

การพัฒนาแบบสอบคู่ขนานวิชาคณิตศาสตร์สำหรับนักเรียนชั้นประถมศึกษาปีที่ 2
โรงเรียนสังกัดกรุงเทพมหานคร โดยการประยุกต์ใช้ทฤษฎีการตอบสนองข้อสอบ
แบบ 2 พารามิเตอร์

Development of Parallel Tests in Mathematics for Students in Grade 2 in
Schools under the Bangkok Metropolitan Administration through an Application
of Two-Parameter Item Response Theory

สาวินา บุญเสน^{1*} จุฑาภรณ์ มาสันเทียะ² และ พนิดา ปานิดวัฒนะ³

Savina Boonsen^{1*}, Chuthaphon Masantiah² and Panida Panidvadana³

(Received: January 1, 2022; Revised: February 28, 2022; Accepted: March 10, 2022)

บทคัดย่อ

การวิจัยนี้มีความมุ่งหมายเพื่อ 1) พัฒนาแบบสอบคู่ขนานวิชาคณิตศาสตร์ระดับชั้นประถมศึกษาปีที่ 2 สาระการวัด และ 2) ตรวจสอบความเป็นคู่ขนานของแบบสอบคู่ขนานวิชาคณิตศาสตร์ระดับชั้นประถมศึกษาปีที่ 2 สาระการวัด กลุ่มตัวอย่างเป็นนักเรียนชั้นประถมศึกษาปีที่ 2 ปีการศึกษา 2562 จำนวน 800 คน จากโรงเรียน สังกัดกรุงเทพมหานคร ในเขตหนองจอก จำนวน 13 โรงเรียน โดยการสุ่มแบบแบ่งชั้น (stratified random sampling) เครื่องมือที่ใช้ คือ แบบสอบแบบปรนัยเรื่องความยาว (1A, 1B) และเรื่องน้ำหนัก (2A, 2B) ซึ่งเป็น ข้อสอบแบบเลือกตอบฉบับละ 20 ข้อ การวิเคราะห์ข้อมูล ได้แก่ ความตรงตามเนื้อหา ความเที่ยง ค่าฟังก์ชัน สารสนเทศของแบบสอบ ค่าความยาก ค่าอำนาจจำแนก ค่าฟังก์ชันสารสนเทศของข้อสอบ และการตรวจสอบ ความเป็นคู่ขนานของข้อสอบและแบบสอบโดยใช้ Root mean square deviation of item information (RMSD_{IF}) และ Root mean square deviation of test information (RMSD_{TF}) ตามลำดับ

ผลการวิจัยพบว่า

1) แบบสอบคู่ขนานวิชาคณิตศาสตร์ทั้ง 4 ฉบับ มีค่า IOC ซึ่งเป็นดัชนีความตรงตามเนื้อหาของแบบสอบ ทุกฉบับ ตั้งแต่ 0.67-1.00 ความเที่ยงมีค่า 0.711, 0.652, 0.718 และ 0.635 ตามลำดับ ค่าฟังก์ชันสารสนเทศของ แบบสอบ 1A และ 1B ให้สารสนเทศสูงสุดสำหรับผู้ที่มีระดับความสามารถระหว่าง -0.60 ถึง -0.80 และแบบสอบ

¹ นักศึกษาปริญญาโท สาขาวิชาการประเมินและการวิจัย คณะศึกษาศาสตร์ มหาวิทยาลัยรามคำแหง

² ผู้ช่วยศาสตราจารย์ สาขาวิชาการประเมินและการวิจัย คณะศึกษาศาสตร์ มหาวิทยาลัยรามคำแหง

³ อาจารย์ สาขาวิชาการประเมินและการวิจัย คณะศึกษาศาสตร์ มหาวิทยาลัยรามคำแหง

¹ Master's degree student, Department of Evaluation and Research, Faculty of Education, Ramkhamhaeng University

² Assistant Professor, Department of Evaluation and Research, Faculty of Education, Ramkhamhaeng University

³ Lecturer, Department of Evaluation and Research, Faculty of Education, Ramkhamhaeng University

* Corresponding Author E-mail: savina_jack@hotmail.com

ฉบับ 2A และ 2B ให้สารสนเทศสูงสุดสำหรับผู้ที่มีระดับความสามารถระหว่าง -0.40 ถึง -0.60 ขณะที่ค่าความยาก (b) อยู่ระหว่าง -2.98 ถึง +0.57 ค่าอำนาจจำแนก (a) อยู่ระหว่าง 0.22 ถึง 1.13 และค่าฟังก์ชันสารสนเทศของข้อสอบทั้ง 4 ฉบับมีค่าอยู่ระหว่าง 0.013 ถึง 0.320

2) ผลการวิเคราะห์ความเป็นคู่ขนานรายข้อและทั้งฉบับของแบบสอบคู่ขนาน เรื่อง ความยาว (1A, 1B) มีค่า $RMSD_{IF}$ ตั้งแต่ 0.008 ถึง 0.167 และค่า $RMSD_{TIF}$ เท่ากับ 0.484 ขณะที่แบบสอบคู่ขนาน เรื่อง น้ำหนัก (2A, 2B) มีค่า $RMSD_{IF}$ ตั้งแต่ 0.020 – 0.110 และค่า $RMSD_{TIF}$ เท่ากับ 0.492

คำสำคัญ: ทฤษฎีการตอบสนองข้อสอบ แบบสอบคู่ขนาน ฟังก์ชันสารสนเทศของข้อสอบ
ฟังก์ชันสารสนเทศของแบบสอบ

Abstract

The purposes of this research were to 1) develop parallel tests in mathematics for grade 2, in the strand of measurement, and 2) verify the parallelism of the parallel tests in mathematics for grade 2, in the strand of measurement. The sample consisted of 800 grade 2 students in the academic year 2019 from 13 schools in Nong Chok District, under the Bangkok Metropolitan Administration, selected by stratified random sampling. The research instruments were 4 objective tests, on length (1A, 1B) and weight (2A, 2B). Each test comprised 20 multiple-choice items. The data were analyzed for content validity, reliability, test information function, item difficulty, discrimination, item information function, and the parallelism of the parallel tests and items using the root mean square deviation of item information ($RMSD_{IF}$) and root mean square deviation of test information ($RMSD_{TIF}$) respectively.

The research findings revealed that:

1) All the 4 parallel tests in mathematics had the index of item objective congruence (IOC), which was the index of content validity, ranging from 0.67 to 1.00; the reliability of 0.711, 0.652, 0.718, and 0.635 respectively; the maximum value of the test information function for the proficiency level was from -0.60 to -0.80 for the 1A and 1B tests and -0.40 to -0.60 for the 2A and 2B tests; the item difficulty (b) was between -2.98 and 0.57; the discrimination (a) was between 0.22 and 1.13, and the item information function of all the 4 tests was between 0.013 and 0.320.

2) The results of the parallelism analysis of the items and tests on length (1A, 1B) revealed that the $RMSD_{IF}$ ranged from 0.008 to 0.167, and $RMSD_{TIF}$ was 0.484; while the parallelism analysis of the items and tests on weight (2A, 2B) revealed that the $RMSD_{IF}$ ranged from 0.020 – 0.110, and $RMSD_{TIF}$ was 0.492.

Keywords: Item Response Theory, parallel test, item information function, test information function

บทนำ

การวัดผลและประเมินผลการเรียนรู้เป็นกระบวนการที่ผู้สอนใช้พัฒนาคุณภาพผู้เรียนเพราะผลจากการวัดจะทำให้ได้สารสนเทศที่แสดงถึงพัฒนาการความก้าวหน้าและความสำเร็จทางการเรียนของผู้เรียน โดยเครื่องมือที่ใช้ในการวัดผลการศึกษาที่มีหลายชนิด ซึ่งเครื่องมือที่นิยมใช้ในการวัดผลสัมฤทธิ์ทางการเรียน ได้แก่ แบบสอบที่ใช้สำหรับวัดความสามารถทางสมองหรือด้านสติปัญญาของผู้เรียนว่ามีความสามารถมากน้อยเพียงใด จากแนวทางการจัดกระบวนการเรียนรู้ของสังกัดกรุงเทพมหานครที่มุ่งเน้นการจัดการศึกษาที่มีคุณภาพ และพัฒนาให้เยาวชนในพื้นที่กรุงเทพฯ มีความรู้ มีจิตสำนึก รักเมืองกรุงเทพฯ ห่วงแค้นและร่วมสืบทอดวัฒนธรรมและอัตลักษณ์ของเมืองกรุงเทพฯ มีคุณธรรม มีทักษะในการดำรงชีวิตและประกอบอาชีพสุจริต (Education Strategy Office Education Bureau, 2016) ครูผู้สอนจึงนำแนวคิดนี้มาพัฒนานักเรียนให้มีความสามารถด้านความรู้ ทักษะ และคุณลักษณะอันพึงประสงค์ ตลอดจนครูต้องมีการวัดและประเมินผลการเรียนรู้ให้สอดคล้องตามสิ่งที่มุ่งหวัง

จากสภาพการวัดและประเมินผลของโรงเรียนสุเหร่าหะยีมิยา ผู้วิจัยในฐานะครูผู้สอนวิชาคณิตศาสตร์ พบปัญหาที่สำคัญ คือ กระบวนการสร้างแบบสอบที่ขาดมาตรฐาน เช่น ขาดการตรวจสอบคุณภาพของข้อสอบ และแบบสอบที่ครูสร้างขึ้น รวมถึงการใช้ข้อสอบซ้ำโดยการนำข้อสอบในปีที่ผ่านมาไปใช้สอบซ้ำ จึงทำให้ข้อสอบมีการเผยแพร่ไปสู่รุ่นต่อไปได้ สอดคล้องกับงานวิจัย Sawangboon (2011) ได้ศึกษาการสร้างแบบทดสอบคู่ขนานตามรูปแบบพาเซทวิชาคณิตศาสตร์ เรื่องการแยกตัวประกอบพหุนามดีกรีสองขั้นมัธยมศึกษาปีที่ 2 โดยใช้แบบทดสอบคู่ขนานเพื่อแก้ปัญหาการจำข้อสอบของผู้เรียน จึงสร้างแบบทดสอบที่มีลักษณะเหมือนกันทั้งในเนื้อหาและค่าสถิติเพื่อนำคะแนนแต่ละครั้งมาเปรียบเทียบกันได้ ผู้สอนมีความสะดวกในการที่จะนำแบบทดสอบฉบับใด ๆ ไปทดสอบกับนักเรียนก่อนก็ได้ และสอดคล้องกับงานวิจัย Kanchanapenkul (2011) ได้ศึกษาการสร้างข้อสอบคู่ขนานอิงโดเมนโดยใช้รูปแบบพาเซท วิชาคณิตศาสตร์ เรื่องวงกลม ชั้นมัธยมศึกษาปีที่ 3 ได้สร้างแบบทดสอบคู่ขนานเพื่อนำมาสลับกันใช้ในแต่ละครั้งที่ดำเนินการสอบไม่ต้องออกข้อสอบใหม่ ทำให้ทุ่นเวลาและป้องกันข้อสอบรั่วได้ด้วย จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้อง พบว่า วิธีการที่สามารถแก้ปัญหาข้างต้นได้ คือ การสร้างแบบสอบคู่ขนาน (Parallel Test) ที่เป็นแบบสอบตั้งแต่ 2 ฉบับขึ้นไป วัดและประเมินในเรื่องเดียวกัน มีค่าความยากง่าย อำนาจจำแนก ความเชื่อมั่น สัมประสิทธิ์สหสัมพันธ์ภายใน ค่าเฉลี่ยและความแปรปรวนของคะแนนใกล้เคียงกัน ตลอดจนมีไค้ลักษณะของข้อสอบ และฟังก์ชันสารสนเทศของข้อสอบและแบบสอบเหมือนกัน (Karnjanawasee, 2012; Pasiphol, 2016) ทั้งนี้จากปัญหาของโรงเรียนที่เกิดขึ้นผู้วิจัยจึงสนใจศึกษาวิธีการแก้ปัญหาการใช้แบบสอบซ้ำโดยการสร้างแบบสอบคู่ขนานซึ่งเป็นแบบสอบที่วัดในเนื้อหาสาระของข้อสอบคุณลักษณะตามเกณฑ์มาตรฐาน

การสร้างแบบสอบคู่ขนาน สามารถแบ่งตามการบริหารการสอบได้เป็น 2 รูปแบบ คือ (1) การสร้างแบบสอบคู่ขนานหลังการสอบ ข้อดีของรูปแบบนี้ คือ ทำให้มีความน่าเชื่อถือและมีความเป็นคู่ขนานของข้อสอบและแบบสอบตามเกณฑ์ความเป็นคู่ขนานที่ชัดเจน (Lin, 2008) และ (2) การสร้างแบบสอบคู่ขนานก่อนการสอบ ข้อดีของรูปแบบนี้ คือ สามารถสร้างข้อสอบในสภาพจริงได้ง่ายเนื่องจากใช้ผู้เชี่ยวชาญที่มีอยู่ในสาขานั้น ๆ และสร้างข้อสอบใหม่ได้ตามโครงสร้างและจุดประสงค์ที่กำหนดโดยไม่ต้องอาศัยค่าสถิติหรือพารามิเตอร์ของข้อสอบเดิมจากคลังข้อสอบ แต่ข้อเสียของวิธีนี้ คือ การพิจารณาคุณภาพของแบบสอบโดยผู้เชี่ยวชาญ ขึ้นอยู่กับระดับ

ความสามารถในการตัดสินใจรายบุคคลของผู้เชี่ยวชาญที่แตกต่างกัน (Pasiphol, 2016) ด้วยเหตุนี้ผู้วิจัยจึงเลือกใช้รูปแบบที่ 1 การสร้างแบบสอบคู่ขนานหลังการสอบ

เทคนิคในการสร้างข้อสอบคู่ขนานมี 5 เทคนิค คือ (1) สลับคำตอบมาเป็นคำถาม เป็นการนำคำตอบของฟอร์ม 1 มาเป็นคำถามของฟอร์ม 2 (2) เปลี่ยนคำถามตรงกันข้ามหรือเป็นปฏิเสธ โดยสร้างคำถามของฟอร์ม 1 เป็นข้อความทางบวก แต่คำถามของฟอร์ม 2 เป็นทางลบ (3) ใช้คำถามเหมือนกันแต่ตัวเลือกแตกต่างกัน โดยสร้างคำถามที่เหมือนกันทั้งฟอร์ม 1 และฟอร์ม 2 แต่ตัวเลือกมีความแตกต่างกันทั้งหมด (4) ใช้คำถามต่างกันแต่ตัวเลือกเหมือนกัน และ (5) เปลี่ยนถ้อยคำในคำถามและคำตอบแต่ถามในประเด็นเดียวกัน ซึ่งใช้ข้อความที่ต่างกัน แต่มีนัยเดียวกันในฟอร์ม 1 และฟอร์ม 2 และตัวเลือกก็มีความแตกต่างกันทั้ง 2 ฟอร์ม (Pasiphol, 2020) จาก 5 เทคนิคนี้ ผู้วิจัยสนใจที่จะประยุกต์ใช้เทคนิคในการสร้างแบบสอบคู่ขนาน 3 เทคนิคคือ (1) สลับคำตอบมาเป็นคำถาม (2) ใช้คำถามเหมือนกันแต่ตัวเลือกแตกต่างกัน และ (3) ใช้คำถามต่างกันแต่ตัวเลือกเหมือนกัน เพราะหากใช้เปลี่ยนคำถามตรงกันข้ามหรือเป็นปฏิเสธ และเปลี่ยนถ้อยคำในคำถามและคำตอบแต่ถามในประเด็นเดียวกัน จะทำให้ข้อสอบมีความยากเกินไปสำหรับผู้เรียนในระดับชั้นประถมศึกษาปีที่ 2

จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องกับการตรวจสอบความเป็นคู่ขนานของแบบสอบที่สร้างขึ้น สามารถสรุปได้ 3 วิธี คือ (1) การตรวจสอบความเป็นคู่ขนานโดยอาศัยผู้เชี่ยวชาญ (2) การตรวจสอบความเป็นคู่ขนานของแบบสอบโดยอาศัยข้อตกลงเบื้องต้นของทฤษฎีการทดสอบแบบดั้งเดิม กำหนดให้แบบสอบคู่ขนาน 2 ฉบับจะต้องมีคะแนนจริงของผู้สอบแต่ละคนและค่าความคลาดเคลื่อนมาตรฐานของประชากรที่สอบทั้ง 2 ฉบับเท่ากัน (Lord & Novick, 1968; Lord, 1980) โดยคะแนนจริงมีค่าเท่ากับคะแนนเฉลี่ยของคะแนนที่ได้จากการวัดซ้ำหลาย ๆ ครั้ง ซึ่งเป็นการพิจารณาค่าสถิติหลังจากการสอบ หากจะนำแบบสอบคู่ขนานนี้ไปใช้ในการสอบครั้งต่อไปต้องใช้แบบสอบซ้ำและความเที่ยงของแบบสอบคู่ขนานอาจเปลี่ยนไปตามประชากรที่แตกต่างกัน ซึ่งขาดสารสนเทศในการพิจารณาความเป็นคู่ขนานของข้อสอบรายข้อด้วย (Dorans et al., 2007) และ (3) การตรวจสอบความเป็นคู่ขนานโดยใช้ทฤษฎีการตอบสนองข้อสอบมาประยุกต์ใช้ในการสร้างแบบสอบคู่ขนาน ซึ่งจากการศึกษางานวิจัยในอดีตที่ผ่านมา พบว่า มีการประยุกต์ใช้ทฤษฎีการตอบสนองข้อสอบในการวิเคราะห์ความเป็นคู่ขนาน จำนวน 3 โมเดล ได้แก่ โมเดล 1 พารามิเตอร์ โมเดล 2 พารามิเตอร์ และ โมเดล 3 พารามิเตอร์ และเมื่อพิจารณาจากขนาดของกลุ่มตัวอย่างของนักเรียนที่อยู่ในสังกัดกรุงเทพมหานคร ในเขตหนองจอก จำนวน 800 คน ซึ่งเป็นจำนวนของกลุ่มตัวอย่างที่เหมาะสมจะวิเคราะห์โดยใช้โมเดล 2 พารามิเตอร์ในการตรวจสอบความเป็นคู่ขนาน เนื่องจาก Foley (2010) ได้กล่าวว่า ขนาดกลุ่มตัวอย่างที่เหมาะสม ลักษณะของตัวเลือก และจำนวนข้อสอบ สำหรับการวิเคราะห์โดยใช้โมเดลแบบ 2 พารามิเตอร์คือ 600 ถึง 1,200 คน

ดังนั้นจากสภาพปัญหาที่ได้กล่าวมาข้างต้น ผู้วิจัยจึงสนใจที่ศึกษาวิธีการแก้ปัญหาการใช้แบบสอบซ้ำ โดยการสร้างแบบสอบคู่ขนานวิชาคณิตศาสตร์ระดับชั้นประถมศึกษาปีที่ 2 สาระการวัด เนื่องจากเป็นสาระที่มีผลสัมฤทธิ์ทางการเรียนค่อนข้างต่ำ และเหตุผลในการเลือกเรื่องความยาว และน้ำหนัก เพราะนักเรียนเกิดความสับสนในหน่วยการวัด โดยรูปแบบที่ใช้ในการวิจัยเป็นการสร้างแบบสอบคู่ขนานหลังการสอบ โดยใช้เทคนิคในการสร้าง 3 เทคนิค (1) สลับคำตอบมาเป็นคำถาม (2) ใช้คำถามเหมือนกันแต่ตัวเลือกแตกต่างกัน และ (3) ใช้คำถามต่างกันแต่ตัวเลือกเหมือนกัน ส่วนการตรวจสอบคุณภาพของแบบสอบคู่ขนานประยุกต์ใช้โมเดลการตอบสนอง

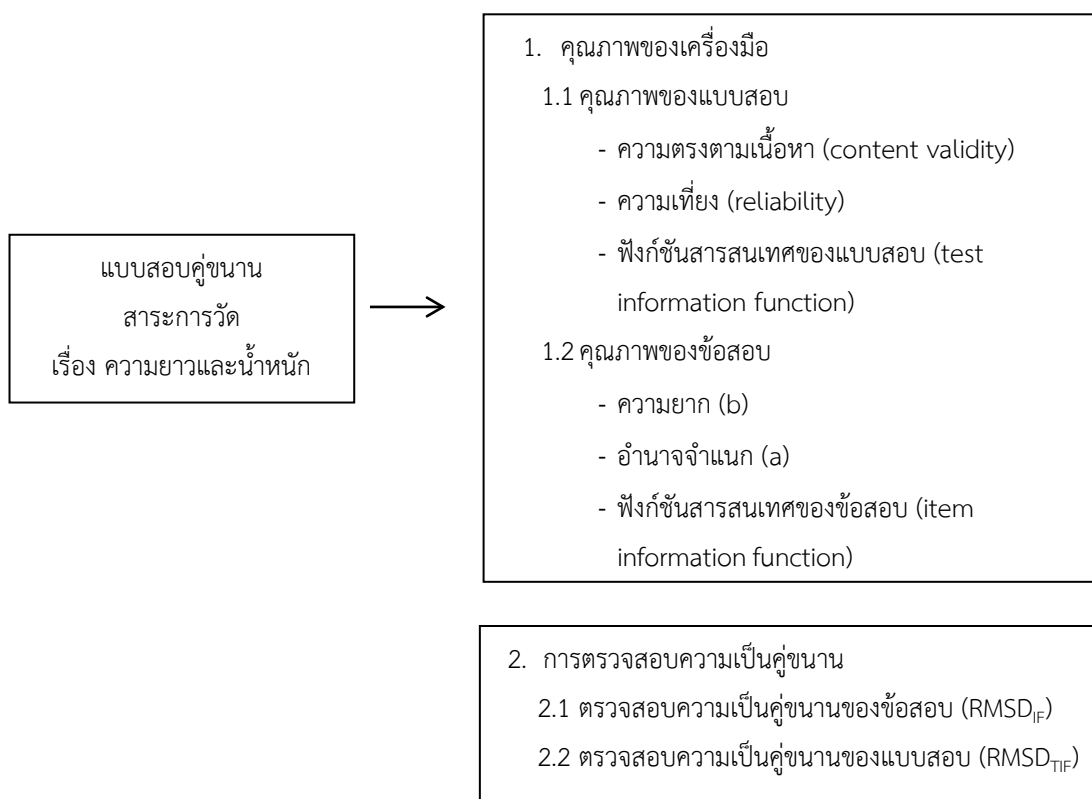
ข้อสอบ 2 พารามิเตอร์ ซึ่งประกอบไปด้วยการวิเคราะห์ คุณภาพของแบบสอบ คุณภาพของข้อสอบ ความเป็นคู่ขนานของข้อสอบ (RMSD_{IF}) และความเป็นคู่ขนานของแบบสอบ (RMSD_{TIF})

ความมุ่งหมายของการวิจัย

1. เพื่อพัฒนาแบบสอบคู่ขนานวิชาคณิตศาสตร์ระดับชั้นประถมศึกษาปีที่ 2 สาระการวัด
2. เพื่อตรวจสอบความเป็นคู่ขนานของแบบสอบคู่ขนานวิชาคณิตศาสตร์ ระดับชั้นประถมศึกษาปีที่ 2 สาระการวัด

กรอบแนวคิดในการวิจัย

จากการศึกษาแนวคิด หลักการและงานวิจัยที่เกี่ยวข้องดังกล่าว สามารถกำหนดเป็นกรอบแนวคิด การวิจัย การพัฒนาแบบสอบคู่ขนานวิชาคณิตศาสตร์ สำหรับนักเรียนชั้นประถมศึกษาปีที่ 2 โรงเรียนสังกัด กรุงเทพมหานครโดยใช้ทฤษฎีการตอบสนองข้อสอบแบบ 2 พารามิเตอร์ ได้ดังภาพ 1



ภาพ 1 กรอบแนวคิดในการวิจัย

วิธีดำเนินการวิจัย

1. ประชากรและกลุ่มตัวอย่าง

ประชากรที่ใช้ในการวิจัยครั้งนี้ เป็นนักเรียนชั้นประถมศึกษาปีที่ 2 ปีการศึกษา 2562 จากโรงเรียน สังกัด กรุงเทพมหานคร ในเขตหนองจอก 37 โรงเรียน จำนวน 1,853 คน

กลุ่มตัวอย่างที่ใช้ในการวิจัย เป็นนักเรียนชั้นประถมศึกษาปีที่ 2 ปีการศึกษา 2562 จากโรงเรียน สังกัด กรุงเทพมหานคร ในเขตหนองจอก 13 โรงเรียน จำนวน 800 คน จำแนกตามขนาดโรงเรียนพบว่าโรงเรียนขนาดใหญ่พิเศษคิดเป็นร้อยละ 24.63 ขนาดใหญ่ 23.12 ขนาดกลาง 22.13 ขนาดเล็ก 30.12 และเมื่อจำแนกตามเพศพบว่าส่วนใหญ่เป็นเพศหญิงคิดเป็นร้อยละ 63.37 มากกว่าเพศชายคิดเป็นร้อยละ 36.63 ซึ่งขนาดตัวอย่างที่เหมาะสมสำหรับการวิเคราะห์โดยใช้โมเดลแบบ 2 พารามิเตอร์ คือ 600 ถึง 1,200 คน (Foley, 2010) โดยตัวอย่างที่ได้มาจากการสุ่มแบบแบ่งชั้น (stratified random sampling)

2. เครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้ในการวิจัยเป็นแบบสอบถามแบบเลือกตอบ 3 ตัวเลือกที่คู่ขนานกันในรายวิชาคณิตศาสตร์ สารการวัดที่มีการให้คะแนนแบบ 2 ค่า หรือ 0,1 ซึ่งแบบสอบถามในการวิจัยครั้งนี้ประกอบด้วยแบบสอบถามจำนวน 4 ฉบับ คือ 1A, 1B, 2A และ 2B โดยจำแนกตามเนื้อหา คือ (1) ความยาว จำนวน 2 ฉบับ คือ 1A และ 1B ฉบับละ 40 ข้อ และคัดเลือกเหลือฉบับละ 20 ข้อ และ (2) น้ำหนัก จำนวน 2 ฉบับ คือ 2A และ 2B ฉบับละ 40 ข้อ และคัดเลือกเหลือฉบับละ 20 ข้อ รายละเอียดดังตาราง 1

ตาราง 1 เครื่องมือที่ใช้ในการวิจัยจำแนกตามเนื้อหา

เนื้อหา	สร้าง (ข้อ)		คัดเลือก (ข้อ)	
	ชุดA	ชุดB	ชุดA	ชุดB
1.ความยาว (1A,1B)	40	40	20	20
2.น้ำหนัก (2A,2B)	40	40	20	20
รวม	80	80	40	40
รวมทั้งหมด	160		80	

3. การเก็บรวบรวมข้อมูล

1. นำแบบสอบถามคู่ขนาน เรื่องการวัด ทั้ง 4 ฉบับ ไปทดสอบกับกลุ่มตัวอย่างทั้ง 800 คน โดยอธิบายให้นักเรียนได้ทราบถึงวัตถุประสงค์ของการสอบในครั้งนี้และขอความร่วมมือในการทำแบบสอบ เพื่อให้ นักเรียนได้เข้าใจถึงวิธีการทำแบบสอบ วิธีการตอบ ก่อนที่จะให้ทุกคนเริ่มต้นทำแบบสอบ โดยทำการสอบ ชุด 1A และ 2A ก่อน และผ่านไปหนึ่งสัปดาห์ แล้วสอบชุด 1B และ 2B

2. ตรวจให้คะแนน โดยมีเกณฑ์สำหรับการให้คะแนน คือ ตอบถูกให้ 1 คะแนน และตอบผิดหรือไม่ตอบให้ 0 คะแนน

3. นำผลที่ได้จากการทดสอบมาวิเคราะห์หาคุณภาพของข้อสอบ แบบสอบ และความเป็นคู่ขนานของข้อสอบและแบบสอบ

4. การวิเคราะห์ข้อมูล

4.1 การวิเคราะห์คุณภาพของแบบสอบ

คุณภาพของแบบสอบ พิจารณาจาก ความตรงตามเนื้อหา (content validity) โดยผู้เชี่ยวชาญจำนวน 3 คน เป็นผู้เชี่ยวชาญด้านการศึกษาคณิตศาสตร์จำนวน 2 คน และด้านการวัดและประเมินผล 1 คน ความเที่ยง

(reliability) ของแบบสอบ ใช้วิธีของ Kuder-Richardson (KR-20) และฟังก์ชันสารสนเทศของแบบสอบ (test information function) โดยใช้โปรแกรม MULTILOG 7.03

4.2 การวิเคราะห์คุณภาพของข้อสอบ

คุณภาพของข้อสอบ พิจารณาจาก ความยาก (b) อำนาจจำแนก (a) ฟังก์ชันสารสนเทศของข้อสอบ (item information function) โดยใช้โปรแกรม MULTILOG 7.03

4.3 การวิเคราะห์ความเป็นคู่ขนาน

การวิเคราะห์ความเป็นคู่ขนานรายข้อ พิจารณาจากค่าส่วนเบี่ยงเบนค่าเฉลี่ยกำลังสองของสารสนเทศของข้อสอบ (Root mean square deviation of item information ; $RMSD_{IF}$) โดยจะถือว่าข้อสอบคู่ขนานกัน ค่า $RMSD_{IF}$ ควรมีค่าไม่เกิน 0.50 และการวิเคราะห์ความเป็นคู่ขนานของแบบสอบ พิจารณาจากค่าส่วนเบี่ยงเบนค่าเฉลี่ยกำลังสองของสารสนเทศของแบบสอบ (Root mean square deviation of test information; $RMSD_{TIF}$) โดยจะถือว่าข้อสอบคู่ขนานกัน ค่า $RMSD_{TIF}$ ควรมีค่าไม่เกิน 0.50 มีสูตรการคำนวณดังนี้

การวิเคราะห์ความเป็นคู่ขนานรายข้อ

$$RMSD_{IF} = \sqrt{\frac{\sum_{j=1}^q [i(\theta_j) - i'(\theta_j)]^2}{q}}$$

การวิเคราะห์ความเป็นคู่ขนานของแบบสอบ

$$RMSD_{TIF} = \sqrt{\frac{\sum_{j=1}^q [I(\theta_j) - I'(\theta_j)]^2}{q}}$$

เมื่อ q แทน จำนวนระดับความสามารถที่ใช้ในการตรวจสอบ

$i(\theta)$ แทน ค่าสารสนเทศของข้อสอบฉบับ A ที่มีระดับความสามารถ θ

$i'(\theta)$ แทน ค่าสารสนเทศของข้อสอบฉบับ B ที่มีระดับความสามารถ θ

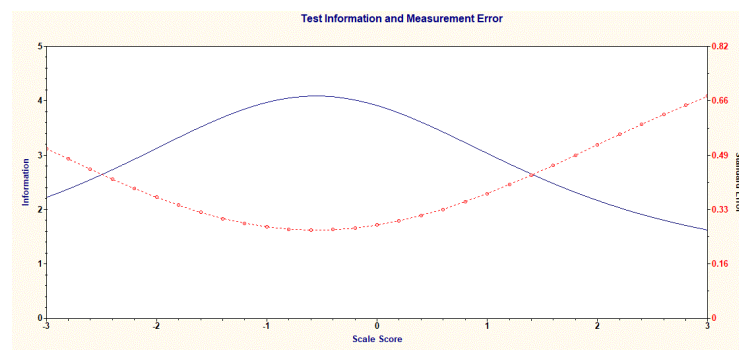
ผลการวิจัย

1. ผลการพัฒนาแบบสอบคู่ขนานวิชาคณิตศาสตร์เมื่อพิจารณาคุณภาพของแบบสอบมีดังนี้

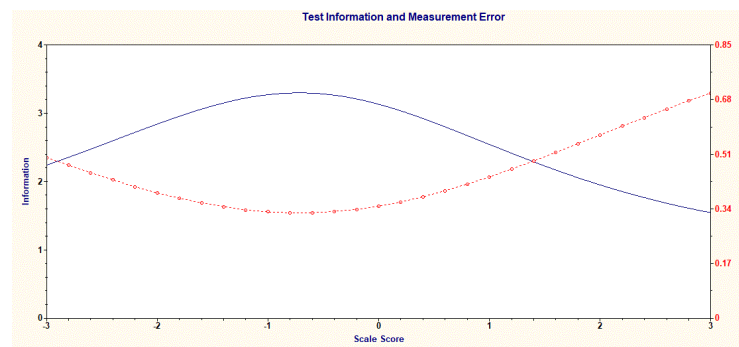
1.1 ความตรงตามเนื้อหาของแบบสอบคู่ขนาน ทั้ง 4 ฉบับ จำนวน 160 ข้อ มีค่า IOC (index of item objective congruence) ตั้งแต่ 0.67 ขึ้นไปถือว่าผ่านเกณฑ์ และได้คัดเลือกให้เหลือ 80 ข้อ

1.2 ความเที่ยงของแบบสอบทั้ง 4 ฉบับที่ผ่านการคัดเลือกเหลือ 80 ข้อ พบว่า ฉบับ 1A มีค่า 0.711, ฉบับ 1B มีค่า 0.652, ฉบับ 2A มีค่า 0.718 และ ฉบับ 2B มีค่า 0.635 ซึ่งถือว่ามีความเที่ยงในระดับปานกลาง

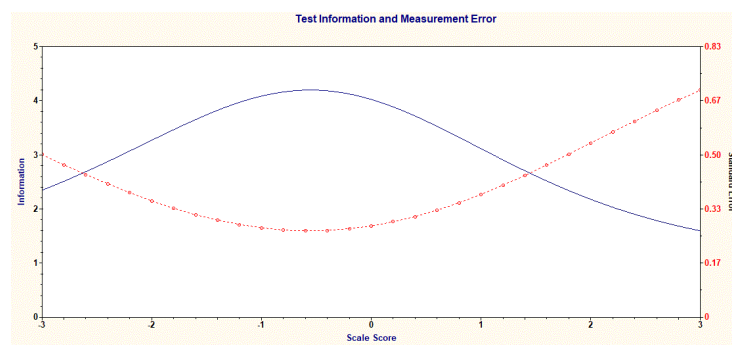
1.3 ค่าฟังก์ชันสารสนเทศของแบบสอบ (Test Information Function) ณ ช่วงความสามารถอยู่ระหว่าง -3 ถึง 3 พบว่า (1) แบบสอบฉบับ 1A มีค่าสูงสุด เท่ากับ 4.083 ณ ระดับความสามารถ เท่ากับ -0.6 (2) แบบสอบฉบับ 1B มีค่าสูงสุด เท่ากับ 3.293 ณ ระดับความสามารถ เท่ากับ -0.8 (3) แบบสอบฉบับ 2A มีค่าสูงสุด เท่ากับ 4.106 ณ ระดับความสามารถ เท่ากับ -0.4 และ (4) แบบสอบฉบับ 2B มีค่าสูงสุด เท่ากับ 3.498 ณ ระดับความสามารถ เท่ากับ -0.6 โดยมีผลการวิเคราะห์ฟังก์ชันสารสนเทศของแบบสอบ ดังภาพ 2 - 5



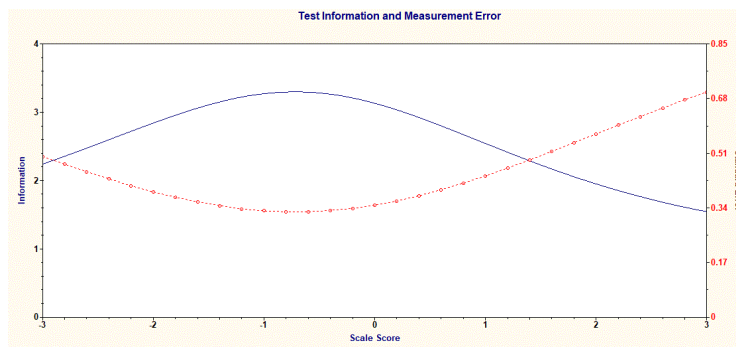
ภาพ 2 ฟังก์ชันสารสนเทศของแบบทดสอบ ฉบับ 1A



ภาพ 3 ฟังก์ชันสารสนเทศของแบบทดสอบ ฉบับ 1B



ภาพ 4 ฟังก์ชันสารสนเทศของแบบทดสอบ ฉบับ 2A



ภาพ 5 ฟังก์ชันสารสนเทศของแบบทดสอบ ฉบับ 2B

1.2 ผลการพัฒนาแบบสอบคู่ขนานวิชาคณิตศาสตร์เมื่อพิจารณาคุณภาพของข้อสอบมีดังนี้

1.2.1 ผลการวิเคราะห์คุณภาพรายข้อ เรื่องความยาว ฉบับ 1A มีค่าความยาก (b) อยู่ระหว่าง -2.16 ถึง 0.57 ค่าอำนาจจำแนก (a) อยู่ระหว่าง 0.30 ถึง 1.13 แปลผลได้ว่า ข้อสอบผ่านเกณฑ์ทุกข้อ ส่วนแบบสอบ ฉบับ 1B มีค่าความยาก (b) อยู่ระหว่าง -2.98 ถึง -0.29 ค่าอำนาจจำแนก (a) อยู่ระหว่าง 0.23 ถึง 1.06 ซึ่งมีข้อสอบ ไม่ผ่านเกณฑ์ 2 ข้อ คือ ข้อที่ 1 มีค่าความยากเท่ากับ -2.98 ข้อที่ 20 มีค่าอำนาจจำแนกน้อยกว่า 0.3

1.2.2 ผลการวิเคราะห์คุณภาพรายข้อ เรื่องน้ำหนัก ฉบับ 2A มีค่าความยาก (b) อยู่ระหว่าง -1.81 ถึง 0.00 ค่าอำนาจจำแนก (a) อยู่ระหว่าง 0.30 ถึง 1.08 แปลผลได้ว่า ข้อสอบผ่านเกณฑ์ทุกข้อ ส่วนแบบสอบ ฉบับ 2B มีค่าความยาก (b) อยู่ระหว่าง -2.90 ถึง -0.73 ค่าอำนาจจำแนก (a) อยู่ระหว่าง 0.22 ถึง 0.86 ซึ่งมีข้อสอบที่ไม่ผ่านเกณฑ์จำนวน 5 ข้อ คือ ข้อที่ 1, 4, 9, 19 และ 20 โดยข้อที่ 1 และข้อที่ 20 มีค่าความยาก เท่ากับ -2.90 และ -2.77 ตามลำดับ ส่วนข้อที่ 4, 9, 19 และ 20 มีค่าอำนาจจำแนกน้อยกว่า 0.3

1.2.3 ผลการวิเคราะห์ค่าฟังก์ชันสารสนเทศของข้อสอบ (item information function) ณ ช่วงความสามารถอยู่ระหว่าง -3 ถึง 3

แบบสอบฉบับ 1A จำนวน 20 ข้อ โดยข้อที่ค่าฟังก์ชันสารสนเทศของข้อสอบมากที่สุดคือ ข้อ 10 มีค่า 0.320 และข้อที่ค่าฟังก์ชันสารสนเทศของข้อสอบน้อยที่สุด คือ ข้อ 4 มีค่า 0.023

แบบสอบฉบับ 1B โดยข้อที่ค่าฟังก์ชันสารสนเทศของข้อสอบมากที่สุดคือ ข้อ 17 มีค่า 0.281 และ ข้อที่ค่าฟังก์ชันสารสนเทศของข้อสอบน้อยที่สุด ข้อ 20 มีค่า 0.013

แบบสอบฉบับ 2A โดยข้อที่ค่าฟังก์ชันสารสนเทศของข้อสอบมากที่สุดคือ ข้อ 14 มีค่า 0.291 และ ข้อที่ค่าฟังก์ชันสารสนเทศของข้อสอบน้อยที่สุด คือ ข้อ 20 มีค่า 0.022

แบบสอบฉบับ 2B โดยข้อที่ค่าฟังก์ชันสารสนเทศของข้อสอบมากที่สุดคือ ข้อ 5 มีค่า 0.186 และ ข้อที่ค่าฟังก์ชันสารสนเทศของข้อสอบน้อยที่สุด คือ ข้อ 4 มีค่า 0.013

2. ผลการตรวจสอบความเป็นคู่ขนานของแบบสอบคู่ขนานวิชาคณิตศาสตร์ระดับชั้นประถมศึกษา ปีที่ 2 สารการวัด

2.1 ผลการวิเคราะห์ความเป็นคู่ขนานรายข้อและทั้งฉบับของแบบสอบเรื่อง ความยาว จำนวน 2 ฉบับ (1A และ 1B) ฉบับละ 20 ข้อ พบว่า ผลการตรวจสอบความเป็นคู่ขนานรายข้อเมื่อพิจารณาจากค่า $RMSD_{IF}$ มีค่า

ตั้งแต่ 0.008 ถึง 0.167 โดยค่า $RMSD_{IF}$ ไม่เกิน 0.50 สะท้อนให้เห็นว่าข้อสอบทั้ง 20 ข้อ มีความเป็นคู่ขนานกัน ขณะที่ผลการตรวจสอบความเป็นคู่ขนานของแบบสอบทั้งฉบับเมื่อพิจารณาจากค่า $RMSD_{TIF}$ มีค่าเท่ากับ 0.484 โดยค่า $RMSD_{TIF}$ มีค่าไม่เกิน 0.50 สะท้อนให้เห็นว่าแบบสอบฉบับ 1A และ 1B มีความเป็นคู่ขนานกัน

2.2 ผลการวิเคราะห์ความเป็นคู่ขนานรายข้อและทั้งฉบับของแบบสอบเรื่อง น้ำหนัก จำนวน 2 ฉบับ (2A และ 2B) ฉบับละ 20 ข้อ พบว่า ผลการตรวจสอบความเป็นคู่ขนานรายข้อเมื่อพิจารณาจากค่า $RMSD_{IF}$ มีค่าตั้งแต่ 0.020 ถึง 0.110 โดยค่า $RMSD_{IF}$ ไม่เกิน 0.50 สะท้อนให้เห็นว่าข้อสอบทั้ง 20 ข้อ มีความเป็นคู่ขนานกันขณะที่ผลการตรวจสอบความเป็นคู่ขนานของแบบสอบทั้งฉบับเมื่อพิจารณาจากค่า $RMSD_{TIF}$ มีค่าเท่ากับ 0.492 โดยค่า $RMSD_{TIF}$ มีค่าไม่เกิน 0.50 สะท้อนให้เห็นว่าแบบสอบฉบับ 2A และ 2B มีความเป็นคู่ขนานกัน รายละเอียดดังตาราง 2

ตาราง 2 ผลการตรวจสอบความเป็นคู่ขนานของแบบสอบคู่ขนานวิชาคณิตศาสตร์ระดับชั้นประถมศึกษา
ปีที่ 2 สาระ การวัด

การตรวจสอบความเป็นคู่ขนาน	รายข้อ ($RMSD_{IF}$)	ทั้งฉบับ ($RMSD_{TIF}$)	การแปลผล
แบบสอบเรื่อง ความยาว (ฉบับ 1A, 1B)	0.008 – 0.167	0.484	มีความคู่ขนานทั้งรายข้อและทั้งฉบับ
แบบสอบเรื่อง น้ำหนัก (ฉบับ 2A, 2B)	0.020 – 0.110	0.492	มีความคู่ขนานทั้งรายข้อและทั้งฉบับ

อภิปรายผล

1. อภิปรายผลการพัฒนาแบบสอบคู่ขนานวิชาคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 2 สาระการวัด

จากผลการวิจัยที่พบว่า ความตรงตามเนื้อหาของแบบสอบคู่ขนาน ทั้ง 4 ฉบับ จำนวน 160 ข้อ มีค่า IOC ตั้งแต่ 0.67 สะท้อนให้เห็นว่าแบบสอบทั้ง 4 ฉบับ ประกอบด้วยข้อสอบที่มีความตรงตามเนื้อหา ทั้งนี้เพราะผู้วิจัยได้ศึกษาโครงสร้างและการสร้างกรอบการประเมินที่ชัดเจน เมื่อนำมาให้ผู้เชี่ยวชาญพิจารณาประเมิน เครื่องมือจึงมีความตรงตามเกณฑ์ที่กำหนด ซึ่งค่า IOC ที่ได้สอดคล้องกับเกณฑ์ที่ใช้ในการตัดสินความตรงตามเนื้อหาควรมีค่าตั้งแต่ 0.50 ขึ้นไป (Karnjanawasee, 2013)

จากผลการวิเคราะห์ค่าความเที่ยงของแบบสอบทั้ง 4 ฉบับที่ผ่านการคัดเลือกเหลือ 80 ข้อ พบว่า ฉบับ 1A มีค่า 0.711, ฉบับ 1B มีค่า 0.652, ฉบับ 2A มีค่า 0.718 และ ฉบับ 2B มีค่า 0.635 ซึ่งมีความเที่ยงในระดับปานกลาง สอดคล้องกับ Katsa & Amonrattanasak (2008); Karnjanawasee (2012) ที่กล่าวว่า แบบสอบที่ใช้ในการทดสอบ ควรมีค่าสัมประสิทธิ์ความเที่ยงสูงที่สุดเท่าที่จะเป็นไปได้ภายใต้สภาพการณ์นั้น อย่างน้อยที่สุดควรมีค่าไม่ต่ำกว่า 0.5 แต่อย่างไรก็ดีแม้ว่าค่าความเที่ยงของแบบสอบในการวิจัยครั้งนี้จะผ่านเกณฑ์แต่ก็ยังไม่สูงมากนัก ทั้งนี้อาจเป็นผลมาจากความยาวของข้อสอบซึ่งประกอบไปด้วยข้อสอบฉบับละ 20 ข้อ ซึ่งความยาวของข้อสอบ จำนวนตัวเลือก 3 ตัวเลือก และข้อสอบชุด B จะมีความเที่ยงต่ำกว่าชุด A อาจเป็นเพราะชุด B ถูกจัดให้ทำหลังชุด A กลุ่มตัวอย่างที่ทำข้อสอบอาจจะเกิดความเหนื่อยล้าในการทำแบบสอบ ซึ่งน่าจะเป็นเป็นปัจจัยที่ส่งผลต่อค่าความเที่ยง

จากผลการวิเคราะห์ค่าฟังก์ชันสารสนเทศของแบบสอบ (test information function) ณ ช่วงความสามารถอยู่ระหว่าง -3 ถึง 3 พบว่า แบบสอบฉบับ 1A มีค่าสูงสุด เท่ากับ 4.083 ณ ระดับความสามารถ เท่ากับ -0.60 และแบบสอบฉบับ 1B มีค่าสูงสุด เท่ากับ 3.293 ณ ระดับความสามารถ เท่ากับ -0.80 สะท้อนให้เห็นว่าแบบสอบ 1A และ 1B ให้สารสนเทศสูงสุดสำหรับผู้ที่มีระดับความสามารถระหว่าง -0.60 ถึง -0.80 หรือแบบสอบทั้ง 2 ฉบับนี้เหมาะสำหรับผู้สอบกลุ่มที่มีความสามารถปานกลางไปจนถึงค่อนข้างอ่อน ขณะที่แบบสอบฉบับ 2A มีค่าสูงสุด เท่ากับ 4.106 ณ ระดับความสามารถ เท่ากับ -0.40 และแบบสอบฉบับ 2B มีค่าสูงสุด เท่ากับ 3.498 ณ ระดับความสามารถ เท่ากับ -0.60 สะท้อนให้เห็นว่าแบบสอบ 2A และ 2B ให้สารสนเทศสูงสุดสำหรับผู้ที่มีระดับความสามารถระหว่าง -0.40 ถึง -0.60 หรือแบบสอบทั้ง 2 ฉบับนี้เหมาะสำหรับผู้สอบกลุ่มที่มีความสามารถปานกลางไปจนถึงค่อนข้างอ่อน ซึ่งในการวิจัยครั้งนี้จะสังเกตได้ว่า ค่าฟังก์ชันสารสนเทศของแบบสอบทั้ง 2 ฉบับมีค่าใกล้เคียงกัน เกิดจากขั้นตอนการสร้างแบบสอบได้มีการศึกษาการสร้างตามโครงสร้างจำแนกระดับพฤติกรรมตรงตามตัวชี้วัดในเนื้อหาการวัดของนักเรียนระดับชั้นประถมศึกษาปีที่ 2 สะท้อนให้เห็นว่าการสร้างข้อสอบครั้งนี้ที่เป็นคู่ขนานทั้งเนื้อหา ค่าความยาก อำนาจจำแนกที่ใกล้เคียงกัน สอดคล้องกับงานวิจัยของ Kantivong (2017); Ali & Rijn (2016) ที่กล่าวว่าค่าสารสนเทศของแบบสอบเกิดจากประสบการณ์และความเชี่ยวชาญในการออกข้อสอบเกิดจากการสร้างแบบสอบให้มีคุณภาพความตรงความถูกต้องสูงขึ้น

จากผลการวิเคราะห์ค่าความยาก (b) ของข้อสอบคู่ขนานวิชาคณิตศาสตร์ สารการวัด ทั้ง 4 ฉบับจำนวน 80 ข้อ พบว่า ค่าความยาก มีค่าอยู่ระหว่าง -2.98 ถึง 0.00 ซึ่งไม่ผ่านเกณฑ์จำนวน 3 ข้อ จากข้อคำถามชุด B เพราะเป็นคำถามที่วัดพฤติกรรมด้านความจำ ดังนั้นต้องสร้างแบบสอบที่วัดพฤติกรรมที่สูงกว่าความจำ ส่วนข้อสอบที่มีความยากผ่านเกณฑ์แต่มีแนวโน้มว่าข้อสอบยากง่ายปานกลางค่อนข้างไปทางง่าย โดยสอดคล้องกับ Karnjanawasee (2013) ที่กล่าวว่า ข้อสอบที่ดีควรมีค่าความยากง่ายอยู่ในระดับปานกลางหรืออยู่ในช่วง -2.50 ถึง 2.50

จากผลการวิเคราะห์ค่าอำนาจจำแนก (a) ของข้อสอบคู่ขนานวิชาคณิตศาสตร์ สารการวัด ทั้ง 4 ฉบับจำนวน 80 ข้อ พบว่า ค่าอำนาจจำแนก มีค่าอยู่ระหว่าง 0.22 ถึง 1.13 จากข้อสอบทั้งหมดที่ได้มีการวิเคราะห์แล้วนั้น มีข้อสอบจำนวน 5 ข้อที่มีค่าอำนาจจำแนกน้อยกว่า 0.3 สอดคล้องกับงานวิจัยของ Sawangsri (2015) ที่พบว่า ค่าอำนาจจำแนกที่ผ่านเกณฑ์ควรมากกว่า 0.3 เพราะข้อสอบที่มีค่าน้อยกว่านี้แสดงว่าข้อสอบนั้นไม่ดีและควรตัดออก ส่วนข้อสอบที่มีค่าอำนาจจำแนกผ่านเกณฑ์ 0.3 ขึ้นไปจำนวน 75 ข้อ สอดคล้องกับ Reckase (2009) ที่กล่าวว่าค่าอำนาจจำแนกควรมีค่าอยู่ระหว่าง 0.30 ถึง 2.00 และสอดคล้องกับ Karnjanawasee (2012) ที่ระบุว่าค่าอำนาจจำแนก (a) ควรมีค่าเป็นบวกและในทางปฏิบัติควรมีค่าระหว่าง 0.50 ถึง 2.50

จากผลการวิเคราะห์ค่าฟังก์ชันสารสนเทศของข้อสอบ (item information function) ณ ช่วงความสามารถอยู่ระหว่าง -3 ถึง 3 แบบสอบฉบับ 1A ข้อที่ค่าฟังก์ชันสารสนเทศของข้อสอบมากที่สุดคือ ข้อ 10 และน้อยที่สุดคือ ข้อ 4 แบบสอบฉบับ 1B ข้อที่ค่าฟังก์ชันสารสนเทศของข้อสอบมากที่สุดคือ ข้อ 17 และน้อยที่สุด คือ ข้อ 20 แบบสอบฉบับ 2A ข้อที่ค่าฟังก์ชันสารสนเทศของข้อสอบมากที่สุดคือ ข้อ 14 และน้อยที่สุด คือ ข้อ 20 และแบบสอบฉบับ 2B ข้อที่ค่าฟังก์ชันสารสนเทศของข้อสอบมากที่สุดคือ ข้อ 5 และน้อยที่สุด คือ ข้อ 4 ทั้งนี้ข้อที่มีค่าฟังก์ชันสารสนเทศของข้อสอบน้อยที่สุดนั้นจะสัมพันธ์กับค่าความยากและอำนาจจำแนกที่ไม่ผ่านเกณฑ์ด้วย เช่น ข้อ 20 ของฉบับ 1B ที่พบว่าค่าฟังก์ชันสารสนเทศของข้อสอบน้อยที่สุด ปรากฏว่าเป็นข้อที่มีอำนาจจำแนก

น้อยกว่า 0.3 ด้วย ทั้งนี้สอดคล้องกับ Karnjanawasee (2012) ที่ได้กล่าวว่าฟังก์ชันสารสนเทศข้อสอบเป็นดัชนีผสมที่สร้างจากดัชนีคุณลักษณะของข้อสอบหลายลักษณะประกอบด้วยค่าพารามิเตอร์ความยาก อำนาจจำแนก และค่าความแปรปรวนของคะแนนรายข้อเพื่อบ่งชี้คุณภาพข้อสอบ สอดคล้องกับงานวิจัยของ Kantivong (2017); Sawangsri (2015) ซึ่งประยุกต์ใช้โมเดลแบบ 2 พารามิเตอร์ ที่มีการพิจารณาค่าความยาก และค่าอำนาจจำแนกของข้อสอบซึ่งสัมพันธ์กับค่าฟังก์ชันสารสนเทศของข้อสอบและแบบสอบ

2.อธิบายผลการตรวจสอบความเป็นคู่ขนานของข้อสอบและแบบสอบ วิชาคณิตศาสตร์ระดับชั้นประถมศึกษาปีที่ 2 สารการวัด

จากผลการวิเคราะห์ความเป็นคู่ขนานรายข้อและทั้งฉบับของแบบสอบเรื่อง ความยาว จำนวน 2 ฉบับ (1A และ 1B) ที่พบว่า ข้อสอบทั้ง 20 ข้อ มีความเป็นคู่ขนานกัน และแบบสอบทั้ง 2 ฉบับ คือ ฉบับ 1A และ 1B คู่ขนานกัน และความเป็นคู่ขนานรายข้อและทั้งฉบับของแบบสอบเรื่อง น้ำหนัก จำนวน 2 ฉบับ (2A และ 2B) ที่พบว่า ข้อสอบทั้ง 20 ข้อ มีความเป็นคู่ขนานกัน และแบบสอบทั้ง 2 ฉบับ คือ ฉบับ 2A และ 2B คู่ขนานกัน สะท้อนให้เห็นว่า ข้อสอบที่ผู้วิจัยสร้างขึ้นนั้นมีความเป็นคู่ขนานในรายข้อส่งผลให้เกิดความคู่ขนานในภาพรวม คือความเป็นคู่ขนานของแบบสอบทั้งฉบับ สอดคล้องกับงานวิจัยของ Kantivong (2017); Debeer et al. (2017) ที่ให้ความสำคัญต่อความเป็นคู่ขนานรายข้อ เพราะจะนำไปสู่ผลรวมของความเป็นคู่ขนานของแบบสอบทั้งฉบับ ซึ่งหากแบบสอบ 2 ฉบับที่คู่ขนานกันในรายข้อมีโอกาที่แบบสอบคู่นั้นจะคู่ขนานกันสูงขึ้น

ข้อเสนอแนะ

1. ข้อเสนอแนะในการนำผลการวิจัยไปใช้

1.1 ได้แบบสอบคู่ขนานวิชาคณิตศาสตร์ ระดับชั้นประถมศึกษาปีที่ 2 สารการวัด ที่ผ่านการตรวจสอบความเที่ยง ความตรงตามเนื้อหา ค่าฟังก์ชันสารสนเทศของข้อสอบและแบบสอบที่ผ่านเกณฑ์จำนวน 75 ข้อ ซึ่งสามารถนำไปใช้ในการจัดสร้างคลังข้อสอบวิชาคณิตศาสตร์ สารการวัด เรื่องความยาวและน้ำหนัก สำหรับนักเรียนชั้นประถมศึกษาปีที่ 2 เพื่อใช้ในการวัดและประเมินผู้เรียนก่อนเรียนและหลังเรียนได้

1.2 จากผลการวิเคราะห์ค่าความเที่ยงของแบบสอบทั้ง 4 ฉบับ ซึ่งมีค่า 0.711, 0.652, 0.718 และ 0.635 โดยมีค่าความเที่ยงในระดับปานกลาง ซึ่งอาจเป็นผลมาจากความยาวของข้อสอบซึ่งประกอบไปด้วยข้อสอบฉบับละ 20 ข้อ ซึ่งความยาวของข้อสอบ เป็นปัจจัยหนึ่งที่ส่งผลต่อค่าความเที่ยง ดังนั้นในการนำผลวิจัยไปใช้อาจพิจารณาเรื่องจำนวนของข้อสอบ เพื่อให้เป็นแนวทางการสร้างแบบสอบคู่ขนานให้มีความเที่ยงมากยิ่งขึ้น

2. ข้อเสนอแนะสำหรับการทำวิจัยครั้งต่อไป

2.1 การวิจัยครั้งนี้ ผู้วิจัยได้นำแบบสอบคู่ขนานไปใช้ในบริบทของการทำแบบสอบกับกระดาษ (paper-based testing) ซึ่งผลการวิจัยที่ได้อาจจะแตกต่างจากการทดสอบแบบออนไลน์ ดังนั้นในการวิจัยครั้งต่อไป ควรมีการสร้างแบบสอบในรูปแบบออนไลน์ เพื่อให้สอดคล้องกับบริบทและสถานการณ์การจัดการเรียนการสอนของยุคปกติใหม่

2.2 การวิจัยครั้งนี้ ผู้วิจัยวิเคราะห์ค่าพารามิเตอร์โดยใช้ทฤษฎีการตอบสนองข้อสอบแบบ 2 พารามิเตอร์ ในการวิจัยครั้งต่อไปควรมีการวิเคราะห์ค่าโอกาสในการเดา (c) โดยการประยุกต์ใช้ทฤษฎีการตอบสนองข้อสอบ

แบบ 3 พารามิเตอร์ เนื่องจากจะได้ข้อสอบแต่ละข้อที่มีความแตกต่างกันทั้งพารามิเตอร์ a, b และ c เพื่อลดความคลาดเคลื่อนของข้อสอบให้ได้มากที่สุด ซึ่งทำให้มีข้อสอบที่มีคุณภาพมากยิ่งขึ้น

2.3 การวิจัยในครั้งนี้เป็นการสร้างและพัฒนาแบบสอบคู่ขนานที่เป็นข้อสอบปรนัยทั้งหมด ดังนั้น ในการวิจัยครั้งต่อไปควรมีการสร้างและพัฒนาแบบสอบคู่ขนานในรูปแบบอัตนัยและแบบรูปผสม เพื่อให้ได้คลังข้อสอบของแบบสอบคู่ขนานที่มีความหลากหลายและมีคุณภาพสำหรับนำไปใช้ในโอกาสที่แตกต่างกัน

2.4 การวิจัยครั้งนี้ใช้เทคนิคในการสร้างข้อสอบ 3 เทคนิค ในการวิจัยครั้งต่อไป ควรมีการศึกษาเพิ่มเติมว่าเทคนิคการสร้างข้อสอบเทคนิคใดมีความเหมาะสมเพื่อยกระดับความเป็นคู่ขนานของแบบสอบ

References

- Ali, U. S., & Rijn, P. W. (2016). An evaluation of different statistical target for assembling parallel form in item response theory. *Applied psychological measurement*, 40(3), 163-179.
- Debeer, D., Ali, U. S., & Rijn, P. W. (2017). Evaluating Statistical Targets for Assembling Parallel Mixed-Format Test Forms. *Journal of Educational Measurement*, 54(2), 218-242.
- Dorans, N. J., Pommerich, M., & Holland, P. W. (2007). *Linking and aligning scores and scales*. Springer-Verlag.
- Foley, B. F. (2010). *Improving IRT item parameter estimates with small sample sizes: Evaluating the efficacy of a new data augmentation technique* [Doctoral dissertation]. University of Nebraska.
- Lin, C. J. (2008). Comparisons between classical theory and item response theory in automated assembly of parallel test forms. *The journal of technology, learning and assessment*.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Erlbaum.
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Addison-Wesley.
- Reckase, M. (2009). *Multidimensional item response theory*. Springer.
- Rupp, A., & Templin, J. (2008). *Effects of Q-matrix misspecification on parameter estimates and misclassification rates in the DINA model*. Educational and Psychological Measurement.

Translate Thai References

- Education Strategy Office Education Bureau. (2016). *Bangkok Basic Education Development Plan issue 2 (2017-2020)*. Agricultural Co-operative Federation of Thailand Press. (in Thai)
- Kanchanapenkul, S. (2011). *A Construction of Parallel Domainre-referenced test on "Circles" by Using Facet Design for Matthayousuksa 3* [Master's thesis]. Srinakharinwirot University. (in Thai)

- Kantivong, A. (2017). *Comparison of Item and Test Parallel Index Between Test Forms that Selected by Different Characteristics of Experts: Application of Formal Cultural Consensus Theory* [Master's thesis]. Chulalongkorn University. (in Thai)
- Karnjanawasee, S. (2012). *New Exam Theory* (4th Ed.). Chulalongkorn University Press. (in Thai)
- Karnjanawasee, S. (2013). *Classical Test Theory* (7th Ed.). Chulalongkorn University Press. (in Thai)
- Katsa, T., & Amonrattanasak, S. (2008). *Educational evaluation*. Ramkhamhaeng University. (in Thai)
- Pasiphol, S. (2016). *Learning Measurement and Evaluation*. Chulalongkorn University Press. (in Thai)
- Pasiphol, S. (2020). *The Development of Parallel Tests: Creation and Parallelism analysis*. Chulalongkorn University Press. (in Thai)
- Sawangboon, N. (2011). *A Construction of Mathematics Parallel Test Using Facet Design on "Factoring of Degree-2 Polynomials" for Matthayomsuksa 2 Students* [Master's thesis]. Mahasarakham University. (in Thai)
- Sawangsri, P. (2015). *Comparison of Psychometric Properties Integrating of Reading Analytical Thinking and Writing Abilities and Indicators of Content Science of Nine Grade Using The Different Item Review Methods* [Master's thesis]. Chulalongkorn University. (in Thai)