

การทำหน้าที่ต่างกันของข้อสอบ Pre O-NET วิชาคณิตศาสตร์ โดยใช้วิธีการ
ตอบสนองข้อสอบ(IRT) แบบ 3 พารามิเตอร์ และวิธีไค-สแควร์ (χ^2)
An Analysis of the Differential Item Functioning for Pre Ordinary
National Educational Test Item in Mathematics Using Item Response
Theory (IRT) in Three-Parameter Procedures and Chi-Square Test (χ^2)

เพ็ญอนงค์ อันเวียง¹ ปณณวิชญ์ ไบกุลารบ²
Penanong Onveang, Pun Nawit Baikularb

บทคัดย่อ (Abstract)

การวิจัยในครั้งนี้มีวัตถุประสงค์เพื่อตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยวิธี
ทฤษฎีการตอบสนองข้อสอบ (IRT) แบบ 3 พารามิเตอร์ด้วยวิธีวัดความแตกต่างของค่าพารามิเตอร์
ความยาก(b) และวิธีไค - สแควร์ (χ^2) โดยวิเคราะห์ระหว่างกลุ่มที่จำแนกตามเพศและขนาด
โรงเรียน กลุ่มตัวอย่างที่ใช้ คือ นักเรียนชั้นมัธยมศึกษาปีที่ 3 ได้ทำการทดสอบด้วยข้อสอบ Pre O-
NET วิชาคณิตศาสตร์ ระหว่างปีการศึกษา 2558 - 2560 ในจังหวัดพิษณุโลก จำนวน 39
โรงเรียน จำนวน 12,696 คน ข้อมูลที่ใช้วิเคราะห์ คือ ข้อมูลทุติยภูมิ(Secondary data) ซึ่งเป็น
แบบบันทึกผลการตอบข้อสอบที่ได้จากสำนักงานเขตพื้นที่การศึกษามัธยมศึกษา เขต 39 วิเคราะห์
การทำหน้าที่ต่างกันของข้อสอบ(DIF) ด้วยโปรแกรมสำเร็จรูป MULTILOG, SPSS และ Excel
ผลการวิจัยพบว่า วิเคราะห์โดยใช้ทฤษฎีการตอบสนองข้อสอบ(IRT) แบบ 3 พารามิเตอร์ เมื่อ
จำแนกผู้สอบตามกลุ่มเพศ พบข้อสอบทำหน้าที่ต่างกัน จำนวน 8 ข้อ โดยลำเอียงเข้าข้างเพศชาย
จำนวน 4 ข้อ และลำเอียงเข้าข้างเพศหญิง จำนวน 4 ข้อ เมื่อจำแนกผู้สอบตามกลุ่มขนาดโรงเรียน
พบข้อสอบทำหน้าที่ต่างกัน จำนวน 18 ข้อ โดยลำเอียงเข้าข้างโรงเรียนขนาดใหญ่จำนวน 10 ข้อ
และลำเอียงเข้าข้างโรงเรียนขนาดเล็ก จำนวน 8 ข้อ และวิเคราะห์ด้วยวิธีไค - สแควร์ (χ^2) เมื่อ
จำแนกผู้สอบตามกลุ่มเพศ พบข้อสอบทำหน้าที่ต่างกันจำนวน 11 ข้อ โดยลำเอียงเข้าข้างเพศชาย
จำนวน 8 ข้อ และลำเอียงเข้าข้างเพศหญิง จำนวน 3 ข้อ เมื่อจำแนกผู้สอบตามกลุ่มขนาดโรงเรียน
พบข้อสอบทำหน้าที่ต่างกัน จำนวน 13 ข้อ โดยลำเอียงเข้าข้างโรงเรียนขนาดใหญ่ จำนวน 10 ข้อ
และลำเอียงเข้าข้างโรงเรียนขนาดเล็กจำนวน 3 ข้อ

Received: 2021-09-10 Revised: 2021-09-27 Accepted: 2021-09-28

¹ นักศึกษาระดับปริญญาโท สาขาการวิจัยและประเมินทางการศึกษา คณะครุศาสตร์ มหาวิทยาลัยราช
ภัฏพิบูลสงคราม, M.Ed. Student in Educational Research and Evaluation, Pibulsongkram Rajabhat
University. Corresponding Author e-mail: penanong_33@hotmail.com

² คณะครุศาสตร์ มหาวิทยาลัยราชภัฏพิบูลสงคราม, Faculty of Education, Pibulsongkram
Rajabhat University

คำสำคัญ(Keywords) :การทำหน้าที่ต่างกันของข้อสอบ ;คณิตศาสตร์; ข้อสอบ Pre O-NET

Abstract

This research aimed to examine the differential item functioning using Item Response Theory (IRT) in three-parameter procedures by measuring the difference of parameters, item difficulty (b), and Chi-Square test (χ^2) by analyzing between groups classified by gender and school size. The sample group of this study consisted of 12,696 ninth-grade students from 39 schools under the secondary educational service area office 39 who have tested with Pre O-NET in Mathematics in the academic year 2015 – 2017. The data used for analysis were secondary data, which was an examination results record form obtained from the secondary educational service area office 39. Analyze the Differential Item Functioning (DIF) with the program MULTILOG, SPSS, and Excel programs. The results showed that: An analysis using Item Response Theory (IRT) in three-parameter procedures found that when distinguished by gender, there was differential item functioning for 8 items. The test was partial male for 4 items and partial female for 4 items. When distinguished by school size groups, there was differential item functioning for 18 items. The test was partial big schools for 10 items and partial small schools for 8 items. An analysis using Chi-Square test (χ^2) found that when distinguished by gender, there was differential item functioning for 11 items. The test was partial male for 8 items and partial female for 3 items. When distinguished by school size groups, there was differential item functioning for 13 items. The test was partial big schools for 10 items and partial small schools for 3 items.

Keywords: Differential Item Functioning; Mathematics; Pre O-NET

บทนำ (Introduction)

สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน(สพฐ.) ได้กำหนดนโยบายและมีมติให้นำผลการทดสอบทางการศึกษาระดับชาติด้านพื้นฐาน(Ordinary National Education : O-NET) มาใช้เป็นองค์ประกอบหนึ่งในการจบหลักสูตรการศึกษาขั้นพื้นฐาน เพื่อเป็นการประเมินคุณภาพการศึกษาขั้นพื้นฐานของนักเรียนแต่ละโรงเรียน และเป็นเกณฑ์การวัดผลสัมฤทธิ์ทางการเรียนของผู้เรียนในระดับชั้นประถมศึกษาปีที่ 6 ชั้นมัธยมศึกษาปีที่ 3 และชั้นมัธยมศึกษาปีที่ 6 ขึ้น การสอบ O-NET จึงมีความสำคัญต่อการเรียนมากขึ้น ประกอบกับหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน

พุทธศักราช 2551 ได้กำหนดให้มีการทดสอบระดับชาติและประเมินผลระดับสถานศึกษา ซึ่งเป็นการประเมินคุณภาพผู้เรียนตามมาตรฐานและตัวชี้วัดหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 และสอดคล้องกับสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน(สพฐ.) ซึ่งมีจุดเน้นด้านผู้เรียนในข้อ 1.1 คือนักเรียนมีสมรรถนะสำคัญสู่มาตรฐานสากล โดยมุ่งยกระดับคุณภาพนักเรียนชั้นประถมศึกษาปีที่ 6 ชั้นมัธยมศึกษาปีที่ 3 และชั้นมัธยมศึกษาปีที่ 6 ให้มีผลสัมฤทธิ์ทางการเรียนจากการทดสอบระดับชาติ(O-NET) ในกลุ่มสาระการเรียนรู้หลักมีค่าเฉลี่ยเพิ่มขึ้นไม่น้อยกว่าร้อยละ 3 ดังนั้นสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน(สพฐ.) ได้มอบหมายให้สำนักทดสอบทางการศึกษาดำเนินการสร้างและพัฒนาข้อสอบมาตรฐานวัดผลสัมฤทธิ์ทางการเรียน Pre O-NET (Pre Ordinary National Education) เพื่อให้เขตพื้นที่การศึกษานำไปทดสอบกับนักเรียน ชั้นประถมศึกษาปีที่ 6 และชั้นมัธยมศึกษาปีที่ 3 ก่อนที่จะมีการทดสอบด้วยข้อสอบ O-NET จริง ซึ่งข้อสอบ Pre O-NET เป็นข้อสอบที่วัดในตัวชี้วัดเช่นเดียวกับข้อสอบ O-NET การจัดสอบ Pre O-NET มีวัตถุประสงค์เพื่อ 1) สร้างความตื่นตัวและชักจูงมทัชกะการทำข้อสอบ O-NET ให้กับนักเรียน 2) นักเรียนได้ทราบผลการทดสอบ Pre O-NET ของตนเองว่าอยู่ในระดับใด เพื่อจะได้นำผลไปใช้ในการปรับปรุงพัฒนาการเรียนของตนเองให้ดีขึ้น 3) นักเรียนได้ฝึกทำข้อสอบ ฝึกการคิดวิเคราะห์ ทบทวนและสรุปรวมองค์ความรู้ทั้งหมด 4) นักเรียนเห็นความสำคัญและให้ความสำคัญกับการเรียนและการสอบมากยิ่งขึ้น 5) ทำให้นักเรียนคุ้นเคยกับรูปแบบข้อสอบที่หลากหลาย 6) โรงเรียนได้ทราบผลการสอบของนักเรียนรายบุคคลและโดยภาพรวม เป็นข้อมูลในการนำไปพัฒนาการจัดการเรียนการสอนก่อนสอบ O-NET (คู่มือการใช้ข้อสอบ Pre O-NET, 2557)

การประเมินผลสัมฤทธิ์ดังกล่าวจึงมีความสำคัญต่อทั้งผู้เรียน โรงเรียน เขตพื้นที่การศึกษา และระดับชาติ ดังนั้นในการทดสอบจึงต้องคำนึงถึงคุณภาพของข้อสอบ ซึ่งดัชนีบ่งชี้คุณภาพของเครื่องมือวัดที่สำคัญแบ่งเป็นสองกลุ่มคือ 1) ดัชนีบ่งชี้คุณภาพรายข้อ ประกอบด้วย ความยากง่าย (difficulty) และอำนาจจำแนก (discrimination) 2) ดัชนีคุณภาพทั้งฉบับ ประกอบด้วย ความตรง(validity) และความเที่ยง(reliability) หัวใจสำคัญของคุณภาพแบบทดสอบที่ต้องคำนึงถึงคือ “ความตรง” เพราะความตรงแสดงถึงความสามารถในการวัดได้ถูกต้องแม่นยำและคามยุติธรรมของข้อสอบ(Item and unfairness) ข้อสอบที่ขาดคุณภาพด้านความตรงในประเด็นของคามยุติธรรมของข้อสอบที่เกิดขึ้น เมื่อผู้สอบกลุ่มย่อยต่างกลุ่มกันและมีลักษณะเฉพาะบางอย่างแตกต่างกัน เช่น เพศ ภาษา อายุ เชื้อชาติ ศาสนา วัฒนธรรม ภูมิลาเนา สังคม ประสบการณ์ เป็นต้น ผู้สอบดังกล่าวอาจไม่ได้รับความยุติธรรมในการทำข้อสอบ โดยข้อสอบบางข้ออาจมีความลำเอียงเข้าข้างผู้สอบกลุ่มย่อยบางกลุ่มของผู้เข้าสอบทั้งหมด ซึ่งทำให้เกิดการได้เปรียบเสียเปรียบระหว่างผู้สอบกลุ่มย่อยด้วยกัน ทั้งๆ ที่สอบด้วยข้อสอบข้อเดียวกันแสดงว่าข้อสอบขาดความตรงสาเหตุดังกล่าวอาจเนื่องมาจากแบบทดสอบไม่ได้วัดความสามารถเป้าหมายที่ต้องการวัดเพียงอย่างเดียว แต่ยังวัดความสามารถแทรกซ้อนที่ไม่ต้องการวัดอีกด้วย ส่วนข้อสอบที่วัดทั้งความสามารถเป้าหมายและความสามารถแทรกซ้อน เรียกว่า “ข้อสอบทำหน้าที่ต่างกัน” ซึ่งถ้าผู้สอบกลุ่มย่อย

ใดมีความสามารถแทรกซ้อนสูงกว่าก็มีโอกาสในการตอบได้ถูกต้องมากกว่า ทั้งๆ ที่ระดับความสามารถเป้าหมายที่ต้องการวัดเท่ากัน จึงส่งผลทำให้ข้อสอบทำหน้าที่ต่างกัน เดิมใช้คำว่า “ความลำเอียงของข้อสอบ”(Item bias) ซึ่งต่อมาได้มีการเปลี่ยนมาใช้คำว่า “การทำหน้าที่ต่างกันของข้อสอบ” (Differential Item Functioning : DIF) (ศิริชัย กาญจนวาสี, 2555)

การทำหน้าที่ต่างกันของข้อสอบ มีการศึกษากันอย่างจริงจังในช่วงปลายทศวรรษ 1960 ได้มี ผู้เสนอวิธีการตรวจสอบอยู่หลายวิธี โดยแยกเป็น 3 ประเภท คือ 1) การวัดความเบี่ยงเบนสัมพัทธ์(Relative Deviation) ของข้อสอบแต่ละข้อจากแนวโน้มเข้าสู่ส่วนกลาง ข้อสอบข้อใดที่เบี่ยงเบนไปจากข้ออื่นๆ มากถือว่าเป็นข้อสอบที่น่าจะลำเอียง ได้แก่ วิธีวิเคราะห์ความแปรปรวนและวิธีการแปลงค่าความยากของแองกอฟฟ์(Angoff) หรือวิธีกำหนดจุดเดลต้า 2) การประเมินความตรงเชิงโครงสร้าง ได้แก่ การวิเคราะห์ตัวประกอบ และ 3)การประมาณค่าโอกาสในการตอบข้อสอบถูกของผู้มีความสามารถในสิ่งที่วัดในระดับเดียวกัน ได้แก่ วิธีไค-สแควร์ วิธีทฤษฎีการตอบสนองข้อสอบ วิธีแมนเทิล-แฮนส์เซล และวิธี SIBTEST ซึ่งแต่ละวิธีจะมีทั้งข้อดีและข้อเสียที่แตกต่างกันไปโดยวิธีวิเคราะห์ความแปรปรวน วิธีแปลงค่าความยาก การวิเคราะห์ตัวประกอบ วิธีไค-สแควร์ และวิธีแมนเทิล-แฮนส์เซล กับวิธี SIBTEST เป็นวิธีใหม่ จะยึดแนวทฤษฎีดั้งเดิมมีข้อดีคือ ใช้กลุ่มตัวอย่างน้อย คำนวณและแปลผลค่อนข้างง่าย ข้อเสียคือ ค่าสถิติต่างๆ ของข้อสอบจะแปรเปลี่ยนไปตามกลุ่มตัวอย่าง สำหรับวิธีที่ยึดแนวทฤษฎีการตอบสนองข้อสอบ คือ วิธีทฤษฎีการตอบสนองข้อสอบที่มี 1 , 2 และ 3 พารามิเตอร์ เป็นวิธีที่กำลังได้รับความนิยม โดยมีประสิทธิภาพของการประมาณค่าและใช้โปรแกรมในการวิเคราะห์ห้ง่าย มีข้อดีคือการไม่แปรเปลี่ยนของค่าสถิติตามกลุ่มตัวอย่างและความถูกต้องตามทฤษฎีมากกว่า และมีข้อเสียคือการคำนวณยุ่งยาก ต้องใช้กลุ่มตัวอย่างที่มีขนาดใหญ่(กาญจนา วัฒนสุนทร, 2538)

การทดสอบในแต่ละครั้งผู้สอบอาจมีลักษณะเฉพาะบางอย่างที่แตกต่างกัน ซึ่งในการทดสอบอาจมีข้อสอบบางข้อที่ทำให้กลุ่มย่อยบางกลุ่มที่มีความสามารถระดับเดียวกันมีโอกาสตอบข้อสอบข้อนั้นได้ถูกต้องไม่เท่ากัน ซึ่งอาจมีสาเหตุจากการใช้ภาษาหรือเนื้อหาบางอย่างที่เอื้อประโยชน์กับผู้สอบกลุ่มใดกลุ่มหนึ่ง จึงแสดงให้เห็นว่าข้อสอบข้อนั้นขาดความตรง โดยเฉพาะแบบสอบที่มีภาษาเข้ามาเกี่ยวข้อง เช่น แบบทดสอบวิชาคณิตศาสตร์เป็นวิชาหนึ่งที่มีลักษณะบางข้อเป็นโจทย์ปัญหาที่ประกอบไปด้วยภาษาและตัวเลขที่ต้องการคำตอบ ซึ่งต้องอาศัยการอ่านโจทย์เพื่อนำไปสู่การแปลงเป็นประโยคสัญลักษณ์ทางคณิตศาสตร์เพื่อหาคำตอบ อาจมีข้อความของโจทย์ปัญหาบางข้อความทำให้ผู้สอบเข้าใจไม่ตรงกัน ส่งผลให้เกิดการได้ประโยชน์และเสียประโยชน์ระหว่างผู้สอบกลุ่มย่อย ซึ่งจากการศึกษางานวิจัยที่ผ่านมา พบว่า เพศเป็นตัวแปรที่มีผลต่อการทำหน้าที่ต่างกันของข้อสอบวิชาคณิตศาสตร์ อาทิเช่น อมลณัฐ อุบลรัตน์ (2552) ได้ศึกษาการทำหน้าที่ต่างกันของข้อสอบวัดผลสัมฤทธิ์ทางการเรียนชั้นมัธยมศึกษาปีที่ 2 สำนักงานเขตพื้นที่การศึกษาปัตตานี เขต 2 พบว่า เพศเป็นตัวแปรที่มีผลต่อการทำหน้าที่ต่างกันของข้อสอบในวิชาคณิตศาสตร์ ข้อสอบที่ทำหน้าที่ต่างกันส่วนใหญ่ลำเอียงเข้าข้างกลุ่มเพศชายมากกว่ากลุ่มเพศหญิง (Gierl & Boughton ,2004)ได้ศึกษาการระบุเนื้อหาและทักษะทางพุทธิปัญญาที่ทำให้เกิดความ

แตกต่างทางเพศที่มีต่อวิชาคณิตศาสตร์ โดยใช้กระบวนการทัศน์ของการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบแบบพหุมิติ พบว่า นักเรียนชายทำคะแนนในส่วนของมิติสัมพันธ์ได้ดีกว่านักเรียนหญิง ในขณะที่นักเรียนหญิงทำคะแนนในส่วนทักษะความจำได้ดีกว่านักเรียนชาย พิรญา สูงเนิน(2551) ได้ศึกษาการทำหน้าที่ต่างกันของข้อสอบวัดความคิดเชิงตรรกศาสตร์ที่มีบริบทเกี่ยวกับการเล่นฟุตบอล อาจทำให้เพศชายมีความคุ้นเคยและมีความรู้เกี่ยวกับฟุตบอลมากกว่าเพศหญิง เจน จันทรแสง (2537) และพัชรี ปิยภักดิ์ (2531) ได้ทำการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ พบว่ามีข้อสอบที่ลำเอียงต่อเพศและกลุ่มที่อยู่ในสภาพแวดล้อมที่ต่างกัน และชัชชัย เผ่าพงษ์ (2527) ได้ทำการศึกษาความลำเอียงของแบบทดสอบวัดความถนัดทางคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 3 ในทุกภาคของประเทศ พบว่าข้อสอบบางข้อมีความลำเอียงต่อกลุ่มนักเรียนเพศใดเพศหนึ่งโดยเฉพาะ จากงานวิจัยดังกล่าวแสดงให้เห็นว่าภูมิศาสตร์และเพศของนักเรียนมีอิทธิพลต่อผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ ดังนั้นการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบจึงเป็นวิธีหนึ่งที่น่าสนใจนำมาใช้ตรวจสอบ เพื่อช่วยพัฒนาข้อสอบให้มีคุณภาพและมีความยุติธรรม

จากเหตุผลที่กล่าวมาข้างต้น ผู้วิจัยมีความสนใจศึกษาและตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ เนื่องจากเพศและขนาดโรงเรียนน่าจะเป็นตัวแปรที่มีผลต่อการทำหน้าที่ของข้อสอบ Pre O-NET วิชาคณิตศาสตร์ ซึ่งวิชาคณิตศาสตร์เป็นวิชาที่พัฒนาทางด้านสติปัญญาเกี่ยวกับความรู้กระบวนการคิดและทักษะการแก้โจทย์ปัญหาที่ประกอบไปด้วยภาษาและตัวเลขที่ต้องการคำตอบต้องอาศัยการอ่านโจทย์เพื่อนำไปสู่การแปลงเป็นประโยคสัญลักษณ์ทางคณิตศาสตร์ โดยทำการตรวจสอบด้วยวิธีทฤษฎีการตอบสนองข้อสอบ (IRT) แบบ 3 พารามิเตอร์ และวิธีไค - สแควร์ (χ^2) ทั้งนี้เนื่องจากวิธีทฤษฎีการตอบสนองข้อสอบเป็นวิธีที่ให้รายละเอียดของค่าพารามิเตอร์ไม่แปรเปลี่ยนไปตามกลุ่มตัวอย่าง มีความถูกต้องและสอดคล้องกับทฤษฎีมากที่สุด และวิธีไค - สแควร์ เป็นวิธีที่คำนวณและแปลความหมายง่าย สามารถคำนวณได้โดยไม่ต้องใช้โปรแกรมคอมพิวเตอร์ ประหยัดเวลาและค่าใช้จ่าย ซึ่งผลที่ได้จากการศึกษาครั้งนี้จะมีประโยชน์ สามารถใช้เป็นแนวทางในการพัฒนาแบบทดสอบให้มีมาตรฐาน มีคุณภาพและมีความยุติธรรม ทำให้การประเมินผลความรู้ความสามารถอย่างแท้จริง เป็นประโยชน์แก่โรงเรียนและเขตพื้นที่การศึกษา เพื่อจะได้มีนโยบายส่งเสริมและแก้ปัญหาทางการเรียนและให้ความช่วยเหลือแก่ผู้เรียนต่อไป

วัตถุประสงค์ของการวิจัย (Research Objectives)

1. เพื่อตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยวิธีการตอบสนองข้อสอบ(IRT) แบบ 3 พารามิเตอร์ด้วยวิธีวัดความแตกต่างของค่าพารามิเตอร์ความยาก (b) โดยวิเคราะห์ระหว่างกลุ่มเพศและขนาดของโรงเรียน
2. เพื่อตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยวิธีไค - สแควร์ (χ^2) โดยวิเคราะห์ระหว่างกลุ่มเพศและขนาดของโรงเรียน

วิธีดำเนินการวิจัย (Research Methods)

1. ประชากรและกลุ่มตัวอย่าง

ประชากร คือ นักเรียนชั้นมัธยมศึกษาปีที่ 3 สำนักงานเขตพื้นที่การศึกษามัธยมศึกษา เขต 39 (พิษณุโลก-อุตรดิตถ์) จำนวน 58 โรงเรียน

กลุ่มตัวอย่าง คือ นักเรียนชั้นมัธยมศึกษาปีที่ 3 สำนักงานเขตพื้นที่การศึกษามัธยมศึกษา เขต 39 (พิษณุโลก-อุตรดิตถ์) โดยเลือกใช้โรงเรียนในจังหวัดพิษณุโลก จำนวน 39 โรงเรียน นักเรียนจำนวน 12,696 คน

2. ข้อมูลและเครื่องมือที่ใช้ในการวิจัย

ข้อมูลที่ใช้ในการวิจัย ได้แก่ ข้อมูลทุติยภูมิ (secondary data) จากสำนักงานเขตพื้นที่การศึกษามัธยมศึกษา เขต 39 ซึ่งเป็นผลการตอบข้อสอบ Pre O-NET วิชาคณิตศาสตร์ ปีการศึกษา 2558 – 2560 ชั้นมัธยมศึกษาปีที่ 3 ของโรงเรียนที่อยู่ในพื้นที่จังหวัดพิษณุโลก จำนวน 39 โรงเรียน เป็นข้อสอบแบบปรนัย 4 ตัวเลือก โดยปีการศึกษา 2558 มีจำนวนข้อสอบที่ใช้ในการวิเคราะห์ 30 ข้อ ปีการศึกษา 2559 มีจำนวนข้อสอบที่ใช้ในการวิเคราะห์ 22 ข้อ และปีการศึกษา 2560 มีจำนวนข้อสอบที่ใช้ในการวิเคราะห์ 16 ข้อ

3. การเก็บรวบรวมข้อมูล

ผู้วิจัยรวบรวมข้อมูลผลการตอบข้อสอบของนักเรียน ซึ่งเป็นข้อมูลทุติยภูมิจากสำนักงานเขตพื้นที่การศึกษามัธยมศึกษา เขต 39 โดยมีขั้นตอนการเก็บรวบรวมข้อมูลดังนี้

1. ขอนหนังสือจากบัณฑิตวิทยาลัย คณะครุศาสตร์ มหาวิทยาลัยราชภัฏพิบูลสงคราม เพื่อขอข้อมูลผลการตอบข้อสอบของนักเรียนโดยใช้แบบทดสอบ Pre O-NET วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ปีการศึกษา 2558 – 2560
2. เมื่อได้รับข้อมูลจากสำนักงานเขตพื้นที่การศึกษามัธยมศึกษา เขต 39 แล้วผู้วิจัยนำมาจำแนกข้อมูลตามกลุ่มเพศและขนาดของโรงเรียน
3. ตรวจสอบความเรียบร้อยของข้อมูลแล้วนำข้อมูลมาวิเคราะห์ตามวัตถุประสงค์ของการวิจัยต่อไป

4. การวิเคราะห์ข้อมูล

การวิเคราะห์ด้วยวิธีการตอบสนองข้อสอบ (IRT) ด้วยวิธีการหาค่าความแตกต่างของค่าพารามิเตอร์ความยาก (b Parameter Difference) การศึกษาการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบครั้งนี้ เป็นการหาประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบของกลุ่มอ้างอิง (เพศชายและโรงเรียนขนาดใหญ่) และกลุ่มเปรียบเทียบ (เพศหญิงและโรงเรียนขนาดเล็ก) มีวิธีการวิเคราะห์ดังนี้

1. จัดทำแฟ้มข้อมูลผลการตอบข้อสอบโดยแยกเป็นกลุ่มอ้างอิงและกลุ่มเปรียบเทียบด้วยโปรแกรม Microsoft Excel
2. นำข้อมูลที่ได้นำมาคำนวณหาค่าพารามิเตอร์ a, b และ c ด้วยโปรแกรม MULTILOG โดยวิเคราะห์แยกระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ

3. หาค่าความแตกต่างของค่าพารามิเตอร์ความยาก (b Parameter Difference) ระหว่าง

กลุ่มอ้างอิง(Reference Group : **R**) และกลุ่มเปรียบเทียบ(Focal Group : **F**) ด้วยดัชนีความแตกต่างระหว่างค่าความยากที่คิดเครื่องหมาย เมื่อมีการควบคุมค่า θ (Signed B Difference Controlling for θ : $SBD - \theta$)

$$SBD - \theta = b_F - b_R$$

$$= \Delta b$$

ค่า Δb เป็นความแตกต่างของค่าพารามิเตอร์ความยากระหว่างกลุ่มอ้างอิง(**R**) กับกลุ่มเปรียบเทียบ(**F**) ถ้า Δb มีค่าเป็นบวก แสดงว่ากลุ่มอ้างอิงทำข้อสอบได้ดีกว่า แต่ถ้า Δb มีค่าเป็นลบ แสดงว่ากลุ่มเปรียบเทียบทำข้อสอบได้ดีกว่า (ศิริชัย กาญจนวาสี, 2555)

สำหรับการทดสอบนัยสำคัญของความแตกต่างของค่าพารามิเตอร์ความยากระหว่างกลุ่มสองกลุ่ม สำหรับทดสอบ $H_0 : \Delta b = 0$ มีสถิติสำหรับการทดสอบดังนี้

$$d = \frac{\Delta b}{S_{\Delta b}}$$

เมื่อ d = สถิติทดสอบซึ่งมีการแจกแจงแบบปกติ (Z)

$$\Delta b = b_F - b_R$$

$$S_{\Delta b} = \sqrt{S_F^2 + S_R^2}$$

S_F = ความคลาดเคลื่อนมาตรฐานของค่าความยากกลุ่มเปรียบเทียบ

S_R = ความคลาดเคลื่อนมาตรฐานของค่าความยากกลุ่มอ้างอิง

เกณฑ์การแปลผล

$d \geq 1.96$ หมายถึง กลุ่มอ้างอิงทำข้อสอบได้ดีกว่าอย่างมีนัยสำคัญที่ .05

$d \leq -1.96$ หมายถึง กลุ่มเปรียบเทียบทำข้อสอบได้ดีกว่าอย่างมีนัยสำคัญที่ .05

การวิเคราะห์ด้วยวิธีการทดสอบไค - สแควร์ (χ^2) วิธีการวิเคราะห์ดังนี้

1. จำแนกข้อมูลในแต่ละกลุ่มตัวอย่างออกเป็นช่วงระดับความสามารถตามคะแนนรวม โดยมี

วิธีการดังนี้

1.1 เรียงคะแนนผู้ที่ได้คะแนนรวมน้อยที่สุดไปหาผู้ที่ได้คะแนนรวมมากที่สุด

1.2 แบ่งคะแนนรวมออกเป็น 5 ช่วงระดับความสามารถ โดยมีหลักการดังนี้

1.2.1 ในแต่ละช่วงระดับความสามารถควรมีความถี่ในการตอบถูกและความถี่ที่คาดหวังมีค่าน้อยที่สุด 5

1.2.2 ช่วงกว้างของคะแนนในแต่ละระดับความสามารถมีจำนวนกลุ่มตัวอย่างใกล้เคียงกันในแต่ละข้อ

1.3 คำนวณค่าสัดส่วนในการตอบถูก (P) จากสูตร

$$P_{ij} = \frac{O_{ij}}{N_{ij}}$$

เมื่อ O_{ij} แทน ความถี่ที่สังเกตได้ในแต่ละชั้นคะแนน

N_{ij} แทน จำนวนนักเรียนรวมในแต่ละชั้นคะแนน

คำนวณความถี่ที่คาดหวัง (E) จากสูตร

$$E_{ij} = P_{ij}N_{ij}$$

เมื่อ P_{ij} แทน สัดส่วนการตอบถูกของกลุ่มตัวอย่างทั้งหมดในแต่ละชั้น

คะแนน

N_{ij} แทน จำนวนนักเรียนรวมในแต่ละชั้นคะแนน

1.4 คำนวณค่า χ^2 จากสูตร

$$\chi^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

$$df = (i - 1)(j - 1)$$

เมื่อ χ^2 แทน ความถี่ที่สังเกตได้ของการตอบถูกของแต่ละกลุ่ม

O_{ij} แทน ความถี่ที่คาดหวังของการตอบถูกของแต่ละกลุ่ม

i แทน จำนวนกลุ่มที่ทำการวิเคราะห์

j แทน จำนวนชั้นคะแนน

วิธีไค-สแควร์ จะถือว่าข้อสอบมีความลำเอียงก็ต่อเมื่อโอกาสในการตอบถูกของผู้สอบ 2 กลุ่มแตกต่างกัน โดยเกณฑ์ในการพิจารณาคือ ถ้าค่าไค-สแควร์ ที่คำนวณได้มีค่ามากกว่าที่เปิดในตาราง แสดงว่า ข้อสอบข้อนั้นมีความลำเอียงและจะลำเอียงเข้าข้างกลุ่มใดนั้นต้องพิจารณาที่สัดส่วนในการตอบถูกของแต่ละกลุ่มในแต่ละระดับความสามารถ(บุญชม ศรีสะอาด,2535)

ผลการวิจัย (Research Results)

การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ เป็นการหาประสิทธิภาพในด้านความยุติธรรมของข้อสอบ ของกลุ่มอ้างอิงกลุ่มและกลุ่มเปรียบเทียบ ในการวิจัยครั้งนี้ผู้วิจัยได้กำหนดเพศชายและโรงเรียนขนาดใหญ่เป็นกลุ่มอ้างอิง และเพศหญิงและโรงเรียนขนาดเล็กเป็นกลุ่มเปรียบเทียบ ผลการวิจัยพบว่า

1. ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบตามทฤษฎีการตอบสนองข้อสอบ (IRT) แบบ 3 พารามิเตอร์ โดยวิธีวัดความแตกต่างของค่าพารามิเตอร์ความยาก (b) พบว่า เมื่อจำแนกผู้สอบตามกลุ่มเพศ ปีการศึกษา 2558 พบข้อสอบทำหน้าที่ต่างกันจำนวน 2 ข้อ โดยลำเอียงเข้าข้างเพศชาย 1 ข้อ ได้แก่ ข้อ 8 และลำเอียงเข้าข้างเพศหญิง 1 ข้อ ได้แก่ ข้อ 29 ปีการศึกษา 2559 พบข้อสอบทำหน้าที่ต่างกัน จำนวน 2 ข้อ โดยลำเอียงเข้าข้างเพศชาย 1 ข้อ ได้แก่ ข้อ 12 และลำเอียงเข้าข้างเพศหญิง 1 ข้อ ได้แก่ ข้อ 19 และปีการศึกษา 2560 พบข้อสอบทำหน้าที่ต่างกัน จำนวน 4 ข้อ โดยลำเอียงเข้าข้างเพศชาย 2 ข้อ ได้แก่ ข้อ 1 , 3 และลำเอียงเข้าข้างเพศหญิง 2 ข้อ ได้แก่ ข้อ 2 และ ข้อ 12 เมื่อจำแนกผู้สอบตามกลุ่มขนาดโรงเรียน ปีการศึกษา 2558 พบข้อสอบทำหน้าที่ต่างกัน จำนวน 9 ข้อ โดยลำเอียงเข้าข้างโรงเรียนขนาดใหญ่ 7 ข้อ ได้แก่ ข้อ 8 , 10 , 11 , 12 , 13 , 18 และ ข้อ 23 และลำเอียงเข้าข้างโรงเรียนขนาดเล็ก 2 ข้อ ได้แก่ ข้อ 2 และ ข้อ 27 ปีการศึกษา 2559 พบข้อสอบทำหน้าที่ต่างกัน จำนวน 5 ข้อ โดยลำเอียงเข้าข้างโรงเรียนขนาดใหญ่ 1 ข้อ ได้แก่ ข้อ 9 และลำเอียงเข้าข้างโรงเรียนขนาดเล็ก 4 ข้อ ได้แก่ ข้อ 12 , 15 , 19 และ ข้อ 21 และปีการศึกษา 2560 พบข้อสอบทำหน้าที่ต่างกัน จำนวน 4 ข้อ โดยลำเอียงเข้าข้างโรงเรียนขนาดใหญ่ 2 ข้อ ได้แก่ ข้อ 3 และ ข้อ 10 และลำเอียงเข้าข้างโรงเรียนขนาดเล็ก 2 ข้อ ได้แก่ ข้อ 6 และ ข้อ 8

2. ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบเป็นรายข้อโดยวิธีไค - สแควร์ (χ^2) พบว่าเมื่อจำแนกผู้สอบตามกลุ่มเพศ ปีการศึกษา 2558 พบข้อสอบทำหน้าที่ต่างกันจำนวน 3 ข้อ โดยลำเอียงเข้าข้างเพศชาย 2 ข้อ ได้แก่ ข้อ 5 และข้อ 6 และลำเอียงเข้าข้างเพศหญิง 1 ข้อ ได้แก่ ข้อ 30 ปีการศึกษา 2559 พบข้อสอบทำหน้าที่ต่างกัน มีจำนวน 3 ข้อ โดยลำเอียงเข้าข้างเพศชาย 2 ข้อ ได้แก่ ข้อ 5 และข้อ 19 และลำเอียงเข้าข้างเพศหญิง 1 ข้อ ได้แก่ ข้อ 12 และปีการศึกษา 2560 พบข้อสอบที่ทำหน้าที่ต่างกัน มีจำนวน 5 ข้อ โดยลำเอียงเข้าข้างเพศชาย 4 ข้อ ได้แก่ ข้อ 3, 6, 12, และ ข้อ 13 และลำเอียงเข้าข้างเพศหญิง 1 ข้อ ได้แก่ ข้อ 10 เมื่อจำแนกผู้สอบตามกลุ่มขนาดโรงเรียน ปีการศึกษา 2558 พบข้อสอบทำหน้าที่ต่างกัน จำนวน 4 ข้อ โดยลำเอียงเข้าข้างโรงเรียนขนาดใหญ่ทั้งหมด ได้แก่ ข้อ 5, 8 , 18 และข้อ 23 ปีการศึกษา 2559 พบข้อสอบทำหน้าที่

ต่างกัน จำนวน 4 ข้อ โดยลำเอียงเข้าข้างโรงเรียนขนาดใหญ่ 2 ข้อ ได้แก่ ข้อ 9 และ 12 และ ลำเอียงเข้าข้างโรงเรียนขนาดเล็ก 2 ข้อ ได้แก่ ข้อ 5 และข้อ 10 และปีการศึกษา 2560 พบข้อสอบ ทำหน้าที่ต่างกันมีจำนวน 5 ข้อ โดยลำเอียงเข้าข้างโรงเรียนขนาดใหญ่ 4 ข้อ ได้แก่ ข้อ 1, 3, 7 และ 13 และลำเอียงเข้าข้างโรงเรียนขนาดเล็ก 1 ข้อ ได้แก่ ข้อ 12

จากผลการวิเคราะห์การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธีการตอบสนอง ข้อสอบ (IRT) แบบ 3 พารามิเตอร์ ด้วยวิธีวัดความแตกต่างของค่าพารามิเตอร์ความยาก (b) และ วิธีไค - สแควร์ (χ^2) เป็นรายข้อของข้อสอบ Pre O-NET วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ระหว่างปีการศึกษา 2558 – 2560 จำแนกตามกลุ่มเพศและกลุ่มขนาดของโรงเรียนสรุปได้ดัง ตาราง

1 และ 2

ตาราง 1 จำนวนข้อสอบที่ตรวจพบการทำหน้าที่ของข้อสอบด้วยวิธีทฤษฎีการตอบสนองข้อสอบ แบบ 3

พารามิเตอร์ (IRT)

ปี การศึกษา	จำแนกตามเพศ		รวม (ข้อ)	จำแนกตามขนาดโรงเรียน		รวม (ข้อ)
	ชาย (ข้อที่)	หญิง (ข้อที่)		ใหญ่ (ข้อที่)	เล็ก (ข้อที่)	
2558	8	29	2	8,10,11,12, 13,18,23	2,27	9
2559	12	19	2	9	12,15,19,21	5
2560	1,3	2,12	4	3,10	6,8	4
รวม	4	4	8	10	8	18

ตาราง 2 จำนวนข้อสอบที่ตรวจพบการทำหน้าที่ของข้อสอบด้วยวิธีไค - สแควร์ (χ^2)

ปี การศึกษา	จำแนกตามเพศ		รวม (ข้อ)	จำแนกตามขนาดโรงเรียน		รวม (ข้อ)
	ชาย (ข้อที่)	หญิง (ข้อที่)		ใหญ่ (ข้อที่)	เล็ก (ข้อที่)	
2558	5,6	30	3	5,8,18,23	-	4
2559	5,19	12	3	9,12	5,10	4
2560	3,6,12,13	10	5	1,3,7,13	12	5
รวม	8	3	11	10	3	13

อภิปรายผลการวิจัย (Research Discussion)

ผู้วิจัยขออภิปรายผลในประเด็นการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ Pre O – NET วิชาคณิตศาสตร์ ปีการศึกษา 2558 – 2560 โดยวิธีการตอบสนองข้อสอบ(IRT) แบบ 3 พารามิเตอร์และ วิธีไค - สแควร์ (χ^2) โดยวิเคราะห์ระหว่างกลุ่มเพศและขนาดของโรงเรียน พบว่า

1. การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบทั้งสองวิธี ตามตัวแปรเพศ พบว่าเพศเป็นตัวแปรสำคัญที่ทำให้ข้อสอบทำหน้าที่ต่างกัน เมื่อวิเคราะห์จากข้อคำถามของข้อสอบ Pre O – NET วิชาคณิตศาสตร์ ข้อสอบมีลักษณะของกิจกรรมที่เหมาะสมกับเพศชาย เพศชายมีโอกาสเจอหรือสัมผัสกับกิจกรรมนั้นอยู่เสมอก็คงจะตอบข้อสอบข้อนั้นได้ดีกว่าเพศหญิง เช่น กิจกรรมงานช่าง การวัด การคำนวณหาพื้นที่ปริมาตร เป็นต้น ข้อสอบวิชาคณิตศาสตร์อาจมีลักษณะบางข้อเป็นโจทย์ปัญหาที่ประกอบไปด้วยภาษาและตัวเลขที่ต้องการคำตอบ ซึ่งต้องอาศัยการอ่านโจทย์เพื่อนำไปสู่การแปลงเป็นประโยคสัญลักษณ์ทางคณิตศาสตร์เพื่อหาคำตอบ มีข้อความของโจทย์ปัญหาบางข้อความทำให้ผู้สอบเข้าใจไม่ตรงกัน เนื้อหาทักษะที่ผู้สอบไม่คุ้นเคย หรือสารสนเทศที่ไม่เป็นการทั่วไปกับกลุ่มผู้สอบ จึงส่งผลให้เกิดการได้เปรียบและเสียเปรียบระหว่างกลุ่มผู้สอบ โดยกลุ่มเพศชายเป็นกลุ่มที่ได้เปรียบในการตอบข้อสอบส่วนกลุ่มเพศหญิงเป็นกลุ่มที่เสียเปรียบ สามารถสรุปได้ว่า ตัวแปรเพศมีผลต่อการทำหน้าที่ต่างกันของข้อสอบวิชาคณิตศาสตร์ สอดคล้องกับผลการวิจัยของ อมลณัฐ อุบลรัตน์(2552) ได้ศึกษาการทำหน้าที่ต่างกันของข้อสอบวัดผลสัมฤทธิ์ทางการเรียน ชั้นมัธยมศึกษาปีที่ 2 สำนักงานเขตพื้นที่การศึกษาปัตตานี เขต 2 พบว่า เพศเป็นตัวแปรที่มีผลต่อการทำหน้าที่ต่างกันของข้อสอบในวิชาคณิตศาสตร์ ข้อสอบที่ทำหน้าที่ต่างกันส่วนใหญ่ลำเอียงเข้าข้างกลุ่มเพศชายมากกว่ากลุ่มเพศหญิง Gierl, Bisanz, Bisanz and Boughton (2003) ได้ศึกษาการระบุเนื้อหาและทักษะทางพุทธิปัญญาที่ทำให้เกิดความแตกต่างทางเพศที่มีต่อวิชาคณิตศาสตร์ โดยใช้กระบวนการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบแบบพหุมิติ พบว่านักเรียนชายทำคะแนนในส่วนของมิติสัมพันธ์ได้ดีกว่านักเรียนหญิง ในขณะที่นักเรียนหญิงทำคะแนนในส่วนทักษะความจำได้ดีกว่านักเรียนชาย พิธญา สูงเนิน(2551) ได้ศึกษาการทำหน้าที่ต่างกันของข้อสอบวัดความคิดเชิงตรรกศาสตร์ที่มีบริบทเกี่ยวกับการเล่นฟุตบอล อาจทำให้เพศชายมีความคุ้นเคยและมีความรู้เกี่ยวกับฟุตบอลมากกว่าเพศหญิง เจน จันทรแสง (2537) และพัชรี ปิยภักดิ์(2531)ได้ทำการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ พบว่ามีข้อสอบมีความลำเอียงต่อเพศและเมื่อวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบจำแนกตามกลุ่มขนาดของโรงเรียน พบว่าข้อสอบลำเอียงเข้าข้างโรงเรียนขนาดใหญ่มากกว่าโรงเรียนขนาดเล็ก ทั้งนี้อาจมีสาเหตุมาจาก ความสามารถของตัวนักเรียนเอง ฐานะทางเศรษฐกิจของนักเรียนที่แตกต่างกัน ความพร้อมของบุคลากรหรือครูเฉพาะสาขาวิชา ซึ่งครูที่อยู่ในเมืองมีความพร้อมและทันสมัยกว่า การรับรู้ข่าวสารและความรู้ต่างๆ มากกว่า ครูที่สอนคณิตศาสตร์ในโรงเรียนขนาดเล็กส่วนใหญ่จะเป็นครูที่ไม่ได้จบสาขาวิชาคณิตศาสตร์มาโดยตรง และอาจเนื่องมาจากความพร้อมของอุปกรณ์ในการจัดการเรียนการสอน สอดคล้องกับงานวิจัย

ของทิพย์รัตน์ ผลบุญ (2540) ได้ศึกษาดัชนีความลำเอียงของข้อสอบวัดผลสัมฤทธิ์ทางการเรียนในวิชาคณิตศาสตร์ ซึ่งเกิดจากตัวแปรประเภทของโรงเรียน พบว่า ข้อสอบลำเอียงเข้าข้างโรงเรียนสามัญศึกษามากกว่าโรงเรียนขยายโอกาส เพ็ญพนา สุขสม(2540) ได้ทำการเปรียบเทียบผลของการวิเคราะห์ความลำเอียงของข้อสอบเมื่อวิเคราะห์ตามที่ตั้งของโรงเรียน พบว่า ข้อสอบที่ลำเอียงส่วนใหญ่จะเข้าข้างกลุ่มผู้สอบที่เรียนอยู่ในเมืองมากกว่ากลุ่มผู้สอบที่เรียนอยู่นอกเมือง

2. การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบทั้งสองวิธี มีจำนวนทั้งหมด 68 ข้อ พบว่า วิธีทฤษฎีการตอบสนองข้อสอบ (IRT) แบบ 3 พารามิเตอร์ตรวจสอบการทำหน้าที่ต่างกันได้มากกว่าวิธีไค - สแควร์(χ^2) โดยวิธีทฤษฎีการตอบสนองข้อสอบ (IRT) ตรวจสอบข้อสอบทำหน้าที่ต่างกัน จำนวน 26 ข้อ และวิธีไค - สแควร์(χ^2) ตรวจสอบข้อสอบทำหน้าที่ต่างกัน จำนวน 24 ข้อ ซึ่งตรวจสอบข้อสอบที่ลำเอียงร่วมกันดังนี้ ปีการศึกษา 2558 ได้แก่ ข้อ 8 , 18 และ 23 โดยข้อสอบ ข้อที่ 8 ลำเอียงเข้าข้างเพศชายและโรงเรียนขนาดใหญ่ ข้อ 18 และข้อ 23 ลำเอียงเข้าข้างโรงเรียนขนาดใหญ่ ปีการศึกษา 2559 ได้แก่ข้อ 9 , 12 และ 19 โดยข้อสอบข้อ 9 ลำเอียงเข้าข้างโรงเรียนขนาดใหญ่ ข้อ 12 ลำเอียงเข้าข้างโรงเรียนขนาดเล็ก และข้อ 19 ลำเอียงเข้าข้างเพศชายและโรงเรียนขนาดเล็ก และปีการศึกษา 2560 ได้แก่ข้อ 1, 3, 6, 10 และ 12 โดยข้อ 1 ลำเอียงเข้าข้างเพศชายและโรงเรียนขนาดใหญ่ ข้อ 3 ลำเอียงเข้าข้างเพศชาย ข้อ 6 ลำเอียงเข้าข้างเพศชายและโรงเรียนขนาดเล็ก ข้อ 10 ลำเอียงเข้าข้างโรงเรียนขนาดใหญ่ และข้อ 12 ลำเอียงเข้าข้างเพศหญิงและโรงเรียนขนาดเล็กซึ่งเป็นกลุ่มเปรียบเทียบ จากผลดังกล่าวทำให้สรุปได้ว่า ในแต่ละวิธีที่ใช้วิเคราะห์ให้ผลการวิเคราะห์ที่สอดคล้องกัน แต่ความไวในการตรวจสอบต่างกัน ซึ่งสอดคล้องกับงานวิจัยของ ทิพย์รัตน์ ผลบุญ (2540) ได้ศึกษาดัชนีความลำเอียงของข้อสอบวัดผลสัมฤทธิ์ทางการเรียนในวิชาคณิตศาสตร์ ด้วยวิธีการวิเคราะห์ 3 วิธี วิธี คือวิธีไค-สแควร์ วิธีทฤษฎีการตอบสนองของข้อสอบที่มี 3 พารามิเตอร์และวิธีแมนเทิล - แฮนส์เซล ผลการวิจัยพบว่า วิธีการวิเคราะห์ความลำเอียงโดยวิธีทฤษฎีการตอบสนองข้อสอบที่มี 3 พารามิเตอร์ ตรวจสอบจำนวนข้อสอบที่ลำเอียงมากที่สุด รองลงมาคือวิธีแมนเทิล-แฮนส์เซล ส่วนวิธีไค-สแควร์ เป็นวิธีที่พบความลำเอียงน้อยที่สุด Baghi&Ferrara(1990)ได้เปรียบเทียบผลจากการสืบค้นความลำเอียงของข้อสอบด้วยทฤษฎีการตอบสนองข้อสอบ 3 พารามิเตอร์ (IRT) และสถิติไค-สแควร์ แบบ MH (MHCS) ผลปรากฏว่า วิธี IRT สืบค้นความลำเอียงของข้อสอบได้ไวกว่าสถิติไค-สแควร์ แบบ MH (MHCS) สรุศักดิ์ อมรรัตนศักดิ์ (2531) และสุพัฒน์ สุขมลสันต์ (2538) พบว่าวิธีทฤษฎีการตอบสนองข้อสอบ(IRT) แบบ 3 พารามิเตอร์ จะมีความไวในการตรวจค้นข้อสอบที่ทำหน้าที่ต่างกันได้ดีที่สุด ทำให้สรุปได้ว่าในการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบนั้นจะเลือกใช้วิธีใดก็ได้ เพราะผลที่ได้ใกล้เคียงกัน ซึ่งการวิจัยครั้งนี้ผู้วิจัยเลือกใช้วิธีที่วิเคราะห์ง่าย สะดวก และประหยัดค่าใช้จ่าย เช่นวิธีไค-สแควร์ เป็นวิธีที่ง่ายต่อการคำนวณ ผลการวิเคราะห์เชื่อถือได้และใช้ได้ในกรณีที่ไม่มีคอมพิวเตอร์ และวิธีทฤษฎีการตอบสนองข้อสอบ เป็นวิธีที่มีความไวต่อการค้นพบข้อสอบที่ทำหน้าที่ต่างกันได้สูงสุด

ข้อเสนอแนะการวิจัย (Research Suggestions)

ข้อเสนอแนะในการนำผลการวิจัยไปใช้

1. นักพัฒนาแบบทดสอบควรตรวจสอบและออกข้อสอบที่อาจเกิดการทำหน้าที่ต่างกันของข้อสอบ โดยเฉพาะลักษณะของขอบเขตเนื้อหา โจทย์ปัญหาและประโยคสัญลักษณ์
2. การพัฒนาลักษณะข้อสอบ ควรมีการพิจารณาการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบเป็นส่วนหนึ่งในการหาคุณภาพข้อสอบนอกเหนือไปจากการหาค่าความยาก ค่าอำนาจจำแนก ค่าความตรงและค่าความเที่ยง

ข้อเสนอแนะในการทำวิจัยครั้งต่อไป

1. ควรศึกษาผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบกับตัวแปรอื่นๆ เช่น รายวิชาอื่น ภาษาถิ่น ศาสนา สังคมและสภาพทางเศรษฐกิจของครอบครัว เป็นต้น และเพื่อใช้เป็นเกณฑ์ในการแบ่งกลุ่มเพราะในแต่ละภูมิภาคจะมีภาษาถิ่น ศาสนา ภูมิลาเนาที่ต่างกัน สามารถทำการวิเคราะห์พร้อมกันหลายๆด้าน เพื่อให้ได้ข้อมูลสารสนเทศมากยิ่งขึ้นและสามารถยืนยันข้อสอบได้
2. ควรมีการศึกษาการทำหน้าที่ต่างกันของข้อสอบอัตนัยหรือข้อสอบที่มีการให้คะแนนแบบหลายค่า(Polytomously Scored Item) นอกเหนือจากข้อสอบที่มีการให้คะแนนแบบสองค่า(Dichotomously Scored Item)
3. จากการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบในครั้งนี้ผู้วิจัยใช้วิธีทฤษฎีการตอบสนองข้อสอบ แบบ 3 พารามิเตอร์ และ วิธีไค - สแควร์ เนื่องจากเป็นวิธีที่คำนวณง่ายและไม่มีข้อจำกัดในเรื่องของขนาดกลุ่มตัวอย่างระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบไม่จำเป็นต้องเท่ากันก็สามารถทำได้ ควรเพิ่มวิธีการศึกษาอื่น เช่น วิธีแมนเทิล-แฮนเซล (MH) ,วิธีซิปเทสท์(SIBTEST) และวิธีการวิเคราะห์องค์ประกอบจำกัด (RFA) เป็นต้น

เอกสารอ้างอิง (References)

- กาญจนา วัฒนสุนทร. (2538). การพัฒนาเกณฑ์ตัดสินข้อสอบลำเอียงทางเพศ. ปรินญาตุษฎบัณฑิตภาควิชาวิจัยการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.
- เจน จันทรแสง. (2537). การวิเคราะห์ความลำเอียงของข้อสอบแบบสอบวัดผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3. วิทยานิพนธ์ศึกษามหาบัณฑิต. มหาวิทยาลัยนเรศวร.
- ซัชชัย เผ่าพงษ์. (2527). การเปรียบเทียบความลำเอียงของข้อสอบจากแบบทดสอบวัดความถนัดทางการเรียนคณิตศาสตร์และภาษาระดับชั้นมัธยมศึกษาตอนต้น. วิทยานิพนธ์ศึกษาศาสตรมหาบัณฑิต บัณฑิตวิทยาลัย. มหาวิทยาลัยขอนแก่น.
- ทิพย์รัตน์ ผลบุญ. (2540). การวิเคราะห์ความลำเอียงของข้อสอบปลายภาควิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 1 ของกลุ่มโรงเรียนมัธยมศึกษาตอนต้น ในจังหวัดกำแพงเพชร. ปรินญาตุษฎศึกษามหาบัณฑิต .มหาวิทยาลัยนเรศวร

- บุญชม ศรีสะอาด. (2535). *การวิจัยทางการวัดผลและประเมินผล*. มหาสารคาม: มหาวิทยาลัยศรีนครินทรวิโรฒ มหาสารคาม.
- พัชรี ปิยพันธ์. (2531). *การวิเคราะห์ความลำเอียงของข้อสอบจากแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนชั้นประถมศึกษาปีที่ 6*. ปริญญาโทศึกษาศาสตร์มหาบัณฑิต. มหาวิทยาลัยศรีนครินทร วิโรฒประสานมิตร.
- พริญา สูงเนิน. (2551). *การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบพหุมิติ : การเปรียบเทียบระหว่างข้อสอบรายข้อกับหมวดข้อสอบโดยวิธีชิปเทสท์*. ปริญญานิพนธ์ วิทยาศาสตร์มหาบัณฑิต. ชลบุรี: มหาวิทยาลัยบูรพา.
- เพ็ญพนา สุขสม. (2540). *การเปรียบเทียบผลของวิธีวิเคราะห์ความลำเอียงของข้อสอบที่แตกต่างกัน 3 วิธี*. วิทยานิพนธ์ปริญญาโทมหาบัณฑิต สาขาวิชาการวัดผลการศึกษา มหาวิทยาลัยนเรศวร.
- ศิริชัย กาญจนาวสี. (2555). *ทฤษฎีการทดสอบแนวใหม่ (Model test theories)*. พิมพ์ครั้งที่ 4. กรุงเทพฯ : โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- สำนักทดสอบการศึกษา. (2557). *คู่มือการใช้ข้อสอบ Pre O-NET*. สืบค้น 20 สิงหาคม 2564, จาก http://db.kkzone.go.th/plus/data/news/photo/26-12-2014-2-10-32_309474594.pdf.
- อมลณัฐ อุบลรัตน์. (2552). *การทำหน้าที่ต่างกันด้วยวิธีโลจิสติกและวิธีแมนเทิล-แฮนเซลสำหรับข้อสอบวัดผลสัมฤทธิ์ทางการเรียนชั้นมัธยมศึกษาปีที่ 2*. วิทยานิพนธ์ปริญญา ศึกษาศาสตร์มหาบัณฑิต. สงขลา: มหาวิทยาลัยสงขลานครินทร์.
- 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