การพัฒนาวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา

Development of Item Pools Strategies by Using Interval a-parameter-Stratification with Content Balancing

ศักดิ์ชัย จันทะแสง[1*] เสรี ชัดแช้ม[1] ปิยะทิพย์ ประดุจพรม[2]

Sakchai Jantasang[1*] Seree Chadcham[1] Piyathip Pradujprom[2]

[1]*College of Research Methodogy and Cognitive Science, Burapha University, Thailand*

[2] *Cognitive Science and Innovation Research Unit: CSIRU, College of Research Methodology and Cognitive Science, Burapha University, Thailand*

บทคัดย่อ

การจัดคลังข้อสอบเป็นการจัดการข้อสอบให้เป็นระบบ หมวดหมู่ พร้อมสำหรับการนำไปใช้ การวิจัยนี้มีวัตถุประสงค์เพื่อ 1) พัฒนาวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา 2) เปรียบเทียบประสิทธิภาพของวิธีการที่พัฒนาขึ้นกับวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบถ่วงน้ำหนักที่มีการบังคับ โดยใช้สถานการณ์จำลองข้อมูล ในด้านการประมาณค่าความสามารถของผู้สอบ ได้แก่ ค่าความลำเอียงเฉลี่ย และค่ารากที่สองของความคลาดเคลื่อนกำลังสอง ด้วยการวิเคราะห์ Wilcoxon และด้านประสิทธิภาพของการใช้ข้อสอบ ได้แก่ ข้อสอบที่มีการแสดงมากเกินไป ข้อสอบที่มีการใช้น้อยเกินไป อัตราการทับซ้อนของข้อสอบ และการแจกแจงอัตราการแสดงข้อสอบ ด้วยการวิเคราะห์ Chi-Square

ผลการวิจัยปรากฏว่า 1) วิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา มี 3 ขั้นตอน ดังนี้ (1) แบ่งคลังข้อสอบเป็น 4 ชั้น ตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วง (2) จัดสมดุลเนื้อหาในแต่ละชั้นของคลังข้อสอบตามสาระการเรียนรู้ที่ออกข้อสอบ และ (3) ควบคุมการแสดงข้อสอบด้วยวิธีการสุ่มตัวอย่างแบบชั้นภูมิ 2) วิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา มีประสิทธิภาพสูงกว่า วิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบถ่วงน้ำหนักที่มีการบังคับ

คำสำคัญ: วิธีการจัดคลังข้อสอบ, ค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วง, การจัดสมดุลเนื้อหา

ABSTRACT

The item pools strategy is a procedure for organizing and categorizing test items to be ready for use. This research aimed to 1) develop the interval a-parameter-stratification with content balancing item pools strategy; 2) compare the efficiency of the developed strategy with the constraint-weighted a-stratification method using simulation scenario of test-takers' estimative efficiency, that is, average bias and root mean square error (RMSE), employing Wilcox Test, and

**Corresponding author. E-mail: amchai.j@gmail.com*

item utilizable efficiency, that is, overexposed items, underutilized items, item overlap rate, and item exposure rate distribution, using Chi-square test analysis.

The results showed that 1) there were three steps of the interval a-parameter-stratification with content balancing item pools strategy: First, dividing Item pools into 4 classes interval a-parameter-stratification. Second, balancing of each class of the item pools according to the subject matter of the test. Third, controlling Item exposure with Stratified Random Sampling; 2) the developed strategy with the test-takers' performance in each competency level of the discriminative power and content balancing method was more efficient than the constraint-weighted a-stratification method.

Keywords: item pools strategies, interval a-parameter test, content balancing

ความนำ

การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์จะต้องสร้างคลังข้อสอบที่มุ่งวัดคุณลักษณะหรือความสามารถของผู้สอบและครอบคลุมระดับความสามารถของผู้สอบที่แตกต่างกัน คลังข้อสอบ (Item pools) เป็นแหล่งรวบรวมข้อสอบหรือคำถามแบบทดสอบ การทดสอบด้วยแบบทดสอบดั้งเดิม ผู้สอบทั้งหมดจะได้รับข้อสอบชุดเดียวกัน สำหรับการทดสอบแบบปรับเหมาะผู้สอบจะได้รับข้อสอบเหมาะสมกับความสามารถของผู้สอบแต่ละคน คลังข้อสอบจึงเป็นเสมือนแหล่งรวบรวมชุดแบบทดสอบจำนวนมากและมีความเป็นคู่ขนานกัน ข้อสอบบรรจุในคลังข้อสอบมีลักษณะค่าความยากกระจายเต็มระดับความสามารถของประชากรผู้สอบ ข้อสอบควรมีจำนวนมากเพียงพอและกระจายทุกระดับความสามารถของผู้สอบ (Thompson & Weiss, 2011; Babcock & Weiss, 2013) การกำหนดขนาดของคลังข้อสอบพิจารณาจากปัจจัยหลายด้าน เช่น วิธีการคัดเลือกข้อสอบ การควบคุมเนื้อหาตามมาตรวัดทางจิตวิทยา กฎการยุติ การทับซ้อนของข้อสอบ การให้คะแนนแบบทดสอบ ความเป็นคู่ขนานกับแบบทดสอบดั้งเดิม จึงเป็นสิ่งสำคัญที่ควรคำนึงถึง (Reckase, 2009)

การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์จะให้ประสิทธิภาพ เมื่อคลังข้อสอบมีจำนวนข้อสอบมากเพียงพอสำหรับจำแนกผู้สอบ คลังข้อสอบที่ดีจะต้องมีจำนวนข้อสอบเท่ากับ เมื่อ n คือ จำนวนข้อสอบที่ต้องการใช้ทดสอบ เช่น ถ้าต้องการทดสอบประมาณ 10 ข้อ ต้องมีข้อสอบในคลัง 1,024 ข้อ แต่ถ้าหากต้องการทดสอบแบบกำหนดจำนวนข้อที่จะใช้คงที่เหมือนกันทุกคน จำนวนข้อสอบในคลังข้อสอบควรมีอย่างน้อย ข้อ เช่น ถ้าต้องทดสอบ 10 ข้อ ต้องมีข้อคำถามในคลัง จำนวน 55 ข้อ (Weiss, 2011) ซึ่ง De Ayla (2009) เสนอว่า คลังข้อสอบควรมีขนาดประมาณ 8–12 เท่า ของจำนวนข้อสอบที่ใช้ในการทดสอบ ถ้าในการทดสอบใช้ข้อสอบประมาณ 25 ข้อ คลังข้อสอบควรมีข้อสอบ 200–300 ข้อ และ Weiss (2011) ได้เสนอให้มีข้อสอบไม่น้อยกว่า 200 ข้อ Thompson and Weiss (2011) เสนอให้มีข้อสอบประมาณ 400 ข้อ และ Embretson and Reise (2000) กล่าวว่า คลังข้อสอบควรประกอบด้วยข้อสอบประมาณ 100 ข้อ ซึ่งหากรูปแบบการตอบข้อคำถามเป็นแบบที่ตรวจให้คะแนนมากกว่า 2 ค่า จำนวนข้อสอบในคลังข้อสอบสามารถมีจำนวนน้อยกว่านี้ได้

การพัฒนาคลังข้อสอบสำหรับการทดสอบแบบปรับเหมาะนั้น มีความคล้ายคลึงกับคลังข้อสอบแบบดั้งเดิม ข้อสอบจะต้องสร้างตามตารางกำหนดเนื้อหาและควรมีการตรวจสอบคุณภาพของเนื้อหาและความไวของแบบทดสอบ (Test sensitivity) Thompson and Weiss (2011) ได้เสนอขั้นตอน ดังนี้ 1) สร้างข้อสอบให้มีจำนวนมากพอที่จะครอบคลุมเนื้อหาทั้งหมดตามกฎเบื้องต้นในการเขียนข้อสอบ 2) ตรวจสอบคุณภาพของข้อสอบเพื่อให้ข้อสอบมีคุณภาพสูงสุด 3) นำข้อสอบที่สร้างขึ้นไปทดลองใช้ 4) คัดเลือกข้อสอบที่มีคุณภาพด้วยการวิเคราะห์ตาม

หลักสถิติจากการทดลองใช้ 5) ตรวจสอบสัดส่วนของข้อสอบในแต่ละเนื้อหาและประเมินกระบวนการทดสอบด้วยการจำลองสถานการณ์การทดสอบ (Simulation) เพื่อพิจารณาว่าข้อสอบครอบคลุมช่วงของคุณลักษณะแฝงที่ต้องการวัดหรือไม่ และ 6) ดำเนินการปรับข้อสอบที่เหมาะสมทั้งหมดให้อยู่ในรูปแบบของคอมพิวเตอร์ตามแนวทางการจัดการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์

การจัดสมดุลเนื้อหาของข้อสอบ (Content balancing) เป็นสิ่งสำคัญอย่างยิ่ง โดยเฉพาะการวัดผลสัมฤทธิ์ทางการศึกษา ในการทดสอบแบบดั้งเดิม แบบทดสอบมาตรฐานส่วนใหญ่ได้สร้างตามรายละเอียดในตารางการกำหนดคุณลักษณะของข้อสอบ (Table of content specification) ซึ่งผู้สอบทุกคนจะได้รับแบบทดสอบชุดเดียวกัน ประกอบด้วยข้อสอบครบทุกคุณลักษณะที่ต้องการวัดตามที่ออกแบบไว้ และสัมพันธ์กับการแสดงข้อสอบ (Item exposure) คือการจัดข้อสอบให้ผู้สอบเป็นส่วนหนึ่งของการทดสอบแบบปรับเหมาะในขั้นตอนการคัดเลือกข้อสอบข้อถัดไปที่มีความเหมาะสมกับระดับความสามารถของผู้สอบ ซึ่งเป็นข้อสอบที่ให้ค่าสารสนเทศสูงสุด ณ ระดับความสามารถของผู้สอบขณะนั้น ค่าสารสนเทศของข้อสอบจะมีค่าเพิ่มขึ้น ใน 2 กรณี คือ 1) ค่าอำนาจจำแนกของข้อสอบมีค่าเพิ่มขึ้นหรือและ 2) ค่าความยากของข้อสอบมีค่าใกล้เคียงกับความสามารถของผู้สอบ

จากเหตุผลข้างต้น ยังไม่มีงานวิจัยที่พัฒนาวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา (Interval a-parameter-Stratification with Content Balancing Item Pools Strategies MCAT Method: IACB) ผู้วิจัยประยุกต์จากวิธี a-Stratified Method: a-STR (Chang & Ying, 1996) ซึ่งได้ขยายวิธีการโดยแบ่งคลังข้อสอบออกเป็น 4 ชั้น ตามช่วงของค่าอำนาจจำแนกข้อสอบ ชั้นที่ 1 (a อยู่ระหว่าง 0.50 ถึง 0.99) ชั้นที่ 2 (a อยู่ระหว่าง 1.00 ถึง 1.49) ชั้นที่ 3 (a อยู่ระหว่าง 1.50 ถึง1.99) และ ชั้นที่ 4 (a อยู่ระหว่าง 2.00 ถึง 2.50) ร่วมกับการจัดสมดุลเนื้อหาภายในชั้นของค่าอำนาจจำแนกของข้อสอบ ได้แบ่งเป็น 4 กลุ่ม ตามสาระการเรียนรู้ที่ออกข้อสอบ O-NET วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 6 ได้แก่ 1) จำนวนและการดำเนินการ 2) การวัด 3) พีชคณิต และ 4) การวิเคราะห์และความน่าจะเป็น และใช้วิธีการควบคุมการแสดงด้วยวิธีการสุ่มตัวอย่างแบบชั้นภูมิ (Stratified random sampling) โดยเปรียบเทียบกับวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบถ่วงน้ำหนักที่มีการบังคับ (Constraint-weighted a-stratification MCAT method: CWA) (Cheng, Chang, Douglas, & Guo, 2009) โดยเปรียบเทียบประสิทธิภาพ 1) ด้านการประมาณค่าความสามารถของผู้สอบ ได้แก่ ค่าความลำเอียงเฉลี่ย และ ค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย ตามแนวทางของ Yan, Von Davier, and Lewis (2016) 2) ด้านประสิทธิภาพของการใช้ข้อสอบ ได้แก่ ข้อสอบที่มีการแสดงมากเกินไป ข้อสอบที่มีการใช้น้อยเกินไป อัตราการทับซ้อนของข้อสอบ และการแจกแจงอัตราการแสดงข้อสอบ ตามแนวทางของ Ozturk and Dogan (2015) ในสถานการณ์จริง

วัตถุประสงค์ของการวิจัย

1. เพื่อพัฒนาวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา

2. เพื่อเปรียบเทียบประสิทธิภาพระหว่างวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหากับวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบถ่วงน้ำหนักที่มีการบังคับ

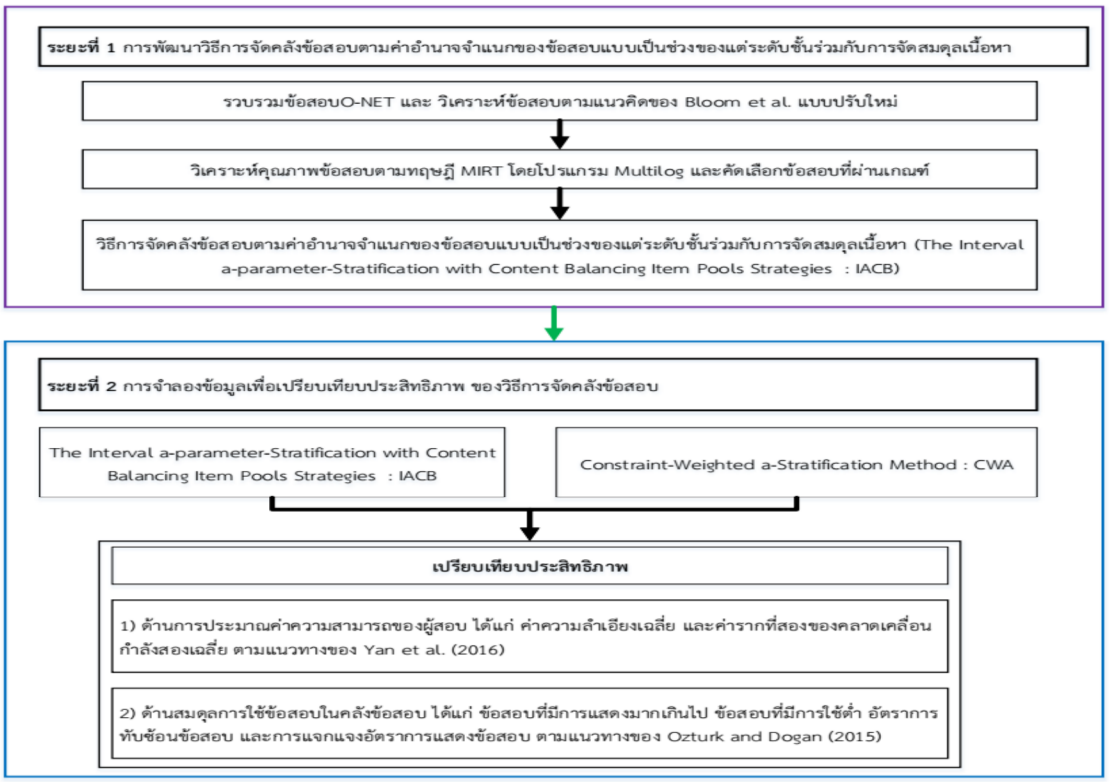
กรอบแนวทางการวิจัย

การวิจัยนี้มุ่งเน้นพัฒนาวิธี IACB โดยใช้ข้อสอบ O-NET วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 6 ระหว่างปีการศึกษา 2551–2553 และปีการศึกษา 2555–2559 ซึ่งเป็นข้อมูลทุติยภูมิ (Secondary data) ที่ได้มาจากสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) โดยนำมาวิเคราะห์ข้อสอบตามแนวคิดของ Bloom, Engelhart, Furst, Hill, and Krathwohl (1956), Anderson et al. (2001) และ Krathwohl (2002) โดยการปรับลำดับขั้น

และคำศัพท์ที่ใช้ในกระบวนการพุทธิปัญญา (Cognitive Process Dimension) มี 6 กระบวนการเหมือนเดิม แต่ 3 กระบวนการแรกเปลี่ยนชื่อเป็น จำ (Remember) เข้าใจ (Understand) และประยุกต์ (Apply) ส่วน 3 กระบวนการหลังเปลี่ยนชื่อที่มีลักษณะเป็นคำนามไปเป็นคำกริยา และสลับที่กับระหว่างกระบวนการที่ 5 กับ 6 และสร้างสรรค์ (Create) เปลี่ยนชื่อมาจาก การสังเคราะห์ (Synthesis) (Anderson et al., 2001; Krathwohl, 2002) และเปลี่ยนโครงสร้างจากมิติเดียวเป็นสองมิติ โดย Anderson et al. (2001) และ Krathwohl (2002) ได้เพิ่มโครงสร้างในมิติด้านความรู้ (Knowledge dimension) เข้ามาในโครงสร้างของจุดประสงค์ทางการศึกษา ด้านพุทธิปัญญา ทำให้โครงสร้างใหม่มีลักษณะเป็น 2 มิติ ประกอบด้วย 1) ด้านกระบวนการพุทธิปัญญา และ 2) ด้านความรู้ และอธิบายความหมายของมิติด้านความรู้ที่เพิ่มเติมเข้ามาได้ (Krathwohl, 2002) โดยมิติด้านความรู้ แบ่งเป็น 4 ส่วน คือ ความรู้เกี่ยวกับข้อเท็จจริง (Factual knowledge) ความรู้เกี่ยวกับมโนทัศน์ (Conceptual knowledge) ความรู้เกี่ยวกับวิธีดำเนินการ (Procedural knowledge) และความรู้เกี่ยวกับอภิปัญญา (Metacognitive knowledge) และมิติด้านกระบวนการพุทธิปัญญา แบ่งเป็น 6 กระบวนการ คือ จำ เข้าใจ ประยุกต์ วิเคราะห์ ประเมินค่า และสร้างสรรค์

การพัฒนาวิธี IACB ผู้วิจัยประยุกต์จากวิธีจัดคลังข้อสอบตามระดับชั้นของค่าอำนาจจำแนกของข้อสอบ (a-Stratified Method: a-STR) ที่พัฒนาโดย Chang and Ying (1996) ซึ่งแบ่งคลังข้อสอบออกเป็น 4 ชั้น ตามช่วงของค่าอำนาจจำแนกข้อสอบ (a) ดังนี้ ชั้นที่ 1 (a อยู่ระหว่าง 0.50 ถึง 0.99) ชั้นที่ 2 (a อยู่ระหว่าง 1.00 ถึง 1.49) ชั้นที่ 3 (a อยู่ระหว่าง 1.50 ถึง1.99) และชั้นที่ 4 (a อยู่ระหว่าง 2.00 ถึง 2.50) ร่วมกับการจัดสมดุลเนื้อหา (Content balancing) ภายในชั้นของค่าอำนาจจำแนก ได้แบ่งเป็น 4 กลุ่ม ตามสาระการเรียนรู้ที่ออกข้อสอบ O-NET วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 6 ได้แก่ 1) จำนวนและการดำเนินการ 2) การวัด 3) พีชคณิต และ 4) การวิเคราะห์และความน่าจะเป็น ดังนั้นวิธีการควบคุมการแสดงข้อสอบ (Controlling item exposure) จึงใช้วิธีการสุ่มตัวอย่างแบบชั้นภูมิ (Stratified random sampling) โดยให้กลุ่มสาระการเรียนรู้เนื้อหาการออกข้อสอบเป็นชั้นภูมิ และให้ค่าความยากของข้อสอบ (b) ที่มีลักษณะเหมือนกัน (Homogenious) เป็นหน่วยการสุ่ม

การตรวจสอบประสิทธิภาพของวิธี IACB ที่ได้พัฒนาขึ้นโดยเปรียบเทียบกับวิธี CWA ซึ่งพิจารณาประสิทธิภาพ ดังนี้ 1) ด้านการประมาณค่าความสามารถของผู้สอบ ได้แก่ ค่าความลำเอียงเฉลี่ย และค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย ตามแนวทางของ Yan et al. (2016) 2) ด้านประสิทธิภาพของการใช้ข้อสอบ ได้แก่ ข้อสอบที่มีการแสดงมากเกินไป ข้อสอบที่มีการใช้น้อยเกินไป อัตราการทับซ้อนของข้อสอบ และการแจกแจงอัตราการแสดงข้อสอบ ตามแนวทางของ Ozturk and Dogan (2015) ซึ่งเป็นโปรแกรมจำลองข้อมูล ตามแนวคิดของ Reckase (2009) และ Thompson and Weiss (2011) ซึ่งได้กำหนดขั้นตอน ดังภาพที่ 1



ภาพที่ 1 กรอบแนวทางการวิจัย

สมมติฐานของการวิจัย

1. การประมาณค่าความสามารถของผู้สอบ ได้แก่ ค่าความลำเอียงเฉลี่ย และค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย ของวิธี IACB มีค่าน้อยกว่าวิธี CWA

2. ประสิทธิภาพของการใช้ข้อสอบในคลังข้อสอบ ได้แก่ ข้อสอบที่มีการแสดงมากเกินไป ข้อสอบที่มีการใช้น้อยเกินไป อัตราการทับซ้อนของข้อสอบ และการแจกแจงอัตราการแสดงข้อสอบ ของวิธี IACB มีค่าน้อยกว่าวิธี CWA

วิธีดำเนินการวิจัย

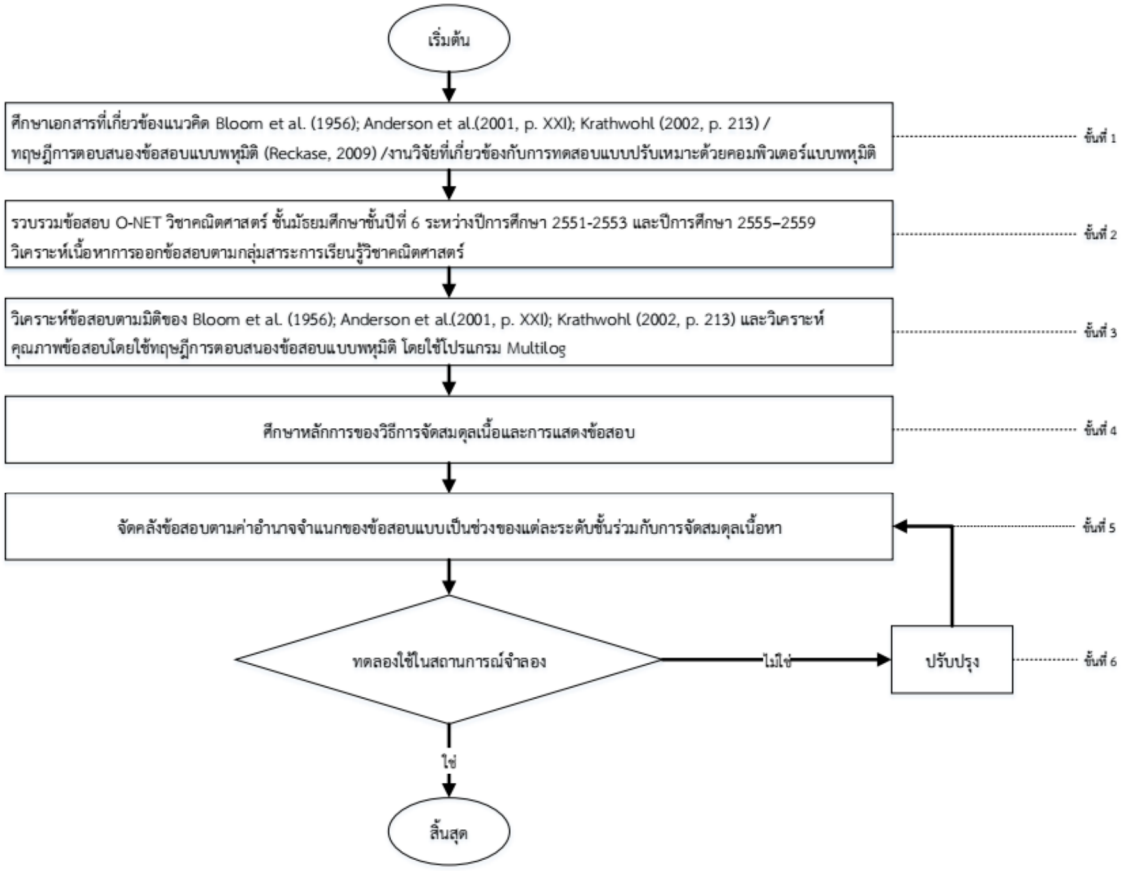
กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 6 ที่ทำการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน (O-NET) วิชาคณิตศาสตร์ ระหว่างปีการศึกษา 2551–2553 และปีการศึกษา 2555–2559 ซึ่งเป็นข้อมูลทุติยภูมิ (Secondary data) ที่ได้มาจากสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน)

การดำเนินการวิจัย แบ่งเป็น 2 ระยะ ดังนี้

ระยะที่ 1 การพัฒนาวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา

ระยะที่ 2 การเปรียบเทียบประสิทธิภาพระหว่างวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหากับวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบถ่วงน้ำหนักที่มีการบังคับ

ระยะที่ 1 การพัฒนาวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละ ระดับชั้นร่วมกับการจัดสมดุลเนื้อหา มีรายละเอียด ดังภาพที่ 2



ภาพที่ 2 วิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา

จากภาพที่ 2 วิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วม กับการจัดสมดุลเนื้อหา แสดงรายละเอียดได้ ดังนี้

ขั้นที่ 1 การศึกษาเอกสารที่เกี่ยวข้องกับแนวคิดของ Bloom et al. (1956), Anderson et al. (2001) และ Krathwohl (2002) ทฤษฎีการตอบสนองข้อสอบแบบพหุมิติและงานวิจัยที่เกี่ยวข้อง

ขั้นที่ 2 การรวบรวมและการวิเคราะห์คุณภาพข้อสอบ O-NET

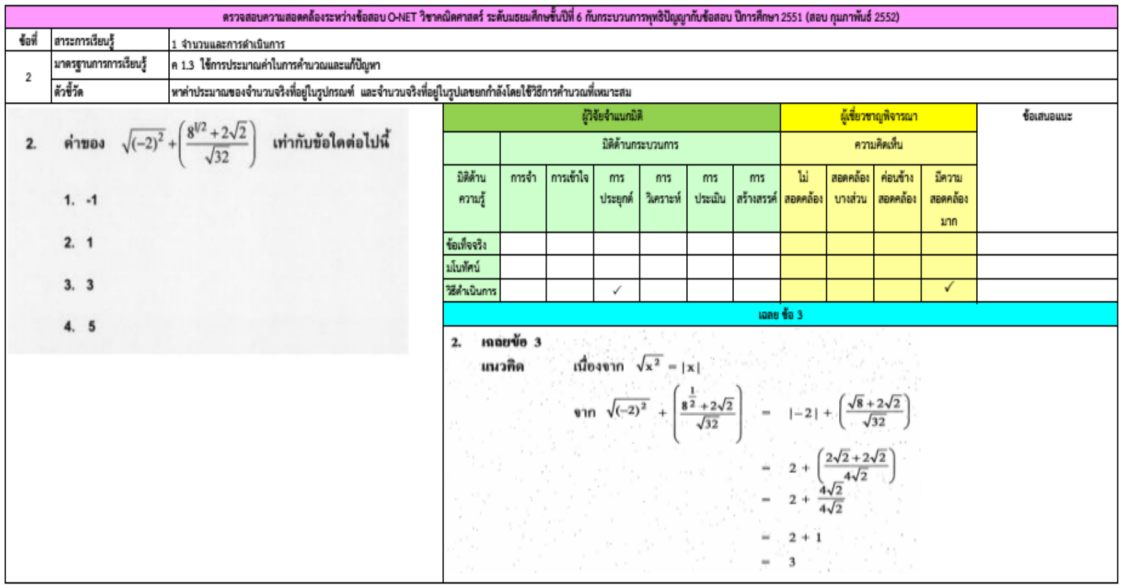
2.1 การรวบรวมข้อสอบ O-NET วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 6 ระหว่างปีการศึกษา 2551-2553 และปีการศึกษา 2555-2559 ที่ได้จากสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) ทั้งหมด 256 ข้อ

2.2 การวิเคราะห์ข้อสอบ O-NET วิชาคณิตศาสตร์ตามสาระการเรียนรู้ที่ออกข้อสอบ ได้แก่ 1) จำนวนและการดำเนินการ 2) การวัด 3) พีชคณิต และ 4) การวิเคราะห์และความน่าจะเป็น

2.3 ผู้เชี่ยวชาญทางด้านคณิตศาสตร์ จำนวน 3 คน พิจารณาความตรงเชิงเนื้อหาด้วยดัชนีความตรงเชิงเนื้อหา (Content Validity Index: CVI) โดยมีค่าดัชนีความตรงเชิงเนื้อหา ตั้งแต่ 0.66 ถึง 1.00 จำนวน 256 ข้อ

ขั้นที่ 3 การวิเคราะห์จุดประสงค์ทางการศึกษาใหม่ตามแนวคิดของ Bloom et al. (1956), Anderson et al. (2001) และ Krathwohl (2002) และการวิเคราะห์คุณภาพข้อสอบ

3.1 ผู้วิจัยนำข้อสอบ O-NET วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 6 ระหว่างปีการศึกษา 2551-2553 และ ปีการศึกษา 2555-2559 จำนวน 256 ข้อ มาวิเคราะห์ตามจุดประสงค์ทางการศึกษาใหม่ตามแนวคิดของ Bloom et al. (1956), Anderson et al. (2001) และ Krathwohl (2002) ซึ่งได้เพิ่มโครงสร้างจากมิติเดียวเป็นสองมิติ ประกอบด้วย มิติด้านกระบวนการพุทธิปัญญาและมิติด้านความรู้ ดังภาพที่ 3



ตรวจสอบความสอดคล้องระหว่างข้อสอบ O-NET วิชาคณิตศาสตร์ ระดับมัธยมศึกษาชั้นปีที่ 6 กับกระบวนการพุทธิปัญญากับข้อสอบ ปีการศึกษา 2551 (สอบ กุมภาพันธ์ 2552)

ข้อที่	สาระการเรียนรู้	1 จำนวนและการดำเนินการ
2	มาตรฐานการการเรียนรู้	ค 1.3 ใช้การประมาณค่าในการคำนวณและแก้ปัญหา
	ตัวชี้วัด	หาค่าประมาณของจำนวนจริงที่อยู่ในรูปกรณฑ์ และจำนวนจริงที่อยู่ในรูปเลขยกกำลังโดยใช้วิธีการคำนวณที่เหมาะสม

2. ค่าของ $\sqrt{(-2)^2}+\left(\frac{8^{1/2}+2\sqrt{2}}{\sqrt{32}}\right)$ เท่ากับข้อใดต่อไปนี้

1. -1
2. 1
3. 3
4. 5

	ผู้วิจัยจำแนกมิติ						ผู้เชี่ยวชาญพิจารณา				ข้อเสนอแนะ
	มิติด้านกระบวนการ						ความคิดเห็น				
มิติด้านความรู้	การจำ	การเข้าใจ	การประยุกต์	การวิเคราะห์	การประเมิน	การสร้างสรรค์	ไม่สอดคล้อง	สอดคล้องบางส่วน	ค่อนข้างสอดคล้อง	มีความสอดคล้องมาก	
ข้อเท็จจริง											
มโนทัศน์											
วิธีดำเนินการ			✓							✓	

เฉลย ข้อ 3

2. เฉลยข้อ 3

แนวคิด เนื่องจาก $\sqrt{x^2}=|x|$

จาก $\sqrt{(-2)^2}+\left(\frac{8^{\frac{1}{2}}+2\sqrt{2}}{\sqrt{32}}\right) = |-2|+\left(\frac{\sqrt{8}+2\sqrt{2}}{\sqrt{32}}\right)$

$= 2+\left(\frac{2\sqrt{2}+2\sqrt{2}}{4\sqrt{2}}\right)$

$= 2+\frac{4\sqrt{2}}{4\sqrt{2}}$

$= 2+1$

$= 3$

ภาพที่ 3 ผลการพิจารณาความตรงเชิงเนื้อหาของผู้เชี่ยวชาญ

3.2 ผู้เชี่ยวชาญทางด้านวัดผลการศึกษา จำนวน 3 คน พิจารณาความตรงเชิงเนื้อหาด้วยดัชนีความตรงเชิงเนื้อหา (Content Validity Index: CVI) โดยมีค่าดัชนีความสอดคล้องตั้งแต่ 0.66 ถึง 1.00 จำนวน 256 ข้อ

ตารางที่ 1 สรุปผลการรวบรวมข้อสอบ จำแนกตามจุดประสงค์ทางการศึกษาด้านพุทธิปัญญาและสาระการเรียนรู้

จุดประสงค์ทางการศึกษาด้านพุทธิปัญญาที่มุ่งวัด	สาระการเรียนรู้ที่ออกข้อสอบ				รวม
	จำนวนและการดำเนินการ (ข้อ)	การวัด (ข้อ)	พีชคณิต (ข้อ)	การวิเคราะห์ข้อมูลและความน่าจะเป็น (ข้อ)	
1. วิเคราะห์เกี่ยวกับข้อเท็จจริง	7	-	4	-	11
2. เข้าใจเกี่ยวกับมโนทัศน์	1	1	2	13	17
3.ประยุกต์เกี่ยวกับมโนทัศน์	1	1	4	4	10
4. วิเคราะห์เกี่ยวกับมโนทัศน์	14	1	41	9	65
5. ประยุกต์เกี่ยวกับวิธีดำเนินการ	22	28	59	39	148
6. วิเคราะห์เกี่ยวกับวิธีดำเนินการ	-	-	5	-	5
รวม	45	31	115	65	256

3.3 การวิเคราะห์คุณภาพของข้อสอบโดยใช้ทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ มีขั้นตอน ดังนี้

1) ตรวจสอบความสมบูรณ์ของข้อสอบ ตัวเลือก และเฉลย ถ้าไม่สมบูรณ์ คัดข้อสอบข้อนั้นออก

2) จัดเตรียมข้อมูลจากผลการตอบข้อสอบการวิเคราะห์คุณภาพของข้อสอบตามหลักการ MIRT

3.4 การคัดเลือกข้อสอบ ใช้เกณฑ์ Urry (1977) ดังนี้ ค่าอำนาจจำแนกของข้อสอบ (a อยู่ระหว่าง 0.50 ถึง 2.50) ค่าความยากของข้อสอบ (b อยู่ระหว่าง -2.50 ถึง 2.50) และค่าโอกาสการเดาของข้อสอบ (c ไม่เกิน .30)

3.5 การวิเคราะห์คุณภาพของข้อสอบ โดยใช้โปรแกรม Multilog มีข้อสอบผ่านเกณฑ์ทั้งสิ้น 193 จากข้อสอบทั้งหมด 256 ข้อ คิดเป็นร้อยละ 75.39 ดังตารางที่ 2

ตารางที่ 2 สรุปผลผลการวิเคราะห์คุณภาพข้อสอบ O-NET วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 6

จุดประสงค์ทางการศึกษาด้านพุทธิปัญญาที่มุ่งวัด	สาระการเรียนรู้ที่ออกข้อสอบ				รวม
	จำนวนและการดำเนินการ (ข้อ)	การวัด (ข้อ)	พีชคณิต (ข้อ)	การวิเคราะห์ข้อมูลและความน่าจะเป็น (ข้อ)	
1. วิเคราะห์เกี่ยวกับข้อเท็จจริง	5	-	3	-	8
2. เข้าใจเกี่ยวกับมโนทัศน์	1	1	1	9	12
3.ประยุกต์เกี่ยวกับมโนทัศน์	1	0	4	4	9
4. วิเคราะห์เกี่ยวกับมโนทัศน์	9	1	32	7	49
5. ประยุกต์เกี่ยวกับวิธีดำเนินการ	19	13	43	35	110
6. วิเคราะห์เกี่ยวกับวิธีดำเนินการ	-	-	5	-	5
รวม	35	15	88	55	193

ขั้นที่ 4 การศึกษาหลักการของวิธีการจัดสมดุลเนื้อหาของข้อสอบ และวิธีการควบคุมการแสดงข้อสอบ

ขั้นที่ 5 การพัฒนาวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ระดับชั้นร่วมกับการจัดสมดุลเนื้อหา

5.1 การพัฒนาคลังข้อสอบเริ่มจากแบ่งชั้นคลังข้อสอบเป็น 4 ชั้นตามระดับค่าอำนาจจำแนกข้อสอบ ดังนี้ ชั้นที่ 1 (a อยู่ระหว่าง 0.50 ถึง 0.99) ชั้นที่ 2 (a อยู่ระหว่าง 1.00 ถึง 1.49) ชั้นที่ 3 (a อยู่ระหว่าง 1.50 ถึง 1.99) และ ชั้นที่ 4 (a อยู่ระหว่าง 2.00 ถึง 2.50)

5.2 การจัดสมดุลเนื้อหาในแต่ละชั้นจัดกลุ่มเนื้อหาข้อสอบ O-NET วิชาคณิตศาสตร์ โดยจำแนกเป็น 4 กลุ่ม ตามสาระการเรียนรู้ที่ออกข้อสอบ ได้แก่ 1) จำนวนและการดำเนินการ 2) การวัด 3) พีชคณิต และ 4) การวิเคราะห์และความน่าจะเป็น

5.3 การควบคุมการแสดงข้อสอบโดยใช้วิธีการสุ่มตัวอย่างแบบชั้นภูมิ (Stratified random sampling) โดยให้กลุ่มสาระการเรียนรู้เป็นชั้นภูมิ และให้ค่าความยากของข้อสอบ (b) ที่มีลักษณะเหมือนกัน (Homogenious) เป็นหน่วยการสุ่ม ดังตารางที่ 3

ตารางที่ 3 สรุปผลการจัดคลังข้อสอบตามระดับชั้น จำแนกตามสาระการเรียนรู้การออกข้อสอบ

คลังข้อสอบชั้นที่	สาระการเรียนรู้ที่ออกข้อสอบ				รวม
	จำนวนและการดำเนินการ (ข้อ)	การวัด (ข้อ)	พีชคณิต (ข้อ)	การวิเคราะห์ข้อมูลและความน่าจะเป็น (ข้อ)	
ชั้นที่ 1 (a อยู่ระหว่าง 0.50 ถึง 0.99)	3	1	9	5	18
ชั้นที่ 2 (a อยู่ระหว่าง 1.00 ถึง 1.49)	5	2	22	20	49
ชั้นที่ 3 (a อยู่ระหว่าง 1.50 ถึง 2.00)	16	4	24	20	64
ชั้นที่ 4 (a อยู่ระหว่าง 2.01 ถึง 2.50)	11	8	33	10	62
รวม	35	15	88	55	193

ขั้นที่ 6 การปรับปรุงและแก้ไขวิธีการจัดคลังข้อสอบแบบ IACB ที่พัฒนาขึ้นไปทดสอบความถูกต้อง โดยการนำไปทดลองใช้ในสถานการณ์จำลอง

ระยะที่ 2 การเปรียบเทียบประสิทธิภาพระหว่างวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหากับวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบถ่วงน้ำหนักที่มีการบังคับ

2.1 การเปรียบเทียบประสิทธิภาพระหว่างวิธี IACB กับวิธี CWA ในครั้งนี้ ใช้การศึกษาในสถานการณ์จำลอง เพื่อให้แน่ใจว่าความแตกต่างของประสิทธิภาพที่เกิดขึ้น (ตัวแปรตาม) เป็นผลมาจากความแตกต่างของวิธีการจัดคลังข้อสอบ (ตัวแปรต้น) ไม่ได้เกิดจากตัวแทรกซ้อนอื่น ๆ เช่น อารมณ์ของผู้สอบ หรือสภาพแวดล้อม ซึ่งจำลองสถานการณ์ที่ใช้วิธีการจัดคลังข้อสอบที่แตกต่างกัน 2 วิธี ดังนี้ 1) วิธี IACB และ 2) วิธี CWA

2.2 การจำลองข้อมูลที่ใช้ศึกษาในสถานการณ์จำลองด้วยการเขียนโปรแกรม JAVA

1) ค่าความสามารถที่แท้จริงของผู้สอบ (True Ability: θ) ถูกจำลองขึ้นด้วยโปรแกรม WinGen3 โดยใช้ค่าพารามิเตอร์ของข้อสอบที่ได้จากสถานการณ์จำลองในแต่ละชั้นของทั้ง 2 วิธี โดย 1 รอบ จะทำ 4 ครั้ง และประมาณค่าความสามารถที่แท้จริงของผู้สอบ จำนวน 500 สถานการณ์ โดยกำหนดให้โปรแกรมสุ่มข้อมูลจากเลขสุ่มที่มีการแจกแจงแบบปกติ N (0,1) และนำรายการตอบที่ได้จากโปรแกรม WinGen3 มาเปรียบเทียบกับสถานการณ์จำลองว่าในรายการตอบเดียวกันค่าประมาณความสามารถที่แท้จริงมีค่าที่เท่ากันหรือใกล้เคียงกันมากที่สุด จำนวน 10 ครั้ง

2) คลังข้อสอบ (Item pools) ขนาด 193 ข้อ โดยข้อสอบแต่ละข้อจะมีค่าพารามิเตอร์ของข้อสอบ 3 ค่า สุ่มจากการแจกแจงแบบยูนิฟอร์ม ดังนี้ ค่าอำนาจจำแนกของข้อสอบ (a อยู่ระหว่าง 0.50 ถึง 2.50) ค่าความยากของข้อสอบ (b อยู่ระหว่าง 2.50 ถึง 2.50) และค่าโอกาสการเดาของข้อสอบ (c ไม่เกิน .30)

3) การเขียนโปรแกรมเพื่อจำลองสถานการณ์การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติที่มีวิธีการจัดคลังข้อสอบแตกต่างกัน 2 วิธี ทุกวิธีทำซ้ำทั้งหมด 10 รอบ รอบละ 500 ค่า แต่ละรอบใช้คลังข้อสอบขนาด 193 ข้อ ซึ่งแบ่งคลังข้อสอบออกเป็น 4 ชั้น ตามค่าอำนาจจำแนกของข้อสอบ ดังนี้ ชั้นที่ 1 (a อยู่ระหว่าง 0.50 ถึง 0.99) ชั้นที่ 2 (a อยู่ระหว่าง 1.00 ถึง 1.49) ชั้นที่ 3 (a อยู่ระหว่าง 1.50 ถึง1.99) และ ชั้นที่ 4 (a อยู่ระหว่าง 2.00 ถึง 2.50) โดยวิธีการ IACB มีการจัดสมดุลเนื้อหาในแต่ละชั้น 4 กลุ่ม ได้แก่ 1) สาระการเรียนรู้จำนวนและการดำเนินการ 2) สาระการเรียนรู้การวัด 3) สาระการเรียนรู้พีชคณิต และ 4) สาระการเรียนรู้การวิเคราะห์และความน่าจะเป็น แต่วิธีการ CWA ไม่ได้จัดสมดุลเนื้อหาภายในชั้น โดยกำหนดรายละเอียดดังนี้

ขั้นที่ 1 คัดเลือกข้อสอบข้อแรก กำหนดให้

Initial ability = 0 การสุ่มข้อสอบข้อแรกนั้น ข้อสอบที่มีค่าความยากของข้อสอบอยู่ในระดับปานกลางที่มีค่าน้อยที่สุด จากทั้ง 4 กลุ่มสาระการเรียนรู้ในคลังข้อสอบชั้นที่ 1 จากนั้นทำการเลือกข้อสอบที่มีค่าความยากของข้อสอบอยู่ในระดับค่าความยากปานกลาง (b อยู่ระหว่าง -0.49 ถึง 0.50) มาพิจารณาค่า b ที่น้อยที่สุด เพื่อใช้เป็นข้อสอบข้อแรกในการทดสอบ

ขั้นที่ 2 วิธีการคัดเลือกข้อสอบข้อถัดไป

2.1) วิธีการ IACB พิจารณาจากค่าความยากของข้อสอบ (b) ข้อก่อนหน้าและวิธีการควบคุมการแสดงข้อสอบด้วยวิธีการสุ่มตัวอย่างแบบชั้นภูมิ การจัดสมดุลเนื้อหาข้อสอบ

2.2) วิธีการ CWA คำนวณดัชนีลำดับความสำคัญ (Maximum Priority Index: MPI) โดยพิจารณาผลคูณของค่าสารสนเทศของข้อสอบกับดัชนีลำดับความสำคัญ ข้อใดให้ผลคูณมากที่สุด จะเป็นข้อสอบข้อถัดไป

ขั้นที่ 3 การประมาณค่าความสามารถของผู้สอบ

3.1) วิธีการ IACB ใช้วิธีการประมาณค่าแบบ Bayesian Estimation (Segall, 2010; Reckase, 2009)

3.2) วิธีการ CWA ใช้วิธีการประมาณค่าแบบ Expected A Posteriori (EAP)

ขั้นที่ 4 เกณฑ์การยุติการทดสอบ โดยทั่วไปจะกำหนดไว้ที่ 0.3 (Thissen, 1990) งานวิจัยที่ผ่านมากำหนดค่า SE ≤ 0.30 (Seo, 2011; โสพส สุขานนท์สวัสดิ์, เสรี ชัดแช้ม และกฤษณะ ชินสาร, 2556)

4.1) วิธีการ IACB ค่าความคลาดเคลื่อนมาตรฐานน้อยกว่า 0.3 หรือ ทำการทดสอบครบ 16 ข้อ

4.2) วิธีการ CWA ค่าความคลาดเคลื่อนมาตรฐานน้อยกว่า 0.3 หรือทำการทดสอบครบ 15 ข้อ

2.3. เปรียบเทียบประสิทธิภาพระหว่างวิธีการจัดคลังข้อสอบ วิธี IACB กับ วิธี CWA รายละเอียดดังนี้

1) ด้านการประมาณค่าความสามารถของผู้สอบ ได้แก่ ค่าความลำเอียงเฉลี่ย และค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย ตามแนวทางของ Yan et al. (2016)

1.1) การคำนวณค่าความลำเอียงเฉลี่ย (Average bias) โดยนำค่าความสามารถที่แท้จริงของผู้สอบ และค่าประมาณความสามารถของผู้สอบที่ได้รับจากการศึกษาในสถานการณ์จำลอง แทนค่าในสมการ

$$Average\ \ Bias = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{\theta}_i - \theta_i\right)$$

เมื่อ n แทน จำนวนผู้สอบทั้งหมด

$\hat{\theta}_i$ แทน ค่าประมาณความสามารถของผู้สอบคนที่ i

θ_i แทน ค่าความสามารถที่แท้จริงของผู้สอบคนที่ i

จากนั้นพิจารณาค่าความลำเอียงเฉลี่ย ของวิธี IACB กับวิธี CWA เข้าใกล้ 0 มากที่สุด แสดงว่ามีความแม่นยำในการประมาณค่าความสามารถของผู้สอบ

1.2) การคำนวณค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Square Error: RMSE) โดยนำค่าความสามารถที่แท้จริงของผู้สอบ (θ) และค่าประมาณความสามารถของผู้สอบ $(\hat{\theta})$ ที่ได้รับจากการศึกษาในสถานการณ์จำลอง มาแทนค่าในสมการ $RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{\theta}_i - \theta_i\right)^2}$

จากนั้นพิจารณาค่า RMSE ของวิธี IACB กับวิธี CWA เข้าใกล้ 0 มากที่สุด แสดงว่ามีประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบ

2) ด้านประสิทธิภาพของการใช้ข้อสอบ ได้แก่ ข้อสอบที่มีการแสดงมากเกินไป ข้อสอบที่มีการใช้น้อยเกินไป อัตราการทับซ้อนของข้อสอบ และการแจกแจงอัตราการแสดงข้อสอบ ตามแนวทางของ Ozturk and Dogan (2015)

2.1) การคำนวณข้อสอบที่มีการแสดงมากเกินไป (Overexposure item) ซึ่งคำนวณได้ดังสมการ โดยนำจำนวนข้อสอบที่ถูกจัดให้ผู้สอบจากการศึกษาในสถานการณ์จำลอง แทนค่าในสมการ

$$\text{อัตราการแสดงข้อสอบ} = \frac{\text{จำนวนครั้งของข้อสอบที่จัดให้แก่ผู้สอบ}}{\text{จำนวนผู้สอบทั้งหมดที่ทำการทดสอบ}}$$

จากนั้นพิจารณาค่าอัตราการแสดงข้อสอบสูงกว่า 0.2 ของวิธี IACB กับ วิธี CWA ถ้าอัตราการแสดงข้อสอบสูงกว่า 0.2 ของวิธีการใดมีจำนวนน้อยกว่า แสดงว่าวิธีการที่มีอัตราการแสดงข้อสอบสูงกว่า 0.2 มีจำนวนน้อยกว่ามีประสิทธิภาพด้านการใช้ข้อสอบที่มีการแสดงมากเกินไป

2.2) ข้อสอบที่มีการใช้น้อยเกินไป (Underutilized item) คำนวนได้จากสมการในข้อ 2.1 จากนั้นพิจารณาข้อสอบที่มีการใช้น้อยกว่า 0.2 ของวิธี IACB กับ วิธี CWA ถ้าข้อสอบที่มีการใช้น้อยกว่า 0.2 ของวิธีการใดมีจำนวนน้อยกว่า แสดงว่าวิธีการที่มีข้อสอบที่มีการใช้น้อยกว่า 0.2 มีจำนวนน้อยกว่ามีประสิทธิภาพด้านข้อสอบที่มีการใช้น้อยเกินไป

2.3) อัตราการทับซ้อนของข้อสอบ (Item Overlap Rate) คำนวณจากชุดข้อสอบที่มีผู้สอบได้รับชุดข้อสอบร่วมกัน (2 คนขึ้นไป) คำนวณได้จากสมการ

$$\text{อัตราการทับซ้อนของข้อสอบ} = \frac{\text{จำนวนผู้สอบที่ใช้ข้อสอบร่วมกัน}}{\text{จำนวนผู้สอบทั้งหมด}} \times 100$$

จากนั้นพิจารณาค่า RMSE ของวิธี IACB กับ วิธี CWA ว่าอัตราการทับซ้อนของข้อสอบของวิธีการใดมีค่าน้อยที่สุด แสดงว่า วิธีการนั้นมีประสิทธิภาพด้านจำนวนข้อสอบที่มีอัตราการทับซ้อนของข้อสอบสูงกว่า

2.4) การแจกแจงอัตราการแสดงข้อสอบ (Item exposure rate distribution) ซึ่งคำนวณจากชุดข้อสอบที่มีผู้สอบได้รับชุดข้อสอบร่วมกัน (2 คนขึ้นไป)

$$\chi^2 = \sum_{j=1}^{n} \left(\frac{er_j - \frac{L}{N}}{\frac{L}{N}} \right)^2$$

เมื่อ er_j แทน อัตราการแสดงข้อสอบสำหรับข้อสอบที่ j

L แทน ความยาวแบบทดสอบ

N แทน ขนาดคลังข้อสอบ

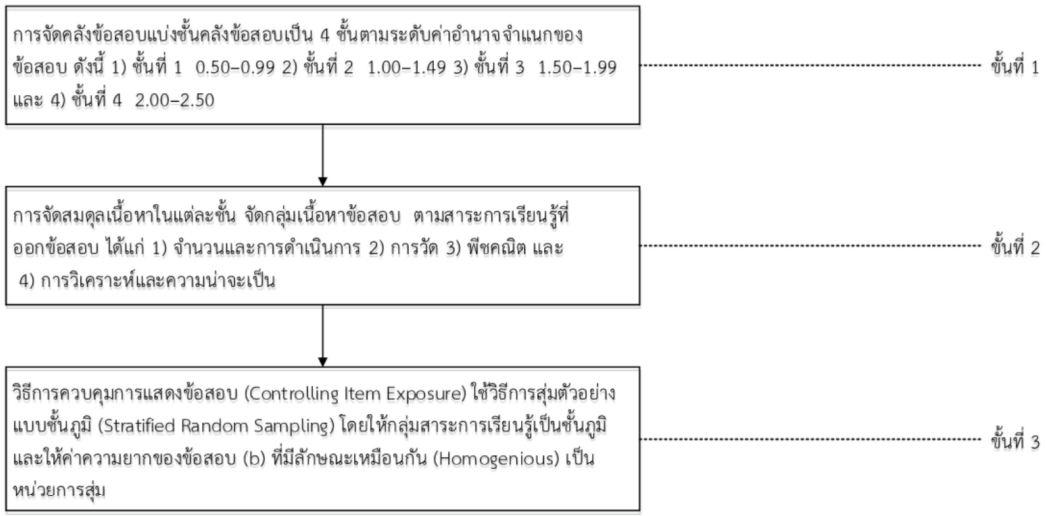
$\frac{L}{N}$ แทน อัตราการแสดงข้อสอบที่ต้องการ

ดังนั้น สมการแสดงถึงความแตกต่างระหว่างค่าสังเกตและค่าคาดหวังของอัตราการแสดงข้อสอบ ถ้าค่า χ^2 เล็กแสดงถึง การใช้ข้อสอบในคลังข้อสอบมีความสมดุลกัน

การเปรียบเทียบการแจกแจงอัตราการแสดงข้อสอบโดยใช้อัตราส่วน F (F ratio) จะพิจารณาจากอัตราส่วนของค่า χ_1^2 ของวิธีที่ 1 หารด้วย χ_2^2 ของวิธีที่ 2 ถ้า $F < 1$ แสดงว่า วิธีที่ 1 ให้สมดุลของอัตราการแสดงข้อสอบโดยรวมดีกว่าวิธีที่ 2

ผลการวิจัย

1. ผลการพัฒนาวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา สามารถสรุปได้ดังภาพที่ 4



ภาพที่ 4 วิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา

จากภาพที่ 4 สามารถอธิบายรายละเอียดได้ดังนี้

การพัฒนาวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ระดับชั้นร่วมกับการจัดสมดุลเนื้อหา โดยข้อสอบที่นำมาบรรจุไว้ในคลังข้อสอบต้องผ่านการวิเคราะห์คุณภาพข้อสอบตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ (MIRT) และผ่านเกณฑ์การคัดเลือกข้อสอบตามที่กำหนดมาแล้ว มีรายละเอียดดังนี้

1) การพัฒนาคลังข้อสอบเริ่มจากแบ่งชั้นของคลังข้อสอบเป็น 4 ชั้น ตามระดับของค่าอำนาจจำแนกของข้อสอบ ดังนี้ 1) ชั้นที่ 1 0.50–0.99 2) ชั้นที่ 2 1.00–1.49 3) ชั้นที่ 3 1.50–1.99 และ 4) ชั้นที่ 4 2.00–2.50

2) การพัฒนาคลังข้อสอบเริ่มจากแบ่งชั้นคลังข้อสอบเป็น 4 ชั้นตามระดับ

ค่าอำนาจจำแนกของข้อสอบร่วมกับการจัดสมดุลเนื้อหาในแต่ละชั้นโดยในแต่ละชั้น ได้จัดกลุ่มเนื้อหาข้อสอบ O-NET วิชาคณิตศาสตร์ โดยจำแนกเป็น 4 กลุ่มสาระการเรียนรู้ ได้แก่ 1) จำนวนและการดำเนินการ 2) การวัด 3) พีชคณิต และ 4) การวิเคราะห์ข้อมูลและความน่าจะเป็น

3) วิธีการควบคุมการแสดงข้อสอบด้วยวิธีการสุ่มตัวอย่างแบบชั้นภูมิ (Stratified random sampling) โดยให้กลุ่มสาระการเรียนรู้เป็นชั้นภูมิ และให้ค่าความยากของข้อสอบ (b) ที่มีลักษณะเหมือนกัน (Homogenious) เป็นหน่วยการสุ่ม ขั้นตอนการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ระดับชั้นร่วมกับการจัดสมดุลเนื้อหาและการควบคุมการแสดงข้อสอบโดยใช้วิธีการสุ่มตัวอย่างแบบชั้นภูมิ

2. ผลการเปรียบเทียบประสิทธิภาพระหว่างวิธีการจัดคลังข้อสอบวิธี IACB กับ วิธี CWA แสดงรายละเอียด ดังนี้

ตารางที่ 4 สรุปผลการจำลองข้อมูลเปรียบเทียบประสิทธิภาพระหว่างวิธีการจัดคลังข้อสอบ

คลังข้อสอบชั้นที่ / ตัวแปร		เปรียบเทียบประสิทธิภาพ				สถิติทดสอบ	
		วิธี		ผลการทดสอบ			
		IACB	CWA	IACB	CWA	Z	χ^2
ด้านการประมาณค่าความสามารถของผู้สอบ							
1	ค่าความลำเอียงเฉลี่ย	0.000	0.003	✓	-	-2.561*	-
	ค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย	0.018	0.056	✓	-	-2.090*	-
2	ค่าความลำเอียงเฉลี่ย	0.000	-0.001	✓	-	-1.328	-
	ค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย	0.024	0.065	✓	-	-2.599**	-
3	ค่าความลำเอียงเฉลี่ย	0.000	-0.002	✓	-	-2.677**	-
	ค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย	0.016	0.046	✓	-	-2.293*	-
4	ค่าความลำเอียงเฉลี่ย	0.000	-0.010	✓	-	-1.546	-
	ค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย	0.016	0.045	✓	-	-2.703**	-
สรุปด้านการประมาณค่าความสามารถของผู้สอบ				8	0		
ด้านประสิทธิภาพของการใช้ข้อสอบ							
1	ข้อสอบที่มีการแสดงมากเกินไป	6.90	10.60	✓	-	-	20.00**
	ข้อสอบที่มีการใช้น้อยเกินไป	11.10	7.40	-	✓	-	20.00**
	อัตราการทับซ้อนของข้อสอบ	97.36	82.28	-	✓	-	20.00
	การแจกแจงอัตราการแสดงข้อสอบ	1532.35	234.29	-	✓	-	20.00
2	ข้อสอบที่มีการแสดงมากเกินไป	3.60	5.30	✓	-	-	13.467**
	ข้อสอบที่มีการใช้น้อยเกินไป	45.40	43.70	-	✓	-	13.467**
	อัตราการทับซ้อนของข้อสอบ	41.64	44.72	✓	-	-	20.00
	การแจกแจงอัตราการแสดงข้อสอบ	180.23	379.23	✓	-	-	20.00
3	ข้อสอบที่มีการแสดงมากเกินไป	1.40	3.40	✓	-	-	13.333**
	ข้อสอบที่มีการใช้น้อยเกินไป	62.60	61.00	-	✓	-	13.467**
	อัตราการทับซ้อนของข้อสอบ	17.84	56.88	✓	-	-	20.00
	การแจกแจงอัตราการแสดงข้อสอบ	102.01	710.74	✓	-	-	20.00
4.	ข้อสอบที่มีการแสดงมากเกินไป	2.90	5.20	✓	-	-	17.000**
	ข้อสอบที่มีการใช้น้อยเกินไป	59.10	56.80	-	✓	-	17.000**
	อัตราการทับซ้อนของข้อสอบ	16.16	58.30	✓	-	-	20.00
	การแจกแจงอัตราการแสดงข้อสอบ	93.60	568.24	✓	-	-	20.00
สรุปด้านประสิทธิภาพของการใช้ข้อสอบ				10	6		

หมายเหตุ: *p<.05; **p<.01

จากตารางที่ 4 ผลการเปรียบเทียบประสิทธิภาพของวิธีการจัดคลังข้อสอบวิธี IACB กับ วิธี CWA ในภาพรวม ปรากฏว่า วิธี IACB มีประสิทธิภาพสูงกว่า วิธี CWA เมื่อพิจารณาจากประเด็น ดังนี้

1. การประมาณค่าความสามารถของผู้สอบ ได้แก่ ค่าความลำเอียงเฉลี่ย และค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย ปรากฏว่า วิธี IACB มีค่าน้อยกว่า วิธี CWA (8:0) แสดงให้เห็นว่า วิธี IACB มีประสิทธิภาพการประมาณค่าความสามารถของผู้สอบสูงกว่า วิธี CWA เป็นไปตามสมมติฐานข้อที่ 1 เมื่อพิจารณารายข้อ ปรากฏว่า

1.1 ค่าความลำเอียงเฉลี่ย ปรากฏว่า วิธี IACB มีค่าน้อยกว่า วิธี CWA (4:0) แสดงให้เห็นว่า วิธี IACB มีความแม่นยำในการประมาณค่าความสามารถของผู้สอบสูงกว่า วิธี CWA ในคลังข้อสอบชั้นที่ 1 อย่างมีนัยสำคัญทางสถิติที่ .05 และในคลังข้อสอบชั้นที่ 3 อย่างมีนัยสำคัญทางสถิติที่ .01 ส่วนในคลังข้อสอบชั้นที่ 2 และ 4 ไม่แตกต่างกัน เป็นไปตามสมมติฐานข้อที่ 1 บางส่วน

1.2 ค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย ปรากฏว่า วิธี IACB มีค่าน้อยกว่า วิธี CWA (4:0) แสดงให้เห็นว่า วิธี IACB มีประสิทธิภาพการประมาณค่าความสามารถของผู้สอบสูงกว่าสูงกว่า วิธี CWA ในคลังข้อสอบชั้นที่ 1 และ 3 อย่างมีนัยสำคัญทางสถิติที่ .05 และในคลังข้อสอบชั้นที่ 2 และ 4 อย่างมีนัยสำคัญทางสถิติที่ .01 เป็นไปตามสมมติฐานข้อที่ 1

2. ประสิทธิภาพของการใช้ข้อสอบ ได้แก่ ข้อสอบที่มีการแสดงมากเกินไป ข้อสอบที่มีการใช้ต่ำ อัตราการทับซ้อนของข้อสอบ และ การแจกแจงอัตราการแสดงข้อสอบ ในภาพรวม ปรากฏว่า วิธี IACB มีค่าน้อย วิธี CWA (10:6) แสดงให้เห็นว่า วิธี IACB มีประสิทธิภาพของการใช้ข้อสอบสูงกว่า วิธี CWA เป็นไปตามสมมติฐานข้อที่ 2 เมื่อพิจารณารายข้อ ปรากฏว่า

2.1 ข้อสอบที่มีการแสดงมากเกินไป ปรากฏว่า วิธี IACB มีค่าน้อยกว่า วิธี CWA (4:0) แสดงให้เห็นว่า วิธี IACB มีประสิทธิภาพด้านการใช้ข้อสอบที่มีการแสดงมากเกินไปสูงกว่า วิธี CWA ในทุกชั้นของคลังข้อสอบอย่างมีนัยสำคัญทางสถิติที่ .01 เป็นไปตามสมมติฐานข้อที่ 2

2.2 ข้อสอบที่มีการใช้น้อยเกินไป ปรากฏว่า วิธี CWA มีค่าน้อยกว่า วิธี IACB (4:0) แสดงให้เห็นว่า วิธี CWA มีประสิทธิภาพด้านการใช้จำนวนข้อสอบที่มีใช้น้อยเกินไปสูงกว่า วิธี IACB ในทุกชั้นของคลังข้อสอบอย่างมีนัยสำคัญทางสถิติที่ .01 ไม่เป็นไปตามสมมติฐานข้อที่ 2

2.3 อัตราการทับซ้อนของข้อสอบ ปรากฏว่า วิธี IACB มีค่าน้อยกว่า วิธี CWA (3:1) แสดงให้เห็นว่า วิธี IACB ประสิทธิภาพด้านจำนวนข้อสอบที่มีอัตราการทับซ้อนข้อสอบสูงกว่า วิธี CWA ในบางชั้นของคลังข้อสอบไม่แตกต่างกัน เป็นไปตามสมมติฐานข้อที่ 2

2.4 การแจกแจงอัตราการแสดงข้อสอบ วิธี IACB มีค่าน้อยกว่า วิธี CWA (3:1) แสดงให้เห็นว่า วิธี IACB มีประสิทธิภาพด้านจำนวนข้อสอบที่มีการแจกแจงอัตราการแสดงข้อสอบสูงกว่า วิธี CWA ในบางชั้นของคลังข้อสอบไม่แตกต่างกัน เป็นไปตามสมมติฐานข้อที่ 2

การอภิปรายผล

1.การพัฒนาวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา

คลังข้อสอบเป็นเสมือนแหล่งรวบรวมชุดแบบทดสอบจำนวนมากและมีความเป็นคู่ขนานกัน ข้อสอบบรรจุในคลังข้อสอบมีลักษณะค่าความยากกระจายเต็มระดับความสามารถของประชากรผู้สอบ ค่าอำนาจจำแนกสูงเพื่อให้ประสิทธิภาพการทดสอบสูงสุด และค่าโอกาสในการเดาเข้าใกล้หรือเท่ากับศูนย์ ข้อสอบควรมีจำนวนมากเพียงพอและกระจายทุกระดับความสามารถของผู้สอบ ดังนั้น ต้องใช้ข้อสอบจำนวนมากสำหรับสร้างคลังข้อสอบ ทำให้คลังข้อสอบสำหรับการทดสอบแบบปรับเหมาะมีขนาดใหญ่กว่าการทดสอบดั้งเดิม (Thompson & Weiss, 2011; Babcock & Weiss, 2013)

การควบคุมการแสดงข้อสอบ ส่วนใหญ่จะคัดเลือกข้อสอบที่มีค่าอำนาจจำแนกข้อสอบสูง มาเป็นข้อสอบข้อถัดไป ดังนั้น ข้อสอบเหล่านี้จึงถูกเลือกใช้ในการทดสอบบ่อยครั้งจนเกินไป เป็นผลให้ผู้สอบรุ่นถัดไป สามารถคาดการณ์ข้อสอบที่ตนเองจะได้รับล่วงหน้าได้ ในการวิจัยนี้

ได้พัฒนาวิธีการควบคุมการแสดงข้อสอบโดยใช้วิธีการสุ่มตัวอย่างแบบชั้นภูมิ (Stratified random sampling) มาควบคุมการแสดงข้อสอบ โดยใช้ข้อสอบที่นำไปใช้ในขั้นตอนการคัดเลือกข้อสอบข้อถัดไป คือ ข้อสอบที่มีค่าอำนาจจำแนกของข้อสอบต่ำถึงปานกลาง และข้อสอบที่มีค่าอำนาจจำแนกของข้อสอบสูง ซึ่งถูกสุ่มมาจากกลุ่มสาระการเรียนรู้ในแต่ละชั้นที่มีการจัดสมดุลเนื้อหาตามกลุ่มสาระการเรียนรู้ที่ออกข้อสอบ O-NET เป็น 4 กลุ่ม ได้แก่ 1) สาระการเรียนรู้จำนวนและการดำเนินการ 2) สาระการเรียนรู้การวัด 3) สาระการเรียนรู้พีชคณิต และ 4) สาระการเรียนรู้การวิเคราะห์ข้อมูลและความน่าจะเป็น ด้วยวิธีการสุ่มตัวอย่างแบบชั้นภูมิ โดยให้กลุ่มสาระเป็นชั้นภูมิ และให้ค่าความยากของข้อสอบ (b) ที่มีลักษณะเหมือนกัน (Homogenious) เป็นหน่วยการสุ่ม ด้วยกระบวนการนี้ผู้สอบที่มีความสามารถเท่ากันหรือใกล้เคียงกัน จึงมีโอกาสได้รับข้อสอบข้อเดียวกันน้อยลง เนื่องจาก ข้อสอบที่มีค่าอำนาจจำแนกของข้อสอบสูงถูกหมุนเวียน (จากการสุ่ม) นำไปใช้ในขั้นตอนการคัดเลือกข้อสอบข้อถัดไป วิธีการนี้จึงสามารถลด จำนวนข้อสอบที่มีการแสดงมากเกินไป จำนวนข้อสอบที่มีการใช้ต่ำ อัตราการทับซ้อนของข้อสอบ และการแจกแจงอัตราการแสดงข้อสอบ ลงได้ (Georgiadou, Triantafillou, & Economides, 2007) และสอดคล้องกับงานวิจัยของ Ozturk and Dogan (2015)

2. การเปรียบเทียบประสิทธิภาพระหว่างวิธีการจัดคลังข้อสอบ

ผลการเปรียบเทียบด้านการประมาณค่าความสามารถของผู้สอบ ได้แก่ ค่าความลำเอียงเฉลี่ย และค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย ปรากฏว่า วิธี IACB มีค่าน้อย วิธี CWA (8:0) แสดงให้เห็นว่า วิธี IACB มีประสิทธิภาพการประมาณค่าความสามารถของผู้สอบสูงกว่า วิธี CWA เป็นไปตามสมมติฐานข้อที่ 1 วิธีการจัดคลังข้อสอบวิธี IACB มีประสิทธิภาพด้านการประมาณค่าความสามารถของผู้สอบสูงกว่าวิธี CWA ที่พิจารณาเพียงค่าสารสนเทศที่ใกล้เคียงกับความสามารถของผู้สอบ ณ ขณะนั้น (Taylor, 2009; Cheng et al., 2009; Finch, 2010; Zhou, 2012) ซึ่งให้ความคลาดเคลื่อนกำลังสองเฉลี่ยและความลำเอียงเฉลี่ย มีแนวโน้มลดลง เหมือนกันเมื่อความยาวแบบทดสอบเพิ่มขึ้น แต่ความคลาดเคลื่อนกำลังสองเฉลี่ยในวิธี IACB ต่ำกว่า วิธี CWA

ผลการเปรียบเทียบประสิทธิภาพของการใช้ข้อสอบ ได้แก่ ข้อสอบที่มีการแสดงมากเกินไป ข้อสอบที่มีการใช้น้อยเกินไป อัตราการทับซ้อนของข้อสอบ และการแจกแจงอัตราการแสดงข้อสอบ ในภาพรวม วิธี IACB มีค่าน้อย วิธี CWA (10:6) แสดงให้เห็นว่า วิธี IACB มีประสิทธิภาพของการใช้ข้อสอบสูงกว่า วิธี CWA เป็นไปตามสมมติฐานข้อที่ 2 ซึ่งสามารถควบคุมประสิทธิภาพของการใช้ข้อสอบ ได้แก่ ข้อสอบที่มีการแสดงมากเกินไป ข้อสอบที่มีการใช้น้อยเกินไป อัตราการทับซ้อนของข้อสอบ และการแจกแจงอัตราการแสดงข้อสอบ ให้มีจำนวนลดลงได้ โดยส่วนใหญ่เป็นข้อสอบที่มีค่าอำนาจจำแนกของข้อสอบระดับปานกลางถึงสูง ซึ่งวิธีการนี้ ทำให้ข้อสอบที่มีความค่าอำนาจจำแนกของข้อสอบปานกลางถึงสูง มีโอกาสถูกเลือกใช้ในการทดสอบสูงขึ้น โดยการสุ่มข้อสอบที่มีค่าอำนาจจำแนกของข้อสอบปานกลางถึงสูง ด้วยวิธีการสุ่มตัวอย่างแบบชั้นภูมิ (Stratified random sampling) มาควบคุมการแสดงข้อสอบ โดยใช้ข้อสอบที่นำไปใช้ในขั้นตอนการคัดเลือกข้อสอบข้อถัดไป คือ ข้อสอบที่มีค่าอำนาจจำแนกของข้อสอบต่ำถึงปานกลาง และข้อสอบที่มีค่าอำนาจจำแนกของข้อสอบสูง ซึ่งถูกสุ่มมาจากกลุ่มสาระการเรียนรู้ในแต่ละชั้นที่มีการจัดสมดุลเนื้อหาตามกลุ่มสาระการเรียนรู้ที่ออกข้อสอบ O-NET เป็น 4 กลุ่ม ได้แก่ 1) จำนวนและการดำเนินการ 2) การวัด 3) พีชคณิต และ 4) การวิเคราะห์ข้อมูลและความน่าจะเป็น ด้วยวิธีการสุ่มตัวอย่างแบบชั้นภูมิ โดยให้กลุ่มสาระเป็นชั้นภูมิ และให้ค่าความยากของข้อสอบ (b) ที่มีลักษณะเหมือนกัน (Homogenious) เป็นหน่วยการสุ่ม ด้วยกระบวนการนี้ ผู้สอบที่มีความสามารถเท่ากันหรือใกล้เคียงกัน จึงมีโอกาสได้รับข้อสอบข้อเดียวกันน้อยลง เนื่องจาก ข้อสอบที่มีค่าอำนาจจำแนกของข้อสอบสูงถูกหมุนเวียน (จากการสุ่ม) นำไปใช้ในขั้นตอนการคัดเลือกข้อสอบข้อถัดไป วิธีการนี้จึงสามารถลดจำนวนข้อสอบที่มีการแสดงมากเกินไป จำนวนข้อสอบที่มีการใช้น้อยเกินไป อัตราการทับซ้อน

ของข้อสอบ และการแจกแจงอัตราการแสดงข้อสอบลงได้ ซึ่งสอดคล้องกับ งานวิจัยของ Ozturk and Dogan (2015) ที่ได้เปรียบเทียบประสิทธิภาพของวิธีการควบคุมการแสดงข้อสอบ ปรากฏว่า วิธีการ Randomesque เป็นวิธีการควบคุมการแสดงข้อสอบด้วยวิธีการสุ่ม สามารถลดจำนวนข้อสอบที่มีการแสดงมากเกินไป จำนวนข้อสอบที่มีการใช้น้อยเกินไป อัตราการทับซ้อนของข้อสอบและการแจกแจงอัตราการแสดงข้อสอบให้น้อยลง

ข้อเสนอแนะการนำผลการวิจัยไปใช้

1. ผู้สนใจพัฒนาวิธีการจัดคลังข้อสอบ สามารถนำวิธีการ IACB ไปประยุกต์ใช้กับวิชาอื่น ๆ ได้ ซึ่งวิธีการนี้มีความเหมาะสมกับข้อสอบที่มีค่าอำนาจจำแนกของข้อสอบอยู่ในระดับปานกลางและระดับสูง

2. ครู อาจารย์ ศึกษานิเทศก์ หรือผู้ที่เกี่ยวข้องสามารถนำข้อสอบที่ผ่านการวิเคราะห์คุณภาพข้อสอบตามทฤษฎีการตอบสนองของสอบแบบพหุมิติ นำมาจัดเก็บในคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบที่เหมาะสม และสามารถนำไปใช้กับโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ หรือนำไปประยุกต์กับการวัดผลการศึกษาในรายวิชาและในระดับอื่น ๆ ได้

ข้อเสนอแนะในการวิจัยต่อไป

1. ควรสร้างคลังข้อสอบที่มีการแบ่งมิติตามทักษะและกระบวนการทางคณิตศาสตร์ตามสาระและมาตรฐานการเรียนรู้หลักสูตรแกนกลางการศึกษาขั้นพื้นฐานของกระทรวงศึกษาธิการ เช่น มิติความสามารถในการแก้ปัญหา มิติความสามารถในการให้เหตุผล มิติความสามารถในการสื่อสาร การสื่อความหมายทางคณิตศาสตร์ และการนำเสนอมิติความสามารถ ในการเชื่อมโยงความรู้ต่าง ๆ ทางคณิตศาสตร์ และเชื่อมโยงคณิตศาสตร์กับศาสตร์อื่น ๆ และมิติการมีความคิดริเริ่มสร้างสรรค์

2. วิธีการควบคุมการแสดงข้อสอบที่พัฒนาขึ้นในการวิจัยนี้ มุ่งเน้นเข้าควบคุมการแสดงข้อสอบที่มีอัตราการใช้ข้อสอบสูง (Over-exposure item) ให้ลดลงได้ แต่ไม่ได้ควบคุมข้อสอบที่ถูกเลือกใช้น้อยหรือไม่ถูกเลือกใช้เลย (Under-utilized item) ให้มีอัตราการใช้ข้อสอบเพิ่มขึ้น จึงควรพัฒนาวิธีการควบคุมการแสดงข้อสอบ ให้ครอบคลุมกรณีข้อสอบที่ถูกเลือกใช้น้อยหรือไม่ถูกเลือกใช้เลย

กิตติกรรมประกาศ

งานวิจัยนี้ได้รับทุนอุดหนุนการวิจัยประเภทบัณฑิตศึกษา จากสำนักงานคณะกรรมการวิจัยแห่งชาติ ประจำปี 2561

เอกสารอ้างอิง

โสพส สุขานนท์สวัสดิ์, เสรี ชัดแช้ม และกฤษณะ ชินสาร (2556). การพัฒนาการคัดเลือกข้อสอบข้อถัดไปโดยใช้ทฤษฎีการตัดสินใจในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์. *วิทยาการวิจัยและวิทยาการปัญญา, 10*(2), 71-85.

Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., & Wittrock, M. C. (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*, Abridged Edition. *White Plains,* New Jersey*:* Longman.

Babcock, B., & Weiss, D. J. (2013). Termination criteria in computerized adaptive tests: Do variable-length CATs provide efficient and effective measurement?. *Journal of Computerized Adaptive Testing*, *1*(1), 1-18.

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of Educational Objetives: the Classification of Educational Goals: Handbook I: Cognitive Domain* (No. 373.19 C734t). New York: D. Mckay.

Chang, H. H., & Ying, Z. (1996). A global information approach to computerized adaptive testing. *Applied Psychological Measurement, 20*(3), 213-229.

Cheng, Y., Chang, H. H., Douglas, J., & Guo, F. (2009). Constraint-weighted a-stratification for computerized adaptive testing with nonstatistical constraints: Balancing measurement efficiency and exposure control. *Educational and Psychological Measurement*, *69*(1), 35-49.

De Ayala, R. J. (2009). *The Theory and Practice of Item Response Theory*. London: The Guilford Press.

Embretson, S. E., & Reise, S. P. (2000). *Item Response Theory for Psychologists*. Mahwah, New Jersey: Lawrence Erlbaum.

Finch, H. (2010). Item parameter estimation for the MIRT model bias and precision of confirmatory factor analysis-based models. *Applied Psychological Measurement, 34*(1), 10-26.

Georgiadou, E. G., Triantafillou, E., & Economides, A. A. (2007). A review of item exposure control strategies for computerized adaptive testing developed from 1983 to 2005. *The Journal of Technology, Learning and Assessment*, *5*(8), 1-39.

Krathwohl, D. R. (2002). A revision of bloom's taxonomy: An overview. *Theory into Practice*, *41*(4), 212-218.

Ozturk, N. B., & Dogan, N. (2015). Investigating item exposure control methods in computerized adaptive testing. *Educational Sciences: Theory and Practice*, *15*(1), 85-98.

Reckase, M. (2009). *Multidimensional Item Response Theory* (Vol. 150). New York: Springer.

Segall, D. O. (2010). Principles of Multidimensional Adaptive Testing. *In Elements of AdaptiveTtesting* (pp. 57-75). New York: Springer.

Seo, D. G. (2011). Application of the Bifactor Model to Computerized Adaptive Testing. *ProQuest*.

Taylor, B. W. (2009). *Introduction to Management Science* (11th ed.). Upper Saddle River, New Jersey: Prentice Hall.

Thissen, David. (1990). Reliability and Measurement Precision. In: Howard Wainer et al.,editors. *Computerized Adaptive Testing: A Primer*. New Jersey: Lawrence Erlbaum.

Thompson, N. A., & Weiss, D. J. (2011). A framework for the development of computerized adaptive tests. *Practical Assessment, Research & Evaluation*, *16*(1), 1-9.

Weiss, D. J. (2011). Better data from better measurements using computerized adaptive testing. *Journal of Methods and Measurement in the Social Sciences*, *2*(1), 1-27.

Urry, V. W. (1977). Tailored testing: A successful application of latent Trait Theory. *Journal of Educational Measurement*, *14*(2), 181-196.

Yan, D., von Davier, A. A., & Lewis, C. (2016). *Computerized Multistage Testing: Theory and Applications.* Princton, New Jersey: CRC Press.

Zhou, X. (2012). Designing P-Optimal Item Pools in Computerized Adaptive Tests with Polytomous Items. *ProQuest LLC.*