแกรม Mplus ในการเขียนคำสั่งไม่ต้องเสียเวลาจำเมทริกซ์พารามิเตอร์ เหมือนกับในโปรแกรม LISREL และนอกจากนี้ ยังมีงานวิจัยที่เกี่ยวข้องของ Finch (2005) ได้ศึกษาการเปรียบเทียบประสิทธิภาพของโมเดล MIMIC กับการทดสอบโดยวิธี แมนเทล - ฮาเอนเซล (Mantel & Haenszel) และวิธี SIBTEST และวิธีการทดสอบ IRT Likelihood Ratio กับความคลาดเคลื่อนประเภทที่ 1 และอำนาจการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ได้แสดงให้เห็นว่าวิธี MIMIC มีค่าสูงขึ้นและค่าความคลาดเคลื่อนประเภทที่ 1 มีค่าลดลงกับจำนวนข้อสอบ 50 ข้อ นอกจากนี้วิธี MIMIC ยังสามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ Uniform DIF ได้ด้วย

ในการวิจัยครั้งนี้ ผู้วิจัยมุ่งศึกษาการวิเคราะห์คุณภาพของข้อสอบตามทฤษฎีการตอบสนอง (IRT) โดยดำเนินการวิเคราะห์วิธีมิมิค (Multiple Indicators and Multiple Causes : MIMIC) โดยประยุกต์ใช้โปรแกรม Mplus ซึ่งโปรแกรมดังกล่าวสามารถวิเคราะห์สถิติขั้นสูง

ได้ดี โดยศึกษาจากแบบทดสอบการทดสอบทางการศึกษาระดับชาติด้านพื้นฐาน (O-NET) ชั้นมัธยมศึกษาปีที่ 6 ประจำปีการศึกษา 2558 รายวิชาคณิตศาสตร์ จำนวน 32 ข้อ โดยผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบครั้งนี้ จะได้ใช้ตัดสินใจว่าสมควรที่จะตัดข้อสอบที่พบว่ามีการทำหน้าที่ต่างกันออกจากข้อสอบ หรือควรปรับปรุงข้อสอบข้อนั้นเสียใหม่ เพื่อรักษาคุณภาพของข้อสอบฉบับนั้นไว้ให้มีความถูกต้องและมีความยุติธรรมมากยิ่งขึ้น เมื่อจำแนกตามเพศ ปีการศึกษา 2558

วัตถุประสงค์

เพื่อตรวจสอบการทำหน้าที่ต่างกันของข้อสอบทางการศึกษาระดับชาติด้านพื้นฐาน (O-NET) รายวิชาคณิตศาสตร์ ปีการศึกษา 2558 โดยใช้วิธี MIMIC

ขอบเขตของการวิจัย

1. ประชากรและกลุ่มตัวอย่าง

ประชากรที่ใช้ในการวิจัยครั้งนี้เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2558 จำนวน 424,430 คน และกลุ่มตัวอย่าง ได้แก่ นักเรียนชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2558 จำนวน 500 คน โดยใช้การสุ่มตัวอย่างแบบหลายขั้นตอน (Multistage Stage sampling)

2. ตัวแปรที่ศึกษา

2.1 ตัวแปรอิสระ กลุ่มที่ศึกษาการทำหน้าที่ต่างกันของข้อสอบ ได้แก่ เพศ คือ เพศชายและเพศหญิง

2.2 ตัวแปรตาม ได้แก่ การทำหน้าที่ต่างกันของข้อสอบโดยใช้วิธีมิมิค (Multiple Indicators and Multiple Causes : MIMIC)

เครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้ในการวิจัยครั้งนี้เป็นแบบทดสอบการทดสอบทางการศึกษาระดับชาติด้านพื้นฐาน (O-NET) ชั้นมัธยมศึกษาปีที่ 6 ประจำปีการศึกษา 2558 รายวิชาคณิตศาสตร์ จำนวน 32 ข้อ โดยมีคะแนนเต็ม 32 คะแนน

การเก็บรวบรวมข้อมูล

ผู้วิจัยรวบรวมข้อมูลผลการตอบของนักเรียนที่เป็นประชากร ซึ่งเป็นข้อมูลทุติยภูมิของนักเรียนที่ทดสอบทางการศึกษาระดับชาติด้านพื้นฐาน (O-NET) ชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2558 รายวิชาคณิตศาสตร์

แล้วนำมาวิเคราะห์ตามวัตถุประสงค์ของการวิจัย

ศึกษาปีที่ 6 ปีการศึกษา 2558 รายวิชาคณิตศาสตร์

2. การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

(DIF) ด้วยวิธี MIMIC ประยุกต์ใช้โปรแกรม Mplus

การวิเคราะห์ข้อมูล

1. คำนวณค่าสถิติพื้นฐาน ได้แก่ ค่าเฉลี่ย (Mean) ค่าส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation) เพื่อบรรยายลักษณะการแจกแจงของคะแนนของแบบทดสอบทางการศึกษาระดับชาติด้านพื้นฐาน (O-NET) ชั้นมัธยม

สรุปผลการวิจัย

ในการศึกษาครั้งนี้ ผู้วิจัยนำเสนอผลการวิเคราะห์ข้อมูลเป็นตามลำดับ ดังนี้

ตาราง 1 จำนวนผู้เข้าสอบ ค่าต่ำสุด ค่าเฉลี่ย ค่าส่วนเบี่ยงเบนมาตรฐาน ค่าความเบ้และค่าความโด่ง จากคะแนนทดสอบทางการศึกษาระดับชาติด้านพื้นฐาน (O-NET) รายวิชาคณิตศาสตร์ ปีการศึกษา 2558 จำแนกตามเพศ

ตัวแปร	ค่าสถิติ						
	<i>n</i>	<i>Min</i>	<i>Max</i>	$\bar{X}$	<i>SD</i>	<i>Skew</i>	<i>Kurt</i>
เพศชาย	227	2.00	31.00	9.89	5.76	1.89	3.44
เพศหญิง	273	4.00	30.00	9.37	4.75	2.13	4.87
รวม	500	2.00	31.00	9.61	5.23	2.03	4.23

*หมายเหตุ ค่าความเบ้ และค่าความโด่งมีค่าอยู่ระหว่าง -1.96 ถึง 1.96 ลักษณะของการแจกแจงเป็นโค้งปกติ

จากตาราง 1 พบว่า เมื่อจำแนกนักเรียนผู้เข้าสอบตามเพศ พบว่า เพศชายมีค่าเฉลี่ยคะแนนสูงกว่าเพศหญิง โดยที่เพศชายมีคะแนนเฉลี่ย 9.89 คะแนน มีส่วนเบี่ยงเบนมาตรฐาน 4.75 และเพศหญิงมีคะแนนเฉลี่ย 9.37 คะแนน มีส่วนเบี่ยงเบนมาตรฐาน 4.75

ตาราง 2 ผลการประมาณค่าพารามิเตอร์ความยากและอำนาจจำแนกของแบบทดสอบการทดสอบทางการศึกษาระดับชาติด้านพื้นฐาน (O-NET) รายวิชาคณิตศาสตร์ ปีการศึกษา 2558

ข้อสอบ	ผลการประมาณค่าพารามิเตอร์ความยากและอำนาจจำแนกของแบบทดสอบ	
	b_i	a_i
1	.69	.50
2	1.10	.65
3	1.75	.41
4	1.59	.56
5	1.07	.59
6	1.50	.72
7	1.30	.45
8	1.33	.63
9	2.68	.36

ตาราง 2 (ต่อ)

ข้อสอบ	ผลการประมาณค่าพารามิเตอร์ความยากและอำนาจจำแนกของแบบทดสอบ	
	b_i	a_i
10	2.03	.39
11	2.34	.22
12	1.44	.52
13	1.89	.43
14	3.27	.19
15	1.24	.55
16	1.00	.63
17	2.25	.41
18	1.30	.37
19	1.25	.49
20	.10	.64
21	.03	.46
22	.89	.27
23	1.97	.35
24	.83	.37
25	1.57	.51
26	.11	.57
27	2.16	.28
28	1.85	.32
29	2.21	.34
30	5.95	.13
31	.40	.73
32	.49	.52

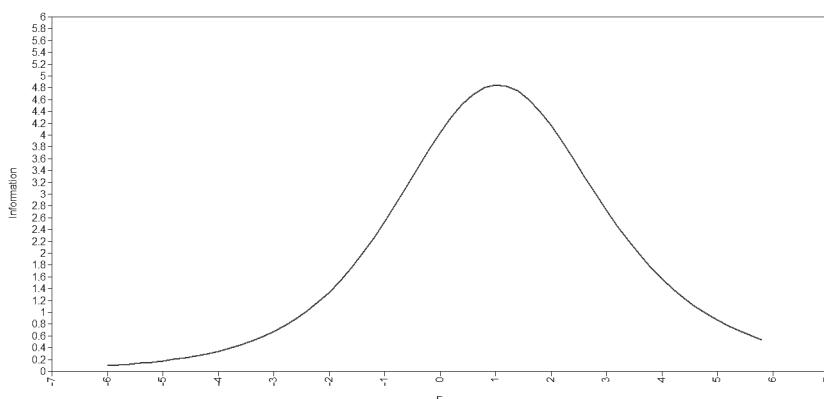
จากตาราง 2 พบว่า ค่าพารามิเตอร์ความยาก 2558 มีค่าความยากน้อยสุดเท่ากับ .03 และสูงสุดเท่ากับ 5.95 ค่าพารามิเตอร์อำนาจจำแนกมีค่าอำนาจจำแนกของแบบทดสอบการทดสอบทางการศึกษาระดับชาติน้อยสุดเท่ากับ .13 และสูงสุด เท่ากับ .73

ตาราง 3 ผลการประมาณค่าระดับความสามารถของผู้สอบที่ได้ทำแบบทดสอบการทดสอบทางการศึกษาระดับชาติน้อยสุดเท่ากับ .13 และสูงสุด เท่ากับ .73

ปีการศึกษา	ผลการประมาณค่าระดับความสามารถของผู้สอบ (θ)		
	n	Min	Max
2558	500	-1.45	3.60

จากตาราง 3 พบว่า ค่าระดับความสามารถของผู้สอบที่ได้ทำแบบทดสอบการทดสอบทางการศึกษาระดับชาติด้านพื้นฐาน (O-NET) รายวิชาคณิตศาสตร์

ปีการศึกษา 2558 มีจำนวนผู้สอบจำนวน 500 คน มีค่าความสามารถของผู้สอบน้อยสุด -1.45 และสูงสุดเท่ากับ 3.60



ภาพประกอบ 1 กราฟแสดงฟังก์ชันสารสนเทศ (Test Information) ของแบบทดสอบการทดสอบทางการศึกษาระดับชาติด้านพื้นฐาน (O-NET) รายวิชาคณิตศาสตร์ ปีการศึกษา 2558

ตาราง 4 ผลการวิเคราะห์ความกลมกลืนของโมเดล CFA วิธี MIMIC ของแบบทดสอบการทดสอบทางการศึกษาระดับชาติด้านพื้นฐาน (O-NET) รายวิชาคณิตศาสตร์ ปีการศึกษา 2558 จำแนกตามเพศ

การวิเคราะห์ความกลมกลืนของโมเดล CFA วิธี MIMIC					
χ^2	<i>df</i>	<i>CFI</i>	<i>TLI</i>	<i>RMSEA</i>	<i>WRMR</i>
1,595.48	780	.56	.54	.04	1.49

จากตาราง 3 พบว่า ผลการวิเคราะห์ความกลมกลืนของโมเดล CFA MIMIC วิชาคณิตศาสตร์จำแนกตามเพศจากการวิเคราะห์ห้อยประกอบเชิงยืนยัน (CFA) ซึ่งพิจารณาจากค่าไคสแควร์ (Chi - square) มีค่าเท่ากับ 1,595.48 ท้องศาอิสระ (df) เท่ากับ 780 ดัชนีวัดระดับความกลมกลืนเปรียบเทียบ (CFI) มีค่าเท่ากับ .56 ดัชนี

The Tucker Lewis Index (TLI) มีค่าเท่ากับ .54 ดัชนีวัดรากที่สองของความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ที่พอเหมาะ (RMSEA) มีค่าเท่ากับ .04 และดัชนี Weighted Root Mean Square Residual (WRMR) มีค่าเท่ากับ 1.49

ตาราง 5 ผลการตรวจสอบการทำหน้าที่ต่างกันของแบบทดสอบทางการศึกษาระดับชาติด้านพื้นฐาน (O-NET) รายวิชาคณิตศาสตร์ ปีการศึกษา 2558 โดยวิธี MIMIC จำแนกตามเพศ

ข้อสอบ	ผลการตรวจสอบการทำหน้าที่ต่างกันของแบบทดสอบ	
	<i>Estimate</i>	<i>P - Value</i>
1	.18	.22
2	.25	.12
3	.33*	.03
4	.23	.16
5	.22	.15
6	.38*	.04
7	.25	.09
8	.26	.10
9	.13	.37
10	.40*	.01
11	.20	.11
12	.24	.12
13	.14	.33
14	.12	.35
15	.52*	.00
16	.10	.55
17	.46*	.00
18	.44*	.00
19	.11	.45
20	.21	.13
21	.23	.09
22	-.02	.89
23	.29*	.04
24	-.10	.43
25	.45*	.01
26	.18	.20
27	.09	.49
28	.03	.85
29	-.03	.85
30	-.12	.35
31	.17	.26
32	.00*	.00

หมายเหตุ * นัยสำคัญทางสถิติที่ระดับ .05

จากตาราง 5 พบว่า ผลการตรวจสอบการทำหน้าที่ต่างกันของแบบทดสอบทางการศึกษาระดับชาติด้านพื้นฐาน (O-NET) รายวิชาคณิตศาสตร์ โดยวิธี MIMIC จำแนกตามเพศ ปีการศึกษา 2558 มีข้อสอบที่ทำหน้าที่ต่างกันของข้อสอบ จำนวน 8 ข้อ ได้แก่ 3, 6, 10, 15, 17, 18, 23, 25 ซึ่งข้อสอบลำเอียงเข้าข้างเพศชาย จำนวน 8 ข้อ และไม่มีข้อสอบลำเอียงเข้าข้างเพศหญิง

อภิปรายผลการวิจัย

จากผลการวิจัย ผู้วิจัยสามารถอภิปรายผลการวิจัยได้ดังนี้

จากการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธี MIMIC จากโปรแกรม Mplus เมื่อจำแนกตามเพศ พบข้อสอบที่ทำหน้าที่ต่างกัน 6 ข้อ คิดเป็นร้อยละ 18.75 ซึ่งถือว่าข้อสอบผิดปกติ จากข้อค้นพบของคลอสเซอร์ที่ว่า แบบทดสอบมาตรฐานที่วัดผลสัมฤทธิ์ทางการเรียนถ้ามีสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันร้อยละ 10 ถึง 15 ถือว่าไม่ปกติ แต่ถ้ามีสัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันร้อยละ 20 ถือว่าเป็นเรื่องผิดปกติอย่างมาก (Clauser, 1993 อ้างใน ศุภวัฒน์ มะลิเผือก, 2548, หน้า 108) เมื่อพิจารณาการทำหน้าที่ต่างกันของข้อสอบแล้วพบว่า ข้อสอบที่ทำหน้าที่ต่างกันเข้าข้างเพศชาย ซึ่งเป็นกลุ่มอ้างอิง ซึ่งสอดคล้องกับงานวิจัยที่ผ่านมา พบว่า Maier และ Casselman (1970 อ้างถึงใน สุพัฒน์ หอมบุปผา, 2556) กล่าวว่าเพศชายจะมีความสามารถในการแก้ปัญหาทางคณิตศาสตร์และวิชาวิทยาศาสตร์เก่งกว่าเพศหญิง ส่วนเพศหญิงจะมีความสามารถในทางภาษาและสังคมศาสตร์เก่งกว่าเพศชาย ซึ่งสอดคล้องกับผลการวิจัยของ ศุภวัฒน์ มะลิเผือก (2548) ที่ได้ศึกษาอิทธิพลของการทำหน้าที่ต่างกันของข้อสอบที่ส่งผลต่อคุณภาพของแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนระดับชาติ

ชั้นประถมศึกษาปีที่ 6 ผลการวิจัยปรากฏว่าพบข้อสอบทำหน้าที่ต่างกันในรูปแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนระดับชาติภาษาไทย เมื่อจำแนกกลุ่มผู้สอบตามตัวแปรเพศ จำนวน 12 ข้อ คิดเป็นร้อยละ 30 และสอดคล้องกับผลการประเมิน PISA 2015 ด้านการอ่าน คณิตศาสตร์ และวิทยาศาสตร์ (สสวท., 2559, หน้า 18) ความแตกต่างระหว่างเพศด้านคณิตศาสตร์ ภาพรวมระดับนานาชาติ ใน PISA 2015 นักเรียนชายมีคะแนนสูงกว่านักเรียนหญิงอย่างมีนัยสำคัญทางสถิติ (8 คะแนน)

ข้อเสนอแนะ

1. ข้อเสนอแนะทั่วไป

1.1 ผู้ที่ออกข้อสอบควรนำผลการวิจัยนี้ไปใช้เป็นข้อมูลในการพิจารณาตัดสินผลการสอบคัดเลือกให้รอบคอบมากยิ่งขึ้น

1.2 การควบคุมการทำหน้าที่ต่างกันของข้อสอบในทางปฏิบัติควรพิจารณาทบทวนข้อสอบ กำจัดความเป็นไปได้ของสาเหตุที่จะทำให้เกิดการทำหน้าที่ต่างกันของข้อสอบ เช่น ภาษาที่ใช้ในการเขียนข้อสอบ ลักษณะเนื้อหาที่นำมาสร้างข้อสอบ ลักษณะเฉพาะของกลุ่มผู้สอบ เป็นต้น

2. ข้อเสนอแนะในการวิจัยครั้งต่อไป

2.1 ควรมีการทำวิจัยเพื่อเปรียบเทียบการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบระหว่างวิธีต่าง ๆ ซึ่งจะตรวจสอบได้ว่าวิธีการใดมีความสามารถในการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบดีกว่ากัน

2.2 ควรมีการศึกษาตัวแปรกลุ่มอื่น ๆ เช่น ภูมิปัญญา ศาสนา เชื้อชาติ ภาษา อายุ เป็นต้น ว่าแบบทดสอบแต่ละวิชาจะส่งผลให้พบข้อสอบทำหน้าที่ต่างกันของข้อสอบ แตกต่างกันหรือไม่

เอกสารอ้างอิง

- นงลักษณ์ วิรัชชัย. (2538). *ความสัมพันธ์โครงสร้างเชิงเส้น (Lisrel) สถิติวิเคราะห์สำหรับการวิจัยทางสังคมศาสตร์ และพฤติกรรมศาสตร์*. กรุงเทพฯ : โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย .
- ศิริชัย กาญจนวาสี. (2555). *ทฤษฎีการทดสอบแนวใหม่ (Model test theories)*. (พิมพ์ครั้งที่ 4). กรุงเทพฯ : โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- ศิริชัย กาญจนวาสี. (2556). *ทฤษฎีการทดสอบแบบดั้งเดิม (Classical test theories)*. (พิมพ์ครั้งที่ 7). กรุงเทพฯ : โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- สุพัฒนา หอมบุปผา. (2556). การเปรียบเทียบการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธี HGLM วิธี MIMIC และวิธี BAYESIAN. *วารสารวิจัยราชภัฏพระนคร*, 8 (2), 10-24.
- ศุภวัฒน์ มะลิเผือก. (2548). *อิทธิพลของการทำหน้าที่ต่างกันของข้อสอบที่ส่งผลต่อคุณภาพของแบบทดสอบ วัดผลสัมฤทธิ์ทางการเรียนระดับชาติ ชั้นประถมศึกษาปีที่ 6. วิทยานิพนธ์ปริญญาโทมหาบัณฑิต สาขาวิชา เทคโนโลยีการวัดทางการศึกษา, มหาวิทยาลัยบูรพา.*
- Woods, C. M., Oltmanns, T. F. & Turkheimer, E. (2009). Illustration of MIMIC-Model DIF Testing with the Schedule for Nonadaptive and Adaptive Personality. *Psychopathol Behav Assess*, p.320-33., C. M. & Grimm, K. J. (2011). *Testing for Nonuniform Differential Item Functioning With Multiple Indicator Multiple Cause Models*. *Applied Psychological Measurement*, p.339-361.
- Finch, Holmes. (2005). The MIMIC Model as a Method for Detecting DIF: Comparison With Mantel - Haenszel, SIBTEST, and the IRT Likelihood Ratio. *Applied Psychological Measurement*, 29, 278-295. DOI: 1.1177/0146621605275728.
- Carroll, L. A. (2014). *MIMIC DIF Testing When the Latent Variable Variance Differs Between Group*. Master's thesis. University of Kansas, Faculty of the Graduate School of Arts.