การพัฒนาวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้น ร่วมกับการจัดสมดุลเนื้อหา สำหรับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ

ศักดิ์ชัย จันทะแสง

ดุษฎีนิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปรัชญาดุษฎีบัณฑิต
สาขาวิชาการวัดและเทคโนโลยีทางวิทยาการปัญญา
วิทยาลัยวิทยาการวิจัยและวิทยาการปัญญา มหาวิทยาลัยบูรพา
กรกฎาคม 2561
ลิขสิทธิ์เป็นของมหาวิทยาลัยบูรพา

คณะกรรมการควบคุมดุษฎีนิพนธ์และคณะกรรมการสอบดุษฎีนิพนธ์ ได้พิจารณาดุษฎีนิพนธ์ของ ศักดิ์ชัย จันทะแสง ฉบับนี้แล้ว เห็นสมควรรับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปรัชญาดุษฎีบัณฑิต สาขาวิชาการวัดและเทคโนโลยีทางวิทยาการปัญญา ของมหาวิทยาลัยบูรพาได้

คณะกรรมการควบคุมดุษฎีนิพนธ์

.....อาจารย์ที่ปรึกษาหลัก
(รองศาสตราจารย์ ดร. เสรี ชัดแช้ม)

.....อาจารย์ที่ปรึกษาร่วม
(ดร. ปิยะทิพย์ ประดุจพรม)

คณะกรรมการสอบดุษฎีนิพนธ์

.....ประธาน
(รองศาสตราจารย์ ดร. กานดา พูนลาภทวี)

.....กรรมการ
(รองศาสตราจารย์ ดร. เสรี ชัดแช้ม)

.....กรรมการ
(ดร. ปิยะทิพย์ ประดุจพรม)

.....กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร. พูลพงศ์ สุขสว่าง)

.....กรรมการ
(ดร. กนก พานทอง)

วิทยาลัยวิทยาการวิจัยและวิทยาการปัญญา อนุมัติให้รับดุษฎีนิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปรัชญาดุษฎีบัณฑิต สาขาวิชาการวัดและเทคโนโลยีทางวิทยาการปัญญา ของมหาวิทยาลัยบูรพา

..... คณบดีวิทยาลัยวิทยาการวิจัย
(ผู้ช่วยศาสตราจารย์ ดร.สุชาดา กรเพชรปาณี) และวิทยาการปัญญา
วันที่ 15 เดือน กรกฎาคม พ.ศ. 2561

ดุษฎีนิพนธ์นี้ได้รับทุนอุดหนุนการวิจัยระดับบัณฑิตศึกษา
จากสำนักงานคณะกรรมการวิจัยแห่งชาติ (วช.)
ประจำปี 2561

กิตติกรรมประกาศ

การทำดุษฎีนิพนธ์ฉบับนี้สำเร็จลุล่วงเป็นอย่างดีจากความกรุณาของ รองศาสตราจารย์ ดร.เสรี ชัดแช้ม อาจารย์ที่ปรึกษาหลัก และ อาจารย์ ดร.ปิยะทิพย์ ประดุจพรม อาจารย์ที่ปรึกษาร่วม ที่ได้ถ่ายทอดความรู้ ให้คำแนะนำในการแก้ไข ตลอดจนคอยกระตุ้นและติดตามการดำเนินการวิจัย จนการทำดุษฎีนิพนธ์ฉบับนี้สำเร็จได้ จึงขอกราบขอบพระคุณอาจารย์ที่ปรึกษาทั้งสองเป็นอย่างสูง

ขอกราบขอบพระคุณสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) ที่อนุเคราะห์ข้อสอบ O-NET วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 6 ระหว่างปีการศึกษา 2551–2553 และปีการศึกษา 2555–2559 รวมทั้งผลการตอบ

ขอกราบขอบพระคุณ อาจารย์ ดร.มิ่ง เทพครเมือง อาจารย์ ดร.ปิยะ ถิรพันธ์เมธี และ อาจารย์รัชกฤช ธนพัฒนดล ที่เป็นผู้เชี่ยวชาญในการประเมินความเหมาะสมของโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ สำหรับการทดสอบ O-NET วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 6 และให้ข้อเสนอแนะในการปรับปรุง จนทำให้โปรแกรมมีความสมบูรณ์ยิ่งขึ้น

ขอกราบขอบพระคุณผู้อำนวยการ คุณครูผู้ประสานงาน และนักเรียนโรงเรียนสาธิตมหาวิทยาลัยราชภัฏบ้านสมเด็จเจ้าพระยา ที่ให้ความร่วมมือในการเก็บรวบรวมข้อมูลจนสำเร็จลุล่วงด้วยดี

ขอกราบขอบพระคุณครอบครัวที่ให้ความช่วยเหลือเป็นกำลังใจตลอดมา ประโยชน์ของดุษฎีนิพนธ์ฉบับนี้ ผู้วิจัยขอมอบเป็นกตัญญูกตเวทิตาแด่ คุณพ่ออร่าม และคุณแม่บัววัน จันทะแสง และผู้มีพระคุณทุกท่าน รวมทั้งผู้มีพระคุณทุกท่านทั้งในอดีตและปัจจุบันที่ทำให้ข้าพเจ้าเป็นผู้มีการศึกษาและประสบความสำเร็จมาจนตราบเท่าทุกวันนี้

สุดท้ายนี้ ผู้วิจัยกราบขอบพระคุณสำนักงานคณะกรรมการวิจัยแห่งชาติ (วช.) เป็นอย่างสูง มา ณ โอกาสนี้ ที่กรุณาให้ทุนอุดหนุนวิจัยระดับบัณฑิตศึกษา ประจำปี 2561

ศักดิ์ชัย จันทะแสง

56810162: สาขาวิชา: การวัดและเทคโนโลยีทางวิทยาการปัญญา;
ปร.ด. (การวัดและเทคโนโลยีทางวิทยาการปัญญา)
คำสำคัญ: วิธีการจัดคลังข้อสอบ/ การควบคุมการแสดงข้อสอบ/ การจัดสมดุลเนื้อหา/
การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ

ศักดิ์ชัย จันทะแสง: การพัฒนาวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา สำหรับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ (DEVELOPMENT OF ITEM POOL STRATEGIES BY USING INTERVAL A-PARAMETER STRATIFICATION WITH CONTENT BALANCING FOR MULTIDIMENSIONAL COMPUTERIZED ADAPTIVE TESTING) คณะกรรมการควบคุมดุษฎีนิพนธ์: เสรี ชัดแช้ม, ค.ด., ปิยะทิพย์ ประดุจพรม, Ph.D., 285 หน้า. ปี พ.ศ. 2561.

การจัดคลังข้อสอบเป็นการจัดการข้อสอบให้เป็นระบบ หมวดหมู่ พร้อมสำหรับการนำไปใช้ การวิจัยนี้มีวัตถุประสงค์เพื่อ 1) พัฒนาวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา 2) เปรียบเทียบประสิทธิภาพของวิธีการที่พัฒนาขึ้นกับวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบถ่วงน้ำหนักที่มีการบังคับ โดยใช้สถานการณ์จำลองข้อมูลในด้านการประมาณค่าความสามารถของผู้สอบ ได้แก่ ค่าความลำเอียงเฉลี่ย และค่ารากที่สองของความคลาดเคลื่อนกำลังสอง ด้วยการวิเคราะห์ Wilcoxon และด้านประสิทธิภาพของการใช้ข้อสอบ ได้แก่ ข้อสอบที่มีการแสดงมากเกินไป ข้อสอบที่มีการใช้น้อยเกินไป อัตราการทับซ้อนของข้อสอบ และการแจกแจงอัตราการแสดงข้อสอบ ด้วยการวิเคราะห์ Chi-Square 3) พัฒนาโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ ในรูปแบบ Web Application และ 4) เปรียบเทียบความสามารถของผู้สอบแต่ละระดับชั้นของค่าอำนาจจำแนกของข้อสอบ สำหรับการทดสอบ O-NET วิชาคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 จำแนกตามเพศและความสามารถด้านการเรียนวิชาคณิตศาสตร์ กลุ่มตัวอย่างเป็นนักเรียนโรงเรียนสาธิตมหาวิทยาลัยราชภัฏบ้านสมเด็จเจ้าพระยา ชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2560 จำนวน 80 คน วิเคราะห์ข้อมูลด้วย Two-way ANOVA

ผลการวิจัยปรากฏว่า 1) วิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา มี 3 ขั้นตอน ดังนี้ (1) แบ่งคลังข้อสอบเป็น 4 ชั้น ตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วง (2) จัดสมดุลเนื้อหาในแต่ละชั้นของคลังข้อสอบตามสาระการเรียนรู้ที่ออกข้อสอบ และ (3) ควบคุมการแสดงข้อสอบด้วยวิธีการสุ่มตัวอย่างแบบชั้นภูมิ 2) วิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา มีประสิทธิภาพสูงกว่าวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบถ่วงน้ำหนักที่มีการบังคับ 3) การพัฒนาโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติในรูปแบบ Web Application สำหรับการทดสอบ O-NET วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 6 มีความเหมาะสมสำหรับการใช้งานในระดับดี และ 4) ความสามารถด้านการเรียนวิชาคณิตศาสตร์มีผลต่อความสามารถของผู้สอบแต่ละระดับชั้นของค่าอำนาจจำแนกของข้อสอบอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ส่วนเพศไม่มีผลต่อความสามารถของผู้สอบ

56810162: MAJOR: MEASUREMENT AND TECHNOLOGY IN COGNITIVE SCIENCE;
Ph.D. (MEASUREMENT AND TECHNOLOGY IN COGNITIVE SCIENCE)
KEYWORDS: ITEM POOLS STRATEGIES/ CONTROLLING ITEM EXPOSURE/ CONTENT
BALANCING/ MULTIDIMENSIONAL COMPUTERIZED ADAPTIVE TESTING
SAKCHAI JANTASANG: DEVELOPMENT OF ITEM POOL STRATEGIES BY USING INTERVAL A-PARAMETER STRATIFICATION WITH CONTENT BALANCING FOR MULTIDIMENSIONAL COMPUTERIZED ADAPTIVE TESTING. ADVISORY COMMITTEE: SEREE CHADCHAM, Ph.D., PIYATHIP PRADUJPROM, Ph.D. 285 P. 2018.

The item pools strategy is a procedure for organizing and categorizing test items so as 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 a simulation scenario of test-takers' estimative efficiency including average bias and root mean square error (RMSE) by using then Wilcox Test, and item utilizable efficiency including overexposed items, underutilized items, item overlap rate, and item exposure rate distribution, using Chi-square tests; (3) develop a multidimensional computerized adaptive testing program in a web application platform; and (4) compare test taker performance in each competency level of the discriminative power of the Grade 12 O-NET mathematics by gender and by mathematical ability. The sample involved 80 Grade 12 students in academic year 2017 of the Secondary Demonstration School of Bansomdejchaopraya Rajabhat University. Two-way ANOVA was used to analyze the data.

The results showed that (1) there were three steps in the interval a-parameter-stratification with content balancing item pools strategy: first, dividing item pools into four class intervals by 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 taker performance in each competency level of the discriminative power and content balancing method was more efficient than the constraint-weighted a-stratification method; (3) the multidimensional computerized adaptive testing program for web application for Grade 12 O-NET Mathematics developed in the present study had a good level of usability; and (4) the comparison of test taker performance in each competency level of the discriminative power of the Grade 12 O-NET was significantly affected by the mathematical ability at the .01 level, but gender did not affect test taker performance.

สารบัญ

		หน้า
บทคัดย่อภาษาไทย		จ
บทคัดย่อภาษาอังกฤษ		ฉ
สารบัญ		ช
สารบัญตาราง		ฌ
สารบัญภาพ		ฌ
บทที่		
1	บทนำ	1
	ความเป็นมาและความสำคัญของปัญหา	1
	วัตถุประสงค์ของการวิจัย	7
	กรอบแนวทางการวิจัย	8
	สมมติฐานของการวิจัย	10
	ประโยชน์ที่คาดว่าจะได้รับจากการวิจัย	11
	ขอบเขตของการวิจัย	11
	นิยามศัพท์เฉพาะ	13
2	เอกสารและงานวิจัยที่เกี่ยวข้อง	16
	ตอนที่ 1 จุดประสงค์ทางการศึกษาด้านพุทธิปัญญาและการทดสอบทางการศึกษาระดับขั้นพื้นฐาน	16
	ตอนที่ 2 ทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ	24
	ตอนที่ 3 การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ	34
	ตอนที่ 4 งานวิจัยที่เกี่ยวข้องกับการจัดคลังข้อสอบ การจัดสมดุลเนื้อหา การควบคุมการแสดงข้อสอบและการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ	71
3	วิธีดำเนินการวิจัย	83
	ระยะที่ 1 การพัฒนาวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา	84
	ระยะที่ 2 การเปรียบเทียบประสิทธิภาพของวิธีการจัดคลังข้อสอบโดยการจำลองข้อมูล	100
	ระยะที่ 3 การพัฒนาโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ โดยใช้คลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา	111

สารบัญ (ต่อ)

บทที่		หน้า
	ระยะที่ 4 การเปรียบเทียบความสามารถของผู้สอบแต่ละระดับชั้นของค่าอำนาจจำแนกของข้อสอบ สำหรับการทดสอบ O-NET วิชาคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 จำแนกตามเพศและความสามารถด้านการเรียนวิชาคณิตศาสตร์	133
4	ผลการวิจัย	135
	ตอนที่ 1 ผลการพัฒนาวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา	136
	ตอนที่ 2 ผลการจำลองเปรียบเทียบประสิทธิภาพของวิธีการจัดคลังข้อสอบ	138
	ตอนที่ 3 ผลการพัฒนาโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ โดยใช้คลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา	160
	ตอนที่ 4 ผลการเปรียบเทียบความสามารถของผู้สอบแต่ละระดับชั้นของค่าอำนาจจำแนกของข้อสอบ สำหรับการทดสอบ O-NET วิชาคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 จำแนกตามเพศและความสามารถด้านการเรียนวิชาคณิตศาสตร์	170
5	สรุปและอภิปรายผล	177
	สรุปผลการวิจัย	177
	อภิปรายผล	180
	ข้อเสนอแนะในการนำผลการวิจัยไปใช้	186
	ข้อเสนอแนะสำหรับการวิจัยต่อไป	187
บรรณานุกรม		189
ภาคผนวก		200
	ภาคผนวก ก รายชื่อผู้เชี่ยวชาญ	201
	ภาคผนวก ข ผลการวิเคราะห์คุณภาพข้อสอบ	203
	ภาคผนวก ค ผลการศึกษาในสถานการณ์จำลองด้านการประมาณความสามารถของผู้สอบ	214
	ภาคผนวก ง ผลการศึกษาในสถานการณ์จำลองด้านประสิทธิภาพของการใช้ข้อสอบ	231
	ภาคผนวก จ คู่มือการใช้โปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติสำหรับการทดสอบ O-NET วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 6	249
	ภาคผนวก ฉ แบบประเมินความเหมาะสมของโปรแกรม สำหรับผู้เชี่ยวชาญ	279
	ภาคผนวก ช แบบประเมินความเหมาะสมของโปรแกรม สำหรับผู้ใช้งาน	282
ประวัติย่อผู้วิจัย		285

สารบัญตาราง

ตารางที่		หน้า
2-1	กระบวนการพุทธิปัญญาของ Bloom et al. (1956) แบบปรับใหม่	17
2-2	โครงสร้างสองมิติของจุดประสงค์ทางการศึกษาของ Bloom et al. (1956) แบบปรับใหม่	18
2-3	ค่าสถิติพื้นฐานผลการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน (O-NET) ชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2558 ระดับประเทศ (คะแนนเต็ม 100 คะแนน)	20
2-4	การเปรียบเทียบคะแนนเฉลี่ยการสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2556 – 2558	21
2-5	ความสัมพันธ์ของขนาดเลขชี้กำลังกับจำนวนมิติจากโมเดลของ Mulaik (p=.5)	28
2-6	ความสัมพันธ์ของขนาดเลขชี้กำลังกับจำนวนมิติจากโมเดลของ Sympson (p=.5)	29
2-7	สรุปกลยุทธการควบคุมการแสดงข้อสอบ (Controlling Item Exposure)	61
3-1	จำนวนข้อสอบ O-NET วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 6 จำแนกตามสาระการเรียนรู้ที่ออกข้อสอบ	88
3-2	จำนวนข้อสอบ O-NET วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 6 จำแนกตามจุดประสงค์ทางการศึกษาด้านพุทธิปัญญาและสาระการเรียนรู้ที่ออกข้อสอบ	91
3-3	จำนวนข้อสอบที่ผ่านเกณฑ์การวิเคราะห์ข้อสอบตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ	92
3-4	สรุปเกณฑ์การเปรียบเทียบประสิทธิภาพด้านการประมาณค่าความสามารถของผู้สอบและด้านประสิทธิภาพการใช้ข้อสอบ	110
3-5	สัญลักษณ์และความหมายที่ใช้ในการวิเคราะห์แผนผังบริบท	112
3-6	สัญลักษณ์และความหมายที่ใช้ในผังการไหลของข้อมูล	113
3-7	แฟ้มข้อมูลข้อสอบ	118
3-8	แฟ้มข้อมูลผู้สอบ	119
3-9	แฟ้มข้อมูลวันที่ทดสอบของนักเรียน	119
3-10	แฟ้มข้อมูลผลการทดสอบของนักเรียน	119
3-11	แฟ้มข้อมูลตารางค่า Theta	120
3-12	แฟ้มข้อมูลแบบประเมิน (สำหรับผู้เชี่ยวชาญ)	121
3-13	แฟ้มข้อมูลแบบประเมิน (สำหรับผู้ใช้งาน)	122
3-14	แฟ้มข้อมูลผู้ดูแลระบบ	123
4-1	ค่าสถิติพื้นฐานของค่าความสามารถที่แท้จริงของผู้สอบของคลังข้อสอบชั้นที่ 1	138
4-2	ค่าสถิติพื้นฐานของค่าความสามารถที่แท้จริงของผู้สอบของคลังข้อสอบชั้นที่ 2	140

สารบัญตาราง (ต่อ)

ตารางที่		หน้า
4-3	ค่าสถิติพื้นฐานของค่าความสามารถที่แท้จริงของผู้สอบของคลังข้อสอบชั้นที่ 3.....	141
4-4	ค่าสถิติพื้นฐานของค่าความสามารถที่แท้จริงของผู้สอบของคลังข้อสอบชั้นที่ 4.....	142
4-5	ค่าสถิติพื้นฐานของค่าพารามิเตอร์ของข้อสอบในคลังข้อสอบจำลอง.....	143
4-6	ค่ารากที่สองของค่าความลำเอียงเฉลี่ย และความคลาดเคลื่อนกำลังสองเฉลี่ย จำแนกตามวิธีการจัดคลังข้อสอบชั้นที่ 1	144
4-7	ผลการวิเคราะห์ความแตกต่างของค่าความลำเอียงเฉลี่ยและค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย จำแนกตามวิธีการจัดคลังข้อสอบชั้นที่ 1.....	144
4-8	ค่ารากที่สองของค่าความลำเอียงเฉลี่ย และความคลาดเคลื่อนกำลังสองเฉลี่ย จำแนกตามวิธีการจัดคลังข้อสอบชั้นที่ 2.....	145
4-9	ผลการวิเคราะห์ความแตกต่างของค่าความลำเอียงเฉลี่ยและค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย จำแนกตามวิธีการจัดคลังข้อสอบชั้นที่ 2.....	146
4-10	ค่ารากที่สองของค่าความลำเอียงเฉลี่ย และความคลาดเคลื่อนกำลังสองเฉลี่ย จำแนกตามวิธีการจัดคลังข้อสอบชั้นที่ 3.....	146
4-11	ผลการวิเคราะห์ความแตกต่างของค่าความลำเอียงเฉลี่ยและค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย จำแนกตามวิธีการจัดคลังข้อสอบชั้นที่ 3.....	147
4-12	ค่ารากที่สองของค่าความลำเอียงเฉลี่ย และความคลาดเคลื่อนกำลังสองเฉลี่ย จำแนกตามวิธีการจัดคลังข้อสอบชั้นที่ 4.....	147
4-13	ผลการวิเคราะห์ความแตกต่างของค่าความลำเอียงเฉลี่ยและค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย จำแนกตามวิธีการจัดคลังข้อสอบชั้นที่ 4.....	148
4-14	จำนวนข้อสอบที่มีการแสดงมากเกินไป (มากกว่า 0.2) และข้อสอบที่มีการใช้น้อยเกินไป (น้อยกว่า 0.2) จำแนกตามวิธีการจัดคลังข้อสอบชั้นที่ 1	149
4-15	ผลการวิเคราะห์ความแตกต่างของจำนวนข้อสอบที่มีการแสดงมากเกินไป (มากกว่า 0.2) และข้อสอบที่มีการใช้น้อยเกินไป (น้อยกว่า 0.2) จำแนกตามวิธีการจัดคลังข้อสอบชั้นที่ 1.....	149
4-16	จำนวนข้อสอบที่มีการแสดงมากเกินไป (มากกว่า 0.2) และข้อสอบที่มีการใช้น้อยเกินไป (น้อยกว่า 0.2) จำแนกตามวิธีการจัดคลังข้อสอบชั้นที่ 2	150
4-17	ผลการวิเคราะห์ความแตกต่างของจำนวนข้อสอบที่มีการแสดงมากเกินไป (มากกว่า 0.2) และข้อสอบที่มีการใช้น้อยเกินไป (น้อยกว่า 0.2) จำแนกตามวิธีการจัดคลังข้อสอบชั้นที่ 2.....	151
4-18	จำนวนข้อสอบที่มีการแสดงมากเกินไป (มากกว่า 0.2) และข้อสอบที่มีการใช้น้อยเกินไป (น้อยกว่า 0.2) จำแนกตามวิธีการจัดคลังข้อสอบชั้นที่ 3	151

สารบัญตาราง (ต่อ)

ตารางที่		หน้า
4-19	ผลการวิเคราะห์ความแตกต่างของจำนวนข้อสอบที่มีการแสดงมากเกินไป (มากกว่า 0.2) และข้อสอบที่มีการใช้น้อยเกินไป (น้อยกว่า 0.2) จำแนกตามวิธีการจัดคลังข้อสอบชั้นที่ 3	152
4-20	จำนวนข้อสอบที่มีการแสดงมากเกินไป (มากกว่า 0.2) และข้อสอบที่มีการใช้น้อยเกินไป (น้อยกว่า 0.2) จำแนกตามวิธีการจัดคลังข้อสอบชั้นที่ 4	152
4-21	ผลการวิเคราะห์ความแตกต่างของจำนวนข้อสอบที่มีการแสดงมากเกินไป (มากกว่า 0.2) และข้อสอบที่มีการใช้น้อยเกินไป (น้อยกว่า 0.2) จำแนกตามวิธีการจัดคลังข้อสอบชั้นที่ 4	153
4-22	จำนวนอัตราการทับซ้อนของข้อสอบ และ การแจกแจงอัตราการแสดงข้อสอบ จำแนกตามวิธีการจัดคลังข้อสอบชั้นที่ 1	153
4-23	ผลการวิเคราะห์ความแตกต่างของจำนวนอัตราการทับซ้อนของข้อสอบ และ การแจกแจงอัตราการแสดงข้อสอบ จำแนกตามวิธีการจัดคลังข้อสอบชั้นที่ 1	154
4-24	จำนวนอัตราการทับซ้อนของข้อสอบ และ การแจกแจงอัตราการแสดงข้อสอบ จำแนกตามวิธีการจัดคลังข้อสอบชั้นที่ 2	154
4-25	ผลการวิเคราะห์ความแตกต่างของจำนวนอัตราการทับซ้อนของข้อสอบ และ การแจกแจงอัตราการแสดงข้อสอบ จำแนกตามวิธีการจัดคลังข้อสอบชั้นที่ 2	155
4-26	จำนวนอัตราการทับซ้อนของข้อสอบ และ การแจกแจงอัตราการแสดงข้อสอบ จำแนกตามวิธีการจัดคลังข้อสอบชั้นที่ 3	156
4-27	ผลการวิเคราะห์ความแตกต่างของจำนวนอัตราการทับซ้อนของข้อสอบ และ การแจกแจงอัตราการแสดงข้อสอบ จำแนกตามวิธีการจัดคลังข้อสอบชั้นที่ 3	156
4-28	จำนวนอัตราการทับซ้อนของข้อสอบ และ การแจกแจงอัตราการแสดงข้อสอบ จำแนกตามวิธีการจัดคลังข้อสอบชั้นที่ 4	157
4-29	ผลการวิเคราะห์ความแตกต่างของจำนวนอัตราการทับซ้อนของข้อสอบ และ การแจกแจงอัตราการแสดงข้อสอบ จำแนกตามวิธีการจัดคลังข้อสอบชั้นที่ 4	158
4-30	สรุปผลการจำลองข้อมูลเปรียบเทียบประสิทธิภาพของวิธีการจัดคลังข้อสอบ	158
4-31	ผลการประเมินความเหมาะสมของโปรแกรม โดยผู้เชี่ยวชาญ	168
4-32	ผลการประเมินความเหมาะสมของโปรแกรม โดยผู้ใช้งาน	169
4-33	ข้อมูลพื้นฐานและผลการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ	170
4-34	การเปรียบเทียบค่าเฉลี่ยของค่าความสามารถของผู้สอบ จำแนกตามเพศและความสามารถด้านการเรียนวิชาคณิตศาสตร์	174
4-35	ผลการวิเคราะห์ความแปรปรวนของค่าความสามารถของผู้สอบ จำแนกตามเพศและความสามารถด้านการเรียนวิชาคณิตศาสตร์	175

สารบัญภาพ

ภาพที่		หน้า
1-1	กรอบแนวทางการวิจัยเรื่อง การพัฒนาวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา	10
2-1	การเปรียบเทียบคะแนนเฉลี่ยการสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2556-2558.....	21
2-2	แนวโน้มคะแนนวิชาคณิตศาสตร์ของนักเรียนไทยกลุ่มสูงและกลุ่มต่ำจาก PISA 2000-2015.....	22
2-3	แนวโน้มคะแนนวิชาคณิตศาสตร์จาก PISA 2000-2015.....	23
2-4	คะแนนวิชาคณิตศาสตร์ของนักเรียนตามกลุ่มโรงเรียนใน PISA 2015.....	23
2-5	การเปลี่ยนแปลงคะแนนวิชาคณิตศาสตร์ของนักเรียนตามกลุ่มโรงเรียนใน PISA 2012 และ PISA 2015.....	23
2-6	ขั้นตอนการดำเนินการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์.....	35
2-7	รูปแบบการทดสอบแบบ 2 ขั้นตอน	36
2-8	โครงสร้างการจัดการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์รูปแบบปรับระดับชั้น	37
2-9	ลักษณะของ Between-items Multidimensionality และ Within-items Multidimensionality	39
2-10	ลักษณะของ Unidimensional และ Multidimensional.....	39
3-1	ระยะการดำเนินการวิจัย.....	84
3-2	ขั้นตอนการพัฒนาวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา.....	85
3-3	ขั้นตอนการรวบรวมข้อสอบและวิเคราะห์เนื้อหาการออกข้อสอบ.....	87
3-4	การวิเคราะห์เนื้อหาการออกข้อสอบ พร้อมระบุสาระการเรียนรู้ มาตรฐานการเรียนรู้และตัวชี้วัดของข้อสอบ.....	88
3-5	ขั้นตอนการวิเคราะห์จุดประสงค์ทางการศึกษาตามแนวคิดของ Bloom et al. (1956) แบบปรับใหม่และการวิเคราะห์คุณภาพข้อสอบ.....	89
3-6	ผลการพิจารณาความตรงเชิงเนื้อหาของผู้เชี่ยวชาญ.....	91
3-7	ขั้นตอนการพัฒนาวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา.....	94
3-8	ตัวอย่างการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา.....	95
3-9	ขั้นตอนการทดสอบในแต่ละชั้นของคลังข้อสอบวิธี IACB-MCAT.....	96
3-10	การเลือกข้อสอบข้อแรก.....	97

สารบัญภาพ (ต่อ)

ภาพที่		หน้า
3-11	การเลือกข้อสอบข้อถัดไป	98
3-12	การเลือกข้อสอบข้อถัดไป เพื่อใช้เป็นข้อสอบข้อแรกในคลังข้อสอบชั้นที่ 2	99
3-13	ขั้นตอนการศึกษาในสถานการณ์จำลองเพื่อเปรียบเทียบประสิทธิภาพของวิธีการจัดคลังข้อสอบ	101
3-14	ขั้นตอนของวิธี IACB	103
3-15	ขั้นตอนของวิธี CWA	105
3-16	ขั้นตอนการพัฒนาโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ	111
3-17	แผนบริบทของโปรแกรมการทดสอบ	112
3-18	การรับส่งข้อมูลของกระบวนการหลักในโปรแกรมการทดสอบ	114
3-19	แผนผังการไหลของข้อมูลระดับ 1 ของกระบวนการจัดคลังข้อสอบ	115
3-20	แผนผังการไหลของข้อมูลระดับ 1 ของกระบวนการจัดการทดสอบด้วยโปรแกรม	116
3-21	ความสัมพันธ์ของตารางข้อมูล	123
3-22	โครงสร้างหน้าจอหลัก	124
3-23	โครงสร้างหน้าจอเมนูผู้ใช้งาน	124
3-24	โครงสร้างหน้าจอเว็บ	124
3-25	โครงสร้างหน้าจอเข้าสู่ระบบตรวจสอบผลการทดสอบ	125
3-26	โครงสร้างหน้าจอผลการทดสอบ	125
3-27	โครงสร้างหน้าจอผลการตอบข้อสอบเป็นรายข้อ	125
3-28	โครงสร้างหน้าจอผลการตอบจำแนกตามจุดประสงค์ทางการศึกษา	126
3-29	โครงสร้างหน้าจอผลการวิเคราะห์ความสามารถของผู้สอบ	126
3-30	โครงสร้างหน้าจอพิมพ์รายงาน	126
3-31	โครงสร้างหน้าจอแบบประเมิน	127
3-32	โครงสร้างหน้าจอติดต่อผู้ดูแลระบบ	127
3-33	โครงสร้างหน้าจอความหมาย IACB-MCAT	127
3-34	โครงสร้างหน้าจอคำชี้แจงของระบบ	127
3-35	โครงสร้างหน้าจอเข้าสู่ระบบ	128
3-36	โครงสร้างหน้าจอลงทะเบียนเข้าใช้งานระบบ	128
3-37	โครงสร้างหน้าจอการทดสอบ	128
3-38	โครงสร้างหน้าจอการทดสอบครั้งสุดท้าย	129
3-39	โครงสร้างหน้าจอเข้าสู่ระบบผู้ดูแลระบบ	129
3-40	โครงสร้างหน้าจอค้นหาข้อสอบ	130

สารบัญภาพ (ต่อ)

ภาพที่		หน้า
3-41	โครงสร้างหน้าจอเมนูดูแลระบบ	130
3-42	โครงสร้างหน้าจอเพิ่มข้อสอบ	130
3-43	โครงสร้างหน้าจอแก้ไขข้อสอบ	131
3-44	ขั้นตอนการทดสอบในระยะที่ 4	133
4-1	วิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา	137
4-2	หน้าจอหลักของโปรแกรมการทดสอบ	161
4-3	หน้าเริ่มต้นเข้าสู่ระบบทดสอบ	162
4-4	หน้าจอลืมรหัสผ่าน	163
4-5	หน้าจอลงทะเบียนเพื่อเข้าใช้งานในระบบ	163
4-6	หน้าจอข้อสอบข้อแรก	164
4-7	หน้าจอรายงานผลการทดสอบ	164
4-8	หน้าจอรายงานผลการตอบข้อสอบเป็นรายข้อจำแนกตามชั้นของค่าอำนาจจำแนกของข้อสอบ	165
4-9	หน้าจอรายงานผลการตอบจำแนกตามจุดประสงค์ทางการศึกษาด้านพุทธิปัญญา	166
4-10	หน้าจอรายงานผลการวิเคราะห์ความสามารถของผู้สอบจำแนกแต่ละระดับชั้นของค่าอำนาจจำแนกข้อสอบ	166
4-11	หน้าจอพิมพ์รายงาน	167
4-12	หน้าจอเกณฑ์การประเมินความสามารถของผู้สอบ	167

บทที่ 1
บทนำ

ความเป็นมาและความสำคัญของปัญหา

การพัฒนาประเทศไทยไปสู่ความมั่นคง มั่งคั่ง และยั่งยืน ให้เกิดขึ้นในอนาคตนั้น จะต้องให้ความสำคัญกับการเสริมสร้างทุนของประเทศที่มีอยู่ให้เข้มแข็ง และมีพลังเพียงพอในการขับเคลื่อนกระบวนการการพัฒนาทั้งในระยะกลางและระยะยาว โดยเฉพาะ “การพัฒนาคน” ให้มีการเตรียมความพร้อมรับการเปลี่ยนแปลงของโลกในศตวรรษที่ 21 สิ่งที่สำคัญที่สุดคือทักษะการเรียนรู้และการเสริมสร้างปัจจัยแวดล้อมที่เอื้อต่อการพัฒนาคุณภาพของคน ไปสู่การปฏิบัติให้เกิดผลสำเร็จตามเป้าหมายนั้น มียุทธศาสตร์สำคัญในการขับเคลื่อนประการหนึ่งคือ ยุทธศาสตร์ที่ 4 การพัฒนาศักยภาพคนทุกช่วงวัยและการสร้างสังคมแห่งการเรียนรู้ มีเป้าหมายหนึ่งให้มีระบบและกลไกการวัดการติดตามและประเมินผลมีประสิทธิภาพ โดยมีแนวทางในการพัฒนาระบบและกลไกการติดตามการวัดและการประเมินผลผู้เรียนให้มีประสิทธิภาพที่ระบุว่า “การพัฒนาระบบและการให้บริการคลังข้อสอบเพื่อการวัดและประเมินผลการเรียนรู้ที่มีคุณภาพและมาตรฐาน ครอบคลุมการวัดความรู้ทักษะ และสมรรถนะของผู้เรียน” (กระทรวงศึกษาธิการ, 2560, หน้า 108-118) แผนพัฒนาการศึกษาของชาติ ปี 2560-2579 ได้กล่าวถึงวิสัยทัศน์ไว้ว่า “คนไทยทุกคนได้รับการศึกษาและเรียนรู้ตลอดชีวิตอย่างมีคุณภาพ ดำรงชีวิตอย่างเป็นสุข สอดคล้องกับหลักปรัชญาของเศรษฐกิจพอเพียงและการเปลี่ยนแปลงของโลกศตวรรษที่ 21” (กระทรวงศึกษาธิการ, 2560, หน้า ฉ)

คลังข้อสอบ (Item Pools, Item Bank) เป็นแหล่งรวบรวมข้อสอบของแบบทดสอบ การทดสอบด้วยแบบทดสอบดั้งเดิม ผู้สอบทั้งหมดจะได้รับข้อสอบชุดเดียวกัน แต่สำหรับการทดสอบแบบปรับเหมาะ ผู้สอบจะได้รับข้อสอบเหมาะสมกับความสามารถของผู้สอบแต่ละคน คลังข้อสอบจึงเป็นเสมือนแหล่งรวบรวมชุดแบบทดสอบจำนวนมากและมีความเป็นคู่ขนานกัน ข้อสอบที่บรรจุในคลังข้อสอบมีลักษณะค่าความยากกระจายตามระดับความสามารถของประชากรผู้สอบ ค่าอำนาจจำแนกสูงเพื่อให้มีประสิทธิภาพการทดสอบสูงสุด และค่าโอกาสในการเดาเข้าใกล้หรือเท่ากับศูนย์ ข้อสอบควรมีจำนวนมากเพียงพอและกระจายทุกระดับความสามารถของผู้สอบ ดังนั้น ต้องใช้ข้อสอบจำนวนมากสำหรับสร้างคลังข้อสอบ ทำให้คลังข้อสอบสำหรับการทดสอบแบบปรับเหมาะมีขนาดใหญ่กว่าการทดสอบแบบดั้งเดิม (Thompson & Weiss, 2011; Babcock & Weiss, 2012) ซึ่งประเภทของคลังข้อสอบ แบ่งตาม Zhang, Chang and Yi (2012) เป็น 2 กลุ่ม ได้แก่ 1) คลังข้อสอบมิติเดียว (Single Item Pools) และ 2) คลังข้อสอบหลายมิติ (Multiple Item Pools) มีรายละเอียดดังนี้ 1) คลังข้อสอบมิติเดียว มีรูปแบบการทดสอบแบบปรับเหมาะที่เหมาะสม ก็ต่อเมื่อข้อสอบมีมิติเด่นมิติเดียว (Dominant Dimension) มีรูปแบบการจัดคลังข้อสอบได้หลายวิธี เช่น 1.1) P-Optimal, D-Optimal ซึ่งมีงานวิจัยของ Kuo, Daud, and Yang (2015); Seo and Weiss (2015); Mao (2014); He and Reckase (2014); Zhang, Chang, and Yi (2012); Zhou (2012); Diao and Reckase (2009) 1.2) Polytomous ซึ่งมีงานวิจัยของ Zhou and Reckase (2014) และ 2) คลังข้อสอบหลายมิติ ลักษณะข้อสอบมีมิติเด่น 2 มิติ หรือข้อสอบพหุมิติมีแนวโน้มว่าให้ค่าอำนาจจำแนก

ของข้อสอบต่ำและเลือกข้อสอบที่มีอำนาจจำแนกของข้อสอบสูง มีรูปแบบการจัดคลังข้อสอบได้หลายวิธี เช่น 2.1) คลังข้อสอบแบบแบ่งชั้น (Stratification Item Pools) ซึ่งมีงานวิจัยของ Barrada, Abad and Olea (2014); Huo (2010); Cheng, Chang, Douglas, and Guo (2009) 2.2) คลังข้อสอบแบบหลายรายการ (Multiple Item Pools) ซึ่งมีงานวิจัยของ Zhang, Chang, and Yi (2012); Lim (2011) การกำหนดขนาดของคลังข้อสอบพิจารณาจากปัจจัยหลายด้าน ซึ่งปัจจัยในการออกแบบคลังข้อสอบที่ส่งผลต่อขนาดของคลังข้อสอบ เช่น วิธีการคัดเลือกข้อสอบ การควบคุมเนื้อหาตามมาตรวัดทางจิตวิทยา การยุติการทดสอบ การทับซ้อนของข้อสอบ การให้คะแนน ความเป็นคู่ขนานกับแบบทดสอบดั้งเดิม ดังนั้น การออกแบบและพัฒนาคลังข้อสอบสำหรับการทดสอบแบบปรับเหมาะ จึงเป็นสิ่งสำคัญที่ควรคำนึงถึง (Reckase, 2009, pp. 336-337)

นักวิชาการจำนวนมากได้ศึกษาในประเด็นเกี่ยวกับขนาดคลังข้อสอบที่มีส่วนช่วยเพิ่มประสิทธิภาพการทดสอบ ได้แก่ Weiss (1985) แนะนำว่า การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์จะให้ประสิทธิภาพ เมื่อคลังข้อสอบมีจำนวนข้อสอบมากเพียงพอสำหรับจำแนกผู้สอบ คลังข้อสอบที่ดีจะต้องมีจำนวนข้อสอบเท่ากับ 2^n เมื่อ n คือ จำนวนข้อสอบที่ต้องการใช้ทดสอบ เช่น ถ้าต้องการใช้ทดสอบประมาณ 10 ข้อ ต้องมีข้อสอบในคลัง 1,024 ข้อ แต่ถ้าหากต้องการทดสอบแบบกำหนดจำนวนข้อที่จะใช้คงที่เหมือนกันทุกคน จำนวนข้อสอบในคลังข้อสอบควรมีอย่างน้อย $n(n+1)/2$ ข้อ เช่น ถ้าต้องการใช้ทดสอบ 10 ข้อ ต้องมีข้อสอบในคลัง จำนวน 55 ข้อ การศึกษาของ Xing and Hambleton (2004) ปรากฏว่า คลังข้อสอบขนาดใหญ่ทำให้สารสนเทศของแบบทดสอบเพิ่มขึ้นตามไปด้วย สอดคล้องกับความคิดเห็นของ Lord (1980) ซึ่งอ้างถึงการศึกษาในสถานการณ์จำลอง เมื่อใช้ความยาวของแบบทดสอบ 25 ข้อ ขนาดคลังข้อสอบ 363 ข้อ และ 183 ข้อ ปรากฏว่า ขนาดคลังข้อสอบ 363 ข้อสามารถให้สารสนเทศได้มากกว่า เพราะคลังข้อสอบขนาดใหญ่มีโอกาสเลือกข้อสอบดีที่สุดจัดให้ผู้สอบได้มากกว่าคลังข้อสอบขนาดเล็ก ต่อมา Segall (2005) เสนอว่า ขนาดคลังข้อสอบควรมีขนาดประมาณ 6-8 เท่าของความยาวแบบทดสอบ ส่วน de Ayla (2009, p. 376) 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, p. 264) กล่าวว่า คลังข้อสอบควรประกอบ ด้วยข้อสอบประมาณ 100 ข้อ ซึ่งหากรูปแบบการตอบข้อสอบเป็นแบบที่มีการตรวจให้คะแนนมากกว่า 2 ค่า จำนวนคำถามในคลังข้อสอบมีจำนวนน้อยกว่านี้ได้

Choppin (1985) และ Rudner (1998, pp. 1-2) ได้สรุปประโยชน์ของคลังข้อสอบไว้ดังนี้ 1) ได้ข้อสอบที่ตรงกับจุดมุ่งหมายและเนื้อหาของแต่ละรายวิชา สามารถเลือกใช้ได้อย่างเหมาะสมกับสถานการณ์ 2) ได้ข้อสอบและแบบทดสอบที่มีคุณภาพน่าเชื่อถือเป็นจำนวนมาก 3) สร้างข้อสอบได้อย่างรวดเร็วและถูกต้อง 4) ใช้ข้อสอบซ้ำได้หลายครั้ง ด้วยการเลือกใช้ข้อสอบที่มีระบบ 5) ประหยัดงบประมาณ ทรัพยากร และเวลา ลดภาระด้านการออกข้อสอบให้แก่ครูผู้สอน 6) ได้ข้อสอบที่มีลักษณะเป็นข้อสอบคู่ขนานได้ 7) สามารถจัดทำแบบทดสอบที่เหมาะกับความสามารถของกลุ่มผู้สอบ และ 8) กระตุ้นครูผู้สอนให้เกิดการพัฒนาข้อสอบให้มีคุณภาพและมีมาตรฐานที่สูงยิ่งขึ้น

การพัฒนาคลังข้อสอบสำหรับการทดสอบแบบปรับเหมาะมีความคล้ายคลึงกับคลังข้อสอบแบบดั้งเดิม ข้อสอบจะต้องสร้างตามตารางกำหนดเนื้อหาและควรมีการตรวจสอบคุณภาพของเนื้อหาและความไวของแบบทดสอบ (Test Sensitivity) และมีการทดลองใช้เพื่อศึกษาคุณลักษณะทางจิตวิทยา (Wainer et al., 2001, pp. 105-112) วิธีการตรวจสอบคุณภาพของข้อสอบสามารถใช้วิธีการทางสถิติของทฤษฎีการทดสอบแบบดั้งเดิม เช่น สัดส่วนของผู้ที่ตอบถูก (Proportion Correct) หรือ สหสัมพันธ์แบบไบซีเรียล (Biserial Correlation) ร่วมกับทฤษฎีการทดสอบแนวใหม่ เช่น พารามิเตอร์ของข้อสอบ (Item Parameters) และค่าสารสนเทศของข้อสอบ (Item Information) (Wainer, 1989) และสิ่งที่สำคัญอีกประการหนึ่ง คือคุณลักษณะของข้อสอบ (Item Characterisitic) เพื่อตรวจสอบว่า ข้อสอบแต่ละข้อเหมาะสมกับโมเดลการวัด Wainer et al., (2001, pp. 37-45) และ Thompson and Weiss (2011) ได้เสนอขั้นตอนการพัฒนาคลังข้อสอบสำหรับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ดังนี้ 1) สร้างข้อสอบให้มีจำนวนมากพอที่จะครอบคลุมเนื้อหาทั้งหมดตามกฎเบื้องต้นในการเขียนข้อสอบ 2) ตรวจสอบคุณภาพของข้อสอบเพื่อให้ข้อสอบมีคุณภาพสูงสุด 3) นำข้อสอบที่สร้างขึ้นไปทดลองใช้ 4) คัดเลือกข้อสอบที่มีคุณภาพด้วยการวิเคราะห์ตามหลักสถิติจากการทดลองใช้ 5) ตรวจสอบสัดส่วนของข้อสอบในแต่ละเนื้อหา และประเมินกระบวนการทดสอบด้วยการจำลองสถานการณ์การทดสอบ (Simulation) เพื่อพิจารณาว่าข้อสอบครอบคลุมช่วงของคุณลักษณะแฝงที่ต้องการวัดหรือไม่ และ 6) ดำเนินการปรับข้อสอบที่เหมาะสมทั้งหมดให้อยู่ในรูปแบบของคอมพิวเตอร์ตามแนวทางการจัดการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์

สำหรับแนวคิดการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ (Computerized Adaptive Testing: CAT) ได้รับความนิยมมากในการทดสอบทางการศึกษาและจิตวิทยา เนื่องจากเป็นการทดสอบที่ตอบสนองความต้องการของผู้สอบในเชิงจิตวิทยา ด้วยการเลือกข้อสอบให้เหมาะกับระดับความสามารถหรือคุณลักษณะของผู้เข้าสอบ (Frey & Seitz, 2009) อีกทั้งเพิ่มประสิทธิภาพของการวัดให้มีความแม่นยำ และสามารถลดระยะเวลาในการทดสอบลงได้ (Segall, 2005; Frey & Seitz, 2009) และได้รับการยอมรับว่า มีประสิทธิภาพมากกว่าการทดสอบแบบดั้งเดิม ทั้งนี้โดยทั่วไปการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์สามารถที่จะลดจำนวนข้อสอบลงจากการทดสอบแบบดั้งเดิมครึ่งหนึ่ง โดยไม่สูญเสียความแม่นยำในการวัด จึงช่วยให้ลดภาระงาน เวลา และค่าใช้จ่ายลงไปอย่างมาก (Frey & Seitz, 2009) มีข้อตกลงเบื้องต้นคือต้องเป็นรูปแบบเอกมิติ (Unidimensional) ซึ่งการใช้ข้อสอบแบบเอกมิติในบางครั้ง จะไม่เหมาะสมกับการทดสอบจริง (Wang & Chen, 2004) เนื่องจากการทดสอบส่วนใหญ่มีลักษณะเป็นพหุมิติ (Multidimensional) ปัจจุบันจึงมีผู้เริ่มศึกษาการทดสอบแบบปรับเหมาะในลักษณะของทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ (Multidimensional Item Response Theory: MIRT) มากขึ้น ส่วนแนวคิดการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ (Multidimensional Computer Adaptive Testing: MCAT) เป็นการขยายแนวคิดจากการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ จากการวัดด้วยข้อสอบที่วัดได้เพียงหนึ่งคุณลักษณะ ไปเป็นการวัดด้วยข้อสอบที่สามารถวัดคุณลักษณะได้มากกว่าหนึ่งคุณลักษณะ แนวคิดนี้ ถ้าใช้จำนวนข้อสอบที่เท่ากัน การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติจะเป็นการทดสอบที่มีประสิทธิภาพสูงกว่าการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์และการทดสอบแบบดั้งเดิม เช่น

การทดสอบด้วยการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติสามารถลดจำนวนข้อสอบลงได้มากกว่าการทดสอบด้วยการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ประมาณ 30-50% และลดจำนวนข้อสอบได้มากกว่าการทดสอบแบบดั้งเดิมประมาณ 70% โดยไม่สูญเสียความแม่นยำในการทดสอบ (Frey & Seitz, 2009)

การจัดสมดุลเนื้อหาของข้อสอบ (Content Balancing) เป็นสิ่งสำคัญอย่างยิ่ง โดยเฉพาะการวัดผลสัมฤทธิ์ทางการศึกษา ในการทดสอบแบบดั้งเดิม แบบทดสอบมาตรฐานส่วนใหญ่ได้สร้างตามรายละเอียดในตารางการกำหนดคุณลักษณะของข้อสอบ (Table of Content Specification) ซึ่งผู้สอบทุกคนจะได้รับแบบทดสอบชุดเดียวกัน ประกอบด้วยข้อสอบครบทุกคุณลักษณะที่ต้องการวัดตามที่ออกแบบไว้ ด้วยเหตุนี้ จึงเป็นข้อจำกัดในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ เนื่องจากการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ จะคัดเลือกข้อสอบที่ให้ค่าสารสนเทศสูงสุดที่ค่าประมาณความสามารถของผู้สอบ ณ ขณะนั้น ซึ่งไม่ได้พิจารณาขอบเขตเนื้อหา (Content Area) ของข้อสอบ เป็นผลให้ผู้สอบบางคนไม่ได้รับข้อสอบทุกขอบเขตเนื้อหาของข้อสอบ เพื่อรับประกันว่าผู้สอบแต่ละคนจะได้รับข้อสอบที่เป็นตัวแทนตามสัดส่วนของแต่ละขอบเขตของเนื้อหาครบถ้วน และเพื่อให้แน่ใจว่าผู้สอบทุกคนได้รับข้อสอบในแต่ละเนื้อหาเท่ากัน ในการคัดเลือกข้อสอบข้อถัดไป จึงต้องคำนึงถึงการจัดสมดุลเนื้อหาของข้อสอบ (Wainer et al., 2001, pp. 120-121; Thompson & Weiss, 2011)

นักวิชาการหลายคนได้ศึกษาการจัดสมดุลเนื้อหาของข้อสอบ ซึ่งวิธีที่นิยมมี 3 วิธี ได้แก่ 1) Kingsbury and Zara (1998) ได้นำเสนอวิธีการทดสอบแบบปรับเหมาะแบบควบคุมสมดุลเนื้อหา (Constrained Computerized Adaptive Testing: CCAT) วิธีนี้เลือกข้อสอบที่เหมาะสมที่สุดจากขอบเขตเนื้อหา โดยพิจารณาอัตราการแสดงข้อสอบครั้งหลังสุด และห่างจากเปอร์เซ็นต์ของการจัดข้อสอบตามกำหนดมากที่สุด ซึ่งมีงานวิจัยที่ใช้วิธีการจัดสมดุลเนื้อหาด้วยวิธี CCAT ได้แก่ Chang et al., (2009); Murphy, Dodd, and Vaughn, (2010); Han, (2012) 2) Leung, Chang and Hau (2003) ได้นำเสนอวิธีการทดสอบแบบปรับเหมาะแบบควบคุมสมดุลเนื้อหาประยุกต์ (Modified Constrained Computerized Adaptive Testing: MCCAT) วิธีนี้ประยุกต์มาจากการทดสอบแบบปรับเหมาะแบบควบคุมสมดุลเนื้อหา เพื่อลดการคาดเดาการเรียงเนื้อหาและให้การควบคุมสมดุลเนื้อหาดียิ่งขึ้น ข้อสอบเลือกมาจากขอบเขตเนื้อหาทั้งหมดที่ยังไม่ได้ใช้เพื่อขจัดผลจากการเรียงลำดับ และ 3) Ankenmann and Spray (1999) ได้นำเสนอการจัดสมดุลเนื้อหารูปแบบมัลติโนเมียลประยุกต์ (Modified Multinomial Model: MMM) วิธีการทดสอบแบบปรับเหมาะแบบควบคุมสมดุลเนื้อหา (CCAT) สามารถคาดเดาอันดับการเรียงเนื้อหา ดังนั้น จึงได้พัฒนารูปแบบมัลติโนเมียลประยุกต์เพื่อต้องการให้เนื้อหามีความสมดุลกัน เมื่อขอบเขตเนื้อหามาถึงเปอร์เซ็นต์กำหนด จะมีการสร้างการแจกแจงมัลติโนเมียลเพื่อปรับเปอร์เซ็นต์ที่ยังไม่เต็มของขอบเขตเนื้อหาที่เหลืออยู่ และมีงานวิจัยที่เปรียบเทียบวิธีการจัดสมดุลเนื้อหาทั้ง 3 วิธี ได้แก่ วิธี CCAT, MCCAT และ MMM ซึ่งมีงานวิจัยของ Leung, Chang and Hua, (2003); Zhou, (2012); He and Reckase, (2014) ผลการศึกษาปรากฏว่า การจัดสมดุลเนื้อหาทั้ง 3 วิธี ให้ความถูกต้องในการประมาณค่าความสามารถของผู้สอบ ความคลาดเคลื่อนกำลังสองเฉลี่ย สัมประสิทธิ์สหสัมพันธ์ ค่าประมาณความสามารถและค่าความสามารถจริงไม่แตกต่างกัน

การแสดงข้อสอบ (Item Exposure) เป็นการจัดข้อสอบที่ผ่านการคัดเลือกแล้วให้ผู้สอบ โดยเป็นส่วนหนึ่งของขั้นตอนการคัดเลือกข้อสอบข้อถัดไป ในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ขั้นตอนการคัดเลือกข้อสอบข้อถัดไป จะคัดเลือกข้อสอบที่มีความเหมาะสมกับระดับความสามารถของผู้สอบ ซึ่งเป็นข้อสอบที่ให้ค่าสารสนเทศสูงสุด ณ ระดับความสามารถของผู้สอบขณะนั้น ค่าสารสนเทศของข้อสอบจะมีค่าเพิ่มขึ้น ใน 2 กรณี คือ 1) ค่าอำนาจจำแนกของข้อสอบมีค่าเพิ่มขึ้น และ 2) ค่าความยากของข้อสอบมีค่าใกล้เคียงกับความสามารถของผู้สอบ ด้วยเหตุนี้ ในขั้นตอนการคัดเลือกข้อสอบข้อถัดไป ส่วนใหญ่จะคัดเลือกข้อสอบที่มีค่าอำนาจจำแนกของข้อสอบสูงไปใช้ในการทดสอบ เป็นผลให้ข้อสอบบางข้อถูกนำไปใช้ในการทดสอบน้อย (Underutilized Item) หรือไม่เคยถูกนำไปใช้ในการทดสอบ ทำให้ข้อสอบบางข้อถูกนำไปใช้ในการทดสอบบ่อยครั้งเกินไป (Overexposed Item) การที่ข้อสอบถูกนำไปใช้ในการทดสอบบ่อยครั้งเกินไป จะส่งผลกระทบต่อความปลอดภัยของการทดสอบ (Test Security) กล่าวคือ เมื่อมีข้อสอบถูกนำไปใช้ในการทดสอบบ่อยครั้งเกินไป จนรู้กันในกลุ่มผู้สอบ เป็นผลทำให้ผู้สอบในรุ่นถัดมาสามารถคาดการณ์ล่วงหน้าได้ว่าตนเองจะได้รับข้อสอบข้อใด ทำให้สามารถตอบข้อสอบข้อดังกล่าวได้ โดยไม่ได้ใช้ความสามารถของตนเอง คะแนนสอบจึงขาดความถูกต้อง ผู้สอบสามารถทำคะแนนได้มากเกินความสามารถของตนเอง ซึ่งส่งผลกระทบต่อความปลอดภัยของการทดสอบและประสิทธิภาพของการแสดงข้อสอบในคลังข้อสอบ

นักวิชาการหลายคนได้ศึกษาวิธีการแสดงข้อสอบ ได้แก่ Georgiadou, Triantafillou, & Econmides, (2007) ได้จำแนกวิธีการแสดงข้อสอบไว้ 5 ประเภท คือ 1) วิธีการเลือกอย่างสุ่ม (Randomized Item-selection Strategies) มีงานวิจัยที่ใช้วิธีการควบคุมการแสดงข้อสอบโดยใช้วิธีการเลือกอย่างสุ่ม เช่น Deng, Ansley, and Chang, (2010); Leroux, Lopez, Hembry, and Dodd, (2013); Ozturk & Dogan, (2015) 2) วิธีการเลือกอย่างมีเงื่อนไข (Conditional Selection Strategies) มีงานวิจัยที่ใช้วิธีการควบคุมการแสดงข้อสอบโดยใช้วิธีการเลือกอย่างมีเงื่อนไข เช่น Belov, Armstrong, and Weissman, (2008); Cheng and Chang, (2009); Han, (2012); Leroux, Lopez, Hembry, and Dodd, (2013); Ozturk and Dogan, (2015) 3) วิธีการเลือกตามระดับชั้น (Stratified Strategies) มีงานวิจัยที่ใช้วิธีการควบคุมการแสดงข้อสอบโดยใช้วิธีการเลือกตามระดับชั้น เช่น Lee and Fuh, (2008); Deng, Ansley, and Chang, (2010); Han, (2012); Zhou, (2012); Huebner, Wang, Quinlan, and Seubert, (2016) 4) วิธีการเลือกแบบรวมกลยุทธ์ (Combined Strategies) มีงานวิจัยที่ใช้วิธีการควบคุมการแสดงข้อสอบโดยใช้วิธีการเลือกแบบรวมกลยุทธ์ เช่น Cheng and Chang, (2009); Leroux, Lopez, Hembry, and Dodd, (2013) และ 5) วิธีการทดสอบหลายขั้นตอน (Multiple Stage Adaptive Test) มีงานวิจัยที่ใช้วิธีการควบคุมการแสดงข้อสอบโดยใช้วิธีการทดสอบหลายขั้นตอน เช่น Murphy, Dodd, and Vaughn, (2010); Ozturk and Dogan, (2015) ส่วนใหญ่เปรียบเทียบประสิทธิภาพของวิธีการควบคุมการแสดงข้อสอบ เพื่อหาวิธีการควบคุมการแสดงข้อสอบที่เหมาะสมกับการทดสอบในรูปแบบที่แตกต่างกัน

ในปัจจุบันการวัดผลทางการศึกษาและจิตวิทยาใช้ข้อสอบเป็นเครื่องมือในการวัดผล ข้อสอบที่สร้างขึ้นมักจะวัดมิติความสามารถที่เป็นเชิงซ้อนมากกว่ามิติความสามารถที่เป็นเชิงเดียว

(Yao & Boughton, 2007) เช่น ข้อสอบวัดทักษะการแก้ปัญหาทางคณิตศาสตร์ และทักษะการคิดคำนวณ (Miller & Hirsch, 1999) ข้อสอบจำเป็นที่จะต้องใช้ทักษะหลายๆ ด้าน ทั้งในด้านทักษะการอ่าน ทักษะการแก้ปัญหา และทักษะการคิดคำนวณ ซึ่งสอดคล้องกับข้อสอบการรู้เรื่องคณิตศาสตร์ประกอบด้วยสถานการณ์และบริบทเนื้อหาทางคณิตศาสตร์และสมรรถนะทางคณิตศาสตร์ การสอนคณิตศาสตร์ในปัจจุบันมองผู้เรียนในฐานะผู้แก้ปัญหา (OECD, 2016) ด้วยเหตุนี้แนวคิดของคนที่รู้คณิตศาสตร์ จึงเปลี่ยนแปลงไป คนที่รู้คณิตศาสตร์ไม่ใช่ผู้ที่เชี่ยวชาญตัวเลขและการทำงานทางคณิตศาสตร์ แต่ต้องสามารถใช้คณิตศาสตร์ในการแก้ปัญหาในบริบทของชีวิตจริง ต้องสามารถนำปัญหาในบริบทไปคิดวิธีการทำให้เป็นปัญหาคณิตศาสตร์ สามารถใช้ ตีความ แปลความเข้าใจถึงบทบาทของคณิตศาสตร์ที่มีในโลกจริง และสามารถตัดสินใจที่เหมาะสมต่อสถานการณ์อย่างพลเมืองที่มีความคิดรับผิดชอบ ซึ่งรวมถึงการใช้เหตุและผลทางคณิตศาสตร์ การใช้กรอบความคิด วิธีการและเครื่องมือทางคณิตศาสตร์ ซึ่งสอดคล้องกับ แนวคิดของ Bloom et al. (1956) และ Anderson et al. (2001, p. XXI); Krathwohl (2002, p. 213) โดยการปรับลำดับขั้นและคำศัพท์ที่ใช้ในกระบวนการพุทธิปัญญา The Cognitive Process Dimension มี6 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, p. XXI; Krathwohl, 2002, p. 213) และเปลี่ยนโครงสร้างจากมิติเดียวเป็นสองมิติ โดย Anderson et al. (2001, p. XXI) และ Krathwohl (2002, p. 213) ได้เพิ่มโครงสร้างในมิติด้านความรู้ (Knowledge Dimension) เข้ามาในโครงสร้างของจุดประสงค์ทางการศึกษา ด้านพุทธิปัญญา ทำให้โครงสร้างใหม่มีลักษณะเป็นสองมิติประกอบด้วยมิติด้านกระบวนการพุทธิปัญญาและมิติด้านความรู้ และอธิบายความหมายของมิติด้านความรู้ที่เพิ่มเติมเข้ามาได้ (Krathwohl, 2002, p. 213) โดยมิติด้านความรู้ แบ่งเป็น 4 ส่วน คือ ความรู้เกี่ยวกับข้อเท็จจริง (Factual Knowledge) ความรู้เกี่ยวกับมโนทัศน์ (Conceptual Knowledge) ความรู้เกี่ยวกับวิธีดำเนินการ (Procedural Knowledge) และความรู้เกี่ยวกับอภิปัญญา (Metacognitive Knowledge) และมิติด้านกระบวนการพุทธิปัญญา แบ่งเป็น 6 กระบวนการ คือ จำ (Remember) เข้าใจ (Understand) ประยุกต์ (Apply) วิเคราะห์ (Analyze) ประเมินค่า (Evaluate) และสร้างสรรค์ (Create) ซึ่งเป็นกระบวนการเรียนรู้ที่เกิดขึ้นในสมองของผู้เรียน สอดคล้องกับจุดประสงค์ทางการศึกษา ผู้เรียนจะต้องได้รับการพัฒนาทางด้านสติปัญญาเกี่ยวกับความรู้ กระบวนการคิด และการแก้ปัญหา ซึ่งวิชาคณิตศาสตร์ยังเป็นเครื่องมือที่สำคัญในการศึกษาทางด้านวิทยาศาสตร์ เทคโนโลยี และศาสตร์อื่นๆ คณิตศาสตร์จึงมีประโยชน์ต่อการดำเนินชีวิต ช่วยพัฒนาคุณภาพให้ดีขึ้น และสามารถอยู่ร่วมกับผู้อื่นได้อย่างมีความสุข (กระทรวงศึกษาธิการ, 2551ข, หน้า 56)

จากเหตุผลดังกล่าวข้างต้นประกอบกับการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องปรากฏว่ายังไม่มีงานวิจัยที่พัฒนาวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา สำหรับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ ผู้วิจัยประยุกต์จากวิธีจัดคลังข้อสอบตามระดับชั้นของค่าอำนาจจำแนกของข้อสอบ (a-Stratified Method: a-STR) ที่พัฒนาโดย Chang and Ying (1996) ซึ่งเดิมแบ่งคลังข้อสอบออกเป็น 4 ชั้น

ตามชั้นของค่าอำนาจจำแนกแบบคงที่ (a = 0.50, 1.00, 1.50, และ 2.00) ผู้วิจัยได้ขยายวิธีการ โดยแบ่งคลังข้อสอบออกเป็น 4 ชั้น ตามช่วงของค่าอำนาจจำแนกข้อสอบ (a) ชั้นที่ 1 0.50–0.99 ชั้นที่ 2 1.00–1.49 ชั้นที่ 3 1.50–1.99 และ ชั้นที่ 4 2.00–2.50 ร่วมกับการจัดสมดุลเนื้อหา (Content Balancing) ภายในแต่ละชั้น ได้แบ่งเป็น 4 กลุ่ม ตามสาระการเรียนรู้เนื้อหาการออกข้อสอบ O-NET วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 6 ได้แก่ 1) สาระการเรียนรู้จำนวนและการดำเนินการ 2) สาระการเรียนรู้การวัด 3) สาระการเรียนรู้พีชคณิต และ 4) สาระการเรียนรู้การวิเคราะห์และความน่าจะเป็น ดังนั้นวิธีการควบคุมการแสดงข้อสอบ (Controlling Item Exposure) จึงใช้วิธีการสุ่มตัวอย่างแบบชั้นภูมิ (Stratified Random Sampling) โดยให้กลุ่มสาระการเรียนรู้เนื้อหาการออกข้อสอบ เป็นชั้นภูมิ และให้ค่าความยากของข้อสอบ (b) ที่มีลักษณะเหมือนกัน (Homogenious) เป็นหน่วยการสุ่ม การตรวจสอบประสิทธิภาพของวิธีการจัดคลังข้อสอบ โดยเปรียบเทียบกับวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบถ่วงน้ำหนักที่มีการบังคับ (Cheng et al., 2009) ที่แบ่งคลังข้อสอบออกเป็น 4 ชั้น ตามชั้นของค่าอำนาจจำแนกของข้อสอบแบบคงที่ (a = 0.50, 1.00, 1.50, และ 2.00) และมีการถ่วงน้ำหนักที่มีการบังคับ โดยเปรียบเทียบประสิทธิภาพ เรื่อง 1) ด้านการประมาณค่าความสามารถของผู้สอบ ได้แก่ ค่าความลำเอียงเฉลี่ย และค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย ตามแนวทางของ Yan, von Davier, and Lewis (2016, p. 115) 2) ด้านประสิทธิภาพของการใช้ข้อสอบ ได้แก่ ข้อสอบที่มีการแสดงมากเกินไป ข้อสอบที่มีการใช้น้อยเกินไป อัตราการทับซ้อนของข้อสอบ และการแจกแจงอัตราการแสดงข้อสอบ ตามแนวทางของ Ozturk and Dogan (2015) ในสถานการณ์จริง และนำวิธีการที่พัฒนาขึ้นนี้ไปพัฒนาโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ ที่ใช้วิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา และเปรียบเทียบความสามารถของผู้สอบแต่ละระดับชั้นของค่าอำนาจจำแนกของข้อสอบสำหรับการทดสอบ O-NET วิชาคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 จำแนกตามเพศและความสามารถด้านการเรียนวิชาคณิตศาสตร์

วัตถุประสงค์ของการวิจัย

1. เพื่อพัฒนาวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา สำหรับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ

2. เพื่อเปรียบเทียบประสิทธิภาพของวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหากับวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบถ่วงน้ำหนักที่มีการบังคับ

3. เพื่อพัฒนาโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติที่ใช้วิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา

4. เพื่อเปรียบเทียบความสามารถของผู้สอบแต่ละระดับชั้นของค่าอำนาจจำแนกของข้อสอบ สำหรับการทดสอบ O-NET วิชาคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 จำแนกตามเพศและความสามารถด้านการเรียนวิชาคณิตศาสตร์

กรอบแนวทางการวิจัย

การวิจัยนี้มุ่งเน้นพัฒนาวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา สำหรับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ ใช้ข้อสอบ O-NET วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 6 ระหว่างปีการศึกษา 2551–2553 และปีการศึกษา 2555–2559 ซึ่งเป็นข้อมูลทุติยภูมิ (Secondary Data) ที่ได้มาจากสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) โดยได้นำมาวิเคราะห์ข้อสอบตามแนวคิดของ Bloom et al. (1956); Anderson et al. (2001, pp. 67-68) และ Krathwohl, 2002, pp. 213-215) โดยการปรับลำดับขั้นและคำศัพท์ที่ใช้ในกระบวนการพุทธิปัญญา (The Cognitive Process Dimension) ยังคงมี 6 กระบวนการเหมือนเดิม แต่ 3 กระบวนการแรกเปลี่ยนชื่อเป็น จำ (Remember) เข้าใจ (Understand) และประยุกต์ (Apply) ส่วน 3 กระบวนการหลังเปลี่ยนชื่อที่มีลักษณะเป็นคำนามไปเป็นคำกริยา และสลับกระบวนการที่ 5 กับ 6 และสร้างสรรค์ (Create) เปลี่ยนชื่อมาจาก การสังเคราะห์ (Synthesis) (Anderson et al., 2001, pp. 67-68; Krathwohl, 2002, pp. 213-215) และเปลี่ยนโครงสร้างจากมิติเดียวเป็นสองมิติ Anderson and Krathwohl ได้เพิ่มโครงสร้างในมิติด้านความรู้ (Knowledge Dimension) เข้ามาในโครงสร้างของจุดประสงค์ทางการศึกษาด้านพุทธิปัญญา ทำให้โครงสร้างใหม่มีลักษณะเป็นสองมิติประกอบด้วยมิติด้านกระบวนการพุทธิปัญญาและมิติด้านความรู้ และอธิบายความหมายของมิติด้านความรู้ที่เพิ่มเติมเข้ามาได้ (Krathwohl, 2002, pp. 212-213) โดยมิติด้านความรู้ 4 ส่วน คือ ความรู้เกี่ยวกับข้อเท็จจริง (Factual Knowledge) ความรู้เกี่ยวกับมโนทัศน์ (Conceptual Knowledge) ความรู้เกี่ยวกับวิธีดำเนินการ (Procedural Knowledge) และความรู้เกี่ยวกับอภิปัญญา (Metacognitive Knowledge) และมิติด้านกระบวนการพุทธิปัญญา 6 กระบวนการ คือ จำ (Remember) เข้าใจ (Understand) ประยุกต์ (Apply) วิเคราะห์ (Analyze) ประเมินค่า (Evaluate) และสร้างสรรค์ (Create) แล้วนำมาวิเคราะห์หาคุณภาพของข้อสอบโดยใช้ทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ (MIRT) แบบ 3 พารามิเตอร์ ประกอบด้วยค่าความยากของข้อสอบ ค่าอำนาจจำแนกของข้อสอบ และค่าโอกาสในการเดาของข้อสอบ

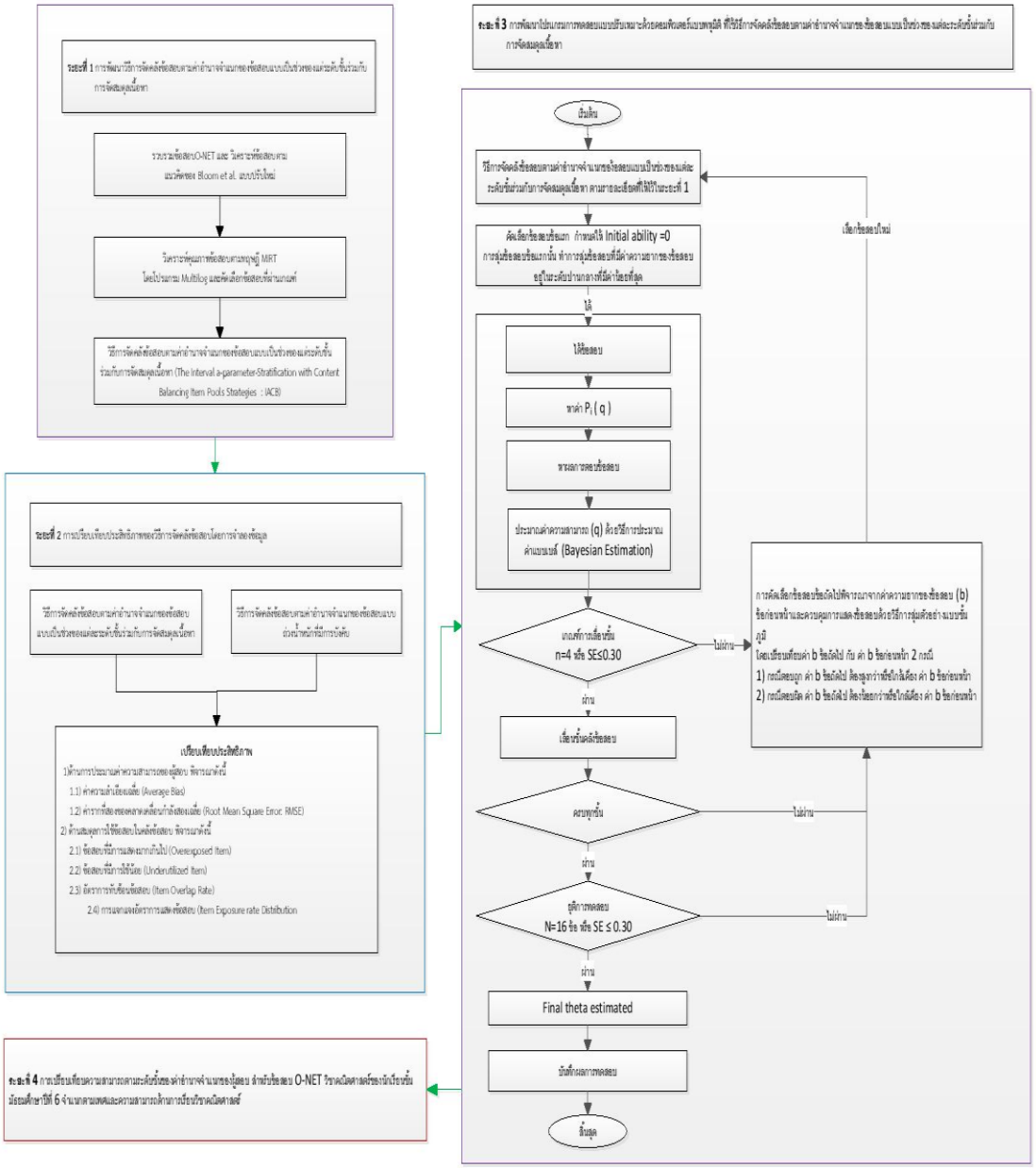
ในส่วนการพัฒนาวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา สำหรับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ ผู้วิจัยประยุกต์จากวิธีจัดคลังข้อสอบตามระดับชั้นของค่าอำนาจจำแนกของข้อสอบ (a-Stratified Method: a-STR) ที่พัฒนาโดย Chang and Ying (1996) ซึ่งแบ่งคลังข้อสอบออกเป็น 4 ชั้น ตามช่วงของค่าอำนาจจำแนกของข้อสอบ 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 ร่วมกับการจัดสมดุลเนื้อหา (Content Balancing) ภายในชั้นของค่าอำนาจจำแนก ได้แบ่งเป็น 4 กลุ่ม ตามสาระการเรียนรู้เนื้อหาการออกข้อสอบ O-NET วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 6 ได้แก่ 1) จำนวนและการดำเนินการ

2) การวัด 3) พีชคณิต และ 4) การวิเคราะห์และความน่าจะเป็น ดังนั้นวิธีการควบคุมการแสดงข้อสอบ (Controlling Item Exposure) จึงใช้วิธีการสุ่มตัวอย่างแบบชั้นภูมิ (Stratified Random Sampling) โดยให้กลุ่มสาระการเรียนรู้เป็นชั้นภูมิ และให้ค่าความยากของข้อสอบ (b) ที่มีลักษณะเหมือนกัน (Homogenious) เป็นหน่วยการสุ่ม

การตรวจสอบประสิทธิภาพของวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหาที่ได้พัฒนาขึ้นโดยเปรียบเทียบกับวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบถ่วงน้ำหนักที่มีการบังคับ โดยเปรียบเทียบประสิทธิภาพ ดังนี้ 1) ด้านการประมาณค่าความสามารถของผู้สอบ ได้แก่ ค่าความลำเอียงเฉลี่ย และค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย 2) ด้านประสิทธิภาพของการใช้ข้อสอบ ได้แก่ ข้อสอบที่มีการแสดงมากเกินไป ข้อสอบที่มีการใช้น้อยเกินไป อัตราการทับซ้อนของข้อสอบ และการแจกแจงอัตราการแสดงข้อสอบ

จากนั้นนำวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา ไปพัฒนาเป็นโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ และนำผลการทดสอบความสามารถของผู้สอบมาเปรียบเทียบความสามารถของผู้สอบแต่ละระดับชั้นของค่าอำนาจจำแนกของข้อสอบ สำหรับการทดสอบ O-NET วิชาคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 จำแนกตามเพศและความสามารถด้านการเรียนวิชาคณิตศาสตร์

ผู้วิจัยได้พัฒนาโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ ตามแนวคิดของ Reckase (2009, pp. 311-338) และ Thompson and Weiss (2011) ซึ่งได้กำหนดขั้นตอนของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ (MCAT) ดังภาพที่ 1-1



ภาพที่ 1-1 กรอบแนวทางการวิจัยเรื่อง การพัฒนาวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา

สมมติฐานของการวิจัย

1. การประมาณค่าความสามารถของผู้สอบ ได้แก่ ค่าความลำเอียงเฉลี่ย และค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย ของวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหามีค่าน้อยกว่าวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบถ่วงน้ำหนักที่มีการบังคับ

2. ประสิทธิภาพของการใช้ข้อสอบในคลังข้อสอบ ได้แก่ ข้อสอบที่มีการแสดงมากเกินไป ข้อสอบที่มีการใช้น้อยเกินไป อัตราการทับซ้อนของข้อสอบ และการแจกแจงอัตราการแสดงข้อสอบของวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหามีค่าน้อยกว่าวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบถ่วงน้ำหนักที่มีการบังคับ

3. ค่าความสามารถของผู้สอบแต่ละระดับชั้นของค่าอำนาจจำแนกของข้อสอบ สำหรับการทดสอบ O-NET วิชาคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 จำแนกตามเพศและความสามารถด้านการเรียนวิชาคณิตศาสตร์มีความแตกต่างกัน

ประโยชน์ที่คาดว่าจะได้รับจากการวิจัย

1. ได้วิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา สำหรับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติที่เหมาะสมกับความสามารถของผู้สอบ

2. ได้โปรแกรมการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติสำหรับการจัดสอบ O-NET วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 6 ที่สามารถนำไปใช้ในการวัดความสามารถของผู้สอบได้อย่างมีประสิทธิภาพ

3. ได้สารสนเทศเกี่ยวกับความสามารถของผู้สอบ ด้านคณิตศาสตร์ โดยการทดสอบด้วยโปรแกรมการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ

4. ได้แนวทางการพัฒนาคลังข้อสอบและโปรแกรมการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ ที่สอดคล้องกับความสามารถของผู้สอบและสามารถนำไปใช้ทดสอบกับการทดสอบอื่นๆ เช่น I-NET N-NET V-NET U-NET และ GAT/PAT

ขอบเขตของการวิจัย

1. ด้านเนื้อหา

การวิจัยนี้เป็นการพัฒนาวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา สำหรับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ โดยใช้ข้อสอบ O-NET วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 6 ระหว่างปีการศึกษา 2551–2553 และปีการศึกษา 2555–2559 ตามสาระการเรียนรู้เนื้อหาการออกข้อสอบ O-NET วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 6 ได้แก่ 1) จำนวนและการดำเนินการ 2) การวัด 3) พีชคณิต และ 4) การวิเคราะห์และความน่าจะเป็น

2. ด้านประชากร

ประชากรที่ใช้ในการวิจัยแบ่งออกเป็น 2 กลุ่ม คือ 1) ประชากรที่ใช้วิเคราะห์ข้อสอบระดับชาติขั้นพื้นฐาน (O-NET) วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 6 ระหว่างปีการศึกษา 2551–2553 และปีการศึกษา 2555–2559 และ 2) ประชากรที่ใช้พัฒนาโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ

1) ประชากรที่ใช้วิเคราะห์ข้อสอบระดับชาติขั้นพื้นฐาน (O-NET) เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 6 ที่ทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน (O-NET) วิชาคณิตศาสตร์ ระหว่างปีการศึกษา 2551–2553 และปีการศึกษา 2555–2559 ซึ่งเป็นข้อมูลทุติยภูมิ ที่ได้มาจากสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน)

2) ประชากรที่ใช้พัฒนาโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 6 โรงเรียนสาธิตมหาวิทยาลัยราชภัฏบ้านสมเด็จเจ้าพระยา ที่กำลังศึกษาอยู่ในภาคเรียนที่ 2 ปีการศึกษา 2560 จำนวน 3 ห้องเรียน รวมทั้งสิ้น 112 คน

3. ตัวแปรที่ใช้ในการวิจัย

3.1 ตัวแปรที่ศึกษาตามวัตถุประสงค์ของการวิจัย ข้อที่ 2

3.1.1 ตัวแปรอิสระ เป็น วิธีการจัดคลังข้อสอบ สำหรับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ 2 วิธี ดังนี้

3.1.1.1 วิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา

3.1.1.2 วิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบถ่วงน้ำหนักที่มีการบังคับ

3.1.2 ตัวแปรตาม เป็น ประสิทธิภาพของวิธีการจัดคลังข้อสอบ ดังนี้

3.1.2.1 ด้านการประมาณค่าความสามารถของผู้สอบ พิจารณาดังนี้

3.1.2.1.1 ค่าความลำเอียงเฉลี่ย (Average Bias) (หน่วยวัดเป็นคะแนนมาตรฐาน)

3.1.2.1.2 ค่ารากที่สองของคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Square Error: RMSE) (หน่วยวัดเป็นคะแนนมาตรฐาน)

3.1.2.2 ด้านประสิทธิภาพของการใช้ข้อสอบ

3.1.2.2.1 ข้อสอบที่มีการแสดงมากเกินไป (Overexposed Item) (หน่วยวัดเป็นจำนวนข้อ)

3.1.2.2.2 ข้อสอบที่มีการใช้น้อยเกินไป (Underutilized Item) (หน่วยวัดเป็นจำนวนข้อ)

3.1.2.2.3 อัตราการทับซ้อนของข้อสอบ (Item Overlap Rate) (หน่วยวัดเป็นจำนวนข้อ)

3.1.2.2.4 การแจกแจงอัตราการแสดงข้อสอบ (Item Exposure rate Distribution (หน่วยวัดเป็นจำนวนข้อ)

3.2 ตัวแปรที่ศึกษาตามวัตถุประสงค์ของการวิจัย ข้อที่ 4

3.2.1 ตัวแปรอิสระ มี 2 ตัวแปร ได้แก่

3.2.1 เพศ จำแนกเป็นเพศชาย และ เพศหญิง

3.2.2 ความสามารถด้านการเรียนวิชาคณิตศาสตร์ จำแนกออกเป็น 2 กลุ่ม ดังนี้

3.2.2.1 กลุ่มสูง (นักเรียนที่มีเกรดเฉลี่ยสะสมวิชาคณิตศาสตร์ตั้งแต่ค่าเฉลี่ย 3.50 ขึ้นไป

3.2.2.2 กลุ่มต่ำ (นักเรียนที่มีเกรดเฉลี่ยสะสมวิชาคณิตศาสตร์น้อยกว่าค่าเฉลี่ย 3.50

3.2.2 ตัวแปรตาม เป็น ความสามารถของผู้สอบแต่ละระดับชั้นของค่าอำนาจจำแนกของข้อสอบ สำหรับการทดสอบ O-NET วิชาคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 (หน่วยวัดเป็นคะแนนมาตรฐาน)

นิยามศัพท์เฉพาะ

การทดสอบแบบปรับเหมาะ (Adaptive Testing) หมายถึง การทดสอบที่ผู้สอบแต่ละคนใช้แบบทดสอบที่ต่างชุดกัน โดยปรับเปลี่ยนไปตามระดับความสามารถของผู้สอบซึ่งมีการคัดเลือกข้อสอบที่มีความยากของข้อสอบเหมาะสมกับระดับความสามารถของผู้สอบ

ทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ (Multidimensional Item Response Theory: MIRT) หมายถึง การขยายทฤษฎีให้สามารถรองรับความสามารถของผู้สอบ (θ) หลายองค์ประกอบ โดยจะถือว่า คุณลักษณะแฝงของบุคคลมีมากกว่า 1 องค์ประกอบ ที่ส่งผลต่อการตอบข้อสอบ ดังนั้นพารามิเตอร์ความสามารถของผู้สอบ จึงมีตั้งแต่ 2 พารามิเตอร์ขึ้นไป ซึ่งทฤษฎีการตอบสนองข้อสอบแบบพหุมิตินี้ เป็นการผ่อนคลายข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบเรื่องความเป็นเอกมิติ

การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ (Multidimensional Computerized Adaptive Testing: MCAT) หมายถึง การผสมผสานแนวคิดของทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ (Multidimensional item response theory: MIRT) กับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ (Computerized Adaptive Testing: CAT) ซึ่งเป็นการทดสอบที่ลดความยาวของข้อสอบและเพิ่มความแม่นยำในการประมาณค่าความสามารถของผู้สอบ

คลังข้อสอบ (Item Pools) หมายถึง แหล่งรวบรวมข้อสอบซึ่งสร้างตามหลักเกณฑ์ที่ถูกต้องและถูกเก็บไว้อย่างเป็นระบบ

วิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา (Item Pool Strategies by using Interval a-parameter Stratification with Content Balancing MCAT method: IACB) หมายถึง วิธีจัดคลังข้อสอบตามระดับชั้นของค่าอำนาจจำแนกของข้อสอบ (a-Stratified Method: a-STR) ซึ่งแบ่งคลังข้อสอบออกเป็น 4 ชั้น ตามช่วงของค่าอำนาจจำแนกข้อสอบ 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 ร่วมกับการจัดสมดุลเนื้อหา (Content Balancing) ภายในชั้นของค่าอำนาจจำแนก ได้แบ่งเป็น 4 กลุ่ม ตามสาระการเรียนรู้เนื้อหาการออกข้อสอบ O-NET วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 6 ได้แก่ 1) จำนวนและการดำเนินการ 2) การวัด 3) พีชคณิต และ 4) การวิเคราะห์และความน่าจะเป็น ดังนั้นวิธีการควบคุมการแสดงข้อสอบ (Controlling Item Exposure) จึงใช้วิธีการสุ่มตัวอย่างแบบชั้นภูมิ (Stratified Random Sampling) โดยให้กลุ่มสาระการเรียนรู้เนื้อหาการออกข้อสอบ เป็นชั้นภูมิ และให้ค่าความยากของข้อสอบ (b) ที่มีลักษณะเหมือนกัน (Homogenious) เป็นหน่วยการสุ่ม

วิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบถ่วงน้ำหนักที่มีการบังคับ (Constraint-weighted a-stratification MCAT method: CWA) หมายถึง การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ที่คัดเลือกแบบทดสอบย่อยโดยการแบ่งคลังข้อสอบออกเป็นชั้นตามลำดับค่าของอำนาจจำแนกของข้อสอบเฉลี่ยของแบบทดสอบย่อย (Subtest) และเลือกแบบทดสอบย่อยจากชั้นที่มีค่าอำนาจจำแนกของข้อสอบต่ำไปหาชั้นที่มีค่าอำนาจจำแนกของข้อสอบสูง การคัดเลือกแบบทดสอบย่อยจะพิจารณาจากผลคูณของค่าสารสนเทศของแบบทดสอบย่อยกับดัชนีลำดับความสำคัญของเงื่อนไขบังคับ วิธีการนี้เป็นแนวคิดของ Cheng et al. (2009)

การแสดงข้อสอบ (Item Exposure) หมายถึง การจัดข้อสอบที่ผ่านการเลือกแล้วให้แก่ผู้สอบ

การควบคุมการแสดงข้อสอบ (Item Exposure Control) หมายถึง การควบคุมจำนวนครั้งของข้อสอบที่จัดให้แก่ผู้สอบตามที่อัตราการแสดงข้อสอบกำหนด

การจัดสมดุลเนื้อหา (Content Balancing) หมายถึง การจัดสมดุลเนื้อหาในการทดสอบให้กลุ่มผู้สอบได้รับชุดแบบทดสอบแตกต่างกัน แต่ได้รับเนื้อหาการทดสอบเดียวกันหรือเป็นคู่ขนานกัน สามารถเปรียบเทียบผลการทดสอบของกลุ่มผู้สอบได้

ค่าความยากของข้อสอบ (Difficulty Parameter: b) หมายถึง ค่าที่แสดงถึงระดับความสามารถของผู้สอบที่จุดเปลี่ยนโค้งคุณลักษณะของข้อสอบมีความชันมากที่สุด มีค่าระหว่าง $-\alpha$ ถึง α ซึ่งค่าความยากของข้อสอบ เท่ากับ -2.50 แสดงว่า ข้อสอบนั้นง่ายมาก และค่าความยากของข้อสอบ เท่ากับ 2.50 แสดงว่า ข้อสอบนั้นยากมาก สำหรับข้อสอบที่คัดเลือกไว้มีค่าความยากของข้อสอบตั้งแต่ -2.50 ถึง 2.50

ค่าอำนาจจำแนกของข้อสอบ (Discrimination Power Parameter: a) หมายถึง ค่าที่เป็นสัดส่วนโดยตรงกับความชันของโค้งคุณลักษณะของข้อสอบ ณ จุดเปลี่ยนโค้งคุณลักษณะของข้อสอบ มีค่าระหว่าง $-\alpha$ ถึง α สำหรับค่าอำนาจจำแนกของข้อสอบที่เป็นลบ แสดงว่า ข้อสอบไม่ดี จำแนกไม่ได้ ต้องตัดข้อสอบข้อนั้นทิ้ง ค่าอำนาจจำแนกของข้อสอบที่เป็นศูนย์ (0) แสดงว่า ข้อสอบไม่มีค่าอำนาจจำแนก และค่าอำนาจจำแนกของข้อสอบที่เป็นบวก แสดงว่า ข้อสอบดี จำแนกได้ สำหรับข้อสอบที่คัดเลือกไว้มีค่าอำนาจจำแนกของข้อสอบ ตั้งแต่ 0.50 ถึง 2.50

ค่าโอกาสในการเดาของข้อสอบ (Guess Parameter: c) หมายถึง ความน่าจะเป็นของผู้สอบที่ไม่มีความสามารถในการตอบข้อสอบนั้น แต่ตอบข้อสอบข้อนั้นได้ถูก หรือค่าที่แสดงโอกาสการตอบข้อสอบถูก โดยไม่มีความรู้ในเรื่องนั้นๆ สำหรับค่าการเดาของข้อสอบมีค่าตั้งแต่ 0 ถึง 1 สำหรับข้อสอบที่คัดเลือกไว้ มีค่าโอกาสในการเดาของข้อสอบ (c) ไม่เกิน 0.3

ความลำเอียงเฉลี่ย (Average Bias) หมายถึง ค่าบอกความเที่ยงตรงของค่าประมาณความสามารถของผู้สอบจำนวนหนึ่ง มีค่าเท่ากับค่าเฉลี่ยของผลต่างระหว่างค่าประมาณความสามารถและค่าความสามารถที่แท้จริงของผู้สอบแต่ละคน ใช้บอกทิศทางการประมาณค่าความสามารถของผู้สอบว่า ให้ผลสูงหรือต่ำกว่าค่าความสามารถที่แท้จริงของผู้สอบมากน้อยเพียงใด

ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Square Error: MSE) หมายถึง ค่าบอกความถูกต้องของค่าประมาณความสามารถของผู้สอบ ซึ่งมีค่าเท่ากับผลต่างกำลังสองเฉลี่ยระหว่างค่าประมาณความสามารถกับค่าความสามารถที่แท้จริงของผู้สอบ ถ้าค่าความคลาดเคลื่อนกำลังสอง

เฉลี่ยมีค่าเข้าใกล้ศูนย์ (0) แสดงว่า ค่าประมาณความสามารถของผู้สอบมีค่าใกล้เคียงกับความสามารถที่แท้จริงของผู้สอบ

ค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Square Error: RMSE) หมายถึง ค่าบอกความถูกต้องของค่าประมาณความสามารถของผู้สอบ ซึ่งมีค่าเท่ากับค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย ถ้าค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ยมีค่าเข้าใกล้ศูนย์ (0) แสดงว่า ค่าประมาณความสามารถของผู้สอบมีค่าใกล้เคียงกับความสามารถที่แท้จริงของผู้สอบ

ประสิทธิภาพของการใช้ข้อสอบ (Item Utilizable Efficiency) หมายถึง ความสามารถของวิธีที่ทำให้ข้อสอบทุกข้อในคลังข้อสอบมีโอกาสนำไปใช้ในการทดสอบ โดยพิจารณาจากข้อสอบที่มีการแสดงมากเกินไป ข้อสอบที่มีการใช้น้อยเกินไป อัตราการทับซ้อนของข้อสอบ และการแจกแจงอัตราการแสดงข้อสอบ

1) ข้อสอบที่มีการแสดงมากเกินไป (Overexposed Item) หมายถึง ข้อสอบใดๆที่มีอัตราการแสดงข้อสอบสูงกว่า 0.2

2) ข้อสอบที่มีการใช้น้อยเกินไป (Underutilized Item) หมายถึง ข้อสอบใดๆ ที่มีอัตราการแสดงข้อสอบเท่ากับหรือน้อยกว่า 0.2

3) อัตราการทับซ้อนของข้อสอบ (Item Overlap Rate) หมายถึง สัดส่วนของข้อสอบที่ใช้ร่วมกันโดยคู่ของผู้สอบต่อข้อสอบที่ใช้โดยคู่ของผู้สอบทั้งหมด

4) การแจกแจงอัตราการแสดงข้อสอบ (Item Exposue Rat Distribution) หมายถึง ลักษณะการกระจายอัตราการแสดงข้อสอบในคลังข้อสอบnบ่งบอกถึงประสิทธิภาพของการใช้ข้อสอบในคลังข้อสอบโดยรวม

ค่าความสามารถของผู้สอบ (Ability of Examiner) หมายถึง ระดับความสามารถของบุคคลนั้น จะทำข้อสอบข้อนั้น ๆ

ได้ถูกต้อง

ความสามารถด้านการเรียนวิชาคณิตศาสตร์ (Mathematics Ability) หมายถึง ผลการเรียนของนักเรียนคำนวณจากเกรดเฉลี่ยสะสมวิชาคณิตศาสตร์ ตั้งแต่ชั้นมัธยมศึกษาปีที่ 4 ภาคเรียนที่ 1 ถึง ชั้นมัธยมศึกษาปีที่ 6 ภาคเรียนที่ 1 รวม 5 ภาคเรียน

บทที่ 2
เอกสารและงานวิจัยที่เกี่ยวข้อง

การวิจัยเรื่องการพัฒนาวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา สำหรับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ นำเสนอแนวคิดหรือทฤษฎีพื้นฐานเกี่ยวกับกระบวนการพุทธิปัญญาของ Bloom et al. (1956) แบบปรับใหม่ (Anderson et al., 2001, pp. 67-68; Krathwohl, 2002, pp. 213-215) ทฤษฎีการตอบสนองข้อสอบ (De Ayala, 2009) ทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ (Reckase, 2009) การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติและงานวิจัยที่เกี่ยวข้อง ซึ่งได้แบ่งการนำเสนอเป็น 4 ตอน ดังนี้

ตอนที่ 1 จุดประสงค์ทางการศึกษาด้านพุทธิปัญญาและการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน

ตอนที่ 2 ทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ

ตอนที่ 3 การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ

ตอนที่ 4 งานวิจัยที่เกี่ยวข้องกับการจัดคลังข้อสอบ การจัดสมดุลเนื้อหา การควบคุมการแสดงข้อสอบและการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ

ตอนที่ 1 จุดประสงค์ทางการศึกษาด้านพุทธิปัญญาและการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐานพื้นฐาน

1. จุดประสงค์ทางการศึกษาด้านพุทธิปัญญา

Anderson et al. (2001, pp. 27-31) ได้ปรับจุดประสงค์ทางการศึกษาใหม่ ตามแนวคิดของ Bloom et al. (1956, p. 18) ซึ่งได้จำแนกจุดประสงค์ทางการศึกษาด้านพุทธิปัญญา (Cognitive Domain) ออกเป็นลำดับขั้นการเรียนรู้ที่เกิดขึ้นในสมอง ซึ่งเป็นการเรียนรู้ทางด้านสติปัญญาเกี่ยวกับความรู้ ความคิด และการแก้ปัญหา โดยแบ่งออกเป็น 6 ขั้น คือ ความรู้ (Knowledge) ความเข้าใจ (Comprehensive) การนำไปใช้ (Application) การวิเคราะห์ (Analysis) การสังเคราะห์ (Synthesis) และการประเมินค่า (Evaluation) ที่ผ่านมาจุดประสงค์ทางการศึกษาด้านพุทธิปัญญาของ Bloom el at. (1956) ได้รับการยอมรับและมีการนำไปใช้อย่างกว้างขวาง (Anderson et al., 2001, p. XXI; Krathwohl, 2002, p. 213) เช่น การนำไปใช้ในขั้นเรียนในฐานะเครื่องมือสำหรับวางแผนการเรียนการสอน การตรวจสอบ ความสอดคล้องของจุดประสงค์การสอน กลยุทธ์การสอน และการประเมินผล ในขณะเดียวกันก็ได้รับข้อเสนอแนะเกี่ยวกับคำที่ใช้ในการจัดหมวดหมู่ที่ไม่คุ้นเคยเนื่องจากเป็นคำที่ใช้ในวงการศึกษา ทำให้ผู้ใช้ไม่สามารถที่จะเข้าใจความหมายของคำเหล่านั้นว่าหมายถึงอะไร (Krathwohl, 2002, p. 213) ต่อมาในช่วง ค.ศ.1990 ถึง 1999 Anderson et al. (2001, p. XXI) ได้ปรับการจำแนกจุดประสงค์ทางการศึกษาด้านพุทธิปัญญาใหม่เพื่อให้สอดคล้องเหมาะสมและเพิ่มความรู้ใหม่เข้าไปให้เกิดความสมบูรณ์ยิ่งขึ้น ในปี ค.ศ. 2001 ได้นำเสนอแนวคิดไว้ในหนังสือเรื่อง “A Taxonomy for Learning,

Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives” โดยปรับเปลี่ยนจุดประสงค์ทางการศึกษาด้านพุทธิปัญญาเป็น 2 ประเด็น คือ ปรับขั้นตอนและคำศัพท์ในกระบวนการพุทธิปัญญา และเพิ่มโครงสร้างจากมิติเดียวเป็นสองมิติ (Krathwohl, 2002, pp. 213-217) ดังนี้

1. การปรับลำดับขั้นและคำศัพท์ที่ใช้ในกระบวนการพุทธิปัญญา ยังคงมี 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, pp. 67-68; Krathwohl, 2002, pp. 213-215) ดังตารางที่ 2-1

ตารางที่ 2-1 กระบวนการพุทธิปัญญาของ Bloom et al. (1956) แบบปรับใหม่

กระบวนการและคำศัพท์เดิม	กระบวนการและคำศัพท์ใหม่
1. ความรู้ (Knowledge)	1. จำ (Remember)
2. ความเข้าใจ (Comprehension)	2. เข้าใจ (Understand)
3. การนำไปใช้ (Application)	3. ประยุกต์ (Apply)
4. การวิเคราะห์ (Analysis)	4. วิเคราะห์ (Analyze)
5. การสังเคราะห์ (Synthesis)	5. ประเมินค่า (Evaluate)
6. การประเมินค่า (Evaluation)	6. สร้างสรรค์ (Create)

จากตารางที่ 2-1 กระบวนการและคำศัพท์ใหม่อธิบายได้ดังนี้

1.1 จำ (Remember) หมายถึง ความสามารถในการดึงเอาความรู้ที่มีอยู่ในหน่วยความจำระยะยาวออกมา แบ่งประเภทย่อยได้ 2 ลักษณะ คือ การจำได้ (Recognizing) และการระลึกได้ (Recalling)

1.2 เข้าใจ (Understand) หมายถึง ความสามารถในการกำหนดความหมายของคำพูด ตัวอักษร และการสื่อสารจากสื่อต่างๆ ที่เป็นผลมาจากการสอน แบ่งประเภทย่อยได้ 7 ลักษณะ คือ การตีความ (Interpreting) การยกตัวอย่าง (Exemplifying) การจำแนกประเภท (Classifying) การสรุป (Summarizing) การอนุมาน (Inferring) การเปรียบเทียบ (Comparing) และการอธิบาย (Explaining)

1.3 ประยุกต์ (Apply) หมายถึง ความสามารถในการดำเนินการหรือใช้ระเบียบวิธีการภายใต้สถานการณ์ที่กำหนดให้ แบ่งประเภทย่อยได้ 2 ลักษณะ คือ การดำเนินงาน (Executing) และการใช้เป็นเครื่องมือ (Implementing)

1.4 วิเคราะห์ (Analyze) หมายถึง ความสามารถในการแยกส่วนประกอบของสิ่งต่างๆ และค้นหาความสัมพันธ์ระหว่างส่วนประกอบ ความสัมพันธ์ระหว่างของส่วนประกอบกับโครงสร้าง

รวมหรือส่วนประกอบเฉพาะ แบ่งประเภทย่อยได้ 3 ลักษณะ คือ การบอกความแตกต่าง (Differentiating) การจัดโครงสร้าง (Organizing) และการระบุคุณลักษณะ (Attributing)

1.5 ประเมินค่า (Evaluate) หมายถึง ความสามารถในการตัดสินใจโดยอาศัยเกณฑ์หรือมาตรฐาน แบ่งประเภทย่อยได้ 2 ลักษณะ คือ การตรวจสอบ (Checking) และการวิพากษ์วิจารณ์ (Critique)

1.6 สร้างสรรค์ (Create) หมายถึง ความสามารถในการรวมส่วนประกอบต่างๆ เข้าด้วยกันด้วยรูปแบบใหม่ๆ ที่มีความเชื่อมโยงกันอย่างมีเหตุผล หรือทำให้ได้ผลิตภัณฑ์ที่เป็นต้นแบบ แบ่งประเภทย่อยได้ 3 ลักษณะ คือ การสร้าง (Generating) การวางแผน (Planning) และผลิต (Producing)

2. เปลี่ยนโครงสร้างจากมิติเดียวเป็นสองมิติ Anderson and Krathwohl ได้เพิ่มโครงสร้างในมิติด้านความรู้ (Knowledge Dimension) เข้ามาในโครงสร้างของจุดประสงค์ทางการศึกษา ด้านพุทธิปัญญา ทำให้โครงสร้างใหม่มีลักษณะเป็นสองมิติประกอบด้วยมิติด้านกระบวนการพุทธิปัญญาและมิติด้านความรู้ และอธิบายความหมายของมิติด้านความรู้ที่เพิ่มเติมเข้ามาได้ (Krathwohl, 2002, pp. 212-213) ดังตารางที่ 2-2

ตารางที่ 2-2 โครงสร้างสองมิติของจุดประสงค์ทางการศึกษาของ Bloom et al. (1956) แบบปรับใหม่

มิติด้านความรู้	มิติด้านกระบวนการ					
	จำ	เข้าใจ	ประยุกต์	วิเคราะห์	ประเมินค่า	สร้างสรรค์
ความรู้เกี่ยวกับข้อเท็จจริง						
ความรู้เกี่ยวกับมโนทัศน์						
ความรู้เกี่ยวกับวิธีดำเนินการ						
ความรู้เกี่ยวกับอภิปัญญา						

จากตารางที่ 2-2 สามารถอธิบายได้ดังนี้

2.1 ความรู้เกี่ยวกับข้อเท็จจริง (Factual Knowledge) หมายถึง ส่วนประกอบพื้นฐานที่นักเรียนจะต้องรู้เพื่อความเข้าใจเกี่ยวกับสาขาวิชาที่เรียน หรือใช้ในการแก้ปัญหาที่เกี่ยวข้อง แบ่งประเภทย่อยได้ 2 ลักษณะ คือ ความรู้เกี่ยวกับศัพท์เฉพาะ (Knowledge of Terminology) และความรู้เฉพาะเกี่ยวกับรายละเอียดหรือส่วนประกอบ (Knowledge of Specific Details and Elements)

2.2 ความรู้เกี่ยวกับมโนทัศน์ (Conceptual Knowledge) หมายถึง ปฏิสัมพันธ์ระหว่างองค์ประกอบพื้นฐานภายใต้โครงสร้างขนาดใหญ่ที่มีการทำงานร่วมกัน แบ่งประเภทย่อยได้ 3 ลักษณะ คือ ความรู้เกี่ยวกับการแบ่งชั้นและจำแนกประเภท (Knowledge of Classifications and Categories) ความรู้เกี่ยวกับหลักการและการสรุปอ้างอิง (Knowledge of Principles and

Generalizations) และความรู้เกี่ยวกับทฤษฎี โมเดล และโครงสร้าง (Knowledge of Theories, Models, and Structures)

2.3 ความรู้เกี่ยวกับวิธีดำเนินการ (Procedural Knowledge) หมายถึง ความรู้เกี่ยวกับวิธีการทำงาน เช่น วิธีการหาความรู้ด้วยการสืบสวนสอบสวน และหลักเกณฑ์ในการใช้ทักษะ ขั้นตอน เทคนิค และวิธีการ แบ่งประเภทย่อยได้ 3 ลักษณะ คือ ความรู้เกี่ยวกับทักษะและขั้นตอนในเนื้อหาเฉพาะ (Knowledge of Subject Specific Skills and Algorithms) ความรู้เกี่ยวกับเทคนิคและวิธีการในเนื้อหาเฉพาะ (Knowledge of Subject-Specifictechniques and Methods) และความรู้เกี่ยวกับหลักเกณฑ์ที่ใช้เป็นแนวทางในกำหนดระเบียบวิธีการทำงานที่เหมาะสม (Knowledge of Criteria for Determining When to Use Appropriate Procedures)

2.4 ความรู้เกี่ยวกับอภิปัญญา (Metacognitive Knowledge) หมายถึง ความรู้เกี่ยวกับความรู้ความเข้าใจทั่วไป มีความตระหนัก และมีความรู้เกี่ยวกับการคิดของตนเอง แบ่งประเภทย่อยได้ 3 ลักษณะ คือ ความรู้เกี่ยวกับยุทธวิธี (Strategic Knowledge) ความรู้เกี่ยวกับความรู้ความเข้าใจในงาน ประกอบด้วยสถานการณ์ที่เหมาะสมในการทำงาน และรู้เงื่อนไขในการทำงาน (Knowledge About Cognitive Tasks, Including Appropriate Contextual and Conditional Knowledge) และความรู้เกี่ยวกับตนเอง (Self-knowledge)

จากแนวคิดจุดประสงค์ทางการศึกษาด้านพุทธิปัญญาข้างต้น สรุปได้ว่า จุดประสงค์ทางการศึกษาด้านพุทธิปัญญา ตามแนวคิดของ Bloom et al. (1956) ซึ่ง Anderson et al. (2001, pp. 67-68) และ Krathwohl (2002, pp. 213-215) ได้เพิ่มโครงสร้างมิติเดียวเป็นสองมิติ ทำให้โครงสร้างใหม่มีลักษณะเป็นสองมิติ ประกอบด้วยมิติด้านกระบวนการพุทธิปัญญา มี 6 กระบวนการ คือ จำ เข้าใจ ประยุกต์ วิเคราะห์ ประเมินค่า และสร้างสรรค์ และมิติด้านความรู้ มี 4 ส่วน คือ ความรู้เกี่ยวกับข้อเท็จจริง มโนทัศน์ วิธีดำเนินการ และอภิปัญญา ที่เป็นกระบวนการเรียนรู้ที่เกิดขึ้นในสมองของผู้เรียน

2. สถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน)

พระราชบัญญัติการศึกษาแห่งชาติ พ.ศ. 2542 แก้ไขเพิ่มเติม (ฉบับที่ 2) พ.ศ. 2545 กำหนดให้การจัดการศึกษาต้องเป็นไปเพื่อพัฒนาคนไทยให้เป็นมนุษย์ที่สมบูรณ์ทั้งร่างกาย จิตใจ สติปัญญา ความรู้และคุณธรรม มีจริยธรรมและวัฒนธรรมในการดำรงชีวิต สามารถอยู่ร่วมกับผู้อื่นได้อย่างมีความสุข โดยการจัดระบบ โครงสร้าง และกระบวนการจัดการศึกษา ยึดหลักมีเอกภาพด้านนโยบาย และมีความหลากหลายในการปฏิบัติ มีการกระจายอำนาจที่มีประสิทธิภาพ และมีการทดสอบที่ได้มาตรฐานไปสู่ระดับชาติ สถานศึกษาและองค์กรปกครองส่วนท้องถิ่น มีการกำหนดมาตรฐานการศึกษา และจัดระบบประกันคุณภาพการศึกษาทุกระดับและทุกประเภทการศึกษา

ดังนั้นจึงจำเป็นที่รัฐต้องตรวจสอบประเมินผลว่าการจัดการศึกษาที่เป็นการให้บริการสาธารณะแก่ประชาชนที่รัฐจะต้องจัดให้ทั่วถึงนั้น มีคุณภาพตามมาตรฐานการศึกษาที่กำหนดและมีคุณภาพเท่าเทียมกันมากน้อยเพียงใด จึงได้จัดตั้งสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) ขึ้น

สถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) เรียกโดยย่อว่า "สทศ." ใช้ชื่อเป็นภาษาอังกฤษว่า "National Institute of Educational Testing Service (Public Organization)"

เรียกโดยย่อว่า "NIETS" จัดตั้งขึ้นเมื่อวันที่ 3 กันยายน พ.ศ. 2548 ตามพระราชกฤษฎีกาจัดตั้งสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) พ.ศ. 2548 ภายใต้การกำกับดูแลของรัฐมนตรีว่าการกระทรวงศึกษาธิการ สถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) จัดตั้งขึ้นเป็นองค์การมหาชนเพื่อให้การทำงานมีประสิทธิภาพสามารถใช้ประโยชน์ทรัพยากรและบุคลากรได้สูงสุด มีความเป็นอิสระ ไม่ขึ้นอยู่กับสายการบริหารของหน่วยงานที่มีหน้าที่ความรับผิดชอบการจัดการศึกษา จึงมีความเป็นกลาง เป็นสถาบันที่มีการกำหนดหลักการ นโยบาย มาตรการและเป้าหมาย โครงสร้างการบริหาร และการดำเนินกิจการ ความสัมพันธ์กับรัฐและหน่วยงานที่เกี่ยวข้อง บุคลากร การเงิน การตรวจสอบ และการประเมินผลที่เป็นระบบและมีประสิทธิภาพ ให้บรรลุตามวัตถุประสงค์การจัดตั้งสถาบันเพื่อบริหารจัดการและดำเนินการเกี่ยวกับการศึกษา วิจัย และให้บริการ ทางด้านการประเมินผลทางการศึกษาและทดสอบทางการศึกษาอย่างต่อเนื่อง รวมทั้งเป็นศูนย์กลางความร่วมมือด้านการทดสอบทางการศึกษาในระดับชาติและนานาชาติ

3. ผลการทดสอบ O-NET วิชาคณิตศาสตร์ ปีการศึกษา 2558

การทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน (O-NET) ชั้นมัธยมศึกษาปีที่ 6 สอบวันที่ 6-7 กุมภาพันธ์ พ.ศ. 2559 ประกาศผลวันที่ 21 มีนาคม พ.ศ. 2559 จำนวนผู้เข้าสอบ 440,592 คน บริหารการจัดสอบตามมาตรฐานการทดสอบทางการศึกษาแห่งชาติ โดยศูนย์สอบซึ่งประกอบด้วยมหาวิทยาลัยของรัฐ รวมทั้งหมดจำนวน 18 ศูนย์ จำนวนสนามสอบ 413 สนาม จำนวนห้องสอบ 14,955 ห้อง และจำนวนโรงเรียน 3,876 โรง การจัดสอบเป็นไปด้วยความเรียบร้อย ประกาศเลขที่นั่งสอบตามรายชื่อผู้มีสิทธิ์สอบที่สถานศึกษานำส่งผ่านระบบได้ตามกำหนด และการประกาศผลมีความเรียบร้อยเป็นไปตามแผนที่กำหนด เป็นไปตามมาตรฐานการทดสอบทางการศึกษาแห่งชาติด้านมาตรการบริหารการทดสอบ ไม่พบปัญหาด้านคุณภาพในการจัดสอบ เช่น การร้องเรียนเกี่ยวกับการจัดสอบ ข้อสอบรั่วหรือทุจริตในการสอบ และนักเรียนทุกคนได้รับการทดสอบที่สถานศึกษาส่งรายชื่อเข้าระบบ O-NET โดยผลการทดสอบสอบแสดงดังตารางที่ 2-3 (สถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน), รายงานประจำปี 2559, หน้า 26)

ตารางที่ 2-3 ค่าสถิติพื้นฐานผลการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน (O-NET)
ชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2558 ระดับประเทศ (คะแนนเต็ม 100 คะแนน)

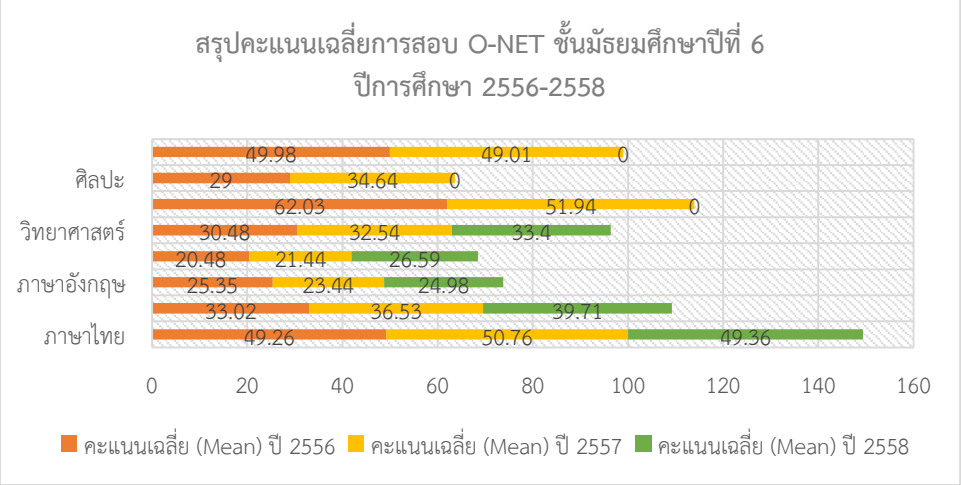
วิชา	จำนวน ผู้เข้าสอบ (N)	คะแนนเฉลี่ย (Mean)	ส่วนเบี่ยงเบนมาตรฐาน (SD)	มัธยฐาน (Median)	ฐานนิยม (Mode)
ภาษาไทย	422,625	49.36	15.57	49.50	49.00
สังคมศึกษา	423,519	39.71	8.30	39.00	38.00
ภาษาอังกฤษ	423,417	24.98	12.46	22.00	19.00
คณิตศาสตร์	423,654	26.59	14.79	22.50	20.00
วิทยาศาสตร์	422,718	33.40	8.43	32.00	31.00

หมายเหตุ: จำนวนผู้เข้าสอบ (N) ไม่รวมเด็กพิเศษ ได้แก่ นักเรียนที่มีความบกพร่องทางด้านต่าง ๆ และพิการซ้ำซ้อน

จากตารางที่ 2-3 ผลการทดสอบ O-NET วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 6 จากคะแนนเต็ม 100 คะแนน ปรากฏว่า วิชาคณิตศาสตร์มีจำนวนผู้เข้าสอบ (N) 423,654 คน มีคะแนนเฉลี่ย(Mean) เท่ากับ 26.59 ส่วนเบี่ยงเบนมาตรฐาน (SD) เท่ากับ 14.79 ค่ามัธยฐาน (Median) เท่ากับ 22.50 และฐานนิยม (Mode) เท่ากับ 20.00 แสดงให้เห็นว่า นักเรียนชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2558 มีคะแนนต่ำกว่าร้อยละ 50 ของคะแนนเต็ม แสดงดังตารางที่ 2-4 และภาพที่ 2-1

ตารางที่ 2-4 การเปรียบเทียบคะแนนเฉลี่ยการสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2556 - 2558

วิชา	คะแนนเฉลี่ย		
	ปีการศึกษา 2556	ปีการศึกษา2557	ปีการศึกษา2558
ภาษาไทย	49.26	50.76	49.36
สังคมศึกษา	33.02	36.53	39.71
ภาษาอังกฤษ	25.35	23.44	24.98
คณิตศาสตร์	20.48	21.44	26.59
วิทยาศาสตร์	30.48	32.54	33.40
สุขศึกษา	62.03	51.94	-
ศิลปะ	29.00	34.64	-
การงานอาชีพ	49.98	49.01	-

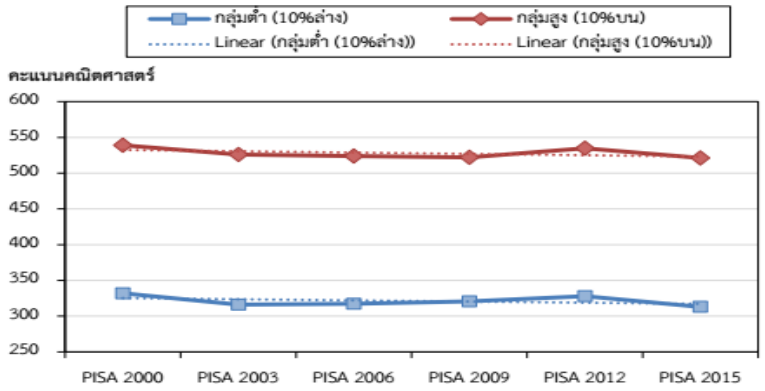


ภาพที่ 2-1 การเปรียบเทียบคะแนนเฉลี่ยการสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2556 - 2558

จากตารางที่ 2-4 และภาพที่ 2-1 เปรียบเทียบคะแนนเฉลี่ยการสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2556 - 2558 จากคะแนนเต็ม 100 คะแนน ปรากฏว่า ผลการทดสอบ O-NET วิชาคณิตศาสตร์ ปีการศึกษา 2558 มีคะแนนสูงสุด (Mean=26.59) รองลงมาคือ ผลการ

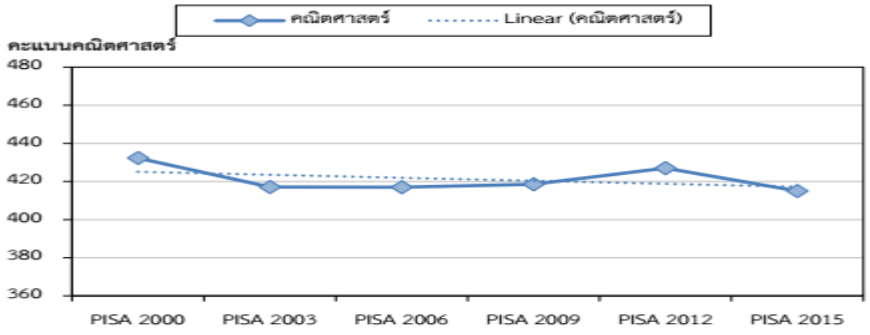
ทดสอบ O-NET วิชาคณิตศาสตร์ ปีการศึกษา 2557 (Mean=21.44) และ ผลการทดสอบ O-NET วิชาคณิตศาสตร์ ปีการศึกษา 2556 (Mean=20.48) ตามลำดับ แสดงให้เห็นว่า นักเรียนชั้นมัธยมศึกษาปีที่ 6 มีคะแนนต่ำกว่าร้อยละ 50 ของคะแนนเต็ม แต่มีแนวโน้มเพิ่มขึ้นถ้าพิจารณาจากคะแนน 3 ปีย้อนหลัง

จากรายงานประจำปี 2559 ของสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) (สทศ.) ผลการทดสอบ O-NET วิชาคณิตศาสตร์ ปีการศึกษา 2558 คะแนนเฉลี่ย (Mean) เท่ากับ 26.59 ซึ่งมีคะแนนต่ำกว่าร้อยละ 50 ของคะแนนเต็ม สอดคล้องกับการสรุปผลการวิจัย PISA 2015 ของสถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี (สสวท.) กระทรวงศึกษาธิการ ร่วมกับ ORGANISATION for ECONOMIC CO-OPERATION and DEVELOPMENT (OECD) (OECD, 2016, pp. 14-18) ปรากฏว่า ผลการประเมินการรู้เรื่องคณิตศาสตร์ คะแนนเฉลี่ย OECD ของคณิตศาสตร์ ใน PISA 2015 คะแนนมาตรฐาน เท่ากับ 490 คะแนน ซึ่งคะแนนเฉลี่ยคณิตศาสตร์ของนักเรียนไทยคือ 415 คะแนนอยู่ในช่วงลำดับที่ 49-55 ซึ่งต่ำกว่าค่าเฉลี่ย OECD มากกว่าหนึ่งระดับ และมีคะแนนคณิตศาสตร์อยู่ในกลุ่มเดียวกับประเทศชิลี ตุรกี มอลโดวา อุรุกวัย มอนเตเนโกร ตรินิแดด โตเบโก และแอลเบเนีย ประเทศในเอเชียที่ร่วมการประเมินและมีคะแนนต่ำกว่าไทยมีเพียงอินโดนีเซีย ทั้งนี้ นักเรียนไทยกลุ่มสูง (กลุ่มที่มีคะแนนอยู่ที่ 10% บน) มีคะแนนคณิตศาสตร์ 521 คะแนน กับ นักเรียนไทยกลุ่มต่ำ (กลุ่มที่มีคะแนนอยู่ที่ 10% ล่าง) มีคะแนนคณิตศาสตร์ 313 คะแนน แสดงได้ดังภาพที่ 2-2



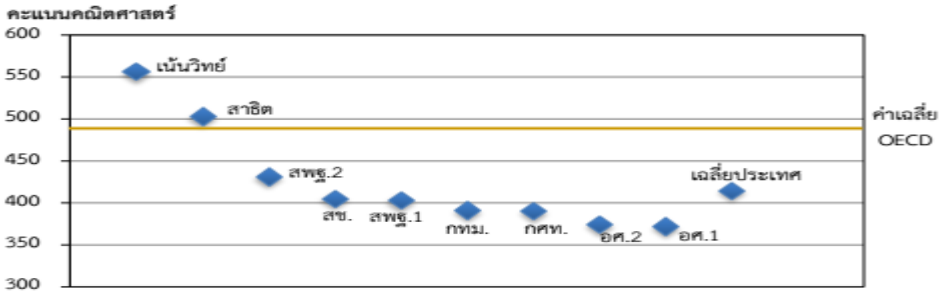
ภาพที่ 2-2 แนวโน้มคะแนนวิชาคณิตศาสตร์ของนักเรียนไทยกลุ่มสูงและกลุ่มต่ำจาก PISA 2000 - 2015 (OECD, 2016, p. 14)

ถ้าพิจารณาแนวโน้มคะแนนคณิตศาสตร์ของนักเรียนไทย เมื่อเทียบกับการประเมินคณิตศาสตร์ที่เป็นวิชาหลักใน PISA 2012 ลดลง 11 คะแนน และใน PISA 2003 ลดลง 2 คะแนน แต่ลดลงอย่างไม่มีนัยสำคัญและคะแนนใกล้เคียงกับการประเมินใน PISA 2006 และ PISA 2009 คะแนน แสดงได้ดังภาพที่ 2-3



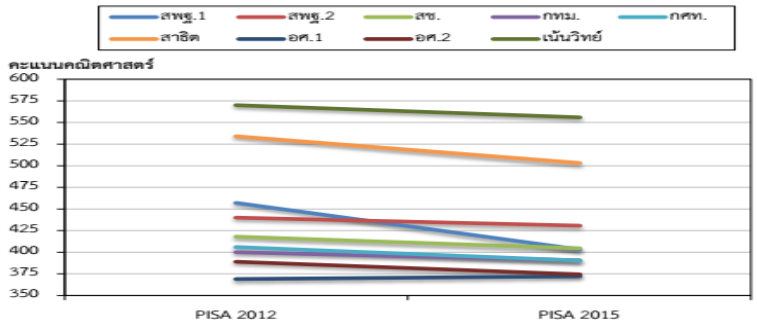
ภาพที่ 2-3 แนวโน้มคะแนนวิชาคณิตศาสตร์จาก PISA 2000 – 2015 (OECD, 2016, p. 15)

แนวโน้มการรู้เรื่องคณิตศาสตร์ของนักเรียนไทยต่างกลุ่มโรงเรียน ปรากฏว่า กลุ่มโรงเรียนที่นักเรียนมีคะแนนคณิตศาสตร์สูงกว่าค่าเฉลี่ย OECD คือนักเรียนจากกลุ่มโรงเรียนเน้นวิทย์ และสาธิต ส่วนกลุ่มอื่น ๆ มีคะแนนต่ำกว่าค่าเฉลี่ย OECD แสดงได้ดังภาพที่ 2-4



ภาพที่ 2-4 คะแนนวิชาคณิตศาสตร์ของนักเรียนตามกลุ่มโรงเรียนใน PISA 2015 (OECD, 2016, p. 15)

เมื่อเปรียบเทียบผลการประเมินคณิตศาสตร์ใน PISA 2012 กับ PISA 2015 ปรากฏว่า เกือบทุกกลุ่มโรงเรียนมีคะแนนคณิตศาสตร์ลดลง แสดงได้ดังภาพที่ 2-5



ภาพที่ 2-5 การเปลี่ยนแปลงคะแนนวิชาคณิตศาสตร์ของนักเรียนตามกลุ่มโรงเรียนใน PISA 2012 และ PISA 2015 (OECD, 2016, p. 16)

ความแตกต่างระหว่างเพศ กับ คะแนนวิชาคณิตศาสตร์ ปรากฏว่า ในภาพรวมระดับนานาชาติ ใน PISA 2015 นักเรียนชายมีคะแนนสูงกว่านักเรียนหญิงอย่างมีนัยสำคัญทางสถิติ (8 คะแนน) สำหรับประเทศไทย นักเรียนชายและนักเรียนหญิงมีคะแนนคณิตศาสตร์ใกล้เคียงกัน โดยนักเรียนหญิงมีคะแนนสูงกว่านักเรียนชาย 3 คะแนน (ชาย 414 คะแนน และหญิง 417 คะแนน) แต่ความแตกต่างนี้ไม่มีนัยสำคัญทางสถิติ ทั้งนี้ความแตกต่างระหว่างเพศของไทยเริ่มมีช่องว่างแคบลงเมื่อเทียบกับการประเมินในรอบที่ผ่านๆ มา ซึ่ง PISA 2012 นักเรียนหญิงมีคะแนนสูงกว่าถึง 14 คะแนน (ชาย 419 คะแนน และหญิง 433 คะแนน) โดยนักเรียนหญิงมีคะแนนลดลงจาก PISA 2012 ถึง 16 คะแนน

ความสัมพันธ์ระหว่างเกรดวิชาคณิตศาสตร์กับความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ มีงานวิจัยของวรรณศิรินทร์ ภูฉายา ณัฐกรานต์ สุระภา ชนิษฐา คลังใหญ่ นิตยา จันตะคุณ และอรุณี จันทร์ศิลา (2561) ได้ศึกษาระดับความสามารถในการแก้ปัญหาทางคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 1-3 โดยใช้หลักการตรรกศาสตร์คลุมเครือ การวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาระดับความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ และปัจจัยที่มีความสัมพันธ์ต่อความสามารถในการแก้ปัญหาคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 1 - 3 จำนวน 119 คน ซึ่งอยู่ในจังหวัดมหาสารคาม ประจำปีการศึกษา 2558 เครื่องมือที่ใช้ในการวิจัยคือ แบบทดสอบและแบบสอบถาม การวิเคราะห์ข้อมูลใช้หลักการตรรกศาสตร์คลุมเครือ สหสัมพันธ์และไค-สแควร์ ผลการวิจัยปรากฏว่า ระดับความสามารถในการแก้ปัญหาทางคณิตศาสตร์นักเรียนส่วนใหญ่ ชั้นมัธยมศึกษาปีที่ 1 อยู่ในระดับต่ำ คิดเป็นร้อยละ 47.22 นักเรียนชั้นมัธยมศึกษาปีที่ 2 อยู่ในระดับปานกลาง คิดเป็นร้อยละ 35.60 และนักเรียนชั้นมัธยมศึกษาปีที่ 3 ในระดับต่ำ คิดเป็นร้อยละ 36.84 และปัจจัยที่มีความสัมพันธ์ต่อความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ คือ เกรดวิชาคณิตศาสตร์ บรรยากาศในชั้นเรียน เจตคติต่อวิชาคณิตศาสตร์ พฤติกรรมการสอนของครูคณิตศาสตร์ ส่วนปัจจัยด้านเพศไม่มีความสัมพันธ์ต่อความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์

จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องสามารถสรุปได้ว่า ตัวแปรเพศ ไม่มีความสัมพันธ์ต่อการรู้เรียนคณิตศาสตร์ ซึ่งส่งผลทำให้ผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์แตกต่างกันอย่างมีนัยสำคัญทางสถิติ

ตอนที่ 2 ทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ

1. ทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ

แนวคิดของทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ (Multidimensional Item Response Theory Model: MIRT Model) ถือได้ว่าเป็นแนวคิดที่ขยายมาจากทฤษฎีการตอบสนองข้อสอบแบบเอกมิติ (Unidimensional Item Response Theory Model: UIRT Model) เพื่อแก้ไขจุดอ่อนของข้อตกลงเบื้องต้นเกี่ยวกับความเป็นเอกมิติ ซึ่งทฤษฎีการตอบสนองข้อสอบแบบพหุมิติจากแนวคิดพื้นฐานที่สำคัญ 2 กลุ่มแนวคิด โดยกลุ่มแรกเป็นการศึกษาโดย Spearman (1927) และ Thurstone (1947) พบว่า โมเดลการตอบสนองข้อสอบแบบพหุมิติมีที่มาจากการวิเคราะห์องค์ประกอบ (Factor Analysis: FA) ลักษณะการวิเคราะห์องค์ประกอบมีความสัมพันธ์โดยตรงกับการวิเคราะห์ทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ และกลุ่มที่สองเป็นการศึกษาของ Lazarsfeld

(1950) กล่าวว่า โมเดลการตอบสนองข้อสอบแบบพหุมิติได้รับอิทธิพลมาจากทฤษฎีการตอบสนองข้อสอบแบบเอกมิติ เช่นเดียวกับการวิเคราะห์องค์ประกอบ แต่แนวคิดเกี่ยวกับโมเดลการตอบสนองข้อสอบแบบเอกมิติ ในช่วงแรกยังไม่เป็นที่นิยมเหมือนเช่นปัจจุบัน จากสองแนวคิด โมเดลการตอบสนองข้อสอบแบบพหุมิติเป็นการวิเคราะห์เชิงสถิติหลายตัวแปร โดยเฉพาะการวิเคราะห์องค์ประกอบหรือโมเดลสมการเชิงโครงสร้าง (Structural Equation Modelling) หรืออาจจะเป็นแนวคิดที่ขยายมาจากทฤษฎีตอบสนองข้อสอบแบบมิติเดียว (Reckase, 2009, p. 63) ดังนี้

1.1 การวิเคราะห์องค์ประกอบ (Factor Analysis)

จากการศึกษาของ Reckase (2009, p. 63) กล่าวว่า นักวิชาการหลายคนและเห็นตรงกันว่า การวิเคราะห์องค์ประกอบ และทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ เป็นระเบียบวิธีทางด้านตัวเลขที่พยายามจะระบุสเกลสมมติฐานเพื่อถอดแบบข้อมูลอันจะนำไปสู่การวิเคราะห์ข้อมูล โดยสเกลการวิเคราะห์องค์ประกอบจะต้องกำหนดจุดกำเนิด (Origin) และหน่วยของการวัด (Unit of Measurement) เพื่อพิจารณาถึงความแตกต่างของลักษณะตัวแปรตั้งต้น เช่น ความแตกต่างของค่าเฉลี่ย (Mean) และส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation) เป็นต้น รวมถึงความเที่ยง (Reliability) ของการวิเคราะห์ข้อมูล เช่น การกำจัดตัวแปรแทรกซ้อนหรือตัวแปรที่ไม่จำเป็นออกจากโมเดลการวิเคราะห์ด้วยการปรับค่าสถิติให้เป็นมาตรฐานโดยจะให้ความสำคัญกับเมตริกซ์ความสัมพันธ์ แต่ไม่ให้ความสำคัญกับความแตกต่างของค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของตัวแปรที่ศึกษา (Herman, 1976 Cited in Reckase, 2009, pp. 63-64) หลักการเบื้องต้นของการวิเคราะห์องค์ประกอบเป็นกระบวนการจัดกระทำกับข้อมูลหรือตัวแปรเพื่อจัดกลุ่มตัวแปรในการจำแนกประเภทหรือองค์ประกอบให้ได้จำนวนน้อยที่สุด วิธีดังกล่าวสามารถทำได้โดยการศึกษาความสัมพันธ์ระหว่างตัวแปรซึ่งเมตริกซ์ความสัมพันธ์ถือเป็นข้อมูลหลักในการวิเคราะห์โดยไม่ได้สนใจกับค่าพารามิเตอร์ของข้อสอบที่เป็นคุณลักษณะที่มีความสัมพันธ์กับข้อสอบหรือตัวแปรตั้งต้นโดยตรง เช่น ค่าอำนาจจำแนกของข้อสอบ ค่าความยากของข้อสอบ และค่าโอกาสในการเดาของข้อสอบ แต่จะพิจารณาตัวแปรแทรกซ้อนอื่นๆ ที่จะตัดออกจากโมเดลการวิเคราะห์เพื่อความถูกต้องมากขึ้น (Carroll, 1945 cited in Reckase, 2009, p. 64)

ในปี ค.ศ. 1981 Bock and Aitkin (1981 Cited in Reckase, 2009, pp. 67-68) ได้นำเสนอแนวคิดที่คล้ายกันระหว่างการวิเคราะห์องค์ประกอบกับทฤษฎีการตอบสนองข้อสอบ และผลที่ได้จากการวิเคราะห์โมเดลการตอบสนองข้อสอบแบบพหุมิติ โดยระบุโมเดลปกติสะสม (Normal Ogive Model) สำหรับการวัดคุณลักษณะพหุมิติ รวมทั้งคุณลักษณะของข้อสอบในโมเดลของการวิเคราะห์องค์ประกอบและทฤษฎีการตอบสนองข้อสอบ ซึ่งส่วนใหญ่จะนำเสนอค่าอำนาจจำแนกของข้อสอบและค่าความยากของข้อสอบเพื่อเป็นตัวแทนของค่าจุดแทนแกน (Intercept) และความชัน (Slopes) แต่ยังขาดในส่วนของโมเดลเต็มรูปแบบโมเดลการตอบสนองข้อสอบแบบพหุมิติใช้ความหมายค่าพารามิเตอร์ข้อสอบว่าเป็นปฏิสัมพันธ์ระหว่างผู้สอบกับข้อสอบ การตีความหมายยังเป็นลักษณะของการวิเคราะห์องค์ประกอบ นั่นคือ พิจารณาในลักษณะขององค์ประกอบ ซึ่ง Bock and Aitkin ได้ขยายแนวคิดในลักษณะ 2 มิติ ของโมเดลปกติสะสมแบบ 2 พารามิเตอร์ ไปสู่โมเดลของการประมาณค่าพารามิเตอร์แบบโมเดลการตอบสนองข้อสอบแบบพหุมิติ Bock and Aitkin ได้

แสดงให้เห็นว่าการวิเคราะห์องค์ประกอบและทฤษฎีการตอบสนองข้อสอบต่างมุ่งศึกษาในสิ่งเดียวกัน โดยผลลัพธ์สุดท้ายคือทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ

อย่างไรก็ตาม แม้ว่าโมเดลที่นำเสนอโดย Bock and Aitkin มีประโยชน์สำหรับการขยายแนวคิดของโมเดลการตอบสนองข้อสอบแบบพหุมิติ แต่จุดมุ่งหมายหลักยังคงมุ่งไปที่การวิเคราะห์องค์ประกอบเป็นสำคัญ ซึ่งคล้ายคลึงกับการศึกษาของ Bock, Gibbons, and Muraki (1988) ที่เน้นไปที่การระบุจำนวนองค์ประกอบมากกว่าที่จะพิจารณาปฏิสัมพันธ์ของผู้สอบและข้อสอบ

1.2 ทฤษฎีการตอบสนองข้อสอบ

เมื่อพิจารณาแนวคิดของทฤษฎีการตอบสนองข้อสอบและการวิเคราะห์องค์ประกอบจะเห็นได้ว่ามีจุดมุ่งหมายในการวิเคราะห์ที่แตกต่างกัน โดยการวิเคราะห์องค์ประกอบมุ่งไปที่การกำหนดองค์ประกอบให้น้อยที่สุดที่ถอดแบบข้อมูลในรูปของเมตริกซ์การตอบข้อสอบ ส่วนโมเดลของทฤษฎีการตอบสนองข้อสอบ ศึกษาความสัมพันธ์ระหว่างผู้สอบและข้อสอบซึ่งการวัดแบบพหุมิติในมุมมองของทฤษฎีการตอบสนองข้อสอบมีนักวิชาการได้นำเสนอแนวคิดตามลำดับ ดังนี้

Rasch (1960, 1962 cited in Reckase, 2009, pp. 68-69) ได้เสนอโมเดลที่ยอมรับโดยทั่วไป รวมถึงความเป็นไปได้ของความสามารถของผู้สอบ โดยนำเสนอในรูปเวคเตอร์มากกว่าค่าสเกลาร์ ซึ่งคำนวณได้ดังสมการ 1 (Reckase, 2009, pp. 68-69)

$$P\left(\mu_{ij}/\theta_j,\eta_i\right)=\frac{1}{\gamma\left(\theta_j,\eta_i\right)}e^{f\left(u_{ij}\right)'\theta_j+g\left(u_{ij}\right)'\eta_i+\theta_j'h\left(u_{ij}\right)\eta_i+\ell\left(u_{ij}\right)} \quad (1)$$

เมื่อ f,g,h และ ℓ แทน ฟังก์ชันของคะแนนในแต่ละข้อ

$\boldsymbol{u}$ แทน ค่าสถิติที่ได้จากการสังเกตจากค่าพารามิเตอร์ของผู้สอบและข้อสอบ

$\boldsymbol{\eta}$ แทน เวคเตอร์ค่าพารามิเตอร์ของข้อสอบ

$\gamma\left(\theta_j,\eta_i\right)$ แทน ฟังก์ชัน Normalizing ที่จะสามารถรับประกันได้ว่าค่าที่อยู่ในฟังก์ชันของ IRT อยู่ในช่วง 0 และ 1

จากการพิจารณาโมเดลในข้างต้น ปรากฏว่า มีการระบุระดับความสามารถของผู้สอบในรูปของเวคเตอร์ค่าพารามิเตอร์ โดยการประมาณค่าพารามิเตอร์ของผู้สอบและข้อสอบมีความเป็นอิสระจากกัน เมื่อพิจารณาส่วนประกอบอื่นๆ เช่น f,g,h ซึ่งเป็นเวคเตอร์ของการให้คะแนนจะต้องทราบข้อมูลในส่วนนี้ก่อนทำการวิเคราะห์ด้วยโมเดลราส์ช ไม่สามารถประมาณค่าได้จากข้อมูลในการตอบข้อสอบของผู้สอบ สำหรับในกรณีของข้อมูลที่มีลักษณะเป็นเอกมิติ โดยที่การให้คะแนนจะมี 2 ค่า คือ 0 และ 1 โดยไม่ได้คำนึงถึงลักษณะข้อสอบ ดังนั้น การพิจารณาเฉพาะจำนวนที่ตอบข้อสอบได้ถูกต้อง ถือเป็นค่าสถิติที่เพียงพอสำหรับการพิจารณา

ถ้าหากมีการตอบสนองข้อสอบในมิติที่ต่างกันสองด้าน โดยคะแนน การตอบข้อสอบในแต่ละข้อของมิติที่ 1 และ 2 คือ μ_1 และ μ_2 ดังนั้นค่าสถิติที่เพียงพอสำหรับประมาณค่า θ_1 และ θ_2

คือ $\eta\mu_1$ และ $\eta\mu_2$ เมื่อ n คือ จำนวนข้อที่ตอบข้อสอบได้ถูกต้อง อย่างไรก็ตามเนื่องจาก μ_1 และ μ_2 เป็นค่าคงที่ข้ามข้อสอบ ดังนั้น คะแนนที่ได้แต่ละมิติเป็นฟังก์ชันของ n เท่านั้น ดังนั้นค่า θ_1 และ θ_2 ที่ประมาณได้ จึงมีความสัมพันธ์กันอย่างสมบูรณ์ นั่นคือโมเดล ที่ได้ยังคงเป็นโมเดลแบบเอกมิตินั่นเอง จากปัญหาดังกล่าวนี้ มีนักวิจัยหลายท่านพยายามที่จะแบ่งคะแนนให้มากกว่า 2 ค่า เช่น จัดกระทำกับข้อสอบจากลักษณะ Dichotomous Items แปลงให้เป็น Single Polytomous Item เพื่อที่จะนำไปสู่การวิเคราะห์แบบ Polytomous Item (Reckase, 2009, pp. 68-69)

Lord and Novick (1968 Cited in Reckase, 2009, pp. 69-70) เสนอโมเดลพื้นฐานทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ ที่บ่งชี้ถึงนิยามของมิติคุณลักษณะแฝงได้อย่างสมบูรณ์ และข้อตกลงเบื้องต้นของความเป็นอิสระในที่นี้ หมายถึงคุณลักษณะของกลุ่มผู้สอบต่างๆ ที่มีค่าเหมือนกันของมิติความสามารถ $\theta_1, \theta_2, ..., \theta_k$ เป็นการกระจายคะแนน ผลการตอบข้อนั้นของผู้สอบแต่ละคนไม่สัมพันธ์กัน เมื่อ k คือ จำนวนของมิติ ส่วน θ ในที่นี้เป็นเวคเตอร์มิติคุณลักษณะที่สมบูรณ์ นอกจากนี้ Lord and Novick ยังแสดงให้เห็นถึงความสัมพันธ์ระหว่างโมเดลปกติสะสม ตามทฤษฎีตอบสนองข้อสอบแบบเอกมิติและโมเดลองค์ประกอบร่วมอย่างชัดเจนมากยิ่งขึ้น ต่อมา Samejima (Samejima, 1974 Cited in Reckase, 2009, p. 70) ได้เสนอสูตรที่ใช้โดยทั่วไป โดยมีข้อตกลงเบื้องต้นว่าการตอบที่มีลักษณะการตรวจให้คะแนน 2 ค่า หรือมากกว่า 2 ค่า เป็นผลมาจากการแบ่งเป็นส่วนย่อยๆ จากการตอบของตัวแปรต่อเนื่อง ซึ่ง Samejima ได้พัฒนาโมเดลการตอบสนองข้อสอบแบบพหุมิติ สำหรับข้อสอบซึ่งมีลักษณะเป็นการตอบแบบต่อเนื่อง ซึ่งคำนวณได้ดังสมการ 2 (Reckase, 2009, p. 70)

$$P_{z_i}(\theta) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a_i'(\theta_j - b_i)} e^{-\frac{t^2}{2}} dt \qquad (2)$$

เมื่อ z_i แทน การตอบสนองแบบต่อเนื่องต่อข้อสอบข้อที่ i

P แทน ความน่าจะเป็นที่จะได้รับคะแนน z หรือดีกว่าจากข้อสอบข้อที่ i

b_i แทน เวกเตอร์ของพารามิเตอร์ความยาก พร้อมกับค่าพารามิเตอร์อื่นๆ ที่นิยามไว้ก่อนหน้านี้

แม้ว่าโมเดลของ Samejima เป็นโมเดลแรกที่ทำให้เกิดแนวคิดเกี่ยวกับโมเดลการตอบสนองข้อสอบแบบพหุมิติ มีความชัดเจน แต่โมเดลดังกล่าวไม่สามารถนำไปประยุกต์ได้จริงในทางปฏิบัติ ซึ่งสาเหตุอาจเกิดจากการตอบข้อสอบ มีการพิจารณาเป็นตัวแปรต่อเนื่อง ซึ่งไม่สอดคล้องกับบริบทของแบบสอบทั้งทางด้านการศึกษาและจิตวิทยา อย่างไรก็ตาม ถ้าประเมินเกี่ยวกับทักษะการปฏิบัติ โมเดลดังกล่าวจะเหมาะในการประยุกต์ได้มากขึ้น (Reskase, 2009, p. 70)

2. พัฒนาการของทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ

ในช่วงปลายปี ค.ศ. 1970 และช่วงต้นปี ค.ศ. 1980 นักวิจัยจำนวนหนึ่งเริ่มพัฒนาแนวคิดเกี่ยวกับโมเดลการตอบสนองข้อสอบแบบพหุมิติให้สามารถนำไปใช้ได้ในทางปฏิบัติ ดังจะเห็นจากการศึกษาเกี่ยวกับโมเดลราช์สแบบพหุมิติ โดย Reckase (1972) ต่อมาก็มีการนำเสนอแนวคิด

เกี่ยวกับปฏิสัมพันธ์ระหว่างผู้สอบกับข้อสอบโดย Mulaik (1972), Sympsom (1978) และ Whitely (1980) ซึ่ง Mulaik (1972) ได้นำเสนอโมเดล ซึ่งคำนวณได้ดังสมการ 3 (Reckase, 2009, p. 71)

$$P\left(u_{ij}\middle|\theta_j,\eta_i\right)=\frac{\sum_{k=1}^{m}e^{(\theta_{jk}-\eta_{ik})u_{ij}}}{1+\sum_{k=1}^{m}e^{(\theta_{jk}-\eta_{ik})u_{ij}}} \qquad (3)$$

จากสัญลักษณ์ที่นิยามไว้ในสมการที่ 1 เมื่อ u_{ij} = 0 หรือ 1 คุณสมบัติที่น่าสนใจของโมเดลคือ มีการกำหนดให้ค่าของเลขชี้กำลังและความน่าจะเป็นในการตอบข้อสอบถูกเพิ่มขึ้นเมื่อจำนวนของมิติเพิ่มขึ้น เช่น ถ้าผู้สอบและพารามิเตอร์ของข้อสอบเท่ากันและกำหนดให้เลขชี้กำลังเท่ากับ 0 ความน่าจะเป็นในการตอบได้ถูกมีค่าเท่ากับ 0.5 ถ้า m = 1 แต่จะเพิ่มขึ้นเป็น 0.67 ถ้า m = 2 และ 0.75 ถ้า m = 3 โดยทั่วไปถ้าทุกมิติมีเลขชี้กำลังเป็น 0 โอกาสของการตอบถูก คือ m/(m+1) ซึ่งถือได้ว่าเป็นคุณสมบัติของโมเดล ถ้าหากความน่าจะเป็นของการตอบถูกได้ กำหนดไว้แล้ว ค่าของตัวเลขชี้กำลังจะมีการเปลี่ยนแปลงตามจำนวนของมิติที่เพิ่มขึ้น ดังตัวอย่างในตารางที่ 2-5 กรณีนี้กำหนดให้ค่าของเลขชี้กำลังเท่ากับ x เมื่อ e^x มีค่าเท่ากันในแต่ละมิติของ m และความน่าจะเป็นของการตอบถูกเท่ากับ 0.5 แสดงว่า เลขชี้กำลังมีการเปลี่ยนแปลงเล็กน้อยตามจำนวนที่เพิ่มขึ้นของมิติ (Reckase, 2009, pp. 71-72)

ตารางที่ 2-5 ความสัมพันธ์ของขนาดเลขชี้กำลังกับจำนวนมิติจากโมเดลของ Mulaik (p=.5)

จำนวนมิติ (Number of Dimensions)	เลขชี้กำลังของ e (Exponent of e)
1	0
2	-.69
3	-1.10
4	-1.39
5	-1.61
6	-1.79
7	-1.95
8	-2.08
9	-2.20
10	-2.30

ต่อมาโมเดลความสัมพันธ์ระหว่างจำนวนมิติและขนาดเลขชี้กำลังที่มีทิศทางตรงกันข้ามกับโมเดลของ Mulaik (1972) ได้ถูกนำเสนอโดย Sympson (1978) และ Whitely (1980) ซึ่งเป็นโมเดลที่กำหนดค่าของเลขชี้กำลังและโอกาสของการตอบถูกลดลงเมื่อจำนวนมิติเพิ่มขึ้น และโมเดลที่นำเสนอโดย Sympson ซึ่งคำนวณได้ดังสมการ 4 (Reckase, 2009, p. 72)

$$P\left(u_{ij}=1\middle|\theta_j,a_i,b_i,c_i\right)=c_i+\left(1-c_i\right)\prod_{k=1}^{m}\frac{e^{a_{ik}\left(\theta_{jk}-b_{ik}\right)}}{1+e^{a_{ik}\left(\theta_{jk}-b_{ik}\right)}} \quad (4)$$

เมื่อ a_i แทน สเกลาร์ค่าพารามิเตอร์อำนาจจำแนกของข้อสอบ

b_i แทน สเกลาร์ค่าพารามิเตอร์ความยากของข้อสอบ

c_i แทน สเกลาร์พารามิเตอร์โอกาสการเดาของข้อสอบ

เมื่อเลขชี้กำลังทั้งหมดเท่ากับ 0 ความน่าจะเป็นในการตอบถูกเท่ากับ $c_i+\left(1-c_i\right)\left(.5\right)^m$ เช่น ในกรณีที่มิติ m เพิ่มขึ้น ค่าที่ได้จะมีค่าเข้าใกล้ c_i (Reckase, 2009, p. 72)

โมเดลของ Mulaik (1972) การประมาณค่าพารามิเตอร์ของข้อสอบเปลี่ยนแปลงไปตามจำนวนมิติ ตารางที่ 2-6 แสดงค่าเลขชี้กำลังของ e เมื่อจำนวนมิติเพิ่มขึ้น ในกรณีที่ c_i = 0 และเลขชี้กำลังทั้งหมดเท่ากัน เนื่องจากความน่าจะเป็นในการตอบถูกขึ้นอยู่กับจำนวนของ m และจำนวนทั้งหมดมีค่าน้อยกว่า 1.0 ซึ่งจำนวนต่างๆ จะต้องเพิ่มค่าขึ้นด้วยถ้าเพิ่มจำนวนของ m ทำให้เกิดโอกาสในการตอบถูกเท่ากัน (Reckase, 2009, pp. 72-73)

ตารางที่ 2-6 ความสัมพันธ์ของขนาดเลขชี้กำลังกับจำนวนมิติจากโมเดลของ Sympson (p=.5)

จำนวนมิติ (Number of Dimensions)	เลขชี้กำลังของ e (Exponent of e)
1	0.0
2	0.9
3	1.3
4	1.7
5	1.9
6	2.1
7	2.3
8	2.4
9	2.5
10	2.6

McKinley and Reckase (1982) ได้ศึกษาความแปรเปลี่ยนของโมเดลทั่วไปของราสช์ (General Rasch Model) และโมเดลโลจิสติกที่เป็นที่นิยมในปัจจุบัน ซึ่งคำนวณได้ดังสมการ 5 (Reckase, 2009, p. 73)

$$P\left(u_{ij}=1\middle|\theta_j,a_i,d_i\right)=\frac{e^{\sum_{k=1}^{m}a_{ik}\theta_{jk}+d_i}}{1+e^{\sum_{k=1}^{m}a_{ik}\theta_{jk}+d_i}} \quad (5)$$

เมื่อ a_i แทน เวคเตอร์ค่าพารามิเตอร์อำนาจจำแนกของข้อสอบ

d_i แทน ค่าพารามิเตอร์ความยากของข้อสอบ

โมเดลนี้เป็นโมเดลแบบพหุ (Multivariate) ที่ขยายแนวคิดมาจากโมเดลโลจิสติกแบบ 2 พารามิเตอร์ ซึ่งมีชื่อเรียกเฉพาะว่า Compensatory Model เพราะว่าการที่ค่า θ มีค่าลดลงในมิติใดมิติหนึ่ง สามารถไปทดแทนให้ค่า θ ในมิติอื่น ๆ สูงขึ้นได้นั่นเอง ต่อมา Spry,Davey, Reckase, Ackerman, and Carlson (1990) ได้พัฒนาโมเดลรูปแบบทั่วไป (Generalized Model) โดยเฉพาะ Compensatory Model และ Partial Compensatory Model

3. ประเภทของทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ

ทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ ประกอบด้วยค่าพารามิเตอร์ของผู้สอบตั้งแต่ 2 ค่าขึ้นไป โดยการศึกษาในหลายมิติ ทำให้โมเดลมีความสอดคล้องกับข้อมูลมากขึ้น สำหรับผู้สอบที่มีความแตกต่างกันอย่างเป็นระบบ เนื่องจากความยากของชุดข้อสอบ โมเดลหลายโมเดลที่เป็นแบบพหุมิติ ค่าอำนาจจำแนกของข้อสอบ จะส่งผลกระทบต่อมิติในแต่ละชุดข้อสอบ ซึ่งสามารถแบ่งโมเดลการตอบสนองข้อสอบแบบพหุมิติ ออกเป็น 2 ประเภท โดยขึ้นอยู่กับคะแนนที่ได้จากแบบวัดเป็นเกณฑ์ คือ คะแนนที่ได้จากแบบวัดที่มี 2 ค่า (Dichotomously Scored) และคะแนนที่ได้จากแบบวัดที่มีหลายค่า (Polytomously Scored) (Ackerman, 1994; Wilson, Adam & Wang, 1997 Cited in Briggs & Wilson, 2003) มีรายละเอียดดังนี้

3.1 ประเภทคะแนนที่ได้จากแบบวัดมี 2 ค่า (Dichotomously Scored)

โมเดลการวัดตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ (Multidimensional Items Response Theory Model: MIRT Model) ได้ถูกจำแนกออกเป็น 2 โมเดล เพื่อใช้อธิบายคะแนนจากการทดสอบที่ถูกแบ่งเป็น 2 ค่า คือ Compensatory Model และ Noncompensatory Model (Ackerman, 1994; Yao & Boughton, 2007) ดังนี้

3.1.1 โมเดลการวัดมิติที่สามารถทดแทนกันได้ (Compensatory Model)

การวัดตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ ความน่าจะเป็นในการตอบข้อสอบข้อที่ i ถูกต้อง อธิบายโดยใช้ Multidimensional Compensatory Two-Parameter Logistic Model: MC2PL (Reckase, 1985 cited in Ackerman, 1994) โมเดลดังกล่าว มีพารามิเตอร์ค่าอำนาจจำแนกของข้อสอบในแต่ละมิติที่กำหนดขึ้น แต่มีพารามิเตอร์ค่าความยากของข้อสอบเพียงค่าเดียว และพารามิเตอร์ความสามารถ (θ) ที่ขาดหายไปในมิติหนึ่งๆ สามารถไปเพิ่มเติมในมิติอื่นๆ ให้สูงขึ้นได้ (Ackerman, 1994; Ackerman, Gierl, & Walker, 2003) ซึ่งคำนวณได้ดังสมการ 6

$$P(X_i = 1) = \frac{e^{\sum_{k=1}^{m} a_{ik}\theta_k + d_i}}{1.0 + e^{\sum_{k=1}^{m} a_{ik}\theta_k + d_i}} \quad (6)$$

เมื่อ X_i แทน คะแนน (0, 1) ของข้อสอบข้อที่ i

a_{ik} แทน เวคเตอร์ค่าพารามิเตอร์อำนาจจำแนกของข้อสอบ

d_i แทน สเกลาร์ค่าพารามิเตอร์ความยากของข้อสอบข้อที่ i

θ_k แทน เวคเตอร์ค่าพารามิเตอร์ความสามารถ

3.1.2 โมเดลการวัดมิติที่ไม่สามารถทดแทนกันได้ (Non compensatory Model)

Sympson (Sympson, 1978 cited in Ackerman, 1994) เสนอ โมเดลการวัดที่มีคุณสมบัติเดียวกันกับโมเดลการวัดมิติที่สามารถทดแทนกันได้ เรียกว่า Multidimensional Noncompensatory Two-Parameter Logistic Model (MNC2PL Model) โดยในโมเดลดังกล่าวนี้จะแสดงพารามิเตอร์ค่าความยากและพารามิเตอร์ค่าอำนาจจำแนกของข้อสอบในแต่ละมิติ และกำหนดค่าของส่วนประกอบในสมการ ซึ่งถ้าค่าความน่าจะเป็นในการตอบข้อสอบถูกลดลง จำนวนมิติจะเพิ่มขึ้น ซึ่งคำนวณได้ดังสมการ 7

$$P(X_i = 1) = \prod_{k=1}^{m} \frac{e^{a_{ik}(\theta_k - b_{ik})}}{1 + e^{a_{ik}(\theta_k - b_{ik})}} \quad (7)$$

เมื่อ X_i แทน คะแนน (0, 1) ของข้อสอบข้อที่ i

a_{ik} แทน เวคเตอร์ค่าพารามิเตอร์อำนาจจำแนกของข้อสอบ

b_{ik} แทน สเกลาร์ค่าพารามิเตอร์ความยากของข้อสอบข้อที่ i

θ_k แทน เวคเตอร์ค่าพารามิเตอร์ความสามารถ

3.2 ประเภทคะแนนที่ได้จากแบบวัดมีหลายค่า (Polytomously Scored)

โมเดลการตอบสนองข้อสอบแบบพหุมิติแบบที่มีการตรวจให้คะแนน 2 ค่า (Dichotomously scored) ที่นำเสนอข้างต้น มีข้อจำกัดที่สำคัญ คือ การตรวจให้คะแนนรายข้อ ต้องเป็นแบบ 2 ค่า (Binary) เช่น การตรวจให้คะแนนแบบ 0 และ 1, ถูกหรือผิด, เห็นด้วย หรือไม่เห็นด้วย เป็นต้น ในขณะที่เครื่องมือทางการศึกษาและจิตวิทยาหลายประเภท แต่ละข้อ คำถามมักมีหลายรายการคำตอบที่กำหนดลำดับหรือ น้ำหนักคะแนนต่างกัน เพราะน่าจะให้สารสนเทศและความเที่ยงจากการตอบที่สูงกว่าการตรวจให้คะแนนแบบ 2 ค่า เช่น แบบวัดเจตคติแบบวัดบุคลิกภาพ แบบวัดความสนใจในอาชีพ เป็นต้น จึงได้มีการพัฒนาโมเดลการตอบสนอง ข้อสอบสำหรับใช้กับการตรวจให้คะแนนรายข้อมากกว่า 2 ค่า ซึ่งปัจจุบันมีรูปแบบของเครื่องมือ ที่นิยมใช้กัน เช่น

มาตรประมาณค่า (Rating Scale) การตรวจให้คะแนนความรู้บางส่วน (Partial Credit) การตรวจให้คะแนนตามลำดับขั้นของรายการหลายคำตอบ (Ordered-response categories) เป็นต้น ซึ่งมีโมเดลดังต่อไปนี้

3.2.1 Multidimensional Random Coefficients Multinomial Logit Model: MRCML

Model เป็นโมเดลที่ได้มีการปรับขยายมาจากโมเดลการตอบสนองข้อสอบของราชส์ (Rasch family of items response model) การประยุกต์ใช้โมเดล ดังกล่าวเริ่มต้นจากนักวิชาการหลายคนที่ทำการศึกษาโมเดลนี้ เช่น Adams (1997) Wang (1999) Wang, Wilson and Adams (1997) มีแนวคิดพื้นฐานการพัฒนาโมเดลจากการสร้างแบบพิมพ์ (Building-block) หมายความว่า สำหรับข้อสอบข้อที่ i ซึ่งเป็นข้อสอบที่สามารถตอบ ได้หลายคำตอบและสามารถให้คะแนนได้หลายค่า (k ค่า) หรืออาจเรียกว่ามิติที่สามารถเป็นไปได้ ทั้งหมดซึ่งมีลักษณะเป็นกลุ่มขนาดใหญ่ โดยกำหนดให้เป็น d (d=1,...,D) ส่วนผู้สอบที่ตอบ ข้อสอบกำหนดให้เป็น p (p=1,...,P) แล้ว ความน่าจะเป็นของผู้สอบที่ตอบข้อสอบข้อที่ i $\left(P_{ik}\right)$ ซึ่ง มีอยู่ k กลุ่ม เมื่อเปรียบเทียบกับ k-1 กลุ่ม $\left(P_{ik-1}\right)$ แล้วจะได้เป็นฟังก์ชันเชิงเส้นตรงของ ความสามารถแฝงบนมิติ $\left(\theta_d\right)$และความยากเชิงสัมพัทธ์ของกลุ่ม k (Briggs & Wilson, 2003) ซึ่งคำนวณได้ดังสมการ 8

$$k\left(\delta_{ik}\right):\log\left(\frac{P_{ik}}{P_{ik-1}}\right)=\theta_d-\delta_{ik} \tag{8}$$

MRCML Model ซึ่งคำนวณได้ดังสมการ 9 ได้ดังนี้ คือ (Wilson, Adam & Wang, 1997 Cited in Briggs & Wilson, 2003)

$$P\left(X_{ik}=1;A,B,\xi/\theta\right)=\frac{\exp\left(b_{ik}\theta+a'_{ik}\xi\right)}{\sum_{k=1}^{k_i}\exp\left(b_{ik}\theta+a'_{ik}\xi\right)} \tag{9}$$

เมื่อ a'_{ik} แทน ค่าพารามิเตอร์อำนาจจำแนกของข้อสอบ

b_{ik} แทน ค่าพารามิเตอร์ความยากของข้อสอบข้อที่ i

θ แทน พารามิเตอร์ความสามารถของผู้สอบ

3.2.2 Multidimensional Partial Credit Model

โมเดล Multidimensional Partial Credit Model (MPC) เป็นโมเดลที่ปรับขยายมาจากโมเดลของราส์ชแบบ 1 พารามิเตอร์สำหรับแบบทดสอบที่ตรวจให้คะแนนหลายค่า ซึ่งคำนวณได้ดังสมการ 10 (Reckase, 2009, p. 33)

$$P\left(u_{ij}=k/\theta_j\right)=\frac{e^{\left[\sum_{u=0}^{k}\left(\theta_j-\delta_{iu}\right)\right]}}{\sum_{v=0}^{mi}e^{\left[\sum_{u=0}^{v}\left(\theta_j-\delta_{iu}\right)\right]}} \quad (10)$$

เมื่อ δ_{iu} แทน ค่าพารามิเตอร์ค่า threshold สำหรับชุดคะแนนความยากของข้อสอบข้อที่ i บนมิติที่ u

3.2.3 Multidimensional Generalized Partial Credit Model

โมเดล Multidimensional Generalized Partial Credit Model (MGPC) เป็นโมเดลที่ปรับขยายมาจากโมเดล Generalized Partial Credit Model (GPC) ซึ่ง ได้รับการออกแบบมาสำหรับอธิบายปฏิสัมพันธ์ของผู้สอบด้วยข้อสอบที่มีคะแนนมากกว่าสองค่า ซึ่งคำนวณได้ดังสมการ 11 (Reckase, 2009, p. 37)

$$P\left(u_{ij}=k/\theta_j\right)=\frac{e^{\left[\sum_{u=1}^{k}Da_i\left(\theta_j-b_i+d_{iu}\right)\right]}}{\sum_{v=1}^{mi}e^{\left[\sum_{u=1}^{v}Da_i\left(\theta_j-b_i+d_{iu}\right)\right]}} \quad (11)$$

เมื่อ k แทน คะแนนบนข้อสอบ (0, 1, 2,...,)

b_i แทน ค่าพารามิเตอร์ความยากของข้อสอบข้อที่ i

d_{iu} แทน พารามิเตอร์ของความสัมพันธ์ของข้อสอบข้อที่ I แต่ละข้อที่สัมพันธ์กับ มิติที่ u

3.2.4 Multidimensional Graded Response Model

โมเดล Multidimensional Graded Response Model เป็นโมเดล สำหรับแบบทดสอบที่ตรวจให้คะแนนแบบหลายค่าซึ่งออกแบบโดย Muraki and Carson (Muraki & Carson, 1993 Cited in Reckase, 2009, p. 39) ซึ่งคำนวณได้ดังสมการ 12 (Reckase, 2009, p. 39)

$$P\left(u_{ij}=k/\theta_j\right)=\frac{1}{\sqrt{2\pi}}\int_{a_i\left(\theta_j-b_i,k+1\right)}^{a_i\left(\theta_j-b_{ik}\right)}e^{\frac{-t^2}{2}}dt \quad (12)$$

เมื่อ k แทน คะแนนบนข้อสอบ (0, 1, 2,..., m_i)

a_i แทน เวคเตอร์ค่าพารามิเตอร์อำนาจจำแนกของข้อสอบ

b_{ik} แทน พารามิเตอร์ของความสัมพันธ์ของผู้สอบแต่ละคนกับขั้นตอนของ การทำข้อสอบแต่ละข้อ

ตอนที่ 3 การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ

1. ความเป็นมาของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ

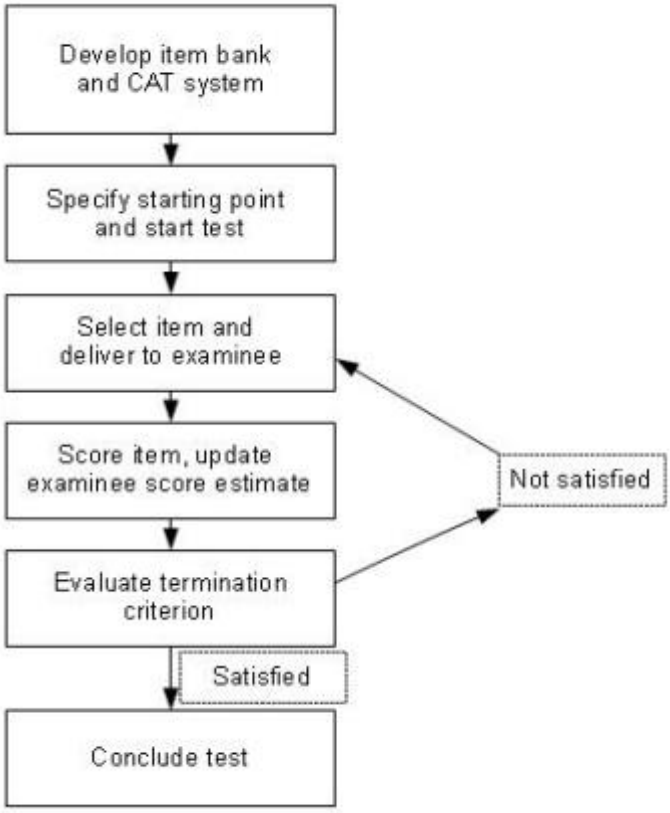
การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ (Multidimensional Computerized Adaptive Testing: MCAT) เป็นวิธีการทดสอบที่มีการขยายแนวคิดมาจากการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบเอกมิติ (Unidimensional Computerized Adaptive Testing: CAT) ซึ่งเป็นวิธีการเฉพาะที่ใช้ในการประเมินความสามารถและคุณลักษณะแฝงอื่นๆ ในการเลือกข้อสอบให้เหมาะกับระดับความสามารถของผู้เข้าสอบ (Frey & Seitz, 2009) เป็นการลดจำนวนข้อสอบให้ผู้สอบตอบข้อสอบและเพิ่มความแม่นยำของการวัด เมื่อเปรียบเทียบกับการทดสอบแบบดั้งเดิมที่มีจำนวนข้อสอบคงที่ในแบบทดสอบ (Fixed Item Test: FIT) และเป็นการขยายแนวคิดจากโมเดลการตอบสนองข้อสอบแบบพหุมิติ (Multidimensional Item Response Theory Model: MIRT) ซึ่งถือว่าคุณลักษณะแฝงของบุคคลมากกว่า 1 องค์ประกอบ ส่งผลต่อการตอบข้อสอบ และการทดสอบแบบปรับเหมาะกำลังได้รับความนิยมอย่างมากในการนำไปปฏิบัติ จึงเกิดการรวมกันของทั้งสองวิธีนี้เรียกว่า การทดสอบปรับเหมาะแบบพหุมิติ (Multidimensional Adaptive Testing: MAT) (Segall, 1996, 2010, p. 62; Reckase, 2010, p. 95) และเมื่อนำคอมพิวเตอร์มาใช้ในกระบวนการทดสอบ จึงเรียกว่า การทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ (Multidimensional Computerized Adaptive Testing: MCAT) ซึ่งวิธีการทดสอบดังกล่าวนี้ เป็นการทดสอบที่มีประสิทธิภาพในการวัดสูง ไม่ว่าจะเป็นความแม่นยำในการวัด การลดความยาวของข้อสอบ และเป็นการวัดที่ตรงกับสภาพจริงของลักษณะข้อสอบ ดังนั้น ในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ จึงสามารถลดจำนวนข้อสอบลงได้มากกว่าการทดสอบด้วยการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์เอกมิติ ประมาณ 30-50% และลดจำนวนข้อสอบได้มากกว่า การทดสอบแบบดั้งเดิม ที่กำหนดจำนวนข้อสอบไว้ประมาณ 70% โดยไม่สูญเสียความแม่นยำ (Frey & Seitz, 2009) และการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติมีประสิทธิภาพ ที่สูงกว่าการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบเอกมิติ ถึง 1.3 เท่า (Frey & Seitz, 2009) อย่างไรก็ตาม ถ้ากระบวนการทดสอบได้ใช้วิธีการเลือกข้อสอบแบบสุ่ม (random) การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ จะส่งผลให้การทดสอบมีประสิทธิภาพสูงกว่าวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบเอกมิติ ถึง 3.7 เท่า (Frey & Seitz, 2009)

การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ เป็นการขยายมิติความสามารถการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบเอกมิติสู่พหุมิติ ซึ่งการนำเสนอสาระสำคัญเกี่ยวกับการประมาณค่าความสามารถของผู้สอบกับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ โดยแบ่งออกเป็น 3 ส่วนคือ โมเดลอยู่ภายใต้ทฤษฎีตอบสนองข้อสอบแบบพหุมิติ องค์ประกอบของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ และการเปรียบเทียบและแนวทางการเลือกวิธีการประมาณค่าความสามารถของผู้สอบในอนาคต

2. หลักการของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ

การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์มีหลักการคัดเลือกข้อสอบที่เหมาะสมกับแต่ละบุคคล ซึ่งตั้งอยู่บนฐานการตอบข้อสอบข้อที่ผ่านมาของผู้สอบ กล่าวคือ เมื่อผู้สอบทำข้อสอบข้อเริ่มต้นหรือชุดเริ่มต้น (ขึ้นอยู่กับการออกแบบ) แล้วจะนำผลการตอบข้อสอบมาวิเคราะห์ หรือ

ประเมินระดับความสามารถของผู้สอบ เพื่อที่จะเลือกข้อสอบที่เหมาะสมข้อถัดไป โดยอาศัยหลักการคัดเลือกข้อสอบว่า ถ้าทำข้อสอบข้อที่ผ่านมาถูกต้อง ข้อสอบข้อถัดไปจะยากมากขึ้นหากทำข้อสอบข้อที่ผ่านมาผิด ข้อสอบข้อถัดไปจะง่ายลง กระบวนการนี้จะดำเนินการไปจนกว่าการประเมินระดับความสามารถของผู้สอบเป็นไปตามเกณฑ์การยุติการทดสอบที่กำหนด (ขึ้นอยู่กับการออกแบบ) การทดสอบจึงยุติลง (Reckase, 2009, pp. 311-313) ซึ่งขั้นตอนการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ แสดงได้ตามภาพที่ 2-6



ภาพที่ 2-6 ขั้นตอนการดำเนินการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ (Thomposon & Weiss, 2011, p. 2)

จากภาพที่ 2.6 แสดงกระบวนการของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์ ซึ่งมี 5 ขั้นตอน สามารถอธิบายได้ ดังนี้ 1) การสร้างคลังข้อสอบ (Develop Item Bank and CAT Sytem) เป็นขั้นตอนที่ต้องการพิจารณาขนาดของคลังข้อสอบและเงื่อนไขต่างที่สอดคล้องกับหลักการของทฤษฎีการตอบสนองข้อสอบ เพื่อคัดเลือกข้อสอบที่ตรงตามเงื่อนที่กำหนดมาจัดเก็บในคลังข้อสอบ 2) การคัดเลือกข้อสอบข้อเริ่มต้น (Staring Point and Start Test) เป็นการคัดเลือกข้อสอบข้อเริ่มต้น ซึ่งเป็นข้อสอบที่มีความยากของข้อสอบระดับปานกลาง กล่าวคือการเลือกข้อสอบที่มีระดับความสามารถที่มีระดับยากของข้อสอบใกล้เคียงกับค่าเฉลี่ยของ ค่าความยากของข้อสอบทั้งหมดหรือการคัดเลือกข้อสอบข้อเริ่มต้นจากการสุ่มในช่วงระดับความยากของข้อสอบที่กำหนดหรือการคัดเลือกข้อสอบข้อเริ่มต้นตามกฎเกณฑ์ที่กำหนด 3) การคัดเลือกข้อสอบข้อถัดไป (Select Item and

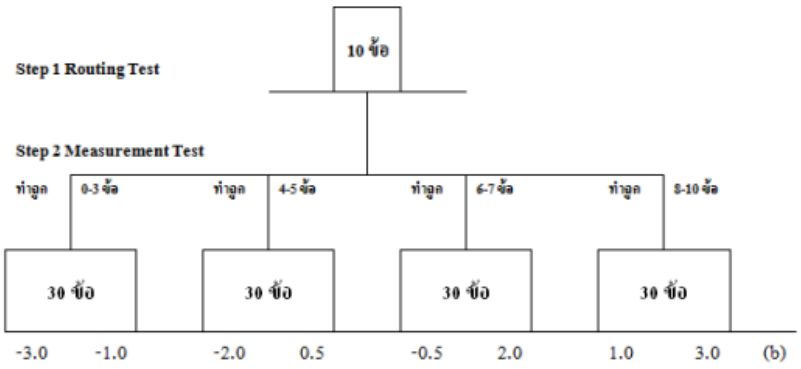
Delliver to Examinee) เป็นขั้นตอนที่ต้องพิจารณาคัดเลือกข้อสอบตามเงื่อนไขที่กำหนด โดยอาศัยผลการตอบข้อสอบข้อก่อนหน้ามาร่วมพิจารณาด้วย 4) การประมาณค่าความสามารถ (Score Item, Updata Examinee Score Estimate) เป็นการประมาณค่าความสามารถของผู้สอบหลังจากตอบข้อสอบแล้ว และ 5) การยุติการทดสอบ (Conctude Test) คือ การสรุปผลการทดสอบหรือกล่าวอีกนัยหนึ่งว่าเป็นการสิ้นสุดการทดสอบ เมื่อผู้สอบทำการทดสอบตามเงื่อนไขหรือเกณฑ์ที่กำหนด

3. ประเภทของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ

การจำแนกประเภทของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์สามารถแบ่งออกได้หลายประเภทขึ้นอยู่กับเกณฑ์ที่นำมาใช้จำแนก Weiss (1974) ได้จำแนกการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ออกเป็น 2 ประเภท ดังนี้

3.1 การทดสอบแบบ 2 ขั้นตอน (Two-stage Strategies)

การทดสอบแบบปรับเหมาะชนิด 2 ขั้นตอน เป็นการทดสอบโดยใช้ข้อสอบ 2 ชุด กล่าวคือการดำเนินการทดสอบเริ่มด้วยให้ผู้สอบทำแบบทดสอบเพื่อแยกทาง (Routing Test) ซึ่งผลจากการทดสอบนี้จะเห็นตัวชี้วัดว่าผู้สอบได้รับแบบทดสอบชุดใดในขั้นที่ 2 สามารถเรียกแบบทดสอบนี้ว่า แบบทดสอบเพื่อวัดผล (Measurement Test) โดยหลักการแล้ว ผู้สอบที่ได้คะแนนน้อยจากการทดสอบแยกทาง แสดงว่าเป็นผู้มีความสามารถต่ำจะได้รับข้อสอบเพื่อวัดผลชุดที่ง่ายลง ส่วนผู้ที่ได้คะแนนสูงจากการทดสอบเพื่อแยกทาง แสดงว่าเป็นผู้มีความสามารถสูงจะได้รับข้อสอบ เพื่อวัดผลที่ยากขึ้น แสดงได้ดังภาพที่ 2-7



ภาพที่ 2-7 รูปแบบการทดสอบแบบ 2 ขั้นตอน (Two-stage Strategies)

3.2 การทดสอบหลายขั้นตอน (Muti-stage Strategies)

การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน เป็นการทดสอบที่มีโครงสร้างของแบบทดสอบเป็นรูปพีระมิดโดยใช้ค่าความยากของข้อสอบเป็นเกณฑ์ ในการเรียงข้อสอบที่มีความยากของข้อสอบในระดับปานกลางจะอยู่ยอดของพีระมิด การดำเนินการทดสอบจะเริ่มจากให้ผู้สอบตอบข้อสอบข้อเริ่มต้นที่มีความยากของข้อสอบในระดับปานกลาง ถ้าผู้สอบตอบถูกข้อสอบข้อต่อไปจะยากขึ้นและถ้าผู้สอบตอบผิดข้อต่อไปก็จะง่ายลง การทดสอบจะดำเนินการเช่นนี้

ไปเรื่อยๆ จนกระทั้งถึงเกณฑ์การยุติการทดสอบตามที่กำหนดไว้การทดสอบจึงสิ้นสุดลง การทดสอบหลายขั้นตอน สามารถแยกเป็น 2 ประเภท ดังนี้

3.2.1 การทดสอบแยกทางแบบคงที่ (Fixed-Branching)

การทดสอบแยกทางแบบคงที่เป็นรูปแบบการทดสอบด้วยข้อสอบที่มีลักษณะเฉพาะตัวที่มีหลายขั้นตอน ซึ่งแต่ละขั้นตอนจะมีข้อเดียวหรือหลายข้อก็ได้ โดยมีรูปแบบหรือเส้นทางการทดสอบที่แน่นอน สามารถสรุปได้เป็น 3 รูปแบบ ดังนี้

ก. รูปแบบพีระมิด (Pyramid Model) เป็นการทดสอบด้วยข้อสอบเฉพาะตัว ซึ่งมีการจัดรูปแบบการทดสอบเป็นรูปสามเหลี่ยม การจัดการทดสอบในรูปพีระมิดนั้น มีการจัดการทดสอบหลายรูปแบบ ดังนี้

1) รูปแบบพีระมิดขนาดชั้นคงที่ (Constant Step Size Pyramidel Model)

2) รูปแบบพีรมิดขนาดขั้นแปรผัน (Variable Step Size Pyramidal Model)

3) รูปแบบพีระมิดแบบโรบิน-มอนโร (Robin-Monro Pyramidal Model)

4) รูปแบบพีระมิดข้งตัด (Truncated Pyramidal Model)

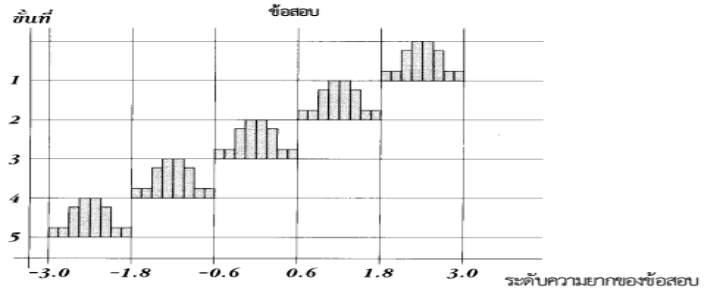
5) รูปแบบพีระมิดหลายข้อในแต่ละชั้น (Multiple-Item Pyramidal Model)

6) รูปแบบพีระมิดแบบให้น้ำหนักตัวเลือกเพื่อแยกทาง (Differential Response Option Branching Pyramidal Model)

ข. รูปแบบยืดหยุ่น (Flevileve Model)

ค. รูปแบบปรับระดับชั้น (Stradaptive Model)

การทดสอบของรูปแบบปรับระดับชั้น (stradptive Model) เป็นรูปแบบการทดสอบที่พัฒนามาจากการทดสอบในรูปแบบยืดหยุ่น (Flexilever Model) การทดสอบรูปแบบนี้จะมีข้อสอบในแต่ละระดับความยากของข้อสอบมากกว่า 1 ข้อ ซึ่งเรียกว่า ระดับชั้น (Steatum) โดยที่ชั้นที่ 1 บรรจุกลุ่มข้อสอบที่มีค่าความยากของข้อสอบน้อยที่สุด ชั้นที่ 2 บรรจุด้วยข้อสอบที่มีค่าความยากของข้อสอบเฉลี่ยสูงกว่าชั้นที่ 1 ชั้นอื่น ๆ บรรจุกลุ่มข้อสอบมีค่าความยากของข้อสอบเฉลี่ยสูงขึ้นเรื่อย ๆ จนถึงชั้นสูงสุด ในการทดสอบผู้เข้าสอบจะเริ่มทำข้อสอบที่ระดับใดก่อนก็ได้ ถ้าผู้เข้าสอบตอบถูก จะทำข้อสอบข้อถัดไปที่ยากมากขึ้น หากตอบผิด จะทำข้อสอบที่ง่ายลงในชั้นนั้น สำหรับการปรับระดับขึ้นลงนั้น ขึ้นอยู่กับเกณฑ์ที่ตั้งไว้และการทดสอบจะยุติลง ก็ต่อเมื่อผู้เข้าสอบไม่สามารถทำข้อสอบในระดับชั้นได้ถูกต้องเลยหรือสิ้นสุดระดับชั้นที่กำหนดไว้ แสดงได้ดังภาพที่ 2-8



ภาพที่ 2-8 โครงสร้างการจัดการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์รูปแบบปรับระดับชั้น (Stradaptive Model)

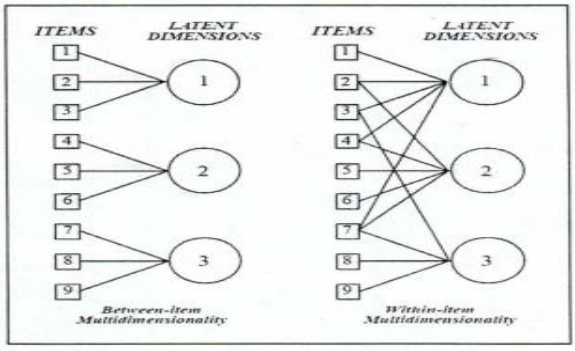
3.2.2 การทดสอบแยกทางแบบแปรผัน (Variable Branching)

การทดสอบแยกทางแบบแปรผันเป็นรูปแบบการทดสอบที่ไม่ได้กำหนดโครงสร้างที่แน่นอนว่า ถ้าตอบข้อสอบถูกหรือตอบข้อสอบผิดจะต้องทำข้อสอบข้อถัดไปในทิศทางใดแต่เป็นการทดสอบเมื่อผู้สอบตอบข้อสอบข้อหนึ่งแล้วจะนำผลการตอบนั้นไปประมาณค่าความสามารถของผู้สอบก่อน จึงเลือกข้อสอบข้อถัดไปที่มีความเหมาะสมกับระดับความสามารถของผู้สอบขณะนั้น การทดสอบจะดำเนินการเช่นนี้จนกว่าจะประเมินระดับความสามารถของผู้สอบเป็นไปตามเกณฑ์การยุติการทดสอบที่กำหนด ขึ้นอยู่กับการออกแบบการทดสอบ

โดยการประมาณความสามารถของผู้สอบที่เหมาะสมกับการทดสอบแบบปรับเหมาะที่มีรูปแบบการทดสอบแยกทางแบบแปรผัน ซึ่งวิธีการประมาณค่าความสามารถที่นิยมใช้มี 2 วิธี ได้แก่ 1) การประมาณค่าความสามารถแบบเบส์ (Bayesian Estimation) และ 2) การประมาณค่าแบบภาวะน่าจะเป็นสูงสุด (Maximum Likelihood Estimation)

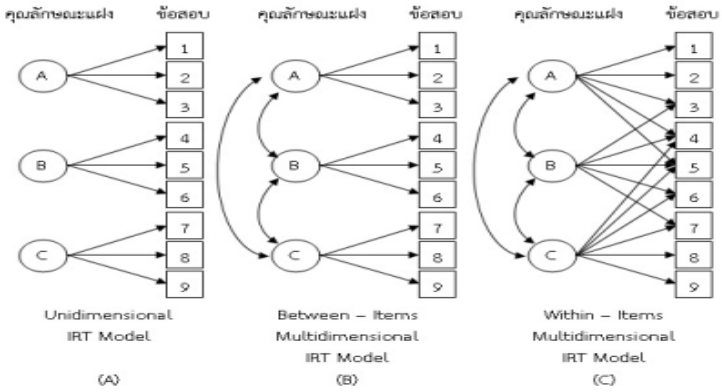
3.3 ประเภทของ Multidimensionality

ประเภทของ multidimensionality แบ่งออกเป็นการทดสอบพหุมิติระหว่างข้อ (between-items Multidimensionality) และการทดสอบพหุมิติภายในข้อ (within-items multidimensionality) (Wilson & Hoskens, 2005; Wu et al., 2007; Wang et al., 2011; Baghaei, 2012) โดยที่ข้อสอบจะ ถูกเรียกว่าเอกมิติ (Unidimensional) เมื่อข้อสอบนั้นมุ่งที่จะวัดคุณลักษณะเดียว และเรียกว่าพหุมิติ (Multidimensional) เมื่อข้อสอบนั้นมุ่งที่จะวัดคุณลักษณะหลายคุณลักษณะ และการทดสอบเป็นการทดสอบพหุมิติระหว่างข้อ (Between-Items Multidimensionality) เมื่อข้อสอบแต่ละข้อไม่ได้วัดหลายคุณลักษณะ และการทดสอบนั้นจะเป็นการทดสอบพหุมิตภายในข้อ (Within-Items Multidimensionality) เมื่อข้อสอบแต่ละข้อวัดหลายคุณลักษณะ แสดงลักษณะของการทดสอบพหุมิติระหว่างข้อและการทดสอบพหุมิติภายในข้อ สำหรับลูกศรสองทางระหว่างคุณลักษณะแฝงบ่งชี้ความสัมพันธ์ของคุณลักษณะแฝงที่ไม่เป็น 0 และจะเห็นได้ว่าการทดสอบพหุมิติระหว่างข้อเป็นลักษณะพิเศษของการทดสอบพหุมิติภายในข้อ โดยที่แต่ละข้อถูกกำหนดให้วัดเพียงคุณลักษณะเดียว การทดสอบพหุมิติระหว่างข้อ และการทดสอบพหุมิติภายในข้อมีลักษณะคล้ายกับโครงสร้างปัจจัย “Simple” และ “Complex” ในการวิเคราะห์ปัจจัย (Factor Analysis) ขั้นตอนการทดสอบแบบปรับเหมาะถูกพัฒนามาจากโมเดลการตอบสนองข้อสอบพหุมิติระหว่างข้อ และภายในข้อจะถูกกล่าวถึงในชื่อ Between-Items MAT และ Within-Items ซึ่งการศึกษาของ Luecht (1996) และ Segall (1996) เป็นตัวอย่างการศึกษาของ Wang and Chen (2004) ที่ศึกษาการทดสอบแบบปรับเหมาะแบบ Between-Items MAT ในลักษณะของการให้คะแนนสองค่า ซึ่ง Li (2006) เรียกรูปแบบนี้ว่า โครงสร้างอย่างง่าย (Simple Structure) แสดงได้ดังภาพที่ 2-9



ภาพที่ 2-9 ลักษณะของ Between-Items Multidimensionality และ Within-Items Multidimensionality (Baghaei, 2012 cited in Wu et al., 2007, p. 239)

จากภาพที่ 2-9 จะเห็นได้ว่าลักษณะของ Between-Items Multidimensionality นั้น มีความคล้ายคลึงกับการทดสอบแบบเอกมิติ เพราะข้อสอบแต่ละข้อวัดเพียงคุณลักษณะเดียวอย่างไรก็ตามความแตกต่างของ Between-Items Multidimensionality และแบบทดสอบเอกมิติ อยู่ที่ความสัมพันธ์ของคุณลักษณะที่นำไปใช้ในการทดสอบ เนื่องจากการทดสอบแบบเอกมิติ คุณลักษณะที่นำมาใช้ในการทดสอบจะไม่มีความสัมพันธ์กัน แต่การทดสอบแบบพหุมิติแบบ Between-Items Multidimensionality คุณลักษณะจะมีความสัมพันธ์กัน ซึ่งแสดงได้ดังภาพที่ 2-10



ภาพที่ 2-10 ลักษณะของ Unidimensional และ Multidimensional (Cheng, Wang, &, Ho, 2009, p. 378)

3.4 โมเดลการตอบสนองข้อสอบแบบพหุมิติ

การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ (Multidimensional Computerized Adaptive Testing : MCAT) สามารถใช้งานร่วมกับโมเดลการตอบสนองข้อสอบแบบพหุมิติ (Multidimensional Item Response Theory Model: MIRT) ได้หลากหลายโมเดล (Frey & Seitz, 2009) เช่น โมเดลโลจิสติกแบบพหุมิติ (Multidimensional Logistic Model) โมเดลปกติสะสมแบบพหุมิติ (Multidimensional Normal Ogive Model) โมเดลสำหรับการวัดมิติ

ที่ไม่สามารถทดแทนได้ (Models for Noncompensatory Dimensional) โมเดลสำหรับการวัดการเรียนรู้และการเปลี่ยนแปลง (Models for Learning and Change) โมเดลในการระบุโครงสร้างของระดับคุณลักษณะ (Models with Specified Trait Level Structures) และโมเดลสำหรับการจำแนกกลุ่มบุคคล (Models for Distinct Classes of Persons) โดยทั่วไปในการระบุความน่าจะเป็นของโมเดลการตอบสนองข้อสอบแบบพหุมิติ สำหรับการตอบสนองกับข้อสอบที่ถูกต้อง i ($u_i = 1$) ขึ้นอยู่กับความสามารถแฝง p เมื่อ $\theta = (\theta_1, \theta_2, ..., \theta_p)$ และคุณลักษณะของข้อสอบที่ครอบคลุมด้วยพารามิเตอร์ข้อสอบหนึ่งพารามิเตอร์หรือมากกว่า (Frey & Seitz, 2009) โดย Segall (1996, 2010, p. 62) นำเสนอ โมเดลการระบุความน่าจะเป็นในการตอบที่ถูกต้องของโมเดลโลจิสติกแบบพหุมิติ ซึ่งคำนวณได้ดังสมการ 13

$$P_i(\theta) \equiv \Pr ob\left(U_{ij} = 1 \middle| \theta_{ij}\right) = c_i + \frac{1 - c_i}{1 + \exp\left[-Da_i'\left(\theta - b_i 1\right)\right]} \tag{13}$$

เมื่อ

$-Da_i\left(\theta - b_i 1\right)$ แทน $-D\sum_{k-1}^{p} a_{ki}\left(\theta_k - b_i\right)$

U_{ij} แทน ตัวแปรสุ่มที่มีสองค่า (0และ1) จากข้อสอบข้อที่ i จากผู้สอบคนที่ j

D แทน ค่าคงที่ เท่ากับ 1.702

a_{ij}' แทน เวกเตอร์ 1 x p ของพารามิเตอร์อำนาจจำแนกของข้อสอบข้อที่ i

b_i แทน พารามิเตอร์ความยากของข้อสอบข้อที่ i

c_i แทน พารามิเตอร์โอกาสการเดาของข้อสอบข้อที่ i

1 แทน เวกเตอร์ p x 1 เมื่อ p คือ จำนวนมิติ

โมเดลการระบุความน่าจะเป็นในการตอบที่ถูกต้องของโมเดลปกติสะสมแบบพหุมิติ ซึ่งคำนวณได้ดังสมการ 14 (Bock & Schilling, 2003; Reckase, 2009, p. 95)

$$\begin{aligned} P_i(\theta) \equiv \Pr ob\left(U_{ij} = 1 \middle| \theta_j\right) &= c_j + \left(1 - c_j\right)\frac{1}{\sqrt{2\pi}}\int_{-z_j(\theta_j)}^{\infty} \exp\left(-t^2/2\right)df \\ &= c_i + \left(1 - c_j\right)F + z_j\left(q_j\right) \end{aligned} \tag{14}$$

เมื่อ $z_i(\theta)$ แทน $d_i + a_i\theta_j = d_i + a_i\theta_{j1} + a_{i2}\theta_{j2} + ... + a_{ip}\theta_{ip}$

F แทน เมทริกซ์ความแปรปรวนร่วมของมิติขนาด p x p

c_i แทน พารามิเตอร์โอกาสการเดาของข้อสอบข้อที่ i

d_i แทน Easiness Intercept ของข้อสอบข้อที่ i

a_i แทน เวกเตอร์อำนาจจำแนกของข้อสอบข้อที่ i

4. องค์ประกอบของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ เป็นการขยายมิติความสามารถของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบเอกมิติสู่พหุมิติ ซึ่งการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติด้วยคอมพิวเตอร์ ประกอบด้วยขั้นตอน 5 ขั้นตอน ได้แก่ 1) การสร้างคลังข้อสอบ (Item pools) 2) การเริ่มต้นการทดสอบ (Initiation/Starting point) 3) การคัดเลือกข้อสอบข้อถัดไป (Item selection algorithm/Item selection) 4) การประมาณความสามารถ (Ability Estimation/Scoring algorithm) และ 5) เกณฑ์การยุติการทดสอบ (Termination criterion/Stopping rules) (Veldkamp & Matteucci, 2013; Thompson & Weiss, 2011) โดยแต่ละขั้นตอนมีรายละเอียด ดังนี้

4.1 คลังข้อสอบ (Item Pools)

คลังข้อสอบ (Item Pools, Item Bank) เป็นแหล่งรวบรวมข้อสอบหรือคำถามแบบทดสอบ การทดสอบด้วยแบบทดสอบดั้งเดิม ผู้สอบทั้งหมดจะได้รับข้อสอบชุดเดียวกัน แต่สำหรับการทดสอบแบบปรับเหมาะผู้สอบจะได้รับข้อสอบเหมาะสมกับความสามารถของผู้สอบแต่ละคน คลังข้อสอบจึงเป็นเสมือนแหล่งรวบรวมชุดแบบทดสอบจำนวนมากและมีความเป็นคู่ขนานกัน ข้อสอบบรรจุในคลังข้อสอบมีลักษณะค่าความยากกระจายเต็มระดับความสามารถของประชากรผู้สอบ ค่าอำนาจจำแนกสูงเพื่อให้ประสิทธิภาพการทดสอบสูงสุด และค่าโอกาสในการเดาเข้าใกล้หรือเท่ากับศูนย์ ข้อสอบควรมีจำนวนมากเพียงพอและกระจายทุกระดับความสามารถของผู้สอบ ดังนั้นต้องใช้ข้อสอบจำนวนมากสำหรับสร้างคลังข้อสอบ ทำให้คลังข้อสอบสำหรับการทดสอบแบบปรับเหมาะมีขนาดใหญ่กว่าการทดสอบดั้งเดิม (Thompson & Weiss, 2011; Babcock & Weiss, 2012)

Molina, Pareja and Sanmartin (2008) ได้นำเสนอขั้นตอนการสร้างคลังข้อสอบที่จะนำมาใช้ในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ เพื่อให้มีประสิทธิภาพมากขึ้น ประกอบด้วย 5 ขั้นตอน ดังนี้ 1) กำหนดกฎเกณฑ์ คำอธิบายต่างๆ รวมถึงคำสั่งที่เกี่ยวข้อง ที่คาดว่าจะส่งผลทาง จิตวิทยาระหว่างดำเนินการทดสอบ 2) คัดเลือกข้อสอบที่เหมาะสมและตรงกฎเกณฑ์ที่กำหนดในคลังข้อสอบที่ได้จาก การบริหารจัดการทดสอบโดยคอมพิวเตอร์หรือการจัดการทดสอบโดยใช้กระดาษคำตอบ 3) พัฒนาในฐานข้อมูลของคลังข้อสอบ จากชุดข้อสอบที่คัดเลือกจากการจัดการทดสอบโดยใช้คอมพิวเตอร์หรือการจัดการทดสอบโดยใช้กระดาษคำตอบ เพื่อนำมาใช้ในการบริหารการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ 4) แยกข้อสอบที่ได้จากการบริหารจัดการทดสอบโดยใช้คอมพิวเตอร์หรือการจัดการทดสอบโดยใช้กระดาษคำตอบ โดยนำข้อสอบที่คัดเลือกแล้วไปจัดเก็บคนละส่วนกันและนำข้อสอบที่คัดเลือกมาแล้วมาตัดสินใจคัดเลือกอีกครั้ง จึงนำไปพัฒนาเป็นคลังข้อสอบที่จะนำไปใช้และ 5) วิเคราะห์ความแตกต่างของผลกระทบทางจิตวิทยาของกลุ่มข้อสอบที่ได้จากการบริหารจัดการทดสอบโดยใช้คอมพิวเตอร์หรือการจัดการทดสอบโดยใช้กระดาษคำตอบกับ

กลุ่มข้อสอบในคลังข้อสอบที่พัฒนาขึ้น ซึ่งสามารถนำไปเป็นข้อมูลในการพัฒนาปรับปรุงคลังข้อสอบใหม่ให้มีคุณภาพมากยิ่งขึ้น

การสร้างข้อสอบจำนวนมากใช้สำหรับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ขั้นตอนดังนี้ ได้แก่ 1) สร้างข้อสอบใหม่ให้มีจำนวนเพียงพอในแต่ละเนื้อหาเน้นที่ครอบคลุมเนื้อหาหลักและกฎเบื้องต้นในการเขียนข้อสอบ 2) ตรวจทานคุณภาพของข้อสอบเพื่อให้ข้อสอบมีคุณภาพสูงสุดเท่าที่เป็นไปได้ 3) ทำการทดสอบเบื้องต้นเพื่อตรวจสอบข้อบกพร่องของข้อสอบใหม่ 4) ทำการเลือกชุดย่อยของข้อสอบใหม่ ตอบข้อสอบ และวิเคราะห์ผลสอบตามทฤษฎีการทดสอบแบบมาตรฐานเดิมและทฤษฎีการตอบข้อสอบ เพื่อขจัดข้อสอบไม่เหมาะสมออกไปหรือทำการปรับปรุงข้อสอบใหม่ 5) เปรียบเทียบสมดุลเนื้อหาของคลังข้อสอบด้วยแบบทดสอบตามทฤษฎีการทดสอบแบบมาตรฐานเดิม เพื่ออ้างไปยังสมดุลเนื้อหาของคลังข้อสอบในการทดสอบแบบปรับเหมาะ ประเมินระบบการทดสอบด้วยการจำลองพฤติกรรมการตอบข้อสอบของผู้สอบระดับความสามารถต่างๆ กัน คะแนนทั้ง 2 การทดสอบนี้ต้องเปรียบเทียบกันได้ และ 6) นำแสดงข้อสอบตามแนวทางการจัดการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ต่อไป (Wainer et al., 2001; Wang, Kuo, & Chao, 2010)

การสร้างคลังข้อสอบมากกว่าหนึ่งคลังข้อสอบหรือแบ่งคลังข้อสอบใหญ่เป็นคลังข้อสอบย่อยสามารถสุ่มเลือกคลังข้อสอบและหมุนเวียนการใช้ชุดแบบทดสอบไปตามคลังข้อสอบต่างๆ ได้ (Wang & Kolen, 2001) การสร้างคลังข้อสอบจำนวนมากต้องทำการศึกษาเปรียบเทียบระหว่างคลังข้อสอบก่อนนำไปใช้ พิจารณาปัจจัยจากต่าง ๆ ดังนี้ 1) ขนาดของคลังข้อสอบ 2) จำนวนคลังข้อสอบเมื่อใช้พร้อม ๆ กัน 3) การทับซ้อนของคลังข้อสอบ 4) วิธีเข้าสู่แต่ละคลังข้อสอบ 5) การหมุนเวียนใช้คลังข้อสอบและเกณฑ์การออกจากคลังข้อสอบ 6) กฎการนำข้อสอบกลับมาใช้ใหม่ 7) การทดสอบข้อสอบก่อนนำมาใช้ วิธีการคำนวณข้อสอบ และวิธีการควบคุมช่วงความสามารถบนเส้นต่อเนื่อง และ 8) วิธีการเลือกข้อสอบเพื่อให้ครอบคลุมแต่ละเนื้อหาวิชา ควบคุมการแสดงข้อสอบอัตราการทับซ้อนและอื่นๆ (Wainer et al., 2001)

วิธีเลือกข้อสอบของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ให้สามารถเปรียบเทียบได้กับการทดสอบใช้กระดาษและดินสอ มีความจำเป็นต้องจัดสมดุลเนื้อหาเพื่อใช้เป็นตัวแทนของแบบทดสอบถ้าการทดสอบใช้กระดาษและดินสอเป็นพหุมิติ แต่การทดสอบแบบปรับเหมาะข้อตกลงกำหนดว่าคลังข้อสอบต้องเป็นมิติเดียว การแก้ปัญหานี้อาจใช้วิธีการแบ่งคลังข้อสอบใหญ่เป็นคลังข้อสอบย่อยคลังข้อสอบย่อยแต่ละคลังแทนแต่ละมิติ คะแนนแต่ละส่วนมาจากแต่ละมิติหรือแต่ละเนื้อหาผลการทดสอบหาโดยคำนวณรวมน้ำหนักของคะแนนต่างๆในแต่ละมิติ แต่วิธีนี้อาจยุ่งยากในการทดสอบและการรายงานผลตามมา (Wang, Kuo, & Chao, 2010)

สำหรับประเภทของคลังข้อสอบ แบ่งตาม Zhang, Chang and Yi (2012) เป็น 2 กลุ่ม ได้แก่ 1) คลังข้อสอบมิติเดียว (Single Item Pools) และ 2) คลังข้อสอบหลายมิติ (Multiple Item Pools) มีรายละเอียดดังนี้

1) คลังข้อสอบมิติเดียว (Single Item Pools) มีรูปแบบการทดสอบแบบปรับเหมาะที่เหมาะสมเกิดได้ ก็ต่อเมื่อข้อสอบมีมิติเด่นมิติเดียว (Dominant Dimension) มีรูปแบบการจัดคลังข้อสอบได้หลายวิธี เช่น 1.1) P-Optimal, D-Optimal ซึ่งมีงานวิจัยของ Kuo, Daud, and Yang , (2015); Seo and Weiss, (2015); Mao, (2014); He and Reckase, (2014); Zhang,

Chang, and Yi, (2012); Zhou, (2012); Diao and Reckase, (2009) 1.2) Polytomous ซึ่งมีงานวิจัยของ Zhou and Reckase, (2014)

2) คลังข้อสอบหลายมิติ (Multiple Item pools) ลักษณะข้อสอบจะมีมิติเด่น 2 มิติ หรือข้อสอบพหุมิติมีแนวโน้มว่าให้ค่าอำนาจจำแนกต่ำ และการเลือกข้อสอบอำนาจจำแนกสูงมีแนวโน้มให้ความเป็นมิติเดียวโดยตัวของมันเองอยู่แล้ว มีรูปแบบการจัดคลังข้อสอบได้หลายวิธี เช่น 2.1) คลังข้อสอบแบบแบ่งชั้น (Stratification Item Pools) ซึ่งมีงานวิจัยของ Barrada, Abad, and Olea, (2014); Huo, (2010); Cheng et al., (2009) 2.2) คลังข้อสอบแบบหลายรายการ (Multiple Item pools) ซึ่งมีงานวิจัยของ Zhang, Chang, and Yi, (2012); Lim, (2011)

สามารถสรุปแนวทางการคัดเลือกข้อสอบที่เหมาะสมสำหรับสร้างคลังข้อสอบได้ดังนี้ (Olino et al., 2012; Pilkonis et al., 2011; Kocalevant et al., 2009; Walter, Becker, & Bjorner, 2007)

1) มิติของมาตรวัด (Dimensionality) ตรวจสอบได้โดยการวิเคราะห์องค์ประกอบเชิงสำรวจ (Exploratory Factor Analysis: EFA) ซึ่งอาจกำหนดค่าน้ำหนักองค์ประกอบ (Factor Loadings) ที่ต้องการเป็นเกณฑ์ในการคัดเลือกข้อสอบ เช่น หากน้ำหนักองค์ประกอบมีค่าน้อยกว่า 0.30 หรือ 0.4 ให้คัดข้อสอบนั้นออกจากคลังข้อสอบ

2) การวิเคราะห์องค์ประกอบเชิงยืนยัน (Confirmatory Factor Analysis: CFA) โดยพิจารณาคัดเลือกข้อคำถามที่มีค่าน้ำหนักองค์ประกอบน้อยกว่า 0.4 ออกจากคลังข้อสอบ

3) ค่าสหสัมพันธ์เศษเหลือ (Residual Correlations) ซึ่งได้จากการวิเคราะห์องค์ประกอบเชิงยืนยัน โดยสามารถกำหนดเกณฑ์ค่าสหสัมพันธ์เศษเหลือที่ยอมรับได้ตามความเหมาะสม โดยทั่วไปนิยมกำหนดเกณฑ์ในการพิจารณาคัดข้อสอบออกจากคลังข้อสอบเมื่อข้อสอบนั้นมีค่าสหสัมพันธ์เศษเหลือมากกว่า 0.20 หรือ 0.25

4) การแจกแจงความถี่ (Frequency Distribution) ของผลคะแนนรวมของผู้สอบโดยเลือกข้อสอบที่มีความเหมาะสมกับการแจกแจงความถี่ของผลคะแนนรวมของผู้สอบ หากข้อสอบใดมีค่าพารามิเตอร์ตำแหน่งอยู่ในช่วงที่มีการแจกแจงความถี่ต่ำ จะพิจารณาคัดออกจากคลังข้อสอบ

5) ค่าสหสัมพันธ์ปรับแก้ระหว่างข้อสอบกับมาตรวัด (Adjusted item-total correlations) ข้อสอบที่มีค่าสหสัมพันธ์ปรับแก้ระหว่างข้อสอบกับมาตรวัดต่ำกว่า 0.4 จะถูกตัดออกจากคลังข้อสอบ

6) โมโนโทนิคซิตี้ (Monotonicity) ซึ่งเป็นสิ่งที่แสดงให้เห็นว่าความน่าจะเป็นในการยืนยันคำตอบที่เพิ่มขึ้น เมื่อระดับคุณลักษณะแฝงสูงขึ้น ซึ่งสามารถตรวจสอบได้โดยใช้ค่าสถิตินอนพาราเมตริก (Nonparametric)

7) ความสามารถของมาตรวัด (Scalability) ตรวจสอบได้โดยคำนวณค่า Loevinger H coefficients ซึ่งหากมีค่าต่ำกว่า 0.30 ให้พิจารณาคัดข้อสอบนั้นออกจากคลังข้อสอบ

การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์จะต้องสร้างคลังข้อสอบที่มุ่งวัดคุณลักษณะหรือความสามารถของผู้สอบ และครอบคลุมระดับความสามารถของผู้สอบที่แตกต่างกันการกำหนดขนาดของคลังข้อสอบพิจารณาจากปัจจัยหลายด้าน ซึ่งปัจจัยในการออกแบบคลังข้อสอบที่ส่งผลต่อขนาดของคลังข้อสอบ เช่น วิธีการคัดเลือกข้อสอบ การควบคุมเนื้อหาตามมาตรวัดทางจิตวิทยา

กฎการยุติ การทับซ้อนของข้อสอบ การให้คะแนนแบบทดสอบ ความเป็นคู่ขนานกับแบบทดสอบแบบดั้งเดิม เป็นต้น แม้ว่าวิธีดำเนินการของการทดสอบปรับเหมาะด้วยคอมพิวเตอร์จะดีเพียงใดหากคลังข้อสอบที่พัฒนาขึ้น ไม่มีคุณภาพ การทดสอบครั้งนี้ก็ถือว่าไม่มีคุณภาพเช่นกัน ดังนั้น การออกแบบและพัฒนาคลังข้อสอบสำหรับการทดสอบแบบปรับเหมาะจึงเป็นสิ่งที่สำคัญที่ควรคำนึงถึง (Reckase, 2009, pp. 336-337)

ขนาดคลังข้อสอบมีส่วนช่วยเพิ่มประสิทธิภาพการทดสอบ Weiss (1985) แนะนำว่าการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ จะให้ประสิทธิภาพเมื่อคลังข้อสอบมีจำนวนข้อสอบมากเพียงพอสำหรับจำแนกผู้สอบ คลังข้อสอบที่ดีจะต้องมีจำนวนข้อสอบเท่ากับ 2^n เมื่อ n คือ จำนวนข้อสอบที่ต้องการใช้ทดสอบ เช่น ถ้าต้องการทดสอบประมาณ 10 ข้อ ต้องมีข้อสอบในคลัง 1,024 ข้อ แต่ถ้าหากต้องการทดสอบแบบกำหนดจำนวนข้อที่จะใช้คงที่เหมือนกันทุกคน จำนวนข้อสอบในคลังข้อสอบควรมีอย่างน้อย $n(n+1)/2$ ข้อ เช่น ถ้าต้องทดสอบ 10 ข้อ ต้องมีข้อคำถามในคลัง จำนวน 55 ข้อ การศึกษาของ Xing and Hambleton (2004) คลังข้อสอบขนาดใหญ่ทำให้สารสนเทศแบบทดสอบเพิ่มขึ้นตามไปด้วย สอดคล้องกับความคิดเห็นของ Lord (1980) ซึ่งอ้างถึงการศึกษาในสถานการณ์จำลอง เมื่อใช้ความยาวแบบทดสอบ 25 ข้อ ขนาดคลังข้อสอบ 363 ข้อ และ 183 ข้อ ปรากฏว่า ขนาดคลังข้อสอบ 363 ข้อสามารถให้สารสนเทศได้มากกว่า เป็นเช่นนี้ เพราะคลังข้อสอบขนาดใหญ่มีโอกาสเลือกข้อสอบดีที่สุดจัดให้ผู้สอบได้มากกว่าคลังข้อสอบขนาดเล็ก ต่อมา Segall (2005) เสนอว่า ขนาดคลังข้อสอบควรมีขนาดประมาณ 6-8 เท่าของความยาวแบบทดสอบ ส่วน de Ayla (2009, p. 376) เสนอว่า คลังข้อสอบควรมีขนาดประมาณ 8–12 เท่า ของจำนวนข้อสอบที่ใช้ในการทดสอบ กล่าวคือ ถ้าในการทดสอบใช้ข้อสอบประมาณ 25 ข้อ คลังข้อสอบควรมีข้อสอบ 200–300 ข้อ และ Weiss (2011) เสนอ ให้มีข้อสอบไม่น้อยกว่า 200 ข้อ Thompson and Weiss (2011) เสนอ ให้มีข้อสอบประมาณ 400 ข้อ และ Embretson and Reise (2000, p. 264) กล่าวว่าคลังข้อสอบควรประกอบด้วยข้อสอบประมาณ 100 ข้อ ซึ่งหากรูปแบบการตอบข้อคำถามเป็นแบบที่มีการตรวจให้คะแนนมากกว่า 2 ค่า จำนวนข้อสอบในคลังข้อสอบสามารถมีจำนวนน้อยกว่านี้ได้

การพัฒนาคลังข้อสอบสำหรับการทดสอบแบบปรับเหมาะนั้น มีความคล้ายคลึงกับคลังข้อสอบแบบดั้งเดิม ข้อสอบจะต้องสร้างตามตารางกำหนดเนื้อหาและควรมีการตรวจสอบคุณภาพของเนื้อหาและความไวของแบบทดสอบ (Test Sensitivity) และมีการทดลองใช้เพื่อศึกษาคุณลักษณะทางจิตวิทยา (Wainer et al., 2000, pp. 105-112) วิธีการตรวจสอบคุณภาพของข้อสอบสามารถใช้วิธีการทางสถิติของทฤษฎีการทดสอบแบบดั้งเดิม เช่น สัดส่วนของผู้ที่ตอบถูก (Proportion Correct) หรือ สหสัมพันธ์แบบไบซีเรียล (Biserial Correlation) ร่วมกับทฤษฎีการทดสอบแนวใหม่ เช่น พารามิเตอร์ของข้อสอบ (Item Parameters) และค่าสารสนเทศของข้อสอบ (Item Information) (Wainer, 1989) และสิ่งที่สำคัญอีกประการหนึ่ง คือ คุณลักษณะของข้อสอบ (Item Characterisitic) เพื่อตรวจสอบว่าข้อสอบแต่ละข้อเหมาะสมกับโมเดลการวัด ซึ่งการพัฒนาคลังข้อสอบสำหรับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ มีขั้นตอนดังนี้ (Wainer et al., 2001, pp. 37-45; Thompson & Weiss, 2011)

ขั้นที่ 1 สร้างข้อสอบให้มีจำนวนมากพอที่จะครอบคลุมเนื้อทั้งหมดตามกฎเบื้องต้นในการเขียนข้อสอบ

ขั้นที่ 2 ตรวจสอบคุณภาพของข้อสอบเพื่อให้ข้อสอบมีคุณภาพสูงสุด

ขั้นที่ 3 นำข้อสอบที่สร้างขึ้นไปทดลองใช้

ขั้นที่ 4 คัดเลือกข้อสอบที่มีคุณภาพด้วยการวิเคราะห์ตามหลักสถิติจากการทดลองใช้

ขั้นที่ 5 ทำการตรวจสอบสัดส่วนขอข้อสอบในแต่ละเนื้อหา และประเมินกระบวนการทดสอบด้วยการจำลองสถานการณ์การทดสอบ (Simlation) เพื่อพิจารณาว่าข้อสอบครอบคลุมช่วงของคุณลักษณะแฝงที่ต้องการวัดหรือไม่

ขั้นที่ 6 ดำเนินการปรับข้อสอบที่เหมาะสมทั้งหมดให้อยู่ในรูปแบบของคอมพิวเตอร์ตามแนวทางการจัดการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์

Choppin (1985) และ, Rudner (1998, pp. 1-2) ได้สรุปประโยชน์ของคลังข้อสอบไว้ดังนี้

1) ทำให้ได้ข้อสอบที่ตรงกับจุดมุ่งหมายและเนื้อหาของแต่ละรายวิชา สามารถเลือกใช้ได้อย่างเหมาะสมกับสถานการณ์

2) ทำให้ได้ข้อสอบและแบบสอบที่มีคุณภาพน่าเชื่อถือเป็นจำนวนมาก ผลที่ได้จากการวัดก็จะมีความถูกต้องแม่นยำมากขึ้น

3) สร้างข้อสอบได้อย่างรวดเร็วและถูกต้อง มีความพร้อมในการสร้างแบบสอบได้ทุกเวลา

4) ใช้ข้อสอบซ้ำได้หลายครั้ง ด้วยการเลือกใช้ข้อสอบที่มีระบบ

5) ประหยัดงบประมาณ ทรัพยากร และเวลา ลดภาระด้านการออกข้อสอบให้แก่ครูผู้สอน

6) ได้ข้อสอบที่มีลักษณะเป็นข้อสอบคู่ขนานได้

7) สามารถจัดทำแบบสอบที่เหมาะกับความสามารถของกลุ่มผู้สอบที่เป็นเป้าหมายได้ตลอดจนสามารถสร้างข้อสอบแบบปรับเหมาะสำหรับบุคคลได้

8) กระตุ้นครูผู้สอนให้เกิดการพัฒนาข้อสอบให้มีคุณภาพและมีมาตรฐานที่สูงยิ่งขึ้น

จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องสามารสรุปได้ว่า คลังข้อสอบ หมายถึง แหล่งรวบรวบข้อสอบซึ่งสร้างตามหลักเกณฑ์ที่ถูกต้องและถูกเก็บไว้อย่างเป็นระบบ โดยมีหลักการในการสร้างคลังข้อสอบสำหรับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ จากการทบทวนเอกสารข้างต้นดังนี้ 1) สร้างข้อสอบให้มีจำนวนมากพอที่จะครอบคลุมเนื้อทั้งหมดตามกฎหลักการสร้างข้อสอบ 2) ตรวจสอบคุณภาพของข้อสอบเพื่อให้ข้อสอบมีความถูกต้องและคุณภาพสูงสุด 3) นำข้อสอบที่สร้างขึ้นไปทดลองใช้และวิเคราะห์คุณภาพข้อสอบตามเกณฑ์ที่กำหนด 4) ทำการตรวจสอบสัดส่วนขอข้อสอบในแต่ละเนื้อหา และประเมินกระบวนการทดสอบด้วยการจำลองสถานการณ์การทดสอบ (Simlation) เพื่อพิจารณาว่าข้อสอบครอบคลุมช่วงของคุณลักษณะแฝงที่ต้องการวัดหรือไม่ และ 5) ดำเนินการจัดข้อสอบที่ผ่านคุณภาพทั้งหมดให้อยู่ในรูปแบบฐานข้อมูล(Data base) ตามแนวทางการจัดการทอดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์

4.2 การเริ่มต้นการทดสอบ (Starting the First Item)

การกำหนดจุดเริ่มต้นการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์มีหลายวิธี โดยทั่วไปจะใช้ข้อสอบข้อแรกทีมีค่าความยากปานกลาง จัดให้กับผู้สอบแต่ละคน เมื่อไม่ทราบค่าสารสนเทศเบื้องต้นหรือระดับความสามารถของผู้สอบ จึงจำเป็นต้องพิจารณาว่าจะใช้เกณฑ์ใดในการเริ่มต้นการ

ทดสอบซึ่งแนวทางในการกำหนดเกณฑ์เริ่มต้นการทดสอบมีหลากหลายแนวทาง ดังนี้ (Becker et al., 2008; Thompson & Weiss, 2011; kirisci et al., 2012)

1) ใช้ข้อสอบที่มีค่าความยากปานกลางจัดให้กับผู้สอบแต่ละคนเมื่อไม่ทราบค่าสารสนเทศเบื้องต้นของผู้สอบ

2) ทำการทดสอบเบื้องต้นก่อนการทดสอบจริงและนำผลมากำหนดจุดเริ่มต้นการทดสอบ

3) หากไม่มีข้อมูลอื่นประกอบการพิจารณา นิยมใช้ค่าเฉลี่ยคุณลักษณะแฝงหรือระดับความสามารถของประชากร (Average Theta) ซึ่งมีค่าเท่ากับ 0.00 เป็นจุดเริ่มต้นในการเลือกข้อสอบข้อแรกให้เหมาะสมกับระดับความสามารถของผู้สอบดังกล่าว

4) ใช้วิธีการสุ่มเลือกข้อสอบด้วยโปรแกรมคอมพิวเตอร์

5) หากมีผลการทดสอบของมาตรวัดนั้นๆ ในครั้งก่อนหน้า อาจเริ่มต้นด้วยค่าคุณลักษณะแฝงหรือระดับความสามารถของผู้สอบที่ได้จากครั้งก่อน

6) พิจารณาจากข้อมูลอื่นๆ เช่น ผลการทดสอบจากมาตรวัดอื่น เกรดเฉลี่ย แรงจูงใจในการทดสอบ และเศรษฐานะทางสังคม เป็นต้น

4.3 การคัดเลือกข้อสอบ (Item Selection)

สำหรับการเลือกข้อสอบจากคลังข้อสอบเพื่อให้ผู้สอบตอบ โดยหลักการคัดเลือกข้อสอบมีเป้าหมายเพื่อลดความแปรปรวนของเวกเตอร์ความสามารถแบบพหุมิติชั่วคราว (Frey & Seitz, 2009) วิธีการคัดเลือกข้อสอบที่ใช้กับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติอาศัยภาวะน่าจะเป็นสูงสุดหรือต่ำสุดจากการประมาณค่าความสามารถ $(\boldsymbol{\theta})$ และมีวิธีการคัดเลือกข้อสอบอยู่หลากหลายวิธี แต่ละวิธีสามารถใช้ร่วมกันกับวิธีการประมาณค่าทุกแบบและยังสามารถใช้วิธีการคัดเลือกข้อสอบหลายวิธีร่วมกันได้อีกด้วย (Reckase, 2009, p. 327) โดยมีวิธีการคัดเลือกที่สำคัญ ดังนี้

4.3.1 Maximize the Determinant of the Fisher Information Matrix

การคัดเลือกข้อสอบด้วยเมทริกซ์สารสนเทศฟิชเชอร์ ดำเนินการภายใต้เงื่อนไขการประมาณค่าเวกเตอร์ความสามารถที่กำหนดจากเวกเตอร์ค่าเฉลี่ยเท่ากับ 0 และเวกเตอร์ความสามารถจริง (Ture $\boldsymbol{\theta}$) มีการแจกแจงแบบปกติหลายตัวแปรด้วยเมทริกซ์ความแปรปรวนและความแปรปรวนร่วม (Variance–Covariance Matrix) ที่เกี่ยวข้องกับเมทริกซ์สารสนเทศของฟิชเชอร์ (Fisher Information Matrix) (Segall, 1996; Frey & Seitz, 2009; Reckase, 2009, p. 330) โครงสร้างทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ มองเมทริกซ์สารสนเทศของฟิชเชอร์เป็นเหมือนการวัดปริมาณของสารสนเทศการตอบสนองของข้อสอบและนำเสนอเกี่ยวกับความสามารถภายหลังการตอบข้อสอบข้อนั้นแล้ว ($\boldsymbol{\theta}$) ในทางคณิตศาสตร์ค่าความคาดหวังของอนุพันธ์ที่สองของความน่าจะเป็นลอกลิฮู้ด ที่เกี่ยวข้องกับความสามารถจะมีค่าเป็นลบ (Frey & Seitz, 2009) สามารถแสดงในวิธีการทำงานแบบพหุมิติ ที่นำเสนอโดย Segall ในปี ค.ศ. 1996 ซึ่งคำนวณได้ดังสมการ 15 (Segall, 1996; Frey & Seitz, 2009; Reckase, 2009, p. 330)

$$I_{rs}\left(\theta,\hat{\theta}\right)=-E\left[\frac{\partial^2 \ln L}{\partial\theta_r,\partial\theta_s}\right] \tag{15}$$

เมื่อ θ แทน ความสามารถจริงของผู้สอบ

$\hat{\boldsymbol{\theta}}$ แทน ค่าประมาณจากภาวะน่าจะเป็นสูงสุด

$\boldsymbol{\theta}_r$ แทน องค์ประกอบของเวกเตอร์ความสามารถ (θ Vecter) ลำดับที่ r

$\boldsymbol{\theta}_s$ แทน องค์ประกอบของเวกเตอร์ความสามารถ (θ Vecter) ลำดับที่ s

L แทน ฟังก์ชันความน่าจะเป็นของเซตการตอบข้อสอบจาก MCAT เนื่องจากแต่ละองค์ประกอบของเมทริกซ์ มาจากระดับข้อสอบ

เมทริกซ์ความแปรปรวน-ความแปรปรวนร่วมในการประมาณค่าหลังจากการทดสอบแบบปรับเหมาะข้อที่ K จะถูกรวมด้วยการผกผันของเมทริกซ์สารสนเทศ ซึ่งคำนวณได้ดังสมการ 16 (Reckase, 2009, p. 330)

$$\sum\nolimits_k \{I(\theta, \hat{\theta}_k\}^{-1} \tag{16}$$

เมื่อ $\sum_k$ แทน ผลรวมของการผกผันของเมทริกซ์สารสนเทศ

$\hat{\theta}_k$ แทน ค่าประมาณความสามารถของผู้สอบหลังจากทำข้อสอบข้อที่ K

เมทริกซ์สารสนเทศที่ใช้ในการคำนวณค่าความสามารถ $\boldsymbol{\theta}$ ในโมเดล MIRT สำหรับเมทริกซ์องค์ประกอบแนวทแยง (the diagonal element of the matrix) ซึ่งคำนวณได้ดังสมการ 17 (Reckase, 2009, p. 330)

$$I_{rr}\left(\theta, \hat{\theta}_k\right) = \sum_{i=1}^{k} \frac{\left[\dfrac{\partial P_i(\theta)}{\partial \theta_i}\right]^2}{P_i(\theta) Q_i(\theta)} \tag{17}$$

หรือ คำนวณได้จากสมการ 18 Segall. (1996)

$$I_{rr}\left(\theta, \hat{\theta}\right) = -D^2 \sum\nolimits_{i \in v} \frac{a_{ri}^2 Q_i(\theta)\left[P_i(\theta) - c_i\right]\left[c_i P_i(\theta) - P_i^2(\theta)\right]}{P_i^2(\theta)\left(1 - c_i\right)^2} \tag{18}$$

เมื่อ D แทน ค่าคงที่ เท่ากับ 1.702

v แทน เวกเตอร์ที่บรรจุข้อสอบแบบปรับเหมาะ

a_{ri} แทน พารามิเตอร์อำนาจจำแนกของข้อสอบข้อที่ I ซึ่งถูกคัดเลือกให้ดำเนินการทดสอบในลำดับที่ r

c_i แทน พารามิเตอร์โอกาสการเดาข้อสอบได้ถูกต้องในการทำข้อสอบข้อที่ i

$P_i(\theta)$ แทน ฟังก์ชันการตอบสนองข้อสอบข้อที่ i

$Q_i(\theta) = 1 - P_i(\theta)$

สำหรับเมทริกซ์สารสนเทศของข้อสอบแสดงได้ด้วยองค์ประกอบทแยงมุม คำนวณได้ดังสมการที่ 19 (segall, 1996)

$$I_{rr}(\theta,u_i)=\frac{\left[\frac{\partial P_i(\theta)}{\partial\theta_r}\right]^2}{P_i(\theta)Q_i(\theta)} \tag{19}$$

และสำหรับองค์ประกอบทแยงมุมแบบปิด (the off diagonal element) คำนวณได้ดังสมการที่ 20 (Reckase, 2009, p. 330)

$$I_{rs}(\theta,\hat{\theta}_k)=\sum_{i=1}^{k}\frac{\left[\frac{\partial P_i(\theta)}{\partial\theta_r}\right]\left[\frac{\partial P_i(\theta)}{\partial\theta_s}\right]}{P_i(\theta)Q_i(\theta)} \tag{20}$$

หรือ คำนวณได้จากสมการ 21 (Segall, 1996)

$$I_{rs}(\theta,\hat{\theta})=-D^2\sum\nolimits_{i\in v}\frac{a_{rs}a_{ri}Q_i(\theta)\left[P_i(\theta)-c_i\right]\left[c_iP_i(\theta)-P_i^2(\theta)\right]}{P_i^2(\theta)(1-c_i)^2} \tag{21}$$

เมื่อ a_{si} แทน พารามิเตอร์อำนาจจำแนกของข้อสอบข้อที่ i ซึ่งถูกคัดเลือกให้ดำเนินการทดสอบในลำดับที่ s

จะได้สมการ 22 (Segall, 1996)

$$I_{rs}(\theta,\hat{\theta})=\sum\nolimits_{i\in v}\frac{\frac{\partial P_i(\theta)}{\partial\theta_r}\times\frac{\partial P_i(\theta)}{\partial\theta_s}}{P_i(\theta)Q_i(\theta)} \tag{22}$$

ลักษณะเฉพาะของเมทริกซ์ความแปรปรวน ความแปรปรวนร่วมที่ใช้ ในการประมาณค่าภายหลังการสอบข้อสอบข้อที่ k จากแบบทดสอบแบบปรับเหมาะ $\sum_k$ เป็นค่าประมาณ จากส่วนกลับของเมทริกซ์สารสนเทศ ซึ่งคำนวณได้ดังสมการ 23 (Segall, 1996)

$$\sum\nolimits_k=\left\{I\left(\theta,\hat{\theta}_k\right)\right\}^{-1}=\left[\sum_{i\in v}I\left(\theta,u_i\right)\right]^{-1} \tag{23}$$

และ $\sum_{k+(k+1)}$ เป็นการประมาณค่าของเมทริกซ์ความแปรปรวนร่วม ชั่วคราวของข้อสอบข้อที่ k รวมกับข้อสอบข้อที่ k+1 ซึ่งคำนวณได้ดังสมการ 24

$$\sum_{k+(k+1)} = \left[I\left(\theta, \hat{\theta}_k\right) + I\left(\theta, \mu_{k+1}\right) \right]^{-1} \quad (24)$$

เกณฑ์การเลือกข้อสอบที่เสนอโดย Segall (1996) จะขึ้นอยู่กับความสัมพันธ์ระหว่างเมทริกซ์สารสนเทศฟิชเชอร์กับช่วงความเชื่อมั่นที่ครอบคลุมการประมาณค่าความสามารถที่กำหนดโดย Anderson (1984) ปรากฏว่า มีความเท่าเทียมกับช่วงความเชื่อมั่นที่ครอบคลุมการประมาณค่าความสามารถจากตำแหน่งอ้างอิงรูปวงรีในพหุมิติ และปริมาณของรูปวงรีนี้ ก็คือ ฟังก์ชัน $\sum_k$ และ Segall (1996) แสดงให้เห็นว่าเมื่อ $\sum_k$ มีค่ามากที่สุดแล้วปริมาณของช่วงความเชื่อมั่นของการประมาณค่าความสามารถบริเวณรูปวงรี จะมีขนาดเล็กที่สุด ซึ่งคำนวณได้ดังสมการ 25

$$\left| I\left(\theta, \left(\hat{\theta}_k\right)\right) + I\left(\theta, \mu_{k+1}\right) \right| \quad (25)$$

จากสมการ 26 เป็นดีเทอร์มิแนนท์ของเมทริกซ์เทอมทางด้านซ้ายมือคือ เมทริกซ์สารสนเทศของข้อสอบที่ได้ทดสอบไปแล้วด้วยข้อสอบข้อที่ k ส่วนเทอมทางด้านขวาเป็น เมทริกซ์สารเทศของข้อสอบข้อที่จะถูกนำไปใช้สอบข้อต่อไป กระบวนการคัดเลือกข้อสอบที่จะ นำไปสอบข้อถัดไปดำเนินการโดยการระบุว่าข้อสอบข้อใดจะมีเมทริกซ์สารสนเทศสูงสุดเมื่อรวมกับ เมทริกซ์สารสนเทศของข้อสอบข้อที่ผ่านมา โดยจะสรุปจากผลรวมของการดีเทอร์มิแนนท์ที่มีค่ามากที่สุด

4.3.2 Largest Decrement in the Volume of the Bayesian Credibility Ellipsoid

วิธีการประมาณค่าแบบเบส์ ที่นำเสนอโดย Segall (1996, 2002, 2010, pp. 65-74) นำไปสู่การคัดเลือกข้อสอบที่จะลดปริมาณของข้อสอบได้อย่างมาก (Frey & Seitz, 2009; Reckase, 2009, pp. 331-332) โดยการเลือกข้อสอบที่มีค่าเมทริกซ์สารสนเทศภายหลังสูงสุด $I_{i|s_{k-1}}$ โดยใช้ข้อมูลจาก 3 แหล่ง คือ อินเวอร์สเมทริกซ์ความแปรปรวนร่วมที่เกิดขึ้นก่อน (Inverse Prior Covariance Matrix: φ^{-1}) ผลรวมของเมทริกซ์ W ที่ได้มาจากข้อสอบข้อที่ผ่านมา $\left(W_{s_{k-1}}\right)$ และเมทริกซ์ W ที่ได้มาจาก ข้อสอบข้อที่แล้ว จะถูกนำมาใช้สอบข้อต่อไป ซึ่งคำนวณได้ดังสมการ 26

$$I_{i|s_{k-1}} = \phi^{-1} + W_{s_{k-1}} + W_i \quad (26)$$

การประมาณค่าความสามารถชั่วคราวของ $\widehat{\theta}_k$ (เมื่อ k = 0) ด้วยการกำหนดเซตของค่าเฉลี่ยของการแจกแจงความสามารถเบื้องต้น โดยทั่วไปมักจะกำหนดให้มีค่าเท่ากับ 0

ขั้นตอนแรก คำนวณหาค่าอินเวอร์สของเมทริกซ์ความแปรปรวนร่วมที่เกิดขึ้นก่อน $\left(\phi^{-1}\right)$ เนื่องจากต้องใช้ในขั้นตอนอื่นๆ ที่จะดำเนินการต่อไป

ขั้นตอนที่สอง คำนวณหาค่า $W_{s_{k-1}}$ ซึ่งคำนวณได้ดังสมการ 27

$$W_{s_{k-1}} = \sum_{j \in s_{k-1}} W_j \tag{27}$$

เมื่อ W_j ของข้อสอบข้อที่ j คำนวณได้จากสมการ และผลรวมของเมทริกซ์ ที่ได้มาจากข้อสอบข้อที่ผ่านมา $\sum_{j \in s_k} W_j$ คำนวณจากข้อสอบทุกข้อที่ได้รับการเลือก มาแล้ว

ขั้นตอนที่สาม คำนวณหาค่าของเมทริกซ์ W ที่ได้มาจากข้อสอบข้อที่จะถูกนำมาใช้สอบข้อต่อไป ซึ่งคำนวณได้ดังสมการ 28

$$W_i = D^2 a_i a_i' w_i^* \tag{28}$$

เมื่อ $w_i^* = \left[\frac{q_i(\theta)}{p_i(\theta)}\right]\left[\frac{p_i(\theta) - c_i}{1 - c_i}\right]^2$

นอกจากนี้ ในการคัดเลือกข้อสอบข้อแรกสามารถเลือกข้อสอบที่มีค่า เมทริกซ์สารสนเทศสูงสุด $\left(I_i\right)$ ซึ่งคำนวณได้ดังสมการ 29 ดังนี้ (Segall, 2010, p. 70)

$$I_i = \phi^{-1} + W_i \tag{29}$$

4.3.3 Maximize Kullback-Leibler Information

วิธีการ Posterior Expected Kullback–Leibler Information ถูกนำเสนอ ให้ใช้ในการคัดเลือกข้อสอบในการทดสอบแบบปรับเหมาะที่เป็นมิติเดียว Chang and Ying (1996) และต่อมา Veldkamp and van der Linden (2002) ได้เสนอทางเลือกให้ใช้ วิธีการ Posterior Expected Kullback–Leibler Information ในการคัดเลือกข้อสอบแบบปรับเหมาะแบบพหุมิติ (MCAT) โดยมีเกณฑ์การคัดเลือกข้อสอบจากระยะห่างของการกระจายของการตอบสนอง จากข้อสอบข้อที่จะถูกเลือกให้ใช้ทดสอบข้อต่อไป เป็นตำแหน่งที่ใช้ประมาณค่าความสามารถของ ผู้สอบกับความสามารถที่แท้จริงของผู้สอบ $(\boldsymbol{\theta})$ ด้วยการคาดหมายเอาจากการตอบข้อสอบ (Wang, Chang, and Boughton, 2011, Mulder, van der Linden, 2009, p. 85; Reckase, 2009, pp. 334-335) กรณีข้อสอบ 1 ข้อ ซึ่งคำนวณได้ดังสมการ 30 และ 31

$$K_i\left(\hat{\theta}, \theta\right) = -E\left[\log \frac{f\left(u_i \middle| \hat{\theta}\right)}{f\left(u_i \middle| \theta\right)}\right] \tag{30}$$

และ

$$K_i\left(\hat{\theta}, \theta\right) = p_i\left(\hat{\theta}\right) \log \frac{P_i\left(\hat{\theta}\right)}{P_i(\theta)} + Q_i\left(\hat{\theta}\right) \log \frac{Q_i\left(\hat{\theta}\right)}{Q_i(\theta)} \tag{31}$$

ให้ $P_j(\theta)$ แทน ฟังก์ชันการตอบสนองข้อสอบสำหรับข้อที่ j ซึ่ง $\boldsymbol{\theta}$ และ θ_n แทน 2 ระดับคุณลักษณะ โดย Chang and Ying สังเกตว่ามีคุณลักษณะที่สำคัญหลายประการของฟังก์ชัน KL ได้แก่

1) $KL_j(\theta\|\theta_0) \neq KL(\theta_0\|\theta)$

2) $KL_j(\theta\|\theta_0) \geq 0$ และ $KL_j(\theta_0\|\theta_0) \geq 0$

3) คล้ายกับผลรวมของค่าฟังก์ชันสารสนเทศของฟิชเชอร์ กล่าวคือ สารสนเทศระดับคะแนนแบบสอบเป็นผลรวมของสารสนเทศระดับข้อสอบ คำนวณได้ดังสมการ 32

$$KL^n(\theta\|\theta_0) = \sum_{j=1}^{n} KL_j(\theta\|\theta_0) \tag{32}$$

ในขณะที่ GWIC คือ ฟังก์ชันถ่วงน้ำหนักที่สามารถประยุกต์ใช้กับ $KL_j(\theta\|\theta_0)$ เพื่อที่จะหาเกณฑ์การคัดเลือกข้อสอบ ซึ่งฟังก์ชันถ่วงน้ำหนักทั่วไป คำนวณได้ดังสมการ 33

$$w(\theta,\theta_0) = \begin{cases} 1, \theta \in (\hat{\theta}-\delta_1, (\theta+\delta_2)) \\ 0, otherwise \end{cases} \text{ and } \theta_0 \in (\hat{\theta}-\delta_3, \hat{\theta}+\delta_4) \tag{33}$$

ซึ่งการประมาณค่าความสามารถ $(\hat{\theta})$ เป็นความสามารถที่แท้จริง $(\boldsymbol{\theta})$ ตามข้อสอบ n ข้อ และพื้นที่ที่ใช้เป็นเกณฑ์ในการคัดเลือกข้อสอบ กล่าวคือ ข้อสอบที่มีพื้นที่สูงสุด ซึ่งเท่ากับสารสนเทศ $KL_j(\theta\|\theta_0)$ ที่เฉลี่ยสูงสุดจะถูกคัดเลือก สามารถคำนวณได้ดังสมการ 34

$$KL_j(\hat{\theta}) = \int_{\theta_l}^{\hat{\theta}_u} KL_j(\theta\|\hat{\theta})d\theta \tag{34}$$

เมื่อ $(\hat{\theta}_l, \hat{\theta}_u) = \left(\hat{\theta} - \frac{z}{\sqrt{n}}, \hat{\theta} + \frac{z}{\sqrt{n}}\right)$ และ $\frac{z}{\sqrt{n}}$ เป็นลิมิตความเชื่อมั่น (Confidence limit)

4.4 การประมาณค่าความสามารถ (Ability Estimation)

การประมาณค่าความสามารถของโมเดลการตอบสนองข้อสอบแบบพหุมิติ (MIRT) มีเกณฑ์การประมาณค่าที่แตกต่างกันหลายแนวทาง เช่น Maximum Likelihood Criterion, Maximum a Posteriori Bayesian Criterion, Newton-Raphson Method และ Least Squares Criterion (Reckase, 2009, p. 139) แต่โดยทั่วไปการประมาณค่าความสามารถของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ (MCAT) ที่นิยมมีอยู่ 2 วิธี คือ วิธีการประมาณค่าแบบภาวะน่าจะเป็นสูงสุด (Maximum Likelihood Estimation) และวิธีการประมาณค่าแบบเบส์ (Bayesian Estimation) (Reckase, 2009, p. 314) และจากข้อตกลงเบื้องต้นเกี่ยวกับความเป็นอิสระของข้อสอบและผู้สอบในทฤษฎีการตอบสนองข้อสอบ ถ้าตัวแปรความสามารถมีค่าคงที่การตอบสนองข้อสอบของผู้สอบต่อข้อสอบแต่ละข้อของแบบทดสอบ ถูกคาดหวังว่าเป็นอิสระจากกัน

ในเชิงสถิติ การประมาณค่าความสามารถด้วยเทคนิคการประมาณค่าแบบภาวะน่าจะเป็นสูงสุด (Maximum Likelihood Estimation) ก่อให้เกิดความไม่ชัดเจนใน การประมาณค่าของเวกเตอร์ความสามารถ ณ จุดเริ่มต้นของการทดสอบ แต่วิธีการประมาณค่าแบบเบส์จะหลีกเลี่ยงข้อบกพร่องนี้ แต่ก็ยังไม่มีข้อสรุปเกี่ยวกับปัญหาใน การประมาณค่าความสามารถ ในทางปฏิบัติจึงมีการใช้วิธีการประมาณค่าทั้งแบบภาวะน่าจะเป็นสูงสุดและแบบเบส์ (Frey & Seitz, 2009) และมีวิธีการประมาณค่าความสามารถที่สำคัญ ดังนี้

4.4.1 วิธีการประมาณค่าแบบภาวะน่าจะเป็นสูงสุด (Maximum Likelihood Estimation) วิธีการประมาณค่าแบบภาวะน่าจะเป็นสูงสุด ซึ่งคำนวณได้ดังสมการ 35 (Segall, 2010, p. 62 ; Reckase, 2009, p. 140)

$$f\left(\mu_{i1}=\mu_{i1},\mu_{i2}=\mu_{12},...,\mu_{in}=\mu_{in}\middle|\theta\right)=L\left(\mu\middle|\theta\right)=\prod_{i=1}^{n}P\left(u_{ij}\middle|\theta_j\right)^{u_{ij}}Q\left(u_{ij}\middle|\theta_j\right)^{1-u_{ij}} \quad (35)$$

เมื่อ

$f\left(\mu_{i1}=\mu_{i1},\mu_{i2}=\mu_{12},...,\mu_{in}=\mu_{in}\middle|\theta\right)$ แทน ฟังก์ชันการตอบข้อสอบ ข้อที่จะนำไปใช้ในการคำนวณในขั้นตอนการคัดเลือกข้อสอบและการให้คะแนน

$L\left(\mu\middle|\theta\right)$ แทน ฟังก์ชันความน่าจะเป็น

$P\left(u_{ij}\middle|\theta_j\right)^{u_{ij}}$ แทน ความน่าจะเป็นในการตอบข้อสอบถูก

$Q\left(u_{ij}\middle|\theta_j\right)^{1-u_{ij}}$ แทน $1-P\left(u_{ij}\middle|\theta_j\right)^{u_{ij}}$

4.4.2 วิธีการประมาณค่าแบบเบส์ (Bayesian Estimation)

วิธีการคำนวณได้ดังสมการ 36 (Segall, 2010, pp. 60-64; Reckase, 2009, pp. 144-145)

$$f\left(\theta|u\right)=\frac{L\left(u|\theta\right)f\left(\theta\right)}{f\left(u\right)} \quad (36)$$

เมื่อ $f\left(\theta|u\right)$ แทน ความน่าจะเป็นที่เกิดขึ้นหลังของ $\boldsymbol{\theta}$ ที่ทำให้เกิดอนุกรมคะแนนรายข้อ

$L\left(u|\theta\right)$ แทน ฟังก์ชันความน่าจะเป็น (likelihood function)

$f\left(\theta\right)$ แทน ความน่าจะเป็นที่เกิดขึ้นก่อนของ $\boldsymbol{\theta}$

$\boldsymbol{f(u)}$ แทน ความน่าจะเป็นภายในขอบเขตของ u ซึ่ง $\boldsymbol{f(u)}$ ซึ่งคำนวณได้จากสมการ 37 และ 38 ดังนี้

$$f\left(u\right)=\int_{-\infty}^{\infty}L\left(u|\theta\right)f\left(\theta\right)d\theta \quad (37)$$

นั่นคือ

$$f\left(\theta|u\right)=\frac{L\left(u|\theta\right)f\left(\theta\right)}{\int_{-\infty}^{\infty}L\left(u|\theta\right)f\left(\theta\right)d\theta} \tag{38}$$

โดยฟังก์ชันความน่าจะเป็น (likelihood function) ซึ่งคำนวณได้จากสมการ 39 ดังนี้

$$L\left(u|\theta\right)\equiv L\left(u_{v_1},u_{v_2},...|\theta=\prod_{i\in v}P_i\left(\theta\right)^{u_i}Q_i\left(\theta\right)^{1-u_i}\right) \tag{39}$$

เมื่อ $P_i\left(\theta\right)$ แทน ฟังก์ชันการตอบสนองข้อสอบข้อที่ *i*

$Q_i\left(\theta\right)=1-P_i\left(\theta\right)$

v แทน เวกเตอร์ที่บรรจุข้อสอบแบบปรับเหมาะ

การประมาณค่าด้วยวิธี ML เป็นสมการที่ดำเนินการพร้อมกัน p มิติ ซึ่งคำนวณได้จากสมการ 40 (Segall, 1996, p. 334)

$$\frac{\partial}{\partial\theta}\ln f\left(\theta|u\right)=0 \tag{40}$$

อนุพันธ์แยกส่วนลำดับที่ 1 ซึ่งคำนวณได้จากสมการ 41 (Segall, 2010, p. 73)

$$\frac{\partial}{\partial\theta}\ln f\left(\theta|u\right)=D\sum_{i\in s}v_1a_i-\phi^{-1}\left(\theta-\mu\right) \tag{41}$$

เมื่อผลรวมรายข้ออยู่ภายใน S ซึ่งคำนวณได้จากสมการ 42

$$v_i=\frac{\left(p_i\left(\theta\right)-c_i\right)\left(u_i-p_i\left(\theta\right)\right)}{\left(1-c_i\right)p_i\left(\theta\right)} \tag{42}$$

วิธีการประมาณค่าแบบเบส์ มีข้อตกลงเบื้องต้นว่ามีการแจกแจงเป็นแบบปกติหลายตัวแปร (Multivariate Normal) ที่มีค่าเฉลี่ยเท่ากับ 0 และมีเมทริกซ์ความแปรปรวน-ความแปรปรวนร่วมเท่ากับ φ ซึ่งคำนวณได้จากสมการ 43 (Segall, 2010, p. 60)

$$f\left(\theta\right)=\left(2\pi\right)^{-\frac{H}{2}}\left|\phi\right|^{-\frac{1}{2}}\exp\left[-\frac{1}{2}\left(\theta-\mu\right)'\phi^{-1}\left(\theta-\mu\right)\right] \tag{43}$$

เมื่อเวกเตอร์ $\mu^a\ \{\mu_1,\mu_2,...,\mu_H\}$ และเมทริกซ์ความแปรปรวนร่วม $H\times H$ ของ φ

การประมาณค่าที่ดีกว่าคือ ทำให้เป็นรูปทั่วไป (Segall, 2010, p. 69) ซึ่งคำนวณได้จากสมการ 44 และ 45 ดังนี้

$$\theta^{(m+1)} = \theta^{(m)} - \delta^{(m)} \tag{44}$$

เมื่อ $\theta^{(m+1)}$ แทน ค่าความสามารถข้อที่ m+1

$\theta^{(m)}$ แทน ค่าความสามารถข้อที่ m

$$\delta^{(m)} = \left[M\left(\theta^{(m)}\right) \right]^{-1} \times \frac{\partial}{\partial \theta} \ln f\left(\theta^{(m)} | u\right) \tag{45}$$

โดย $f\left(\theta^{(m)} | u\right)$ แทน ความน่าจะเป็นที่เกิดขึ้นหลังของ $\boldsymbol{\theta}$ ที่ทำให้เกิดอนุกรมคะแนนรายข้อ

$M\left(\theta^{(m)}\right)$ แทน เทริกซ์ทั้งเมทริกซ์ของอนุพันธ์แยกส่วนลำดับที่ 2 ของ $J(\theta)$ ซึ่งคำนวณได้จากสมการ 46 และ 47 ดังนี้

$$J_s\left(\theta\right) \equiv \frac{\partial^2}{\partial \theta \partial \theta'} \ln f\left(\theta | u\right) = D^2 \sum_{i \in s} a_i a_i' w_i - \phi^{-1} \tag{46}$$

เมื่อ D แทน ค่าคงที่ เท่ากับ 1.702

a_i แทน เวกเตอร์พารามิเตอร์อำนาจจำแนกของข้อที่ i

a_i' แทน เวกเตอร์ $1 \times p$ ของพารามิเตอร์อำนาจจำแนกข้อที่ i

$$w_i = \frac{Q_i(\theta)\left[P_i(\theta) - c_i\right]\left[c_i u_i - P_i^2(\theta)\right]}{P_i^2(\theta)\left(1 - c_i\right)^2} \tag{47}$$

โดย c_i แทน พารามิเตอร์โอกาสการเดาข้อสอบได้ถูกต้องในการทำข้อสอบข้อที่ i

u_i แทน ผลการตอบข้อสอบข้อที่ i ซึ่ง $u_i = 1$ เมื่อตอบถูก และ $u_i = 0$ เมื่อตอบผิด

$P_i\left(\theta\right)$ แทน ฟังก์ชันการตอบสนองข้อสอบข้อที่ i

$Q_i\left(\theta\right) = 1 - P_i\left(\theta\right)$

4.4.3 วิธีการประมาณค่าแบบนิวตันราฟ (Newton-Raphson Method)

วิธีการนี้ใช้ควบคู่ไปกับการประมาณค่าวิธีแบบภาวะน่าจะเป็นสูงสุด ในกรณีที่ไม่สามารถประมาณค่าความสามารถได้ (Segall, 1996; Chen, 2009, pp. 1-2; Frey & Seitz, 2009) ซึ่งคำนวณได้ดังสมการ 48 (Segall, 1996; Chen, 2009, pp. 1-2)

$$\theta^{(j+1)} = \theta^{(j)} - \delta^{(i)} \tag{48}$$

เมื่อ $\delta^{(j)}$ แทน เวกเตอร์ p x 1 และคำนวณค่าได้ดังสมการ 49

$$\delta^{(j)} = \left[H\left(q^{(j)}\right) \right]^{-1} \left(\left(\frac{\partial}{\partial_q} 1nL\left(u|\theta\right) \right) \right) \tag{49}$$

เมื่อ $H\left(\theta^{(j)}\right)$ แทน เมทริกซ์ p x p เป็นค่าของอนุพันธ์อันดับสองของฟังก์ชันความน่าจะเป็นของ และ แสดงได้ดังสมการ 50

$$H\left(\theta\right) = \begin{bmatrix} \frac{\partial^2 1nL}{\partial\theta_1^2} & \frac{\partial^2 1nL}{\partial\theta_1\partial\theta_2} & \frac{\partial^2 1nL}{\partial\theta_1\partial\theta_p} \\ \frac{\partial^2 1nL}{\partial\theta_2\partial\theta_1} & \frac{\partial^2 1nL}{\partial\theta_2^2} & \frac{\partial^2 1nL}{\partial\theta_2\partial\theta_p} \\ \frac{\partial^2 1nL}{\partial\theta_p\partial\theta_1} & \frac{\partial^2 1nL}{\partial\theta_q\partial\theta_2} & \frac{\partial^2 1nL}{\partial\theta_p^2} \end{bmatrix} \tag{50}$$

และ $\frac{\partial}{\partial_q} 1nL\left(u|\theta\right)$ เป็นค่าของอนุพันธ์อันดับหนึ่งของฟังก์ชันความน่าจะเป็นของ $\theta^{(j)}$ ซึ่งคำนวณได้ดังสมการ 51

$$\frac{\partial}{\partial_q} 1nL\left(u|\theta\right) = \begin{bmatrix} \frac{\partial}{\partial\theta_1} 1nL\left(m|\theta\right) \\ \frac{\partial}{\partial\theta_2} 1nL\left(u|\theta\right) \\ M \\ \frac{\partial}{\partial\theta_p} 1nL\left(u|\theta\right) \end{bmatrix} \tag{51}$$

4.4.4 วิธีการประมาณค่าแบบกำลังสองน้อยที่สุด (Least Squares Estimation)

วิธีการนี้พบได้ไม่บ่อยนักในการประมาณค่าพารามิเตอร์ของทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ แต่ก็เป็นวิธีที่ใช้ในโปรแกรม NOHARM ซึ่งคำนวณได้ดังสมการ 52 (Reckase, 2009, pp. 147-148)

$$SS_\theta = \sum_{i=1}^{n} \left(u_i - P\left(u_i = 1|\theta, a_i, d_i\right) \right)^2 \tag{52}$$

เมื่อ SS_θ แทน ผลรวมของความแตกต่างยกกำลังของแต่ละเวกเตอร์ θ

u_i แทน คะแนนที่ได้จากข้อสอบข้อที่ i (มีค่าเป็น 0 และ 1)

4.5 เกณฑ์การยุติการทดสอบ (Stopping Rule)

กฎการยุติการทดสอบของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์โดยทั่วไปใช้

กฎการยุติการทดสอบ มี 2 ลักษณะได้แก่ กฎความยาวคงที่ (Fixed-Length) และ จำนวนข้อสอบผันแปร (variable length) (Reckase, 2009, p. 335) และการพิจารณาจากค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่า (Standard error of estimation: SE) (Thissen, 1990) แต่การกำหนดเกณฑ์การยุติการทดสอบขึ้นอยู่กับวัตถุประสงค์ในการทดสอบ ลักษณะคลังข้อสอบ และการควบคุมการปฏิบัติงานต่างๆเกณฑ์ในการพิจารณาการยุติการทดสอบมีหลากกลายเกณฑ์ให้เลือกใช้ (Wainer et al., 2000, p. 113; Thompson & Weiss, 2011; Kuo et al., 2015) ดังนี้

4.5.1 ยุติการทดสอบเมื่อข้อสอบในคลังข้อสอบหมด

4.5.2 ยุติการทดสอบเมื่อการประมาณค่าคุณลักษณะแฝงมีการเปลี่ยนแปลงเล็กน้อยหรือไม่เปลี่ยนแปลงเลย

4.5.3 กฎความยาวคงที่ (Fixed-Length) และ จำนวนข้อสอบผันแปร (variable length) (Reckase, 2009, p. 335) สำหรับงานวิจัยการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ (MCAT) ที่ผ่านมานั้น ส่วนใหญ่จะใช้การจำกัดจำนวนข้อสอบที่ 10, 20 ,30 ,40 ,50 และ 60 ข้อ (Chen, 2009; Diao & Reckase, 2009; Finch, 2010) ซึ่งงานวิจัยของ Diao and Reckase (2009) ปรากฏว่า การใช้ข้อสอบจำนวน 50 ข้อ ความลำเอียงจะน้อยกว่าการใช้ข้อสอบจำนวน 20 ข้อ แต่ยังระบุไม่ได้ว่าจำนวนข้อสอบเท่าใดจึงจะเหมาะสมสำหรับการทดสอบปรับเหมาะแบบพหุมิติ

4.5.4 การพิจารณาจากค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่า (Standard Error of estimation: SE) โดยทั่วไปจะใช้กฎการยุติการทดสอบจากค่าความคลาดเคลื่อนมาตรฐานกำหนดไว้ที่ 0.3 (Thissen, 1990) ส่วนงานวิจัยที่ผ่านมาจะกำหนด SE ≤ .55, .50 และ 0.30 (Seo, 2011; Thissen, 1990; โสพส สุขานนท์สวัสดิ์, เสรี ชัดแช้ม และกฤษณะ ชินสาร, 2556) ซึ่งคำนวณได้ดังสมการ 53

$$SE(\theta) = \frac{1}{\sqrt{I(\theta)}} \tag{53}$$

เมื่อ $SE(\theta)$ แทน ความคลาดเคลื่อนมาตรฐานของการประมาณค่า θ

$I(\theta)$ แทน สารสนเทศของแบบสอบที่ให้สำหรับผู้มีความสามารถ θ

เมื่อค่าฟังก์ชันสารสนเทศของแบบสอบ (test information: $I(\theta)$) เกิดจากผลรวมเชิงพีชคณิตของค่าฟังก์ชันสารสนเทศของข้อสอบแต่ละข้อรวมเข้าด้วยกันทั้งฉบับ ณ ตำแหน่ง θ เดียวกัน สารสนเทศแบบสอบจะมีคุณภาพการทดสอบสูง ซึ่งคำนวณได้ดังสมการ 54 และ 55 (Hambleton and Cook, 1977, p. 66)

$$I(\theta) = \sum_{i=1}^{k} I_i(\theta) \text{ , } i = 1,2,3,\ldots,k \tag{54}$$

และ

$$I_i(\theta) = \frac{2.89a_i^2(1-c_i)}{\left[c_i + \exp(Da_i(\theta - b_i))\right]\left[1+\exp\left(-Da_i(\theta-b_i)\right)\right]^2} \quad ,i = 1,2,3,\ldots,k \qquad (55)$$

เมื่อ $I_i(\theta)$ แทน ค่าสารสนเทศที่ได้จากข้อสอบข้อที่ I สำหรับผู้ตอบที่มีความสามารถ θ

D แทน ค่าคงที่ เท่ากับ 1.702

a_i แทน ค่าพารามิเตอร์อำนาจจำแนกของข้อสอบข้อที่ i

b_i แทน ค่าพารามิเตอร์ความยากของข้อสอบข้อที่ i

c_i แทน ค่าพารามิเตอร์โอกาสการเดาของข้อสอบข้อที่ i

ฟังก์ชันสารสนเทศของข้อสอบ เมื่อพิจารณาเทียบกับค่าพารามิเตอร์อำนาจจำแนก (a) ความยาก (b) และค่าโอกาสการเดา (c) จะสรุปได้ดังนี้

1. ค่าสารสนเทศของข้อสอบจะสูงขึ้น สำหรับผู้สอบมีความสามารถ θ ใกล้กับค่าพารามิเตอร์ b ของข้อสอบ และค่าสารสนเทศของข้อสอบจะลดลง สำหรับผู้สอบที่มีความสามารถ θ ไกลกับค่าพารามิเตอร์ b ของข้อสอบ

2. ค่าสารสนเทศของข้อสอบโดยทั่วไปจะมีค่าสูง ถ้าค่าพารามิเตอร์ a ของข้อสอบมีค่ามากขึ้น

3. ค่าสารสนเทศของข้อสอบมีค่าสูงขึ้น เมื่อค่าพารามิเตอร์ c ของข้อสอบเข้าใกล้ 0

4. $I_i(\theta)$ จะมีค่าสูงสุด ณ ตำแหน่ง $\theta_{\max}$ ถ้า $c_i = 0, \theta_{\max} = b$ แต่ถ้า $c_i > 0, \theta_{\max} > b$

จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องสามารสรุปได้ว่า การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ มีองค์ประกอบคล้ายกับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบเอกมิติ เช่น โมเดลอยู่ภายใต้ทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ การพัฒนาคลังข้อสอบที่ครอบคลุมเนื้อหาที่ต้องการวัด มีการประมาณค่าความสามารถของผู้สอบ การคัดเลือกข้อสอบและเกณฑ์การยุติการทดสอบ ในส่วนที่แตกต่างกัน คือ ข้อสอบแต่ละข้อของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติสามารถวัดความสามารถของผู้สอบได้หลายอย่างแต่การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบเอกมิติข้อสอบแต่ละข้อมุ่งวัดความสามารถของผู้สอบเพียงด้านเดียว ซึ่งการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติสามารถให้สารสนเทศได้มากกว่าการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบเอกมิติเมื่อใช้ข้อสอบจำนวนข้อสอบเท่ากัน

4.6 การแสดงข้อสอบ

การแสดงข้อสอบ (Item Eposure) หมายถึง การจัดข้อสอบที่ผ่านการคัดเลือกแล้วให้ผู้สอบโดยเป็นส่วนหนึ่งของขั้นตอนการคัดเลือกข้อสอบข้อถัดไป ในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ขั้นตอนการคัดเลือกข้อสอบข้อถัดไป จะคัดเลือกข้อสอบทีมีความเหมาะสมกับระดับความสามารถของผู้สอบ ซึ่งเป็นข้อสอบที่ให้ค่าสารสนเทศสูงสุด ณ ระดับความสามารถของผู้สอบขณะนั้น

ค่าสารสนเทศของข้อสอบจะมีค่าเพิ่มขึ้น ใน 2 กรณี คือ 1) ค่าอำนาจจำแนกของข้อสอบมีค่าเพิ่มขึ้น หรือ 2) ค่าความยากของข้อสอบมีค่าใกล้เคียงกับความสามารถของผู้สอบ ด้วยเหตุนี้ ในขั้นตอนการคัดเลือกข้อสอบข้อถัดไป ส่วนใหญ่จะคัดเลือกข้อสอบที่มีค่าอำนาจจำแนกของข้อสอบสูงไปใช้ในการทดสอบ เป็นผลให้ข้อสอบบางข้อถูกนำไปใช้ในการทดสอบน้อย หรือไม่เคยถูกนำไปใช้ในการทดสอบ เป็นผลให้ข้อสอบบางข้อถูกนำไปใช้ในการทดสอบบ่อยครั้งเกินไป (Overexposed Item) ขณะเดียวกัน ข้อสอบบางข้อถูกนำไปใช้ในการทดสอบน้อย หรือไม่เคยถูกนำไปใช้ในการทดสอบ (Underutilized Item) การมีข้อสอบที่ถูกนำไปใช้ในการทดสอบบ่อยครั้งเกินไป จะส่งผลกระทบต่อความปลอดภัยของการทดสอบ (Test Security) กล่าวคือ เมื่อมีข้อสอบที่ถูกนำไปใช้ในการทดสอบบ่อยครั้งเกินไป จนรู้กันในกลุ่มผู้สอบเป็นผลทำให้ผู้สอบในรุ่นถัดมาสามารถคาดการณ์ล่วงหน้าได้ว่าตนเองจะได้รับข้อสอบข้อใด ทำให้สามารถตอบข้อสอบข้อดังกล่าวได้ โดยไม่ได้ใช้ความสามารถของตนเอง คะแนนสอบจึงขาดความถูกต้อง ผู้สอบสามารถทำคะแนนได้มากเกินความสามารถของตนเอง ซึ่งส่งผลกระทบต่อความปลอดภัยของการทดสอบและประสิทธิภาพของการแสดงข้อสอบในคลังข้อสอบ (Item pools) ขณะเดียวกัน กรณีที่ข้อสอบถูกจัดให้ผู้สอบใช้น้อยหรือไม่ถูกจัดให้ผู้สอบใช้เลย ทำให้เสียค่าใช้จ่ายในการจัดเตรียมคลังข้อสอบโดยเปล่าประโยชน์

ดังนั้นจึงมีการศึกษาเกี่ยวกับวิธีการแสดงข้อสอบหลายวิธี โดย Georgiadou et al. (2007) ได้จำแนกวิธีการแสดงข้อสอบไว้ 5 ประเภท คือ วิธีการเลือกอย่างสุ่ม (Randomized Item-Selection Strategies) วิธีการเลือกอย่างมีเงื่อนไข (Conditional Selection Strategies) วิธีการเลือกตามระดับชั้น (Stratified Strategies) วิธีการเลือกแบบรวมกลยุทธ์ (Combined Strategies) และ วิธีการทดสอบหลายขั้นตอน (Multiple Stage Adaptive Test) รายละเอียดดังนี้

2.1 วิธีการเลือกอย่างสุ่ม (Randomized Item-Selection Strategies)

วิธีนี้การเลือกกลุ่มข้อสอบที่เหมาะสมที่สุดจากชุดข้อสอบ แทนการเลือกข้อสอบสารสนเทศสูงสุดเพียงข้อเดียว และสุ่มเลือกข้อสอบข้อหนึ่งจากกลุ่มข้อสอบจัดให้กับผู้สอบ วิธีการเลือกอย่างสุ่มไม่รับรองว่าการแสดงข้อสอบจะควบคุมได้ตามเกณฑ์ที่ต้องการ ตัวอย่างวิธีนี้ เช่นวิธี 5-4-3-2-1 (5-4-3-2-1 Procedure) (McBride & Martin, 1983; Hetter & Sympson, 1997 cited in Davis & Dodd, 2003) วิธีนี้เน้นการเลือกข้อสอบในระยะเริ่มต้นของการทดสอบ เมื่อผู้สอบมีแนวโน้มว่าเริ่มต้นการทดสอบที่ระดับความสามารถเท่าเทียมกัน ต่อมา Kingsbury and Zara (1989) ได้เสนอวิธีสุ่มเลือกจากกลุ่มข้อสอบ (Randomesque Procedure) วิธีนี้จะสุ่มเลือกข้อสอบเพียงข้อเดียวจากกลุ่มข้อสอบสารสนเทศสูงสุดจำนวนเท่าๆ กัน (Kingsbury & Zara, 1989 cited in Boyd, 2003) Thomason (1998) ได้เสนอ วิธีสุ่มเลือกข้อสอบหนึ่งในสาม (Choose One of Three) โดยใช้วิธีสุ่มเลือกข้อสอบจากข้อสอบสารสนเทศสูงสุด 3 ข้อ ข้อสอบอีก 2 ข้อ ที่ไม่ได้รับเลือกจะนำกลับไปจัดชุดใหม่เพื่อทำการเลือกครั้งต่อไป (Thomasson, 1998 cited in Davis & Dodd, 2003) ถัดมา Lunz and Stahl (1998) นำเสนอ วิธีสุ่มเลือกโลจิทส์ (Within .10 Logits Procedure) โดยวิธีนี้ข้อสอบได้รับเลือกจากการสุ่มในกลุ่มข้อสอบความยากเดียวกันและจับคู่ระหว่างค่าความสามารถของผู้สอบกับค่าความยากของข้อสอบ ถ้าไม่ได้ข้อสอบความยากตามที่ต้องการ จะสุ่มเลือกข้อสอบความยากใกล้เคียงค่าความยากเป้าหมายแทน (Lunz & Stahl, 1998 cited in Davis & Dodd, 2003) และ Revuelta and Ponsoda (1998) นำเสนอ วิธีสุ่มเลือกแบบก้าวหน้า

(Progressive Procedure) เป็นวิธีรวมการสุ่มเลือกและการเลือกจากสารสนเทศสูงสุดเข้าไว้ด้วยกัน ในระยะแรกใช้วิธีการสุ่มเลือกมากกว่าการเลือกจากสารสนเทศสูงสุด แต่เมื่อการทดสอบดำเนินต่อไป สารสนเทศจะมีผลต่อการเลือกข้อสอบมากกว่าวิธีการสุ่มเลือก (Revuelta & Ponsoda, 1998)

2.2 วิธีการเลือกอย่างมีเงื่อนไข (Conditional Selection Strategies) วิธีการเลือกอย่างมีเงื่อนไขเป็นวิธีควบคุมความน่าจะเป็นของข้อสอบที่ได้รับการเลือกแล้วให้มีการแสดงข้อสอบเป็นไปตามกำหนด วิธีนี้รับรองว่าอัตราการแสดงข้อสอบสูงสุดเป็นไปตามกำหนด แต่ขั้นตอนมีความซับซ้อน และต้องศึกษาในสถานการณ์จำลองในด้านจำนวนครั้งสำหรับการใช้ปฏิบัติ เมื่อองค์ประกอบต่างๆในการทดสอบเปลี่ยนแปลงไป จำเป็นต้องศึกษาในสถานการณ์จำลองก่อนใช้ในสถานการณ์จริง มีวิธีต่างๆ ได้แก่

2.2.1 วิธีซิมสัน-เฮตเทอร์ (Sympson-Hetter Procedure: SH) วิธีนี้เป็นที่รู้จักกันแพร่หลาย เสนอโดย Sympson and Hetter (1985) โดยใช้ความถี่ของข้อสอบที่จัดให้กับกลุ่มผู้สอบจำนวนมากในสถานการณ์จำลอง เปรียบเทียบกับอัตราการแสดงข้อสอบเป้าหมายเพื่อหาพารามิเตอร์การแสดงข้อสอบ ขบวนการนี้ทำซ้ำๆ กันจนพารามิเตอร์การแสดงข้อสอบแต่ละข้อมีค่าระหว่าง 0 ถึง 1 พารามิเตอร์นี้ใช้สำหรับการเลือกข้อสอบในการทดสอบสถานการณ์จริงโดยสร้างตัวเลขสุ่มจากการแจกแจงเหมือนกัน (0 - 1) เปรียบเทียบกับพารามิเตอร์การแสดงข้อสอบ ถ้าพารามิเตอร์การแสดงข้อสอบมากกว่าตัวเลขสุ่มจะจัดข้อสอบให้ แต่ถ้าพารามิเตอร์การแสดงข้อสอบน้อยกว่าตัวเลขสุ่มจะทำการเลือกข้อสอบข้อใหม่แทน ข้อดีของวิธีนี้คือ ยอมให้มีการกำหนดอัตราการแสดงข้อสอบล่วงหน้าเพื่อยืนยันว่าการแจกแจงอัตราการแสดงข้อสอบควรมีลักษณะเดียวกัน วิธีนี้ยังเป็นวิธีพื้นฐานประยุกต์ไปสู่วิธีการเลือกอย่างมีเงื่อนไขอื่นๆ อีกด้วย (Sympson & Hetter, 1985 cited in Stocking & Lewis, 2000)

2.2.2 วิธีซิมสัน-เฮตเทอร์แบบมีเงื่อนไข (The Conditional Sympson-Hetter Procedure) วิธีนี้พัฒนามาจากวิธีซิมสัน-เฮตเทอร์ เสนอโดย Stocking and Lewis (1995) โดยใช้พารามิเตอร์การแสดงข้อสอบได้จากการประมาณระดับค่าความสามารถแทนการแจกแจงค่าความสามารถของผู้สอบ ระหว่างขั้นตอนสถานการณ์จำลอง จะมีการสร้าง n × m เมทริกส์ เมื่อแถวนอน n คือจำนวนข้อสอบในคลังข้อสอบ และแถวตั้ง m คือ จำนวนจุดค่าความสามารถที่กระจายบนช่วงการแจกแจงค่าความสามารถ ดังนั้น ข้อสอบแต่ละข้อจะมีพารามิเตอร์การแสดงข้อสอบที่แต่ละจุดค่าความสามารถของผู้สอบ (Stocking & Lewis, 1995 cited in Boyd, 2003)

2.2.3 วิธีดาเวย์-พาร์เชลล์ (Davey-Parshall Procedure) เสนอโดย Davey and Parshall (1995) วิธีนี้เน้นการลดการทับซ้อนข้อสอบระหว่างกลุ่มผู้สอบ โดยสร้างพารามิเตอร์การแสดงข้อสอบจาก n × n เมทริกส์ เมื่อ n คือ จำนวนข้อสอบในคลังข้อสอบ เงื่อนไขพารามิเตอร์ในไดอะโกนอลเมทริกส์ (Diagonal Matrix) ควบคุมความถี่ของข้อสอบจากการเปรียบเทียบรายคู่ (Pairwise Comparisons) ในสถานการณ์จำลองเพื่อควบคุมการทับซ้อนข้อสอบ ไดอะโกนอลเมทริกส์มีพารามิเตอร์การแสดงข้อสอบปฏิบัติคล้ายพารามิเตอร์การแสดงข้อสอบในวิธี ซิมสัน-เฮตเทอร์ (Davey & Parshall, 1995 cited in Boyd, 2003)

2.2.4 วิธีโดยรวมเงื่อนไข (Triconditional Procedure) เสนอโดย Parshall, Hogerty and Kromrey (1999) วิธีนี้รวมวิธีซิมสัน- เฮตเทอร์(Sympson-Hetter Technique)

วิธีซิมสัน-เฮตเทอร์แบบมีเงื่อนไข (The Conditional Sympson-Hetter Procedure) และ วิธีดาเวย์-พาร์เชลล์ (Davey-Parshall Procedure) เข้าไว้ด้วยกัน (Parshall, Hogerty & Kromrey, 1999 cited in Boyd, 2003)

2.2.5 วิธีมัลติโนเมียลแบบไม่มีเงื่อนไข (Unconditional Multinomial Procedure) เสนอโดย Stocking and Lewis (1995) วิธีนี้พารามิเตอร์การแสดงข้อสอบได้จากวิธีซิมสัน-เฮทเทอร์ในช่วงระยะแรก ระยะที่ 2 ใช้รูปแบบมัลติโนเมียลในการเลือกข้อสอบ วิธีนี้ประยุกต์ไปสู่การควบคุมการแสดงข้อสอบมัลติโนเมียลแบบมีเงื่อนไข (Conditional Multinomial Procedure) ลักษณะต่างๆ (Stocking & Lewis, 1998) (Stocking & Lewis, 1995 cited in Boyd, 2003)

2.2.6 วิธีจำกัดสารสนเทศสูงสุด (Restricted Maximum Information Procedure) เสนอโดย Revuelta and Ponsoda (1998) วิธีนี้กำหนดอัตราการแสดงข้อสอบสูงสุดไว้ล่วงหน้าและจำกัดจำนวนครั้งการจัดข้อสอบให้ผู้สอบในชุดแบบทดสอบทั้งหมด การเลือกข้อสอบพิจารณาจากสารสนเทศสูงสุด ข้อสอบที่มีอัตราการแสดงข้อสอบครั้งหลังสุดต่ำกว่าอัตราการแสดงข้อสอบสูงสุดจะนำไปรวมไว้ในคลังข้อสอบ ถ้าข้อสอบมีอัตราการแสดงข้อสอบมากกว่าอัตราการแสดงข้อสอบสูงสุด ข้อสอบจะถูกขับออกจากคลังข้อสอบและไม่นำมาพิจารณาในการเลือกข้อสอบ

2.3 วิธีการเลือกตามระดับชั้น (Stratified Strategies)

2.3.1 วิธีเลือกตามระดับชั้นของค่าอำนาจจำแนก (a-Stratified Method: a-STR) เสนอโดย Chang and Ying (1996) ประยุกต์มาจากรูปแบบปรับระดับชั้นของ Weiss (1974) เป็นวิธีการเลือกข้อสอบและให้ประสิทธิภาพในการควบคุมการแสดงข้อสอบโดยอัตโนมัติ เริ่มจากการแบ่งชั้นคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบ และแบ่งแบบทดสอบเข้าในแต่ละชั้นคลังข้อสอบ ชั้นแรกบรรจุข้อสอบอำนาจจำแนกต่ำสุดชั้นต่อไปบรรจุข้อสอบอำนาจจำแนกสูงกว่าชั้นแรก ทำเช่นนี้เรื่อยๆ จนกระทั่งชั้นสุดท้ายบรรจุข้อสอบค่าอำนาจจำแนกสูงสุด การทดสอบจะดำเนินจากชั้นแรกจนถึงชั้นสุดท้าย วิธีนี้กระจายการใช้ข้อสอบในคลังข้อสอบให้เท่าเทียมกัน

2.3.2 วิธีเลือกตามระดับชั้นของค่าอำนาจจำแนกและค่าความยาก (a-Stratified Method with b Blocking: ab-STR) เสนอโดย Chang, Qian and Ying (2001) ประยุกต์มาจากวิธี a-SRT สำหรับวิธี ab-STR คลังข้อสอบจะถูกแบ่ง 4 ระดับ เหมือนกัน แต่จัดบล็อก 90 บล็อกด้วยกลุ่มค่าความยาก (b) เหมือนกัน แต่ละกลุ่มประกอบด้วย ข้อสอบ 4 ข้อที่มีค่า b เหมือนกันมากที่สุด ดังนั้น กลุ่มที่ 1 ประกอบด้วยข้อสอบ 4 ข้อ ที่มีค่า b ต่ำที่สุดและกลุ่มที่ 90 บรรจุข้อสอบ 4 ข้อ ที่มีค่า b สูงสุด ข้อสอบค่า a ต่ำสุดมีอยู่ในทุกๆ กลุ่มระดับแรก ข้อสอบค่า a ต่ำเป็นอันดับ 2 มีอยู่ในทุกๆ กลุ่มระดับที่ 2 ค่า a อื่นๆ ถูกสร้างในทำนองเดียวกัน

2.3.3 วิธีการออกแบบ 0-1-SRT เสนอโดย Chang and van der Linden (2001) ใช้วิธีการที่แตกต่างกันไปที่อยู่การแบ่งชั้นคลังข้อสอบ ขึ้นอยู่กับเทคนิคของการเขียนโปรแกรมเชิงเส้น 0-1 (LP) นี้วิธี stratifies Item pool ได้อย่างดีที่สุดโดยร่วมกับการรวมกันของ ค่าอำนาจจำแนก (a) และ ค่าความยาก (b) โดยเฉพาะการกำหนดเป้าหมายค่าสำหรับแต่ละชั้น

2.4 วิธีการเลือกแบบรวมกลยุทธ์ (Combined Strategies) มีนักวิชาการหลายท่านได้นำเอาวิธีการควบคุมการแสดงข้อสอบวิธีต่างมารวมกันมากมายและเป็นที่นิยมและง่ายต่อการประยุกต์วิธีการใหม่ อาทิ

2.4.1. วิธีจำกัดความก้าวหน้า (The Progressive Restricted Procedure) เสนอโดย Revuelta and Ponsoda (1998) โดยรวมวิธีจำกัดสารสนเทศสูงสุด (Restricted Maximum Information Procedure) และวิธีสุ่มเลือกแบบก้าวหน้า (Progressive Procedure) ไว้ด้วยกัน วิธีจำกัดสารสนเทศสูงสุดนำมาใช้ เพื่อไม่ให้ข้อสอบมีอัตราการแสดงข้อสอบเกินอัตราการแสดงข้อสอบสูงสุดขณะที่วิธีสุ่มเลือกแบบก้าวหน้าใช้สำหรับเลือกข้อสอบให้ผู้สอบ

2.4.2 วิธีเลือกตามระดับชั้นของค่าอำนาจจำแนกร่วมกับวิธีซิมสันและเฮตเทอร์ (a-Stratified with Sympson-Hetter Method: a-STR-SH) เสนอโดย Leung, Chang, and Hau (2002) โดยวิธีนี้จัดการกับปัญหาที่เกิดจากขั้นตอนการแบ่งชั้น โดยทำการจำลองการศึกษาแล้วปรากฏว่า วิธี a-SRT-SH มีประสิทธิภาพมากกว่าการใช้วิธี SH หรือวิธี a-SRT อย่างเดียว โดยอัตราการทับซ้อนของข้อสอบและการควบคุมการแสดงข้อสอบที่มากเกินไปลดลง

2.4.3 วิธีซิมสันและเฮตเทอร์ร่วมกับการหมุนคลังข้อสอบ (SH with rotating item pools) เสนอโดย Barrada, Olea, and Abda (2008) โดยรวมขั้นตอน SH และกลยุทธ์ของการหมุนคลังข้อสอบ โดยอัตราการเปิดรับรายการสูงสุดในการหมุนคลังข้อสอบ แสดงให้เห็นว่าวิธีการที่ร่วมกันดำเนินการในทำนองเดียวกันนี้อย่างใดอย่างหนึ่งกับขั้นตอน SH หรือกลยุทธ์การหมุนคลังข้อสอบแสดงผลดีในแง่ของอัตราอัตราการทับซ้อนกันลดลง (ขึ้นอยู่กับการกระจายของค่าพารามิเตอร์คือค่าอำนาจจำแนก (a) และค่าความยาก (b) ในคลังข้อสอบด้วย)

2.4.4 วิธีขยายช่วงระดับชั้น (Enhanced Stratified Procedure) เสนอโดย Leung, Chang and Hua (2003) เป็นการประยุกต์รวมวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกและวิธีควบคุมการแสดงข้อสอบซิมสัน-เฮตเทอร์เข้าไว้ด้วยกัน

2.5 วิธีการทดสอบหลายขั้นตอน (Multiple Stage Adaptive Test) เป็นรูปแบบการทดสอบที่เสนอโดย Krathworth and Huyser (Krathworth & Huyser, 1969 cited in Weiss, 1974) ซึ่งการทดสอบรูปแบบนี้จะมีจำนวนข้อสอบในแต่ละขั้นมากกว่า 1 ข้อ เพื่อที่จะแยกทางการทำข้อสอบข้อถัดไปได้ถูกต้องมากขึ้น สมมติว่ามีจำนวนข้อสอบในแต่ละชุดมี 3 ข้อ ถ้าผู้สอบตอบข้อสอบในขั้นที่ 1 (ข้อ 1 ถึง ข้อ 3) ถูก 0 ถึง 1 ข้อ ข้อสอบที่จะทำชุดถัดไปในขั้นที่ 2 จะเป็นข้อสอบข้อ 4 ถึงข้อ 6 หากตอบข้อสอบในขั้นที่ 1 (ข้อ 1 ถึง ข้อ 3) ถูก 2 ถึง 3 ข้อ จะไปทำข้อสอบข้อ 7 ถึงข้อ 9 ในขั้นที่ 2 เป็นชุดถัดไป สำหรับในขั้นอื่น ๆ จะมีวิธีการดำเนินการเช่นเดียวกัน

ตารางที่ 2-7 สรุปกลยุทธการควบคุมการแสดงข้อสอบ (Controlling Item Exposure) Georgiadou, Triantafillou, and Economides (2007)

ประเภทกลยุทธ์	วิธีการควบคุมการแสดงข้อสอบ	แหล่งอ้างอิง
วิธีการเลือกอย่างสุ่ม (Randomized Item-Selection Strategies)	5-4-3-2-1 Procedure	Davis and Dodd, 2003
	Randomesque Strategy	Kingsbury and Zara, 1989
	Choose One of Three	Davis and Dodd, 2003
	Within .10 Logits Procedure	Davis and Dodd, 2003
	Progressive Strategy	Revuelta and Ponsoda, 1998

	จุดเด่น	เป็นวิธีที่ง่ายในการนำไปใช้
	ข้อจำกัด	เนื่องจากเป็นการใช้กระบวนการสุ่มจึงไม่รับประกันในเรื่องของอัตราการใช้ข้อสอบซ้ำที่เกิดขึ้นและสารสนเทศได้ต่ำเพราะเป็นเลือกข้อสอบอย่างสุ่ม

ตารางที่ 2-7 (ต่อ)

ประเภทกลยุทธ์	วิธีการควบคุมการแสดงข้อสอบ	แหล่งอ้างอิง
วิธีการเลือกอย่างมีเงื่อนไข (Conditional Selection Strategies)	Sympson-Hetter (SH) Strategy	Stocking, 1993
	The Conditional Sympson-Hetter Procedure	Stocking and Lewis, 1995
	Davey-Parshall Procedure	Davey and Parshall, 1995
	Triconditional Procedure	Parshall, Hogerty and Kromrey, 1999
	Unconditional Multinomial Procedure	Stocking and Lewis, 1995
	Restricted Maximum Information Procedure	Revuelta and Ponsoda, 1998
	Stocking and Lewis Conditioning on Estimated Ability	Stocking and Lewis, 2002
	Chen and Lei strategy	Chen and Lei, 2005
	Shadow Test approach	van der Linden and Veldkamp, 2005
	จุดเด่น	1. ควบคุมความน่าจะเป็นของข้อสอบที่ได้รับการเลือกแล้ว ให้มีการใช้ข้อสอบเป็นไปตาม กำหนด
	ข้อจำกัด	2. วิธีนี้รับรองว่าอัตราการใช้ข้อสอบสูงสุดเป็นไปตามกำหนดล่วงหน้า ขั้นตอนมีความซับซ้อนและต้องศึกษาใน สถานการณ์จำลองโดยการทำซ้ำจำนวนมากและในการใช้งานจริง เมื่อองค์ประกอบต่างๆ ในการทดสอบเปลี่ยนแปลงไปจำเป็นต้องศึกษาในสถานการณ์จำลองใหม่ทุกครั้งเพื่อกำหนดพารามิเตอร์ควบคุมอัตราการใช้ข้อสอบซ้ำ
วิธีการเลือกตามระดับชั้น (Stratified Strategies)	a-Stratified Strategy (a-STR)	Chang and Ying, 1999
	a-STR with b-Blocking	Chang, Qian and Ying, 2001
	0-1 Stratification Strategy	Chang and van der Linden, 2003
	a-STR with Content Balancing	Leung et al., 2003
	จุดเด่น	เพิ่มอัตราการใช้ข้อสอบซ้ำในข้อสอบที่มีค่าอำนาจจำแนกต่ำและทำให้อัตราการใช้ข้อสอบซ้ำของข้อสอบทั้งคลังข้อสอบมีความเท่าเทียมกัน
	ข้อจำกัด	บรรจุภายในแต่ละชั้นควรมีค่าอำนาจจำแนกใกล้เคียงกัน และมีช่วงของค่าความยากของข้อสอบครอบคลุมช่วงความสามารถของผู้สอบ
วิธีการเลือกแบบรวมกลยุทธ์ (Combined Strategies)	Progressive Restricted Strategy	Revuelta and Ponsoda, 1998

	a-Stratified with Sympson-Hetter Method: a-STR-SH	Leung, Chang and Hau, 2002
	SH with Rotating Item Pools	Barrada, Olea, Ponsoda and Abda, 2008

ตารางที่ 2-7 (ต่อ)

ประเภทกลยุทธ์	วิธีการควบคุมการแสดงข้อสอบ	แหล่งอ้างอิง
	Enhanced Stratified Procedure	Leung et al., 2002
	Constraints CAT using Shadow Test	van der Linden, 2005
	Computerized Adaptive Sequential Testing	Luecht and Nungester, 1998
จุดเด่น	เป็นการนำเอาลักษณะเด่นของแต่ละวิธีมารวมกันทำใหม่ ความสามารถมากกว่ากลยุทธ์เดี่ยว จากงานวิจัยในอดีต ปรากฏว่า มีการนำเทคนิคมาบรูณาการเข้าด้วยกันเป็นจำนวนมาก	
ข้อจำกัด	ข้อจำกัดที่มีจะได้รับสืบทอดมาจากวิธีการต้นแบบที่นำมาใช้	
วิธีการทดสอบหลายขั้นตอน (Multiple Stage Adaptive Test)	Adaptive Multi-stage Item Bundles	Leucht, 2003
	Multiple Forms Structures	Armstrong, Jones, Koppel and Pashley, 2004
	Testlet-Based Adaptive Mastery Testing	Vos and Glas, 2002
	Two-Phase ISP for Flexible Content Balancing	Cheng et al., 2007
จุดเด่น	มีการนำเทคนิคมาบรูณาการเข้าด้วยกันเป็นจำนวนมาก ทำให้เกิดลักษณะที่เด่นในการทดสอบ	
ข้อจำกัด	ขั้นตอนมีความซับซ้อน และต้องศึกษาในสถานการณ์จำลองโดยการทำซ้ำจำนวนมาก	

4.7 การจัดสมดุลเนื้อหา

การจัดสมดุลของข้อสอบ (Content Balancing) เป็นสิ่งสำคัญอย่างยิ่งโดยเฉพาะการวัดผลสัมฤทธิ์ทางการศึกษา ในการทดสอบแบบดั้งเดิม แบบทดสอบมาตรฐานส่วนใหญ่ถูกสร้างตามรายละเอียดในตารางการกำหนดคุณลักษณะของข้อสอบ (Table of Content Specification) ผู้สอบทุกคนจะได้รับแบบทดสอบชุดเดียวกัน ประกอบด้วยข้อสอบครบทุกคุณลักษณะที่ต้องการวัดตามที่ออกแบบไว้ ด้วยเหตุนี้ จึงเป็นข้อจำกัดในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ เนื่องจากการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์จะคัดเลือกข้อสอบที่ให้ค่าสารสนเทศสูงสุดที่ค่าประมาณความสามารถของผู้สอบ ณ ขณะนั้น ซึ่งไม่ได้พิจารณาขอบเขตเนื้อหา (Content Area) ของข้อสอบ เป็นผลให้ผู้สอบบางคนไม่ได้รับข้อสอบทุกขอบเขตเนื้อหาของข้อสอบ ยกตัวอย่างเช่น ใน

การวัดความสามารถด้านการคำนวณ การบวก การลบ การคูณ และการหาร ในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ผู้สอบอาจได้รับข้อสอบเฉพาะ การบวก การลบ แต่ไม่ได้รับข้อสอบการคูณและการหาร ส่งผลให้ผู้ที่ไม่มีความรู้เรื่องการคูณและการหาร จะได้รับการประเมินความสามารถสูงเกินความจริง ในขณะเดียวกัน ผู้สอบที่มีความรู้เรื่องการคูณและการหารจะถูกประเมินต่ำเกินจริง ดังนั้น เพื่อรับประกันว่าผู้สอบแต่ละคนจะได้รับข้อสอบที่เป็นตัวแทนตามสัดส่วนของแต่ละขอบเขตของเนื้อหาครบถ้วน เพื่อให้แน่ใจว่าผู้สอบทุกคนได้รับข้อสอบในแต่ละเนื้อหาเท่ากัน ในการคัดเลือกข้อสอบข้อถัดไปจึงต้องคำนึงถึงการจัดสมดุลเนื้อหาของข้อสอบ (Wainer et al., 2001, pp. 120-121; Thompson & Weiss, 2011) ดังนั้น การเลือกข้อสอบในการทดสอบแบบปรับเหมาะนอกจากจะต้องรักษาประสิทธิภาพการประมาณค่าความสามารถของผู้สอบและควบคุมการแสดงข้อสอบแล้ว จำเป็นต้องจัดสมดุลเนื้อหาไปพร้อมๆ กันด้วย การจัดสมดุลเนื้อหากล่าวได้ว่าไม่ใช่การวัดทางจิตวิทยา จึงไม่เกี่ยวข้องกับรูปแบบการวัดต่างๆ ถือว่า เป็นการควบคุมไม่ใช้สถิติ (Non-Statistical Constraint) หรือเรียกว่า การควบคุมเนื้อหา (Content Constraint) แนวคิดเกี่ยวกับการจัดสมดุลเนื้อหานี้ Stocking and Swanson (1993) อธิบายลักษณะของข้อสอบไม่ใช้สถิติสำหรับการจำแนกและเลือกข้อสอบ ดังนี้

1. คุณสมบัติภายในข้อสอบ (Intrinsic Item Properties) กล่าวถึงลักษณะของข้อสอบด้านขอบเขตเนื้อหา

2. การควบคุมการทับซ้อน (Overlap Constraints) กล่าวถึงข้อสอบมุ่งวัดสิ่งอื่นนอกเหนือจากลักษณะที่ต้องการวัด

3. การควบคุมชุดข้อสอบ (Item Set Constraints) กล่าวถึงการยอมให้กลุ่มข้อสอบแบ่งสิ่งเร้าหรือจัดสิ่งมุ่งหมายร่วมกัน

นักวิชาการหลายคนได้ศึกษาการจัดสมดุลเนื้อหาของข้อสอบ ซึ่งวิธีที่นิยมมี 3 วิธี ได้แก่ 1) Kingsbury and Zara (1998) ได้นำเสนอวิธีการทดสอบแบบปรับเหมาะแบบควบคุมสมดุลเนื้อหา (Constrained Computerized Adaptive Testing: CCAT) วิธีนี้เลือกข้อสอบที่เหมาะสมที่สุดจากขอบเขตเนื้อหา โดยพิจารณาอัตราการแสดงข้อสอบครั้งหลังสุด และห่างจากเปอร์เซ็นต์ของการจัดข้อสอบตามกำหนดมากที่สุด งานวิจัยที่ใช้วิธีการจัดสมดุลเนื้อหาด้วยวิธี CCAT ได้แก่ Chang et al., (2009); Murphy, Dodd, and Vaughn, (2010); Han, (2012) 2) Leung, Chang, and Hau, (2003) ได้นำเสนอวิธีการทดสอบแบบปรับเหมาะแบบควบคุมสมดุลเนื้อหาประยุกต์ (Modified Constrained Computerized Adaptive Testing: MCCAT) วิธีนี้ประยุกต์มาจากการทดสอบแบบปรับเหมาะแบบควบคุมสมดุลเนื้อหา เพื่อลดการคาดเดาการเรียงเนื้อหาและให้การควบคุมสมดุลเนื้อหาให้ดียิ่งขึ้น ข้อสอบเลือกมาจากขอบเขตเนื้อหาทั้งหมดที่ยังไม่ได้ใช้เพื่อขจัดผลจากการเรียงลำดับ และ 3) Ankenmann and Spray (1999) ได้นำเสนอการจัดสมดุลเนื้อหารูปแบบมัลติโนเมียลประยุกต์ (Modified Multinomial Model: MMM) วิธีการทดสอบแบบปรับเหมาะแบบควบคุมสมดุลเนื้อหา (CCAT) สามารถคาดเดาอันดับการเรียงเนื้อหา ดังนั้น จึงได้พัฒนารูปแบบมัลติโนเมียลประยุกต์เพื่อต้องการให้เนื้อหามีความสมดุลกัน เมื่อขอบเขตเนื้อหามาถึงเปอร์เซ็นต์กำหนดจะมีการสร้างการแจกแจงมัลติโนเมียลเพื่อปรับเปอร์เซ็นต์ที่ยังไม่เต็มของขอบเขตเนื้อหาที่เหลืออยู่ และมีงานวิจัยที่เปรียบเทียบวิธีการจัดสมดุลเนื้อหาทั้ง 3 วิธี ได้แก่ วิธี CCAT, MCCAT และ MMM

ซึ่งมีงานวิจัยของ Leung, Chang, and Hua, (2003); Zhou, (2012); He and Reckase, (2014) ผลการศึกษาปรากฏว่า การจัดสมดุลเนื้อหาทั้ง 3 วิธี ให้ความถูกต้องในการประมาณค่าความสามารถของผู้สอบ ความคลาดเคลื่อนกำลังสองเฉลี่ย สัมประสิทธิ์สหสัมพันธ์ ค่าประมาณความสามารถและค่าความสามารถจริงไม่แตกต่างกัน

4.8 การประเมินประสิทธิภาพการประมาณค่าความสามารถของผู้สอบ

ประสิทธิภาพการประมาณค่าความสามารถของผู้สอบ หมายถึง ความสามารถของวิธีกำหนดในการประมาณค่าความสามารถของผู้สอบให้ใกล้เคียงหรือเท่ากับค่าความสามารถที่แท้จริงของผู้สอบมากที่สุด การประมาณค่าความสามารถว่าใกล้เคียงหรือเท่ากับค่าความสามารถที่แท้จริงของผู้สอบหรือไม่ สามารถพิจารณาได้จากค่าความคลาดเคลื่อนกำลังสองเฉลี่ย ค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย หรือค่าความลำเอียงเฉลี่ย ซึ่งมีนักวิชาการหลายท่านที่นำค่าทั้ง 3 ค่า นี้มาใช้เปรียบเทียบประสิทธิภาพการประมาณค่าความสามารถของผู้สอบของวิธีการต่างๆ ในการดำเนินการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ไม่ว่าจะเป็นวิธีการคัดเลือกข้อสอบข้อถัดไป วิธีการควบคุมการแสดงข้อสอบ หรือวิธีการควบคุมการจัดสมดุลเนื้อของข้อสอบ นักวิชาการเหล่านั้นได้แก่ Belov, Armstrong, and Weissman (2008); Cheng and Chang (2009); Cheng, Chang, Douglas, and Guo (2009); Hau and Chang (2001) และ Leung, Chang, and Hau (2002, 2003)

ค่าที่ใช้ประมาณความสามารถของผู้สอบ มีดังนี้

1) ค่าความคลาดเคลื่อนกำลังสอง (Mean Square Error: MSE) หมายถึง ค่าบอกความถูกต้องของค่าประมาณความสามารถของผู้สอบ แสดงเป็นผลต่างกำลังสองเฉลี่ยระหว่างค่าประมาณความสามารถและค่าความสามารถที่แท้จริงของผู้สอบ ซึ่งค่า MSE ขนาดเล็กบอกถึงการประมาณค่าความสามารถมีความแตกต่างจากค่าความสามารถที่แท้จริงของผู้สอบน้อยมาก สะท้อนให้เห็นถึงประสิทธิภาพของวิธีการคัดเลือกข้อสอบข้อถัดไป ซึ่งคำนวณได้ดังสมการ 56 (Yan, von Davier, & Lewis, 2016, p. 115)

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{\theta}_i - \theta_i\right)^2 \qquad (56)$$

เมื่อ n แทน จำนวนผู้สอบทั้งหมด

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

θ_i แทน ค่าความสามารถที่แท้จริงของผู้สอบคนที่ i

2) ค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Square Error: RMSE) หมายถึง ค่าบอกความถูกต้องของค่าประมาณความสามารถของผู้สอบ แสดงเป็นค่ารากที่สองของผลต่างกำลังสองเฉลี่ยระหว่างค่าประมาณความสามารถและค่าความสามารถที่แท้จริงของผู้สอบ นั่นคือ ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย ดังนั้น ค่า RMSE ขนาดเล็กบอกถึงการประมาณค่าความสามารถมีความแตกต่างจากค่าความสามารถที่แท้จริงของผู้สอบน้อยมาก ซึ่งคำนวณได้ดังสมการ 57 (Yan, von Davier, & Lewis, 2016, p. 115)

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{\theta}_i - \theta_i\right)^2} \quad (57)$$

เมื่อ n แทน จำนวนผู้สอบทั้งหมด

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

θ_i แทน ค่าความสามารถที่แท้จริงของผู้สอบคนที่ i

3) ค่าความลำเอียงเฉลี่ย (Average Bias) หมายถึง ค่าบอกความเที่ยงตรงของค่าประมาณความสามารถของผู้สอบ แสดงเป็นผลต่างเฉลี่ยระหว่างค่าประมาณความสามารถและค่าความสามารถที่แท้จริงของผู้สอบ สามารถบอกทิศทางการประมาณค่าว่าให้ผลสูงหรือต่ำกว่าค่าความสามารถที่แท้จริงของผู้สอบ ค่าความลำเอียงเฉลี่ยขนาดเล็กสะท้อนให้เห็นถึงประสิทธิภาพของวิธีการคัดเลือกข้อสอบข้อถัดไป ซึ่งคำนวณได้ดังสมการ 58 (Yan, von Davier, & Lewis, 2016, p. 115)

$$Average \quad Bias = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{\theta}_i - \theta_i\right) \quad (58)$$

เมื่อ n แทน จำนวนผู้สอบทั้งหมด

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

θ_i แทน ค่าความสามารถที่แท้จริงของผู้สอบคนที่ i

4.9 การประเมินประสิทธิภาพของการใช้ข้อสอบในคลังข้อสอบ

ประสิทธิภาพของการใช้ข้อสอบ หมายถึง ความสามารถของวิธีที่ทำให้ข้อสอบทุกข้อในคลังข้อสอบมีโอกาสนำไปใช้ในการทดสอบ (Ozturk & Dogan, 2015) พิจารณาจาก

1) ข้อสอบที่มีการแสดงมากเกินไป (Overexposure Item) หมายถึง ข้อสอบใดๆ ที่มีอัตราการแสดงสูงกว่า 0.2 เมื่ออัตราการแสดงข้อสอบ หมายถึง จำนวนครั้งของข้อสอบที่จัดให้แก่ผู้สอบต่อจำนวนผู้สอบทั้งหมดที่ทำการทดสอบ ซึ่งคำนวณได้ดังสมการ 59

$$\text{อัตราการแสดงข้อสอบ} = \frac{\text{จำนวนครั้งของข้อสอบที่จัดให้แก่ผู้สอบ}}{\text{จำนวนผู้สอบทั้งหมดที่ทำการทดสอบ}} \quad (59)$$

ถ้าข้อสอบในคลังข้อสอบมีอัตราการแสดงสูงจำนวนมาก แสดงถึงแนวโน้มข้อสอบเหล่านั้นนำไปใช้ในการทดสอบมากเกินไป เป็นที่รู้จักแพร่หลายในกลุ่มผู้สอบ ส่งผลกระทบต่อความปลอดภัยของข้อสอบ ด้านประสิทธิภาพของการใช้ข้อสอบ และสะท้อนถึงประสิทธิภาพวิธีเลือกข้อสอบไม่เหมาะสม

2) ข้อสอบที่มีการใช้น้อยเกินไป (Underutilized Item) หมายถึง ข้อสอบใดๆ ที่มีอัตราการแสดงต่ำกว่า 0.2 ถ้าข้อสอบที่มีการใช้น้อยเกินไปมีจำนวนมากแสดงถึงข้อสอบในคลังข้อสอบมี

โอกาสนำไปใช้น้อย ด้านประสิทธิภาพของการใช้ข้อสอบ สะท้อนถึงประสิทธิภาพวิธีเลือกข้อสอบไม่เหมาะสม และขาดประสิทธิภาพในการบริหารต้นทุนในการพัฒนาข้อสอบ

3) อัตราการทับซ้อนของข้อสอบ (Item Overlap Rate) หมายถึง สัดส่วนของข้อสอบที่ใช้ร่วมกันหารจำนวนผู้เข้าสอบทั้งแล้วคูณด้วย100 อัตราการทับซ้อนของข้อสอบคำนวณได้ดังสมการ 60

$$\text{อัตราการทับซ้อนของข้อสอบ} = \frac{\text{จำนวนผู้สอบที่ใช้ข้อสอบร่วมกัน}}{\text{จำนวนผู้สอบทั้งหมด}} \times 100 \qquad (60)$$

อัตราการทับซ้อนของข้อสอบเป็นดัชนีสรุปสำหรับการควบคุมการแสดงข้อสอบ ใช้กำหนดขนาดและส่วนประกอบของคลังข้อสอบในการทดสอบแบบปรับเหมาะ ให้ภาพรวมว่าชุดข้อสอบถูกจัดให้บ่อยครั้งเพียงไรและคาดว่าจำนวนข้อสอบใช้ร่วมกันระหว่างคู่ของผู้สอบต้องมีจำนวนน้อยที่สุด ถ้าอัตราการทับซ้อนของข้อสอบสูงจะคุกคามความปลอดภัยของข้อสอบ

4) การแจกแจงอัตราการแสดงข้อสอบ (Item Exposure Rate Distribution) หมายถึง ลักษณะการกระจายอัตราการแสดงของข้อสอบในคลังข้อสอบ บ่งบอกถึงประสิทธิภาพของการใช้ข้อสอบโดยรวม พิจารณาจากสถิติไค-สแควร์ ซึ่งคำนวณได้ดังสมการ 61

$$\chi^2 = \sum_{j=1}^{n} \left(\frac{er_j - \frac{L}{N}}{\frac{L}{N}} \right)^2 \qquad (61)$$

เมื่อ er_j แทน อัตราการแสดงข้อสอบสำหรับข้อสอบที่ j

L แทน ความยาวแบบทดสอบ

N แทน ขนาดคลังข้อสอบ

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

er_j เป็นค่าสังเกต ขณะที่ $\frac{L}{N}$ เป็นค่าคาดหวังว่าข้อสอบทุกข้อมีอัตราการแสดงข้อสอบเท่าเทียมกัน ดังนั้น สมการแสดงถึงความแตกต่างระหว่างค่าสังเกตและค่าคาดหวังของอัตราการแสดงข้อสอบ ถ้าค่า χ^2 เล็ก แสดงถึง การใช้ข้อสอบมีความสมดุลกัน

การเปรียบเทียบการแจกแจงอัตราการแสดงข้อสอบ 2 วิธี ได้แก่ วิธีที่ 1 และวิธีที่ 2 โดยใช้อัตราส่วน F (F - ratio) พิจารณาจากอัตราส่วนของค่า χ^2 วิธีที่ 1 หารด้วย χ^2 วิธีที่ 2 คำนวณได้ดังสมการ 62

$$F_{1.2} = \frac{\chi_1^2}{\chi_2^2} \qquad (62)$$

เมื่อ $F_{1.2}$ แทน อัตราส่วนการแจกแจงอัตราการแสดงข้อสอบ 2 วิธีหรือวิธีที่1หารด้วยวิธีที่ 2

χ_1^2 แทน การแจกแจงอัตราการแสดงข้อสอบวิธีที่ 1

χ_2^2 แทน การแจกแจงอัตราการแสดงข้อสอบวิธีที่ 2

ถ้า $F < 1$ แสดงว่า วิธีที่ 1 ให้สมดุลของอัตราการแสดงข้อสอบโดยรวมดีกว่าวิธีที่ 2

4.10 วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบแบ่งตามค่าอำนาจจำแนกของข้อสอบแบบถ่วงน้ำหนักที่มีการบังคับ

Cheng and Chang (2009) เสนอวิธีดัชนีลำดับความสำคัญสูงสุด (Maximum Priority Index: MPI) สำหรับการคัดเลือกข้อสอบที่มีการควบคุมเงื่อนไขบังคับอย่างเข้มงวดในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ซึ่งนำเทคนิคการคัดเลือกข้อสอบแบบสองขั้น (Two-phase Item Selection) ที่เสนอโดย Cheng et al. (2007) มาประยุกต์ใช้ร่วมด้วยเพื่อทำให้วิธี MIP สามารถจัดการกับเงื่อนไขบังคับได้อย่างยืดหยุ่นโดยสามารถกำหนดให้เลือกข้อสอบที่ตรงตามเงื่อนไขบังคับ ภายในช่วงขอบเขตบนและขอบเขตล่างแทนที่การกำหนดเป็นค่าคงที่ และนำมาเปรียบเทียบกับวิธี Weighted Deviation Modeling (WDM) ซึ่งเป็นวิธีที่มีรูปแบบการคัดเลือกข้อสอบอย่างเป็นลำดับเหมือนกัน ผลการศึกษาปรากฏว่า วิธี MPI มีประสิทธิภาพดีกว่า WDM อย่างเห็นได้ชัดในการจัดการเงื่อนไขข้อบังคับ อีกทั้งยังมีความถูกต้องแม่นยำของการประมาณค่าความสามารถของผู้สอบสูง และสามารถควบคุมอัตราการใช้ข้อสอบซ้ำให้อยู่ในระดับที่กำหนดไว้ได้ แต่มีสัดส่วนของข้อสอบที่ไม่ถูกนำมาใช้สูงถึงร้อยละ 50 ทำให้วิธี MPI มีจุดอ่อนในเรื่องการขาดความสมดุลในการใช้คลังข้อสอบ

Cheng and Chang (2009) จึงให้ข้อเสนอแนะว่า ควรนำมาใช้ร่วมกับวิธีการแบ่งคลังข้อสอบ ออกเป็นชั้น ๆ ตามค่าอำนาจจำแนกที่พัฒนาโดย Chang and Ying (1999) เพื่อเพิ่มประสิทธิภาพของการใช้คลังข้อสอบ หลังจากนั้น Cheng, Wang, and Ho (2009) ศึกษาถึงประโยชน์ของวิธีการแบ่งคลังข้อสอบออกเป็นชั้นๆ ตามค่าอำนาจจำแนก โดยนำมาใช้ร่วมกับวิธี MPI เพื่อให้เกิดความยืดหยุ่นในการควบคุมเงื่อนไขบังคับที่ไม่ใช่ทางสถิติและสามารถใช้คลังข้อสอบได้อย่างมีประสิทธิภาพ ผลจาก การศึกษาในสถานการณ์จำลอง ปรากฏว่า วิธีแบ่งกลุ่มค่าอำนาจจำแนกแบบถ่วงน้ำหนักที่มีการบังคับ (Constraint-Weighted a-Stratification Method: CWA) สามารถสร้างความสมดุลในการใช้คลังข้อสอบ โดยสูญเสียความถูกต้องแม่นยำของการประมาณค่าความสามารถของผู้สอบน้อยที่สุด รวมทั้งยัง ควบคุมเงื่อนไขบังคับได้อย่างดีเยี่ยม โดยไม่มีการฝ่าฝืนเงื่อนไขบังคับ

วิธีดัชนีลำดับความสำคัญจะพิจารณาความแปรปรวนของค่าสารสนเทศโดยการนำดัชนีไปเป็นตัวคูณให้กับค่าสารสนเทศของข้อสอบ ดังนั้นในการคัดเลือกข้อสอบแทนที่จะพิจารณาจากค่าสารสนเทศเพียงอย่างเดียว วิธีการนี้จะนำลำดับความสำคัญซึ่งกำหนดโดยผู้เชี่ยวชาญที่พิจารณาร่วมด้วย การคัดเลือกข้อสอบจะพิจารณาจากผลคูณของค่าสารสนเทศของข้อสอบกับดัชนีลำดับความสำคัญ ข้อใดให้ผลคูณมากที่สุด ข้อนั้นจะได้รับการคัดเลือกให้นำไปใช้กับผู้สอบ ดังนั้น วิธี

Priority Index (PI) เกี่ยวข้องกับเมตริกซ์เงื่อนไขบังคับ ซึ่งจะบอกว่าข้อสอบแต่ละข้อประกอบด้วยเงื่อนไขบังคับใดบ้าง

โดยกำหนดให้เมตริกซ์ดังกล่าว คือ เมตริกซ์ C.C เมื่อ c_{ik} = 1 แสดงว่า ข้อสอบข้อที่ j เกี่ยวข้องกับเงื่อนไขบังคับที่ K

เมื่อ C.C แทน เมตริกซ์ที่เกี่ยวข้องกับข้อบังคับ และมีขนาด J x K

j แทน จำนวนข้อสอบในคลังข้อสอบ

K แทน จำนวนเงื่อนไขบังคับทั้งหมด

k แทน ข้อบังคับเนื้อหาจะเกี่ยวข้องกับค่าน้ำหนัก (Weight: w_k)

c_i แทน 1 บ่งชี้ว่า ข้อบังคับ k เกี่ยวข้องกับข้อสอบที่ j

c_j แทน 0 บ่งชี้ว่า ข้อบังคับ k ไม่เกี่ยวข้องกับข้อสอบที่ j

โดยทั่วไป Matrix C.C จะถูกระบุก่อนการเลือกข้อสอบโดยผู้เชี่ยวชาญด้านเนื้อหาหรือผู้เชี่ยวชาญทางด้านจิตมิติ โดยที่แต่ละข้อบังคับ k จะเกี่ยวข้องกับค่าน้ำหนัก ซึ่งการ ทดสอบส่วนใหญ่จะใส่ค่าน้ำหนักขนาดใหญ่ให้กับข้อบังคับที่สำคัญและค่าน้ำหนักที่น้อยกว่ากับ ข้อบังคับอื่นๆ ดังนั้น Priority Index ของข้อสอบข้อที่ j คำนวณได้ดังสมการ 63

$$PI_j = I_j \prod_{k=1}^{k} \left(w_k f_k \right)^{c_{jk}} \tag{63}$$

เมื่อ PI แทน Fisher information ของข้อสอบข้อที่ J ที่ถูกประมาณค่าที่ระดับความสามารถปัจจุบัน

X_k แทน ข้อสอบจากขอบเขตเนื้อหาที่แน่นอน

x_k แทน ข้อสอบที่จะถูกเลือก

f_{jk} แทน โควตาที่มีอยู่ของข้อบังคับ k

และผลของโควตาที่เหลืออยู่คำนวณได้ดังสมการ 64

$$f_k = \frac{\left(X_k - x_k \right)}{X_k} \tag{64}$$

เมื่อ C_{jk} = 0 หมายความว่า ข้อสอบข้อที่ j ไม่ถูกควบคุมด้วยข้อบังคับ k

สมมติว่า ต้องการข้อบังคับ k' ดังนั้นอัตราการแสดงของแต่ละข้อเป็นค่าที่ต่ำกว่าหรือเท่ากับ r และระหว่างผู้สอบทั้งหมด N คน ผู้ซึ่งทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ มีผู้สอบ n คน ที่ได้เห็นข้อสอบ j แล้ว $f_{jk'}$ ซึ่งคำนวณได้ดังสมการ 65

$$f_k = \frac{\left(r - \left(\frac{n}{N} \right) \right)}{r} \tag{65}$$

จากสมการที่ 65 สามารถคำนวณดัชนีลำดับความสำคัญ (Maximum Priority Index: MPI) สำหรับข้อสอบทุกข้อที่อยู่ในคลังข้อสอบ จากนั้นจึงเลือกใช้ข้อสอบที่มีค่าดัชนีลำดับความสำคัญสูงสุด แทนที่ข้อสอบข้อที่มีค่าสารสนเทศของฟิชเชอร์สูงสุด (Maximum Fisher Information) วิธี MPI ในรูปแบบปัจจุบันถูกจำกัดข้อบังคับในรูปของขอบเขตบน อย่างไรก็ตาม โปรแกรมการทดสอบมักจะมีข้อบังคับที่เกี่ยวข้องกับขอบเขตล่างดังตัวอย่างผังข้อสอบสำหรับแบบทดสอบคณิตศาสตร์อาจจะระบุไว้เพียงไม่เกิน 15 ข้อ (Upper Bound) แต่ก็ไม่น้อยกว่า 10 ข้อ (Lower Bound) ที่เป็นเนื้อหาทางด้านพีชคณิต

ดังนั้นข้อบังคับจะเรียกว่า ข้อบังคับความสมดุลทางเนื้อหาที่ยืดหยุ่น (Flexible Content Balancing Constraint) (Cheng, Chang, & Yi, 2007) เมื่อข้อบังคับความสมดุลทางเนื้อหาที่ยืดหยุ่นได้ถูกนำเสนอ วิธี MPI ข้างต้นจึงได้รับการปรับปรุงแก้ไขโดยใช้ร่วมกันกับวิธี Two-phase item Selection Strategy (Cheng et al., 2007) เพื่อจัดการกับข้อบังคับความสมดุลทางเนื้อหาที่ยืดหยุ่น

`สำหรับกรอบแนวคิดการใช้ MPI ร่วมกับ two-phase item selection แต่ละข้อบังคับความสมดุลของเนื้อหาที่ยืดหยุ่นซึ่งเกี่ยวข้องกับขอบเขตล่าง และขอบเขตบน คำนวณได้ดังสมการ 66

$$l_k \mu_k \le u_k \quad \text{และ}$$
$$\sum_{k=1}^{k} \mu_k = L \tag{66}$$

เมื่อ μ_k แทน จำนวนของข้อสอบที่ถูกเลือกจากขอบเขตเนื้อหา k

l_k แทน ขอบเขตล่างของข้อบังคับเนื้อหา (Lower Bound)

u_k แทน ขอบเขตบนของข้อบังคับเนื้อหา (Upper Bound)

K แทน จำนวนของขอบเขตเนื้อหาทั้งหมด

L แทน ความยาวของแบบทดสอบ

แนวคิดของการคัดเลือกข้อสอบแบบ two-phase เป็นการจัดการกับ lower bounds ในช่วงระยะที่ 1 และจัดการกับขอบเขตบน ระยะที่ 2 ดังนั้นในระยะที่ 1 ของการคัดเลือกจะเกี่ยวข้องกับข้อสอบจำนวน L_1 เมื่อ $L_1 = \sum_{k=1}^{k} l_k$ และในระยะที่ 2 ของการคัดเลือกเมื่อ $L_2 = L - L_1$ โดยระยะที่ 1 ขอบเขตล่าง ถูกทำให้เป็น ขอบเขตบน

ในทั้ง 2 ระยะ การคำนวณ Priority index ของแต่ละข้อยังใช้สูตรการคำนวณตามสมการที่ 63 อย่างไรก็ตาม f_k คำนวณแตกต่างกัน ซึ่งในระยะที่ 1 คำนวณได้ดังสมการ 67

$$f_k = \frac{(l_k - x_k)}{l_k} \tag{67}$$

สมการที่ 67 แตกต่างจากสมการที่ 64 เพียงการแทนตำแหน่งของ X_k ด้วย l_k เมื่อข้อบังคับ k ไปถึงขอบเขตล่างแล้ว f_k เป็น 0 และ Priority Index ที่เกี่ยวข้องเปลี่ยนไปเป็น 0 ดังนั้น

Priority Index จึงได้รับการจัดระดับความสำคัญต่ำกว่าข้อสอบข้ออื่นๆ ในคลังข้อสอบซึ่งมี ค่าดัชนีความสำคัญเป็นบวก ตัวอย่างเช่น สมมติให้แบบทดสอบมีเพียง 2 ข้อบังคับ และข้อบังคับแรกไปถึงขอบเขตล่างเรียบร้อยแล้ว ขอบเขตเนื้อหานั้นจะหยุดนิ่ง และไม่มีข้อสอบที่จะสามารถถูกเลือกจากขอบเขตเนื้อหานั้นจนกระทั่งขอบเขตเนื้อหาอื่นมาถึง เนื่องจากขอบเขตล่างทั้งหมดจะพบกันที่จุดสุดท้ายของระยะที่ 1 ในระยะที่ 2 f_k คำนวณได้ดังสมการ 68

$$f_k = \frac{(u_k - x_k)}{u_k} \tag{68}$$

เมื่อข้อบังคับ k มาถึงขอบเขตบน u_k, f_k จะเป็น 0 และไม่มีข้อสอบที่สามารถเลือกเพื่อนำไปใช้กับผู้สอบได้จากในขอบเขตเนื้อหานั้น ดังนั้นข้อสอบจะสามารถเลือกออกมาใช้ได้ก็ต่อเมื่อมีขอบเขตเนื้อหาอื่นที่ยังไม่ถูกนำมาใช้จนเต็มขอบเขตบน

วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก (a-Stratified Method: a-STR) ดำเนินการตามขั้นตอนดังนี้

1. แบ่งคลังข้อสอบเป็นชั้นตามค่าอำนาจจำแนกข้อสอบ (พารามิเตอร์ a) ชั้นแรกบรรจุข้อสอบอำนาจจำแนกต่ำสุด ชั้น 2 บรรจุข้อสอบอำนาจจำแนกสูงกว่าชั้นแรก และเพิ่มขึ้นเรื่อยๆ ตามลำดับชั้น จนกระทั่งชั้นสุดท้ายบรรจุข้อสอบอำนาจจำแนกสูงสุด

2. แบ่งขั้นการทดสอบหรือความยาวแบบทดสอบตามการแบ่งชั้นของคลังข้อสอบ

3. ค่าความสามารถของผู้สอบ ณ ตำแหน่งจุดเริ่มต้นการทดสอบ ข้อสอบที่มีค่าความยากใกล้เคียงกับค่าความสามารถของผู้สอบมากที่สุดจะถูกเลือกและนำไปใช้กับผู้สอบ เมื่อผู้สอบตอบตอบข้อสอบจะถูกนำไปคำนวณหาค่าประมาณความสามารถ จากนั้นจึงนำค่าความสามารถของ ผู้สอบ ณ ตำแหน่งปัจจุบันไปใช้เลือกข้อสอบข้อต่อไป เมื่อทดสอบจนครบตามความยาว แบบทดสอบย่อยในขั้นนั้น จึงเลื่อนการทดสอบขึ้นไปยังขั้นและชั้นคลังข้อสอบต่อไป

ตอนที่ 4 งานวิจัยที่เกี่ยวข้องกับการจัดคลังข้อสอบ การจัดสมดุลเนื้อหา การควบคุมการแสดงข้อสอบ และการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ

งานวิจัยที่เกี่ยวข้อง แบ่งได้ 4 ส่วน ดังนี้

1. งานวิจัยที่เกี่ยวข้องกับวิธีการจัดคลังข้อสอบ ในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ มีดังนี้

Cheng and Chang (2009) ได้เสนอวิธีดัชนีลำดับความสำคัญสูงสุด (Maximum Priority Index: MPI) สำหรับการคัดเลือกข้อสอบที่มีข้อบังคับอย่างเข้มงวด การวิจัยนี้ศึกษาผ่านการจำลองข้อมูลเพื่อเปรียบเทียบวิธีการจัดการข้อบังคับ 2 วิธี ได้แก่ 1) วิธีดัชนีลำดับความสำคัญสูงสุด 2) วิธี Weighted Deviation Modeling (WMD) โดยใช้วิธีการคัดเลือกข้อสอบแบบ Maximum Information และวิธี Randomized เป็นฐานในการเปรียบเทียบโดยใช้ข้อมูลจากคลังข้อสอบของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ของการสอบเพื่อจัดตำแหน่งในมหาวิทยาลัย (College

Placement CAT) ในวิชาพื้นฐานพีชคณิต คลังข้อสอบประกอบด้วยข้อสอบจาก 3 กลุ่มเนื้อหาวิชา แต่ละกลุ่มเนื้อหาวิชาแบ่งออกเป็น 4-10 ขอบเขตเนื้อหา ผลการศึกษาปรากฏว่า 1) วิธี MPI และวิธี WDM ลดการฝ่าฝืนข้อบังคับในการเปรียบเทียบกับวิธี MI และ Randomized ผลการทดสอบแสดงให้เห็นอย่างชัดเจนว่า MPI มีประสิทธิภาพดีกว่า WDM ทุกด้าน โดยมีอัตราการใช้ข้อสอบซ้ำสูงสุด การใช้ข้อสอบซ้ำเกิน และอัตราการทับซ้อนของแบบสอบต่ำกว่าวิธี WDM นอกจากนี้ยังปรากฏว่า ทุกวิธียกเว้น วิธี Randomized มีสัดส่วนของ Never Exposed สูงมากกว่า 50% แสดงว่าข้อสอบในคลังข้อสอบถูกใช้อย่างไม่มีประสิทธิภาพ ซึ่งปัญหานี้สามารถทำให้ลดลงได้โดยการนำไปรวมเข้ากับวิธี a-Stratified Design ของ Chang and Ying (1999) จากข้อค้นพบของการศึกษาครั้งนี้ วิธี MPI สามารถปรับให้เข้ากับข้อบังคับที่ไม่ใช่ทางสถิติที่หลากหลายและสามารถควบคุมได้หลายอย่างพร้อมๆ กัน เช่น ความสมดุลของเนื้อหา การควบคุมการใช้ข้อสอบซ้ำ ความสมดุลของคำตอบ โดยวิธีการนี้สามารถนำไปปรับใช้กับโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ที่มีอยู่ในปัจจุบันได้สะดวก โดยไม่จำเป็นต้องปรับค่าน้ำหนักความสัมพันธ์ระหว่างข้อบังคับและสารสนเทศ

Cheng, Cheng, Douglas, and Guo (2009) ได้เสนอวิธีดัชนีลำดับความสำคัญสูงสุด มาใช้ร่วมกับการแบ่งชั้นพารามิเตอร์อำนาจจำแนก (a-Stratification) จึงถูกเรียกชื่อใหม่ว่าวิธีการคัดเลือกข้อสอบแบบแบ่งค่าอำนาจจำแนกที่มีการถ่วงน้ำหนักเงื่อนไขบังคับ (Constraint-Weighted a-Stratification: CWASRT) ผลการศึกษาปรากฏว่า วิธี CWASRT มีประสิทธิภาพมากในเรื่องของความสมดุลของเนื้อหาแต่มีอัตราความแปรปรวนของความคลาดเคลื่อนในการประมาณค่าความสามารถมากกว่าวิธีการคัดเลือกข้อสอบที่ให้สารสนเทศสูงสุด แต่วิธีการคัดเลือกข้อสอบที่ให้สารสนเทศสูงสุดและวิธีการคัดเลือกอย่างสุ่มมีการฝ่าฝืนในเรื่องของข้อบังคับด้านความสมดุลของเนื้อหา นอกจากนี้ยังปรากฏว่า การจัดเรียงข้อสอบในแต่ละชั้นตามวิธีการ Ascending-a ดีกว่าวิธีการ Descending-a ในด้านประสิทธิภาพในการประมาณค่าความสามารถแต่ ด้านประสิทธิภาพของการควบคุมการใช้ข้อสอบซ้ำมีประสิทธิภาพดีใกล้เคียงกัน สนับสนุน ผลการวิจัยของ Hau and Chang, (2001); Chang and Ying, (1999) ที่พบว่า วิธี a-STR และ วิธี a-DSTR ให้ความคลาดเคลื่อนกำลังสองเฉลี่ยและความลำเอียงเฉลี่ยลดลง เหมือนกันเมื่อความยาวแบบทดสอบเพิ่มขึ้น แต่ความคลาดเคลื่อนกำลังสองเฉลี่ยในวิธี a-STR ต่ำกว่า วิธี a-DSTR

Zhang, Chang, and Yi (2012) ได้ศึกษาการเปรียบเทียบคลังข้อสอบแบบเดี่ยวและแบบหลายขั้นตอนโดยทดสอบความปลอดภัยในการทดสอบด้วยคอมพิวเตอร์ ด้วยการแบ่งปันการตรวจระหว่างผู้ตรวจสอบกับคอมพิวเตอร์โดยการใช้สถานการณ์จำลองเพื่อดำเนินการเปรียบเทียบระหว่างคลังข้อสอบที่แตกต่างกัน โดยใช้วิธีการเลือกข้อสอบข้อถัดไปวิธี Maximum Item Information กับวิธี Sympson-Hetter ในการควบคุมการแสดงข้อสอบและการจัดสมดุลเนื้อหา ผลจากการศึกษาปรากฏว่า การจำลองข้อมูลแสดงให้เห็นว่าการออกแบบคลังข้อสอบแบบ 2 คลัง มีความปลอดภัยและมีประสิทธิภาพในด้านของความแม่นยำในการวัดและความสามารถในการประมาณค่าดีขึ้น คลังข้อสอบแบบหลายขั้นตอนสามารถลดจำนวนข้อสอบลงเมื่อเทียบกับการใช้คลังข้อสอบแบบเดี่ยว และการตรวจสอบภายใต้วิธีการคัดเลือกข้อสอบแบบสุ่ม และการทดสอบโดยใช้โปรแกรมคอมพิวเตอร์ในการคัดเลือกข้อสอบแบบสุ่ม ปรากฏว่า การจัดคลังข้อสอบแบบหลายขั้นตอนมีความปลอดภัยและ

การจัดสมดุลเนื้อหาและการควบคุมการแสดงข้อสอบ โดยใช้วิธี a-Stratified-With-b-Blocking Method

Zhou (2012) ได้ศึกษาการออกแบบคลังข้อสอบแบบ p-Optimal ในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์กับข้อสอบแบบพหุวิภาค การศึกษานี้มีวัตถุประสงค์เพื่อการออกแบบคลังข้อสอบแบบ p-Optimal ในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์กับข้อสอบแบบพหุวิภาค โดยขยายวิธี p-Optimal (2007) โดยการจำลองข้อมูล ผลการวิจัยปรากฏว่าข้อจำกัดในทางปฏิบัติของการควบคุมการแสดงข้อสอบวิธีแบ่งชั้น (a-Stratified) และการจัดสมดุลเนื้อหาไม่ได้มีผลต่อขนาดคลังข้อสอบขนาดใหญ่ อย่างไรก็ตามการควบคุมการแสดงข้อสอบวิธีแบ่งชั้น ส่งผลกระทบต่อลักษณะของคลังข้อสอบมาก ข้อสอบที่รวมอยู่ในคลังข้อสอบจำลองที่มีการควบคุมการแสดงข้อสอบวิธีแบ่งชั้น ที่มีการควบคุมมีขนาดใหญ่ พารามิเตอร์และให้ข้อมูลที่สูงกว่าค่าเฉลี่ย ในทางตรงกันข้ามสมดุลเนื้อหานำไปใช้ในการศึกษาครั้งนี้มีผลกระทบเพียงเล็กน้อยในการออกแบบคลังข้อสอบและการดำเนินงานคลังข้อสอบมีความสัมพันธ์อย่างใกล้ชิดกับลักษณะคลังข้อสอบ เมื่อควบคุมการแสดงข้อสอบวิธีแบ่งชั้น ให้ผลที่สอดคล้องกัน ได้แก่ 1) ข้อมูลการทดสอบมีค่าเฉลี่ยต่ำกว่าที่ได้ไม่จำกัด 2) RMSE สูงขึ้นและความสัมพันธ์ระหว่างความจริงและความสามารถในการประเมินต่ำ 3) ร้อยละของการจัดหมวดหมู่ที่ถูกต้องสำหรับความสำเร็จในระดับสูงสุดที่ต่ำ แต่สำหรับทุกคลังข้อสอบที่ดีที่สุดจำลองข้อมูลการทดสอบอย่างต่อเนื่องเหนือระดับเป้าหมาย 10.0 ก็สรุปได้ว่าวิธีการควบคุมการแสดงข้อสอบวิธีแบ่งชั้น ส่งผลให้การใช้งานที่มีประสิทธิภาพของรายการจำแนกน้อยกับการลดลงของขนาดเล็กในการวัดความแม่นยำ

He and Reckase (2014) ได้ศึกษาการออกแบบคลังข้อสอบสำหรับการดำเนินงานตามตัวแปรความยาวกับการทดสอบแบบปรับด้วยคอมพิวเตอร์ โดยมีวัตถุประสงค์ของการวิจัยเพื่อพัฒนาคลังข้อสอบที่เหมาะสมและตรวจสอบประสิทธิภาพการทำงานของคลังข้อสอบ สำหรับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ โดยการขยาย Bin-and-union Method ของ Reckase (2003) โดยมีการควบคุมการแสดงข้อสอบแบบไม่ควบคุมการแสดงข้อสอบและการจัดสมดุลเนื้อหา ผลการศึกษาปรากฏว่า กลไกที่ใช้ในการระบุคุณลักษณะที่พึงประสงค์คลังข้อสอบที่ได้ทำหน้าที่อย่างดีและคลังข้อสอบสองแบบที่พัฒนาขึ้นมีความปลอดภัยในการใช้งาน

Mao (2014) ได้ศึกษาการออกแบบคลังข้อสอบแบบ P-Optimal สำหรับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติโดยคำนึงถึงความจำเป็น และการออกแบบคลังข้อสอบที่เหมาะสม มีการจัดสมดุลข้อสอบเพื่อให้ได้ความแม่นยำในการทดสอบและลดจำนวนข้อสอบและค่าใช้จ่ายลง โดยการพัฒนาวิธี P-Optimal ซึ่งพัฒนาโดย Reckase (2003 & 2007) สำหรับ CAT แบบมิติเดี่ยวขยายวิธีของ Reckase โดยมีจุดมประสงค์เพื่อพัฒนาการออกแบบคลังข้อบแบบ P-Optimal สำหรับการทดสอบแบบ MCAT แบบหลายมิติ และขยายการสร้างข้อสอบแบบหลายมิติแบบ D-Optimal โดยใช้ค่า MDIFF-bin เพื่ออธิบายข้อมูลแบบหลายมิติ สำหรับการออกแบบคลังข้อสอบแบบ P-Optimal ในบริบทแบบหลายมิติ ได้ออกแบบคลังข้อสอบออกเป็น 24 รายการและพัฒนาเงื่อนไขในการทดสอบที่แตกต่างกันโดยมีความสัมพันธ์ระหว่างมิติข้อมูลกับขนาดของคลังข้อสอบที่แตกต่างกันภายใต้เงื่อนไขที่มีและมีการควบคุมการแสดงข้อสอบเพื่อหาประสิทธิภาพของคลังข้อสอบแบบ P-Optimal และ Baseline Pools ปรากฏว่า ข้อสอบแบบ P-Optimal มีความถูก

ต้องของการวัดเหมือนกัน เช่นการลดจำนวนข้อสอบและการทำงานได้ดีขึ้นในด้านความปลอดภัยและสมรรถนะของคลังข้อสอบแบบ P-Optimal มีประสิทธิภาพสูงขึ้น ผลกระทบดังกล่าวมีปัจจัยมาจากข้อกำหนดที่เกี่ยวกับความสัมพันธ์ระหว่างขนาดของคลังข้อสอบและการควบคุมการแสดงข้อสอบ โดยผลการศึกษาสามารถนำไปพัฒนาใช้การการออกแบบคลังข้อสอบแบบ P-Optimail สำหรับโปรแกรม MCAT อื่นๆที่มีคุณสมบัติและจุดประสงค์ที่แตกต่างกันได้ นอกจากการออกแบบคลังข้อสอบแบบ P-Optimail ได้จัดทำคู่มือการพัฒนาคลังข้อสอบและการจัดการโปรแกรมคลังข้อสอบ

Kuo, Daud, and Yang (2015) ได้ศึกษาการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติสำหรับวิชาชีววิทยาในโรงเรียนมัธยมศึกษาตอนต้นของประเทศอินโดนีเซีย โดยสร้างข้อสอบจำนวน 300 ข้อที่แตกต่างกัน จากนั้นนำไปทดสอบกับนักเรียนจำนวน 2,238 คน โดยนำแบบจำลองที่เรียนว่า Multidimensional Random Coefficients Multinomial Logit Model (MRCMLM) มาใช้เพื่อพัฒนาสเกลแบบพหุมิติ ในวิชาชีววิทยาประกอบด้วยเนื้อหาหลัก 6 ด้าน โดยศึกษาวิธีการประมาณค่าความสามารถ 3 วิธี ได้แก่ วิธี MLE วิธี MAP และวิธี EAP ซึ่งประเมินจากข้อมูลจริงเพื่อประเมินการทำงานของระบบ Biology-MCAT System ผลการศึกษาแสดงให้เห็นว่าวิธี MAP มีการประมาณค่าสูงสุด มีประสิทธิภาพมากกว่า วิธ MLE และ วิธี EAP ค่าความคลาดเคลื่อนมาตรฐานและค่าความคลาดเคลื่อนเอนเอียง Bias and Standard Error) ของระบบ Biology-MCAT นั้นสามารถยอมรับได้ โดยชี้ให้เห็นว่า ระบบการประเมินออนไลน์นี้สามารถนำไปใช้แบบจำลองในการสอบระดับประเทศของชาวอินโดนีเซียในอนาคตได้

Seo and Weiss (2015) ได้ศึกษาการออกแบบที่ดีที่สุดสำหรับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติด้วยโมเดลสององค์ประกอบ ในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์(CAT) ส่วนใหญ่มีกรอบการศึกษาโดยใช้ทฤษฎีการตอบสนองข้อสอบแบบมิติเดียว แต่ตัวแปรทางจิตวิทยาต้องวัดหลายตัวและหลายมิติ และอาจได้รับประโยชน์จากการใช้วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ การศึกษานี้ได้ศึกษาการตรวจสอบความถูกต้องความตรงและประสิทธิภาพของมิติข้อมูลแบบเต็มรูปแบบตามขั้นตอนการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติโดยใช้รูปแบบ Bifactor โดยจำลองข้อมูลด้วยการเลือกวิธีการคัดเลือกข้อสอบ 4 วิธีโดยใช้แบบจำลอง bifactor 3 รูปแบบ และ 2 แบบจำลองตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติเพื่อเปรียบเทียบการคัดเลือกข้อสอบและการประมาณค่าความสามารถโดยใช้การทดสอบความยาวคงที่และการคัดเลือกข้อสอบแบบ D - A optimality โดยการประมาณค่าดีขึ้นเกี่ยวกับปัจจัยทั่วไปภายใต้เงื่อนไข 3 รูปแบบตามโมเดล bifactor ตามแบบจำลองทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ แบบ 2 พารามิเตอร์ ที่มีค่าอำนาจจำแนก (a) และ ค่าโอกาสการเดา (c) โดยใช้การประมาณค่าความสามารถแบบ MAP และวิธี EAP ปรากฏว่า วิธี MAP มีความถูกต้องในการประมาณค่าความสามารถที่มีความผิดพลาดที่ต่ำกว่าวิธี EAP ภายใต้เงื่อนไขที่กำหนด ยกเว้นเงื่อนไขทั่วไปในการคัดเลือกข้อสอบแบบ D-Optimality

จากการศึกษางานวิจัยที่เกี่ยวข้องกับวิธีการจัดคลังข้อสอบ ปรากฏว่า ในช่วงเริ่มแรกเป็นงานวิจัยที่พัฒนาวิธีการขึ้นใหม่ และปรับปรุงประสิทธิภาพของวิธีการจัดคลังข้อสอบ ที่มีผู้พัฒนาขึ้นแล้ว ให้มีประสิทธิภาพเพิ่มมากขึ้น ในระยะต่อมางานวิจัยส่วนใหญ่เป็นการศึกษาในสถานการณ์

จำลองร่วมกับการควบคุมการแสดงข้อสอบและการจัดสมดุลเนื้อหาเพื่อเปรียบเทียบประสิทธิภาพของวิธีการจัดคลังข้อสอบ ทีนำไปใช้ในสถานการณ์ต่างๆ เพื่อทราบข้อดีหรือข้อจำกัดของวิธีการนั้น ๆ

การทบทวนเอกสารและงานวิจัยที่เกี่ยวข้องกับวิธีการจัดคลังข้อสอบ ในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ ปรากฏว่า ยังไม่พบงานวิจัยที่ศึกษาในสถานการณ์จริงที่พัฒนาวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา สำหรับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ ดังนั้นผู้วิจัยจึงสนใจพัฒนาวิธีจัดคลังข้อสอบตามระดับชั้นของค่าอำนาจจำแนกของข้อสอบ (a-Stratified Method: a-STR) โดยประยุกต์วิธีการของ Chang and Ying (1996) ซึ่งแบ่งคลังข้อสอบออกเป็น 4 ชั้น ตามช่วงของค่าอำนาจจำแนกข้อสอบ (a) ชั้นที่ 1 0.50–0.99 ชั้นที่ 2 1.00–1.49 ชั้นที่ 3 1.50–1.99 และ ชั้นที่ 4 2.00–2.50 ร่วมกับการจัดสมดุลเนื้อหา (Content Balancing) ภายในชั้นของค่าอำนาจจำแนก ได้แบ่งเป็น 4 กลุ่ม ตามสาระการเรียนรู้เนื้อหาการออกข้อสอบ O-NET วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 6 ได้แก่ 1) สาระการเรียนรู้จำนวนและการดำเนินการ 2) สาระการเรียนรู้การวัด 3) สาระการเรียนรู้พีชคณิต และ 4) สาระการเรียนรู้การวิเคราะห์และความน่าจะเป็น ดังนั้นวิธีการควบคุมการแสดงข้อสอบ (Controlling Item Exposure) จึงใช้วิธีการสุ่มตัวอย่างแบบชั้นภูมิ (Stratified Random Sampling) โดยให้กลุ่มสาระการเรียนรู้เนื้อหาการออกข้อสอบเป็นชั้นภูมิ และให้ค่าความยากของข้อสอบ (b) ที่มีลักษณะเหมือนกัน (Homogenious) เป็นหน่วยการสุ่ม โดยจะพัฒนาให้วิธีการนี้มีประสิทธิภาพ 1) ด้านการประมาณค่าความสามารถของผู้สอบ ได้แก่ ค่าความลำเอียงเฉลี่ย และค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย 2) ด้านประสิทธิภาพของการใช้ข้อสอบ ได้แก่ ข้อสอบที่มีการแสดงมากเกินไป ข้อสอบที่มีการใช้น้อยเกินไป อัตราการทับซ้อนของข้อสอบ และการแจกแจงอัตราการแสดงข้อสอบ

2. งานวิจัยที่เกี่ยวข้องกับการควบคุมการแสดงข้อสอบในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ มีดังนี้

Lee and Fuh (2008) ได้ศึกษายุทธวิธีในการควบคุมการใช้ข้อสอบซ้ำในการทดสอบแบบปรับเหมาะแบบพหุมิติโดยวิธีการที่นำเสนอใช้พื้นฐานของการแบ่งชั้นในความสอดคล้องด้วยฟังก์ชันของเวคเตอร์พารามิเตอร์อำนาจจำแนก ซึ่งสามารถนำไปใช้ในการคำนวณ การศึกษาเปรียบเทียบอัตราการใช้ข้อสอบซ้ำโดยวิธี ASTR และ D-optimality โดยใช้เกณฑ์การคัดเลือกข้อสอบด้วยวิธี Match-b เงื่อนไขการเปรียบเทียบประกอบด้วย 1) ความยาวของข้อสอบ 30 ข้อ และ 60 ข้อ 2) วิธีการ คือ ASTR, D-optimality และ Match-b 3) ความแตกต่างของค่าความสัมพันธ์ระหว่าง θ_1 และ θ_2 ประกอบด้วย $\rho = 0,.3$ และ .6 ปรากฏว่า การเปรียบเทียบด้วยการคัดเลือกข้อสอบวิธี Match-b ผลการวิจัยชี้ว่า วิธีการ ASTR ค่อนข้างมีประสิทธิภาพ และข้อสอบที่มีความยาวมากกว่าจะมีประสิทธิภาพมากกว่า โดยพิจารณาจากค่า Mean Square Error (MSE) ซึ่งมีค่าน้อยกว่า เพื่อพิจารณาที่ความสัมพันธ์ของ θ_1 และ θ_2 เมื่อไม่สนใจวิธีการ ปรากฏว่า ความสัมพันธ์ในระดับสูงจะมีประสิทธิภาพน้อยกว่า ในทางตรงข้าม ค่า ρ ไม่มีผลต่อการประมาณค่าความสามารถ θ เมื่อแบบทดสอบมีความยาวมากกว่า 60 ข้อ

Deng, Ansley, and Chang (2010) ได้ศึกษาเปรียบเทียบประสิทธิภาพของวิธีการ 3 วิธีที่ใช้ในการควบคุมการใช้ข้อสอบ คือ 1) a-Stratified Adaptive Testing (ASTR) 2) การคัดเลือก

อย่างสุ่ม (RAN) และ 3) Low a-Stratified (USTR) โดยพิจารณาจากค่าความแปรปรวนของความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบ และค่าความน่าเชื่อถือของค่าประมาณความสามารถของผู้สอบ ผลจากการศึกษาปรากฏว่า วิธี USTR สามารถลดความแปรปรวนของค่าความคลาดเคลื่อนในการประมาณค่าความสามารถได้เมื่อคลังข้อสอบมีขนาดเล็ก และการประมาณค่าความสามารถของผู้สอบทั้ง 3 วิธี มีประสิทธิภาพไม่แตกต่างกัน

Leroux, Lopez, Hembry, and Dodd (2013) ได้เปรียบเทียบประสิทธิภาพของวิธีการควบคุมการแสดงข้อสอบระหว่างวิธีการ Progressive-restricted Standard Error (PR-SE) กับวิธีการควบคุมการใช้ข้อสอบทั่วไป คือวิธีการ Randomesque วิธีการ Sympson-Hetter (SH) และวิธีการที่ไม่มีการควบคุมการแสดงข้อสอบ ภายใต้โมเดลโลจิสติก 3 พารามิเตอร์ และเปรียบเทียบในกรณีคลังข้อสอบขนาดเล็กเทียบกับคลังข้อสอบขนาดใหญ่ ผลการศึกษาปรากฏว่า วิธีการ PR-SE สามารถเข้าควบคุมการแสดงข้อสอบของข้อสอบในคลังได้เกือบทุกข้อ แต่วิธีการRandomesque กับวิธี SH สามารถควบคุมการแสดงข้อสอบของข้อสอบในคลังได้เพียง 52% และ 80% ในกรณีที่คลังข้อสอบมีขนาดใหญ่และเล็ก ตามลำดับ

Ozturk and Dogan (2015) ได้เปรียบเทียบประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบและการควบคุมการแสดงข้อสอบระหว่างวิธีการควบคุมการใช้ข้อสอบ 3 วิธี ได้แก่ 1) วิธี Randomesque 2) วิธี Sympson-Hetter และ 3) วิธี Fade-Away โดยใช้คลังข้อสอบจำลอง ด้วยการสุ่มค่าอำนาจจำแนกของข้อสอบ (a) จากการแจกแจงแบบยูนิฟอร์มในช่วง 0.50 ถึง 2.00 สุ่มค่าโอกาสในการเดาของข้อสอบ (c) จากการแจกแจงแบบยูนิฟอร์มในช่วง 0.05 ถึง 0.20 ส่วนการจำลองค่าความยากของข้อสอบ (b) แบ่งได้ 2 ส่วน คือ 1) คลังข้อสอบที่มีความยากของข้อสอบปานกลาง ด้วยการสุ่มค่า b จากการแจกแจงแบบยูนิฟอร์มในช่วง -3.00 ถึง 3.00 และ 2) คลังข้อสอบที่มีค่าความยากของข้อสอบสูงด้วยการสุ่มค่า b จากการแจกแจงปกติ N(2, 1.5) ผลการศึกษาปรากฏว่า ประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบทั้ง 3 วิธี ไม่มีความแตกต่างกัน แต่วิธี Fade-Away สามารถควบคุมการแสดงข้อสอบได้ดีกว่าวิธี Randomesque และวิธี Sympson-Hetter

Yao (2016) ได้ศึกษาวิธีการคัดเลือกข้อสอบข้อถัดไปโดยใช้คะแนนโดเมนและคะแนนคอมโพสิท สำหรับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติโดยมีการควบคุมอัตราการแสดงข้อสอบและการจัดสมดุลเนื้อหา โดยมีวัตถุประสงค์เพื่อค้นหาวิธีการคัดเลือกข้อสอบในกรอบทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ (MIRT) ที่มีความแม่นยำทั้งในส่วนของคะแนนโดเมน (Domain Scores) และคะแนนคอมโพสิท (Composite Scores) ใช้ประโยชน์จากคลังข้อสอบ (Item Pool) และควบคุมอัตราการแสดงข้อสอบ โดยทำศึกษาและเปรียบเทียบกระบวนการทั้งห้าอย่าง ในการเลือกรายการด้วยการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ (MCAT) (กระบวนการใช้มุมขั้นต่ำ กระบวนการใช้ปริมาณ กระบวนการใช้ความคลาดเคลื่อนขั้นต่ำของผลรวมเชิงเส้น กระบวนการใช้ความคลาดเคลื่อนขั้นต่ำของคะแนนคอมโพสิทคำนวณด้วยค่าน้ำหนักที่เหมาะสมที่สุด และวิธี Kullback Leibler Information กับวิธีการสองวิธีในการควบคุมอัตราการแสดงข้อสอบ (Sympson-Hetter) และวิธีการแบบอัตราตายตัว ซึ่งวิธีหลังการทดสอบจำกัดอัตราการแสดงข้อสอบคงที่) ซึ่งใช้ข้อมูลที่จำลองขึ้นมา โดยมีการใช้ดัชนีความสำคัญสูงสุดกับการจัดสมดุล

เนื้อหา ผลการวิจัยแสดงให้เห็นว่า วิธี Sympson-Hetter มีความแม่นยำมากกว่า แต่มีการใช้คลังข้อสอบน้อยกว่าและใช้เวลามากกว่าวิธีการแบบอัตราตายตัว วิธีการคัดเลือกข้อสอบทั้งห้าวิธี มีประสิทธิภาพใกล้เคียงกันภายใต้วิธี Sympson-Hetter สำหรับวิธีการแบบอัตราตายตัวมีความแม่นยำในการเลือกมากกว่า แต่ก็ต้องใช้คลังรายการมากกว่าวิธีการอื่นด้วย วิธีการทั้งห้าอย่างมีรูปแบบที่แตกต่างกันไป ปรากฏว่า (1) วิธี Kullback Leibler Information ความแม่นยำมากกว่าแต่มีการใช้คลังข้อสอบน้อยกว่าวิธีการอื่น (2) วิธี Minimum Agle and Volume มีความแม่นยำและการใช้คลังข้อสอบที่สมดุลกัน และ (3) วิธีการทั้งสองวิธีในการลดความผันผวนของความคลาดเคลื่อนใช้คลังข้อสอบได้ดีที่สุด และมีการรักษาคะแนนโดยรวมไว้ในระดับที่สูงพอประมาณ แต่มีความแม่นยำน้อยในบางโดเมน งานวิจัยนี้มีการจัดทำดัชนีความสำคัญสำหรับการจัดสมดุลเนื้อหาและการควบควบคุมอัตราการแสดงข้อสอบ

Huebner, Wang, Quinlan, and Seubert (2016) ได้ศึกษาการควบคุมการแสดงข้อสอบสำหรับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติโดยใช้วิธี MLE และวิธี EAP โดยแบ่งคลังข้อสอบแบบแบ่งชั้น โดยพิจารณาประสิทธิภาพการต่อต้านข้อสอบและการแสดงข้อสอบมากเกินไป สองวิธีในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบมิติเดียวและแบบพหุมิติ อย่างไรก็ตามการแบ่งชั้นของคลังข้อสอบไม่สามารถบ่งบอกว่าคลังข้อสอบนั้นๆ จะมีการคุมควบการแสดงข้อสอบ ซึ่งการควบคุมการแสดงข้อสอบมากเกินไป จะกำหนดเงื่อนไขไว้ล่วงหน้า ในการศึกษามีการสร้างคลังข้อสอบแบบแบ่งชั้นสำหรับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ ร่วมกับวิธีการคัดเลือกข้อสอบโดยใช้เทคนิคการควบคุมการแสดงข้อสอบที่มากเกินไป สำหรับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์โดยพิจารณาประสิทธิภาพของวิธีการตรวจสอบความถูกต้องโดยการจำลองข้อมูลและเปรียบเทียบกับวิธีการคัดเลือกข้อสอบและการควบคุมการทับซ้อนข้อสอบ และวิธี MLE และวิธี EAP สำหรับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ ผลการจำลองข้อมูลปรากฏว่า วิธีที่นำเสนอมีประสิทธิภาพในการควบคุมอัตราการแสดงข้อสอบมากเกินไปที่เกิดความผิดพลาดน้อยและมีความถูกต้องแม่นยำ เมื่อเทียบกับวิธี MLE กับ วิธี EAP มีความผิดพลาดเล็กน้อยในสภาวะการจำลองข้อมูล

Mao et al. (2016) ได้ปรียบเทียบกลยุทธ์การควบคุมการแสดงข้อสอบในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติโดยการประเมินดัชนีการควบคุมการแสดงข้อสอบ 4 วิธี ที่มีและไม่มีการควบคุมการแสดงข้อสอบในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ ดัชนีการเลือกข้อสอบข้อถัดไป 4 วิธี ได้แก่ D-Optimal, Posterior Expectation , Kullback-Leibler Information (KLP) ความแปรปรวนของข้อผิดพลาดที่น้อยที่สุดของคะแนนรวมเชิงเส้นที่มีน้ำหนักเท่ากัน (V1) และความแปรปรวนของข้อผิดพลาดที่น้อยที่สุดของคะแนนคอมโพสิตที่มีน้ำหนักเหมาะสมที่สุด (V2) และวิธี Maximum Priority Index (MPI) สำหรับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติและวิธีการควบคุมการแสดงข้อสอบแบบ Restrictive Threshold (RT), Restrictive Progressive (RPG) และ Originally Proposed for Cognitive Diagnostic CAT ผลการวิจัยปรากฏว่า 1) วิธี KLP, D-Optimail และ V1 มีประสิทธิภาพในการกู้คืนข้อมูลโดเมนได้ดีกว่า V2 ในด้านความแม่นยำในการวัด 2) KLP, D-Optimal, V1 และ V2 มีอัตราการควบคุมการแสดงข้อสอบในคลังข้อสอบเพิ่มขึ้น และ 3) กลยุทธ์การควบคุมการแสดงข้อสอบทั้งหมดช่วยเพิ่ม

ประสิทธิภาพความสม่ำเสมอในการควบคุมการแสดงข้อสอบและลดจำนวนข้อสอบลงและมีความแม่นยำในการวัด

จากการศึกษางานวิจัยที่เกี่ยวข้องกับการควบคุมการแสดงข้อสอบ ปรากฏว่า ในช่วงเริ่มแรกเป็นงานวิจัยที่พัฒนาวิธีการขึ้นใหม่ และปรับปรุงประสิทธิภาพของวิธีการควบคุมการแสดงข้อสอบที่มีผู้พัฒนาขึ้นแล้ว ให้มีปประสิทธิภาพเพิ่มมากขึ้น ในระยะต่อมางานวิจัยส่วนใหญ่เป็นการศึกษาในสถานการณ์จำลองเพื่อเปรียบเทียบประสิทธิภาพของวิธีการควบคุมการแสดงข้อสอบที่นำไปใช้ในสถานการณ์ต่างๆ เพื่อทราบข้อดีหรือข้อจำกัดของวิธีการนั้นๆ

การทบทวนเอกสารและงานวิจัยที่เกี่ยวข้องกับวิธีการควบคุมการแสดงข้อสอบในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ ปรากฏว่า ยังไม่พบงานวิจัยใดที่นำวิธีการสุ่มตัวอย่างแบบชั้นภูมิมาประยุกต์กับวิธีการควบคุมการแสดงข้อสอบ ดังนั้นผู้วิจัยจึงมีความสนใจในการพัฒนาวิธีการควบคุมการแสดงข้อสอบด้วยวิธีการสุ่มตัวอย่างแบบชั้นภูมิ โดยพัฒนาให้วิธีการนี้มีประสิทธิภาพ 1) ด้านการประมาณค่าความสามารถของผู้สอบ ได้แก่ ค่าความลำเอียงเฉลี่ย และค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย และ 2) ด้านประสิทธิภาพของการใช้ข้อสอบ ได้แก่ ข้อสอบที่มีการแสดงมากเกินไป ข้อสอบที่มีการใช้น้อยเกินไป อัตราการทับซ้อนของข้อสอบ และการแจกแจงอัตราการแสดงข้อสอบ

3. งานวิจัยที่เกี่ยวข้องกับการจัดสมดุลเนื้อหา ในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ มีดังนี้

Belov, Armstrong, and Weissman (2008) ได้เสนอวิธีการใหม่สำหรับการควบคุมเนื้อหา (Content Constraints) โดยเปรียบเทียบวิธี Shadow CAT กับวิธี Monte Carlo CAT ที่ใช้ควบคุมเนื้อหา การศึกษาในครั้งนี้ใช้คลังข้อสอบจากการสอบเข้าศึกษาต่อในด้านกฎหมาย (Law School Admission Test: LSAT) ผลการศึกษาปรากฏว่า 1) Monte Carlo CAT มีอัตราการใช้ข้อสอบซ้ำสูงสุดต่ำกว่า และมีการใช้ประโยชน์ของคลังข้อสอบดีกว่า โดยเฉพาะอย่างยิ่งอัตราการใช้ข้อสอบซ้ำสูงสุดสำหรับ Monte Carlo CAT มีค่าต่ำกว่าวิธี Shadow CAT 2) เมื่อมีโอกาสในการเดาเกิดขึ้นทั้ง Monte Carlo CAT และ Shadow CAT มี Bias เพิ่มขึ้นทั้งคู่ และ MSE ของ Monte Carlo CAT แทบจะไม่มีการเปลี่ยนแปลงขณะที่ของ Shadow CAT เพิ่มขึ้น อย่างไรก็ตามการเปลี่ยนแปลงเหล่านี้ไม่ได้เพิ่มขึ้นมาก และ3) Monte Carlo CAT สามารถรองรับผู้สอบทั้งหมด 2,723 คนที่ทำข้อสอบพร้อมๆ กัน ไม่เกิน 5 นาที ในการนำข้อสอบออกมาใช้ ขณะที่ Shadow CAT ใช้เวลามากกว่า 20 นาที กับผู้สอบเพียง 134 คน โดยสรุปแล้ว Monte Carlo CAT ให้ผลดีกว่าอย่างเห็นได้ชัดเจนทั้งในเรื่องอัตราการใช้ข้อสอบซ้ำ Bias และ MSE ดังนั้น Monte Carlo CAT จึงเหมาะสมกับการ นำไปใช้ในทางปฏิบัติเพื่อบริหารการสอบ

Han (2012) ได้พัฒนาวิธีการคัดเลือกข้อสอบข้อถัดไปโดยใช้ประสิทธิภาพของข้อมูลเพื่อหาวิธีการคัดเลือกข้อสอบที่มีการรักษาความสมดุลของเนื้อหาโดยใช้การเปรียบเทียบ 2 วิธี คือ 1) วิธีการคัดเลือกข้อสอบข้อถัดไปโดยใช้ค่าสารสนเทศของฟิชเชอร์สูงสุด (Maximum Fisher Information: MFI) และ 2) วิธีการคัดเลือกข้อสอบข้อถัดไปโดยวิธีการแบ่งชั้น (Stratification Methods) ผลการศึกษาปรากฏว่า วิธีการคัดเลือกข้อสอบข้อถัดไปโดยใช้ประสิทธิภาพของข้อมูลมีการรักษาความสมดุลของเนื้อหามากกว่าวิธีการคัดเลือกข้อสอบข้อถัดไปโดยใช้สารสนเทศของ

ฟิชเชอร์สูงสุดและวิธีการคัดเลือกข้อสอบข้อถัดไปโดยวิธีการแบ่งชั้น

Culbertson (2014) ได้ศึกษาการสร้างสมดุลของข้อมูลด้วยคลังข้อสอบที่ขาดสมดุลสำหรับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ ในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ (MCAT) หากคลังข้อสอบให้ข้อมูลในด้านหนึ่งมากกว่าด้านอื่นๆ อัลกอริทึ่มในการคัดเลือกข้อสอบข้อถัดไปจะให้ความสำคัญกับด้านที่มีข้อมูลมาก งานวิจัยนี้ได้ประเมินหลักเกณฑ์การเลือกสองอย่างโดยอิงตามข้อมูลที่ใช้ร่วมกัน ซึ่งมีวัตถุประสงค์ในการปรับความแม่นยำในการประมาณความสามารถในด้านต่างๆ ให้เท่ากัน

Sari and Huggins-Manley (2017) ได้ศึกษาการตรวจสอบการจัดสมดุลเนื้อหาในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ โดยศึกษาแบบจำลองเพื่อทดสอบความแม่นยำของผลการทดสอบในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ (CAT) และการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบหลายขั้นตอน (ca-MST) เมื่อจำนวนพื้นที่เนื้อหาแตกต่างกันไปตามช่วงความยาวที่หลากหลาย โดยได้เปรียบเทียบ CAT และการออกแบบ ca-MST สองรูปแบบ (แบบ 1-3 และ 1-3-3) ในเงื่อนไขต่างๆ รวมถึงความยาวของแบบทดสอบทั้งหมด (ความยาว 24 ข้อ และ 48 ข้อ) และจำนวนพื้นที่เนื้อหาที่มีการควบคุม ระดับเนื้อหา 5 ระดับของพื้นที่เนื้อหาประกอบด้วยพื้นที่เนื้อหาที่เป็นศูนย์ (ไม่มีการจัดสมดุลเนื้อหา), 2, 4, 6 และ 8 ได้วางเงื่อนไขทั้งหมดที่ได้รับและการจัดการภายใน CAT และ CA-MST จำลองข้อมูลสร้างผู้สอบ 4000 คนจาก N (0,1) ภายใต้เงื่อนไขของแบบจำลอง IRT เพื่อพิจารณาอัตราการสัมผัสของ CAT และ ca-MSTs ผลการวิจัยปรากฏว่า ความยาวของแบบและรูปแบบการบริหารแบบทดสอบมีผลต่อผลลัพธ์มากกว่าจำนวนพื้นที่เนื้อหา ถ้าไม่คำนึงถึงสภาพการศึกษาใด ๆ CAT มีประสิทธิภาพดีกว่า ca-MSTs และ ca - MSTs ได้เปรียบเรื่องผลลัพธ์ในการเชื่อมต่อกับการควบคุมการออกแบบการทดสอบเนื้อหา การทดสอบมีประสิทธิภาพคุ้มค่าใช้จ่ายและการใช้ข้อสอบ

จากการศึกษางานวิจัยที่เกี่ยวข้องกับการจัดสมดุลเนื้อหา ปรากฏว่า ในช่วงเริ่มแรกเป็นงานวิจัยที่พัฒนาวิธีการขึ้นใหม่ และปรับปรุงประสิทธิภาพของการจัดสมดุลเนื้อหา ที่มีผู้พัฒนาขึ้นแล้ว ให้มีปประสิทธิภาพเพิ่มมากขึ้น ในระยะต่อมางานวิจัยส่วนใหญ่เป็นการศึกษาในสถานการณ์จำลองเพื่อเปรียบเทียบประสิทธิภาพของการจัดสมดุลเนื้อหา ที่นำไปใช้ในสถานการณ์ต่างๆ เพื่อทราบข้อดีหรือข้อจำกัดของวิธีการนั้นๆ

การทบทวนเอกสารและงานวิจัยที่เกี่ยวข้องกับการจัดสมดุลเนื้อหาในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ ปรากฏว่า ยังไม่พบงานวิจัยใดที่นำวิธีการสุ่มตัวอย่างแบบชั้นภูมิมาประยุกต์กับวิธีการควบคุมการแสดงข้อสอบและการจัดสมดุลเนื้อหา ตามสาระการเรียนรู้ที่ออกข้อสอบ สำหรับการทดสอบ O-NET วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 6 ดังนั้นผู้วิจัยจึงมีความสนใจที่จะพัฒนาวิธีการสุ่มตัวอย่างแบบชั้นภูมิมาประยุกต์กับวิธีการควบคุมการแสดงข้อสอบและการจัดสมดุลเนื้อหา โดยจะพัฒนาให้วิธีการนี้มีประสิทธิภาพ 1) ด้านการประมาณค่าความสามารถของผู้สอบ ได้แก่ ค่าความลำเอียงเฉลี่ย และค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย 2) ด้านประสิทธิภาพของการใช้ข้อสอบ ได้แก่ ข้อสอบที่มีการแสดงมากเกินไป ข้อสอบที่มีการใช้น้อยเกินไป อัตราการทับซ้อนของข้อสอบ และการแจกแจงอัตราการแสดงข้อสอบ

4. งานวิจัยที่เกี่ยวข้องกับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ มีดังนี้

Finch (2010) ได้ศึกษาความลำเอียงและความแม่นยำในการประมาณค่าพารามิเตอร์ของข้อสอบจากโมเดลการตอบสนองข้อสอบแบบพหุมิติโดยอาศัยโมเดลการวิเคราะห์องค์ประกอบเชิงยืนยัน มีวัตถุประสงค์เพื่อศึกษาความลำเอียง (Bias) ความคลาดเคลื่อนมาตรฐาน (SE) และรากที่สองของความคลาดเคลื่อนเฉลี่ยกำลังสอง (RMSE) โดยสร้างข้อมูลจำลองด้วยเทคนิคมอนติคาโล (Monte Carlo) เป็นข้อมูลชนิดสองค่าที่มีความหลากหลายทั้งจำนวนของผู้สอบ (250, 500, 1,000 และ 2,000) ความยาวของข้อสอบ (15, 30 และ 60) การแจกแจงของคุณลักษณะแฝง (Normal, Skewed) ความสัมพันธ์ขององค์ประกอบร่วม (0, 0.3, 0.5 และ 0.8) ชนิดของโมเดล (2 พารามิเตอร์ และ 3 พารามิเตอร์ ที่มีค่าโอกาสการเดาของข้อสอบ) ใช้วิธีการประมาณค่า 2 วิธี คือ วิธีกำลังสองน้อยที่สุดไม่ถ่วงน้ำหนัก (Unweighted Least Squares: ULS) และวิธีกำลังสองน้อยที่สุดถ่วงน้ำหนัก (Robust Weighted Least Squares : RWLS) ซึ่งวิธี ULS เป็นการประมาณค่าโดยใช้โปรแกรมสำเร็จรูป NOHARM และวิธี RWLS เป็นการประมาณค่าโดยใช้โปรแกรม Mplus ทั้งสองโปรแกรมนี้เป็นการวิเคราะห์พารามิเตอร์ของข้อสอบในลักษณะพหุมิติ และยังได้วิเคราะห์พารามิเตอร์ของข้อสอบในลักษณะเอกมิติด้วยโปรแกรม BILOGMG ผลการวิจัย ปรากฏว่า มีความสัมพันธ์ที่ซับซ้อนระหว่างตัวแปรที่ศึกษาและความแม่นยำในการประมาณค่าของวิธีการ จากการศึกษางานวิจัยทั้งในและต่างประเทศจะเห็นได้ว่าการนำการทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์ไปใช้เพื่อการประเมินเชิงวินิจฉัยยังไม่แพร่หลาย และเป็นงานวิจัยที่ประยุกต์ใช้กับทฤษฎีการตอบสนองข้อสอบแบบมิติเดียว ส่วนการประยุกต์ใช้กับทฤษฎีการตอบสนองข้อสอบแบบพหุมิติยังปรากฏว่ามีน้อยมาก ดังนั้น การประยุกต์ใช้ทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ และการทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์เพื่อการประเมินวินิจฉัยข้อบกพร่องทางการเรียนของนักเรียน จึงยังต้องการการดำเนินการวิจัยที่หลากหลายเพื่อให้ได้ข้อสรุปที่ชัดเจนขึ้น

Seo (2011) ได้ศึกษาวิธีการประมาณค่าความสามารถและวิธีการคัดเลือกข้อสอบของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ (MBICAT) ตามโมเดลการตอบสนองข้อสอบแบบ 2 พารามิเตอร์ (2PL) โดยการจำลองข้อมูล (Simulation Data) การเปรียบเทียบวิธีการประมาณค่าความสามารถ 3 วิธี ได้แก่ วิธี MLE วิธี MAP และ EAP โดยใช้วิธีการคัดเลือกข้อสอบแบบ Maximizing the Determinant of the Posterior Information โดยเปรียบเทียบวิธีการประมาณจากรากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) และความคลาดเคลื่อนมาตรฐานที่สังเกตได้ (OSE) เกณฑ์การยุติการทดสอบเป็นแบบแปรผัน (Variable Length) โดยกำหนดความคลาดเคลื่อนมาตรฐานที่สังเกตได้ เท่ากับ .50 และ .55 ผลการศึกษา ปรากฏว่า วิธีการประมาณค่าความสามารถของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ (MBICAT) มีประสิทธิภาพและความแม่นยำกว่าวิธีการประมาณค่าความสามารถของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบเอกมิติ (BICAT) กรณี Group Factor

Wang and Chang (2011) ได้ศึกษาเปรียบเทียบวิธีการคัดเลือกข้อสอบสำหรับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ โดยศึกษาจากการจำลองข้อมูลใน 2 มิติ ในลักษณะของ Within Item Multidimensionality โดยได้เปรียบเทียบวิธีการคัดเลือกข้อสอบระหว่าง

D-optimality (Fisher Information), KL Information Index (KL), Continuous Entropy Method (CEM) และ mutual information (MI) ผลวิจัยปรากฏว่า mutual information ไม่เพียงแต่จะปรับปรุงความแม่นยำในการประมาณค่าแต่ยังให้ค่าความคลาดเคลื่อนต่ำที่สุด Diao and Reckase (2009) ศึกษาเปรียบเทียบวิธีการคัดเลือกข้อสอบสำหรับการทดสอบแบบพหุมิติ ด้วยวิธี D-optimality, A-optimality และ KL ผลวิจัยปรากฏว่าทั้งสามวิธีให้ผลไม่แตกต่างกัน

Zhang (2012) ได้เปรียบเทียบวิธีการประมาณค่าพารามิเตอร์ข้อสอบที่มีลักษณะเป็นโครงสร้างง่าย ๆ (เอกมิติ) กับโครงสร้างผสม (พหุมิติภายในข้อ) ระหว่างการวิเคราะห์แบบเอกมิติและพหุมิติ โดยใช้การประมาณค่าด้วยวิธีการ Joint Maximum Likelihood และ Marginal Maximum Likelihood กับโมเดลแบบเอกมิติ 2 และ 3 พารามิเตอร์ และโมเดลพหุมิติแบบ 2 และ 3 พารามิเตอร์ วิเคราะห์ข้อมูลที่จำลองขึ้นด้วยโปรแกรม BILOG , PARSCALE และ ASSEST ผลการวิจัยปรากฏว่า เมื่อจำนวนสอบมีน้อยการวิเคราะห์แบบพหุมิติมีความถูกต้องในการประมาณค่าพารามิเตอร์ของข้อสอบมากกว่าในขณะที่การวิเคราะห์แบบเอกมิติจะดีกว่าเมื่อจำนวนข้อสอบในแต่ละมิติมีมากพอ นอกจากนี้ยังปรากฏว่าการละเมิดข้อตกลงด้านความเป็นเอกมิติ โดยข้อสอบเป็นพหุมิติแต่วิเคราะห์แบบเอกมิติ โมเดลจะประมาณค่าไม่ถูกต้อง และสัมประสิทธิ์ความสัมพันธ์ระหว่างความสามารถจะมีค่าสูงเกินไป

Tseng (2016) ได้พัฒนาโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์สำหรับวัดคำศัพท์ภาษาอังกฤษ เพื่อทดแทนการทดสอบแบบกระดาษและดินสอ ซึ่งมีความยุ่งยากและเสียค่าใช้จ่ายมาก การดำเนินการพัฒนาแบ่งเป็น 3 ส่วน ได้แก่ ส่วนที่ 1 การพัฒนาคลังข้อสอบ ประกอบด้วยข้อสอบจำนวนทั้งสิ้น 240 ข้อ มีลักษณะเป็นรายการคำตอบแบบ 4 ตัวเลือก แบ่งเป็นชุดข้อสอบแบบคู่ขนานได้ 5 ชุด โดย 4 ชุดแรกมีข้อสอบชุดละ 45 ข้อ และชุดสุดท้ายเป็นข้อสอบพื้นฐาน 15 ข้อ วิเคราะห์ข้อสอบโดยใช้โมเดลโลจิสติกแบบหนึ่งพารามิเตอร์ หรือ โมเดลราส์ช (One-Parameter Logistic Model or Rasch Model) กลุ่มตัวอย่างที่ใช้ในการวิจัยเป็นนักเรียนระดับมัธยมปลายในไต้หวัน จำนวน 1,536 คน ผลการวิเคราะห์ความเที่ยง ปรากฏว่า ข้อสอบในแต่ละชุดมีความเที่ยงสูง และโมเดลราส์ชมีความสอดคล้องกับข้อมูลเชิงประจักษ์ ส่วนที่ 2 การพัฒนาขั้นตอนการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ประกอบด้วยวิธีการคัดเลือกข้อสอบข้อถัดไปโดยการสุ่มข้อสอบข้อหนึ่ง จากกลุ่มข้อสอบที่มีค่าความยากของข้อสอบแตกต่างจากค่าประมาณความสามารถของผู้สอบในช่วง -0.5 ถึง 0.5 ใช้วิธีของ Bayesian ในการประมาณค่าความสามารถของผู้สอบ และใช้เกณฑ์ยุติการทดสอบ 2 แบบ คือ เกณฑ์ยุติการทดสอบแบบความยาวคงที่ (Fixed-length) ที่ 30 ข้อ ร่วมกับเกณฑ์ยุติการทดสอบแบบความยาวยืดหยุ่น (Variable-length) ที่มีค่าความคลาดเคลื่อนมาตรฐานของการประมาณค่าความสามารถ น้อยกว่า 0.3 และส่วนที่ 3 การจำแนกกลุ่มผู้สอบผ่านกับไม่ผ่าน จากการประเมินผลการใช้งานโปรแกรม ปรากฏว่า โปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์สำหรับวัดคำศัพท์ภาษาอังกฤษ ใช้ข้อสอบ เพียง 1 ใน 3 ของการทดสอบแบบกระดาษและดินสอ รวมทั้งการประมาณค่าความสามารถของผู้สอบมีความแม่นยำสูง

Martin-Fernandez and Revuelta (2017) ได้เปรียบเทียบการทำงานของ Algorithms ด้านการประมาณค่า 2 แบบ สำหรับการใช้งานแบบใหม่ ได้แก่ Metropolis-Hastings Robins-Monro (MHRM) และ Hamiltonian (HMC) ด้วย Alogorithms แบบรวม 2 Algorithms

(Consolidated Alogorithms) การศึกษานี้ได้ประเมินการทำงานสำหรับการคืนตัวของพารามิเตอร์ผ่านการศึกษาแบบจำลอง 3 แบบ จากวิธี Bayesian โดยแบบจำลองที่ 1 อาศัยโมเดลแบบมิติเดียวอย่างง่าย เพื่อประเมินผลกระทบของการแจกแจงก่อนหน้าที่กระจายตัวและที่รวมกันสำหรับการคืนตัว ส่วนแบบจำลองที่ 2 เป็นการเปรียบเทียบ Algorithms ระหว่าง MHRM กับ MML-EM และ MCMC ในการประมาณค่าโมเดลการตอบสนองข้อสอบกับจำนวนมิติที่มียากของ ความสัมพันธ์ในระดับปานกลาง แบบจำลองที่ 3 เป็นการประเมินการทำงานของ MHRM, MML-EM และ MCMC ในการประมาณค่าโมเดลการตอบสนองข้อสอบแบบพหุมิติ ผลการทดลองแสดงให้เห็นว่า MML-EM สูญเสียความเที่ยงตรงกับโมเดลที่มีมิติสูง ในขณะที่ 3 Algorithms ที่เหลือช่วยคืนตัวพารามิเตอร์ที่แท้จริงด้วยความตรง ซึ่งมีลักษณะคล้ายคลึงกัน นอกจากนี้แล้วความแตกต่างที่สำคัญระหว่าง Algorithms ต่าง ๆ ได้แก่ 1) เวลาการประมาณค่า MHRM สั้นกว่าวิธีอื่น ๆ 2) MHRM มีความตรงดีที่สุดในทุกเงื่อนไขและได้รับผลกระทบน้อยกว่าจากการแจกแจงก่อนหน้านี้ 3) การแจกแจงก่อนหน้าสำหรับความชัน (Slopes) ใน MCMC และ HMC กำหนดไว้อย่างระมัดระวังเพื่อหลีกเลี่ยงปัญหาทางด้านปัจจัย กล่าวโดยสรุปคือ วิธี MHRM และ HMC ลดความยุ่งยากของ Algorithms แบบเดิมโดยมีการมาบรรจบกัน (Converging) ได้เร็วกว่าและทำให้ได้ผลลัพธ์ที่มีความตรงและแม่นยำกว่า

จากการศึกษางานวิจัยที่เกี่ยวข้องกับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิต ปรากฏว่า ในช่วงเริ่มแรกเป็นงานวิจัยเปรียบเทียบวิธีการประมาณค่าความสามารถและวิธีการคัดเลือกข้อสอบข้อถัดไป และพัฒนาเกณฑ์การยุติการทดสอบ ในระยะต่อมางานวิจัยส่วนใหญ่เป็นการศึกษาในสถานการณ์จำลองเพื่อเปรียบเทียบวิธีการประมาณค่าความสามารถและวิธีการคัดเลือกข้อสอบข้อถัดไป และเกณฑ์การยุติการทดสอบ ที่นำไปใช้ในสถานการณ์ต่างๆ เพื่อทราบข้อดีหรือข้อจำกัดของวิธีการนั้นๆ

การทบทวนเอกสารและงานวิจัยที่เกี่ยวข้องกับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ ปรากฏว่า ยังไม่พบงานวิจัยใดที่นำพัฒนาวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา และวิธีการควบคุมการแสดงข้อสอบ (Controlling Item Exposure) จึงใช้วิธีการสุ่มตัวอย่างแบบชั้นภูมิ (Stratified Random Sampling) ดังนั้นผู้วิจัยจึงสนใจพัฒนาวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา และวิธีการควบคุมการแสดงข้อสอบ โดยพัฒนาให้วิธีการนี้มีประสิทธิภาพ 1) ด้านการประมาณค่าความสามารถของผู้สอบ ได้แก่ ค่าความลำเอียงเฉลี่ย และค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย 2) ด้านประสิทธิภาพของการใช้ข้อสอบ ได้แก่ ข้อสอบที่มีการแสดงมากเกินไป ข้อสอบที่มีการใช้น้อยเกินไป อัตราการทับซ้อนของข้อสอบ และการแจกแจงอัตราการแสดงข้อสอบ

บทที่ 3
วิธีดำเนินการวิจัย

การวิจัยนี้มีวัตถุประสงค์เพื่อ 1) พัฒนาวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา สำหรับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ 2) เปรียบเทียบประสิทธิภาพของวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ระดับชั้นร่วมกับการจัดสมดุลเนื้อหา กับวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบถ่วงน้ำหนักที่มีการบังคับ โดยเปรียบเทียบประสิทธิภาพจาก (1) ด้านการประมาณค่าความสามารถของผู้สอบ ได้แก่ ค่าความลำเอียงเฉลี่ย และค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (2) ด้านประสิทธิภาพของการใช้ข้อสอบ ได้แก่ ข้อสอบที่มีการแสดงมากเกินไป ข้อสอบที่มีการใช้น้อยเกินไป อัตราการทับซ้อนของข้อสอบ และการแจกแจงอัตราการแสดงข้อสอบ 3) พัฒนาโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ ที่ใช้วิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา และ 4) เปรียบเทียบความสามารถของผู้สอบแต่ละระดับชั้นของค่าอำนาจจำแนกของข้อสอบ สำหรับการทดสอบ O-NET วิชาคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 จำแนกตามเพศและความสามารถด้านการเรียนวิชาคณิตศาสตร์ ข้อมูลที่ใช้ในการวิจัยเป็นข้อมูลทุติยภูมิ ใช้ข้อสอบ O-NET วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 6 ระหว่างปีการศึกษา 2551–2553 และปีการศึกษา 2555–2559 ที่ได้มาจากสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) วิธีดำเนินการวิจัยแบ่งเป็น 4 ระยะ ดังนี้

ระยะที่ 1 การพัฒนาวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา

ระยะที่ 2 การเปรียบเทียบประสิทธิภาพของวิธีการจัดคลังข้อสอบโดยการจำลองข้อมูล

ระยะที่ 3 การพัฒนาโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติโดยใช้คลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา

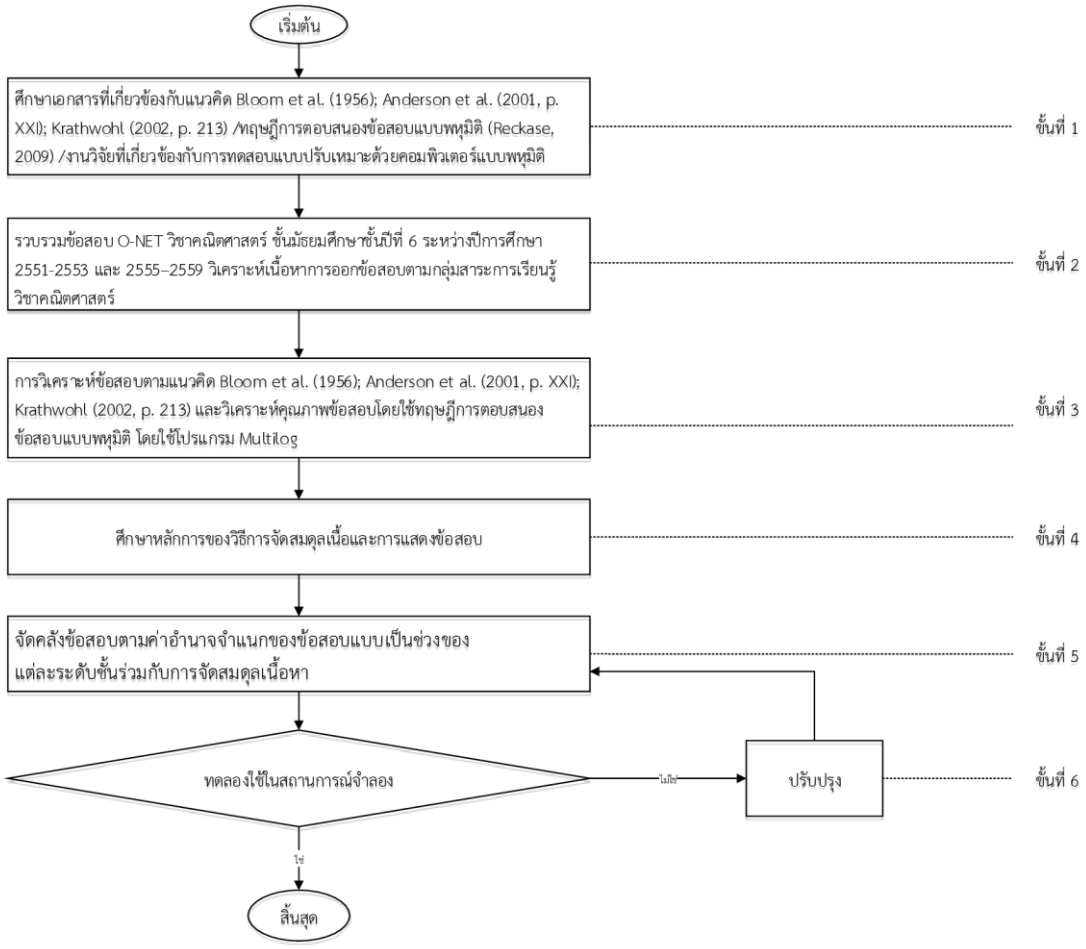
ระยะที่ 4 การเปรียบเทียบความสามารถของผู้สอบแต่ละระดับชั้นของค่าอำนาจจำแนกของข้อสอบ สำหรับการทดสอบ O-NET วิชาคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 จำแนกตามเพศและความสามารถด้านการเรียนวิชาคณิตศาสตร์

การพัฒนาวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา สำหรับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ มีการดำเนินการวิจัย 4 ระยะ ดังภาพที่ 3-1

ภาพที่ 3-1 ระยะการดำเนินการวิจัย

ระยะที่ 1 การพัฒนาวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา

1. การพัฒนาวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา มีขั้นตอนการดำเนินการ ตามภาพที่ 3-2



ภาพที่ 3-2 ขั้นตอนการพัฒนาวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา

จากภาพที่ 3-2 ขั้นตอนการพัฒนาวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา โดยประยุกต์โมเดลการตอบสนองข้อสอบแบบพหุมิติ มีรายละเอียดในการดำเนินการดังนี้

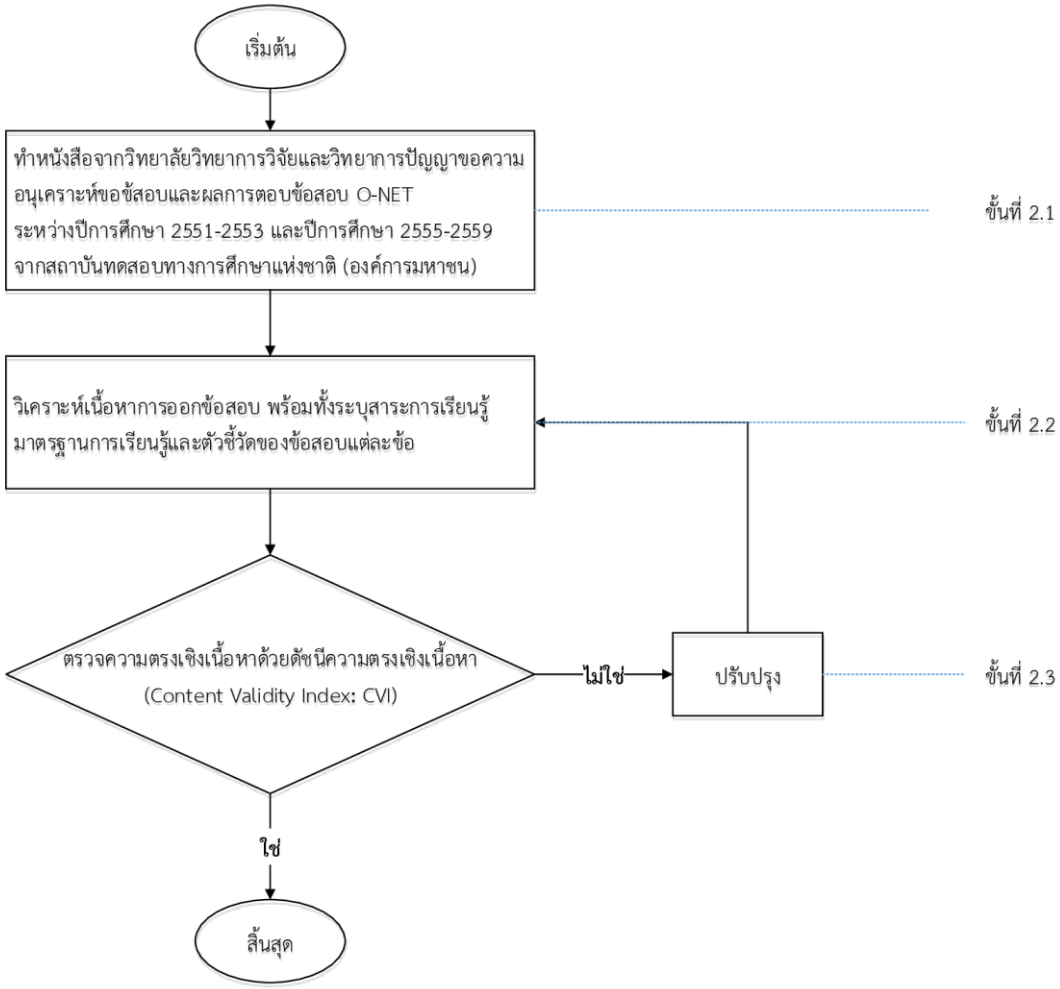
ขั้นที่ 1 การศึกษาเอกสารที่เกี่ยวข้องกับแนวคิดของ Bloom et al. (1956); Anderson et al. (2001, p. XXI) และ Krathwohl (2002, p. 213) ทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ (Reckase, 2009) และงานวิจัยที่เกี่ยวข้องกับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ สรุปได้ดังนี้

1.1 ข้อสอบ O-NET วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 6 ระหว่างปีการศึกษา 2551–2553 และปีการศึกษา 2555–2559 ตามเนื้อหาสาระการเรียนรู้ที่การออกข้อสอบ O-NET วิชาคณิตศาสตร์ ได้แก่ 1) จำนวนและการดำเนินการ 2) การวัด 3) พีชคณิต และ 4) การวิเคราะห์ข้อมูลและความน่าจะเป็น ที่ได้มาจากสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน)

1.2 โมเดลการตอบสนองข้อสอบแบบพหุมิติเป็นโมเดลการวัดที่ถือว่าการตอบสนองข้อสอบของผู้สอบเกิดจากคุณลักษณะแฝงของบุคคลมากกว่า 1 องค์ประกอบ ซึ่งการพิจารณาความสามารถของคนจากหลายองค์ประกอบ จะช่วยให้โมเดลมีความสอดคล้องกับข้อมูลได้ดียิ่งขึ้น ทั้งนี้โมเดลการตอบสนองข้อสอบแบบพหุมิติมีอยู่หลากหลายโมเดล เช่น โมเดลโลจิสติกแบบพหุมิติ โมเดลปกติสะสมแบบพหุมิติ โมเดลสำหรับการวัดมิติที่ไม่สามารถทดแทนได้ โมเดลสำหรับการวัดการเรียนรู้และการเปลี่ยนแปลง โมเดลในการระบุโครงสร้างของระดับคุณลักษณะ และโมเดลสำหรับการจำแนกกลุ่มบุคคล ในการวิเคราะห์ข้อมูลทางสถิติ สามารถแบ่งโมเดลดังกล่าวได้เป็น 2 ลักษณะคือ Exploratory MIRT Model และ Confirmatory MIRT Model ในการวิจัยนี้โมเดลที่ใช้คือ โมเดลปกติสะสมแบบพหุมิติ เป็นโมเดลที่มีการประมาณค่าพารามิเตอร์ได้ใกล้เคียงกับโมเดลโลจิสติกแบบพหุมิติ (ค่าพารามิเตอร์มีความแตกต่างกันไม่เกิน .01) และมีโปรแกรมคอมพิวเตอร์สำหรับใช้ในการประมาณค่าที่เก่าแก่และน่าเชื่อถือหลายโปรแกรม ส่วนโมเดลโลจิสติกแบบพหุมิติ โปรแกรมในการประมาณค่าค่อนข้างจะหายาก และวิธีที่เหมาะสมสำหรับโมเดลปกติสะสมแบบพหุมิติ คือ วิธีการวิเคราะห์ข้อมูลแบบ Confirmatory MIRT Model มีความเหมาะสมในการยืนยันความสัมพันธ์ระหว่างข้อสอบ (ตัวแปรสังเกตได้) กับมิติด้านความรู้ แบ่งเป็น 4 ส่วน คือ ความรู้เกี่ยวกับข้อเท็จจริง (Factual Knowledge) ความรู้เกี่ยวกับมโนทัศน์ (Conceptual Knowledge) ความรู้เกี่ยวกับวิธีดำเนินการ (Procedural Knowledge) และความรู้เกี่ยวกับอภิปัญญา (Metacognitive Knowledge) และมิติด้านกระบวนการพุทธิปัญญา แบ่งเป็น 6 กระบวนการ คือ จำ (Remember) เข้าใจ (Understand) ประยุกต์ (Apply) วิเคราะห์ (Analyze) ประเมินค่า (Evaluate) และสร้างสรรค์ (Create)

ขั้นที่ 2 การรวบรวมข้อสอบ O-NET วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาชั้นปีที่ 6 ระหว่างปีการศึกษา 2551-2553 และ2555–2559 วิเคราะห์เนื้อหาการออกข้อสอบตามกลุ่มสาระการเรียนรู้วิชาคณิตศาสตร์

ผู้วิจัยรวบรวมข้อสอบ O-NET วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 6 ระหว่างปีการศึกษา 2551–2553 และปีการศึกษา 2555-2559 ที่ได้จากสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) จำนวน 256 ข้อ รายละเอียดดังภาพที่ 3-3



ภาพที่ 3-3 ขั้นตอนการรวบรวมข้อสอบและวิเคราะห์เนื้อหาการออกข้อสอบ

จากภาพที่ 3-3 แสดงขั้นตอนการรวบรวมข้อสอบและวิเคราะห์เนื้อหาการออกข้อสอบ O-NET วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 6 ดังนี้

2.1. ผู้วิจัยทำหนังสือขอความอนุเคราะห์ขอข้อมูลเพื่อการวิจัยจากวิทยาลัยวิทยาการวิจัยและวิทยาการปัญญา ถึง ผู้อำนวยการสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) โดยมีความประสงค์ขอความอนุเคราะห์ในการขอข้อสอบ O-NET วิชาคณิตศาสตร์ ระหว่างปีการศึกษา 2551-2553 และปีการศึกษา 2555-2559 พร้อมผลการตอบและเฉลย ส่วนข้อสอบปีการศึกษา 2554 ไม่มีข้อมูลเนื่องจากเป็นปีที่น้ำท่วมใหญ่กรุงเทพมหานครและปริมณฑล

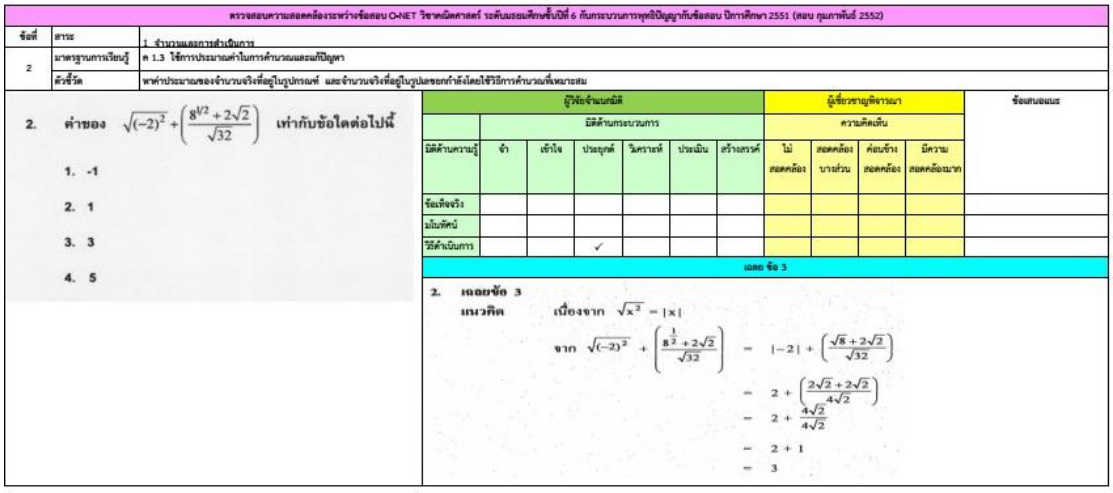
2.2. ผลการรวบรวมข้อสอบ O-NET วิชาคณิตศาสตร์ ระหว่างปีการศึกษา 2551-2553 และปีการศึกษา 2555-2559 ได้ทั้งหมด 256 ข้อ แสดงดังตารางที่ 3-1

ตารางที่ 3-1 จำนวนข้อสอบ O-NET วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 6 จำแนกตามสาระการเรียนรู้การออกข้อสอบ

ปีการศึกษา	สาระการเรียนรู้ที่ออกข้อสอบ				จำนวนข้อ	จำนวนตัวเลือก
	จำนวนและการดำเนินการ	การวัด	พีชคณิต	การวิเคราะห์ข้อมูลและความน่าจะเป็น		
2551	6	3	15	16	40	4
2552	7	5	10	14	36	4
2553	3	3	12	2	20	4
2555	7	4	15	16	32	5
2556	4	4	17	7	32	5
2557	6	6	13	7	32	5
2558	6	3	17	6	32	5
2559	7	3	15	7	32	5
รวม	46	31	114	65	256	

จากตารางที่ 3-1 ข้อสอบ O-NET วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 6 ระหว่างปีการศึกษา 2551–2553 และปีการศึกษา 2555-2559 จำนวน 256 ข้อ จำแนกตามสาระการเรียนรู้ที่ออกข้อสอบ ได้ดังนี้ 1) จำนวนและการดำเนินการ จำนวน 46 ข้อ 2) การวัด จำนวน 31 ข้อ 3) พีชคณิต จำนวน 114 ข้อ และ 4) การวิเคราะห์ข้อมูลและความน่าจะเป็น จำนวน 65 ข้อ ตามลำดับ

จากนั้นวิเคราะห์เนื้อหาการออกข้อสอบ พร้อมระบุสาระการเรียนรู้ มาตรฐานการเรียนรู้ และตัวชี้วัดของข้อสอบทั้ง 256 ข้อ ตัวอย่างการวิเคราะห์ข้อสอบ แสดงดังภาพที่ 3-4



ภาพที่ 3-4 การวิเคราะห์เนื้อหาการออกข้อสอบ พร้อมระบุสาระการเรียนรู้ มาตรฐานการเรียนรู้ และตัวชี้วัดของข้อสอบ

2.3. การตรวจสอบความตรงเชิงเนื้อหาโดยผู้เชี่ยวชาญทางด้านคณิตศาสตร์ จำนวน 3 คน ประกอบด้วย

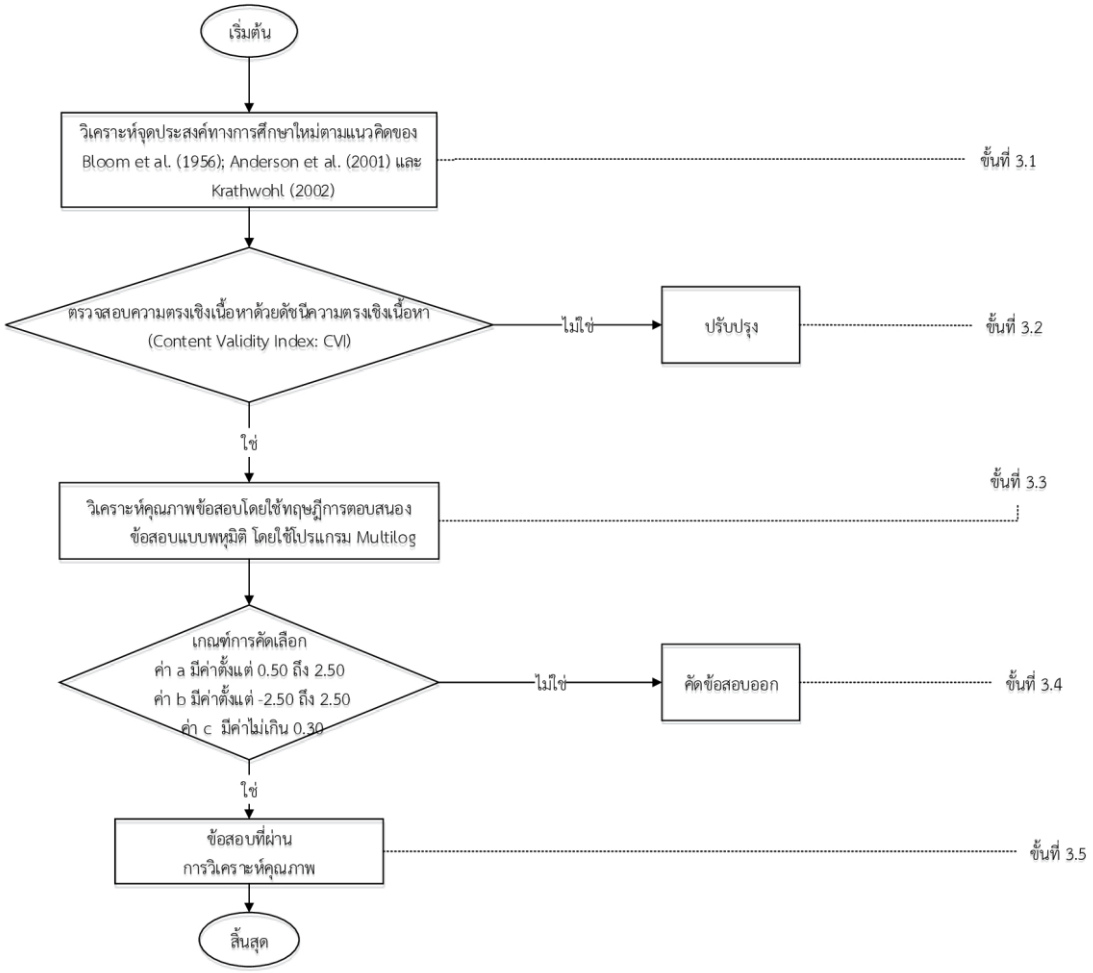
1) นายปศาสตร์ ต้นทอง ครู คศ.2 โรงเรียนทวีธาภิเศก บางขุนเทียน กรุงเทพมหานคร

2) นางสาวจริยา ทศพร ครู คศ.3 โรงเรียนทวีธาภิเศก บางขุนเทียน กรุงเทพมหานคร

3) นายธเนตร โชควิเศษชัยสิทธิ์ รองผู้อำนวยการ โรงเรียนทวีธาภิเศก บางขุนเทียน กรุงเทพมหานคร

ผู้เชี่ยวชาญได้พิจารณาความตรงเชิงเนื้อหาด้วยดัชนีความตรงเชิงเนื้อหา (Content Validity Index: CVI) ว่าแต่ละข้อวัดได้ตรงตามที่ต้องการหรือไม่ ผู้วิจัยคัดเลือกข้อสอบที่ผ่านเกณฑ์และปรับแก้ตามผู้เชี่ยวชาญเสนอแนะ ค่าดัชนีความตรงเชิงเนื้อหา มีตั้งแต่ 0.66 ถึง 1.00 และได้คัดเลือกไว้ จำนวน 256 ข้อ

ขั้นที่ 3 การวิเคราะห์ข้อสอบตามแนวของ Bloom et al. (1956); Anderson et al. (2001, p. XXI) และ Krathwohl (2002, p. 213) วิเคราะห์คุณภาพข้อสอบโดยใช้ทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ โดยใช้โปรแกรม Multilog ดังภาพที่ 3-5



ภาพที่ 3-5 ขั้นตอนการวิเคราะห์จุดประสงค์ทางการศึกษา และการวิเคราะห์คุณภาพข้อสอบ

จากภาพที่ 3-5 แสดงขั้นตอนการวิเคราะห์จุดประสงค์ทางการศึกษา และการวิเคราะห์คุณภาพข้อสอบ ดังนี้

3.1. ผู้วิจัยนำข้อสอบ O-NET วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 6 ระหว่างปีการศึกษา 2551-2553 และปีการศึกษา 2555-2559 จำนวน 256 ข้อ มาวิเคราะห์ตามจุดประสงค์ทางการศึกษาใหม่ตามแนวคิดของ Bloom et al. (1956); Anderson et al. (2001, p. XXI) และ Krathwohl (2002, p. 213) โดยการปรับลำดับขั้นและคำศัพท์ที่ใช้ในกระบวนการพุทธิปัญญา ซึ่งมี 6 กระบวนการเหมือนเดิม แต่ 3 กระบวนการแรกเปลี่ยนชื่อเป็น จำ (Remember) เข้าใจ (Understand) และประยุกต์ (Apply) ส่วน 3 กระบวนการหลังเปลี่ยนชื่อที่มีลักษณะเป็นคำนามไปเป็นคำกริยา และสลับที่กระบวนการที่ 5 กับ 6 และสร้างสรรค์ (Create) เปลี่ยนชื่อมาจาก การสังเคราะห์ (Synthesis) (Anderson et al., 2001, p. XXI; Krathwohl, 2002, p. 213) และเปลี่ยนโครงสร้างจากมิติเดียวเป็นสองมิติ โดยได้เพิ่มโครงสร้างในมิติด้านความรู้ (Knowledge Dimension) เข้ามาในโครงสร้างของจุดประสงค์ทางการศึกษา ด้านพุทธิปัญญา ทำให้โครงสร้างใหม่มีลักษณะเป็นสองมิติประกอบด้วยมิติด้านกระบวนการพุทธิปัญญาและมิติด้านความรู้ รวมทั้งอธิบายความหมายของมิติด้านความรู้ โดยมิติด้านความรู้ แบ่งเป็น 4 ส่วน คือ ความรู้เกี่ยวกับข้อเท็จจริง (Factual Knowledge) ความรู้เกี่ยวกับมโนทัศน์ (Conceptual Knowledge) ความรู้เกี่ยวกับวิธีดำเนินการ (Procedural Knowledge) และความรู้เกี่ยวกับอภิปัญญา (Metacognitive Knowledge) และมิติด้านกระบวนการพุทธิปัญญา แบ่งเป็น 6 กระบวนการ คือ จำ (Remember) เข้าใจ (Understand) ประยุกต์ (Apply) วิเคราะห์ (Analyze) ประเมินค่า (Evaluate) และสร้างสรรค์ (Create)

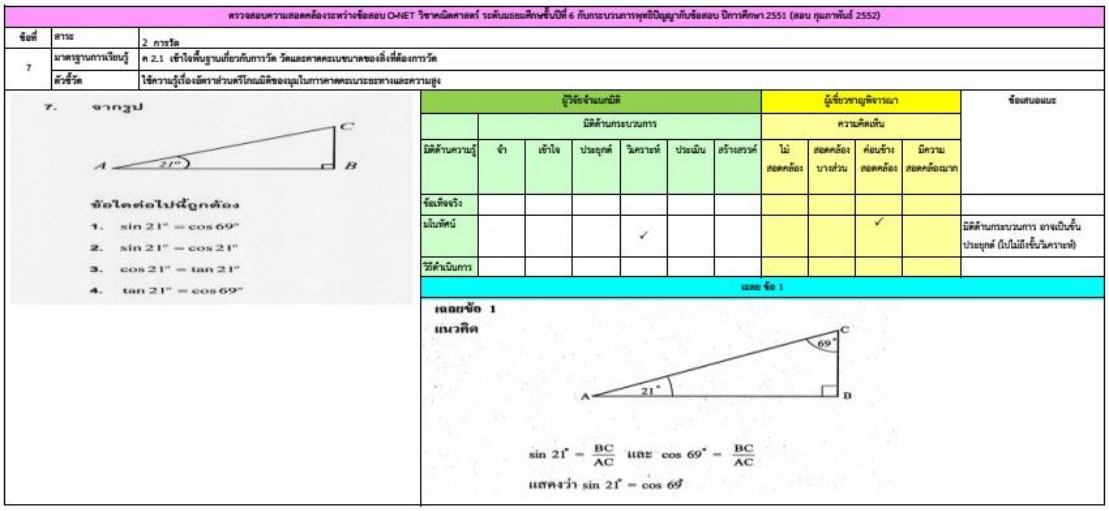
3.2. การตรวจสอบความตรงเชิงเนื้อหาโดยผู้เชี่ยวชาญทางด้านวัดผลการศึกษา จำนวน 3 คน ประกอบด้วย

1) อาจารย์ ดร.กนก พานทอง อาจารย์ประจำสาขาวิชาการวัดและเทคโนโลยีทางวิทยาการปัญญา วิทยาลัยวิทยาการวิจัยและวิทยาการปัญญา มหาวิทยาลัยบูรพา

2) อาจารย์ ดร.หนึ่งฤทัย เมฆวฑัต อาจารย์ประจำสาขาวิชาวิจัยทางการศึกษา คณะครุศาสตร์ มหาวิทยาลัยราชภัฏราชนครินทร์

3) อาจารย์ปานวาสน์ มหาลวเลิศ อาจารย์ประจำสำนักทดสอบทางการศึกษาและจิตวิทยา มหาวิทยาลัยศรีนครินทรวิโรฒ

ผู้เชี่ยวชาญพิจารณาความตรงเชิงเนื้อหาด้วยดัชนีความตรงเชิงเนื้อหา (Content Validity Index: CVI) แต่ละข้อวัดได้ตรงตามที่ต้องการหรือไม่ ผลการพิจารณาพบว่า ข้อสอบทั้ง 256 ข้อ ผ่านเกณฑ์ ค่าดัชนีความสอดคล้องมีตั้งแต่ 0.66 ถึง 1.00 แสดงดังภาพที่ 3-6 และตารางที่ 3-2



ภาพที่ 3-6 ผลการพิจารณาความตรงเชิงเนื้อหาของผู้เชี่ยวชาญ

ตารางที่ 3-2 จำนวนข้อสอบ O-NET วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 6 จำแนกตามจุดประสงค์ทางการศึกษาด้านพุทธิปัญญาและสาระการเรียนรู้ที่ออกข้อสอบ

จุดประสงค์ทางการศึกษาด้านพุทธิปัญญาที่มุ่งวัด	สาระการเรียนรู้ที่ออกข้อสอบ				รวม
	จำนวนและการดำเนินการ	การวัด	พีชคณิต	การวิเคราะห์ข้อมูลและความน่าจะเป็น	
1. วิเคราะห์เกี่ยวกับข้อเท็จจริง	7	0	4	0	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	0	0	5	0	5
รวม	45	31	115	65	256

จากตารางที่ 3-2 ข้อสอบ O-NET วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 6 ระหว่างปีการศึกษา 2551–2553 และปีการศึกษา 2555-2559 จำนวน 256 ข้อ จำแนกตามจุดประสงค์ทางการศึกษาด้านพุทธิปัญญาที่มุ่งวัดและสาระการเรียนรู้ที่ออกข้อสอบ ได้ดังนี้ 1) วิเคราะห์เกี่ยวกับข้อเท็จจริง จำนวน 11 ข้อ จำแนกเป็นสาระการเรียนรู้จำนวนและการดำเนินการ 7 ข้อ และพีชคณิต 4 ข้อ 2) เข้าใจเกี่ยวกับมโนทัศน์ จำนวน 17 ข้อ จำแนกเป็น สาระการเรียนรู้จำนวนและการดำเนินการ 1 ข้อ การวัด 1 ข้อ พีชคณิต 2 ข้อ และการวิเคราะห์ข้อมูลและความน่าจะเป็น 13 ข้อ 3) ประยุกต์เกี่ยวกับมโนทัศน์ จำนวน 10 ข้อ จำแนกเป็นสาระการเรียนรู้จำนวนและการดำเนินการ 1 ข้อ การวัด 1 ข้อ พีชคณิต 4 ข้อ และการวิเคราะห์ข้อมูลและความน่าจะเป็น 4 ข้อ 4) วิเคราะห์เกี่ยวกับมโนทัศน์ จำนวน 65 ข้อ จำแนกเป็นสาระการเรียนรู้จำนวนและการดำเนินการ 14 ข้อ การวัด 1 ข้อ พีชคณิต 41 ข้อ และการวิเคราะห์ข้อมูลและความน่าจะเป็น 9 ข้อ 5) ประยุกต์เกี่ยวกับ

วิธีดำเนินการ จำนวน 148 ข้อ จำแนกเป็นสาระการเรียนรู้จำนวนและการดำเนินการ 22 ข้อ การวัด 28 ข้อ พีชคณิต 59 ข้อ และการวิเคราะห์ข้อมูลและความน่าจะเป็น 39 ข้อ และ 6) วิเคราะห์เกี่ยวกับวิธีดำเนินการ จำนวน 5 ข้อ จำแนกเป็นสาระการเรียนรู้พีชคณิต 5 ข้อ

3.3. วิเคราะห์คุณภาพข้อสอบโดยใช้ทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ รายละเอียดมีดังนี้

1) ตรวจสอบความสมบูรณ์ของข้อสอบ O-NET วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 6 ทั้งข้อคำถาม ตัวเลือก และเฉลย

2) จัดเตรียมข้อมูลจากผลการตอบข้อสอบ การวิเคราะห์คุณภาพของข้อสอบตามหลักการของทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ แบบ 3 พารามิเตอร์

3.4 วิเคราะห์คุณภาพของข้อสอบตามหลักการของทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ แบบ 3 พารามิเตอร์ โดยโปรแกรม Multilog โดยมีการคัดเลือกข้อสอบ (1) ค่าอำนาจจำแนกของข้อสอบ (a) มีค่าตั้งแต่ 0.50 ถึง 2.50 (2) ค่าความยากของข้อสอบ (b) มีค่าตั้งแต่ -2.50 ถึง 2.50 และ (3) ค่าโอกาสการเดาของข้อสอบ (c) มีค่าไม่เกิน 0.30

สำหรับเกณฑ์ระดับความยากเฉลี่ยของข้อสอบ มีดังนี้

ค่าความยากมากกว่า	2.0000	หมายถึง	ข้อสอบยากมาก
ค่าความยากตั้งแต่	1.0001 ถึง 2.0000	หมายถึง	ข้อสอบยาก
ค่าความยากตั้งแต่	0.5001 ถึง 1.0000	หมายถึง	ข้อสอบค่อนข้างยาก
ค่าความยากตั้งแต่	-0.4999 ถึง 0.5000	หมายถึง	ข้อสอบปานกลาง
ค่าความยากตั้งแต่	-0.9999 ถึง -0.5000	หมายถึง	ข้อสอบค่อนข้างง่าย
ค่าความยากตั้งแต่	-2.0000 ถึง -1.0000	หมายถึง	ข้อสอบง่าย
ค่าความยากน้อยกว่า	2.0000	หมายถึง	ข้อสอบง่ายมาก

3.5 ผลการวิเคราะห์ปรากฏว่า ข้อสอบผ่านเกณฑ์ทั้งสิ้น 193 จากข้อสอบทั้งหมด 256 ข้อ คิดเป็นร้อยละ 75.39 ดังตารางที่ 3-3

ตารางที่ 3-3 จำนวนข้อสอบที่ผ่านเกณฑ์จากการวิเคราะห์ข้อสอบตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ

จุดประสงค์ทางการศึกษาด้านพุทธิปัญญาที่มุ่งวัด	สาระการเรียนรู้ที่ออกข้อสอบ				รวม
	จำนวนและการดำเนินการ	การวัด	พีชคณิต	การวิเคราะห์ข้อมูลและความน่าจะเป็น	
1. วิเคราะห์เกี่ยวกับข้อเท็จจริง	5	0	3	0	**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	0	0	5	0	**5**
รวม	**35**	**15**	**88**	**55**	**193**

จากตารางที่ 3-3 ผลการวิเคราะห์คุณภาพข้อสอบตามหลักการของทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ แบบ 3 พารามิเตอร์ โดยใช้โปรแกรม Multilog มีจำนวนข้อสอบที่ผ่านเกณฑ์ 193 ข้อ จำแนกตามจุดประสงค์ทางการศึกษาด้านพุทธิปัญญาและสาระการเรียนรู้ที่ออกข้อสอบได้ดังนี้ 1) วิเคราะห์เกี่ยวกับข้อเท็จจริง จำนวน 8 ข้อ ประกอบด้วย สาระการเรียนรู้จำนวนและการดำเนินการ จำนวน 5 ข้อ และพีชคณิต จำนวน 3 ข้อ 2) เข้าใจเกี่ยวกับมโนทัศน์ จำนวน 12 ข้อ ประกอบด้วย สาระการเรียนรู้จำนวนและการดำเนินการ จำนวน 1 ข้อ การวัด จำนวน 1 ข้อ พีชคณิต จำนวน 1 ข้อ และการวิเคราะห์ข้อมูลและความน่าจะเป็น จำนวน 9 ข้อ 3) ประยุกต์เกี่ยวกับมโนทัศน์ จำนวน 9 ข้อ ประกอบด้วย สาระการเรียนรู้จำนวนและการดำเนินการ จำนวน 1 ข้อ พีชคณิต จำนวน 4 ข้อ และการวิเคราะห์ข้อมูลและความน่าจะเป็น จำนวน 4 ข้อ 4) วิเคราะห์เกี่ยวกับมโนทัศน์ จำนวน 49 ข้อ ประกอบด้วย สาระการเรียนรู้จำนวนและการดำเนินการ จำนวน 9 ข้อ การวัด จำนวน 1 ข้อ พีชคณิต จำนวน 32 ข้อ และการวิเคราะห์ข้อมูลและความน่าจะเป็น จำนวน 7 ข้อ 5) ประยุกต์เกี่ยวกับวิธีดำเนินการ จำนวน 110 ข้อ ประกอบด้วย สาระการเรียนรู้จำนวนและการดำเนินการ จำนวน 19 ข้อ การวัด จำนวน 13 ข้อ จำนวน 43 ข้อ และการวิเคราะห์ข้อมูลและความน่าจะเป็น จำนวน 35 ข้อ และ 6) วิเคราะห์เกี่ยวกับวิธีดำเนินการ จำนวน 5 ข้อ ประกอบด้วย สาระการเรียนรู้พีชคณิต จำนวน 5 ข้อ (รายละเอียดแสดงในภาคผนวก ข)

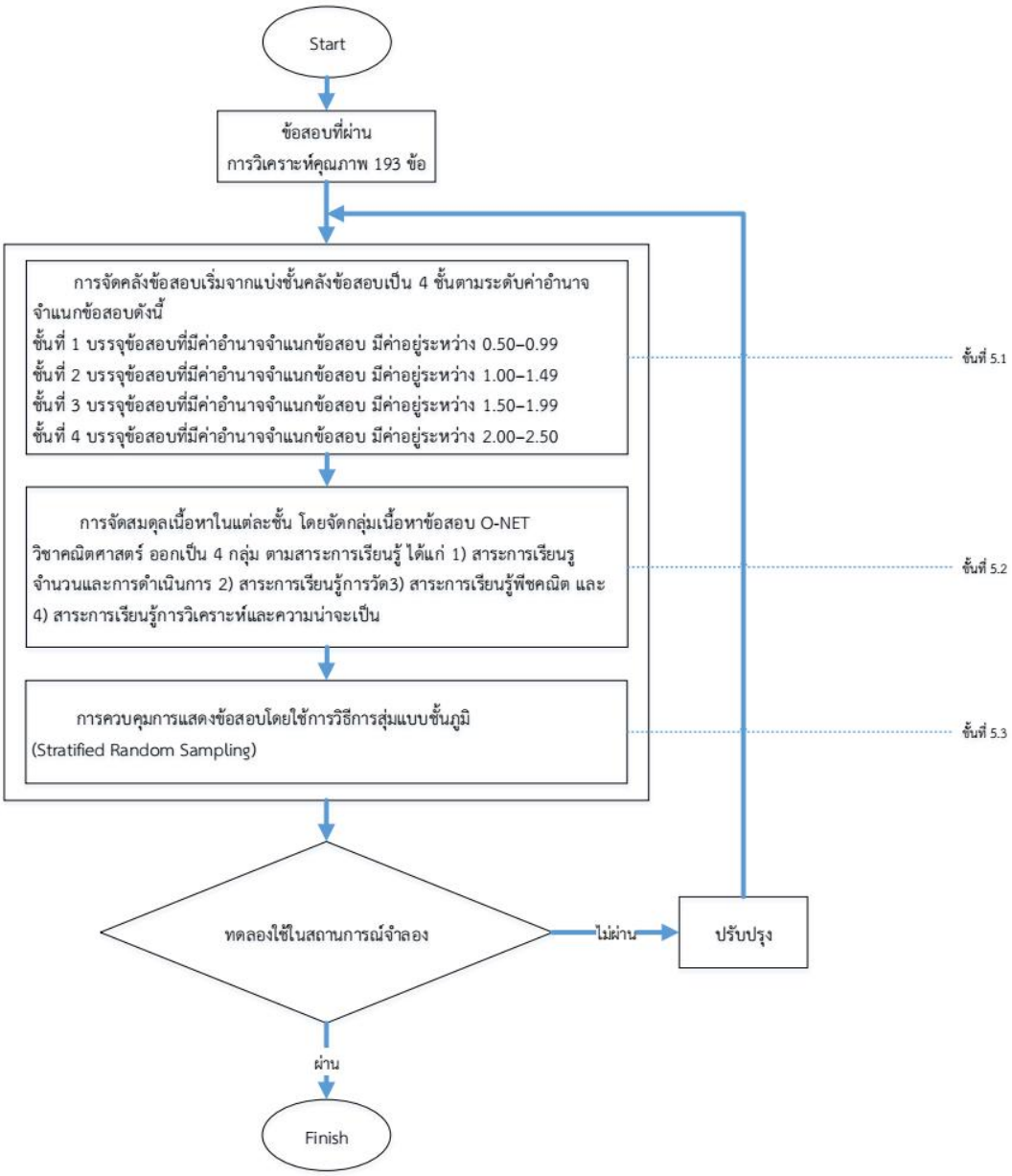
ขั้นที่ 4 การศึกษาหลักการของวิธีการจัดสมดุลเนื้อหาและการแสดงข้อสอบ

การจัดสมดุลเนื้อหาของข้อสอบ (Content Balancing) เป็นสิ่งสำคัญอย่างยิ่ง โดยเฉพาะการวัดผลสัมฤทธิ์ทางการศึกษา ในการทดสอบแบบดั้งเดิม แบบทดสอบมาตรฐานส่วนใหญ่ได้สร้างตามรายละเอียดในตารางการกำหนดคุณลักษณะของข้อสอบ (Table of Content Specification) ผู้สอบทุกคนจะได้รับแบบทดสอบชุดเดียวกัน ประกอบด้วยข้อสอบครบทุกคุณลักษณะที่ต้องการวัดตามที่ออกแบบไว้ ด้วยเหตุนี้ จึงเป็นข้อจำกัดในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ เนื่องจากการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ จะคัดเลือกข้อสอบที่ให้ค่าสารสนเทศสูงสุดที่ค่าประมาณความสามารถของผู้สอบ ณ ขณะนั้น ซึ่งไม่ได้พิจารณาขอบเขตเนื้อหา (Content Area) ของข้อสอบ เป็นผลให้ผู้สอบบางคนไม่ได้รับข้อสอบทุกขอบเขตเนื้อหาของข้อสอบ ตัวอย่างเช่น ในการวัดความสามารถด้านการคำนวณ การบวก การลบ การคูณ และการหาร ในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ผู้สอบอาจได้รับข้อสอบเฉพาะการบวก การลบ แต่ไม่ได้รับข้อสอบการคูณและการหาร ส่งผลให้ผู้ที่ไม่มีความรู้เรื่องการคูณและการหาร จะได้รับการประเมินความสามารถสูงเกินความจริง ในขณะเดียวกัน ผู้สอบที่มีความรู้เรื่องการคูณและการหาร จะถูกประเมินต่ำเกินจริง ดังนั้น เพื่อรับประกันว่าผู้สอบแต่ละคนจะได้รับข้อสอบที่เป็นตัวแทนตามสัดส่วนของแต่ละขอบเขตของเนื้อหาครบถ้วน

วิธีการควบคุมการแสดงข้อสอบ (Controlling Item Exposure) เป็นการศึกษาเอกสารและวรรณกรรมที่เกี่ยวข้องกับวิธีการควบคุมการแสดงข้อสอบ เพื่อเป็นฐานความรู้มาใช้กับวิธีการควบคุมการแสดงข้อสอบ ซึ่งจากการเปรียบเทียบข้อดีของวิธีการควบคุมการแสงข้อสอบทั้ง 5 ประเภท ได้แก่ วิธีการเลือกอย่างสุ่ม วิธีการเลือกอย่างมีเงื่อนไข การเลือกตามระดับชั้น วิธีการเลือกแบบรวมกลยุทธ์ และวิธีการทดสอบหลายขั้นตอน ปรากฏว่า แนวคิดและการนำไปใช้ของวิธีการแสดงข้อสอบในประเภทวิธีการคัดเลือกอย่างสุ่ม (Randomization Strategies) มีความยุ่งยาก

ซับซ้อนน้อยกว่าเมื่อเทียบกับวิธีการแสดงข้อสอบประเภทอื่นๆ เพราะใช้เพียงการสุ่มข้อสอบ ทำให้ข้อสอบข้อใดข้อหนึ่งไม่ถูกเลือกใช้ในการทดสอบบ่อยครั้งจนเกินไป (Georgiadou et al., 2007) โดยอาศัยแนวคิดของการสุ่มข้อสอบนี้ จึงนำวิธีการสุ่มตัวอย่างแบบชั้นภูมิ (Stratified Random Sampling) โดยให้กลุ่มสาระเป็นชั้นภูมิ และให้ค่าความยากของข้อสอบ (b) ที่มีลักษณะเหมือนกัน (Homogenious) เป็นหน่วยการสุ่ม มาใช้กับวิธีการแสดงข้อสอบ

ขั้นที่ 5 จัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ระดับชั้นร่วมกับการจัดสมดุลเนื้อหา ดังภาพที่ 3-7



ภาพที่ 3-7 ขั้นตอนการพัฒนาวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา

จากภาพที่ 3-7 วิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา มีขั้นตอนดังนี้

5.1 การพัฒนาคลังข้อสอบเริ่มจากแบ่งชั้นคลังข้อสอบเป็น 4 ชั้นตามระดับค่าอำนาจจำแนกของข้อสอบ โดยบรรจุข้อสอบที่มีค่าอำนาจจำแนกของข้อสอบ 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

5.2 การพัฒนาคลังข้อสอบเริ่มจากแบ่งชั้นคลังข้อสอบเป็น 4 ชั้นตามระดับค่าอำนาจจำแนกข้อสอบร่วมกับการจัดสมดุลเนื้อหาในแต่ละชั้น โดยในแต่ละชั้นได้จัดกลุ่มเนื้อหาข้อสอบ O-NET วิชาคณิตศาสตร์ โดยจำแนกเป็น 4 กลุ่มสาระการเรียนรู้ ได้แก่ 1) จำนวนและการดำเนินการ 2) การวัด 3) พีชคณิต และ 4) การวิเคราะห์ข้อมูลและความน่าจะเป็น ดังนั้นจำนวนข้อสอบแต่ละชั้นคลังข้อสอบเป็นไปตามวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นร่วมกับการจัดสมดุลเนื้อหาแบบใหม่ ดังภาพที่ 3-8

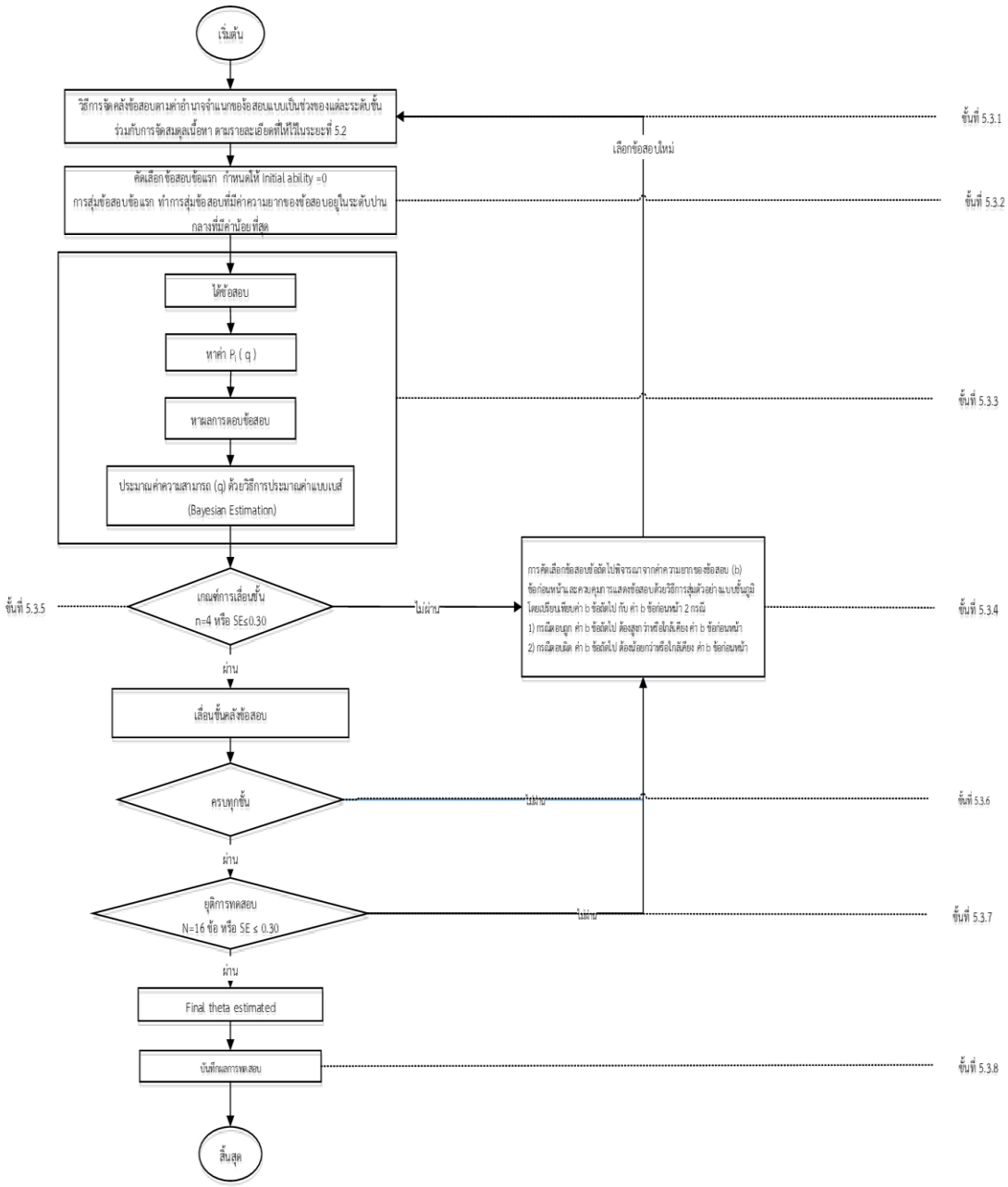
ชั้น	ช่วงของค่าอำนาจจำแนกข้อสอบ	รหัสข้อสอบ	A0	สาระการเรียนรู้คณิตศาสตร์											
				จำนวนและการดำเนินการ			การวัด			พีชคณิต			การวิเคราะห์และความน่าจะเป็น		
				ค่า a	ค่า b	ค่า c	ค่า a	ค่า b	ค่า c	ค่า a	ค่า b	ค่า c	ค่า a	ค่า b	ค่า c
1	0.50-0.99	Q_2551_1	A1	0.90	1.63	0.06	-	-	-	-	-	-	-	-	-
		Q_2559_10	A2	-	-	-	0.61	0.36	0.08	-	-	-	-	-	-
		Q_2557_1	A3	-	-	-	-	-	-	0.89	1.26	0.27	-	-	-
		Q_2556_28	A4	-	-	-	-	-	-	-	-	-	0.63	0.99	0.08
2	1.00-1.49	Q_2556_3	A1	1.47	1.44	0.15	-	-	-	-	-	-	-	-	-
		Q_2558_15	A2	-	-	-	1.49	1.90	0.22	-	-	-	-	-	-
		Q_2555_2	A3	-	-	-	-	-	-	1.46	1.80	0.11	-	-	-
		Q_2556_32	A4	-	-	-	-	-	-	-	-	-	1.28	1.38	0.18
3	1.50-1.99	Q_2556_5	A1	1.55	1.60	0.18	-	-	-	-	-	-	-	-	-
		Q_2558_14	A2	-	-	-	1.90	1.30	0.16	-	-	-	-	-	-
		Q_2558_9	A3	-	-	-	-	-	-	1.60	1.23	0.15	-	-	-
		Q_2559_30	A4	-	-	-	-	-	-	-	-	-	1.71	0.72	0.16
4	2.00-2.50	Q_2555_7	A1	2.28	1.22	0.20	-	-	-	-	-	-	-	-	-
		Q_2551_8	A2	-	-	-	2.12	0.80	0.28	-	-	-	-	-	-
		Q_2558_8	A3	-	-	-	-	-	-	2.00	1.65	0.22	-	-	-
		Q_2551_27	A4	-	-	-	-	-	-	-	-	-	2.26	0.48	0.27

หมายเหตุ ; A0 แทน จุดประสงค์ด้านพุทธิปัญญาตามแนวคิดของ Bloom et al. (1956) ซึ่งปรับใหม่โดย Anderson et al. (2001, pp.27-31)
A1 แทน ข้อสอบวัดการประยุกต์เกี่ยวกับมโนทัศน์
A2 แทน ข้อสอบวัดการประยุกต์เกี่ยวกับวิธีดำเนินการ
A3 แทน ข้อสอบวัดการวิเคราะห์เกี่ยวกับมโนทัศน์
A4 แทน ข้อสอบวัดการเข้าใจเกี่ยวกับมโนทัศน์

ภาพที่ 3-8 ตัวอย่างการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา

5.3 การควบคุมการแสดงข้อสอบโดยใช้วิธีการสุ่มตัวอย่างแบบชั้นภูมิ (Stratified Random Sampling) จากวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ระดับชั้น แบ่งออกเป็น 4 ชั้น แต่ละชั้นประกอบด้วย 4 สาระการเรียนรู้คณิตศาสตร์ ได้แก่ 1) จำนวนและการดำเนินการ 2) การวัด 3) พีชคณิต และ 4) การวิเคราะห์ข้อมูลและความน่าจะเป็น ดังนั้นวิธีการควบคุมการจัดสมดุลเนื้อหาด้วยวิธีการสุ่มตัวอย่างแบบชั้นภูมิ โดยให้กลุ่มสาระเป็นชั้นภูมิ และให้ค่าความยากของข้อสอบ (b) ที่มีลักษณะเหมือนกัน (Homogenious) เป็น

หน่วยการสุ่ม ขั้นตอนการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ระดับชั้นร่วมกับการจัดสมดุลเนื้อหาและการควบคุมการแสดงข้อสอบโดยใช้วิธีการสุ่มตัวอย่างแบบชั้นภูมิ ดังภาพที่ 3-9

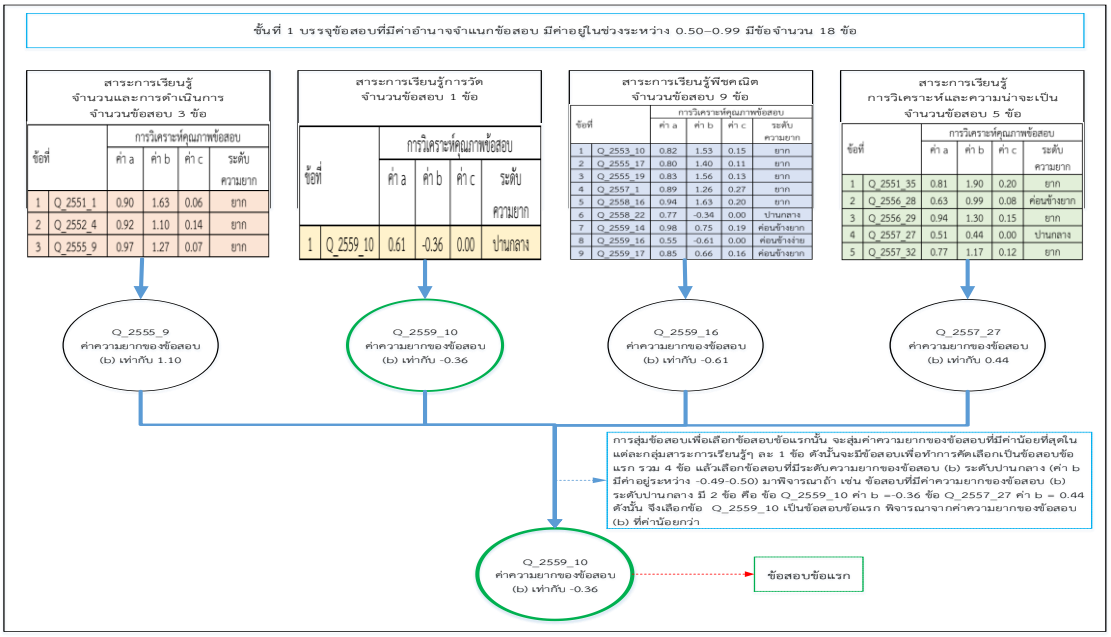


ภาพที่ 3-9 ขั้นตอนการทดสอบในแต่ละชั้นของคลังข้อสอบวิธี IACB-MCAT

จากภาพที่ 3-9 ขั้นตอนการทดสอบในแต่ละชั้นของคลังข้อสอบวิธี IACB มีดังนี้

5.3.1 จัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ระดับชั้นร่วมกับการจัดสมดุลเนื้อหา ตามรายละเอียดที่ให้ไว้ในระยะที่ 5.2

5.3.2 การคัดเลือกข้อสอบข้อแรก กำหนดให้ Initity Ability = 0 ตัวอย่างแสดงดังภาพที่ 3-10



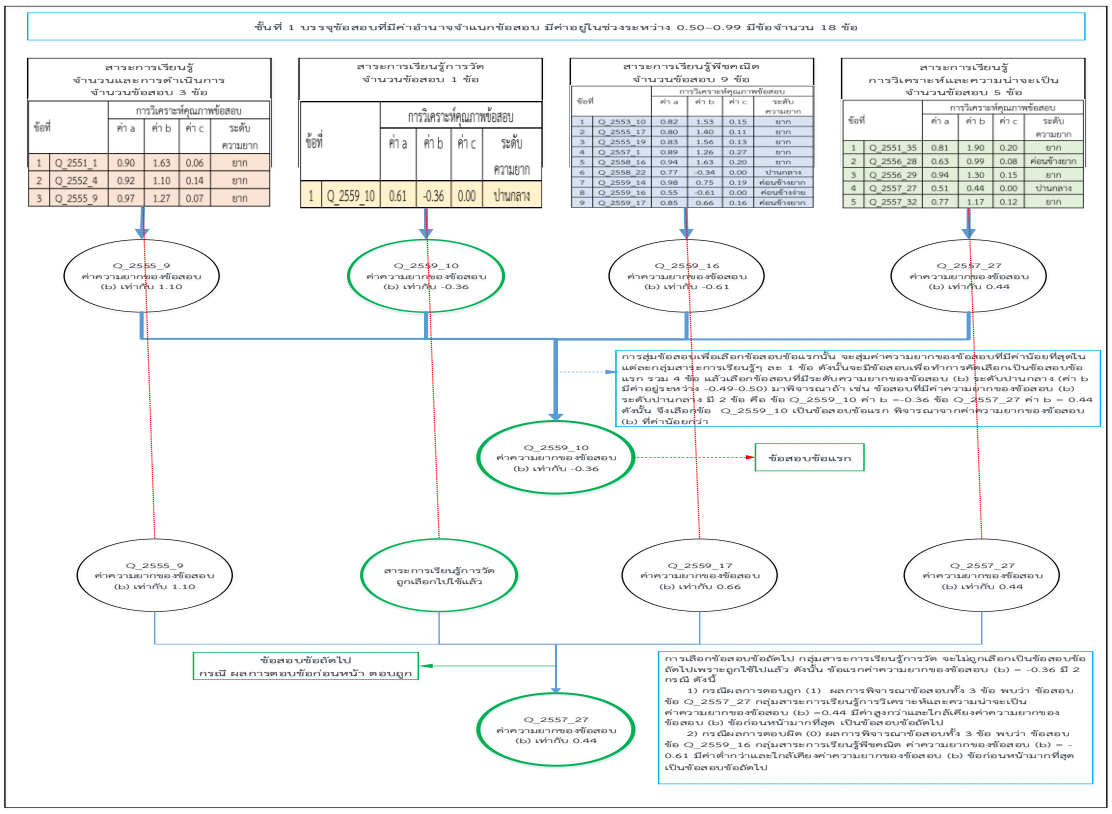
ภาพที่ 3-10 การเลือกข้อสอบข้อแรก

จากภาพที่ 3-10 อธิบายวิธีการคัดเลือกข้อสอบข้อแรก กำหนดให้ Initity Ability = 0 โดยสุ่มข้อสอบที่มีความสามารถใกล้เคียงกับความสามารถของผู้สอบ ซึ่งในคลังข้อสอบชั้นที่ 1 บรรจุข้อสอบที่มีค่าอำนาจจำแนกข้อสอบ มีค่าอยู่ในช่วงระหว่าง 0.50–0.99 การสุ่มข้อสอบเพื่อเลือกข้อสอบข้อแรกนั้น จะสุ่มค่าความยากของข้อสอบที่มีค่าน้อยที่สุดในแต่ละกลุ่มสาระการเรียนรู้ๆ ละ 1 ข้อ ดังนั้นจะมีข้อสอบเพื่อคัดเลือกเป็นข้อสอบข้อแรก รวม 4 ข้อ แล้วเลือกข้อสอบที่มีระดับความยากของข้อสอบ (b) ระดับปานกลาง (ค่า b มีค่าอยู่ระหว่าง -0.49-0.50) มาพิจารณาถ้า เช่น ข้อสอบที่มีค่าความยากของข้อสอบ (b) ระดับปานกลาง มี 2 ข้อ คือ ข้อ Q_2559_10 ค่า b =-0.36 ข้อ Q_2557_27 ค่า b = 0.44 ดังนั้น จึงเลือกข้อ Q_2559_10 เป็นข้อสอบข้อแรก พิจารณาจากค่าความยากของข้อสอบ (b) ที่ค่าน้อยกว่า

5.3.3 เมื่อได้ข้อสอบข้อแรก หาผลการตอบข้อสอบ ซึ่งข้อสอบจะให้ผลการตอบข้อสอบเป็น 0 และ 1 ถ้าผู้สอบตอบถูกผลการตอบจะเป็น 1 ถ้าผู้สอบตอบผิดผลการตอบจะเป็น 0 แล้วคำนวณหาค่า $P_i(\theta)$ จากสมการ (14) และประมาณค่า θ ด้วยวิธีการประมาณค่าแบบเบส์ (Bayesian Estimation)

5.3.4 การคัดเลือกข้อสอบข้อถัดไปพิจารณาจากค่าความยากของข้อสอบ (b) ข้อก่อนหน้าโดยใช้วิธีควบคุมการแสดงข้อสอบด้วยวิธีการสุ่มตัวอย่างแบบชั้นภูมิที่คำนึงถึงการจัดสมดุลเนื้อหาข้อสอบ โดยจะให้กลุ่มสาระการเรียนรู้คณิตศาสตร์เป็นชั้นภูมิ และให้ค่าความยาก

ของข้อสอบ (b) ที่มีลักษณะเหมือนกัน (Homogenious) เป็นหน่วยการสุ่ม โดยการบังคับกลุ่มสาระการเรียนรู้ที่ถูกเลือกไปให้ผู้สอบแล้วไม่ให้ถูกเลือกขึ้นมาซ้ำ ตัวอย่างแสดงดังภาพที่ 3-11



ภาพที่ 3-11 การเลือกข้อสอบข้อถัดไป

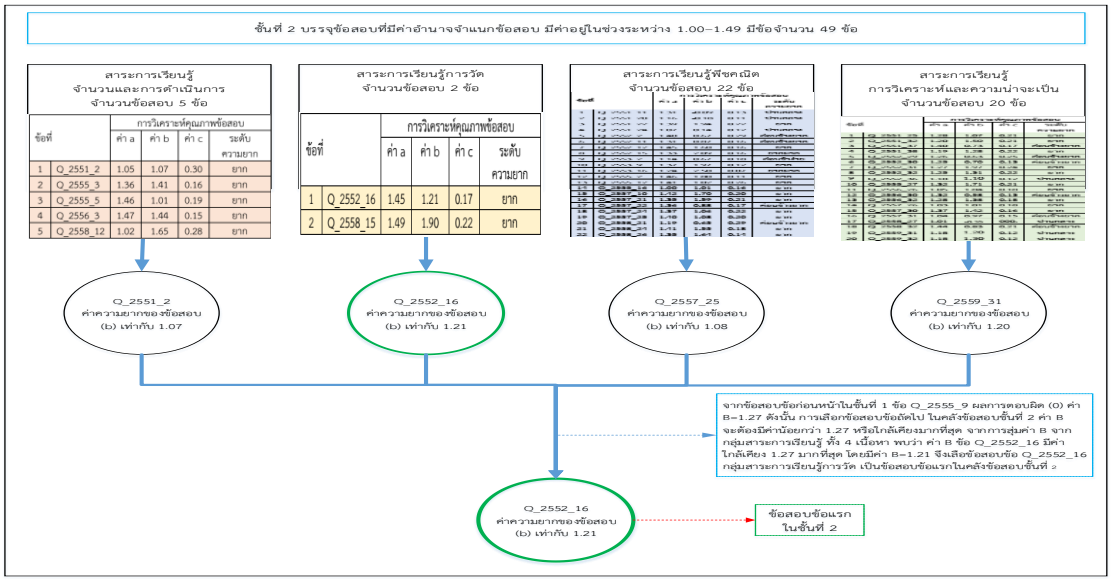
จากภาพที่ 3-11 อธิบายวิธีการคัดเลือกข้อสอบข้อถัดไป กลุ่มสาระการเรียนรู้การวัด จะไม่ถูกเลือกเป็นข้อสอบข้อถัดไปเพราะถูกใช้ไปแล้ว ถ้าข้อสอบข้อก่อนหน้า Q_2559_10 ความยากของข้อสอบ (b) = -0.36 ผลการตอบถูก ค่าความยากของข้อสอบ (b) ข้อถัดไป ต้องมีค่าสูงกว่าหรือใกล้เคียง -0.36 ตัวอย่างเช่น ข้อสอบทั้ง 3 ข้อ ได้แก่ 1) สาระการเรียนรู้จำนวนและการดำเนินการ Q_2559_9 ค่าความยากของข้อสอบ (b) =1.10 2) สาระการเรียนรู้พีชคณิต Q_2559_16 ค่าความยากของข้อสอบ (b) = -0.61 และ 3) สาระการเรียนรู้การวิเคราะห์ข้อมูลและความน่าจะเป็น ค่าความยากของข้อสอบ (b) = 0.44 แต่การเลือกข้อสอบข้อถัดไปจะพิจารณาค่าความยากของข้อสอบ (b) ที่มีค่าสูงกว่าหรือต่ำกว่าและใกล้เคียงค่าความยากของข้อสอบ (b) ข้อก่อนหน้ามากที่สุด มี 2 กรณี ดังนี้

1. กรณีผลการตอบถูก (1) ผลการพิจารณาข้อสอบทั้ง 3 ข้อ ปรากฏว่า ข้อสอบข้อ Q_2557_27 กลุ่มสาระการเรียนรู้การวิเคราะห์ข้อมูลและความน่าจะเป็น ค่าความยากของข้อสอบ (b) =0.44 มีค่าสูงกว่าและใกล้เคียงค่าความยากของข้อสอบ (b) ข้อก่อนหน้ามากที่สุด เป็นข้อสอบข้อถัดไป

2. กรณีผลการตอบผิด (0) ผลการพิจารณาข้อสอบทั้ง 3 ข้อ ปรากฏว่า ข้อสอบข้อ Q_2559_16 กลุ่มสาระการเรียนรู้พีชคณิต ค่าความยากของข้อสอบ (b) = -0.61 มีค่าต่ำกว่าและใกล้เคียงค่าความยากของข้อสอบ (b) ข้อก่อนหน้ามากที่สุด เป็นข้อสอบข้อถัดไป

5.3.5 เกณฑ์การยุติการทดสอบในแต่ละชั้นและเลื่อนชั้นการทดสอบ ผู้สอบจะต้องทดสอบในชั้นนั้นๆ ให้ครบ 4 ข้อ ตามเนื้อหาสาระการเรียนรู้ที่ออกข้อสอบ O-NET วิชาคณิตศาสตร์ หรือค่าความคลาดเคลื่อนมาตรฐาน น้อยกว่าหรือเท่ากับ 0.30 (n=4, หรือ SE ≤ 0.30) และรวบรวมผลการทดสอบในแต่ละชั้น

การเลื่อนชั้นการทดสอบไปทดสอบในคลังข้อสอบในชั้นที่ 2, 3 และ 4 จะต้องคำนึงถึงค่าความยากของข้อสอบ (b) ของข้อสอบข้อก่อนหน้าในชั้นที่แล้วเป็นเกณฑ์ ตัวอย่างแสดงดังภาพที่ 3-12



ภาพที่ 3-12 การเลือกข้อสอบข้อถัดไป เพื่อใช้เป็นข้อสอบข้อแรกในคลังข้อสอบชั้นที่ 2

จากภาพที่ 3-12 อธิบายการเลือกข้อสอบข้อถัดไป เพื่อใช้เป็นข้อสอบข้อแรกในคลังข้อสอบชั้นที่ 2 การพิจารณาในการคัดเลือกข้อสอบข้อถัดไปในคลังข้อสอบชั้นต่อมา ยกตัวอย่าง เช่น ข้อสอบ Q_2555_9 เป็นข้อสอบข้อก่อนหน้าในชั้นที่ 1 ค่า b = 1.27 ผลการตอบผิด (0) ดังนั้นข้อสอบข้อถัดไปโดยใช้คลังข้อสอบในชั้นที่ 2 ซึ่งจำนวนข้อสอบ 49 ข้อ สุ่มข้อสอบที่มีค่าความยากของข้อสอบ (b) ต้องมีค่าต่ำกว่าและใกล้เคียง 1.27 ตัวอย่างเช่น ข้อสอบทั้ง 4 ข้อ ได้แก่ 1) สาระการเรียนรู้จำนวนและการดำเนินการ Q_2551_2 ค่าความยากของข้อสอบ (b) =1.07 2) สาระการเรียนรู้การวัด Q_2552_16 ค่าความยากของข้อสอบ (b) = 1.21 3) สาระการเรียนรู้พีชคณิต Q_2559_31 ค่าความยากของข้อสอบ (b) = 1.08 และ 4) สาระการเรียนรู้การวิเคราะห์ข้อมูลและความน่าจะเป็น ค่าความยากของข้อสอบ (b) = 0.1.20 แต่การเลือกข้อสอบข้อถัดไปจะ

พิจารณาค่าความยากของข้อสอบ (b) ที่มีค่าต่ำกว่าหรือสูงและใกล้เคียงค่าความยากของข้อสอบ (b) ข้อก่อนหน้ามากที่สุด มี 2 กรณี ดังนี้

1. กรณีผลการตอบถูก (1) ผลการพิจารณาข้อสอบทั้ง 4 ข้อ ปรากฏว่า ไม่มีข้อสอบที่มีค่าความยากของข้อสอบสูงกว่าหรือใกล้เคียง ค่าความความยากของข้อก่อนหน้า จึงไปพิจารณาค่าความยากของข้อสอบทั้ง 4 กลุ่มสาระการเรียนรู้ใหม่ ปรากฏว่า ข้อสอบข้อ Q_2552_31 กลุ่มสาระการเรียนรู้การวิเคราะห์ข้อมูลและความน่าจะเป็น ค่าความยากของข้อสอบ (b) = 1.31 มีค่าสูงกว่าและใกล้เคียงค่าความยากของข้อสอบ (b) ข้อก่อนหน้ามากที่สุด เป็นข้อสอบข้อถัดไป

2. กรณีผลการตอบผิด (0) ผลการพิจารณาข้อสอบทั้ง 4 ข้อ ปรากฏว่า ข้อสอบข้อ Q_2552_16 กลุ่มสาระการเรียนรู้การวัด ค่าความยากของข้อสอบ (b) = 1.21 มีค่าต่ำกว่าและใกล้เคียงค่าความยากของข้อสอบ (b) ข้อก่อนหน้ามากที่สุด เป็นข้อสอบข้อถัดไป

5.3.6 ในกรณีที่ไม่สามารถหาข้อสอบจัดให้ได้และไม่มีข้อสอบอื่น จะทำซ้ำในขั้นที่ 3 จนกระทั่งหาข้อสอบข้อใดข้อหนึ่งจัดให้

5.3.7 ขั้นที่ 4 ถึงขั้นที่ 5 จะทำซ้ำ ๆ จนกระทั่งความยาวแบบทดสอบย่อยครบตามกำหนดที่ชั้นการทดสอบนั้น การทดสอบจะต่อเนื่องไปยังขั้นต่อไป

5.3.8 ขั้นที่ 4 ถึงขั้นที่ 5 จะทำซ้ำ ๆ จนครบทุกชั้นและครบทุกชั้นจึงยุติการทดสอบ

ขั้นที่ 6 การปรับปรุงและแก้ไขวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ระดับชั้นร่วมกับการจัดสมดุลเนื้อหา เป็นการทดสอบความถูกต้อง หรือนำไปพัฒนาวิธีการให้มีประสิทธิภาพในสถานการณ์จำลอง ซึ่งมีรายละเอียดดังนี้

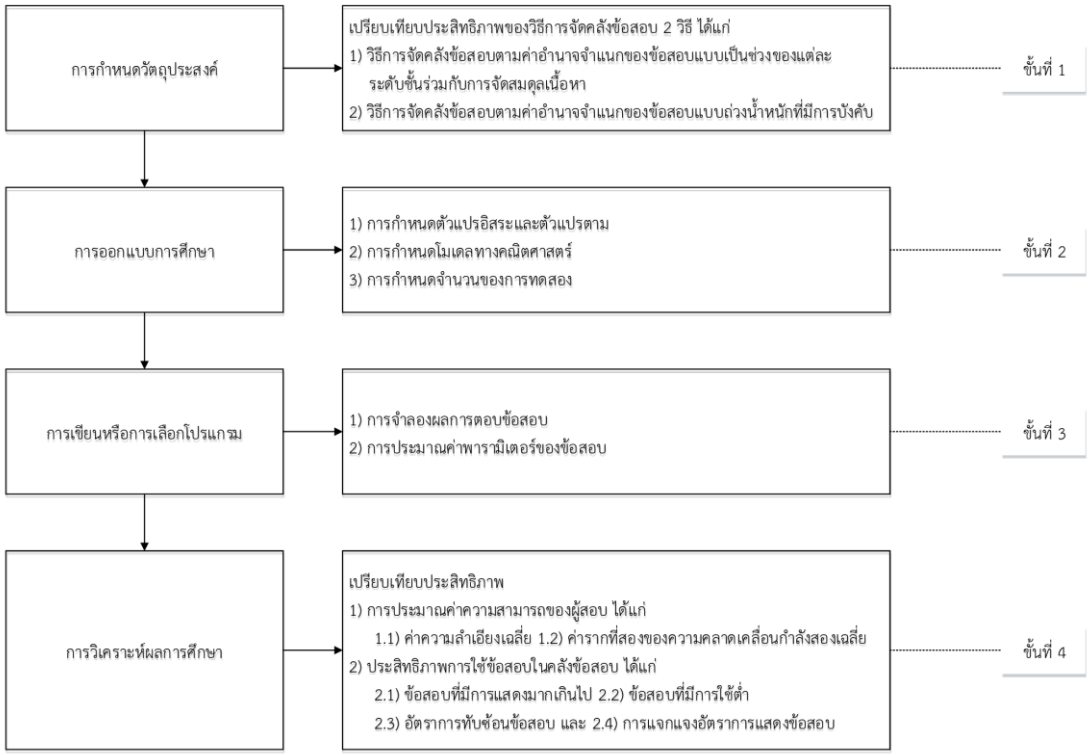
1. ข้อผิดพลาดที่พบจากการเขียนโปรแกรมมี 2 ลักษณะ ได้แก่ 1) ข้อผิดพลาดจากการเขียนคำสั่งของโปรแกรมผิดรูปแบบไวยกรณ์ของภาษา สามารถตรวจสอบได้ง่ายขณะประมวลผล โดยโปรแกรมจะรายงานข้อผิดพลาดที่เกิดขึ้น ส่วนใหญ่เกิดขึ้นจากการพิมพ์ผิด แก้ไขโดยการพิมพ์ให้ถูกต้อง และ 2) ข้อผิดพลาดจากการใช้คำสั่งผิดตรรกะหรือผิดเงื่อนไข ซึ่งทำให้ผลลัพธ์จาการการกระทำตามเงื่อนไขนั้นเกิดความผิดพลาด เช่น การตรวจสอบเงื่อนไข ใช้คำว่า “and” ซึ่งข้อผิดพลาดลักษณะนี้ตรวจสอบโดยพิจารณาค่าของตัวแปรที่ได้รับ หรือพิจารณาผลลัพธ์จากการประมวลผลโปรแกรมที่ไม่เป็นไปตามที่คาดไว้ หากพบจะตรวจสอบผลลัพธ์จากการประมวลผลที่ละคำสั่ง แล้วแก้ไขปรับปรุง

2. การปรับปรุงและแก้ไขวิธีการให้มีประสิทธิภาพมากขึ้น เป็นการนำแนวคิดจากหลายส่วน เช่น ทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ การจัดสมดุลเนื้อหา และการควบคุมการแสดงข้อสอบและวิธีการสุ่มตัวอย่างในระเบียบวิธีการวิจัย มาปรับปรุงวิธีควบคุมการแสดงข้อสอบใช้ในการคัดเลือกข้อสอบข้อถัดไป จนกระทั้งได้วิธีการที่มีประสิทธิภาพ ในช่วงเริ่มต้นการพัฒนาได้กำหนดให้ใช้ค่าความยากของข้อสอบ (b) แต่เมื่อทดลองใช้ในสถานการณ์จำลอง ปรากฏว่าผลการประมาณความสามารถของผู้สอบ ค่ามีความคลาดเคลื่อนจากค่าความสามารถที่แท้จริงของผู้สอบมาก จึงปรับปรุงโดยกำหนดค่าผลต่างระหว่างค่าความยากของข้อสอบกับความสามารถของผู้สอบ แล้วทดลองในสถานการ์จำลอง ปรากฏว่า การประมาณค่าความสามารถของผู้สอบมีความคลาดเคลื่อนลดลงจากค่าความสามารถที่แท้จริงของผู้สอบเป็นอย่างมาก

ระยะที่ 2 การเปรียบเทียบประสิทธิภาพของวิธีการจัดคลังข้อสอบโดยการจำลองข้อมูล

การเปรียบเทียบประสิทธิภาพของวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา กับวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบถ่วงน้ำหนักที่มีการบังคับ ในครั้งนี้ศึกษาในสถานการณ์จำลอง เพื่อให้แน่ใจว่าความแตกต่างของประสิทธิภาพที่เกิดขึ้น (ตัวแปรตาม) เป็นผลมาจากความแตกตามของวิธีการจัดคลังข้อสอบ (ตัวแปรต้น) ไม่ได้เกิดจากตัวแทรกซ้อนอื่นๆ เช่น อารมณ์ของผู้สอบ หรือสภาพแวดล้อมขณะทดสอบ ซึ่งเป็นการควบคุมตัวแปรแทรกซ้อนให้คงที่ตลอดการศึกษาในสถานการณ์จำลอง นอกจากนี้ การวัดประสิทธิภาพด้านการประมาณค่าความสามารถของผู้สอบ จำเป็นต้องทราบค่าความสามารถที่แท้จริงของผู้สอบ ซึ่งในสถานการณ์จริงไม่สามารถหาได้

การศึกษาในสถานการณ์จำลอง ดำเนินการตามขั้นตอนที่เสนอโดย Harwell et al. (1996) ซึ่งสามารถแสดงขั้นตอนการศึกษาได้ตามดังภาพที่ 3-13



ภาพที่ 3-13 ขั้นตอนการศึกษาในสถานการณ์จำลองเพื่อเปรียบเทียบประสิทธิภาพของวิธีการจัดคลังข้อสอบ

จากภาพที่ 3-13 ขั้นตอนการศึกษาในสถานการณ์จำลองเพื่อเปรียบเทียบประสิทธิภาพของวิธีการจัดคลังข้อสอบ มีรายละเอียดดังนี้

ขั้นตอนที่ 1 การกำหนดวัตถุประสงค์ของการศึกษาในสถานการณ์จำลองเป็นการศึกษาเปรียบเทียบประสิทธิภาพระหว่างวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็น

ช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา กับวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบถ่วงน้ำหนักที่มีการบังคับ โดยเปรียบเทียบประสิทธิภาพ ดังนี้

1. การประมาณค่าความสามารถของผู้สอบ ได้แก่ ค่าความลำเอียงเฉลี่ย และค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย

2. ประสิทธิภาพของการใช้ข้อสอบ ได้แก่ ข้อสอบที่มีการแสดงมากเกินไป ข้อสอบที่มีการใช้น้อยเกินไป อัตราการทับซ้อนของข้อสอบ และการแจกแจงอัตราการแสดงข้อสอบ

ขั้นตอนที่ 2 การออกแบบการศึกษา ดำเนินการดังนี้

1. กำหนดตัวแปรอิสระและตัวแปรตาม จากวัตถุประสงค์ของการศึกษาในสถานการณ์จำลอง

1.1 ตัวแปรอิรสะ เป็น วิธีการจัดคลังข้อสอบ 2 วิธี ดังนี้

1.1.1 วิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา

1.1.2 วิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบถ่วงน้ำหนักที่มีการบังคับ

1.2 ตัวแปรตาม เป็น ประสิทธิภาพของวิธีการจัดคลังข้อสอบ พิจารณาดังนี้

1.2.1 ด้านการประมาณค่าความสามารถของผู้สอบ พิจารณาดังนี้

1.2.1.1 ค่าความลำเอียงเฉลี่ย (Average Bias) (หน่วยวัดเป็นคะแนนมาตรฐาน)

1.2.1.2 ค่ารากที่สองของคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Square Error: RMSE) (หน่วยวัดเป็นคะแนนมาตรฐาน)

1.2.2 ด้านประสิทธิภาพของการใช้ข้อสอบ

1.2.2.1 ข้อสอบที่มีการแสดงมากเกินไป (Overexposed Item) (หน่วยวัดเป็นจำนวนข้อ)

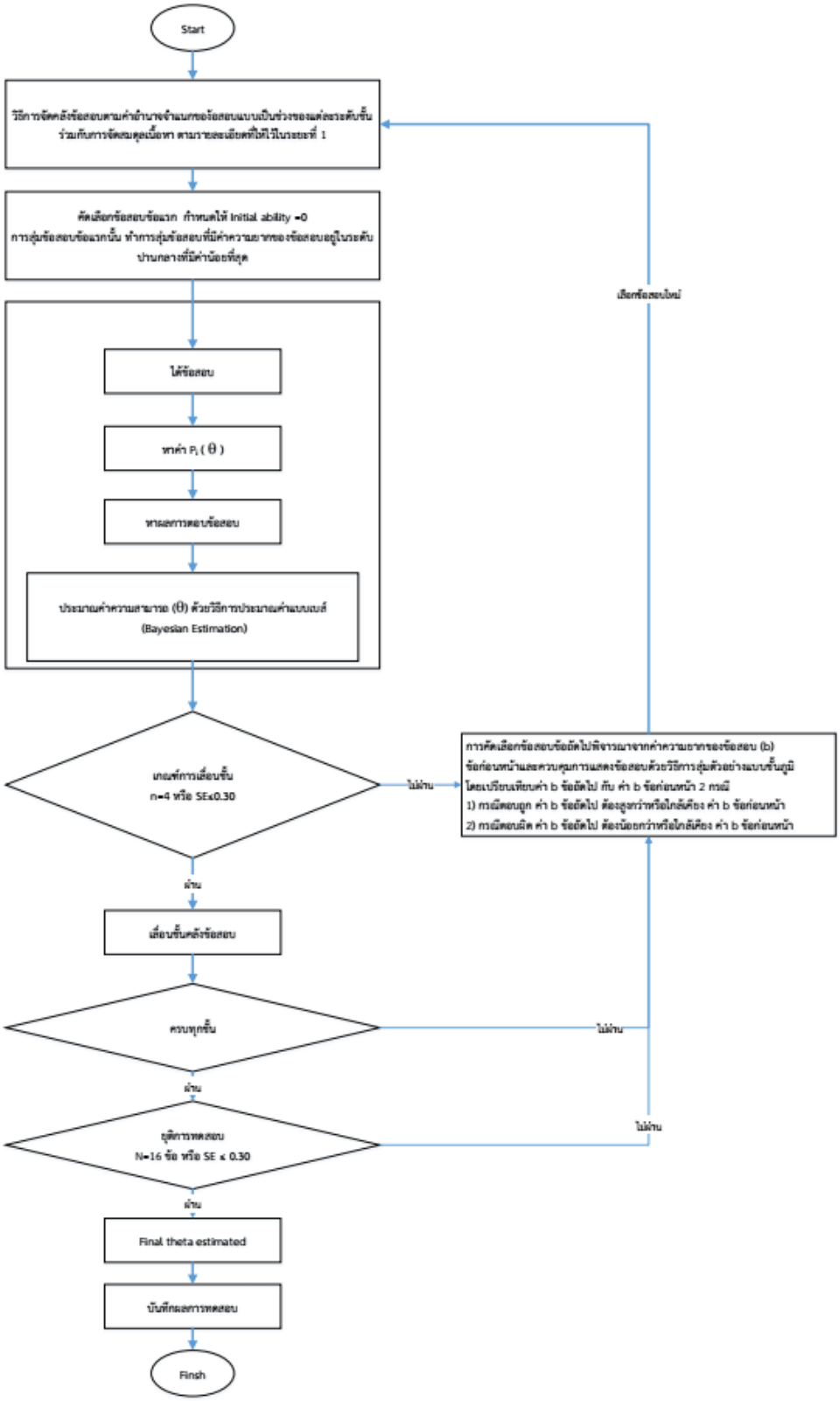
1.2.2.2 ข้อสอบที่มีการใช้น้อยเกินไป (Underutilized Item) (หน่วยวัดเป็นจำนวนข้อ)

1.2.2.3 อัตราการทับซ้อนของข้อสอบ (Item Overlap Rate) (หน่วยวัดเป็นจำนวนข้อ)

1.2.2.4 การแจกแจงอัตราการแสดงข้อสอบ (Item Exposure Rate Distribution (หน่วยวัดเป็นจำนวนข้อ)

2. การกำหนดโมเดลทางคณิตศาสตร์

ขั้นที่ 2.1 การจำลองข้อมูลตามวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา โดยการจำลองข้อมูลจริงเป็นข้อสอบ O-NET วิชาคณิตศาสตร์ ระหว่างปีการศึกษา 2551-2553 และปีการศึกษา 2555-2559 โดยใช้คลังข้อสอบที่จัดตาม ระยะที่ 1 แล้วพัฒนาโปรแกรมการจำลองการทดสอบ มีขั้นตอนแสดงดังภาพที่ 3-14



ภาพที่ 3-14 ขั้นตอนของวิธี IACB

1) วิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา ตามรายละเอียดที่ให้ไว้ในระยะที่ 1

2) การคัดเลือกข้อสอบข้อแรก กำหนดให้ Initial ability = 0 การสุ่มข้อสอบข้อแรกนั้นสุ่มข้อสอบที่มีค่าความยากของข้อสอบอยู่ในระดับปานกลางที่มีค่าน้อยที่สุด จากทั้ง 4 กลุ่มสาระการเรียนรู้ ในคลังข้อสอบชั้นที่ 1 จากนั้นเลือกข้อสอบที่มีค่าความยากของข้อสอบอยู่ในระดับค่าความยากปานกลาง (ค่า b มีค่าอยู่ระหว่าง -0.49-0.50) มาพิจารณาค่า b ที่น้อยที่สุด เพื่อใช้เป็นข้อสอบข้อแรกในการทดสอบ รายละเอียดแสดงในระยะที่ 1 ในขั้นที่ 5.3.1

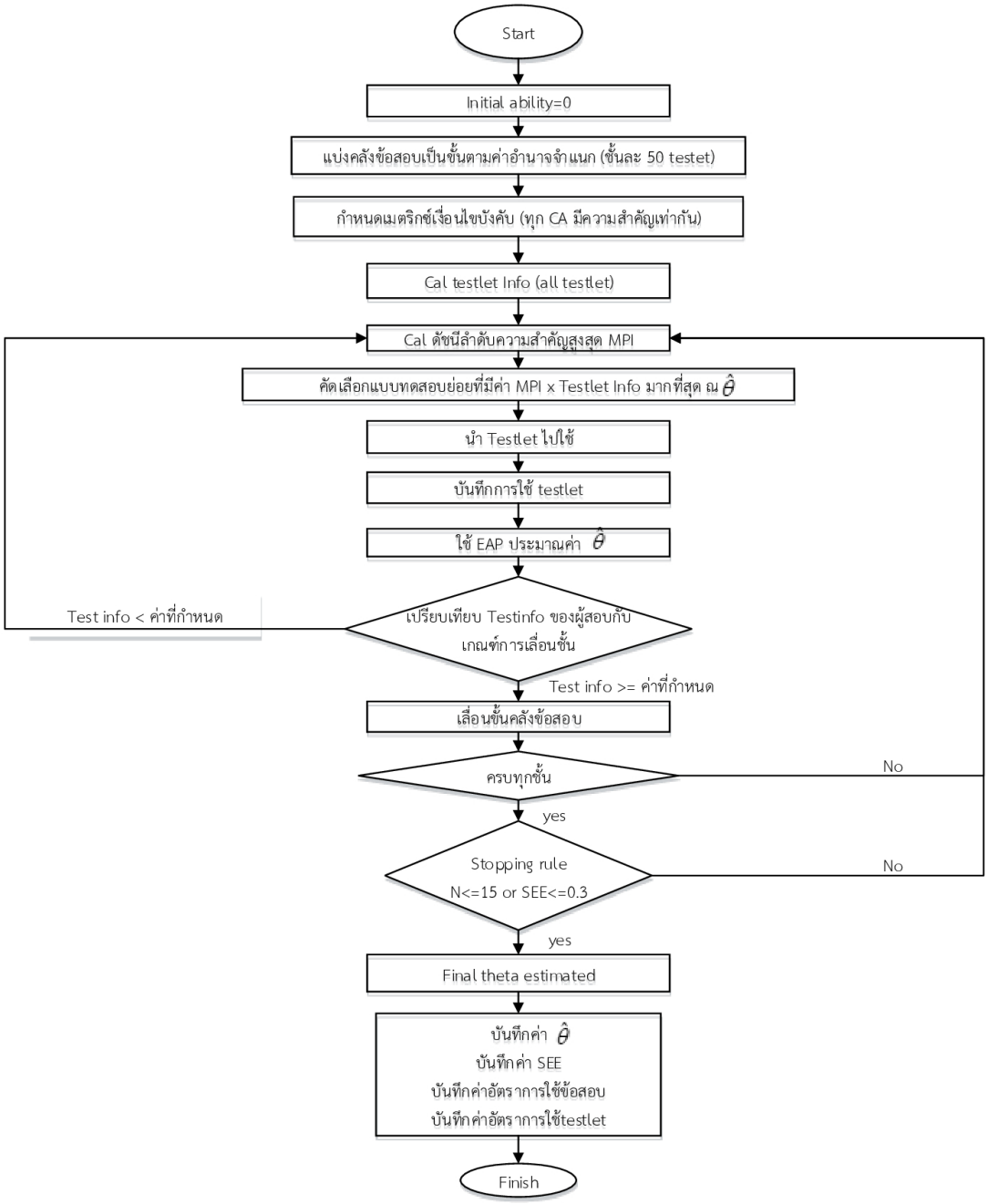
3) เมื่อได้ข้อสอบข้อแรก หาผลการตอบข้อสอบ ซึ่งข้อสอบจะให้ผลการตอบเป็น 0 และ 1 ถ้าผู้ตอบถูกผลการตอบถูกจะเป็น 1 ถ้าผลการตอบผิดจะเป็น 0 แล้วคำนวณหาค่า $P_i(\theta)$ จากสมการ (14) และประมาณค่าความสามารถ (θ) ด้วยวิธีการประมาณค่าแบบเบส์ (Bayesian Estimation) (Segall, 2010, pp. 60-64; Reckase, 2009, pp. 144-145) จากสมการ (37)

4) การคัดเลือกข้อสอบข้อถัดไปพิจารณาจากค่าความยากของข้อสอบ (b) ข้อก่อนหน้าและวิธีการควบคุมการแสดงข้อสอบด้วยวิธีการสุ่มตัวอย่างแบบชั้นภูมิที่คำนึงถึงการจัดสมดุลเนื้อหาข้อสอบ รายละเอียดแสดงในระยะที่ 1 ในขั้นที่ 5.3.3

5) การเปรียบเทียบค่าสารสนเทศของแบบทดสอบกับเกณฑ์ที่ใช้พิจารณาเลื่อนชั้นของคลังของข้อสอบ ถ้าค่าสารสนเทศของแบบสอบมีค่าน้อยกว่าเกณฑ์ที่ตั้งไว้ ข้อสอบจะถูกเลือกจากชั้นถัดไปของคลังข้อสอบ เกณฑ์ที่ใช้กำหนดให้เลื่อนชั้นของคลังข้อสอบใช้ค่าความคลาดเคลื่อนมาตรฐานเป็นไปตามเกณฑ์ (SE < 0.3) เมื่อผู้สอบทำข้อสอบครบ 4 ข้อ ในแต่ละชั้น ก็เลื่อนชั้นและยุติการทดสอบเมื่อทดสอบครบ 16 ข้อ

6) การนับจำนวนการจำลองข้อมูล ทำซ้ำตั้งแต่ข้อ 1-4 จนกว่าจำนวนข้อมูลที่ต้องการจำลอง ครบ 500 ชุด

ขั้นที่ 2.2 การจำลองข้อมูลตามวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบถ่วงน้ำหนักที่มีการบังคับ โดยการจำลองข้อมูลจริงเป็นข้อสอบ O-NET วิชาคณิตศาสตร์ ระหว่างปีการศึกษา 2551-2553 และปีการศึกษา 2555-2559 โดยแบ่งคลังข้อสอบออกเป็น 4 ชั้น ตามช่วงของค่าอำนาจจำแนกในระยะที่ 1 ไม่มีการจัดสมดุลเนื้อหาของข้อสอบมีขั้นตอนแสดงดังภาพที่ 3-15



ภาพที่ 3-15 ขั้นตอนของวิธี CWA

จากภาพที่ 3-15 มีรายละเอียดดังนี้

1) แบ่งคลังข้อสอบเป็นชั้นตามค่าอำนาจจำแนกของข้อสอบ โดยบรรจุข้อสอบที่มีค่าอำนาจจำแนกของข้อสอบ ดังนี้ ชั้นที่ 1 0.50–0.99 ชั้นที่ 2 1.00–1.49 ชั้นที่ 3 1.50–1.99 และชั้นที่ 4 2.00–2.50

ในการจำลองสถานการณ์ ผู้วิจัยได้ใช้ข้อมูลจริงในการจัดคลังข้อสอบโดยคลังข้อสอบชั้นที่ 1 มีจำนวนข้อสอบ 18 ข้อ มีค่าความยากของข้อสอบ อยู่ระหว่าง -0.61-1.90 คลังข้อสอบชั้นที่

2 มีจำนวนข้อสอบ 49 ข้อ มีค่าความยากของข้อสอบ อยู่ระหว่าง -0.35-2.50 คลังข้อสอบชั้นที่ 3 มีจำนวนข้อสอบ 64 ข้อ มีค่าความยากของข้อสอบ อยู่ระหว่าง 0.66-2.49 และคลังข้อสอบชั้นที่ 4 มีจำนวนข้อสอบ 62 ข้อ มีค่าความยากของข้อสอบ อยู่ระหว่าง 0.39-2.14 ภายในแต่ละชั้นของคลังข้อสอบจะมีค่าความยากเฉลี่ยกระจายตัวแบบยูนิฟอร์ม

2) คัดเลือกข้อสอบข้อแรก กำหนดให้ Initial ability = 0

3) กำหนดเมตริกซ์เงื่อนไขบังคับ ซึ่งจะบอกว่าข้อสอบแต่ละข้อประกอบด้วยเงื่อนไขบังคับใดบ้าง โดยกำหนดให้เมตริกซ์ดังกล่าวคือ เมตริกซ์ C.C เมื่อ $c_{jk}=1$ แสดงว่าข้อสอบข้อที่ j เกี่ยวข้องกับเงื่อนไขบังคับที่ k

4) คำนวณดัชนีลำดับความสำคัญ (Maximum Priority Index: MPI) โดยนำค่าลำดับความสำคัญซึ่งกำหนดโดยผู้วิจัยมาพิจารณาร่วมในการคัดเลือกข้อสอบ ซึ่งพิจารณาจากผลคูณของค่าสารสนเทศของข้อสอบกับดัชนีลำดับความสำคัญ ข้อใดให้ผลคูณมากที่สุด ข้อนั้นจะได้รับการคัดเลือกให้นำไปใช้เป็นข้อสอบข้อถัดไป

5) คัดเลือกข้อสอบที่มี MPI มากที่สุด ณ ตำแหน่งค่าความสามารถของผู้สอบเมื่อข้อสอบถูกเลือกและนำไปใช้กับผู้สอบ เมื่อผู้สอบตอบ ผลการตอบข้อสอบจะถูกนำไปคำนวณหาค่าประมาณความสามารถด้วยวิธีการประมาณค่าแบบ Expected A Posteriori (EAP) จากนั้นจึงนำค่าความสามารถของผู้สอบ ณ ตำแหน่งปัจจุบันไปใช้เลือกข้อสอบขัดถัดไป

6) เปรียบเทียบค่าสารสนเทศของแบบทดสอบกับเกณฑ์ที่ใช้พิจารณาเลื่อนชั้นของคลังข้อสอบ ถ้าค่าสารสนเทศของแบบทดสอบมีค่าน้อยกว่าเกณฑ์ที่ตั้งไว้ ข้อสอบจะถูกเลือกจากชั้นถัดไปของคลังข้อสอบ เกณฑ์ที่ใช้กำหนดให้เลื่อนชั้นของคลังข้อสอบใช้ค่าความคลาดเคลื่อนมาตรฐานเป็นไปตามเกณฑ์ (SE < 0.3) และจำนวนข้อสอบเกิน 15 ข้อ

7) นับจำนวนการจำลองข้อมูล ทำซ้ำตั้งแต่ข้อ 1-4 จนกว่าจำนวนข้อมูลที่ต้องการจำลองครบ 500 ชุด

ขั้นที่ 2.3 การกำหนดจำนวนของการทดลอง เทียบได้กับการกำหนดขนาดกลุ่มตัวอย่างของการศึกษาในสถานการณ์จริง แต่การศึกษาในสถานการณ์จำลองยังไม่มีกฎเกณฑ์แน่นอนในการกำหนดจำนวนของการทดลองซ้ำ

ขั้นตอนที่ 3 การเขียนหรือเลือกใช้โปรแกรม การศึกษาในสถานการณ์จำลองนี้เป็นการจำลองสถานการณ์การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์โดยใช้โปรแกรม JAVA ผู้วิจัยได้เขียนคำสั่งให้โปรแกรมดำเนินการตามขั้นตอนในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ เหมือนในสถานการณ์จริงทุกประการ (รายละเอียดแสดงในขั้นตอนที่ 2) แต่การศึกษาในสถานการณ์จำลองนี้ ทราบค่าความสามารถที่แท้จริงของผู้สอบ ผลการตอบข้อสอบและคลังข้อสอบจริง จึงต้องจำลองขึ้น ค่าความสามารถที่แท้จริงของผู้สอบสามารถจำลองขึ้นโดยใช้เลขสุ่ม (รายละเอียด แสดงใน 2.3) ส่วนผลการตอบข้อสอบและคลังข้อสอบ สามารถจำลองขึ้นได้ โดยมีรายละเอียดใน การจำลองผลการตอบข้อสอบและการประมาณค่าพารามิเตอร์ของข้อสอบ ดังนี้

1. การจำลองผลการตอบข้อสอบ เนื่องจากในสถานการณ์จำลองไม่มีผู้สอบที่แท้จริงจึงไม่มีผลการตอบข้อสอบแต่ละข้อ ดังนั้นในการเขียนโปรแกรมจำลองสถานการณ์การทดสอบ จึงต้องจำลองผลการตอบข้อสอบแต่ละข้อของผู้สอบ ซึ่งดำเนินการโดยให้โปรแกรมสุ่มเลขที่มีการแจกแจง

แบบยูนิฟอร์ม U(0,1) มา 1 ค่า แล้วนำเลขสุ่มนี้มาเปรียบเทียบกับค่าความน่าจะเป็นที่ผู้สอบที่มีความสามารถ θ ตอบข้อสอบข้อที่ i ได้ถูกต้อง $(P_i(\theta))$ตามโมเดลโลจิสติกสามพารามิเตอร์ ซึ่งคำนวณได้จากสมการที่ 14 แล้วผลการตอบข้อสอบจะกำหนดได้จากเงื่อนไขดังนี้ (Thompson & Weiss, 2011)

กรณีที่ ค่า $P_i(\theta) \geq$ เลขสุ่ม ให้กำหนดผลการตอบข้อสอบเป็น 1 (ตอบถูก)

กรณีที่ ค่า $P_i(\theta) <$ เลขสุ่ม ให้กำหนดผลการตอบข้อสอบเป็น 0 (ตอบผิด)

2. การประมาณค่าพารามิเตอร์ของข้อสอบ เนื่องจากการศึกษาในสถานการณ์จำลอง ผู้วิจัยใช้คลังข้อสอบที่แท้จริง จากการดำเนินงานในระยะที่ 1 ดังนั้น คลังข้อสอบทั้งหมดมีจำนวน 193 ข้อ และแบ่งคลังข้อสอบออกเป็น 4 ชั้น ตามค่าอำนาจจำแนกของข้อสอบ ดังนี้ ชั้นที่ 1 0.50–0.99 ชั้นที่ 2 1.00–1.49 ชั้นที่ 3 1.50–1.99 และชั้นที่ 4 2.00–2.50

ข้อสอบแต่ละข้อจะมีเพียงค่าพารามิเตอร์ของข้อสอบ 3 ค่า ได้แก่ ค่าความยากของข้อสอบ ค่าอำนาจจำแนกของข้อสอบและค่าโอกาสการเดาของข้อสอบ ซึ่งวิเคราะห์ข้อสอบโดยใช้โปรแกรม Multilog รายละเอียดมีดังนี้

1) ค่าความยากของข้อสอบ (b-parameter) สุ่มจากการแจกแจงแบบยูนิฟอร์มมีค่าอยู่ในช่วงระหว่าง -2.50 ถึง 2.50

2) ค่าอำนาจจำแนกของข้อสอบ (a-parameter) สุ่มจากการแจกแจงแบบยูนิฟอร์มมีค่าอยู่ในช่วงระหว่าง 0.50 ถึง 2.50

3) ค่าการเดาของข้อสอบ (c-parameter) สุ่มจากการแจกแจงแบบยูนิฟอร์มมีค่าไม่เกิน 0.30

การศึกษาในสถานการณ์จำลอง การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์กำหนดให้แต่ละวิธีการจัดคลังข้อสอบ ทดสอบซ้ำ 10 รอบ โดยทุกรอบใช้คลังข้อสอบจริง ในระยะที่ 1 จำนวน 193 ข้อ โดยมีวิธีการจัดคลังข้อสอบที่แตกต่างกัน 2 วิธี คือ วิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา ที่ผู้วิจัยพัฒนาขึ้นกับวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบถ่วงน้ำหนักที่มีการบังคับ ทดลองซ้ำ 10 รอบ โดยทุกรอบใช้คลังข้อสอบจริงชุดเดียวกันของแต่ละวิธี แล้วในแต่ละรอบได้เก็บข้อมูลได้แก่ ค่าการประมาณค่าความสามารถของผู้สอบ ($\hat{\theta}$) และจำนวนข้อสอบที่ใช้ในการทดสอบ เพื่อใช้เปรียบเทียบประสิทธิภาพ

ขั้นตอนที่ 4 การวิเคราะห์ผลการศึกษาดำเนินการโดยรวบรวมผลการศึกษาในสถานการณ์จำลองการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ที่ใช้วิธีการจัดคลังข้อสอบ 2 วิธี คือ วิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ระดับชั้นร่วมกับการจัดสมดุลเนื้อหาที่ผู้วิจัยพัฒนาขึ้น กับวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบถ่วงน้ำหนักที่มีการบังคับ โดยทุกวิธีได้ดำเนินการซ้ำ 10 รอบ แล้วนำข้อมูลทั้งหมดไปใช้เพื่อเปรียบเทียบประสิทธิภาพระหว่างวิธีการจัดคลังข้อสอบ วิธี IACB กับ วิธี CWA รายละเอียดมีดังนี้

1. ด้านการประมาณค่าความสามารถของผู้สอบ ได้แก่ ค่าความลำเอียงเฉลี่ย และค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย ตามแนวทางของ Yan et al. (2016, p. 115)

1.1 การคำนวณค่าความลำเอียงเฉลี่ย (Average Bias) โดยนำค่าความสามารถที่แท้จริงของผู้สอบ (θ) และค่าประมาณความสามารถของผู้สอบ $(\hat{\theta})$ ที่ได้รับจากการศึกษาในสถานการณ์จำลอง คำนวณได้จากสมการ

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

เมื่อ n แทน จำนวนผู้สอบทั้งหมด

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

θ_i แทน ค่าความสามารถที่แท้จริงของผู้สอบคนที่ i

จากนั้นพิจารณาค่าความลำเอียงเฉลี่ยของวิธี IACB กับวิธี CWA ถ้าค่าเข้าใกล้ 0 มากที่สุด แสดงว่ามีความแม่นยำในการประมาณค่าความสามารถของผู้สอบ

1.2 การคำนวณค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Square Error: RMSE) โดยนำค่าความสามารถที่แท้จริงของผู้สอบ $\left(\theta\right)$ และค่าประมาณความสามารถของผู้สอบ $(\hat{\theta})$ ที่ได้รับจากการศึกษาในสถานการณ์จำลอง คำนวณได้จากสมการ

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{\theta}_i - \theta_i\right)^2}$$

จากนั้นพิจารณาค่า RMSE ของวิธี IACB กับวิธี CWA วิธีการใด มีค่า RMSE น้อยที่สุด แสดงว่ามีประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบ

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

2.1 การคำนวณข้อสอบที่มีการแสดงมากเกินไป (Overexposure Item) คำนวณได้ดังสมการโดยนำจำนวนข้อสอบที่ถูกจัดให้ผู้สอบจากการศึกษาในสถานการณ์จำลอง คำนวณได้จากสมการ

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

จากนั้นพิจารณาค่าอัตราการแสดงข้อสอบสูงกว่า 0.2 ของวิธี IACB กับ วิธี CWA ถ้าอัตราการแสดงข้อสอบสูงกว่า 0.2 ของวิธีการใดมีจำนวนน้อยกว่า แสดงว่าวิธีการที่มีประสิทธิภาพด้านการใช้ข้อสอบที่มีการแสดงมากเกินไป

2.2 ข้อสอบที่มีการใช้น้อยเกินไป (Underutilized Item) คำนวนได้จากสมการในข้อ 2.1 จากนั้นพิจารณาข้อสอบที่มีการใช้น้อยกว่า 0.2 ของวิธี IACB กับ วิธี CWA ถ้าข้อสอบที่มีการใช้น้อย กว่า 0.2 ของวิธีการใดมีจำนวนน้อยกว่า แสดงว่าวิธีการนั้นมีประสิทธิภาพด้านข้อสอบที่มีการใช้น้อยเกินไป

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) คำนวณจากชุดข้อสอบที่มีผู้สอบได้รับชุดข้อสอบร่วมกัน (2 คนขึ้นไป) คำนวณได้จากสมการ

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

เมื่อ er_j คือ อัตราการแสดงข้อสอบสำหรับข้อสอบที่ j

L คือ ความยาวแบบทดสอบ

N คือ ขนาดคลังข้อสอบ

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

ดังนั้น สมการแสดงถึงความแตกต่างระหว่างค่าสังเกตและค่าคาดหวังของอัตราการแสดงข้อสอบ ถ้าค่า χ^2 เล็ก แสดงถึง การใช้ข้อสอบในคลังข้อสอบมีความสมดุลกัน

การเปรียบเทียบการแจกแจงอัตราการแสดงข้อสอบ โดยใช้อัตราส่วน *F* (*F* ratio) พิจารณาจากอัตราส่วนของค่า χ_1^2 ของวิธีที่ 1 หารด้วย χ_2^2 ของวิธีที่ 2 ถ้า $F < 1$ แสดงว่า วิธีที่ 1 ให้สมดุลของอัตราการแสดงข้อสอบโดยรวมดีกว่าวิธีที่ 2

โดยวิเคราะห์ด้วยสถิติ ดังนี้

1) วิเคราะห์ด้วยสถิติพื้นฐาน ได้แก่ ค่าเฉลี่ย (Mean) และส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation)

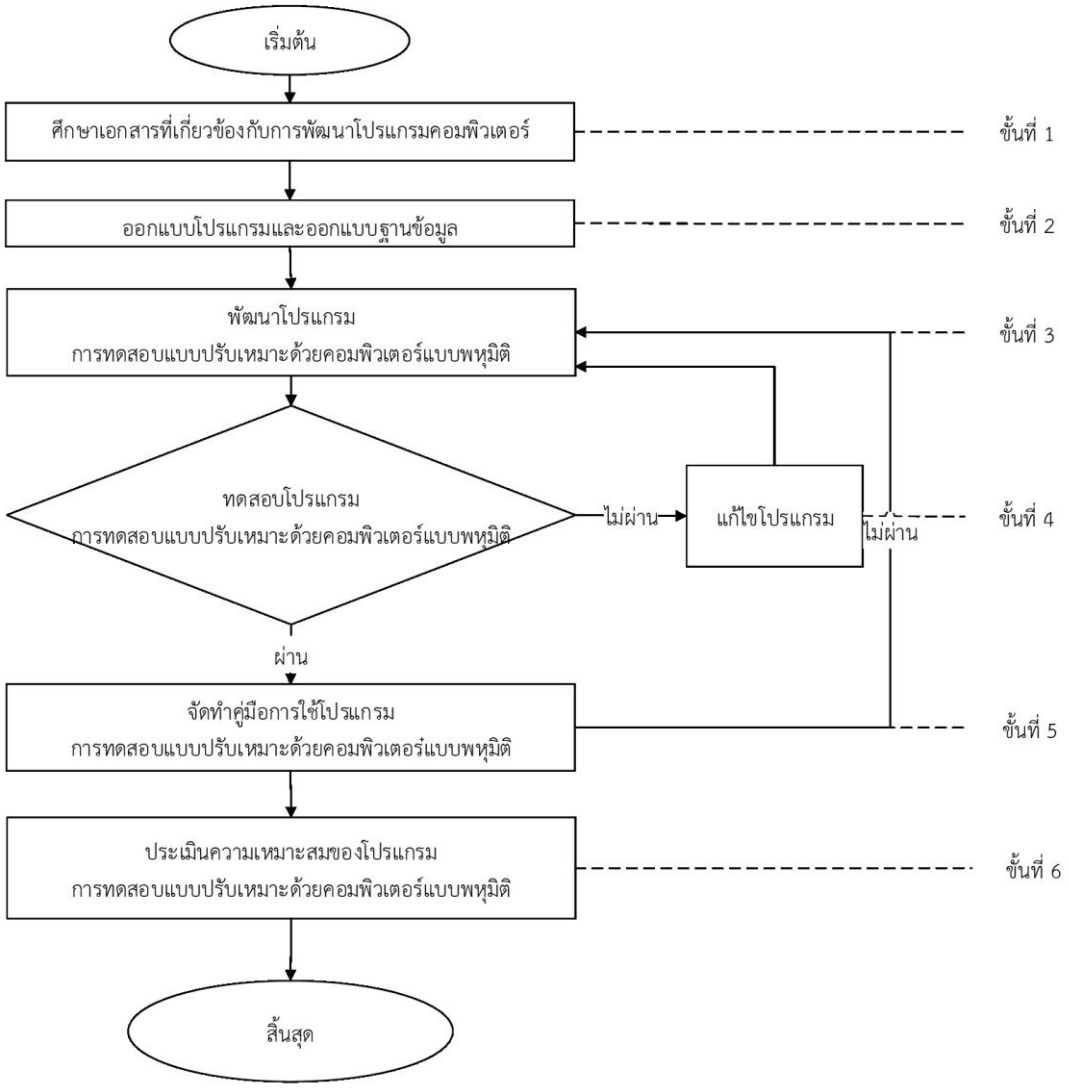
2) การประมาณค่าความสามารถของผู้สอบได้แก่ ค่าความลำเอียงเฉลี่ย และค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย ซึ่งเป็นข้อมูลแบบต่อเนื่อง (Continuous Data) ใช้สถิติทดสอบนอนพาราเมตริก วิเคราะห์ข้อมูลด้วยการวิเคราะห์ Wilcoxon Test รายละเอียดแสดงใน ภาคผนวก ค1

3) ประสิทธิภาพของการใช้ข้อสอบ ได้แก่ ข้อสอบที่มีการแสดงมากเกินไป ข้อสอบที่มีการใช้น้อยเกินไป อัตราการทับซ้อนของข้อสอบ และการแจกแจงอัตราการแสดงข้อสอบ ซึ่งเป็นข้อมูลไม่ต่อเนื่อง (Discrete Data) ใช้สถิติทดสอบพาราเมตริก วิเคราะห์ข้อมูลด้วยการวิเคราะห์ Chi-Square รายละเอียดแสดงใน ภาคผนวก ง1– ง4

ตารางที่ 3-4 สรุปเกณฑ์การเปรียบเทียบประสิทธิภาพด้านการประมาณค่าความสามารถของผู้สอบ และด้านประสิทธิภาพการใช้ข้อสอบ

ตัวแปร	เกณฑ์การพิจารณา	งานวิจัย
ด้านการประมาณค่าความสามารถของผู้สอบ		
1. ค่าความลำเอียงเฉลี่ย	ค่าเข้าใกล้ 0 มากที่สุด (แสดงว่ามีความแม่นยำในการประมาณค่าความสามารถของผู้สอบ)	Yan et al. (2016, p. 115)
2. ค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย	มีค่าน้อยที่สุด (แสดงว่ามีประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบ)	Yan et al. (2016, p. 115)
ด้านประสิทธิภาพของการใช้ข้อสอบ		
3. ข้อสอบที่มีการแสดงมากเกินไป	มีจำนวนน้อยกว่า (แสดงว่ามีประสิทธิภาพด้านข้อสอบที่มีการแสดงมากเกินไป)	Ozturk and Dogan (2015)
4. ข้อสอบที่มีการใช้น้อยเกินไป	มีจำนวนน้อยกว่า (แสดงว่ามีประสิทธิภาพด้านข้อสอบที่มีการใช้น้อยเกินไป)	Ozturk and Dogan (2015)
5. อัตราการทับซ้อนของข้อสอบ	มีจำนวนน้อยกว่า (แสดงว่ามีประสิทธิภาพด้านอัตราการทับซ้อนของข้อสอบ)	Ozturk and Dogan (2015)
6.การแจกแจงอัตราการแสดงข้อสอบ	อัตราส่วนของค่า χ_1^2 ของวิธีที่ 1 หารด้วย χ_2^2 ของวิธีที่ 2 ถ้า $F < 1$ (แสดงว่า วิธีที่ 1 มีประสิทธิภาพของการแจกแจงอัตราการแสดงข้อสอบโดยรวมดีกว่าวิธีที่ 2)	Ozturk and Dogan (2015)

ระยะที่ 3 การพัฒนาโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติที่ใช้วิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา



ภาพที่ 3-16 ขั้นตอนการพัฒนาโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ

จากภาพที่ 3-16 ขั้นตอนการพัฒนาโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติที่ใช้วิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา มีดังนี้ คือ

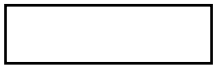
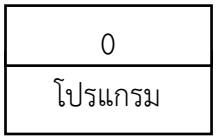
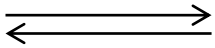
ขั้นที่ 1 การศึกษาเอกสารที่เกี่ยวข้องกับการพัฒนาโปรแกรมคอมพิวเตอร์

ขั้นที่ 2 การออกแบบโปรแกรมและการออกแบบฐานข้อมูล ประกอบด้วย

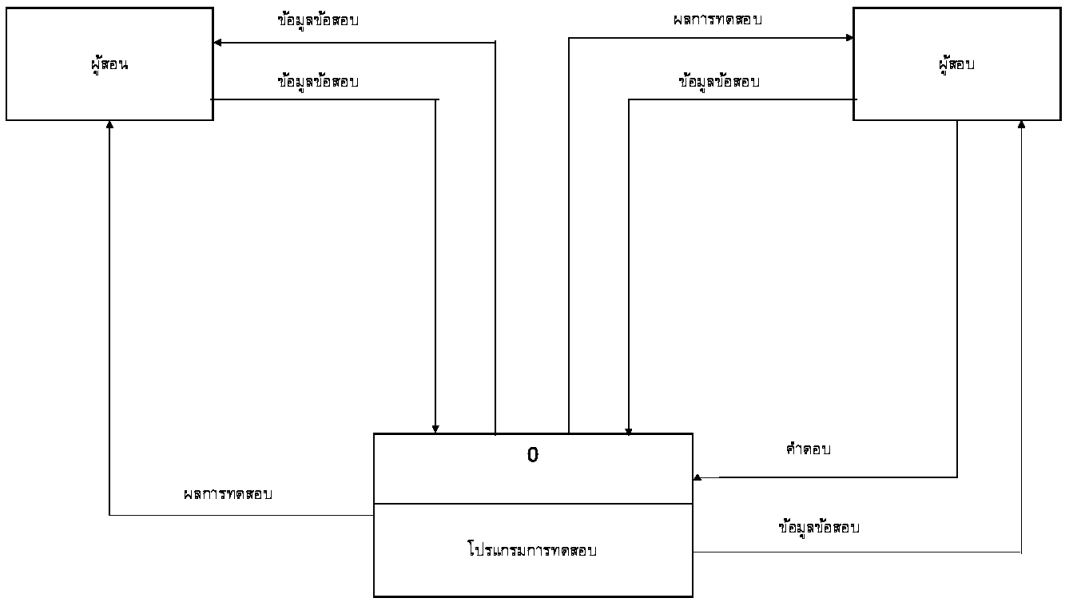
1. การออกแบบโปรแกรม มีรายละเอียดดังนี้

1.1 วิเคราะห์แผนผังบริบท (Context Diagram) เป็นการวิเคราะห์เพื่อหา Source Destination ที่เกี่ยวข้องกับโปรแกรมการทดสอบโดยกำหนดสัญลักษณ์ที่ใช้ในการวิเคราะห์แผนผังบริบทดังนี้

ตารางที่ 3-5 สัญลักษณ์และความหมายที่ใช้ในการวิเคราะห์แผนผังบริบท

สัญลักษณ์	ความหมาย
	ผู้ใช้ที่มีความเกี่ยวข้องกับโปรแกรม ไม่ว่าจะเป็นผู้ป้อนข้อมูลเข้าสู่โปรแกรม หรือผู้รับข้อมูลออกจากโปรแกรม
	โปรแกรมที่พัฒนา
	ทิศทางการไหล หรือการติดต่อของข้อมูลในโปรแกรม

Source Destination ที่เกี่ยวข้องกับโปรแกรมการทดสอบ ได้แก่ ผู้สอน นักเรียน ซึ่งสามารถเขียนแผนผังบริบทแสดงทิศทางการรับส่งข้อมูลในโปรแกรมได้ ดังภาพที่ 3-17



ภาพที่ 3-17 แผนบริบทของโปรแกรมการทดสอบ

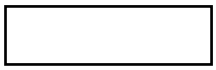
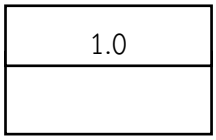
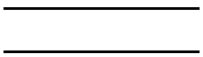
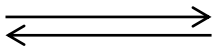
จากภาพที่ 3-17 แสดงการรับส่งข้อมูลระหว่างโปรแกรมกับ Source Destination ซึ่ง Source Destination มี 2 แหล่ง ได้แก่

ผู้สอน หมายถึง ครู อาจารย์ หรือบุคลากรที่มีหน้าที่ในการจัดการเรียนการสอนให้กับผู้สอบโดยเป็นผู้ที่สามารถเพิ่ม ลบ หรือแก้ไขข้อสอบในคลังข้อสอบได้ รวมทั้งสามารถตรวจสอบผลการทดสอบได้

ผู้สอบ หมายถึง นักเรียน หรือบุคคลอื่นๆ ในสถานศึกษาที่มีหน้าที่เรียนรู้และเข้าทดสอบเพื่อวัดความรู้ในเรื่องนั้นตามที่โปรแกรมการทดสอบมุ่งหมายที่จะวัด

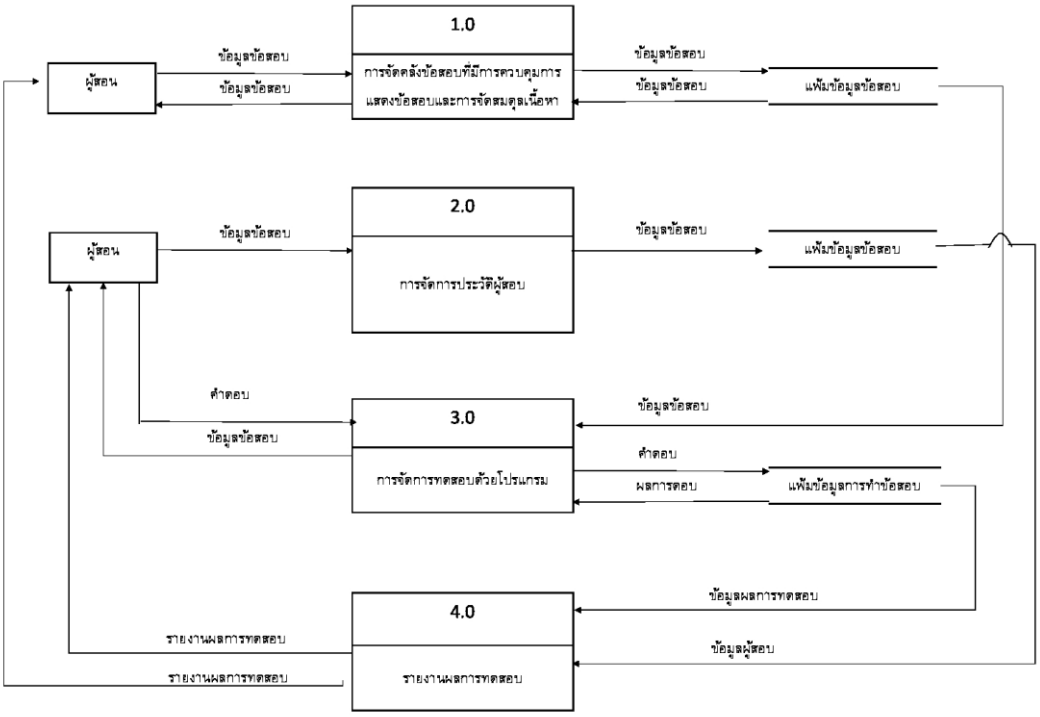
1.2 การจัดทำผังการไหลของข้อมูล (Data Flow Diagram: DFD) เป็นการวิเคราะห์ให้เห็นภาพรวมในการทำงานของโปรแกรม โดยแสดงรายละเอียดการรับหรือส่งข้อมูลภายในโปรแกรม ซึ่งมีสัญลักษณ์ที่ใช้ในการแสดงผังการไหลของข้อมูลดังนี้

ตารางที่ 3-6 สัญลักษณ์และความหมายที่ใช้ในผังการไหลของข้อมูล

สัญลักษณ์	ความหมาย
	ผู้ใช้ที่มีความเกี่ยวข้องกับโปรแกรม ไม่ว่าจะเป็นผู้ป้อนข้อมูลเข้าสู่โปรแกรม หรือผู้รับข้อมูลออกจากโปรแกรม
	การประมวลผลข้อมูลที่เกิดขึ้นในโปรแกรม
	แหล่งเก็บข้อมูล เช่น ไฟล์ ตารางจัดเก็บข้อมูล
	ทิศทางการไหล หรือการติดต่อของข้อมูลในโปรแกรม

การจัดทำผังการไหลของข้อมูลสามารถแบ่งขั้นตอนย่อยตามระดับได้ดังนี้

1. แผนผังการไหลของข้อมูลระดับ 0 (Data Flow Diagram Level 0) เป็นแผนผังแสดงการรับส่งข้อมูลที่อยู่ในกระบวนการหลักของแผนผังบริบท ได้แก่ การจัดคลังข้อสอบที่มีการจัดสมดุลเนื้อหาและการควบคุมการแสดงข้อสอบ การจัดประวัติผู้สอบ การจัดทดสอบด้วยโปรแกรม และรายงานผลการทดสอบ ซึ่งแต่ละกระบวนการมีรายละเอียด ดังภาพที่ 3-18



ภาพที่ 3-18 การรับส่งข้อมูลของกระบวนการหลักในโปรแกรมการทดสอบ

จากภาพที่ 3-18 แสดงการรับส่งข้อมูลของกระบวนการหลักในโปรแกรมการทดสอบซึ่งแต่ละกระบวนการหลักมีรายละเอียดดังนี้

กระบวนการหลักที่ 1 การจัดคลังข้อสอบ เป็นกระบวนการที่ใช้จัดเก็บข้อมูลซึ่งเป็นข้อสอบที่ใช้ในการวัดความรู้ของผู้สอบโดยข้อสอบที่เก็บเข้าคลังข้อสอบนี้ ต้องเป็นข้อสอบแบบ 4 และ 5 ตัวเลือก และผ่านการวิเคราะห์คุณภาพของข้อสอบตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ แบบตรวจให้คะแนน 2 ค่า (Dichotomous MIRT) แบบ 3 พารามิเตอร์ ข้อสอบแต่ละข้อต้องมีค่าความยากของข้อสอบ (b) มีตั้งแต่ -2.50 ถึง 2.50 ค่าอำนาจจำแนกของข้อสอบ (a) มีตั้งแต่ 0.50 ถึง 2.50 และมีค่าโอกาสในการเดาของข้อสอบไม่เกิน 0.30 (Urry, 1977) ร่วมกับการจัดสมดุลเนื้อหา (Content Balancing) โดยประยุกต์ใช้วิธีการสุ่มตัวอย่างแบบชั้นภูมิ (Stratified Random Sampling) ในการควบคุมการแสดงข้อสอบ (Item Exposure)

กระบวนการหลักที่ 2 การจัดการประวัติผู้สอบ เป็นกระบวนการที่ใช้จัดเก็บข้อมูล ซึ่งเป็นประวัติทั่วไปของผู้สอบ

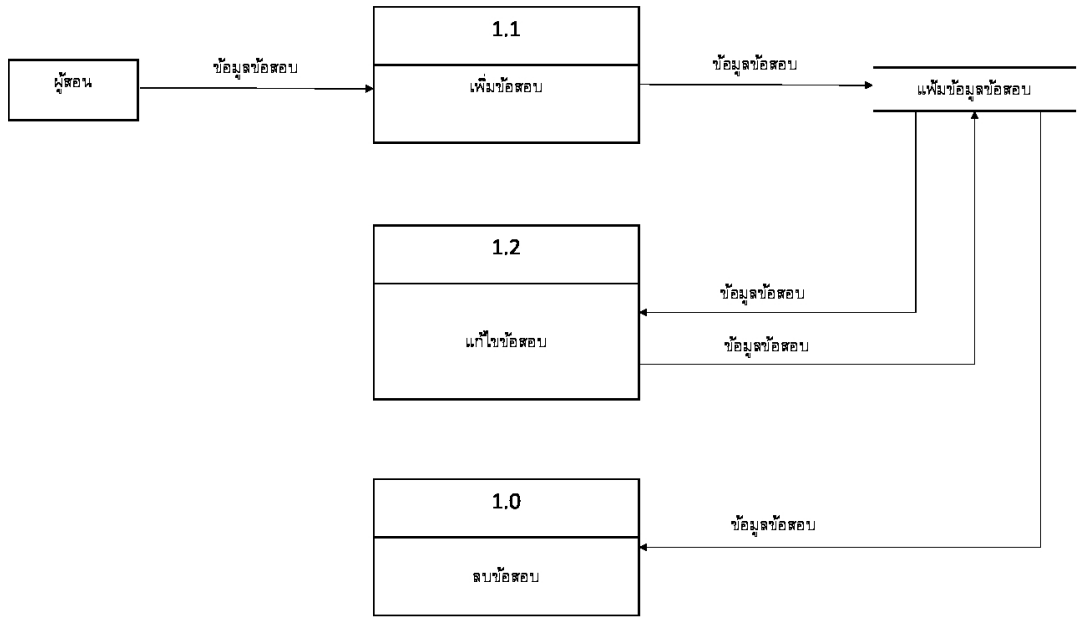
กระบวนการหลักที่ 3 การจัดการทดสอบด้วยโปรแกรม เป็นกระบวนการดำเนินการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ

กระบวนการหลักที่ 4 การรายงานผลการสอบ เป็นกระบวนการที่นำข้อมูลซึ่งเป็นผลการสอบ มาจัดทำเป็นรายงานแล้วนำเสนอให้ผู้สอบ หรือผู้เกี่ยวข้องทราบ

2. แผนผังการไหลของข้อมูลระดับ 1 (Data Flow Diagram Level 1) เป็นแผนผังที่แสดงรายละเอียดการรับส่งของข้อมูลในกระบวนการหลัก ในที่นี้มีกระบวนการหลัก 2 กระบวน ได้แก่

กระบวนการหลักที่ 1 การจัดการคลังข้อสอบ มี 3 กระบวนการย่อย คือ 1) การเพิ่มข้อสอบ 2) การแก้ไขข้อสอบ และ 3) การลบข้อสอบ แสดงดังภาพที่ 3-19

.



ภาพที่ 3-19 แผนผังการไหลของข้อมูลระดับ 1 ของกระบวนการจัดคลังข้อสอบ

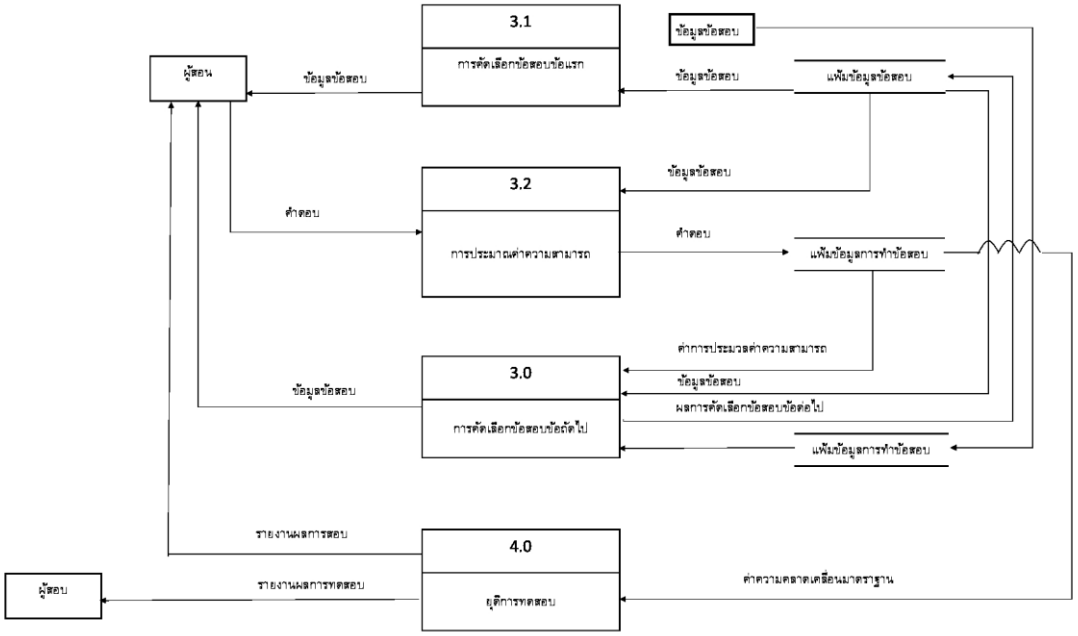
จากภาพที่ 3-19 แสดงแผนผังการรับส่งข้อมูลในกระบวนการจัดการคลังข้อสอบซึ่งมีกระบวนการย่อย 3 กระบวนการ คือ

1) การเพิ่มข้อสอบ เป็นการเพิ่มข้อสอบใหม่ลงในคลังข้อสอบโดยข้อสอบที่เพิ่มลงไปต้องเป็นข้อสอบแบบ 4 และ 5 ตัวเลือก และผ่านการวิเคราะห์คุณภาพของข้อสอบตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ แบบตรวจให้คะแนน 2 ค่า (Dichotomous MIRT) แบบ 3 พารามิเตอร์ ข้อสอบแต่ละข้อต้องมีค่าความยากของข้อสอบ (b) มีตั้งแต่ -2.50 ถึง 2.50 ค่าอำนาจจำแนกของข้อสอบ (a) มีตั้งแต่ 0.50 ถึง 2.50 และมีค่าโอกาสในการเดาของข้อสอบไม่เกิน 0.30 (Urry, 1977) ร่วมกับการจัดสมดุลเนื้อหา (Content Balancing) โดยประยุกต์ใช้วิธีการสุ่มตัวอย่างแบบชั้นภูมิ (Stratified Random Sampling) ในการควบคุมการแสดงข้อสอบ (Item Exposure)

2). การแก้ไขข้อสอบ เป็นการแก้ไขข้อสอบเดิมที่มีอยู่ในคลังข้อสอบโดยข้อสอบที่ได้รับการแก้ไขแล้ว ต้องเป็นข้อสอบแบบ 5 ตัวเลือก และผ่านการวิเคราะห์คุณภาพของข้อสอบตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ แบบตรวจให้คะแนน 2 ค่า (Dichotomans MIRT) แบบ 3 พารามิเตอร์ ข้อสอบแต่ละข้อต้องมีค่าความยากของข้อสอบ (b) มีตั้งแต่ -2.50 ถึง 2.50 ค่าอำนาจจำแนกของข้อสอบ (a) มีตั้งแต่ 0.50 ถึง 2.50 และมีค่าโอกาสในการเดาของข้อสอบไม่เกิน 0.30 (Urry, 1977) ร่วมกับการจัดสมดุลเนื้อหา (Content Balancing) โดยประยุกต์ใช้วิธีการสุ่มตัวอย่างแบบชั้นภูมิ (Stratified Random Sampling) ในการควบคุมการแสดงข้อสอบ (Item Exposure)

3). การลบข้อสอบ เป็นการลบข้อสอบที่ไม่ต้องการออกจากคลังข้อสอบ

กระบวนการหลักที่ 3 การจัดการทดสอบด้วยโปรแกรม เป็นกระบวนการจัดการทดสอบ มีกระบวนการย่อย ได้แก่ 3.1) การคัดเลือกข้อสอบข้อแรก 3.2) การประมาณค่าความสามารถของผู้สอบ 3.3) การคัดเลือกข้อสอบข้อถัดไป และ 3.4) ยุติการทดสอบ ดังภาพที่ 3-20



ภาพที่ 3-20 แผนผังการไหลของข้อมูลระดับ 1 ของกระบวนการจัดการทดสอบด้วยโปรแกรม

จากภาพที่ 3-20 แสดงการรับส่งข้อมูลในกระบวนการจัดการทดสอบด้วยโปรแกรม ซึ่งมีกระบวนการย่อย 5 กระบวนการ เป็นการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ (Multidimensional Computerized Adaptive Testing: MCAT) ซึ่งรายละเอียดมีดังนี้

ขั้นที่ 1 การศึกษาเอกสารที่เกี่ยวข้องแนวคิด Bloom et al. (1956); Anderson et al. (2001, p. XXI); Krathwohl (2002, p. 213) ทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ (Reckase, 2009) และงานวิจัยที่เกี่ยวข้องกับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ

ขั้นที่ 2 การสร้างข้อสอบ/การรวบรวมข้อสอบ O-NET ให้ครอบคลุมเนื้อทั้งหมดตามกฎเบื้องต้นในการเขียนข้อสอบ

ขั้นที่ 3 การวิเคราะห์จุดประสงค์ทางการศึกษาใหม่ตามแนวคิดของ Bloom et al. (1956); Anderson et al. (2001, p. XXI); Krathwohl (2002, p. 213) และการวิเคราะห์คุณภาพข้อสอบตามทฤษฎีการตอบสนองของข้อสอบแบบพหุมิติ

ชั้นที่ 4 การศึกษาหลักการของวิธีการจัดสมดุลเนื้อหาและการแสดงข้อสอบ

ขั้นที่ 5 การพัฒนาวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ระดับชั้นร่วมกับการจัดสมดุลเนื้อหา มีรายละเอียดดังนี้

ขั้นที่ 5.1 การพัฒนาคลังข้อสอบเริ่มจากแบ่งชั้นคลังข้อสอบเป็น 4 ชั้นตามระดับ

ค่าอำนาจจำแนกข้อสอบ 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

ขั้นที่ 5.2 การพัฒนาคลังข้อสอบเริ่มจากแบ่งชั้นคลังข้อสอบเป็น 4 ชั้นตามระดับค่าอำนาจจำแนกข้อสอบร่วมกับการจัดสมดุลเนื้อหาในแต่ละชั้น โดยในแต่ละชั้นได้จัดกลุ่มเนื้อหาข้อสอบ O-NET วิชาคณิตศาสตร์ โดยจำแนกเป็น 4 กลุ่มสาระการเรียนรู้ที่ออกข้อสอบ ได้แก่ 1) จำนวนและการดำเนินการ 2) การวัด3) พีชคณิต และ 4) สาระการเรียนรู้การวิเคราะห์ข้อมูลข้อมูลและความน่าจะเป็น

ขั้นที่ 5.3 วิธีการควบคุมการจัดสมดุลเนื้อหาด้วยวิธีการสุ่มตัวอย่างแบบชั้นภูมิ (Stratified Random Sampling) โดยให้กลุ่มสาระเป็นชั้นภูมิ และให้ค่าความยากของข้อสอบ (b) ที่มีลักษณะเหมือนกัน (Homogenious) เป็นหน่วยการสุ่ม ขั้นตอนการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ระดับชั้นร่วมกับการจัดสมดุลเนื้อหาและการควบคุมการแสดงข้อสอบโดยใช้วิธีการสุ่มตัวอย่างแบบชั้นภูมิ มีรายละเอียดดังนี้

1) จัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ระดับชั้นร่วมกับการจัดสมดุลเนื้อหา ตามรายละเอียดที่ให้ไว้ในระยะที่ 5.2

2) การคัดเลือกข้อสอบข้อแรก กำหนดให้ Initity Ability = 0

3) เมื่อได้ข้อสอบข้อแรก หาผลการตอบข้อสอบ ซึ่งข้อสอบจะให้ผลการตอบข้อสอบเป็น 0 และ 1 ถ้าผู้สอบตอบถูกผลการตอบจะเป็น 1 ถ้าผู้สอบตอบผิดผลการตอบจะเป็น 0 แล้วคำนวณหาค่า $P_i(\theta)$ จากสมการ (14) และประมาณค่า θ ด้วยวิธีการประมาณค่าแบบเบส์ (Bayesian Estimation)

4) การคัดเลือกข้อสอบข้อถัดไปพิจารณาจากค่าความยากของข้อสอบ (b) ข้อก่อนหน้าโดยใช้วิธีควบคุมการแสดงข้อสอบด้วยวิธีการสุ่มตัวอย่างแบบชั้นภูมิที่คำนึงถึงการจัดสมดุลเนื้อหาข้อสอบโดยจะให้กลุ่มสาระการเรียนรู้คณิตศาสตร์เป็นชั้นภูมิ และให้ค่าความยากของข้อสอบ (b) ที่มีลักษณะเหมือนกัน (Homogenious) เป็นหน่วยการสุ่ม โดยการบังคับกลุ่มสาระการเรียนรู้ที่ถูกเลือกไปให้ผู้สอบแล้วไม่ให้ถูกเลือกขึ้นมาซ้ำ

5) เกณฑ์การยุติการทดสอบในแต่ละชั้นและเลื่อนชั้นการทดสอบ ผู้สอบจะต้องทำการทดสอบสอบในชั้นนั้นๆให้ครบ 4 ข้อ ตามเนื้อหาสาระการเรียนรู้ที่ออกข้อสอบ O-NET วิชาคณิตศาสตร์ หรือค่าความคลาดเคลื่อนมาตรฐาน น้อยกว่าหรือเท่ากับ 0.30 หรือ ค่า SE ≤ 0.30 และรวบรวมผลการทดสอบในแต่ละชั้น

6) ในกรณีที่ไม่สามารถหาข้อสอบจัดให้ได้และไม่มีข้อสอบอื่น จะทำซ้ำในขั้นที่ 3 จนกระทั่งหาข้อสอบข้อใดข้อหนึ่งจัดให้

7) ขั้นที่ 4 ถึงขั้นที่ 5 จะทำซ้ำๆ จนกระทั่งความยาวแบบทดสอบย่อยครบตามกำหนดที่ชั้นการทดสอบนั้น การทดสอบจะต่อเนื่องไปยังขั้นต่อไป

8) ขั้นที่ 4 ถึงขั้นที่ 5 จะทำซ้ำๆ จนกระทั่งครบทุกขั้นการทดสอบและครบทุกชั้นคลังข้อสอบ การทดสอบจึงยุติ

2. การออกแบบฐานข้อมูล

การออกฐานข้อมูลของโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติในการวิจัยครั้งนี้ มี 8 แฟ้มข้อมูล คือ 1) แฟ้มข้อมูลข้อสอบ 2) แฟ้มข้อมูลผู้สอบ 3) แฟ้มข้อมูลวันที่ทำการทดสอบของนักเรียน 4) แฟ้มข้อมูลผลการทดสอบของนักเรียน 5) แฟ้มข้อมูลตารางค่า Theta 6) แฟ้มข้อมูลแบบประเมิน (สำหรับผู้เชี่ยวชาญ) 7) แฟ้มข้อมูลแบบประเมิน (สำหรับผู้ใช้งาน) และ 8) แฟ้มข้อมูลผู้ดูแลระบบ มีรายละเอียดดังนี้

2.1 แฟ้มข้อมูลข้อสอบ (Examination) มีทั้งหมด 17 Fields ดังตารางที่ 3-7

ตารางที่ 3-7 แฟ้มข้อมูลข้อสอบ

key	Field Name	Field Type	Size	Note	Sample
PK	EXAM_SEQNO	Int	11	Running Number	1
	EXAM_GROUP_STRATIFIED	Varchar	2	คลังข้อสอบชั้นที่	01
	EXAM_GROUP_NO	Varchar	5	ชุดของข้อสอบ	00001
	EXAM_SECTION	Varchar	400	เนื้อหาของข้อสอบ	จำนวนและการดำเนินการ การวัด พีชคณิต การวิเคราะห์และความน่าจะเป็น
	EXAM_QUESTION	Varchar	20	โจทย์	10 -3 =?
	EXAM_CHOICE_A	Varchar	20	ตัวเลือกที่ 1	A
	EXAM_CHOICE_B	Varchar	20	ตัวเลือกที่ 2	B
	EXAM_CHOICE_C	Varchar	20	ตัวเลือกที่ 3	C
	EXAM_CHOICE_D	Varchar	20	ตัวเลือกที่ 4	D
	EXAM_CHOICE_E	Varchar	20	ตัวเลือกที่ 5	E
	EXAM_A_PARAMETER	Double	(1,3)	ค่าอำนาจจำแนก	0.123
	EXAM_B_PARAMETER	Double	(1,3)	ค่าความยาก	0.265
	EXAM_C_PARAMETER	Double	(1,3)	ค่าโอกาสในการเดา	0.12
	EXAM_DIFFICULTY_LEVEL	Int	11	ระดับความยาก	3
	EXAM_COLLECT_CHOICE	Varchar	20	คำตอบที่ถูก	B
	EXAM_ADD_DATE	Date		วันที่ทำการบันทึกข้อมูล	11 may 2016
	EXAM_STATUS	Varchar	1	สถานะการใช้งานข้อสอบ	A = ใช้งาน S = ไม่ได้ใช้งาน

2.2 แฟ้มข้อมูลผู้สอบ (Student_info) มีทั้งหมด 9 Fields ดังตารางที่ 3-8

ตารางที่ 3-8 แฟ้มข้อมูลผู้สอบ

key	Field Name	Field Type	Size	Note	Sample
PK	STUDENT_SEQNO	Int	11	Running Number	1
	STUDENT_ID_CARD	Varchar	13	รหัสบัตรประชาชน	3120100686889
	STUDENT_SCHOOL	Varchar	100	ชื่อโรงเรียน	รร.หอวัง
	STUDENT_PROVINCE	Varchar	20	จังหวัด	ม.6/2
	STUDENT_NAME	Varchar	100	ชื่อ-นามสกุลนักเรียน	มานะ เรียนดี
	STUDENT_GENDER	Varchar	1	เพศ	M = ชาย, F = หญิง
	STUDENT_ADD_DATE	Datatime	-	วันที่ทำการบันทึกข้อมูล	11 may 2016
	STUDENT_STATUS	Varchar	1	สถานะผู้ทำการทดสอบ	A = ใช้งาน S = ไม่ได้ใช้งาน
	STUDENT_PASSWORD	Varchar	20	รหัสผ่าน	Password

2.3 แฟ้มข้อมูลวันที่ทำการทดสอบของนักเรียน (Student_Exam_Version) มีทั้งหมด 5 Fields ดังตารางที่ 3-9

ตารางที่ 3-9 แฟ้มข้อมูลวันที่ทดสอบของนักเรียน

key	Field Name	Field Type	Size	Note	Sample
PK	VERSION_SEQNO	Int	11	Running Number	1
FK	STUDENT_SEQNO	Int	11	Running Number ของแฟ้มข้อมูลผู้สอบ	1
	VERSION_NO	Int	11	ครั้งที่ทำการทดสอบ	1
	VERSION_START	Timestamp	-	วันที่เริ่มทำข้อสอบ	2017-11-01 12:40:33
	VERSION_END	Timestamp	-	วันที่ทำข้อสอบเสร็จ	2017-11-01 12:59:27

2.4 แฟ้มข้อมูลผลการทดสอบของนักเรียน (Student_Exam_Result) มีทั้งหมด 14 Fields ดังตารางที่ 3-10

ตารางที่ 3-10 แฟ้มข้อมูลผลการทดสอบของนักเรียน

key	Field Name	Field Type	Size	Note	Sample
PK	RESULT_SEQNO	Int	11	Running Number	1
FK	EXAM_SEQNO	Int	11	Running Number ของแฟ้มข้อสอบ	1

ตารางที่ 3-10 (ต่อ)

key	Field Name	Field Type	Size	Note	Sample
FK	STUDENT_SEQNO	Int	11	Running Number ของแฟ้มข้อมูลผู้สอบ	1
FK	VERSION_SEQNO	Int	11	Running Number ของแฟ้มวันที่ทำการทดสอบ	1
	RESULT_ANSWER	Varchar	10	คำตอบที่เลือก	A
	RESULT_CORRECT	Varchar	1	ผลการตอบ	Y = ตอบถูก N = ตอบผิด
	RESULT_THETA	Double	(16,5)	ค่า Theta สะสม	-0.20515
	RESULT_ADD_DATE	Date		วันที่ทำการบันทึกข้อมูล	11 may 2016
	RESULT_STATUS	Varchar	1	สถานะแฟ้มข้อมูลผลการทดสอบของนักเรียน	A = ใช้งาน S = ไม่ได้ใช้งาน
	RESULT_KULL_BACK	Double	(16,5)	ค่าประมาณความสามารถ	-0.01306
	RESULT_STANDARD_ERROR	Double	(16,5)	เกณฑ์การยุติ	0.81314
	RESULT_LEVEL	Int	11	ชั้นของข้อสอบ	1
	RESULT_THETA_LEVEL	Double	(16,5)	ค่า Theta รายข้อ	-0.20515
	RESULT_QUESTION	Int	11	คำถามข้อที่ได้ทำ	16

2.5 แฟ้มข้อมูลตารางค่า Theta (Theta_Examination) มีทั้งหมด 6 Fields ดังตารางที่ 3-11

ตารางที่ 3-11 แฟ้มข้อมูลตารางค่า Theta

key	Field Name	Field Type	Size	Note	Sample
PK	THETA_SEQNO	Int	11	Running Number	1
FK	EXAM_SEQNO	Int	11	Running Number ของแฟ้มข้อสอบ	123
	THETA_LENGTH	Double	(16,2)		2.2
	THETA_AXR	Double	(16,5)		0.00443
	THETA_LXR_CORRECT	Double	(16,5)		0.89076
	THETA_LXR_WRONG	Double	(16,5)		0.10924

2.6 แฟ้มข้อมูลแบบประเมิน (สำหรับผู้เชี่ยวชาญ) (Evaluation_Specialist) มีทั้งหมด 38 Fields ดังตารางที่ 3-12

ตารางที่ 3-12 แฟ้มข้อมูลแบบประเมิน (สำหรับผู้เชี่ยวชาญ)

key	Field Name	Field Type	Size	Note	Sample
PK	EVALUATION_SP_SEQNO	Int	11	Running Number	1
	EVALUATION_SP_DATE	Date	-	วันที่บันทึกแบบประเมิน	2017-11-01
	EVALUATION_SP_RESULT11	Int	11	ตอนที่ 1 ข้อที่ 1	4
	EVALUATION_SP_RESULT12	Int	11	ตอนที่ 1 ข้อที่ 2	3
	EVALUATION_SP_RESULT13	Int	11	ตอนที่ 1 ข้อที่ 3	2
	EVALUATION_SP_RESULT14	Int	11	ตอนที่ 1 ข้อที่ 4	1
	EVALUATION_SP_RESULT15	Int	11	ตอนที่ 1 ข้อที่ 5	1
	EVALUATION_SP_RESULT21	Int	11	ตอนที่ 2 ข้อที่ 1	2
	EVALUATION_SP_RESULT22	Int	11	ตอนที่ 2 ข้อที่ 2	3
	EVALUATION_SP_RESULT23	Int	11	ตอนที่ 2 ข้อที่ 3	4
	EVALUATION_SP_RESULT24	Int	11	ตอนที่ 2 ข้อที่ 4	5
	EVALUATION_SP_RESULT25	Int	11	ตอนที่ 2 ข้อที่ 5	5
	EVALUATION_SP_RESULT26	Int	11	ตอนที่ 2 ข้อที่ 6	5
	EVALUATION_SP_RESULT27	Int	11	ตอนที่ 2 ข้อที่ 7	5
	EVALUATION_SP_RESULT28	Int	11	ตอนที่ 2 ข้อที่ 8	5
	EVALUATION_SP_RESULT29	Int	11	ตอนที่ 2 ข้อที่ 9	4
	EVALUATION_SP_RESULT210	Int	11	ตอนที่ 2 ข้อที่ 10	3
	EVALUATION_SP_RESULT31	Int	11	ตอนที่ 3 ข้อที่ 1	2
	EVALUATION_SP_RESULT32	Int	11	ตอนที่ 3 ข้อที่ 2	2
	EVALUATION_SP_RESULT33	Int	11	ตอนที่ 3 ข้อที่ 3	3
	EVALUATION_SP_RESULT34	Int	11	ตอนที่ 3 ข้อที่ 4	3
	EVALUATION_SP_RESULT35	Int	11	ตอนที่ 3 ข้อที่ 5	4
	EVALUATION_SP_RESULT36	Int	11	ตอนที่ 3 ข้อที่ 6	3
	EVALUATION_SP_RESULT37	Int	11	ตอนที่ 3 ข้อที่ 7	2
	EVALUATION_SP_RESULT38	Int	11	ตอนที่ 3 ข้อที่ 8	3
	EVALUATION_SP_RESULT39	Int	11	ตอนที่ 3 ข้อที่ 9	2
	EVALUATION_SP_RESULT310	Int	11	ตอนที่ 3 ข้อที่ 10	1
	EVALUATION_SP_RESULT41	Int	11	ตอนที่ 4 ข้อที่ 1	2
	EVALUATION_SP_RESULT42	Int	11	ตอนที่ 4 ข้อที่ 2	2
	EVALUATION_SP_RESULT43	Int	11	ตอนที่ 4 ข้อที่ 3	2
	EVALUATION_SP_RESULT44	Int	11	ตอนที่ 4 ข้อที่ 4	2
	EVALUATION_SP_RESULT45	Int	11	ตอนที่ 4 ข้อที่ 5	2
	EVALUATION_SP_RESULT51	Int	11	ตอนที่ 5 ข้อที่ 1	2
	EVALUATION_SP_RESULT52	Int	11	ตอนที่ 5 ข้อที่ 2	2
	EVALUATION_SP_RESULT53	Int	11	ตอนที่ 5 ข้อที่ 3	2

ตารางที่ 3-12 แฟ้มข้อมูลแบบประเมิน (สำหรับผู้เชี่ยวชาญ)

key	Field Name	Field Type	Size	Note	Sample
	EVALUATION_SP_RESULT54	Int	11	ตอนที่ 5 ข้อที่ 4	3
	EVALUATION_SP_RESULT55	Int	11	ตอนที่ 5 ข้อที่ 5	3
	EVALUATION_SP_COMMENT	Varchar	500	คำแนะนำเพิ่มเติม	

2.7 แฟ้มข้อมูลแบบประเมิน (สำหรับผู้ใช้งาน) (Evaluation_Student) มีทั้งหมด 19 Fields ดังตารางที่ 3-13

ตารางที่ 3-13 แฟ้มข้อมูลแบบประเมิน (สำหรับผู้ใช้งาน)

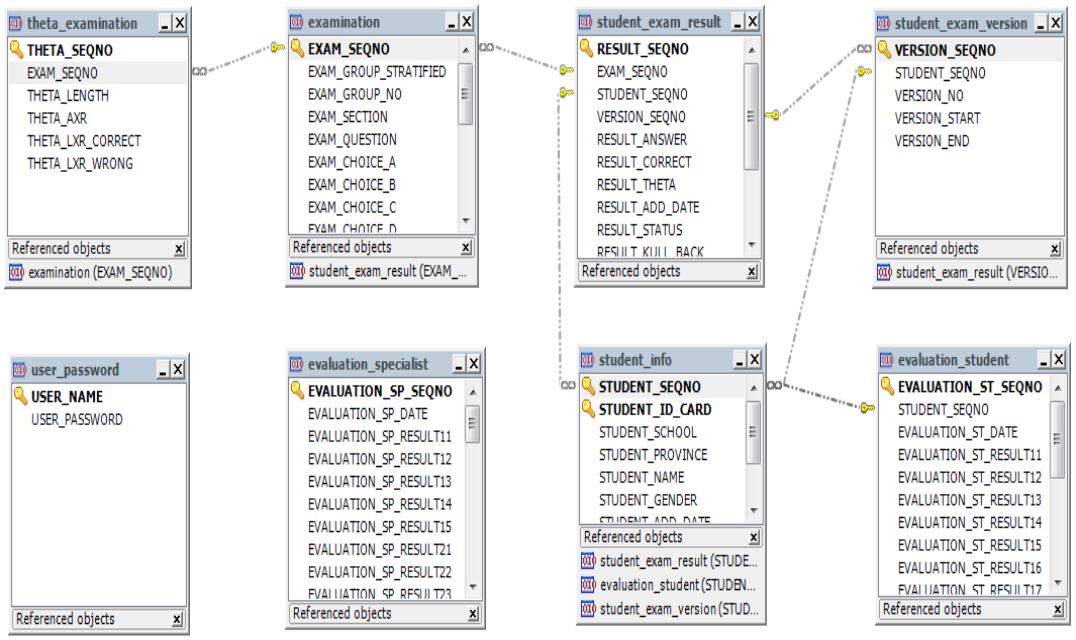
key	Field Name	Field Type	Size	Note	Sample
PK	EVALUATION_ST_SEQNO	Int	11	Running Number	1
	STUDENT_SEQNO	Int	11	Running Number ของแฟ้มข้อมูลผู้สอบ	1
	EVALUATION_ST_DATE	Date	-	วันที่บันทึกแบบประเมิน	2017-11-01
	EVALUATION_ST_RESULT11	Int	11	ตอนที่ 1 ข้อที่ 1	4
	EVALUATION_ST_RESULT12	Int	11	ตอนที่ 1 ข้อที่ 2	3
	EVALUATION_ST_RESULT13	Int	11	ตอนที่ 1 ข้อที่ 3	2
	EVALUATION_ST_RESULT14	Int	11	ตอนที่ 1 ข้อที่ 4	1
	EVALUATION_ST_RESULT15	Int	11	ตอนที่ 1 ข้อที่ 5	1
	EVALUATION_ST_RESULT16	Int	11	ตอนที่ 1 ข้อที่ 6	2
	EVALUATION_ST_RESULT17	Int	11	ตอนที่ 1 ข้อที่ 7	3
	EVALUATION_ST_RESULT18	Int	11	ตอนที่ 1 ข้อที่ 8	4
	EVALUATION_ST_RESULT19	Int	11	ตอนที่ 1 ข้อที่ 9	5
	EVALUATION_ST_RESULT110	Int	11	ตอนที่ 1 ข้อที่ 10	5
	EVALUATION_ST_RESULT21	Int	11	ตอนที่ 2 ข้อที่ 1	5
	EVALUATION_ST_RESULT22	Int	11	ตอนที่ 2 ข้อที่ 2	5
	EVALUATION_ST_RESULT23	Int	11	ตอนที่ 2 ข้อที่ 3	5
	EVALUATION_ST_RESULT24	Int	11	ตอนที่ 2 ข้อที่ 4	4
	EVALUATION_ST_RESULT25	Int	11	ตอนที่ 2 ข้อที่ 5	3
	EVALUATION_ST_COMMENT	Varchar	500	คำแนะนำเพิ่มเติม	

2.8) แฟ้มข้อมูลผู้ดูแลระบบ (User_Password) มีทั้งหมด 2 Fields ดังตารางที่ 3-14

ตารางที่ 3-14 แฟ้มข้อมูลผู้ดูแลระบบ

key	Field Name	Field Type	Size	Note	Sample
PK	USER_NAME	Varchar	20	รหัสผู้ใช้งาน	mcatadmin
	USER_PASSWORD	Varchar	20	รหัสเข้าใช้งาน	P@ssw0rd

ความสัมพันธ์ของตารางข้อมูล สามารถแสดงดังภาพที่ 3-21



ภาพที่ 3-21 ความสัมพันธ์ของตารางข้อมูล

ขั้นที่ 3 การพัฒนาโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ

การพัฒนาโปรแกรมในครั้งนี้เป็นการพัฒนาโปรแกรมในรูปแบบ Website โดยใช้ภาษา JAVA ในการพัฒนาโปรแกรม ใช้ MySQL เป็นฐานข้อมูล และใช้ภาษา SQL เป็นคำสั่งในการเชื่อมโยงในฐานข้อมูล โครงสร้างหน้าจอทั่วไปของโปรแกรมการทดสอบเป็นการแจ้งรายละเอียดต่างๆ ที่เกี่ยวข้องกับโปรแกรมการทดสอบ

1. โครงสร้างหน้าจอหลัก (Home Screen) แสดงรายละเอียดต่าง ๆ ได้แก่ Header: Logo Program, User Menu: หัวข้อเมนูของผู้ทำการสอบ, Admin Login Screen: เมนูสำหรับ ผู้ดูแลระบบ, IACB-MCAT Detail Screen: ปุ่มกดไปหน้าความหมายของ IACB-MCAT, Program Instruction Screen: ปุ่มกดไปหน้าคำชี้แจงของระบบ และ Start Exam Screen: ปุ่มกดไปหน้าเข้าสู่ระบบ ดังภาพที่ 3-22

<table>
<tr><td colspan="3">Header</td></tr>
<tr><td colspan="2">User Menu</td><td>Admin Login Screen</td></tr>
<tr><td>IACB-MCAT Detail Screen</td><td>Program Instruction Screen</td><td>Start Exam Screen</td></tr>
</table>

ภาพที่ 3-22 โครงสร้างหน้าจอหลัก

2. โครงสร้างหน้าจอเมนูผู้ใช้งาน (User Menu) แสดงรายละเอียดต่างๆ ได้แก่ Home Screen: กลับไปหน้าเริ่มต้น, About Web Screen: หน้าจอความหมายของคลังข้อสอบ, Test Result Login Screen: หน้าจอเข้าสู่ระบบตรวจสอบผลการทดสอบตามบุคคล และ Contact Us Screen: หน้าจอข้อมูลผู้จัดทำระบบ ดังภาพที่ 3-23

<table>
<tr><td>Home Screen</td><td>About Web Screen</td><td>Test Result Login Screen</td><td>Contact Us Screen</td></tr>
</table>

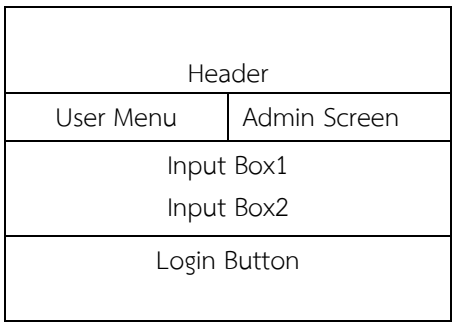
ภาพที่ 3-23 โครงสร้างหน้าจอเมนูผู้ใช้งาน

3. โครงสร้างหน้าจอเว็บ (About Web Screen) แสดงรายละเอียดต่างๆ ได้แก่ Detail: รายละเอียด ของคลังข้อสอบ ดังภาพที่ 3-24

<table>
<tr><td colspan="2">Header</td></tr>
<tr><td>User Menu</td><td>Admin Screen</td></tr>
<tr><td colspan="2">Detail</td></tr>
</table>

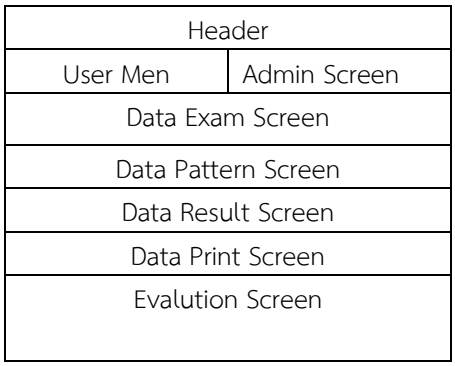
ภาพที่ 3-24 โครงสร้างหน้าจอเว็บ

4. โครงสร้างหน้าจอเข้าสู่ระบบตรวจสอบผลการทดสอบ (Test Result Login Screen) แสดงรายละเอียดต่าง ๆ ได้แก่ Input Box1: กรอกตัวเลขบัตรประชาชน, Input Box2: กรอกรหัสผ่านเข้าสู่ระบบ และ Login Button: ปุ่มกดไปสู่หน้า รายงานผลการทดสอบ ดังภาพที่ 3-25



ภาพที่ 3-25 โครงสร้างหน้าจอเข้าสู่ระบบตรวจสอบผลการทดสอบ

5. โครงสร้างหน้าจอผลการทดสอบ (Test Result Screen) แสดงรายละเอียดต่างๆ ได้แก่ Data Exam Screen: เปิดหน้า, Data Pattern Screen: เปิดหน้า ผลการตอบจำแนกตามจุดประสงค์ทางการศึกษาด้านพุทธิปัญญา, Data Result Screen: เปิดหน้า ผลการวิเคราะห์ความสามารถของผู้สอบจำแนกตามชั้นของค่าอำนาจจำแนกข้อสอบ, Data Print Screen: เปิดหน้าพิมพ์รายงาน และ Evalution Screen: เปิดหน้า แบบประเมิน ดังภาพที่ 3-26



ภาพที่ 3-26 โครงสร้างหน้าจอผลการทดสอบ

6. โครงสร้างหน้าจอผลการตอบข้อสอบเป็นรายข้อ (Data Exam Screen) แสดงรายละเอียดต่าง ๆ ได้แก่ Detail: ข้อมูลผลการตอบข้อสอบเป็นรายข้อจำแนกตามชั้นของค่าอำนาจจำแนกของข้อสอบ ดังภาพที่ 3-27

ภาพที่ 3-27 โครงสร้างหน้าจอผลการตอบข้อสอบเป็นรายข้อ

7. โครงสร้างหน้าจอผลการตอบจำแนกตามจุดประสงค์ทางการศึกษา (Data Pattern Screen) แสดงรายละเอียดต่าง ๆ ได้แก่ Detail: ข้อมูลผลการตอบจำแนกตามจุดประสงค์ทางการศึกษาด้านพุทธิปัญญา ดังภาพที่ 3-28

ภาพที่ 3-28 โครงสร้างหน้าจอผลการตอบจำแนกตามจุดประสงค์ทางการศึกษา

8. โครงสร้างหน้าจอผลการวิเคราะห์ความสามารถของผู้สอบ (Data Result Screen) แสดงรายละเอียดต่าง ๆ ได้แก่ Detail: ข้อมูลผลการวิเคราะห์ความสามารถของผู้สอบจำแนกตามชั้นของค่าอำนาจจำแนกข้อสอบ ดังภาพที่ 3-29

ภาพที่ 3-29 โครงสร้างหน้าจอผลการวิเคราะห์ความสามารถของผู้สอบ

9. โครงสร้างหน้าจอพิมพ์รายงาน (Data Print Screen) แสดงรายละเอียดต่าง ๆ ได้แก่ Detail: ข้อมูลพิมพ์รายงาน ดังภาพที่ 3-30

ภาพที่ 3-30 โครงสร้างหน้าจอพิมพ์รายงาน

10. โครงสร้างหน้าจอแบบประเมิน (Evaluation Screen) แสดงรายละเอียดต่าง ๆ ได้แก่ Detail: ข้อมูลพิมพ์รายงาน, Radio Button1: ระดับ 5 หมายถึง โปรแกรมมีความเหมาะสมมากที่สุด , Radio Button2: ระดับ 4 หมายถึง โปรแกรมมีความเหมาะสมมาก, Radio Button3: ระดับ 3 หมายถึง โปรแกรมมีความเหมาะสมปานกลาง, Radio Button4: ระดับ 2 หมายถึง โปรแกรมมีความเหมาะสมน้อย, Radio Button5: ระดับ 1 หมายถึง โปรแกรมมีความเหมาะสมน้อยที่สุด และ Save Button: ปุ่มกดบันทึกแบบประเมิน ดังภาพที่ 3-31

Detail	Radio Button1	Radio Button2	Radio Button3	Radio Button4	Radio Button5
Save Button					

ภาพที่ 3-31 โครงสร้างหน้าจอแบบประเมิน

11. โครงสร้างหน้าจอติดต่อผู้ดูแลระบบ (Contact Us Screen) แสดงรายละเอียดต่าง ๆ ได้แก่ Detail: ข้อมูลผู้จัดทำระบบ ดังภาพที่ 3-32

Header	
User Menu	Admin Screen
Detail	

ภาพที่ 3-32 โครงสร้างหน้าจอติดต่อผู้ดูแลระบบ

12. โครงสร้างหน้าจอความหมาย IACB-MCAT (IACB -MCAT Detail Screen) แสดงรายละเอียดต่าง ๆ ได้แก่ Detail: ความหมายของ IACB-MCAT ดังภาพที่ 3-33

Detail

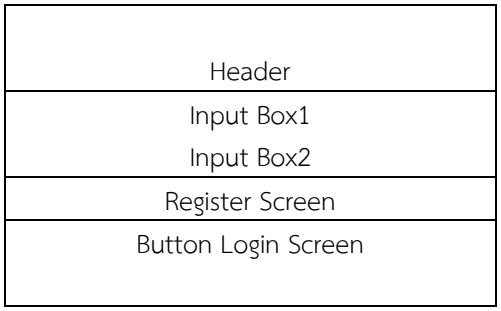
ภาพที่ 3-33 โครงสร้างหน้าจอความหมาย IACB-MCAT

13. โครงสร้างหน้าจอคำชี้แจงของระบบ (Program Instruction Screen) แสดงรายละเอียดต่าง ๆ ได้แก่ Detail: คำชี้แจงของระบบ ดังภาพที่ 3-34

Detail

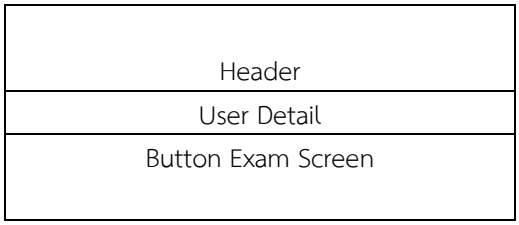
ภาพที่ 3-34 โครงสร้างหน้าจอคำชี้แจงของระบบ

14. โครงสร้างหน้าจอเข้าสู่ระบบ (Start Exam Screen) แสดงรายละเอียดต่าง ๆ ได้แก่ Input Box1: รหัสผู้ใช้งานระบบ (กรณีเคยละทะเบียนไว้แล้ว), Input Box2: รหัสผ่านเข้าใช้งานระบบ (กรณีเคยละทะเบียนไว้แล้ว), Register Screen: ลงทะเบียนเข้าใช้งานระบบ และ Button Login Screen: ปุ่มกดเข้าสู่หน้าเริ่มทำข้อสอบ ดังภาพที่ 3-35



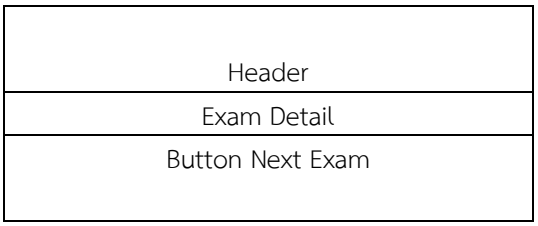
ภาพที่ 3-35 โครงสร้างหน้าจอเข้าสู่ระบบ

15. โครงสร้างหน้าจอลงทะเบียนเข้าใช้งานระบบ (Register Screen) แสดงรายละเอียดต่าง ๆ ได้แก่ User Detail: ข้อมูลรายละเอียดผู้ทำการสอบ และ Button Exam Screen: ปุ่มเริ่มทำการข้อสอบ ดังภาพที่ 3-36



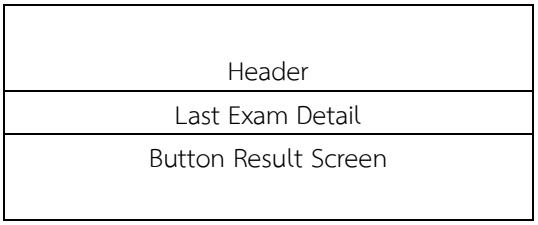
ภาพที่ 3-36 โครงสร้างหน้าจอลงทะเบียนเข้าใช้งานระบบ

16. โครงสร้างหน้าจอการทดสอบ (Exam Screen) แสดงรายละเอียดต่าง ๆ ได้แก่ Exam Detail: ข้อมูล โจทย์ และคำตอบให้เลือก และ Button Next Exam: ทำโจทย์ข้อต่อไป ดังภาพที่ 3-37



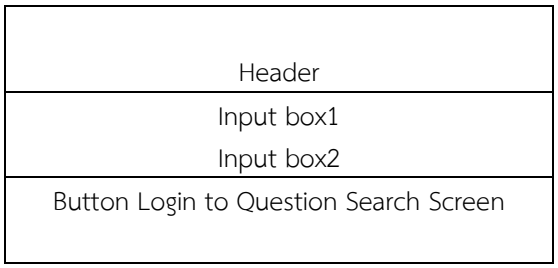
ภาพที่ 3-37 โครงสร้างหน้าจอการทดสอบ

17. โครงสร้างหน้าจอการทดสอบครั้งสุดท้าย (Last Exam Screen) แสดงรายละเอียดต่าง ๆ ได้แก่ Last Exam Detail: โจทย์ข้อสุดท้าย และ Button Result Screen: ปุ่มประมวลผลข้อสอบที่ได้ทำไป ไปต่อที่หน้า ผลการทดสอบ ดังภาพที่ 3-38



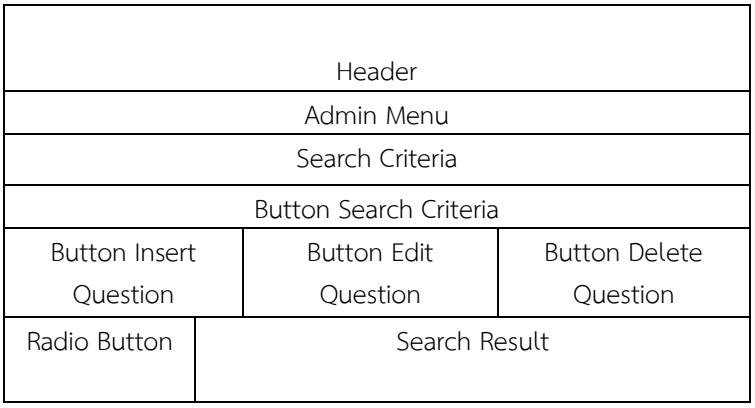
ภาพที่ 3-38 โครงสร้างหน้าจอการทดสอบครั้งสุดท้าย

18. โครงสร้างหน้าจอเข้าสู่ระบบผู้ดูแลระบบ (Admin Login Screen) แสดงรายละเอียดต่าง ๆ ได้แก่ Input box1: กรอกรหัสผู้ใช้งาน, Input box2: กรอกรหัสผ่านเพื่อเข้าใช้งาน และ Button Login to Question Search Screen: กดปุ่มยืนยันเพื่อเข้าใช้งาน ดังภาพที่ 3-39



ภาพที่ 3-39 โครงสร้างหน้าจอเข้าสู่ระบบผู้ดูแลระบบ

19. โครงสร้างหน้าจอค้นหาข้อสอบ (Question Search Screen) แสดงรายละเอียดต่าง ๆ ได้แก่ Admin Menu: เมนูของผู้ดูแลระบบ, Search Criteria: ตัวเลือกการค้นหาข้อสอบ, Button Search Criteria: ปุ่มกดค้นหาข้อสอบ, Button Insert Question: ปุ่มกดเพิ่มข้อสอบ, Button Edit Question: ปุ่มกดแก้ไขข้อสอบ, Button Delete Question: ปุ่มกดลบข้อสอบ, Radio Button: ปุ่มเลือกข้อสอบที่จะทำการ แก้ไข หรือ ลบ และ Search Result: ผลการค้นหาข้อสอบ ดังภาพที่ 3-40

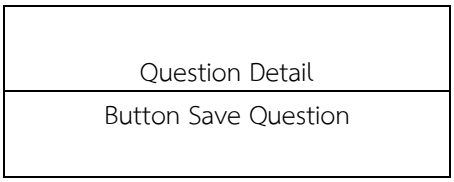


ภาพที่ 3-40 โครงสร้างหน้าจอค้นหาข้อสอ

20. โครงสร้างหน้าจอเมนูดูแลระบบ (Admin Menu) แสดงรายละเอียดต่าง ๆ ได้แก่ Home Screen: กลับไปที่หน้าเริ่มต้น ดังภาพที่ 3-41

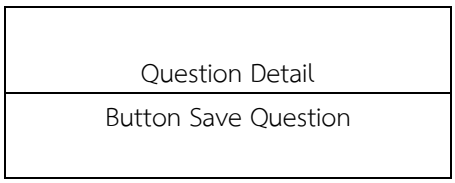
ภาพที่ 3-41 โครงสร้างหน้าจอเมนูดูแลระบบ

21. โครงสร้างหน้าจอเพิ่มข้อสอบ (Add Question Screen) แสดงรายละเอียดต่าง ๆ ได้แก่ Question Detail: รายละเอียดของข้อสอบ และ Button Save Question: บันทึกข้อสอบ ดังภาพที่ 3-42



ภาพที่ 3-42 โครงสร้างหน้าจอเพิ่มข้อสอบ

22. โครงสร้างหน้าจอแก้ไขข้อสอบ (Edit Question Screen) แสดงรายละเอียดต่าง ๆ ได้แก่ Question Detail: รายละเอียดของข้อสอบที่แก้ไขข้อมูลได้ และ Button Save Question: บันทึกข้อสอบ ดังภาพที่ 3-43



ภาพที่ 3-43 โครงสร้างหน้าจอแก้ไขข้อสอบ

ขั้นที่ 4 การทดสอบและแก้ไขโปรแกรม

เมื่อพัฒนาโปรแกรมการทดสอบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติเรียบร้อยแล้ว ผู้วิจัยร่วมกับโปรแกรมเมอร์ได้ตรวจสอบโปรแกรมการทดสอบเพื่อหาจุดบกพร่อง ได้แก่ เครื่องหมายการดำเนินการทางคณิตศาสตร์ผิด (+ - x ÷) พิมพ์ตัวแปรผิด รวมทั้งการใช้คำสั่งในการเขียนโปรแกรมผิดพลาด ซึ่งได้ตรวจสอบทีละคำสั่ง แล้วแก้ไขโปรแกรมให้ถูกต้อง นอกจากนี้ เพื่อให้มั่นใจว่าโปรแกรมการทดสอบที่พัฒนาขึ้นมีความถูกต้อง สามารถใช้งานได้จริง จึงประมวลผลโปรแกรมการทดสอบ แล้วนำผลการทดสอบเทียบกับผลการทดสอบที่ได้จากการศึกษาในสถานการณ์จำลอง ปรากฏว่า โปรแกรมการทดสอบที่พัฒนาขึ้นให้ผลการทดสอบเท่ากับผลการทดสอบจากการศึกษาในสถานการณ์จำลอง จึงมั่นใจได้ว่าโปรแกรมการทดสอบที่พัฒนาขึ้นไม่มีความผิดพลาด จากนั้นจึงนำโปรแกรมเสนอต่ออาจารย์ที่ปรึกษาเพื่อขอความคิดเห็น แล้วนำข้อคิดเห็นที่ได้มาปรับปรุงโปรแกรมให้มีความสมบูรณ์

ขั้นที่ 5 การจัดทำคู่มือการใช้โปรแกรม

หลังจากทดสอบและแก้ไขโปรแกรมเป็นที่พอใจแล้ว ได้จัดทำคู่มือในการใช้โปรแกรมการทดสอบนี้เพื่ออธิบายวิธีการใช้โปรแกรม โดยแบ่งออกเป็น 2 ส่วน ได้แก่ ส่วนที่ 1 การจัดการข้อสอบ ประกอบด้วย การตั้งรหัสผู้สอบ การจัดการข้อสอบ การรายงานผลสอบ (ต่ออาจารย์) และ การออกจากระบบ และส่วนที่ 2 การจัดการสอบ ประกอบด้วย เริ่มต้นการทดสอบ กรอกข้อมูลผู้สอบ การดำเนินการสอบ และการรายงานผลการสอบ (ต่อผู้สอบ) รายละเอียดแสดงใน ภาคผนวก จ

ขั้นที่ 6 การประเมินความเหมาะสมของโปรแกรม

การประเมินความเหมาะสมของโปรแกรม ดำเนินการหลังการตรวจสอบ ปรับปรุงแก้ไข และจัดทำคู่มือการใช้โปรแกรมเป็นที่เรียบร้อยแล้ว โดยผู้วิจัยนำโปรแกรมและคู่มือการใช้งานไปประเมินความเหมาะสม ซึ่งแบ่งเป็น 2 ส่วน คือ ประเมินความเหมาะสมโดยผู้เชี่ยวชาญ และการประเมินความเหมาะสมโดยผู้ใช้งาน โดยมีรายละเอียดดังนี้

1. การประเมินความเหมาะสมโดยผู้เชี่ยวชาญ

การประเมินความเหมาะสมของโปรแกรมการทดสอบโดยผู้เชี่ยวชาญ ได้กำหนดเกณฑ์ว่าผู้เชี่ยวชาญต้องมีความรู้ทางด้านการวัดผล ด้านคณิตศาสตร์ หรือด้านการพัฒนาโปรแกรมคอมพิวเตอร์ ไม่น้อยกว่า 10 ปี และมีวุฒิการศึกษาไม่ต่ำกว่าระดับปริญญาปริญญาโท ผู้เชี่ยวชาญ 3 คน ประกอบด้วย

1.1 อาจารย์ ดร.มิ่ง เทพครเมือง อาจารย์ประจำสาระการเรียนรู้วิชาคณิตศาสตร์ โรงเรียนสาธิตมหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร มหาวิทยาลัยศรีนครินทรวิโรฒ

1.2 อาจารย์ ดร.ปิยะ ถิรพันธ์เมธี อาจารย์ประจำสาขาวิชาคอมพิวเตอร์และเทคโนโลยีสารสนเทศ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยเทคโนโลยีราชมงคลกรุงเทพ

1.3 อาจารย์รัชกฤช ธนพัฒนดล อาจารย์ประจำสาขาวิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยอีสเทิร์นเอเชีย

ผู้เชี่ยวชาญได้ประเมินความเหมาะสมของโปรแกรมทั้ง 5 ด้าน ได้แก่ 1) ด้านความต้องการของผู้ใช้โปรแกรม 2) ด้านการทำงานของโปรแกรม 3) ด้านการใช้งาน 4) ด้านการรักษาความปลอดภัยของการเข้าถึงข้อมูล 5) ด้านความชัดเจนของคู่มือการใช้งาน และข้อเสนอแนะเพื่อการพัฒนาโปรแกรม โดยใช้แบบประเมินด้วยวิธี Black-Box Testing ลักษณะมาตรประมาณค่า (Rating Scale) 5 ระดับ (รายละเอียดแสดงในภาคผนวก ฉ)

2. การประเมินความเหมาะสมโดยผู้ใช้งาน

การประเมินความเหมาะสมโดยผู้ใช้งาน เป็นการประเมินความคิดเห็นของผู้ใช้งานโปรแกรมทั้ง 2 ด้าน ได้แก่ 1) ด้านการใช้งาน 2) ด้านความชัดเจนของคู่มือการใช้งาน และข้อเสนอแนะเพื่อการพัฒนาโปรแกรม โดยกลุ่มตัวอย่างที่ทดลองใช้เป็นนักเรียนมัธยมศึกษาชั้นปีที่ 6 โรงเรียนบางกะปิ จำนวน 30 คน คัดเลือกด้วยวิธีการเลือกตัวอย่างตามสะดวก โดยใช้แบบประเมินด้วยวิธี Black-Box Testing ลักษณะมาตรประมาณค่า (Rating Scale) 5 ระดับ (รายละเอียดแสดงในภาคผนวก ช) โดยในข้อ 6.1 และ 6.2 มีเกณฑ์การให้คะแนน ดังนี้

5 หมายถึง เหมาะสมมากที่สุด

4 หมายถึง เหมาะสมมาก

3 หมายถึง เหมาะสมปานกลาง

2 หมายถึง เหมาะสมน้อย

1 หมายถึง เหมาะสมน้อยที่สุด

เกณฑ์การแปลความหมายค่าเฉลี่ย ดังนี้

ค่าเฉลี่ยตั้งแต่ 4.51 ถึง 5.00 หมายถึง โปรแกรมมีความเหมาะสมมากที่สุด

ค่าเฉลี่ยตั้งแต่ 3.51 ถึง 4.50 หมายถึง โปรแกรมมีความเหมาะสมมาก

ค่าเฉลี่ยตั้งแต่ 2.51 ถึง 3.50 หมายถึง โปรแกรมมีความเหมาะสมปานกลาง

ค่าเฉลี่ยตั้งแต่ 1.51 ถึง 2.50 หมายถึง โปรแกรมมีความเหมาะสมน้อย

ค่าเฉลี่ยตั้งแต่ 0.00 ถึง 1.50 หมายถึง โปรแกรมมีความเหมาะสมน้อยที่สุด

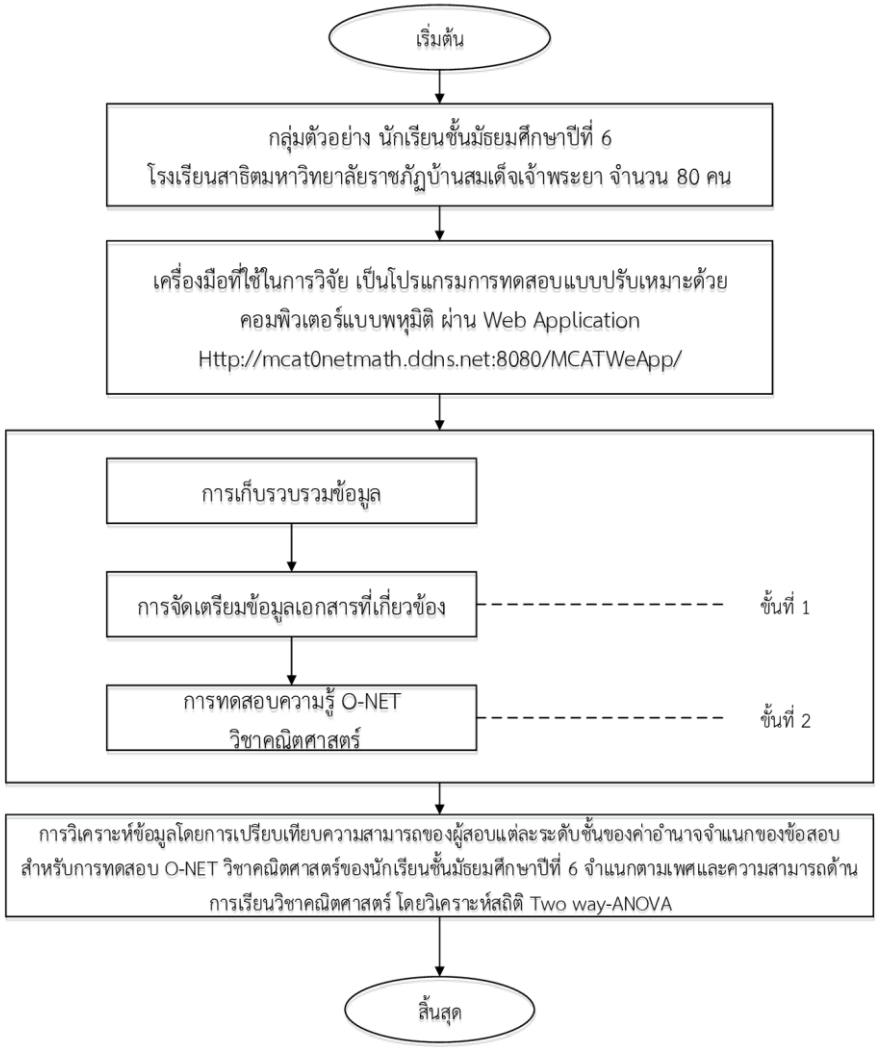
การตรวจสอบประสิทธิภาพโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ ดังนี้

1) ข้อมูลจากผู้ทดลองใช้ที่เป็นนักเรียน วิเคราะห์ด้วยสถิติพื้นฐาน ได้แก่ ค่าเฉลี่ย (Mean) และส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation) และ วิธี Black-Box Testing

2) ข้อมูลจากผู้เชี่ยวชาญ วิเคราะห์ด้วยดัชนีความตรงเชิงเนื้อหา (Content Validity Index: CVI) และ วิธี Black-Box Testing

ระยะที่ 4 การเปรียบเทียบความสามารถของผู้สอบแต่ละระดับชั้นของค่าอำนาจจำแนกของข้อสอบ สำหรับการทดสอบ O-NET วิชาคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 จำแนกตามเพศและความสามารถด้านการเรียนคณิตศาสตร์

การวิจัยในระยะนี้ใช้เทคนิคการวิจัยแบบ Factorial Design โดยใช้แบบแผนการทดลองแบบ 2 x 2 Factorial Design (Edmonds & Kennedy, 2017, p. 77) ซึ่งมีขั้นตอนการทดสอบดังภาพที่ 3-44



ภาพที่ 3-44 ขั้นตอนการทดสอบในระยะที่ 4

จากภาพที่ 3-44 การทดสอบในระยะที่ 4 ที่ทดลองใช้โปรแกรมจริงกับนักเรียนชั้นมัธยมศึกษาปีที่ 6 มาเปรียบเทียบความสามารถของผู้สอบแต่ละชั้นของค่าอำนาจจำแนกของข้อสอบวิชาคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 โรงเรียนสาธิตมหาวิทยาลัยราชภัฏบ้านสมเด็จเจ้าพระยา จำนวน 112 คน โดยจำแนกตาม เพศ และความสามารถด้านการเรียนวิชาคณิตศาสตร์

1. กลุ่มตัวอย่างที่ใช้ในการวิจัยครั้งนี้ เป็นนักเรียนชั้นมัธยมศึกษาชั้นปีที่ 6 ที่เรียนวิชาคณิตศาสตร์ ในภาคเรียนที่ 2 ปีการศึกษา 2560 โรงเรียนสาธิตมหาวิทยาลัยราชภัฏบ้านสมเด็จเจ้าพระยา จำนวน 80 คน ใช้วิธีการสุ่มตัวอย่างง่าย (Simple Random Sampling)

2. เครื่องมือที่ใช้ในการวิจัย เป็นโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ สำหรับการทดสอบ O-NET วิชาคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 6

3. การเก็บรวบรวมข้อมูล ในการวิจัยครั้งนี้แบ่งเป็น 2 ขั้นตอน คือ ขั้นตอนที่ 1 การจัดเตรียมข้อมูลเอกสารที่เกี่ยวข้องในการจัดเก็บข้อมูล ขั้นตอนที่ 2 การทดสอบความรู้วิชาคณิตศาสตร์จากคลังข้อสอบตามระดับชั้น มีรายละเอียดดังนี้

ขั้นที่ 1 การจัดเตรียมข้อมูลเอกสารที่เกี่ยวข้อง มีขั้นตอนการดำเนินการ ดังนี้

1. ทำหนังสือจากวิทยาลัยวิทยาการวิจัยและวิทยาการปัญญา มหาวิทยาลัยบูรพา ถึงผู้อำนวยการโรงเรียนสาธิตมหาวิทยาลัยราชภัฏบ้านสมเด็จเจ้าพระยา เพื่อขอความอนุเคราะห์และความสะดวกในการเก็บข้อมูล

2. เตรียมเอกสารชี้แจงการใช้งานโปรแกรมและแบบประเมินความเหมาะสมของโปรแกรมที่ผู้วิจัยพัฒนาขึ้น ผ่าน http://mcatonetmath.ddns.net:8080/MCATWebApp/

3. นำหนังสือขอความร่วมมือไปติดต่อโรงเรียนสาธิตมหาวิทยาลัยราชภัฏบ้านสมเด็จเจ้าพระยาเพื่อขอความอนุเคราะห์เก็บข้อมูล

ขั้นที่ 2 การทดสอบความรู้ O-NET วิชาคณิตศาสตร์ สำหรับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ มีขั้นตอนการดำเนินการดังนี้

1. เก็บข้อมูลจากนักเรียนมัธยมศึกษาชั้นปีที่ 6 โรงเรียนสาธิตมหาวิทยาลัยราชภัฏบ้านสมเด็จเจ้าพระยา วันพฤหัสบดีที่ 18 มกราคม พ.ศ. 2561 ที่เรียนวิชาคณิตศาสตร์ ภาคเรียนที่ 2 ปีการศึกษา 2560 จำนวน 80 คน ทดสอบด้วยโปรแกรม ผ่าน http://mcatonetmath.ddns.net:8080/MCATWebApp/

2. อธิบายเงื่อนไขหลักการของการทดสอบด้วยโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติกับนักเรียนอย่างละเอียดพร้อมกับแจกวิธีการดำเนินการทดสอบให้กับนักเรียน

3. ดำเนินการทดสอบด้วยโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติกับนักเรียนซึ่งจัดการทดสอบผ่านทาง Website นักเรียนสามารถทำการทดสอบที่ใดก็ได้

หลังจากเสร็จสิ้นการทดสอบตามวันที่กำหนด ผู้วิจัยนำผลการทดสอบของนักเรียน มาวิเคราะห์ความสามารถแต่ละชั้นของค่าอำนาจจำแนกของข้อสอบวิชาคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 โดยจำแนกตาม เพศ และเกรดเฉลี่ยสะสมของนักเรียน

4. การวิเคราะห์ข้อมูลโดยการเปรียบเทียบความสามารถของผู้สอบแต่ละระดับชั้นของค่าอำนาจจำแนกของข้อสอบ สำหรับการทดสอบ O-NET วิชาคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 จำแนกตามเพศและความสามารถด้านการเรียนวิชาคณิตศาสตร์ วิเคราะห์ข้อมูลโดยใช้ Two Way-ANOVA

บทที่ 4
ผลการวิจัย

การวิจัยนี้มีวัตถุประสงค์เพื่อ 1) พัฒนาวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา สำหรับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ 2) เปรียบเทียบประสิทธิภาพของวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา กับวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบถ่วงน้ำหนักที่มีการบังคับ โดยเปรียบเทียบประสิทธิภาพ ดังนี้ (1) การประมาณค่าความสามารถของผู้สอบ ได้แก่ ค่าความลำเอียงเฉลี่ย และค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (2) ประสิทธิภาพของการใช้ข้อสอบ ได้แก่ ข้อสอบที่มีการแสดงมากเกินไป ข้อสอบที่มีการใช้น้อยเกินไป อัตราการทับซ้อนของข้อสอบ และการแจกแจงอัตราการแสดงข้อสอบ 3) พัฒนาโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติที่ใช้วิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา และ 4) เปรียบเทียบความสามารถของผู้สอบแต่ละระดับชั้นของค่าอำนาจจำแนกของข้อสอบ สำหรับการทดสอบ O-NET วิชาคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 จำแนกตามเพศและความสามารถด้านการเรียนวิชาคณิศาสตร์ของนักเรียน ข้อมูลที่ใช้ในการวิจัยเป็นข้อมูลทุติยภูมิ โดยใช้ข้อสอบ O-NET วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 6 ระหว่างปีการศึกษา 2551–2553 และปีการศึกษา 2555–2559 ที่ได้มาจากสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) ผลการวิจัยแบ่งเป็น 4 ตอน ดังนี้

ตอนที่ 1 ผลการพัฒนาวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา

ตอนที่ 2 ผลการจำลองเปรียบเทียบประสิทธิภาพของวิธีการจัดคลังข้อสอบ

ตอนที่ 3 การพัฒนาโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติโดยใช้คลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา

ตอนที่ 4 ผลการเปรียบเทียบความสามารถของผู้สอบแต่ละระดับชั้นของค่าอำนาจจำแนกของข้อสอบ สำหรับการทดสอบ O-NET วิชาคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 จำแนกตามเพศและความสามารถด้านการเรียนวิชาคณิศาสตร์

สัญลักษณ์ที่ใช้ในการเสนอผลการวิเคราะห์ข้อมูล

ในการนำเสนอผลการวิเคราะห์ข้อมูลในการวิจัยนี้ ได้กำหนดสัญลักษณ์ที่ใช้แทนค่าตัวแปรและค่าสถิติต่าง ๆ ไว้ดังนี้

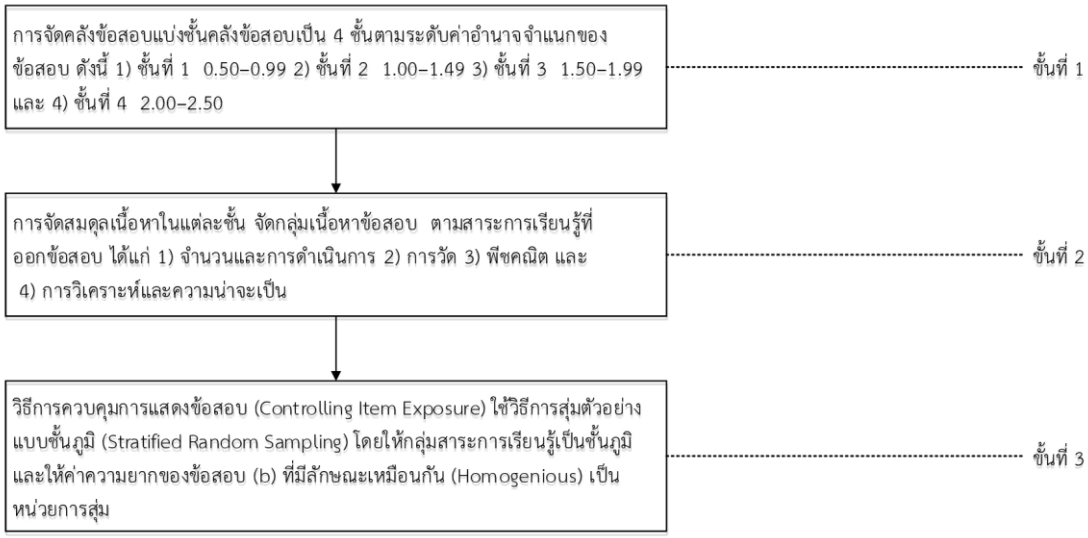
n	หมายถึง	ขนาดกลุ่มตัวอย่าง
Mean	หมายถึง	ค่าเฉลี่ยคณิต
SD	หมายถึง	ส่วนเบี่ยงเบนมาตรฐาน
Min	หมายถึง	ค่าต่ำสุด
Max	หมายถึง	ค่าสูงสุด

θ	หมายถึง	ความสามารถที่แท้จริงของผู้สอบ
$\hat{\theta}$	หมายถึง	ค่าประมาณความสามารถของผู้สอบ
a	หมายถึง	ค่าอำนาจจำแนกของข้อสอบ
b	หมายถึง	ค่าความยากของข้อสอบ
c	หมายถึง	ค่าโอกาสการเดาของข้อสอบ
F	หมายถึง	ค่าอัตราส่วนความแปรปรวน (ค่าสถิติ F)
MS	หมายถึง	ค่าเฉลี่ยผลรวมคะแนนเบี่ยงเบนยกกำลังสอง
SS	หมายถึง	ผลรวมของคะแนนเบี่ยงเบนยกกำลังสอง
F_{ratio}	หมายถึง	การเปรียบเทียบการแจกแจงอัตราการแสดงข้อสอบ
χ^2	หมายถึง	ค่าไคสแควร์
Wilcoxon Test	หมายถึง	ค่าสถิตินอนพาราเมตริกซ์วิลคอกซัน
z	หมายถึง	สถิติ z ของค่าสถิติทดสอบวิลคอกซัน
p-value	หมายถึง	ระดับนัยสำคัญ
df	หมายถึง	องศาอิสระ
*	หมายถึง	มีนัยสำคัญทางสถิติที่ระดับ .05
**	หมายถึง	มีนัยสำคัญทางสถิติที่ระดับ .01
IACB	หมายถึง	วิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา
CWA	หมายถึง	วิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบถ่วงน้ำหนักที่มีการบังคับ

ผลการวิจัย ปรากฏดังนี้

ตอนที่ 1 ผลการพัฒนาวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา

ผลการพัฒนาวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา โดยประยุกต์โมเดลการตอบสนองข้อสอบแบบพหุมิติ รายละเอียดมี ดังภาพที่ 4-1



ภาพที่ 4-1 วิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา

จากภาพที่ 4-1 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 ชั้นตามระดับค่าอำนาจจำแนกของข้อสอบร่วมกับการจัดสมดุลเนื้อหาในแต่ละชั้น โดยในแต่ละชั้นได้จัดกลุ่มเนื้อหาข้อสอบ O-NET วิชาคณิตศาสตร์ โดยจำแนกเป็น 4 กลุ่มสาระการเรียนรู้ที่ออกข้อสอบ ได้แก่ 1) จำนวนและการดำเนินการ 2) การวัด 3) พีชคณิต และ 4) การวิเคราะห์ข้อมูลและความน่าจะเป็น

3. วิธีการควบคุมการแสดงข้อสอบด้วยวิธีการสุ่มตัวอย่างแบบชั้นภูมิ (Stratified Random Sampling) โดยให้กลุ่มสาระการเรียนรู้เป็นชั้นภูมิ และให้ค่าความยากของข้อสอบ (b) ที่มีลักษณะเหมือนกัน (Homogenious) เป็นหน่วยการสุ่ม ขั้นตอนการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ระดับชั้นร่วมกับการจัดสมดุลเนื้อหาและการควบคุมการแสดงข้อสอบโดยใช้วิธีการสุ่มตัวอย่างแบบชั้นภูมิ

ตอนที่ 2 ผลการจำลองเปรียบเทียบประสิทธิภาพของวิธีการจัดคลังข้อสอบ

การเปรียบเทียบประสิทธิภาพวิธีการจัดคลังข้อสอบในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ เป็นการนำวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกข้อสอบ การจัดสมดุลเนื้อหาและการควบคุมการแสดงข้อสอบที่ผู้วิจัยพัฒนาขึ้น มาเปรียบเทียบประสิทธิภาพกับวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกข้อสอบที่คล้ายคลึงกัน 2 วิธี ได้แก่

1) วิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา

2) วิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบถ่วงน้ำหนัก

ประสิทธิภาพของวิธีการจัดคลังข้อสอบ พิจารณาจาก 1) การประมาณค่าความสามารถของผู้สอบ ได้แก่ ค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) และค่าความลำเอียงเฉลี่ย (Average Bias) 2) ประสิทธิภาพของการใช้ข้อสอบ ได้แก่ ข้อสอบที่มีการแสดงมากเกินไป (Overexposed Item) ข้อสอบที่มีการใช้น้อยเกินไป (Underutilized Item) อัตราการทับซ้อนของข้อสอบ (Item Overlap Rate) และการแจกแจงอัตราการแสดงข้อสอบ (Item Exposue Rat Distribution) การดำเนินการด้วยการศึกษาในสถานการจำลอง โดยจำลองสถานการณ์การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติที่ใช้วิธีการจัดคลังข้อสอบที่คล้ายคลึงกัน 2 วิธีข้างต้น การทดสอบในแต่ละวิธีการ ดำเนินการซ้ำทั้งหมด 10 รอบ และใช้สถานการณ์จำลอง 500 สถานการณ์ โดยทุกรอบของแต่ละวิธีใช้ข้อสอบจริงที่แบ่งคลังข้อสอบตามชั้นของค่าอำนาจออกเป็น 4 ชั้น ได้สถานการณ์จำลองในแต่ละชั้นและค่าประมาณความสามารถของผู้สอบในแต่ละชั้นของแต่ละวิธีที่แตกต่างกัน

ค่าความสามารถที่แท้จริงของผู้สอบ ถูกจำลองขึ้นด้วยโปรแกรม WinGen3 โดยใช้ค่าพารามิเตอร์ของข้อสอบที่ได้จากสถานการณ์จำลองในแต่ละชั้นของทั้ง 2 วิธี โดย 1 รอบ ทำ 4 ครั้ง และประมาณค่าความสามารถที่แท้จริงของผู้สอบ จำนวน 500 สถานการณ์ โดยกำหนดให้โปรแกรมสุ่มข้อมูลจากเลขสุ่มที่มีการแจกแจงแบบปกติ N (0,1) และนำรายการตอบที่ได้จากโปรแกรม winGen3 มาเปรียบเทียบกับสถานการณ์จำลองว่าในรายการตอบเดียวกันค่าประมาณความสามารถที่แท้จริงมีค่าเท่ากันหรือใกล้เคียงกันมากที่สุด จำนวน 10 ครั้ง แสดงดังตารางที่ 4-1

ตารางที่ 4-1 ค่าสถิติพื้นฐานของค่าความสามารถที่แท้จริงของผู้สอบในคลังข้อสอบชั้นที่ 1

ค่าความสามารถที่แท้จริง (θ)	*n*	Min	Max	Mean	*SD*	Skewness	Kurtosis
วิธี IACB							
รอบที่ 1	500	-0.508	2.461	1.005	0.481	0.127	-0.212
รอบที่ 2	500	-0.279	2.648	1.014	0.494	0.133	0.011
รอบที่ 3	500	-0.497	2.638	0.974	0.494	0.086	0.224
รอบที่ 4	500	-0.159	2.321	0.992	0.497	0.116	-0.400
รอบที่ 5	500	-0.614	2.435	1.028	0.505	-0.157	-0.003
รอบที่ 6	500	-0.240	2.500	1.048	0.486	0.090	-0.089

ตารางที่ 4-1 (ต่อ)

ค่าความสามารถที่แท้จริง (θ)	*n*	Min	Max	Mean	*SD*	Skewness	Kurtosis
วิธี IACB							
รอบที่ 7	500	-0.430	2.290	0.990	0.457	0.100	-0.213
รอบที่ 8	500	-0.500	2.430	1.014	0.521	-0.067	-0.136
รอบที่ 9	500	-0.290	2.560	0.998	0.504	0.080	-0.130
รอบที่ 10	500	-0.260	2.290	0.998	0.462	0.139	-0.143
วิธี CWA							
รอบที่ 1	500	-0.714	2.356	1.011	0.492	-0.181	0.077
รอบที่ 2	500	-0.645	2.439	0.969	0.523	0.116	-0.193
รอบที่ 3	500	-0.285	2.708	1.017	0.487	0.113	-0.057
รอบที่ 4	500	-0.668	2.292	0.978	0.495	-0.066	-0.053
รอบที่ 5	500	-0.474	2.425	1.025	0.517	-0.125	-0.033
รอบที่ 6	500	-0.570	2.310	0.958	0.489	-0.166	0.021
รอบที่ 7	500	-0.440	2.640	1.018	0.488	0.195	0.006
รอบที่ 8	500	-0.190	2.370	1.023	0.499	0.193	-0.186
รอบที่ 9	500	-0.720	2.520	1.048	0.515	-0.145	0.171
รอบที่ 10	500	-0.590	2.360	0.998	0.488	0.043	-0.172

จากตารางที่ 4-1 แสดงค่าสถิติพื้นฐานของค่าความสามารถที่แท้จริงของผู้สอบ (θ) ในคลังข้อสอบชั้นที่ 1 โดยใช้ค่าพารามิเตอร์ของข้อสอบที่ได้จากโปรแกรมจำลองที่ผู้วิจัยพัฒนาขึ้นทั้ง 2 วิธี ไปจำลองค่าความสามารถที่แท้จริงของผู้สอบ ด้วยโปรแกรม WinGen3 จำนวน 500 ค่า แล้วจำลองรายการตอบของผู้สอบที่เหมือนกัน เพื่อพิจารณาว่าค่าความสามารถที่แท้จริงของผู้สอบ ค่าที่ได้ต้องมีค่าเหมือนหรือใกล้เคียงกับโปรแกรมที่ผู้วิจัยพัฒนา จำนวน 10 รอบ ปรากฏว่า วิธี IACB ค่าความสามารถที่แท้จริงของผู้สอบที่จำลองขึ้นทั้ง 10 รอบ มีลักษณะเป็นโค้งระฆังคว่ำ (การแจกแจงแบบปกติ) มีค่าเฉลี่ยเข้าใกล้ 1 ค่าความเบ้ (Skewness) และค่าความโด่ง (Kurtosis) เข้าใกล้ 0 ส่วนวิธี CWA ค่าความสามารถที่แท้จริงของผู้สอบที่จำลองขึ้นทั้ง 10 รอบ มีลักษณะเป็นโค้งระฆังคว่ำ (การแจกแจงแบบปกติ) มีค่าเฉลี่ยเข้าใกล้ 1 ค่าความเบ้ (Skewness) และค่าความโด่ง (Kurtosis) เข้าใกล้ 0 (รายละเอียดค่าความสามารถที่แท้จริงของผู้สอบชั้นที่ 1 แสดงในภาคผนวก ค1)

ตารางที่ 4-2 ค่าสถิติพื้นฐานของค่าความสามารถที่แท้จริงของผู้สอบในคลังข้อสอบชั้นที่ 2

ค่าความสามารถที่แท้จริง (θ)	*n*	Min	Max	Mean	*SD*	Skewness	Kurtosis
วิธี IACB							
รอบที่ 1	500	-0.256	2.294	0.997	0.462	0.139	-0.143
รอบที่ 2	500	-0.379	2.283	1.028	0.479	0.038	-0.290
รอบที่ 3	500	-0.452	2.644	1.015	0.496	0.106	0.409
รอบที่ 4	500	-0.611	2.598	1.010	0.488	-0.017	0.051
รอบที่ 5	500	-0.434	2.256	0.995	0.465	0.016	-0.296
รอบที่ 6	500	-0.490	2.670	1.019	0.518	0.214	0.072
รอบที่ 7	500	-0.540	2.420	1.027	0.505	0.098	-0.195
รอบที่ 8	500	-0.470	2.330	0.996	0.483	-0.003	-0.164
รอบที่ 9	500	-0.630	2.770	1.018	0.536	0.156	0.229
รอบที่ 10	500	-0.570	2.340	0.999	0.495	-0.270	0.123
วิธี CWA							
รอบที่ 1	500	-1.018	2.654	0.978	0.482	-0.052	0.266
รอบที่ 2	500	-0.586	2.363	0.998	0.488	0.043	-0.172
รอบที่ 3	500	-0.677	2.409	0.989	0.483	-0.043	-0.240
รอบที่ 4	500	-0.653	2.535	1.007	0.506	0.066	-0.178
รอบที่ 5	500	-0.595	2.759	1.025	0.503	0.173	0.355
รอบที่ 6	500	-0.750	2.710	0.988	0.495	-0.038	0.237
รอบที่ 7	500	-0.990	1.280	-0.275	0.574	0.150	-0.262
รอบที่ 8	500	-0.420	2.580	1.000	0.515	-0.081	-0.145
รอบที่ 9	500	-0.420	2.570	1.010	0.522	-0.059	-0.163
รอบที่ 10	500	-0.670	2.290	0.978	0.495	-0.066	-0.053

จากตารางที่ 4-2 แสดงค่าสถิติพื้นฐานของค่าความสามารถที่แท้จริงของผู้สอบ (θ) ในคลังข้อสอบชั้นที่ 2 ปรากฏว่า วิธี IACB ค่าความสามารถที่แท้จริงของผู้สอบที่จำลองขึ้นทั้ง 10 รอบ มีลักษณะเป็นโค้งระฆังคว่ำ (การแจกแจงแบบปกติ) มีค่าเฉลี่ยเข้าใกล้ 1 ค่าความเบ้ (Skewness) ส่วนค่าความโด่ง (Kurtosis) เข้าใกล้ 0 และวิธี CWA ค่าความสามารถที่แท้จริงของผู้สอบที่จำลองขึ้นทั้ง 10 รอบ มีลักษณะเป็นโค้งระฆังคว่ำ (การแจกแจงแบบปกติ) มีค่าเฉลี่ยเข้าใกล้ 1 ค่าความเบ้ (Skewness) และค่าความโด่ง (Kurtosis) เข้าใกล้ 0

ตารางที่ 4-3 ค่าสถิติพื้นฐานของค่าความสามารถที่แท้จริงของผู้สอบในคลังข้อสอบชั้นที่ 3

ค่าความสามารถที่แท้จริง (θ)	*n*	Min	Max	Mean	*SD*	Skewness	Kurtosis
วิธี IACB							
รอบที่ 1	500	-0.593	3.232	0.991	0.490	0.093	0.294
รอบที่ 2	500	-0.567	2.449	0.991	0.486	-0.020	0.267
รอบที่ 3	500	-0.213	2.790	1.005	0.501	0.192	0.292
รอบที่ 4	500	-0.567	2.343	0.999	0.495	-0.170	0.123
รอบที่ 5	500	-0.842	2.575	1.025	0.500	-0.084	0.370
รอบที่ 6	500	-0.900	2.700	1.003	0.504	0.002	0.127
รอบที่ 7	500	-0.680	2.410	1.000	0.493	-0.029	-0.365
รอบที่ 8	500	-0.250	2.440	1.003	0.487	0.138	-0.013
รอบที่ 9	500	-0.490	2.750	1.019	0.476	0.104	0.313
รอบที่ 10	500	-0.840	2.580	1.025	0.500	-0.084	0.370
วิธี CWA							
รอบที่ 1	500	-0.362	2.528	1.024	0.487	0.035	-0.051
รอบที่ 2	500	-0.699	2.287	0.998	0.488	-0.130	-0.011
รอบที่ 3	500	-0.542	2.599	1.029	0.471	0.026	0.003
รอบที่ 4	500	-0.471	2.571	0.982	0.502	-0.050	-0.018
รอบที่ 5	500	-0.631	2.365	0.943	0.506	-0.086	-0.145
รอบที่ 6	500	-0.310	2.400	0.959	0.501	0.127	-0.322
รอบที่ 7	500	-0.560	2.490	1.019	0.520	0.100	0.007
รอบที่ 8	500	-0.500	2.360	0.980	0.507	-0.044	-0.179
รอบที่ 9	500	-0.350	2.380	1.037	0.514	-0.138	-0.408
รอบที่ 10	500	-0.450	2.520	0.981	0.488	0.039	0.097

จากตารางที่ 4-3 แสดงค่าสถิติพื้นฐานของค่าความสามารถที่แท้จริงของผู้สอบ (θ) ในคลังข้อสอบชั้นที่ 3 ปรากฏว่า วิธี IACB ค่าความสามารถที่แท้จริงของผู้สอบที่จำลองขึ้นทั้ง 10 รอบ มีลักษณะเป็นโค้งระฆังคว่ำ (การแจกแจงแบบปกติ) มีค่าเฉลี่ยเข้าใกล้ 1 ค่าความเบ้ (Skewness) ส่วนค่าความโด่ง (Kurtosis) เข้าใกล้ 0 และวิธี CWA ค่าความสามารถที่แท้จริงของผู้สอบที่จำลองขึ้นทั้ง 10 รอบ มีลักษณะเป็นโค้งระฆังคว่ำ (การแจกแจงแบบปกติ) มีค่าเฉลี่ยเข้าใกล้ 1 ค่าความเบ้ (Skewness) และค่าความโด่ง (Kurtosis) เข้าใกล้ 0

ตารางที่ 4-4 ค่าสถิติพื้นฐานของค่าความสามารถที่แท้จริงของผู้สอบในคลังข้อสอบชั้นที่ 4

ค่าความสามารถที่แท้จริง (θ)	*n*	Min	Max	Mean	*SD*	Skewness	Kurtosis
วิธี IACB							
รอบที่ 1	500	-0.572	2.538	0.987	0.484	-0.006	0.044
รอบที่ 2	500	-0.522	2.335	0.990	0.473	-0.033	0.149
รอบที่ 3	500	-0.597	2.669	0.997	0.483	-0.164	0.129
รอบที่ 4	500	-0.244	2.670	1.003	0.484	-0.023	-0.237
รอบที่ 5	500	-0.516	2.275	0.969	0.488	-0.046	-0.217
รอบที่ 6	500	-0.560	2.370	0.995	0.496	-0.109	-0.003
รอบที่ 7	500	-0.370	2.540	0.966	0.519	0.173	-0.066
รอบที่ 8	500	-0.330	2.460	1.034	0.490	0.026	-0.386
รอบที่ 9	500	-0.510	2.950	0.996	0.522	0.206	0.160
รอบที่ 10	500	-0.450	2.520	0.981	0.488	0.039	0.097
วิธี CWA							
รอบที่ 1	500	-0.308	2.699	1.034	0.505	0.087	-0.102
รอบที่ 2	500	-0.751	2.462	1.007	0.471	-0.146	0.295
รอบที่ 3	500	-0.495	2.635	1.001	0.492	-0.060	0.273
รอบที่ 4	500	-0.807	2.582	0.981	0.518	0.047	0.393
รอบที่ 5	500	-0.451	2.516	0.981	0.488	0.039	0.097
รอบที่ 6	500	-0.770	2.570	1.002	0.514	-0.042	0.026
รอบที่ 7	500	-0.400	2.450	0.987	0.494	-0.056	0.094
รอบที่ 8	500	-1.060	2.450	0.996	0.530	-0.174	0.197
รอบที่ 9	500	-0.240	3.080	0.988	0.506	0.158	0.105
รอบที่ 10	500	-0.560	2.490	1.019	0.520	0.100	0.007

จากตารางที่ 4-4 แสดงค่าสถิติพื้นฐานของค่าความสามารถที่แท้จริงของผู้สอบ (θ) ในคลังข้อสอบชั้นที่ 4 ปรากฏว่า วิธี IACB ค่าความสามารถที่แท้จริงของผู้สอบที่จำลองขึ้นทั้ง 10 รอบ มีลักษณะเป็นโค้งระฆังคว่ำ (การแจกแจงแบบปกติ) มีค่าเฉลี่ยเข้าใกล้ 1 ค่าความเบ้ (Skewness) ส่วนค่าความโด่ง (Kurtosis) เข้าใกล้ 0 และวิธี CWA ค่าความสามารถที่แท้จริงของผู้สอบที่จำลองขึ้นทั้ง 10 รอบ มีลักษณะเป็นโค้งระฆังคว่ำ (การแจกแจงแบบปกติ) มีค่าเฉลี่ยเข้าใกล้ 1 ค่าความเบ้ (Skewness) และค่าความโด่ง (Kurtosis) เข้าใกล้ 0

คลังข้อสอบจำลอง ถูกจำลองขึ้นจากการสุ่มตัวเลขด้วยโปรแกรม WinGen3 เพื่อจำลองค่าพารามิเตอร์ของข้อสอบ จำนวน 500 ชุด แต่ละชุดประกอบด้วย ค่าพารมมิเตอร์ 3 ค่า ได้แก่ ค่าอำนาจจำแนกของข้อสอบ (a) ค่าความยากของข้อสอบ (b) และค่าโอกาสการเดาของข้อสอบ (c) รายละเอียด ดังตารางที่ 4-5

ตารางที่ 4-5 ค่าสถิติพื้นฐานของค่าพารามิเตอร์ของข้อสอบในคลังข้อสอบจำลอง

ค่าพารามิเตอร์ของข้อสอบ		*n*	Min	Max	Mean	*SD*
ค่าอำนาจจำแนกของข้อสอบ (a)	ชั้นที่ 1	18	0.51	0.98	0.81	0.14
	ชั้นที่ 2	49	1.00	1.49	1.28	0.14
	ชั้นที่ 3	64	1.50	1.98	1.73	0.13
	ชั้นที่ 4	62	1.68	2.50	2.23	0.17
ค่าความยากของข้อสอบ (b)	ชั้นที่ 1	18	0.61	1.90	0.96	0.74
	ชั้นที่ 2	49	-0.35	2.50	1.11	0.66
	ชั้นที่ 3	64	0.66	2.49	1.27	0.33
	ชั้นที่ 4	62	0.39	2.14	1.25	0.37
ค่าโอกาสการเดาของข้อสอบ (c)	ชั้นที่ 1	18	0.00	0.27	0.11	0.08
	ชั้นที่ 2	49	0.00	0.30	0.18	0.06
	ชั้นที่ 3	64	0.04	0.30	0.18	0.06
	ชั้นที่ 4	62	0.05	0.28	0.19	0.05

จากตารางที่ 4-5 แสดงค่าสถิติพื้นฐานของค่าพารามิเตอร์ของข้อสอบในคลังข้อสอบจากการวิเคราะห์คุณภาพข้อสอบในระยะที่ 1 ประกอบด้วยค่าอำนาจจำแนกของข้อสอบ (a) มีค่าอยู่ระหว่าง 0.51 ถึง 2.50 ค่าความยากของข้อสอบ (b) มีค่าอยู่ระหว่าง -0.35 ถึง 2.50 และค่าโอกาสการเดาของข้อสอบ (c) มีค่าอยู่ระหว่าง 0.00 ถึง 0.30 ซึ่งมีความสอดคล้องกับข้อกำหนด คือ ค่าอำนาจจำแนกของข้อสอบมีค่าอยู่ระหว่าง 0.50 ถึง 2.50 ค่าความยากของข้อสอบมีค่าอยู่ระหว่าง -2.50 ถึง 2.50 และค่าโอกาสการเดาของข้อสอบมีค่าไม่เกิน 0.30

ผลการเปรียบเทียบประสิทธิภาพของวิธีการคัดเลือกข้อสอบข้อถัดไป แบ่งได้ 2 ประเด็น ดังนี้

1. ผลการเปรียบเทียบประสิทธิภาพด้านการประมาณค่าความสามารถของผู้สอบ

การเปรียบเทียบประสิทธิภาพด้านการประมาณค่าความสามารถของผู้สอบ เป็นการเปรียบเทียบความแม่นยำในการประมาณความสามารถของผู้สอบ $(\hat{\theta})$ ที่ได้จากการศึกษาในสถานการณ์จำลองการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติที่มีวิธีการจัดคลังข้อสอบที่แตกต่างกัน 2 วิธี ได้แก่ วิธีการ IACB และ วิธี CWA ซึ่งทั้ง 2 วิธีถูกทำซ้ำ 10 รอบ โดยใช้คลังข้อสอบจริงที่แบ่งออกเป็น 4 ชั้น โดยคลังข้อสอบชั้นที่ 1 มีข้อสอบ 18 ข้อ ชั้นที่ 2 มีข้อสอบ 49 ข้อ ชั้นที่ 3 มีข้อสอบ 64 ข้อ และ ชั้นที่ 4 มีข้อสอบ 62 ข้อ ชุดเดียวกัน แต่ค่าความสามารถที่แท้จริงของผู้สอบได้จากการนำรายการตอบจริงจากสถานการณ์จำลองการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ ในแต่ละรอบของแต่ละชั้นมาจำลอง แล้วนำมาคำนวณค่าความลำเอียงเฉลี่ย (Average Bias) และค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) ดังตารงที่ 4-6

ตารางที่ 4-6 ค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย และค่าความลำเอียงเฉลี่ย จำแนกตามวิธีการจัดคลังข้อสอบชั้นที่ 1

รอบที่	*n*	Average Bias		RMSE	
		วิธี IACB	วิธี CWA	วิธี IACB	วิธี CWA
รอบที่ 1	500	0.000	0.000	0.000	0.005
รอบที่ 2	500	0.000	-0.002	0.002	0.049
รอบที่ 3	500	0.001	-0.002	0.030	0.041
รอบที่ 4	500	-0.001	-0.005	0.016	0.104
รอบที่ 5	500	0.000	-0.004	0.006	0.081
รอบที่ 6	500	0.000	-0.003	0.003	0.064
รอบที่ 7	500	0.001	-0.004	0.018	0.100
รอบที่ 8	500	-0.002	-0.001	0.051	0.031
รอบที่ 9	500	0.002	-0.001	0.040	0.013
รอบที่ 10	500	-0.001	-0.003	0.017	0.068
Mean		0.000	0.003	0.018	0.056
SD		0.001	0.002	0.017	0.034

จากตารางที่ 4-6 แสดงค่าความลำเอียงเฉลี่ย และค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ยของวิธีการจัดคลังข้อสอบ ปรากฏว่า ค่าความลำเอียงเฉลี่ย เข้าใกล้ 0 มากที่สุด คือ วิธี IACB (Mean = 0.000) มีค่าน้อยกว่า วิธี CWA (Mean = 0.003) แสดงให้เห็นว่า วิธี IACB มีความแม่นยำในการประมาณค่าความสามารถของผู้สอบ สำหรับคลังข้อสอบในชั้นที่ 1 สูงกว่า วิธี CWA เมื่อพิจารณาค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย ปรากฏว่า วิธี IACB (Mean = 0.018) มีค่า RMSE น้อยกว่า วิธี CWA (Mean = 0.056) แสดงให้เห็นว่า วิธี IACB มีประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบ สำหรับคลังข้อสอบในชั้นที่ 1 สูงกว่า วิธี CWA (รายละเอียดค่าประมาณค่าความสามารถของผู้สอบ การคำนวณค่าความลำเอียงเฉลี่ย และค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ยแสดงในภาคผนวก ค2 - ค3)

ตารางที่ 4-7 ผลการวิเคราะห์ความแตกต่างของค่าความลำเอียงเฉลี่ย และค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย จำแนกตามวิธีการจัดคลังข้อสอบชั้นที่ 1

	วิธี	*n*	กลุ่ม	Mean Rank	Z	*p*-value
ค่าความลำเอียงเฉลี่ย	CWA	10	Control	5.50	-2.561*	<.05
	IACB	10	experiment	1.00		
ค่ารากที่สองของความ	CWA	10	Control	3.50	-2.090*	<.05
คลาดเคลื่อนกำลังสองเฉลี่ย	IACB	10	experiment	6.00		

จากตารางที่ 4-7 แสดงผลการวิเคราะห์ความแตกต่างของค่ารากที่สองของความลำเอียงเฉลี่ย (Average Bias) และค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) ด้วยการวิเคราะห์ Wilcoxon test เมื่อพิจารณาค่าความลำเอียง ปรากฏว่า วิธีการจัดคลังข้อสอบวิธี IACB และวิธี CWA มีค่าความลำเอียงเฉลี่ยแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ .01 เมื่อพิจารณาค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย ปรากฏว่า วิธีการจัดคลังข้อสอบวิธี IACB และวิธี CWA มีค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ยแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ .05

ตารางที่ 4-8 ค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย และค่าความลำเอียงเฉลี่ย จำแนกตามวิธีการจัดคลังข้อสอบชั้นที่ 2

รอบที่	*n*	Average Bias		RMSE	
		วิธี IACB	วิธี CWA	วิธี IACB	วิธี CWA
รอบที่ 1	500	-0.001	-0.002	0.014	0.045
รอบที่ 2	500	0.001	-0.004	0.012	0.095
รอบที่ 3	500	-0.001	0.009	0.024	0.009
รอบที่ 4	500	-0.001	-0.002	0.017	0.045
รอบที่ 5	500	-0.001	-0.004	0.033	0.100
รอบที่ 6	500	0.002	-0.002	0.035	0.039
รอบที่ 7	500	0.000	0.003	0.010	0.068
รอบที่ 8	500	0.000	-0.004	0.011	0.084
รอบที่ 9	500	0.002	-0.004	0.039	0.083
รอบที่ 10	500	-0.002	-0.004	0.040	0.085
Mean		0.000	-0.001	0.024	0.065
SD		0.001	0.004	0.12	0.029

จากตารางที่ 4-8 แสดงค่าความลำเอียงเฉลี่ย และค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ยของวิธีการจัดคลังข้อสอบ ปรากฏว่า ค่าความลำเอียงเฉลี่ย เข้าใกล้ 0 มากที่สุด คือ วิธี IACB (Mean = 0.000) มีค่าน้อยกว่า วิธี CWA (Mean = -0.001) แสดงให้เห็นว่า วิธี IACB มีความแม่นยำในการประมาณค่าความสามารถของผู้สอบ สำหรับคลังข้อสอบในชั้นที่ 2 สูงกว่า วิธี CWA เมื่อพิจารณาค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย ปรากฏว่า วิธี IACB (Mean = 0.024) มีค่า RMSE น้อยกว่า วิธี CWA (Mean = 0.065) แสดงให้เห็นว่า วิธี IACB มีประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบ สำหรับคลังข้อสอบในชั้นที่ 2 สูงกว่า วิธี CWA

ตารางที่ 4-9 ผลการวิเคราะห์ความแตกต่างของค่าความลำเอียงเฉลี่ย และค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย จำแนกตามวิธีการจัดคลังข้อสอบชั้นที่ 2

	วิธี	*n*	กลุ่ม	Mean Rank	Z	*p*-value
ค่าความลำเอียงเฉลี่ย	CWA	10	Control	5.06	-1.328	.184
	IACB	10	experiment	7.25		
ค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย	CWA	10	Control	2.00	-2.599**	<.01
	IACB	10	experiment	5.89		

จากตารางที่ 4-9 แสดงผลการวิเคราะห์ความแตกต่างของค่ารากที่สองของความลำเอียงเฉลี่ย (Average Bias) และค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) ด้วยการวิเคราะห์ Wilcoxon test เมื่อพิจารณาค่าความลำเอียง ปรากฏว่า วิธีการจัดคลังข้อสอบวิธี IACB และวิธี CWA มีค่าความลำเอียงเฉลี่ยไม่แตกต่างกัน เมื่อพิจารณาค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย ปรากฏว่า วิธีการจัดคลังข้อสอบวิธี IACB และวิธี CWA มีค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ยแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ .01

ตารางที่ 4-10 ผลการวิเคราะห์ความแตกต่างของค่าความลำเอียงเฉลี่ย และค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย จำแนกตามวิธีการจัดคลังข้อสอบชั้นที่ 3

รอบที่	*n*	Average Bias		RMSE	
		วิธี IACB	วิธี CWA	วิธี IACB	วิธี CWA
รอบที่ 1	500	0.000	-0.002	0.004	0.034
รอบที่ 2	500	0.001	-0.003	0.019	0.061
รอบที่ 3	500	-0.001	-0.002	0.030	0.044
รอบที่ 4	500	0.000	0.000	0.003	0.007
รอบที่ 5	500	0.000	-0.001	0.010	0.011
รอบที่ 6	500	0.000	-0.001	0.008	0.019
รอบที่ 7	500	0.002	-0.001	0.034	0.017
รอบที่ 8	500	0.001	-0.004	0.019	0.093
รอบที่ 9	500	0.001	-0.005	0.028	0.114
รอบที่ 10	500	0.000	-0.003	0.007	0.059
Mean		0.000	-0.002	0.016	0.046
SD		0.001	0.002	0.011	0.036

จากตารางที่ 4-10 แสดงค่าความลำเอียงเฉลี่ย และค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ยของวิธีการจัดคลังข้อสอบ ปรากฏว่า ค่าความลำเอียงเฉลี่ย เข้าใกล้ 0 มากที่สุด คือ วิธี IACB (Mean = 0.000) มีค่าน้อยกว่า วิธี CWA (Mean = -0.002) แสดงให้เห็นว่า วิธี IACB มีความแม่นยำในการประมาณค่าความสามารถของผู้สอบ สำหรับคลังข้อสอบในชั้นที่ 3 สูงกว่า

วิธี CWA เมื่อพิจารณาค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย ปรากฏว่า วิธี IACB (Mean = 0.016) มีค่า RMSE น้อยกว่า วิธี CWA (Mean = 0.046) แสดงให้เห็นว่า วิธี IACB มีประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบ สำหรับคลังข้อสอบในชั้นที่ 3 สูงกว่า วิธี CWA

ตารางที่ 4-11 ผลการวิเคราะห์ความแตกต่างของค่าความลำเอียงเฉลี่ย และค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย จำแนกตามวิธีการจัดคลังข้อสอบชั้นที่ 3

	วิธี	*n*	กลุ่ม	Mean Rank	Z	*p*-value
ค่าความลำเอียงเฉลี่ย	CWA	10	Control	5.00	-2.677**	<.01
	IACB	10	experiment	.00		
ค่ารากที่สองของความ	CWA	10	Control	5.00	-2.293*	<.05
คลาดเคลื่อนกำลังสองเฉลี่ย	IACB	10	experiment	5.56		

จากตารางที่ 4-11 แสดงผลการวิเคราะห์ความแตกต่างของค่ารากที่สองของความลำเอียงเฉลี่ย (Average Bias) และ ค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) ด้วยการวิเคราะห์ Wilcoxon test เมื่อพิจารณาค่าความลำเอียง ปรากฏว่า วิธีการจัดคลังข้อสอบวิธี IACB และวิธี CWA มีค่าความลำเอียงเฉลี่ยแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ .01 เมื่อพิจารณาค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย ปรากฏว่า วิธีการจัดคลังข้อสอบวิธี IACB และวิธี CWA มีค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ยแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ .01

ตารางที่ 4-12 ค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย และค่าความลำเอียงเฉลี่ย จำแนกตามวิธีการจัดคลังข้อสอบชั้นที่ 4

รอบที่	*n*	Average Bias		RMSE	
		วิธี IACB	วิธี CWA	วิธี IACB	วิธี CWA
รอบที่ 1	500	0.000	0.000	0.010	0.009
รอบที่ 2	500	0.001	-0.003	0.015	0.063
รอบที่ 3	500	-0.002	-0.001	0.034	0.040
รอบที่ 4	500	-0.001	0.002	0.013	0.055
รอบที่ 5	500	0.000	-0.003	0.002	0.062
รอบที่ 6	500	0.001	0.001	0.021	0.023
รอบที่ 7	500	0.000	-0.001	0.007	0.022
รอบที่ 8	500	0.000	-0.002	0.010	0.045
รอบที่ 9	500	-0.001	-0.003	0.015	0.063
รอบที่ 10	500	0.002	-0.003	0.037	0.064
Mean		0.000	-0.010	0.016	0.045
SD		0.001	0.002	0.011	0.020

จากตารางที่ 4-12 แสดงค่าความลำเอียงเฉลี่ย และค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ยของวิธีการจัดคลังข้อสอบ ปรากฏว่า ค่าความลำเอียงเฉลี่ย เข้าใกล้ 0 มากที่สุด คือ วิธี IACB (Mean = 0.000) มีค่าน้อยกว่า วิธี CWA (Mean = -0.001) แสดงให้เห็นว่า วิธี IACB มีความแม่นยำในการประมาณค่าความสามารถของผู้สอบ สำหรับคลังข้อสอบในชั้นที่ 4 สูงกว่า วิธี CWA เมื่อพิจารณาค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย ปรากฏว่า วิธี IACB (Mean = 0.016) มีค่า RMSE น้อยกว่า วิธี CWA (Mean = 0.045) แสดงให้เห็นว่า วิธี IACB มีประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบ สำหรับคลังข้อสอบในชั้นที่ 4 สูงกว่า วิธี CWA

ตารางที่ 4-13 ผลการวิเคราะห์ความแตกต่างของค่าความลำเอียงเฉลี่ยและค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย จำแนกตามวิธีการจัดคลังข้อสอบชั้นที่ 4

	วิธี	*n*	กลุ่ม	Mean Rank	Z	*p*-value
ค่าความลำเอียงเฉลี่ย	CWA	10	Control	4.83	-1.546	.122
	IACB	10	experiment	3.50		
ค่ารากที่สองของความ	CWA	10	Control	1.00	-2.703**	<.01
คลาดเคลื่อนกำลังสองเฉลี่ย	IACB	10	experiment	6.00		

จากตารางที่ 4-13 แสดงผลการวิเคราะห์ความแตกต่างของค่ารากที่สองของความลำเอียงเฉลี่ย (Average Bias) และ ค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) ด้วยการวิเคราะห์ Wilcoxon test เมื่อพิจารณาค่าความลำเอียง ปรากฏว่า วิธีการจัดคลังข้อสอบวิธี IACB และวิธี CWA มีค่าความลำเอียงเฉลี่ยไม่แตกต่างกัน เมื่อพิจารณาค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย ปรากฏว่า วิธีการจัดคลังข้อสอบวิธี IACB และวิธี CWA มีค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ยแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ .01

2. ผลการเปรียบเทียบประสิทธิภาพของการใช้ข้อสอบ

การเปรียบเทียบประสิทธิภาพของการใช้ข้อสอบ จากการศึกษาในสถานการณ์จำลองการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติที่มีวิธีการจัดคลังข้อสอบที่แตกต่างกัน 2 วิธี ได้แก่ วิธีการ IACB และ วิธี CWA ได้แก่ ข้อสอบที่มีการแสดงมากเกินไป ข้อสอบที่มีการใช้น้อยเกินไป อัตราการทับซ้อนของข้อสอบ และการแจกแจงอัตราการแสดงข้อสอบ

2.1 การเปรียบเทียบประสิทธิภาพจำนวนข้อสอบที่มีการแสดงมากเกินไป และข้อสอบที่มีการใช้น้อยเกินไป

จากการศึกษาในสถานการณ์จำลองการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติที่มีวิธีการจัดคลังข้อสอบที่แตกต่างกัน 2 วิธี ได้แก่ วิธี IACB และ วิธี CWA ซึ่งทั้ง 2 วิธีถูกทำซ้ำ 10 รอบ โดยใช้คลังข้อสอบจริงที่แบ่งออกเป็น 4 ชั้น โดยคลังข้อสอบชั้นที่ 1 มีข้อสอบ 18 ข้อ ชั้นที่ 2 มีข้อสอบ 49 ข้อ ชั้นที่ 3 มีข้อสอบ 64 ข้อ และ ชั้นที่ 4 มีข้อสอบ 62 ข้อ ชุดเดียวกัน ที่ทำการทดสอบแล้วนำผลการทดสอบจากสถานการณ์จำลองการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติในแต่ละรอบของแต่ละชั้นมาจำลอง แล้วนำมาคำนวณอัตราการใช้ข้อสอบที่มีการแสดงมากเกินไป อัตราการใช้ข้อสอบที่มีการใช้น้อยเกินไป ดังตารางที่ 4-14

ตารางที่ 4-14 จำนวนข้อสอบที่มีการแสดงมากเกินไป และข้อสอบที่มีการใช้น้อยเกินไป จำแนกตามวิธีการจัดคลังข้อสอบชั้นที่ 1

รอบที่	จำนวนข้อสอบที่มีการแสดงมากเกินไป (มากกว่า 0.2)		จำนวนข้อสอบที่มีการใช้น้อยเกินไป (น้อยกว่า 0.2)	
	วิธี IACB	วิธี CWA	วิธี IACB	วิธี CWA
รอบที่ 1	6	11	12	7
รอบที่ 2	7	10	11	8
รอบที่ 3	7	11	11	7
รอบที่ 4	7	10	11	8
รอบที่ 5	8	10	10	8
รอบที่ 6	6	11	12	7
รอบที่ 7	7	11	11	7
รอบที่ 8	7	11	11	7
รอบที่ 9	7	10	11	8
รอบที่ 10	7	11	11	7
Mean	6.90	10.60	11.10	7.40
SD	.568	.516	.568	.516

จากตารางที่ 4-14 แสดงจำนวนข้อสอบที่มีการแสดงมากเกินไป และข้อสอบที่มีการใช้น้อยเกินไป ของวิธีการจัดคลังข้อสอบ ปรากฏว่า จำนวนข้อสอบที่มีการแสดงมากเกินไป คือ วิธี CWA (Mean = 10.60) มีค่ามากกว่า วิธี IACB (Mean = 6.90) แสดงให้เห็นว่า วิธี IACB มีประสิทธิภาพการใช้ข้อสอบที่มีการแสดงมากเกินไป สำหรับคลังข้อสอบในชั้นที่ 1 สูงกว่า วิธี CWA เมื่อพิจารณาจำนวนข้อสอบที่มีการใช้น้อยเกินไป คือ วิธี CWA (Mean = 7.40) มีค่าน้อยกว่า วิธี IACB (Mean = 11.10) แสดงให้เห็นว่า วิธี CWA มีประสิทธิภาพการใช้จำนวนข้อสอบที่มีใช้น้อยเกินไป สำหรับคลังข้อสอบในชั้นที่ 1 สูงกว่า วิธี IACB (รายละเอียด การคำนวณค่าจำนวนข้อสอบที่มีการแสดงมากเกินไป และข้อสอบที่มีการใช้น้อยน้อยเกินไป แสดงในภาคผนวก ง1)

ตารางที่ 4-15 ผลการวิเคราะห์ความแตกต่างของจำนวนข้อสอบที่มีการแสดงมากเกินไป และข้อสอบที่มีการใช้น้อยเกินไป จำแนกตามวิธีการจัดคลังข้อสอบชั้นที่ 1

	วิธี	*n*	Mean	*SD*	χ^2	p-value
ข้อสอบที่มีการแสดงมากเกินไป	IACB	10	6.90	.568	20.00**	*p*<.01
(มากกว่า 0.2)	CWA	10	10.60	.516		
ข้อสอบที่มีการใช้น้อยเกินไป	IACB	10	11.10	.568	20.00**	*p*<.01
(น้อยกว่า 0.2)	CWA	10	7.40	.516		

จากตารางที่ 4-15 แสดงผลการวิเคราะห์ความแตกต่างของจำนวนข้อสอบที่มีการแสดงมากเกินไป และข้อสอบที่มีการใช้น้อยเกินไป ด้วยการวิเคราะห์ ไค-สแควร์ $\left(\chi^2\right)$ ปรากฏว่า วิธีการจัดคลังข้อสอบ วิธี IACB และวิธี CWA มีจำนวนข้อสอบที่มีการแสดงมากเกินไป และข้อสอบที่มีการใช้น้อยเกินไป สำหรับคลังข้อสอบในชั้นที่ 1 แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ .01

ตารางที่ 4-16 จำนวนข้อสอบที่มีการแสดงมากเกินไป และข้อสอบที่มีการใช้น้อยเกินไป จำแนกตามวิธีการจัดคลังข้อสอบชั้นที่ 2

รอบที่	จำนวนข้อสอบที่มีการแสดงมากเกินไป (มากกว่า 0.2)		จำนวนข้อสอบที่มีการใช้น้อยเกินไป (น้อยกว่า 0.2)	
	วิธี IACB	วิธี CWA	วิธี IACB	วิธี CWA
รอบที่ 1	3	5	46	44
รอบที่ 2	3	6	46	43
รอบที่ 3	4	5	45	44
รอบที่ 4	3	5	46	44
รอบที่ 5	3	4	46	45
รอบที่ 6	4	5	45	44
รอบที่ 7	4	6	45	43
รอบที่ 8	5	6	44	43
รอบที่ 9	4	5	45	44
รอบที่ 10	3	6	46	43
Mean	3.60	5.30	45.40	43.70
SD	.699	.675	.699	.675

จากตารางที่ 4-16 แสดงจำนวนข้อสอบที่มีการแสดงมากเกินไป และข้อสอบที่มีการใช้น้อยเกินไป ของวิธีการจัดคลังข้อสอบ ปรากฏว่า จำนวนข้อสอบที่มีการแสดงมากเกินไป คือ วิธี CWA (Mean = 5.30) มีค่ามากกว่า วิธี IACB (Mean = 3.60) แสดงให้เห็นว่า วิธี IACB มีประสิทธิภาพการใช้ข้อสอบที่มีการแสดงมากเกินไป สำหรับคลังข้อสอบในชั้นที่ 2 สูงกว่า วิธี CWA เมื่อพิจารณาจำนวนข้อสอบที่มีการใช้น้อยเกินไป คือ วิธี CWA (Mean = 43.70) มีค่าน้อยกว่า วิธี IACB (Mean = 45.40) แสดงให้เห็นว่า วิธี CWA มีประสิทธิภาพการใช้จำนวนข้อสอบที่มีใช้น้อยเกินไป สำหรับคลังข้อสอบในชั้นที่ 2 สูงกว่า วิธี IACB

ตารางที่ 4-17 ผลการวิเคราะห์ความแตกต่างของจำนวนข้อสอบที่มีการแสดงมากเกินไป และข้อสอบที่มีการใช้น้อยเกินไป จำแนกตามวิธีการจัดคลังข้อสอบชั้นที่ 2

	วิธี	*n*	Mean	*SD*	χ^2	p-value
ข้อสอบที่มีการแสดงมากเกินไป	IACB	10	3.60	.699	13.467**	*p*<.01
(มากกว่า 0.2)	CWA	10	5.30	.675		
ข้อสอบที่มีการใช้น้อยเกินไป	IACB	10	45.40	.699	13.467**	*p*<.01
(น้อยกว่า 0.2	CWA	10	43.70	.675		

จากตารางที่ 4-17 แสดงผลการวิเคราะห์ความแตกต่างของจำนวนข้อสอบที่มีการแสดงมากเกินไป และข้อสอบที่มีการใช้น้อยเกินไป ด้วยการวิเคราะห์ ไค-สแควร์ $\left(\chi^2\right)$ ปรากฏว่า วิธีการจัดคลังข้อสอบ วิธี IACB และวิธี CWA มีจำนวนข้อสอบที่มีการแสดงมากเกินไป และข้อสอบที่มีการใช้น้อยเกินไป สำหรับคลังข้อสอบในชั้นที่ 2 แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ .01

ตารางที่ 4-18 จำนวนข้อสอบที่มีการแสดงมากเกินไป และข้อสอบที่มีการใช้น้อยเกินไป จำแนกตามวิธีการจัดคลังข้อสอบชั้นที่ 3

รอบที่	จำนวนข้อสอบที่มีการแสดงมากเกินไป (มากกว่า 0.2)		จำนวนข้อสอบที่มีการใช้น้อยเกินไป (น้อยกว่า 0.2)	
	วิธี IACB	วิธี CWA	วิธี IACB	วิธี CWA
รอบที่ 1	1	4	63	60
รอบที่ 2	1	3	63	61
รอบที่ 3	2	4	62	60
รอบที่ 4	1	3	63	61
รอบที่ 5	1	4	63	60
รอบที่ 6	1	2	63	62
รอบที่ 7	2	4	62	60
รอบที่ 8	1	4	63	64
รอบที่ 9	1	2	63	62
รอบที่ 10	3	4	61	60
Mean	1.400	3.400	62.600	61.00
SD	.699	.843	.699	.843

จากตารางที่ 4-18 แสดงจำนวนข้อสอบที่มีการแสดงมากเกินไป และข้อสอบที่มีการใช้น้อยเกินไป ของวิธีการจัดคลังข้อสอบ ปรากฏว่า จำนวนข้อสอบที่มีการแสดงมากเกินไป คือ วิธี CWA (Mean = 3.40) มีค่ามากกว่า วิธี IACB (Mean = 1.400) แสดงให้เห็นว่า วิธี IACB มีประสิทธิภาพการใช้ข้อสอบที่มีการแสดงมากเกินไป สำหรับคลังข้อสอบในชั้นที่ 3 สูงกว่า วิธี CWA เมื่อพิจารณาจำนวนข้อสอบที่มีการใช้น้อยเกินไป คือ วิธี CWA (Mean = 61.00) มีค่าน้อยกว่า วิธี

IACB (Mean = 62.000) แสดงให้เห็นว่า วิธี CWA มีประสิทธิภาพการใช้จำนวนข้อสอบที่มีใช้น้อยเกินไป สำหรับคลังข้อสอบในชั้นที่ 3 สูงกว่า วิธี IACB

ตารางที่ 4-19 ผลการวิเคราะห์ความแตกต่างของจำนวนข้อสอบที่มีการแสดงมากเกินไป และข้อสอบที่มีการใช้น้อยเกินไป จำแนกตามวิธีการจัดคลังข้อสอบชั้นที่ 3

	วิธี	*n*	Mean	*SD*	χ^2	p-value
ข้อสอบที่มีการแสดงมาก	IACB	10	1.40	.699	13.333**	*p*<.01
เกินไป (มากกว่า 0.2)	CWA	10	3.40	.843		
ข้อสอบที่มีการใช้น้อย	IACB	10	62.60	.699	13.467**	*p*<.01
(น้อยกว่า 0.2)	CWA	10	61.00	.843		

จากตารางที่ 4-19 แสดงผลการวิเคราะห์ความแตกต่างของจำนวนข้อสอบที่มีการแสดงมากเกินไป และข้อสอบที่มีการใช้น้อยเกินไป ด้วยการวิเคราะห์ ไค-สแควร์ $\left(\chi^2\right)$ ปรากฏว่า วิธีการจัดคลังข้อสอบ วิธี IACB และวิธี CWA มีจำนวนข้อสอบที่มีการแสดงมากเกินไป และข้อสอบที่มีการใช้น้อยเกินไปสำหรับคลังข้อสอบในชั้นที่ 3 แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ .01

ตารางที่ 4-20 จำนวนข้อสอบที่มีการแสดงมากเกินไป และข้อสอบที่มีการใช้น้อยเกินไป จำแนกตามวิธีการจัดคลังข้อสอบชั้นที่ 4

รอบที่	จำนวนข้อสอบที่มีการแสดงมากเกินไป (มากกว่า 0.2)		จำนวนข้อสอบที่มีการใช้น้อยเกินไป (น้อยกว่า 0.2)	
	วิธี IACB	วิธี CWA	วิธี IACB	วิธี CWA
รอบที่ 1	3	5	59	57
รอบที่ 2	2	6	60	56
รอบที่ 3	4	5	58	57
รอบที่ 4	2	6	60	56
รอบที่ 5	3	5	59	57
รอบที่ 6	2	5	60	57
รอบที่ 7	4	5	58	57
รอบที่ 8	3	6	59	56
รอบที่ 9	4	4	58	58
รอบที่ 10	2	5	60	57
Mean	2.90	5.20	59.10	56.80
SD	.876	.632	.876	.632

จากตารางที่ 4-20 แสดงจำนวนข้อสอบที่มีการแสดงมากเกินไป และข้อสอบที่มีการใช้น้อยเกินไป ของวิธีการจัดคลังข้อสอบ ปรากฏว่า จำนวนข้อสอบที่มีการแสดงมากเกินไป คือ วิธี CWA (Mean = 5.20) มีค่ามากกว่า วิธี IACB (Mean = 2.90) แสดงให้เห็นว่า วิธี IACB มี

ประสิทธิภาพการใช้ข้อสอบที่มีการแสดงมากเกินไป สำหรับคลังข้อสอบในชั้นที่ 4 สูงกว่า วิธี CWA เมื่อพิจารณาจำนวนข้อสอบที่มีการใช้น้อยเกินไป คือ วิธี CWA (Mean = 56.80) มีค่าน้อยกว่า วิธี IACB (Mean = 59.10) แสดงให้เห็นว่า วิธี CWA มีประสิทธิภาพการใช้จำนวนข้อสอบที่มีใช้น้อยเกินไป สำหรับคลังข้อสอบในชั้นที่ 4 สูงกว่า วิธี IACB

ตารางที่ 4-21 ผลการวิเคราะห์ความแตกต่างของจำนวนข้อสอบที่มีการแสดงมากเกินไป และข้อสอบที่มีการใช้น้อยเกินไป จำแนกตามวิธีการจัดคลังข้อสอบชั้นที่ 4

	วิธี	*n*	Mean	*SD*	χ^2	p-value
ข้อสอบที่มีการแสดงมาก	IACB	10	2.90	.876	17.00**	*p*<.01
เกินไป (มากกว่า 0.2)	CWA	10	5.20	.632		
ข้อสอบที่มีการใช้น้อยเกินไป	IACB	10	59.10	.876	17.00**	*p*<.01
(น้อยกว่า 0.2)	CWA	10	56.80	.632		

จากตารางที่ 4-21 แสดงผลการวิเคราะห์ความแตกต่างของจำนวนข้อสอบที่มีการแสดงมากเกินไป และข้อสอบที่มีการใช้น้อยเกินไป ด้วยการวิเคราะห์ ไค-สแควร์ $\left(\chi^2\right)$ ปรากฏว่า วิธีการจัดคลังข้อสอบ วิธี IACB และวิธี CWA มีจำนวนข้อสอบที่มีการแสดงมากเกินไป และข้อสอบที่มีการใช้น้อยเกินไป สำหรับคลังข้อสอบในชั้นที่ 4 แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ .01

2.2 ผลการเปรียบเทียบประสิทธิภาพด้านจำนวนข้อสอบที่มีอัตราการทับซ้อนของข้อสอบ และการแจกแจงอัตราการแสดงข้อสอบ

ตารางที่ 4-22 จำนวนอัตราการทับซ้อนของข้อสอบ และการแจกแจงอัตราการแสดงข้อสอบ จำแนกตามวิธีการจัดคลังข้อสอบชั้นที่ 1

รอบที่	อัตราการทับซ้อนของข้อสอบ (ร้อยละ)		การแจกแจงอัตราการแสดงข้อสอบ		
	วิธี IACB	วิธี CWA	วิธี IACB $\left(\chi_1^2\right)$	วิธี CWA $\left(\chi_2^2\right)$	*F-ratio*
รอบที่ 1	98.80	84.20	1666.240	216.509	7.696
รอบที่ 2	98.00	82.40	1611.711	272.818	5.908
รอบที่ 3	98.40	81.80	1553.196	256.473	6.056
รอบที่ 4	98.80	82.40	1666.240	233.991	7.121
รอบที่ 5	95.20	82.60	1355.920	277.691	4.883
รอบที่ 6	98.80	78.20	1666.240	210.087	7.931
รอบที่ 7	97.20	83.60	1511.04	235.942	6.404
รอบที่ 8	97.40	81.40	1511.54	184.669	8.185
รอบที่ 9	95.20	83.40	1386.18	232.647	5.958
รอบที่ 10	95.80	82.80	1395.22	221.649	6.295
Mean	97.36	82.28	1532.353	234.288	6.540
SD	1.47	1.658	121.425	28.654	

จากตารางที่ 4-22 แสดงจำนวนอัตราการทับซ้อนของข้อสอบ และ การแจกแจงอัตราการแสดงข้อสอบ ของวิธีการจัดคลังข้อสอบ ปรากฏว่า จำนวนอัตราการทับซ้อนของข้อสอบมากสุด คือ วิธี IACB (Mean = 97.36) มีค่ามากกว่า วิธี CWA (Mean = 82.28) แสดงให้เห็นว่า วิธี CWA มีประสิทธิภาพด้านจำนวนข้อสอบที่มีอัตราการทับซ้อนของข้อสอบ สำหรับคลังข้อสอบในชั้นที่ 1 สูงกว่า วิธี IACB เมื่อพิจารณาการแจกแจงอัตราการแสดงข้อสอบ อัตราส่วน F (F - ratio) พิจารณาจากอัตราส่วนของค่า χ_1^2 วิธี IACB หารด้วย χ_2^2 วิธี CWA ปรากฏว่า ค่าF_{ratio} >1 แสดงว่า วิธี CWA มีประสิทธิภาพด้านจำนวนข้อสอบที่มีการแจกแจงอัตราการแสดงข้อสอบ สำหรับคลังข้อสอบในชั้นที่ 1 สูงกว่า วิธี IACB (รายละเอียดการคำนวณค่าจำนวนอัตราการทับซ้อนของข้อสอบและการแจกแจงอัตราการแสดงข้อสอบ แสดงในภาคผนวก ง2-ง3)

ตารางที่ 4-23 ผลการวิเคราะห์ความแตกต่างของจำนวนอัตราการทับซ้อนของข้อสอบ และการแจกแจงอัตราการแสดงข้อสอบ จำแนกตามวิธีการจัดคลังข้อสอบชั้นที่ 1

	วิธี	*n*	Mean	*SD*	χ^2	*p*-value
อัตราการทับซ้อนของข้อสอบ	IACB	10	97.36	1.47	20.00	.172
	CWA	10	82.28	1.68		
การแจกแจงอัตราการแสดงข้อสอบ	IACB	10	1532.353	234.288	20.00	.274
	CWA	10	234.288	28.654		

จากตารางที่ 4-23 แสดงผลการวิเคราะห์ความแตกต่างของจำนวนอัตราการทับซ้อนของข้อสอบ และการแจกแจงอัตราการแสดงข้อสอบ ด้วยการวิเคราะห์ ไค-สแควร์$\left(\chi^2\right)$ ปรากฏว่า วิธีการจัดคลังข้อสอบวิธี IACB และวิธี CWA มีจำนวนอัตราการทับซ้อนของข้อสอบ และการแจกแจงอัตราการแสดงข้อสอบ สำหรับคลังข้อสอบในชั้นที่ 1 ไม่แตกต่างกัน

ตารางที่ 4-24 จำนวนอัตราการทับซ้อนของข้อสอบ และการแจกแจงอัตราการแสดงข้อสอบ จำแนกตามวิธีการจัดคลังข้อสอบชั้นที่ 2

รอบที่	อัตราการทับซ้อนของข้อสอบ (ร้อยละ)		การแจกแจงอัตราการแสดงข้อสอบ		
	วิธี IACB	วิธี CWA	วิธี IACB $\left(\chi_1^2\right)$	วิธี CWA$\left(\chi_2^2\right)$	F_{ratio}
รอบที่ 1	42.60	52.60	173.843	526.489	0.330
รอบที่ 2	42.00	56.20	177.330	671.026	0.264
รอบที่ 3	42.60	43.80	170.904	325.549	0.525
รอบที่ 4	41.60	44.80	177.885	349.519	0.509
รอบที่ 5	41.80	42.40	180.997	301.399	0.601
รอบที่ 6	41.80	39.20	184.917	327.726	0.564
รอบที่ 7	42.00	44.20	179.209	324.096	0.553

ตารางที่ 4-24 (ต่อ)

รอบที่	อัตราการทับซ้อนของข้อสอบ (ร้อยละ)		การแจกแจงอัตราการแสดงข้อสอบ		
	วิธี IACB	วิธี CWA	วิธี IACB $\left(\chi_1^2\right)$	วิธี CWA $\left(\chi_2^2\right)$	F_{ratio}
รอบที่ 8	40.40	40.60	186.656	333.338	0.560
รอบที่ 9	40.20	38.80	186.484	268.011	0.696
รอบที่ 10	41.40	44.60	184.574	365.109	0.506
Mean	41.64	44.72	180.226	379.226	0.475
SD	.804	5.607	5.421	123.387	

จากตารางที่ 4-24 แสดงจำนวนอัตราการทับซ้อนของข้อสอบ และการแจกแจงอัตราการแสดงข้อสอบ ของวิธีการจัดคลังข้อสอบ ปรากฏว่า จำนวนอัตราการทับซ้อนของข้อสอบมากสุด คือ วิธี CWA (Mean = 44.72) มีค่ามากกว่า วิธี IACB (Mean = 41.64) แสดงให้เห็นว่า วิธี IACB มีประสิทธิภาพด้านจำนวนข้อสอบที่มีอัตราการทับซ้อนของข้อสอบ สำหรับคลังข้อสอบในชั้นที่ 2 สูงกว่า วิธี CWA เมื่อพิจารณาการแจกแจงอัตราการแสดงข้อสอบ อัตราส่วน F (F - ratio) พิจารณาจากอัตราส่วนของค่า χ_1^2 วิธี IACB หารด้วย χ_2^2 วิธี CWA ปรากฏว่า ค่า F_{ratio} <1 แสดงว่า วิธี IACB มีประสิทธิภาพด้านจำนวนข้อสอบที่มีการแจกแจงอัตราการแสดงข้อสอบ สำหรับคลังข้อสอบในชั้นที่ 2 สูงกว่าวิธี CWA

ตารางที่ 4-25 ผลการวิเคราะห์ความแตกต่างของจำนวนอัตราการทับซ้อนของข้อสอบ และการแจกแจงอัตราการแสดงข้อสอบ จำแนกตามวิธีการจัดคลังข้อสอบชั้นที่ 2

	วิธี	*n*	Mean	*SD*	χ^2	*p*-value
อัตราการทับซ้อนของข้อสอบ	IACB	10	41.64	.804	20.00	.220
	CWA	10	44.72	5.607		
การแจกแจงอัตราการแสดงข้อสอบ	IACB	10	180.226	5.421	20.00	.395
	CWA	10	379.226	123.387		

จากตารางที่ 4-25 แสดงผลการวิเคราะห์ความแตกต่างของจำนวนอัตราการทับซ้อนของข้อสอบ และการแจกแจงอัตราการแสดงข้อสอบ ด้วยการวิเคราะห์ ไค-สแควร์ $\left(\chi^2\right)$ ปรากฏว่า วิธีการจัดคลังข้อสอบวิธี IACB และวิธี CWA มีจำนวนอัตราการทับซ้อนของข้อสอบ และการแจกแจงอัตราการแสดงข้อสอบ สำหรับคลังข้อสอบในชั้นที่ 2 ไม่แตกต่างกัน

ตารางที่ 4-26 จำนวนอัตราการทับซ้อนของข้อสอบ และ การแจกแจงอัตราการแสดงข้อสอบ จำแนกตามวิธีการจัดคลังข้อสอบชั้นที่ 3

รอบที่	อัตราการทับซ้อนของข้อสอบ (ร้อยละ)		การแจกแจงอัตราการแสดงข้อสอบ		
	วิธี IACB	วิธี CWA	วิธี IACB $\left(\chi_1^2\right)$	วิธี CWA $\left(\chi_2^2\right)$	F_{ratio}
รอบที่ 1	18.80	59.20	94.493	700.400	0.135
รอบที่ 2	17.80	55.20	82.990	526.828	0.158
รอบที่ 3	18.60	59.00	96.750	720.828	0.134
รอบที่ 4	17.60	54.20	109.255	647.945	0.169
รอบที่ 5	18.80	55.80	94.493	652.738	0.145
รอบที่ 6	18.20	56.40	101.265	792.353	0.128
รอบที่ 7	17.00	59.40	105.85	693.210	0.153
รอบที่ 8	17.60	54.20	101.638	590.408	0.172
รอบที่ 9	17.00	57.80	113.468	992.558	0.114
รอบที่ 10	17.00	57.60	119.930	790.078	0.152
Mean	17.84	56.88	102.013	710.735	0.144
SD	.729	2.005	10.660	128.454	

จากตารางที่ 4-26 แสดงจำนวนอัตราการทับซ้อนของข้อสอบ และ การแจกแจงอัตราการแสดงข้อสอบ ของวิธีการจัดคลังข้อสอบ ปรากฏว่า จำนวนอัตราการทับซ้อนของข้อสอบมากสุด คือ วิธี CWA (Mean = 56.88) มีค่ามากกว่า วิธี IACB (Mean = 17.84) แสดงให้เห็นว่า วิธี IACB มีประสิทธิภาพด้านจำนวนข้อสอบที่มีอัตราการทับซ้อนของข้อสอบ สำหรับคลังข้อสอบในชั้นที่ 3 สูงกว่า วิธี CWA เมื่อพิจารณาการแจกแจงอัตราการแสดงข้อสอบ อัตราส่วน F (F - ratio) พิจารณาจากอัตราส่วนของค่า χ_1^2 วิธี IACB หารด้วย χ_2^2 วิธี CWA ปรากฏว่า ค่า F_{ratio} <1 แสดงว่า วิธี IACB มีประสิทธิภาพด้านจำนวนข้อสอบที่มีการแจกแจงอัตราการแสดงข้อสอบ สำหรับคลังข้อสอบในชั้นที่ 3 สูงกว่าวิธี CWA

ตารางที่ 4-27 ผลการวิเคราะห์ความแตกต่างของจำนวนอัตราการทับซ้อนของข้อสอบ และ การแจกแจงอัตราการแสดงข้อสอบ จำแนกตามวิธีการจัดคลังข้อสอบชั้นที่ 3

	วิธี	*n*	Mean	*SD*	χ^2	*p*-value
อัตราการทับซ้อนของข้อสอบ	IACB	10	17.84	.729	20.00	.130
	CWA	10	56.88	2.005		
การแจกแจงอัตราการแสดงข้อสอบ	IACB	10	102.013	10.660	20.00	.333
	CWA	10	710.735	128.454		

จากตารางที่ 4-27 แสดงผลการวิเคราะห์ความแตกต่างของจำนวนอัตราการทับซ้อนของข้อสอบ และการแจกแจงอัตราการแสดงข้อสอบ ด้วยการวิเคราะห์ ไค-สแควร์ $\left(\chi^2\right)$ ปรากฏว่า วิธีการจัดคลังข้อสอบ วิธี IACB และวิธี CWA มีจำนวนอัตราการทับซ้อนของข้อสอบ และการแจกแจงอัตราการแสดงข้อสอบ สำหรับคลังข้อสอบในชั้นที่ 3 ไม่แตกต่างกัน

ตารางที่ 4-28 จำนวนอัตราการทับซ้อนของข้อสอบ และ การแจกแจงอัตราการแสดงข้อสอบ จำแนกตามวิธีการจัดคลังข้อสอบชั้นที่ 4

รอบที่	อัตราการทับซ้อนของข้อสอบ (ร้อยละ)		การแจกแจงอัตราการแสดงข้อสอบ		
	วิธี IACB	วิธี CWA	วิธี IACB $\left(\chi_1^2\right)$	วิธี CWA $\left(\chi_2^2\right)$	F_{ratio}
รอบที่ 1	17.20	59.00	87.221	642.669	0.136
รอบที่ 2	15.60	52.60	71.692	464.445	0.154
รอบที่ 3	16.80	58.80	92.852	619.018	0.150
รอบที่ 4	16.60	57.00	91.327	521.752	0.175
รอบที่ 5	15.20	58.60	101.736	524.752	0 .194
รอบที่ 6	14.60	56.80	113.347	503.061	0.225
รอบที่ 7	15.80	62.00	102.589	664.892	0.154
รอบที่ 8	17.20	61.60	87.221	625.030	0.140
รอบที่ 9	16.60	57.60	95.047	585.265	0.162
รอบที่ 10	16.00	59.00	92.954	531.585	0.175
Mean	16.16	58.30	93.599	568.240	0.165
SD	.868	2.642	11.088	97.717	

จากตารางที่ 4-28 แสดงจำนวนอัตราการทับซ้อนของข้อสอบ และการแจกแจงอัตราการแสดงข้อสอบ ของวิธีการจัดคลังข้อสอบ ปรากฏว่า จำนวนอัตราการทับซ้อนของข้อสอบมากสุด คือ วิธี CWA (Mean 58.30) มีค่ามากกว่า วิธี IACB (Mean = 16.16) แสดงให้เห็นว่า วิธี IACB มีประสิทธิภาพด้านจำนวนข้อสอบที่มีอัตราการทับซ้อนของข้อสอบ สำหรับคลังข้อสอบในชั้นที่ 3 สูงกว่า วิธี CWA เมื่อพิจารณาการแจกแจงอัตราการแสดงข้อสอบ อัตราส่วน F (F - ratio) พิจารณาจากอัตราส่วนของค่า χ_1^2 วิธี IACB หารด้วย χ_2^2 วิธี CWA ปรากฏว่า ค่า F_{ratio} <1 แสดงว่า วิธี IACB มีประสิทธิภาพด้านจำนวนข้อสอบที่มีการแจกแจงอัตราการแสดงข้อสอบ สำหรับคลังข้อสอบในชั้นที่ 4 สูงกว่าวิธี CWA

ตารางที่ 4-29 ผลการวิเคราะห์ความแตกต่างของจำนวนอัตราการทับซ้อนของข้อสอบ และการแจกแจงอัตราการแสดงข้อสอบ จำแนกตามวิธีการจัดคลังข้อสอบชั้นที่ 4

	วิธี	*n*	Mean	*SD*	χ^2	*p*-value
อัตราการทับซ้อนของข้อสอบ	IACB	10	16.16	.868	20.00	.220
	CWA	10	58.30	2.642		
การแจกแจงอัตราการแสดงข้อสอบ	IACB	10	93.599	11.088	20.00	.333
	CWA	10	568.240	97.717		

จากตารางที่ 4-29 แสดงผลการวิเคราะห์ความแตกต่างของจำนวนอัตราการทับซ้อนของข้อสอบ และการแจกแจงอัตราการแสดงข้อสอบ ด้วยการวิเคราะห์ ไค-สแควร์ $\left(\chi^2\right)$ ปรากฏว่า วิธีการจัดคลังข้อสอบ วิธี IACB และวิธี CWA มีจำนวนอัตราการทับซ้อนของข้อสอบ และการแจกแจงอัตราการแสดงข้อสอบ สำหรับคลังข้อสอบในชั้นที่ 4 ไม่แตกต่างกัน

สรุปผลเปรียบเทียบประสิทธิภาพของวิธีการจัดคลังข้อสอบ 2 วิธี โดยใช้เกณฑ์การพิจารณาตามตารางที่ 3-4 สรุปเกณฑ์การเปรียบเทียบประสิทธิภาพด้านการประมาณค่าความสามารถของผู้สอบและด้านประสิทธิภาพการใช้ข้อสอบ (หน้า 114) ดังตารางที่ 4-30

ตารางที่ 4-30 สรุปผลการเปรียบเทียบประสิทธิภาพของวิธีการจัดคลังข้อสอบด้านการประมาณค่าความสามารถของผู้สอบ และด้านประสิทธิภาพการใช้ข้อสอบ

คลังข้อสอบชั้นที่	ตัวแปร	ผลการเปรียบเทียบประสิทธิภาพวิธีการจัดคลังข้อสอบ	
		วิธี IACB	วิธี CWA
ด้านการประมาณค่าความสามารถของผู้สอบ			
ชั้นที่ 1	ค่าความลำเอียงเฉลี่ย	✓(0.000)	(0.003)
	ค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย	✓(0.018)	(0.056)
ชั้นที่ 2	ค่าความลำเอียงเฉลี่ย	✓(0.000)	(0.001)
	ค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย	✓(0.024)	(0.065)
ชั้นที่ 3	ค่าความลำเอียงเฉลี่ย	✓(0.000)	(0.002)
	ค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย	✓(0.016)	(0.046)
ชั้นที่ 4	ค่าความลำเอียงเฉลี่ย	✓(0.000)	(0.010)
	ค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย	✓(0.016)	(0.045
	รวม	**8**	**0**
ด้านประสิทธิภาพของการใช้ข้อสอบ			
ชั้นที่ 1	ข้อสอบที่มีการแสดงมากเกินไป	✓(6.90)	(10.60)
	ข้อสอบที่มีการใช้น้อยเกินไป	(11.10)	✓(7.40)
	อัตราการทับซ้อนของข้อสอบ	(97.36)	✓(82.28)
	การแจกแจงอัตราการแสดงข้อสอบ	(1532.35)	✓(234.29)

ตารางที่ 4-30 (ต่อ)

คลังข้อสอบชั้นที่	ตัวแปร	ผลการเปรียบเทียบประสิทธิภาพวิธีการจัดคลังข้อสอบ	
		วิธี IACB	วิธี CWA
ชั้นที่ 2	ข้อสอบที่มีการแสดงมากเกินไป	✓(3.60)	(5.30)
	ข้อสอบที่มีการใช้น้อยเกินไป	(45.40)	✓(43.70)
	อัตราการทับซ้อนของข้อสอบ	✓(41.64)	(44.72)
	การแจกแจงอัตราการแสดงข้อสอบ	✓(180.23)	(379.23)
ชั้นที่ 3	ข้อสอบที่มีการแสดงมากเกินไป	✓(1.40)	(3.40)
	ข้อสอบที่มีการใช้น้อยเกินไป	(62.60)	✓(61.00)
	อัตราการทับซ้อนของข้อสอบ	✓(17.84)	(56.88)
	การแจกแจงอัตราการแสดงข้อสอบ	✓(102.01)	(710.74)
ชั้นที่ 4	ข้อสอบที่มีการแสดงมากเกินไป	✓(2.90)	(5.20)
	ข้อสอบที่มีการใช้น้อยเกินไป	(59.10)	✓(56.80)
	อัตราการทับซ้อนของข้อสอบ	✓(16.16)	(58.30)
	การแจกแจงอัตราการแสดงข้อสอบ	✓(93.60)	(568.24)
	รวม	**10**	**6**

จากตารางที่ 30 ผลการเปรียบเทียบประสิทธิภาพของวิธีการจัดคลังข้อสอบวิธี IACB กับวิธี CWA ในภาพรวม ปรากฏว่า วิธี IACB มีประสิทธิภาพสูงกว่า วิธี CWA เมื่อพิจารณาประสิทธิภาพรายด้านปรากฏว่า

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

1.1 ค่าความลำเอียงเฉลี่ย ปรากฏว่า วิธี IACB มีค่าน้อย 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 มีค่าน้อย 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 มีค่าน้อย วิธี CWA (4:0) แสดงให้เห็นว่า วิธี IACB มีประสิทธิภาพด้านการใช้ข้อสอบที่มีการแสดงมากเกินไปสูงกว่า วิธี CWA ในทุกชั้นของคลังข้อสอบอย่างมีนัยสำคัญทางสถิติที่ .01 เป็นไปตามสมมติฐานข้อที่ 2

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

2.3 อัตราการทับซ้อนของข้อสอบ ปรากฏว่า วิธี IACB มีค่าน้อย วิธี CWA (3:1) แสดงให้เห็นว่า วิธี IACB ประสิทธิภาพด้านจำนวนข้อสอบที่มีอัตราการทับซ้อนข้อสอบสูงกว่า วิธี CWA ในบางชั้นของคลังข้อสอบ ไม่แตกต่างกัน เป็นไปตามสมมติฐานข้อที่ 2

2.4 การแจกแจงอัตราการแสดงข้อสอบ วิธี IACB มีค่าน้อย วิธี CWA (3:1) แสดงให้เห็นว่า วิธี IACB มีประสิทธิภาพด้านจำนวนข้อสอบที่มีการแจกแจงอัตราการแสดงข้อสอบสูงกว่า วิธี CWA ในบางชั้นของคลังข้อสอบ ไม่แตกต่างกัน เป็นไปตามสมมติฐานข้อที่ 2

ตอนที่ 3 การพัฒนาโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติโดยใช้คลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา

ผลการพัฒนาโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติในการวิจัยนี้ เป็นการนำวิธีการจัดคลังข้อสอบตามระดับชั้นร่วมกับการจัดสมดุลเนื้อหาแบบใหม่ มาพัฒนาโปรแกรมการทดสอบสำหรับ O-NET วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 6 ระหว่างปีการศึกษา 2551–2553 และ2555–2559 ในรูปแบบของ Web Application การนำเสนอผลการพัฒนาโปรแกรม แบ่งเป็น 2 ประเด็น ดังนี้

1. ผลการพัฒนาโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ ได้แก่ การจัดการสอบ และการรายงานผลการทดสอบ

2. ผลการประเมินความเหมาะสมของโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ ได้แก่ ผลการประเมินความเหมาะสมโดยผู้เชี่ยวชาญ และผลการประเมินความเหมาะสมโดยผู้ใช้งาน

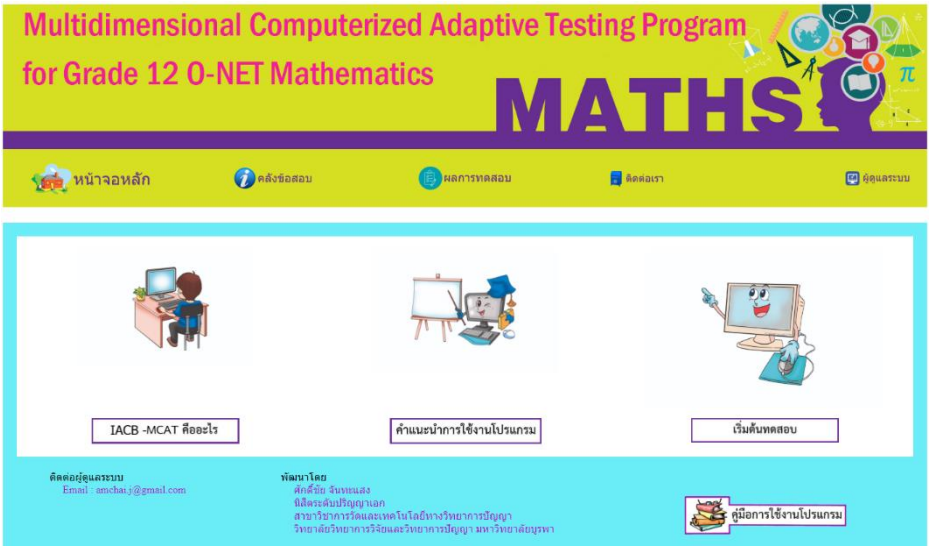
ในแต่ละประเด็นมีรายละเอียดดังนี้

1. ผลการพัฒนาโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ

โปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ สำหรับทดสอบ O-NET วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 6 นี้ การพัฒนาคลังข้อสอบเริ่มจากแบ่งชั้นคลังข้อสอบเป็น 4 ชั้นตามระดับค่าอำนาจจำแนกของข้อสอบ 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

การพัฒนาคลังข้อสอบเริ่มจากแบ่งชั้นคลังข้อสอบเป็น 4 ชั้นตามระดับค่าอำนาจจำแนกข้อสอบร่วมกับการจัดสมดุลเนื้อหาแบบใหม่ในแต่ละชั้น โดยในแต่ละชั้นได้จัดกลุ่มเนื้อหาข้อสอบ O-NET วิชาคณิตศาสตร์ โดยจำแนกเป็น 4 กลุ่มสาระการเรียนรู้ ได้แก่ จำนวนและการดำเนินการ

การวัด พีชคณิต และการวิเคราะห์ข้อมูลและความน่าจะเป็น ผู้ใช้งานเข้าใช้โปรแกรมได้จาก http://mcatonetmath.ddns.net:8080/MCATWebApp/ แสดงดังภาพที่ 4-2



ภาพที่ 4-2 หน้าจอหลักของโปรแกรมการทดสอบ

จากภาพที่ 4-2 แสดงหน้าจอหลักของโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ หน้าหลักของโปรแกรมประกอบด้วยเมนูหลัก 5 เมนู ดังนี้

1) หน้าจอหลัก เป็นเมนูที่แสดงหน้าจอหลักของโปรแกรมการทดสอบ เมื่อกดปุ่มนี้ “หน้าจอหลัก” จะกลับมาหน้าจอหลักของโปรแกรมการทดสอบ ซึ่งแบ่งเนื้อหาเป็น 4 ส่วนย่อย ได้แก่ IACB-MCAT คืออะไร คำแนะนำการใช้โปรแกรม เริ่มต้นทดสอบ และคู่มือการใช้งานโปรแกรม

2) คลังข้อสอบ เป็นเมนูแสดงหน้าคลังข้อสอบ เมื่อกดปุ่มนี้ “คลังข้อสอบ” จะปรากฏลิงค์ไปยังหน้าคลังข้อสอบ ซึ่งแบ่งเนื้อหาเป็น 3 ส่วยย่อยได้แก่ ความหมายคลังข้อสอบ รูปแบบการจัดคลังข้อสอบ และเกณฑ์การประเมินความสามารถของผู้สอบ

3) ผลการทดสอบ เป็นเมนูสำหรับการค้นหาผลการทดสอบที่ผ่านมาของผู้ใช้งาน เมื่อกดปุ่มนี้ “ผลการทดสอบ” จะปรากฏลิงค์ไปยังหน้าเริ่มต้นเข้าสู่ระบบแสดงผลการทดสอบ ซึ่งผู้ใช้งานต้องใส่ข้อมูลเลขประจำตัวประชาชน กับ รหัสผ่าน เพื่อค้นหาผลการทดสอบ

4) ติดต่อเรา เป็นเมนูสำหรับแสดงข้อมูลที่เกี่ยวข้องกับผู้วิจัยและสถาบันของผู้วิจัย เมื่อกดปุ่มนี้ “ติดต่อเรา” จะปรากฏลิงค์ไปยังหน้าแสดงช่องทางการติดต่อผู้วิจัย

5) ผู้ดูแลระบบ เป็นเมนูสำหรับเข้าสู่ระบบของผู้ดูแลระบบ เมื่อกดปุ่มนี้ ”ผู้ดูแลระบบ” จะปรากฏลิงค์ไปยังหน้าเริ่มต้นเข้าสู่ระบบสำหรับผู้ดูแลระบบ ผู้ดูแลระบบต้องใส่ข้อมูลชื่อผู้ใช้งานกับ รหัสผ่านถูกต้องแล้ว แล้วกดเข้าสู่ระบบ

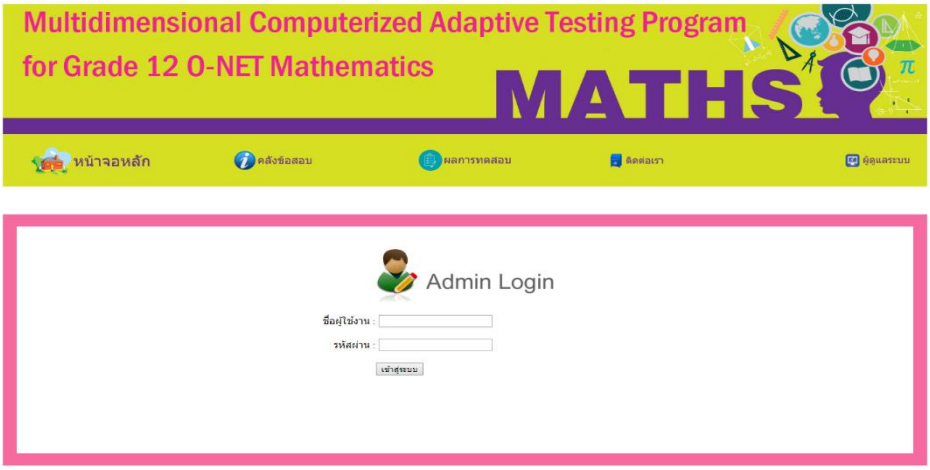
รายละเอียดของโปรแกรมการทดสอบนี้ แบ่งได้ 2 ส่วน ดังนี้

1.1 การจัดการสอบ

การจัดการสอบนี้ เป็นส่วนที่ให้นักเรียนเข้าใช้โปรแกรมเพื่อทดสอบข้อสอบ O-NET วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 6 ซึ่งการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ มีขั้นตอนการทดสอบดังต่อไปนี้

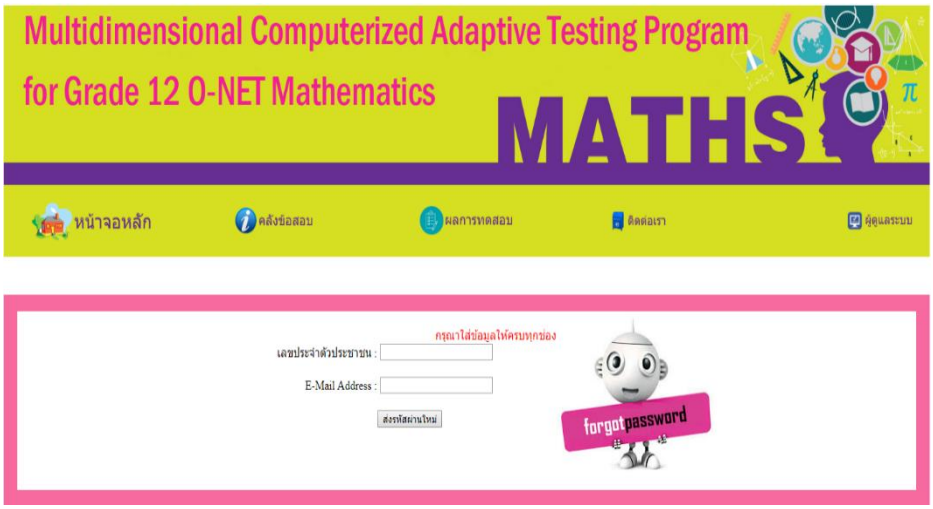
1.1.1 เริ่มต้นการทดสอบ

เมื่อผู้สอบต้องการเริ่มต้นการทดสอบ ให้กดปุ่มนี้ “ เริ่มต้นทดสอบ ” จะปรากฏลิงค์ไปยังหน้าเริ่มต้นเข้าสู่ระบบทดสอบ แสดงได้ดังภาพที่ 4-3



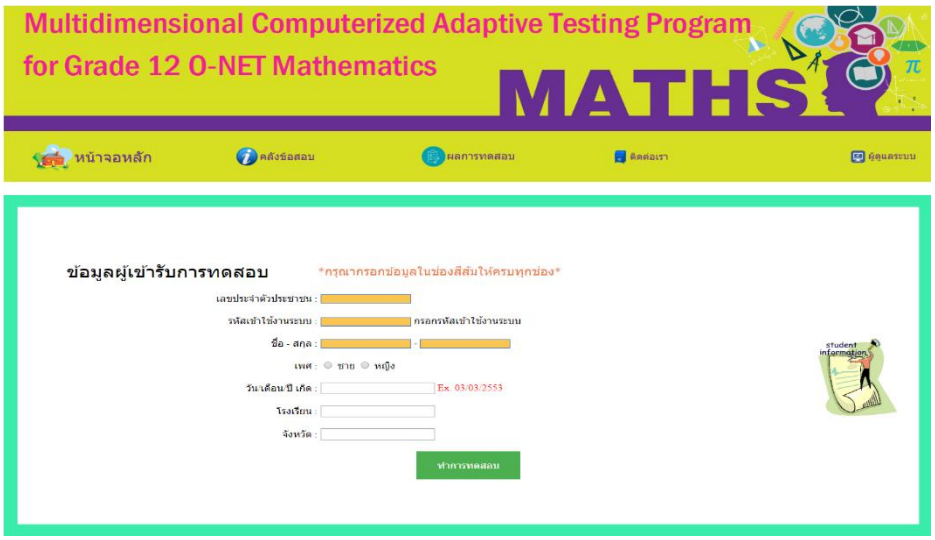
ภาพที่ 4-3 หน้าเริ่มต้นเข้าสู่ระบบทดสอบ

จากภาพที่ 4-3 สำหรับผู้เริ่มต้นใหม่ ให้กดที่ข้อความ “ลงทะเบียนเพื่อเข้าใช้งานระบบ” ถ้าใครเคยลงทะเบียนแล้ว สามารถใส่เลขประจำตัวประชาชนกับและรหัสผ่านเพื่อเข้าสู่หน้าทำการทดสอบได้เลยและกรณีจำรหัสผ่านไม่ได้ สามารถกดปุ่มลืมรหัสให้กดข้อความ “ ลืมรหัสผ่าน? ” จะปรากฏลิงค์ไปยังหน้าจอลงทะเบียนเพื่อเข้าใช้งานในระบบ ดังภาพที่ 4-4 แต่สำหรับคนที่ยังไม่ลงทะเบียน ให้กดปุ่ม “ ลงทะเบียนเพื่อเข้าใช้งานระบบ ”จะปรากฏลิงค์ไปยังหน้าจอลงทะเบียนเพื่อเข้าใช้งานในระบบ แสดงดังภาพที่ 4-4



ภาพที่ 4-4 หน้าจอลืมรหัสผ่าน

จากภาพที่ 4-4 กรอกข้อมูลให้ครบทุกช่อง เลขประจำตัวประชาชน และ E-Mail Address หลักจากนั้นให้กดปุ่ม “ ส่งรหัสผ่านใหม่ ” ระบบจะส่งรหัสผ่านใหม่ไปที่ E-Mail Address ที่ผู้ใช้งานกรอกข้อมูลไว้

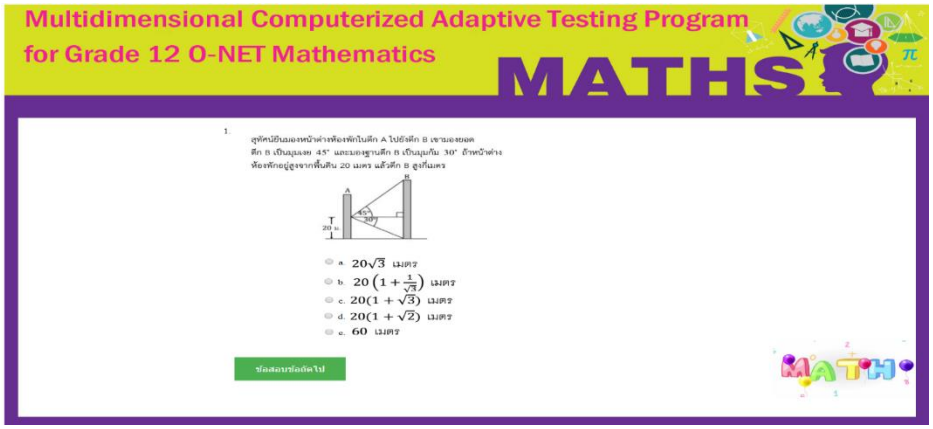


ภาพที่ 4-5 หน้าจอลงทะเบียนเพื่อเข้าใช้งานในระบบ

จากภาพที่ 4-5 กรอกข้อมูลให้ครบทุกช่อง เลขประจำตัวประชาชน รหัสเข้าใช้งานระบบ ชื่อ-สกุล เพศ วัน/เดือน/ปีเกิด โรงเรียน จังหวัด หลักจากนั้นให้กดปุ่ม “ทำการทดสอบ” เพื่อดำเนินการทดสอบต่อไป

1.2 การดำเนินการสอบ

หลังจากนั้นให้กดปุ่ม “ทำการทดสอบ” เพื่อดำเนินการทดสอบเรียบร้อยแล้ว จะเข้าสู่การดำเนินการทดสอบ แสดงได้ดังภาพที่ 4-6

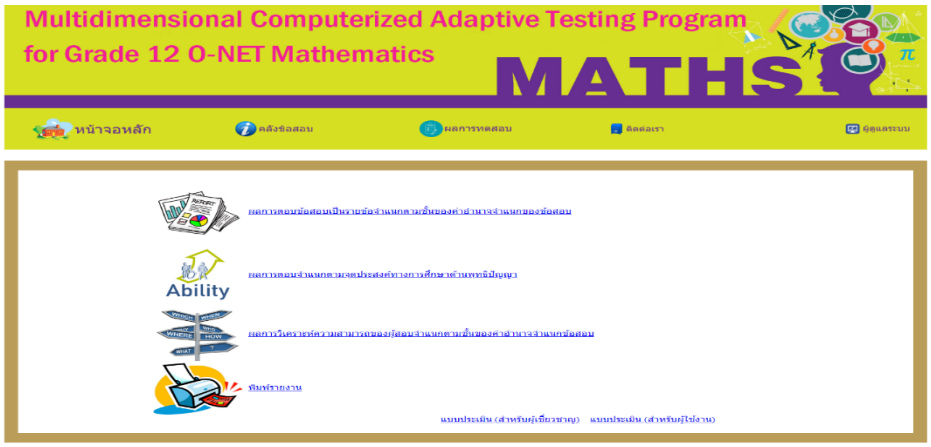


ภาพที่ 4-6 หน้าจอข้อสอบข้อแรก

จากภาพที่ 4-6 หน้าจอข้อสอบข้อแรกโดยโปรแกรมจะสุ่มเลือกข้อสอบขึ้นมาให้ ขึ้นอยู่กับความสามารถของผู้สอบโดยเรียงตามชั้นของค่าอำนาจจำแนกข้อสอบ หลังจากนั้นให้ผู้สอบเลือกคำตอบ แล้วกดปุ่ม “ข้อสอบข้อถัดไป” เพื่อรับข้อสอบข้อถัดไป การดำเนินการทดสอบจะดำเนินต่อไป จนสิ้นสุดการทดสอบ

1.3 รายงานผลารทดสอบ

หลังจากผู้สอบดำเนินการทดสอบเสร็จสิ้นแล้ว โปรแกรมจะให้กดปุ่ม “สรุปผลการทดสอบ” จะมีลิงค์ไปยังหน้าจอรายงานผลการทดสอบ แสดงดังภาพที่ 4-7



ภาพที่ 4-7 หน้าจอรายงานผลการทดสอบ

จากภาพที่ 4-7 หน้าจอรายงานผลการทดสอบ จะมีลิงค์ไปยังหน้าจอแสดงผลการทดสอบ 4 ส่วน ดังนี้

1) ผลการตอบข้อสอบเป็นรายข้อจำแนกตามชั้นของค่าอำนาจจำแนกของข้อสอบ

เมื่อกดปุ่มนี้จะแสดงรายงาน ผลการตอบข้อสอบเป็นรายข้อจำแนกตามชั้นของค่าอำนาจจำแนกของข้อสอบ

2) ผลการตอบจำแนกตามจุดประสงค์ทางการศึกษาด้านพุทธิปัญญา

เมื่อกดปุ่มนี้จะแสดงรายงาน ผลการตอบจำแนกตามจุดประสงค์ทางการศึกษาด้านพุทธิปัญญา

3) ผลการวิเคราะห์ความสามารถของผู้สอบจำแนกตามชั้นของค่าอำนาจจำแนกข้อสอบ

เมื่อกดปุ่มนี้จะแสดงรายงาน ผลการวิเคราะห์ความสามารถของผู้สอบจำแนกตามชั้นของค่าอำนาจจำแนกข้อสอบ

4) พิมพ์รายงาน

เมื่อกดปุ่มนี้จะแสดงรายงาน ผลในรูปแบบการพิมพ์ทางเครื่องพิมพ์

1.3.1 หน้าจอรายงานผลการตอบข้อสอบเป็นรายข้อจำแนกตามชั้นของค่าอำนาจจำแนกของข้อสอบ เมื่อกดปุ่มนี้ “ ผลการตอบข้อสอบเป็นรายข้อจำแนกตามชั้นของค่าอำนาจจำแนกของข้อสอบ ” จะปรากฏลิงค์ไปยัง หน้าจอรายงานผลการตอบข้อสอบเป็นรายข้อจำแนกตามชั้นของค่าอำนาจจำแนกของข้อสอบ แสดงดังภาพที่ 4-8

ผลการตอบข้อสอบเป็นรายข้อจำแนกตามชั้นและสรุปรวม

ชั้น	ข้อสอบ	รหัสข้อสอบ	ผลการตอบ
1	1	Q_2559_10.png	ผิด
	2	Q_2555_9.png	ผิด
	3	Q_2557_27.png	ผิด
	4	Q_2559_17.png	ผิด
2	5	Q_2556_3.png	ผิด
	6	Q_2553_2.png	ถูก
	7	Q_2556_32.png	ผิด
	8	Q_2552_16.png	ผิด
3	9	Q_2558_14.png	ผิด
	10	Q_2552_23.png	ถูก
	11	Q_2556_25.png	ผิด
	12	Q_2559_2.png	ถูก
4	13	Q_2557_5.png	ผิด
	14	Q_2556_7.png	ถูก
	15	Q_2556_18.png	ผิด
	16	Q_2551_26.png	ผิด
ผลการตอบถูก	4 ข้อ	คิดเป็นร้อยละ	25.0

ภาพที่ 4-8 หน้าจอรายงานผลการตอบข้อสอบเป็นรายข้อจำแนกตามชั้นของค่าอำนาจจำแนกของข้อสอบ

จากภาพที่ 4-8 แสดงหน้าจอรายงานผลการตอบข้อสอบเป็นรายข้อจำแนกตามชั้นของค่าอำนาจจำแนกของข้อสอบ โดยแสดงรายละเอียดเกี่ยวกับจำนวนข้อสอบในแต่ละชั้น รหัสข้อสอบ ผลการตอบ จำนวนข้อสอบที่ตอบถูกและร้อยละของผลการตอบถูก

1.3.2 หน้าจอรายงานผลการตอบจำแนกตามจุดประสงค์ทางการศึกษาด้านพุทธิปัญญา เมื่อกดปุ่มนี้ “ ผลการตอบจำแนกตามจุดประสงค์ทางการศึกษาด้านพุทธิปัญญา ” จะปรากฏลิงค์ไปยังหน้าจอรายงานผลการตอบจำแนกตามจุดประสงค์ทางการศึกษาด้านพุทธิปัญญา แสดงได้ดังภาพที่ 4-9

ผลการวิเคราะห์ชั้นของคลังข้อสอบกับกระบวนการพุทธิปัญญาภาพรวม

ชั้น	ข้อสอบ	กระบวนการพุทธิปัญญา: การวิเคราะห์เกี่ยวกับข้อเท็จจริง	การเข้าใจเกี่ยวกับมโนทัศน์	การประยุกต์เกี่ยวกับมโนทัศน์	การวิเคราะห์เกี่ยวกับมโนทัศน์	การประยุกต์เกี่ยวกับวิธีดำเนินการ	การวิเคราะห์เกี่ยวกับวิธีดำเนินการ	ผลการตอบ
1	1					✓		ผิด
	2					✓		ผิด
	3					✓		ผิด
2	4				✓			ผิด
	5				✓			ผิด
	6				✓			ถูก
	7		✓					ผิด
	8					✓		ผิด
3	9					✓		ผิด
	10					✓		ถูก
	11				✓			ผิด
	12					✓		ถูก
4	13	✓						ผิด
	14					✓		ถูก
	15					✓		ผิด
	16					✓		ผิด
รวม		1	1	0	4	10	0	

ภาพที่ 4-9 หน้าจอรายงานผลการตอบจำแนกตามจุดประสงค์ทางการศึกษาด้านพุทธิปัญญา

จากภาพที่ 4-9 แสดงหน้าจอรายงานผลการตอบจำแนกตามจุดประสงค์ทางการศึกษาด้านพุทธิปัญญา โดยแสดงรายละเอียดเกี่ยวกับจำนวนข้อสอบในแต่ละชั้น กระบวนการพุทธิปัญญา และผลการตอบ

1.3.3 หน้าจอรายงานผลการวิเคราะห์ความสามารถของผู้สอบจำแนกตามชั้นของค่าอำนาจจำแนกข้อสอบ เมื่อกดปุ่มนี้ “ ผลการวิเคราะห์ความสามารถของผู้สอบจำแนกตามชั้นของค่าอำนาจจำแนกข้อสอบ ” จะปรากฏลิงค์ไปยัง หน้าจอรายงานผลการวิเคราะห์ความสามารถของผู้สอบจำแนกตามชั้นของค่าอำนาจจำแนกข้อสอบ แสดงดังภาพที่ 4-10

ผลการวิเคราะห์ความสามารถของผู้สอบจำแนกตามชั้นและภาพรวม

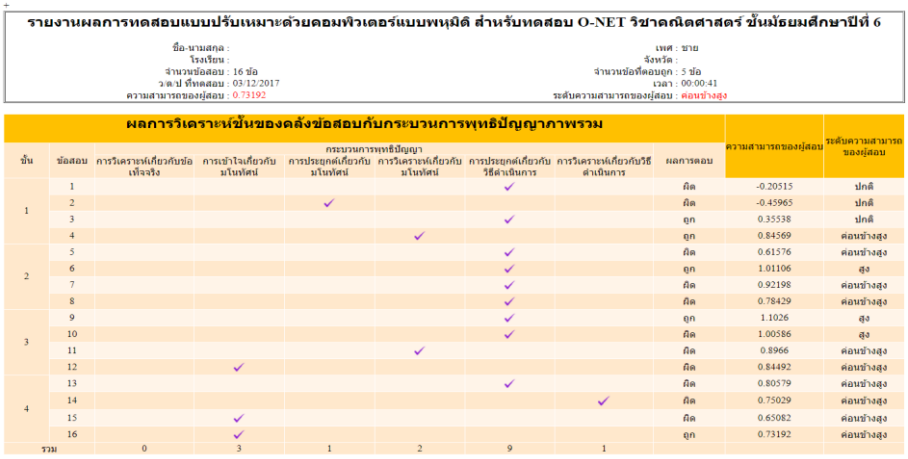
ชั้น	จำนวนข้อทั้งหมด	ข้อสอบที่ตอบถูก	ร้อยละ	ความสามารถของผู้สอบ	ระดับความสามารถของผู้สอบ
1	4	2	50.0	0.62414	ต่อนข้างสูง
2	4	2	50.0	1.14971	ค่อนข้างสูง
3	4	2	50.0	1.29234	สูง
4	4	2	50.0	1.31256	สูง
รวม	16	8	50.0		

กราฟความสามารถของผู้สอบ

ภาพที่ 4-10 หน้าจอรายงานผลการวิเคราะห์ความสามารถของผู้สอบจำแนกตามชั้นของค่าอำนาจจำแนกข้อสอบ

จากภาพที่ 4-10 แสดงหน้าจอรายงานผลการวิเคราะห์ความสามารถของผู้สอบจำแนกตามชั้นของค่าอำนาจจำแนกข้อสอบ โดยแสดงรายละเอียดเกี่ยวกับชั้นของข้อสอบ จำนวนข้อทั้งหมด ข้อสอบที่ตอบถูก ร้อยละของการตอบถูกในแต่ละชั้น ค่าความสามารถของผู้สอบ และระดับความสามารถของผู้สอบ

1.3.4 หน้าจอพิมพ์รายงาน เมื่อกดปุ่มนี้ “ ” จะปรากฏลิงค์ไปยัง หน้าจอพิมพ์รายงาน ดังภาพที่ 4-11



รายงานผลการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ สำหรับทดสอบ O-NET วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 6

ชื่อ-นามสกุล :
โรงเรียน :
จำนวนข้อสอบ : 16 ข้อ
ว/ด/ป ที่ทดสอบ : 03/12/2017
ความสามารถของผู้สอบ : 0.73192

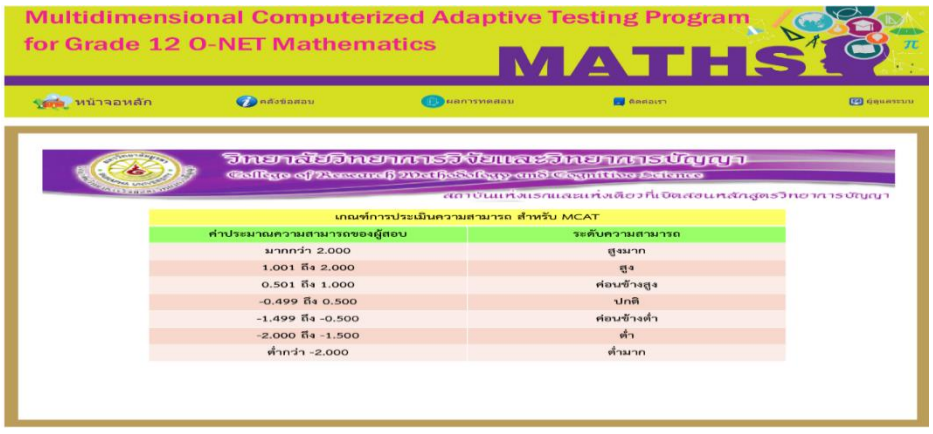
เพศ : ชาย
จังหวัด :
จำนวนข้อที่ตอบถูก : 5 ข้อ
เวลา : 00:00:41
ระดับความสามารถของผู้สอบ : ค่อนข้างสูง

ผลการวิเคราะห์ชั้นของคลังข้อสอบกับกระบวนการพุทธิปัญญาภาพรวม

ชั้น	ข้อสอบ	กระบวนการพุทธิปัญญา: การวิเคราะห์เกี่ยวกับข้อเท็จจริง	การเข้าใจเกี่ยวกับมโนทัศน์	การประยุกต์เกี่ยวกับมโนทัศน์	การวิเคราะห์เกี่ยวกับมโนทัศน์	การประยุกต์เกี่ยวกับวิธีดำเนินการ	การวิเคราะห์เกี่ยวกับวิธีดำเนินการ	ผลการตอบ	ความสามารถของผู้สอบ	ระดับความสามารถของผู้สอบ
1	1					✓		ผิด	-0.20515	ปกติ
	2			✓				ผิด	-0.45965	ปกติ
	3					✓		ถูก	0.35538	ปกติ
	4				✓			ถูก	0.84569	ค่อนข้างสูง
2	5					✓		ผิด	0.61576	ค่อนข้างสูง
	6					✓		ถูก	1.01106	สูง
	7					✓		ผิด	0.92198	ค่อนข้างสูง
	8					✓		ผิด	0.78429	ค่อนข้างสูง
3	9					✓		ถูก	1.1026	สูง
	10					✓		ผิด	1.00586	สูง
	11				✓			ผิด	0.8966	ค่อนข้างสูง
	12		✓					ผิด	0.84492	ค่อนข้างสูง
4	13					✓		ผิด	0.80579	ค่อนข้างสูง
	14						✓	ผิด	0.75029	ค่อนข้างสูง
	15		✓					ผิด	0.65082	ค่อนข้างสูง
	16		✓					ถูก	0.73192	ค่อนข้างสูง
รวม		0	3	1	2	9	1			

ภาพที่ 4-11 หน้าจอพิมพ์รายงาน

จากภาพที่ 4-11 แสดงหน้าจอพิมพ์รายงาน หน้าจอพิมพ์รายงานการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ โดยแสดงรายละเอียดผู้สอบ ได้แก่ ชื่อ-สกุล เพศ โรงเรียน จังหวัด จำนวนข้อสอบทั้งหมด จำนวนข้อสอบที่ตอบถูก วัน/เดือน/ปี เวลาที่ใช้ในการทดสอบ ค่าประมาณความสามารถของผู้สอบ (Ability Estimate) และระดับความสามารถของผู้สอบ (Ability Level) ทั้งนี้มีเกณฑ์การประเมินระดับความสามารถของผู้สอบมีรายละเอียดแสดงดังภาพที่ 4-12



เกณฑ์การประเมินความสามารถ สำหรับ MCAT

ค่าประมาณความสามารถของผู้สอบ	ระดับความสามารถ
มากกว่า 2.000	สูงมาก
1.001 ถึง 2.000	สูง
0.501 ถึง 1.000	ค่อนข้างสูง
-0.499 ถึง 0.500	ปกติ
-1.499 ถึง -0.500	ค่อนข้างต่ำ
-2.000 ถึง -1.500	ต่ำ
ต่ำกว่า -2.000	ต่ำมาก

ภาพที่ 4-12 หน้าจอเกณฑ์การประเมินความสามารถของผู้สอบ

2. ผลการประเมินความเหมาะสมของโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ

ผลการประเมินความเหมาะสมของโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ สำหรับทดสอบ O-NET วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 6 นำเสนอ 2 ส่วน ดังนี้

2.1 ผลการประเมินความเหมาะสมโดยผู้เชี่ยวชาญ

การประเมินความเหมาะสมของโปรแกรมการทดสอบโดยผู้เชี่ยวชาญ ผู้วิจัยได้นำโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ พร้อมคู่มือการใช้โปรแกรมส่งให้ผู้เชี่ยวชาญตรวจสอบ โดยผู้เชี่ยวชาญทั้ง 3 คน ได้ประเมินความเหมาะสมของโปรแกรมใน 5 ด้าน ได้แก่ 1) ความต้องการของผู้ใช้โปรแกรม 2) การทำงานของโปรแกรม 3) การใช้งาน 4) การรักษาความปลอดภัยของการเข้าถึงข้อมูล 5) ความชัดเจนของคู่มือการใช้งาน และข้อเสนอแนะเพื่อการพัฒนาโปรแกรม ประเมินในระหว่างวันที่ 1–10 ธันวาคม พ.ศ. 2560 ผลการประเมินความเหมาะสมแสดงดังตารางที่ 4-31

ตารางที่ 4-31 ผลการประเมินความเหมาะสมของโปรแกรมโดยผู้เชี่ยวชาญ

ด้านการประเมิน	Mean	*SD*	ระดับความเหมาะสม
1. ความต้องการของผู้ใช้โปรแกรม	4.73	0.23	มากที่สุด
2. การทำงานของโปรแกรม	4.17	0.87	มาก
3. การใช้งาน	4.23	0.35	มาก
4. การรักษาความปลอดภัยของการเข้าถึงข้อมูล	3.67	0.42	มาก
5. ความชัดเจนของคู่มือการใช้งาน	4.67	0.48	มากที่สุด
สรุป	4.27	0.41	มาก

จากตารางที่ 4-31 แสดงผลการประเมินความเหมาะสมของโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ สำหรับทดสอบ O-NET วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 6 โดยผู้เชี่ยวชาญ แสดงให้เห็นว่า โปรแกรมในภาพรวมมีความเหมาะสมอยู่ในระดับมาก (Mean=4.27) เมื่อพิจารณาเป็นรายด้าน ปรากฏว่า มี 2 ด้าน ที่มีความเหมาะสมอยู่ในระดับมากที่สุด ได้แก่ ด้านความต้องการของผู้ใช้โปรแกรม (Mean = 4.73) และด้านความชัดเจนของคู่มือการใช้งาน (Mean = 4.67) รองลงมาได้แก่ ด้านการใช้งาน มีความเหมาะสมอยู่ในระดับมาก (Mean = 4.23) ด้านการทำงานของโปรแกรม มีความเหมาะสมอยู่ในระดับมาก (Mean = 4.17) และด้านการรักษาความปลอดภัยของการเข้าถึงข้อมูล มีความเหมาะสมอยู่ในระดับมาก (Mean = 3.67)

นอกจากนี้ ผู้เชี่ยวชาญยังมีข้อเสนอแนะเพิ่มเติม ดังนี้

1. เลขประจำตัวประชาชน ตรวจสอบไม่ได้ว่าถูกต้องหรือไม่ ควรมีการเพิ่มช่องวันเดือนปีเกิด วันเดือนปีเกิดนำมาประมวลผลเพื่อป้องกันผู้ที่อายุเกินมาทำแบบทดสอบ

2. เรื่องรหัสเข้าใช้งาน ควรใช้ได้ทั้งตัวเลขอย่างเดียว หรือ ตัวเลขผสมตัวอักษร และหากผู้ใช้ลืมเปลี่ยนตัวอักษรจากภาษาไทยเป็นภาษาอังกฤษ

3. ควรมีช่องทางสำหรับ Reset Password ของผู้ใช้ กรณีที่จำ Password ไม่ได้ ให้ใช้ Email โดย User กรอก Mail แล้วให้ระบบส่ง Password แล้วค่อยไปแก้ Password

4. ถ้า user คนหนึ่งทำหลายๆ ครั้ง ระบบควรเก็บข้อมูลการทำหลายครั้งไว้เพื่อเก็บสถิติและดูพัฒนาการ

5. เมนู คู่มือการใช้งานโปรแกรมเมื่อทำการโหลดไฟล์ pdf ขึ้นมาแล้วควรให้ขึ้นหน้าต่างใหม่พร้อมทั้งค้างแท็ปหน้าเพจหลักไว้อีกหนึ่งแท็ปด้วย

6. เมนูความหมายของคลังข้อสอบ รูปแบบการจัดคลังข้อสอบ เกณฑ์การประเมินความสามารถของผู้สอบ เมื่อคลิกแล้วข้อมูลที่แสดงตัวอักษรไม่คมชัด มี

2.2 ผลการประเมินความเหมาะสมโดยผู้ใช้งาน

การประเมินความเหมาะสมของโปรแกรมการทดสอบโดยผู้ใช้งาน เป็นการประเมินความคิดเห็นของผู้ใช้งานโปรแกรมทั้ง 2 ด้าน ด้แก่ 1) การใช้งาน 2) ความชัดเจนของคู่มือการใช้งาน และข้อเสนอแนะเพื่อการพัฒนาโปรแกรม กลุ่มตัวอย่างที่ทดลองใช้เป็นนักเรียนมัธยมศึกษาชั้นปีที่ 6 โรงเรียนบางกะปิ จำนวน 30 คน คัดเลือกด้วยวิธีการเลือกตัวอย่างตามสะดวก ซึ่งประเมินในระหว่างวันที่ 15–29 ธันวาคม พ.ศ. 2560 ผลการประเมินความเหมาะสมของโปรแกรมแสดงดังตารางที่ 4-32

ตารางที่ 4-32 ผลการประเมินความเหมาะสมของโปรแกรม โดยผู้ใช้งาน

ด้านการประเมิน	Mean	*SD*	ระดับความเหมาะสม
1. การใช้งาน	4.17	0.46	มาก
2. ความชัดเจนของคู่มือการใช้งาน	4.47	0.51	มาก
สรุป	4.27	0.38	มาก

จากตารางที่ 4-32 แสดงผลการประเมินความเหมาะสมของโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ สำหรับทดสอบ O-NET วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 6 โดยผู้ใช้งาน แสดงให้เห็นว่า โปรแกรมในภาพรวมมีความเหมาะสมอยู่ในระดับมาก (Mean=4.27) เมื่อพิจารณาเป็นรายด้าน ปรากฏว่า ด้านความชัดเจนของคู่มือการใช้งานมีความเหมาะสมอยู่ในระดับมาก (Mean = 4.47) รองลงมาได้แก่ ด้านการใช้งานมีความเหมาะสมอยู่ในระดับมาก (Mean = 4.18)

นอกจากนี้ ผู้ใช้งานมีข้อเสนอแนะเพิ่มเติม ดังนี้

1. ขนาดตัวอักษรในเว็บไซต์ไม่ค่อยชัดเจน
2. ปรับปรุงโปรแกรมให้สามารถกลับไปแก้ไขคำตอบได้
3. ปรับปรุงให้มีรูปภาพที่น่าสนใจมากกว่านี้
4. อยากให้โปรแกรมแจ้งเตือนทันทีว่าตอบถูกหรือผิดก่อนไปข้อถัดไป

จากผลการประเมินความเหมาะสมของโปรแกรมทั้งผู้เชี่ยวชาญและผู้ใช้งาน ผู้วิจัยได้นำข้อเสนอแนะมาปรับปรุงแก้ไขโปรแกรมให้มีความเหมาะสมยิ่งขึ้น จากผลการประเมินความเหมาะสมโปรแกรม แสดงให้เห็นว่า โปรแกรมมีความเหมาะสมเพียงพอที่จะนำไปใช้ทดสอบจริงได้

ตอนที่ 4 ผลการเปรียบเทียบความสามารถของผู้สอบแต่ละระดับชั้นของค่าอำนาจจำแนกของข้อสอบ สำหรับการทดสอบO-NET วิชาคณิตศาสตร์ ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 จำแนกตามเพศ และความสามารถด้านการเรียนวิชาคณิตศาสตร์

ผลการศึกษาการเปรียบเทียบความสามารถของผู้สอบแต่ละระดับชั้นของค่าอำนาจจำแนกของข้อสอบ สำหรับการทดสอบ O-NET วิชาคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 จำแนกตามเพศและความสามารถด้านการเรียนวิชาคณิตศาสตร์ กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาชั้นปีที่ 6 โรงเรียนสาธิตมหาวิทยาลัยราชภัฏบ้านสมเด็จเจ้าพระยา จำนวน 80 คน ใช้วิธีการสุ่มอย่างง่าย (Simple Random Sampling) ดำเนินการเก็บข้อมูลในวันพฤหัสบดีที่ 18 มกราคม พ.ศ. 2561 โดยใช้โปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ สำหรับทดสอบ O-NET วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 6 เป็นเครื่องมือที่ใช้ในการวิจัย ผลการวิจัย แสดงดังตารางที่ 4-33

ตารางที่ 4-33 ข้อมูลพื้นฐานและผลการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ

คนที่	เพศ	GPAX (Math) สูง : ต่ำ	จำนวนที่ตอบถูก (ข้อ)	เวลาที่ใช้ในการทดสอบ (นาที)	ความสามารถของผู้สอบตามระดับชั้นของค่าอำนาจจำแนกของข้อสอบ (a)			
					ชั้นที่ 1	ชั้นที่ 2	ชั้นที่ 3	ชั้นที่ 4
1	ชาย	สูง	5	23:08	0.021	0.526	0.683	0.983
2	ชาย	สูง	6	22:01	0.763	1.042	0.944	0.823
3	ชาย	สูง	7	24:24	1.487	1.144	1.323	1.088
4	ชาย	สูง	7	24:09	1.242	0.787	1.102	0.972
5	ชาย	สูง	7	22:01	1.952	0.824	1.069	0.978
6	ชาย	สูง	8	24:12	0.624	0.684	1.486	1.243
7	ชาย	สูง	8	23:17	0.756	1.527	1.186	1.284
8	ชาย	สูง	8	30:34	0.055	1.319	1.329	1.350
9	ชาย	สูง	8	19:44	0.775	0.655	1.319	1.331
10	ชาย	สูง	8	26:30	1.132	1.767	1.319	1.099
11	ชาย	สูง	8	27:03	0.725	1.261	1.308	1.321
12	ชาย	สูง	8	22:04	1.971	1.642	1.206	1.046
13	ชาย	สูง	8	26:14	0.725	1.532	1.150	1.170
14	ชาย	สูง	8	27;56	1.971	0.843	1.119	1.257
15	ชาย	สูง	9	28:32	1.460	1.457	1.412	1.355
16	ชาย	สูง	9	25:02	1.952	1.610	1.456	1.224
17	ชาย	สูง	9	30;15	1.460	1.819	1.360	1.354
18	ชาย	สูง	10	29:58	0.972	1.638	1.790	1.517
19	ชาย	สูง	11	30:01	1.971	1.109	1.526	1.573
20	ชาย	สูง	12	25:34	1.165	1.424	1.895	1.878
21	หญิง	สูง	6	27:05	0.725	0.687	1.047	0.958
22	หญิง	สูง	6	23:52	0.012	0.516	0.607	1.152

ตารางที่ 4-33 (ต่อ)

คนที่	เพศ	GPAX (Math) สูง : ต่ำ	จำนวนที่ตอบถูก (ข้อ)	เวลาที่ใช้ในการทดสอบ (นาที)	ความสามารถของผู้สอบตามระดับชั้นของค่าอำนาจจำแนกของข้อสอบ (a)			
					ชั้นที่ 1	ชั้นที่ 2	ชั้นที่ 3	ชั้นที่ 4
23	หญิง	สูง	6	16:34	0.734	0.645	1.018	0.838
24	หญิง	สูง	6	26:55	-0.033	0.385	0.928	1.165
25	หญิง	สูง	7	23:16	1.132	0.260	0.555	1.171
26	หญิง	สูง	7	19:37	1.487	0.938	0.920	1.011
27	หญิง	สูง	7	22:47	1.487	1.955	1.354	0.949
28	หญิง	สูง	7	23:54	0.731	0.714	1.092	1.090
29	หญิง	สูง	7	22:41	1.211	0.951	1.109	1.066
30	หญิง	สูง	7	20:17	-0.635	0.678	1.261	1.189
31	หญิง	สูง	7	26:04	1.460	1.322	1.360	0.924
32	หญิง	สูง	7	26:13	0.539	1.459	1.181	1.008
33	หญิง	สูง	7	25:54	1.487	1.349	1.351	0.924
34	หญิง	สูง	8	24:01	0.756	1.150	1.279	1.223
35	หญิง	สูง	8	27:21	0.756	0.664	1.036	1.287
36	หญิง	สูง	9	22:07	1.952	1.619	1.497	1.231
37	หญิง	สูง	9	28:10	1.242	0.848	1.114	1.389
38	หญิง	สูง	9	19:50	1.427	1.090	1.491	1.367
39	หญิง	สูง	10	23:08	1.132	1.330	1.531	1.472
40	หญิง	สูง	11	25:09	1.460	1.530	1.669	1.706
41	ชาย	ต่ำ	2	20:23	-0.855	-0.210	-0.317	0.251
42	ชาย	ต่ำ	3	23:23	0.314	-0.250	0.273	0.550
43	ชาย	ต่ำ	3	18:05	0.127	-0.222	-0.327	0.611
44	ชาย	ต่ำ	3	21:45	-0.720	-0.871	-0.033	0.680
45	ชาย	ต่ำ	3	23:45	-0.600	0.788	0.897	0.538
46	ชาย	ต่ำ	3	20:02	0.049	-0.315	0.331	0.511
47	ชาย	ต่ำ	3	18:26	0.178	0.162	0.502	0.285
48	ชาย	ต่ำ	3	20:29	-0.645	0.165	0.001	0.806
49	ชาย	ต่ำ	3	18:21	-0.616	-0.779	-0.814	0.804
50	ชาย	ต่ำ	3	20:11	0.854	0.165	0.522	0.342
51	ชาย	ต่ำ	4	23:35	0.754	-0.111	0.353	0.541
52	ชาย	ต่ำ	4	22:03	0.31	0.579	1.063	0.774
53	ชาย	ต่ำ	4	22:36	-0.245	0.242	0.827	0.550
54	ชาย	ต่ำ	4	19:02	0.245	0.530	0.685	0.666
55	ชาย	ต่ำ	4	19:46	-0.674	-0.268	0.750	0.740
56	ชาย	ต่ำ	4	19:06	1.06	0.339	0.604	0.774
57	ชาย	ต่ำ	4	20:06	0.251	0.990	0.590	0.672
58	ชาย	ต่ำ	4	20:04	-0.693	-0.886	0.839	0.792

ตารางที่ 4-33 (ต่อ)

คนที่	เพศ	GPAX (Math) สูง : ต่ำ	จำนวนที่ตอบถูก (ข้อ)	เวลาที่ใช้ในการทดสอบ (นาที)	ความสามารถของผู้สอบตามระดับชั้นของค่าอำนาจจำแนกของข้อสอบ (a)			
					ชั้นที่ 1	ชั้นที่ 2	ชั้นที่ 3	ชั้นที่ 4
59	ชาย	ต่ำ	4	18:23	0.205	0.958	0.867	0.627
60	ชาย	ต่ำ	5	27:06	0.299	1.057	0.868	0.873
61	หญิง	ต่ำ	2	23:23	-0.645	-0.805	0.507	0.344
62	หญิง	ต่ำ	2	22:06	-0.578	-0.776	-0.817	0.644
63	หญิง	ต่ำ	2	19:22	-0.656	-0.931	-0.062	0.359
64	หญิง	ต่ำ	3	22:30	-0.605	-0.902	-0.014	0.651
65	หญิง	ต่ำ	3	20:23	-0.577	-0.769	0.635	0.608
66	หญิง	ต่ำ	3	21:45	0.178	0.448	0.657	0.503
67	หญิง	ต่ำ	3	21:04	-0.233	-0.498	-0.590	0.504
68	หญิง	ต่ำ	3	25:29	0.029	0.261	0.069	0.421
69	หญิง	ต่ำ	3	21:11	0.303	0.874	0.498	0.285
70	หญิง	ต่ำ	4	19:07	0.132	1.009	0.551	0.715
71	หญิง	ต่ำ	4	19:25	0.318	0.553	0.705	0.782
72	หญิง	ต่ำ	4	23:01	-0.253	1.260	0.742	0.550
73	หญิง	ต่ำ	4	20:03	0.825	0.246	0.769	0.597
74	หญิง	ต่ำ	4	24:35	0.734	0.131	-0.081	0.692
75	หญิง	ต่ำ	4	20:03	0.994	0.252	0.084	0.595
76	หญิง	ต่ำ	5	19:05	-0.577	0.230	1.150	1.080
77	หญิง	ต่ำ	5	27:01	0.763	0.760	1.058	0.665
78	หญิง	ต่ำ	5	25:09	0.103	1.432	1.130	0.822
79	หญิง	ต่ำ	5	20:06	0.838	0.179	0.489	0.806
80	หญิง	ต่ำ	5	21:36	0.275	0.434	0.928	0.845
Min (ต่ำ)	-	-	2	18:05	-0.855	-0.931	-0.817	0.251
Max (ต่ำ)	-	-	5	27:06	1.060	1.432	1.150	1.080
Mean(ต่ำ)	-	-	3.58	21:21	0.024	0.136	0.422	0.621
SD (ต่ำ)	-	-	0.87	2:34	0.56	0.66	0.51	0.18
Min (สูง)	-	-	5	16:34	-0.635	0.260	0.555	0.823
Max (สูง)	-	-	12	30:34	1.971	1.955	1.1895	1.878
Mean(สูง)	-	-	7.88	24:53	1.056	1.118	1.235	1.199
SD (สูง)	-	-	1.51	3:21	0.63	0.44	0.28	0.23
Min (รวม)	-	-	2	16:34	-0.855	-0.931	-0.817	0.251
Max (รวม)	-	-	12	30:34	1.971	1.955	1.895	1.878
Mean(รวม)	-	-	5.73	22:87	0.540	0.627	0.828	0.910
SD (รวม)	-	-	2.49	3:25	0.787	0.744	0.581	0.357

หมายเหตุ: ค่า Mean ของ GPAX วิชาคณิตศาสตร์ เท่ากับ 3.49 จัดกลุ่ม GPAX วิชาคณิตศาสตร์ ออกเป็น 2 กลุ่ม ได้แก่ 1) สูง หมายถึง $(GPAX_{math} \geq 3.50)$ 2) ต่ำ หมายถึง $(GPAX_{math} < 3.50)$

จากตารางที่ 4-33 แสดงข้อมูลพื้นฐานของผู้สอบและผลการทดสอบที่ได้จากการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ ในภาพรวม ปรากฏว่า นักเรียนที่ทดสอบด้วยโปรแกรมจำนวน 80 คน จำแนกเป็นเพศชาย 40 คน เพศหญิง 40 คน และจำแนกตามความสามารถด้านการเรียนวิชาคณิตศาสตร์ โดยใช้เกรดเฉลี่ยสะสม (GPAX) วิชาคณิตศาสตร์ ในการแบ่ง2 กลุ่ม ได้แก่ กลุ่มสูง $(GPAX_{math} \geq 3.50)$ และกลุ่มต่ำ $(GPAX_{math} < 3.50)$ กลุ่มละ 40 คน โดยจัดกลุ่มชายหญิงภายในกลุ่มสูงและกลุ่มต่ำ กลุ่มละ 20 คน มีจำนวนข้อสอบที่ตอบถูกเฉลี่ยเท่ากับ 5.73 ข้อ และใช้เวลาในการทดสอบเฉลี่ยเท่ากับ 22.87 นาที ความสามารถของผู้สอบ $(\hat{\theta})$ ตามระดับชั้นของค่าอำนาจจำแนกทั้ง 4 ชั้น ดังนี้

การประมาณค่าความสามารถของผู้สอบ $(\hat{\theta})$ ในภาพรวม ปรากฏว่า ชั้นที่ 1 มีค่าความสามารถของผู้สอบเฉลี่ย 0.540 หมายความว่า ผู้สอบที่ทดสอบความสามารถในคลังข้อสอบชั้นที่ 1 มีความสามารถโดยเฉลี่ยในระดับความสามารถค่อนข้างสูง ชั้นที่ 2 มีค่าความสามารถของผู้สอบเฉลี่ย 0.627 หมายความว่า ผู้สอบที่ทดสอบความสามารถในคลังข้อสอบชั้นที่ 2 มีความสามารถโดยเฉลี่ยในระดับความสามารถค่อนข้างสูง ชั้นที่ 3 มีค่าความสามารถของผู้สอบเฉลี่ย 0.828 หมายความว่า ผู้สอบที่ทดสอบความสามารถในคลังข้อสอบชั้นที่ 3 มีความสามารถโดยเฉลี่ยในระดับความสามารถค่อนข้างสูง และชั้นที่ 4 มีค่าความสามารถของผู้สอบเฉลี่ย 0.910 หมายความว่า ผู้สอบที่ทดสอบความสามารถในคลังข้อสอบชั้นที่ 4 มีความสามารถโดยเฉลี่ยในระดับความสามารถค่อนข้างสูง (เกณฑ์การประเมิน แสดงดังภาพที่ 4-12)

ข้อมูลพื้นฐานของผู้สอบและผลการทดสอบที่ได้จากการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ จำแนกตามความสามารถด้านการเรียนวิชาคณิตศาสตร์กลุ่มสูง $(GPAX_{math} \geq 3.50)$ ปรากฏว่า นักเรียนที่ทดสอบด้วยโปรแกรมจำนวน 40 คน จำแนกเป็นเพศชาย 20 คน เพศหญิง 20 คน มีจำนวนข้อสอบที่ตอบถูกเฉลี่ยเท่ากับ 7.88 ข้อ และใช้เวลาในการทดสอบเฉลี่ยเท่ากับ 24.53 นาที ความสามารถของผู้สอบ $(\hat{\theta})$ ตามระดับชั้นของค่าอำนาจจำแนกทั้ง 4 ชั้น ดังนี้

การประมาณค่าความสามารถของผู้สอบ $(\hat{\theta})$ จำแนกตามความสามารถด้านการเรียนวิชาคณิตศาสตร์กลุ่มสูง $(GPAX_{math} \geq 3.50)$ ปรากฏว่า ชั้นที่ 1 มีค่าความสามารถของผู้สอบเฉลี่ย 1.056 หมายความว่า ผู้สอบที่ทดสอบความสามารถในคลังข้อสอบชั้นที่ 1 มีความสามารถโดยเฉลี่ยในระดับความสามารถสูง ชั้นที่ 2 มีค่าความสามารถของผู้สอบเฉลี่ย 1.118 หมายความว่า ผู้สอบที่ทดสอบความสามารถในคลังข้อสอบชั้นที่ 2 มีความสามารถโดยเฉลี่ยในระดับความสามารถสูง ชั้นที่ 3 มีค่าความสามารถของผู้สอบเฉลี่ย 1.235 หมายความว่า ผู้สอบที่ทดสอบความสามารถในคลังข้อสอบชั้นที่ 3 มีความสามารถโดยเฉลี่ยในระดับความสามารถสูง และชั้นที่ 4 มีค่าความสามารถของผู้สอบเฉลี่ย 1.199 หมายความว่า ผู้สอบที่ทดสอบความสามารถในคลังข้อสอบชั้นที่ 4 มีความสามารถโดยเฉลี่ยในระดับความสามารถสูง

ข้อมูลพื้นฐานของผู้สอบและผลการทดสอบที่ได้จากการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ จำแนกตามความสามารถด้านการเรียนวิชาคณิตศาสตร์กลุ่มต่ำ $(GPAX_{math} < 3.50)$ ปรากฏว่า นักเรียนที่ทดสอบด้วยโปรแกรมจำนวน 40 คน จำแนกเป็นเพศชาย 20

คน เพศหญิง 20 คน มีจำนวนข้อสอบที่ตอบถูกเฉลี่ยเท่ากับ 3.58 ข้อ และใช้เวลาในการทดสอบเฉลี่ยเท่ากับ 21.21 นาที ความสามารถของผู้สอบ $(\hat{\theta})$ ตามระดับชั้นของค่าอำนาจจำแนกทั้ง 4 ชั้น ดังนี้

การประมาณค่าความสามารถของผู้สอบ $(\hat{\theta})$ จำแนกตามความสามารถด้านการเรียนวิชาคณิตศาสตร์กลุ่มต่ำ $(GPAX_{math} < 3.50)$ ปรากฏว่า ชั้นที่ 1 มีค่าความสามารถของผู้สอบเฉลี่ย 0.024 หมายความว่า ผู้สอบที่ทดสอบความสามารถในคลังข้อสอบชั้นที่ 1 มีความสามารถโดยเฉลี่ยในระดับความสามารถปกติ ชั้นที่ 2 มีค่าความสามารถของผู้สอบเฉลี่ย 0.136 หมายความว่า ผู้สอบที่ทดสอบความสามารถในคลังข้อสอบชั้นที่ 2 มีความสามารถโดยเฉลี่ยในระดับความสามารถปกติ ชั้นที่ 3 มีค่าความสามารถของผู้สอบเฉลี่ย 0.422 หมายความว่า ผู้สอบที่ทดสอบความสามารถในคลังข้อสอบชั้นที่ 3 มีความสามารถโดยเฉลี่ยในระดับความสามารถปกติ และชั้นที่ 4 มีค่าความสามารถของผู้สอบเฉลี่ย 0.621 หมายความว่า ผู้สอบที่ทดสอบความสามารถในคลังข้อสอบชั้นที่ 4 มีความสามารถโดยเฉลี่ยในระดับความสามารถค่อนข้างสูง

ผลการวิเคราะห์ความแปรปรวนของคะแนนความสามารถของผู้สอบ $(\hat{\theta})$ โดยใช้ตัวแปรอิสระ 2 ตัว คือ เพศ (ชายและหญิง) และความสามารถด้านการเรียนวิชาคณิตศาสตร์ $(GPAX_{math} < 3.50)$ และ $(GPAX_{math} \geq 3.50)$ ตามระดับชั้นของค่าอำนาจจำแนกทั้ง 4 ชั้น ดังนี้

ตารางที่ 4-34 ความสามารถของผู้สอบแต่ละระดับชั้นของค่าอำนาจจำแนกของข้อสอบ จำแนกตามเพศและความสามารถด้านการเรียนวิชาคณิตศาสตร์

ระดับชั้นของค่าอำนาจจำแนกของข้อสอบ	เพศ	GPAX (Math) สูง : ต่ำ	*n*	Mean	*SD*
ชั้นที่ 1 (ค่า a อยู่ระหว่าง 0.50-0.99)	ชาย	สูง	20	1.159	0.62
		ต่ำ	20	-0.020	0.58
	หญิง	สูง	20	0.953	0.63
		ต่ำ	20	0.068	0.56
ชั้นที่ 2 (ค่า a อยู่ระหว่าง 1.00-1.49)	ชาย	สูง	20	1.231	0.40
		ต่ำ	20	0.103	0.60
	หญิง	สูง	20	1.005	0.45
		ต่ำ	20	0.169	0.73
ชั้นที่ 3 (ค่า a อยู่ระหว่าง 1.50-1.99)	ชาย	สูง	20	1.299	0.27
		ต่ำ	20	0.424	0.50
	หญิง	สูง	20	1.170	0.29
		ต่ำ	20	0.420	0.54
ชั้นที่ 4 (ค่า a อยู่ระหว่าง 2.00-2.50)	ชาย	สูง	20	1.242	0.24
		ต่ำ	20	0.619	0.18
	หญิง	สูง	20	1.156	0.21
		ต่ำ	20	0.623	0.19

จากตารางที่ 4-34 ปรากฏว่า ชั้นที่ 1 กลุ่มสูง นักเรียนชาย มีค่าเฉลี่ยความสามารถของผู้สอบสูงกว่านักเรียนหญิง และ ถ้าพิจารณาจากความสามารถด้านการเรียนวิชาคณิตศาสตร์ กลุ่มต่ำ นักเรียนชาย มีค่าเฉลี่ยความสามารถของผู้สอบน้อยกว่านักเรียนหญิง

ตารางที่ 4-35 ผลการวิเคราะห์ความแปรปรวนของค่าความสามารถของผู้สอบแต่ละระดับชั้นของค่าอำนาจจำแนกของข้อสอบ จำแนกเพศและความสามารถด้านการเรียนวิชาคณิตศาสตร์

ระดับชั้นของค่าอำนาจจำแนกของข้อสอบ	แหล่งความแปรปรวน	*SS*	*df*	*MS*	*F*	*p*-value
ชั้นที่ 1	เพศ	0.069	1	.069	20.328	.662
(ค่า a อยู่ระหว่าง 0.50-0.99)	$GPAX_{Math}$	21.295	1	21.295	65.287**	*p*<.01
	เพศ* $GPAX_{Math}$	.433	1	.433	.193	.274
	ความคลาดเคลื่อน	27.165	76	.357	59.579	
	รวม	72.298	80		1.212	
ชั้นที่ 2	เพศ	.128	1	.128	.405	.526
(ค่า a อยู่ระหว่าง 1.00-1.49)	$GPAX_{Math}$	19.256	1	19.256	61.193**	*p*<.01
	เพศ* $GPAX_{Math}$	.427	1	.427	1.357	.248
	ความคลาดเคลื่อน	23.915	76	.315		
	รวม	75.165	80			
ชั้นที่ 3	เพศ	.088	1	.088	.504	.480
(ค่า a อยู่ระหว่าง 1.50-1.99)	$GPAX_{Math}$	13.197	1	13.197	75.515**	*p*<.01
	เพศ* $GPAX_{Math}$	.079	1	.079	.450	.504
	ความคลาดเคลื่อน	13.282	76	.175		
	รวม	81.544	80			
ชั้นที่ 4	เพศ	.034	1	.034	.780	.380
(ค่า a อยู่ระหว่าง 2.00-2.50)	$GPAX_{Math}$	6.676	1	6.676	153.957**	*p*<.01
	เพศ* $GPAX_{Math}$	.041	1	.041	.941	.335
	ความคลาดเคลื่อน	3.296	76	.043		
	รวม	76.333	80			

จากตารางที่ 4-35 ปรากฏว่า เพศมีผลต่อความสามารถของผู้สอบไม่แตกต่างกัน ผลร่วมกันระหว่างเพศกับความสามารถด้านการเรียนวิชาคณิตศาสตร์ของนักเรียนมีผลต่อความสามารถของผู้สอบไม่แตกต่างกัน แต่ปรากฏว่า นักเรียนที่มีความสามารถด้านการเรียนวิชาคณิตศาสตร์คณิตศาสตร์กลุ่มสูง มีความสามารถของผู้สอบสูงกว่า นักเรียนที่มีความสามารถด้านการเรียนวิชาคณิตศาสตร์กลุ่มต่ำ อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ในทุกระดับชั้นของค่าอำนาจจำแนกของข้อสอบ

สรุปผลการเปรียบเทียบความสามารถของผู้สอบแต่ละระดับชั้นของค่าอำนาจจำแนกของข้อสอบ สำหรับการทดสอบ O-NET วิชาคณิตศาสตร์ ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 จำแนกตามเพศและความสามารถด้านการเรียนวิชาคณิตศาสตร์

1. นักเรียนที่ทดสอบด้วยโปรแกรมจำนวน 80 คน จำแนกตาม เพศ เป็นเพศชาย 40 คน เพศหญิง 40 คน และจำแนกตามความสามารถด้านการเรียนวิชาคณิตศาสตร์ เป็นกลุ่มสูง $(GPAX_{math} \geq 3.50)$ และ กลุ่มต่ำ $(GPAX_{math} < 3.50)$ กลุ่มละ 40 คน โดยจัดกลุ่มชายหญิงภายในกลุ่มสูงและกลุ่มต่ำ กลุ่มละ 20 คน มีจำนวนข้อสอบที่ตอบถูกเฉลี่ยเท่ากับ 5.73 ข้อ และใช้เวลาในการทดสอบเฉลี่ยเท่ากับ 22.87 นาที

2. ผลการเปรียบเทียบความสามารถของผู้สอบแต่ละระดับชั้นของค่าอำนาจจำแนกของข้อสอบ โดยใช้โปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ สำหรับการทดสอบ O-NET วิชาคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 จำแนกตามเพศและความสามารถด้านการเรียนวิชาคณิตศาสตร์ ปรากฏว่า

2.1 เพศมีผลต่อความสามารถของผู้สอบไม่แตกต่างกัน

2.2 ผลร่วมกันระหว่างเพศกับความสามารถด้านการเรียนวิชาคณิตศาสตร์ของนักเรียนมีผลต่อความสามารถของผู้สอบไม่แตกต่างกัน

2.3 นักเรียนที่มีความสามารถด้านการเรียนวิชาคณิตศาสตร์กลุ่มสูง มีความสามารถของผู้สอบสูงกว่า นักเรียนที่มีความสามารถด้านการเรียนวิชาคณิตศาสตร์กลุ่มต่ำ อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ในทุกระดับชั้นของค่าอำนาจจำแนกของข้อสอบ

ซึ่งตัวแปรที่สอดคล้องกับสมมติฐานคือความสามารถด้านการเรียนวิชาคณิตศาสตร์ แต่ตัวแปรเพศไม่สอดคล้องกับสมมติฐานที่ตั้งไว้

บทที่ 5
สรุปและอภิปรายผล

การวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา สำหรับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ การดำเนินการวิจัยแบ่งเป็น 4 ระยะ ได้แก่ 1) การพัฒนาวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา 2) การเปรียบเทียบประสิทธิภาพของวิธีการจัดคลังข้อสอบโดยการจำลองข้อมูล โดยใช้ข้อมูลจริงเป็นฐานข้อมูลในการจำลองข้อมูล และใช้โปรแกรม WinGen3 โดยการนำเข้าค่าพารามิเตอร์ของมูลจริงแต่ละชั้นไปจำลองค่าความสามารถที่แท้จริงของผู้สอบ 500 ค่า จำลอง 10 รอบ วิเคราะห์ข้อมูลด้วยค่าเฉลี่ยเลขคณิต ส่วนเบี่ยงเบนมาตรฐาน ด้านการประมาณค่าความสามารถของผู้สอบ ได้แก่ ค่าความลำเอียงเฉลี่ย และค่ารากที่สองของความคลาดเคลื่อนกำลังสอง วิเคราะห์ข้อมูลด้วย Wilcoxon Test และประสิทธิภาพของการใช้ข้อสอบในคลังข้อสอบ ได้แก่ ข้อสอบที่มีการแสดงมากเกินไป ข้อสอบที่มีการใช้น้อยเกินไป อัตราการทับซ้อนของข้อสอบ และการแจกแจงอัตราการแสดงข้อสอบ วิเคราะห์ข้อมูลด้วย Chi-Square Test 3) การพัฒนาโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ และ 4) การเปรียบเทียบความสามารถของผู้สอบแต่ละระดับชั้นของค่าอำนาจจำแนกของข้อสอบ สำหรับการทดสอบ O-NET วิชาคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 จำแนกตามเพศและความสามารถด้านการเรียนวิชาคณิตศาสตร์ โดยใช้กลุ่มตัวอย่างเป็นนักเรียนโรงเรียนสาธิตมหาวิทยาลัยราชภัฏบ้านสมเด็จเจ้าพระยา ชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2560 ซึ่งคัดเลือกตัวอย่างด้วยวิธีการสุ่มอย่างง่าย (Simple Random Sampling) จำนวน 80 คน โดยวิเคราะห์ข้อมูลด้วย Two-Way ANOVA

สรุปผลการวิจัย

1. ผลการพัฒนาวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา เริ่มจากการรวบรวมหรือการสร้างข้อสอบ จนถึงวิธีการจัดคลังข้อสอบ

ผลการพัฒนาวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ระดับชั้นร่วมกับการจัดสมดุลเนื้อหา โดยข้อสอบที่นำมาบรรจุไว้ในคลังข้อสอบต้องผ่านการวิเคราะห์คุณภาพข้อสอบตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ (MIRT) และผ่านเกณฑ์การคัดเลือกข้อสอบตามที่กำหนดมาแล้ว มีรายละเอียดดังนี้

1.1 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

1.2 การพัฒนาคลังข้อสอบเริ่มจากแบ่งชั้นคลังข้อสอบเป็น 4 ชั้นตามระดับ

ค่าอำนาจจำแนกของข้อสอบร่วมกับการจัดสมดุลเนื้อหาในแต่ละชั้น โดยในแต่ละชั้นได้จัดกลุ่มเนื้อหาข้อสอบ O-NET วิชาคณิตศาสตร์ โดยจำแนกเป็น 4 กลุ่มสาระการเรียนรู้ที่ออกข้อสอบ ได้แก่ 1) จำนวนและการดำเนินการ 2) การวัด 3) พีชคณิต และ 4) การวิเคราะห์ข้อมูลและความน่าจะเป็น

1.3 วิธีการควบคุมการแสดงข้อสอบด้วยวิธีการสุ่มตัวอย่างแบบชั้นภูมิ (Stratified Random Sampling) โดยให้กลุ่มสาระการเรียนรู้เป็นชั้นภูมิ และให้ค่าความยากของข้อสอบ (b) ที่มีลักษณะเหมือนกัน (Homogenious) เป็นหน่วยการสุ่ม ขั้นตอนการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ระดับชั้นร่วมกับการจัดสมดุลเนื้อหาและการควบคุมการแสดงข้อสอบโดยใช้วิธีการสุ่มตัวอย่างแบบชั้นภูมิ

2. ผลการเปรียบเทียบประสิทธิภาพของวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา (IACB) กับวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบถ่วงน้ำหนักที่มีการบังคับ (CWA) สรุปผลการวิจัยได้ดังนี้

ผลการเปรียบเทียบประสิทธิภาพของวิธีการจัดคลังข้อสอบ วิธี IACB กับ วิธี CWA ในภาพรวม ปรากฏว่า วิธี IACB มีประสิทธิภาพสูงกว่า วิธี CWA เมื่อพิจารณาประสิทธิภาพด้านปรากฏว่า

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

2.1.1 ค่าความลำเอียงเฉลี่ย ปรากฏว่า วิธี IACB มีค่าน้อย 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

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

2.2 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

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

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

2.2.3 อัตราการทับซ้อนของข้อสอบ ปรากฏว่า วิธี IACB มีค่าน้อย วิธี CWA (3:1) แสดงให้เห็นว่า วิธี IACB ประสิทธิภาพด้านจำนวนข้อสอบที่มีอัตราการทับซ้อนข้อสอบสูงกว่า วิธี CWA ในบางชั้นของคลังข้อสอบ ไม่แตกต่างกัน เป็นไปตามสมมติฐานข้อที่ 2

2.2.4 การแจกแจงอัตราการแสดงข้อสอบ วิธี IACB มีค่าน้อย วิธี CWA (3:1) แสดงให้เห็นว่า วิธี IACB มีประสิทธิภาพด้านจำนวนข้อสอบที่มีการแจกแจงอัตราการแสดงข้อสอบสูงกว่า วิธี CWA ในบางชั้นของคลังข้อสอบ ไม่แตกต่างกัน เป็นไปตามสมมติฐานข้อที่ 2

3. ผลการพัฒนาโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติสำหรับการทดสอบ O-NET วิชาคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 6

3.1 ผลการพัฒนาโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติสำหรับการทดสอบ O-NET วิชาคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 พัฒนาอยู่ในรูปของ Web Application จึงสามารถใช้งานผ่านอุปกรณ์ทั่วไปที่เชื่อมต่ออินเทอร์เน็ตได้ ทั้งนี้โปรแกรมแบ่งการใช้งานได้เป็น 3 ส่วน คือ 1) ข้อมูลทั่วไปเกี่ยวกับโปรแกรม เป็นส่วนที่อธิบายส่วนประกอบต่าง ๆ ของโปรแกรม 2) การจัดการของผู้ดูแลระบบ เป็นส่วนที่ครู อาจารย์ หรือผู้ที่เกี่ยวข้องในการจัดการข้อสอบ นำข้อสอบที่ผ่านการวิเคราะห์คุณภาพข้อสอบตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ (Multidimensional Item Response Theory: MIRT) และจัดข้อสอบตามชั้นของค่าอำนาจจำแนกของข้อสอบและตามกลุ่มเนื้อหาสาระการเรียนรู้ มาใส่ในโปรแกรมเพื่อใช้เป็นคลังข้อสอบ รวมทั้งการลบ และการแก้ไขข้อสอบเดิมที่มีอยู่ในคลังข้อสอบ นอกจากนี้ยังมีรายงานผลการทดสอบของผู้สอบทั้งหมด เพื่อให้ครู อาจารย์ และผู้เกี่ยวข้องนำผลการทดสอบไปใช้ในการประเมินผลการเรียนรู้ต่อไป 3) การจัดการสอบ เป็นส่วนที่ให้ผู้สอบใช้โปรแกรมนี้เพื่อทดสอบความสามารถทางวิชาคณิตศาสตร์ เมื่อผู้สอบทดสอบเสร็จสิ้น จะรายงานผลการทดสอบให้ผู้สอบทราบทันที

3.2 ผลการประเมินความเหมาะสมของโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ สำหรับการทดสอบ O-NET วิชาคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 มีดังนี้

3.2.1 ผลการประเมินความเหมาะสมของโปรแกรมโดยผู้เชี่ยวชาญ ปรากฏว่า โปรแกรมมีความเหมาะสมอยู่ในระดับมากที่สุด 2 ด้าน ได้แก่ ความต้องการของผู้ใช้โปรแกรม และความชัดเจนของคู่มือการใช้งาน ส่วนความเหมาะสมของโปรแกรม 3 ด้าน ได้แก่ การทำงานของโปรแกรม การใช้งาน และการรักษาความปลอดภัยมีความเหมาะสมอยู่ในระดับมาก

3.2.2 ผลการประเมินความเหมาะสมของโปรแกรมโดยผู้ใช้งาน ปรากฏว่า โปรแกรมมีความเหมาะอยู่ในระดับมากที่สุดในด้านความชัดเจนของคู่มือการใช้งาน ส่วนด้านการใช้งาน โปรแกรมมีความเหมาะสมอยู่ในระดับมาก

4. ผลการเปรียบเทียบความสามารถของผู้สอบแต่ละระดับชั้นของค่าอำนาจจำแนกของข้อสอบ โดยใช้โปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ สำหรับการทดสอบ O-NET วิชาคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 จำแนกตามเพศและความสามารถด้านการเรียนวิชาคณิตศาสตร์

4.1 นักเรียนที่ทดสอบด้วยโปรแกรมจำนวน 80 คน จำแนกตามตัวแปรเพศ เป็นเพศชาย 40 คน เพศหญิง 40 คน และจำแนกตามความสามารถด้านการเรียนวิชาคณิตศาสตร์ เป็นกลุ่มสูง ($GPAX_{math} \geq 3.50$) และ กลุ่มต่ำ ($GPAX_{math} < 3.50$) กลุ่มละ 40 คน โดยจัดกลุ่มชายหญิงภายในกลุ่มสูงและกลุ่มต่ำ กลุ่มละ 20 คน มีจำนวนข้อสอบที่ตอบถูกเฉลี่ยเท่ากับ 5.73 ข้อ และใช้เวลาในการทดสอบเฉลี่ยเท่ากับ 22.87 นาที

4.2 ผลการเปรียบเทียบความสามารถของผู้สอบแต่ละระดับชั้นของค่าอำนาจจำแนกของข้อสอบ โดยใช้โปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ สำหรับการทดสอบ O-NET วิชาคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 จำแนกตามเพศและความสามารถด้านการเรียนวิชาคณิตศาสตร์ ปรากฏว่า

4.2.1 เพศมีผลต่อความสามารถของผู้สอบไม่แตกต่างกัน

4.2.2 ผลร่วมกันระหว่างเพศกับความสามารถด้านการเรียนวิชาคณิตศาสตร์ของนักเรียนมีผลต่อความสามารถของผู้สอบไม่แตกต่างกัน

4.2.3 นักเรียนที่มีความสามารถด้านการเรียนวิชาคณิตศาสตร์กลุ่มสูง มีความสามารถของผู้สอบสูงกว่า นักเรียนที่มีความสามารถด้านการเรียนวิชาคณิตศาสตร์กลุ่มต่ำ อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ในทุกระดับชั้นของค่าอำนาจจำแนกของข้อสอบ

ซึ่งตัวแปรที่สอดคล้องกับสมมติฐานคือความสามารถด้านการเรียนวิชาคณิตศาสตร์ แต่ตัวแปรเพศไม่สอดคล้องกับสมมติฐานที่ตั้งไว้

อภิปรายผล

การอภิปรายผลการวิจัยแบ่งได้ 4 ประเด็น ได้แก่ 1) การพัฒนาวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา 2) การเปรียบเทียบประสิทธิภาพของวิธีการจัดคลังข้อสอบ 3) การพัฒนาโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ และ 4) การเปรียบเทียบความสามารถของผู้สอบแต่ละระดับชั้นของค่าอำนาจจำแนกของข้อสอบ สำหรับการทดสอบ O-NET วิชาคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 จำแนกตามเพศและความสามารถด้านการเรียนวิชาคณิตศาสตร์ รายละเอียดมีดังนี้

1. การพัฒนาวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา

การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ (Multidimensional Computerized Adaptive Testing: MCAT) เป็นการทดสอบที่มีการจัดข้อสอบให้เหมาะสมกับความสามารถของผู้สอบ ซึ่งพิจารณาจากผลการตอบข้อสอบในข้อที่ผ่านมา กล่าวคือเมื่อผู้สอบทำข้อสอบข้อแรกหรือชุดเริ่มต้นเสร็จลง จะนำผลการตอบมาวิเคราะห์ระดับความสามารถของผู้สอบเพื่อใช้คัดเลือกข้อสอบข้อถัดไปที่มีความเหมาะสมกับความสามารถของผู้สอบ ข้อสอบต้องสร้างตาม

ตารางกำหนดเนื้อหาและควรมีการตรวจสอบคุณภาพของเนื้อหาและความไวของแบบทดสอบ (Test Sensitivity) และมีการทดลองใช้เพื่อศึกษาคุณลักษณะทางจิตวิทยา (Wainer et al., 2001, pp. 105-112) วิธีการตรวจสอบคุณภาพของข้อสอบสามารถใช้วิธีการทางสถิติของทฤษฎีการทดสอบแบบดั้งเดิม เช่น สัดส่วนของผู้ที่ตอบถูก (Proportion Correct) หรือ สหสัมพันธ์แบบไบซีเรียล (Biserial Correlation) ร่วมกับทฤษฎีการทดสอบแนวใหม่ เช่น พารามิเตอร์ของข้อสอบ (Item Parameters) และค่าสารสนเทศของข้อสอบ (Item Information) (Wainer, 1989) และสิ่งที่สำคัญอีกประการหนึ่ง คือ คุณลักษณะของข้อสอบ (Item Characterisitic) เพื่อตรวจสอบว่าข้อสอบแต่ละข้อเหมาะสมกับโมเดลการวัด ซึ่งการพัฒนาคลังข้อสอบสำหรับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ (Wainer et al., 2001, pp. 37-45; Thompson & Weiss, 2011) โดยใช้ทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ (Multidimensional Item Response Theory: MIRT) เป็นพื้นฐาน ขั้นตอนการทดสอบแบบ MCAT มี 5 ขั้นตอน คือ ขั้นตอนที่ 1 การสร้างคลังข้อสอบ (Create Item Pools) เป็นการจัดทำทำข้อสอบที่ใช้ในการทดสอบ ซึ่งต้องสอดคล้องกับทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ (MIRT) ขั้นตอนที่ 2 การคัดเลือกข้อสอบข้อแรก (First Item Selection) เป็นการคัดเลือกข้อสอบข้อแรกโดยส่วนใหญ่ใช้ข้อสอบที่มีค่าความยากของข้อสอบระดับปานกลาง ขั้นตอนที่ 3 การคัดเลือกข้อสอบข้อถัดไป (Next Item Selection) โดยคัดเลือกข้อสอบให้มีความเหมาะสมกับระดับความสามารถของผู้สอบ ซึ่งพิจารณาจากผลการตอบข้อสอบในข้อที่ผ่านมา ขั้นตอนที่ 4 การประมาณค่าความสามารถของผู้สอบ (Calculate Possible Ability Level) โดยอาศัยผลการตอบข้อสอบในข้อที่ผ่านมา และขั้นตอนที่ 5 เกณฑ์การยุติการทดสอบ (Termination Criterion) เป็นการกำหนดเกณฑ์สิ้นสุดการทดสอบ โดยการทดสอบการทดสอบจะดำเนินการซ้ำตั้งแต่ขั้นตอนที่ 3 ถึง ขั้นตอนที่ 5 จนกระทั่งการทดสอบเป็นไปตามเกณฑ์ยุติการทดสอบที่กำหนดไว้ ซึ่งคลังข้อสอบเป็นเสมือนแหล่งรวบรวมชุดแบบทดสอบจำนวนมากและมีความเป็นคู่ขนานกัน ข้อสอบบรรจุในคลังข้อสอบมีลักษณะค่าความยากกระจายเต็มระดับความสามารถของประชากรผู้สอบ ค่าอำนาจจำแนกสูงเพื่อให้ประสิทธิภาพการทดสอบสูงสุด และค่าโอกาสในการเดาเข้าใกล้หรือเท่ากับศูนย์ ข้อสอบควรมีจำนวนมากเพียงพอและกระจายทุกระดับความสามารถของผู้สอบ ดังนั้น ต้องใช้ข้อสอบจำนวนมากสำหรับสร้างคลังข้อสอบทำให้คลังข้อสอบสำหรับการทดสอบแบบปรับเหมาะมีขนาดใหญ่กว่าการทดสอบดั้งเดิม (Thompson & Weiss, 2011; Babcock & Weiss, 2012)

การควบคุมการแสดงข้อสอบ ส่วนใหญ่จะคัดเลือกข้อสอบที่มีค่าอำนาจจำแนกข้อสอบสูงมาเป็นข้อสอบข้อถัดไป ดังนั้น ข้อสอบเหล่านี้จึงถูกเลือกใช้ในการทดสอบบ่อยครั้งจนเกินไป เป็นผลให้ผู้สอบรุ่นถัดไป สามารถคาดการณ์ข้อสอบที่ตนเองจะได้รับล่วงหน้าได้ ในการการวิจัยนี้ ได้พัฒนาวิธีการควบคุมการแสดงข้อสอบโดยใช้วิธีการสุ่มตัวอย่างแบบชั้นภูมิ (Stratified Random Sampling) มาควบคุมการแสดงข้อสอบ โดยใช้ข้อสอบที่นำไปใช้ในขั้นตอนการคัดเลือกข้อสอบข้อถัดไป คือ ข้อสอบที่มีค่าอำนาจจำแนกของข้อสอบต่ำถึงปานกลาง และข้อสอบที่มีค่าอำนาจจำแนกของข้อสอบสูง ซึ่งถูกสุ่มมาจากกลุ่มสาระการเรียนรู้ในแต่ละชั้นที่มีการจัดสมดุลเนื้อหาตามกลุ่มสาระการเรียนรู้ที่ออกข้อสอบ O-NET เป็น 4 กลุ่มสาระการเรียนรู้ ได้แก่ 1) จำนวนและการดำเนินการ 2) การวัด3) พีชคณิต และ 4) การวิเคราะห์ข้อมูลและความน่าจะเป็น ด้วยวิธีการสุ่มตัวอย่างแบบชั้นภูมิ โดยให้กลุ่มสาระเป็นชั้นภูมิ และให้ค่าความยากของข้อสอบ (b) ที่มีลักษณะเหมือนกัน

(Homogenious) เป็นหน่วยการสุ่ม ด้วยกระบวนการนี้ ผู้สอบที่มีความสามารถเท่ากันหรือใกล้เคียงกัน จึงมีโอกาสได้รับข้อสอบข้อเดียวกันน้อยลง เนื่องจาก ข้อสอบที่มีค่าอำนาจจำแนกของข้อสอบสูงถูกหมุนเวียน (จากการสุ่ม) นำไปใช้ในขั้นตอนการคัดเลือกข้อสอบข้อถัดไป วิธีการนี้จึงสามารถลด จำนวนข้อสอบที่มีการแสดงมากเกินไป จำนวนข้อสอบที่มีการใช้น้อยเกินไป อัตราการทับซ้อนของข้อสอบ และ การแจกแจงอัตราการแสดงข้อสอบลงได้ (Georgiadou, Triantafillou and Economides, 2007) สอดคล้องกับงานวิจัยของ Ozturk and Dogan (2015) ที่ได้เปรียบเทียบประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบและการควบคุมการแสดงข้อสอบระหว่างวิธีการควบคุมการใช้ข้อสอบ 3 วิธี ได้แก่ 1) วิธี Randomesque 2) วิธี Sympson-Hetter และ 3) วิธี Fade-Away โดยใช้คลังข้อสอบจำลอง ด้วยการสุ่มค่าอำนาจจำแนกของข้อสอบ (a) จากการแจกแจงแบบยูนิฟอร์มในช่วง 0.50 ถึง 2.00 สุ่มค่าโอกาสในการเดาของข้อสอบ (c) จากการแจกแจงแบบยูนิฟอร์มในช่วง 0.05 ถึง 0.20 ส่วนการจำลองค่าความยากของข้อสอบ (b) แบ่งได้ 2 ส่วน คือ 1) คลังข้อสอบที่มีความยากของข้อสอบปานกลาง ด้วยการสุ่มค่า b จากการแจกแจงแบบยูนิฟอร์มในช่วง -3.00 ถึง 3.00 และ 2) คลังข้อสอบที่มีค่าความยากของข้อสอบสูงด้วยการสุ่มค่า b จากการแจกแจงปกติ N(2, 1.5) ผลการศึกษาปรากฏว่า ประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบทั้ง 3 วิธี ไม่มีความแตกต่างกัน แต่วิธี Fade-Away สามารถควบคุมการแสดงข้อสอบได้ดีกว่าวิธี Randomesque และวิธี Sympson-Hetter

ประเด็นหนึ่งที่สำคัญในการพัฒนาวิธีการควบคุมการแสดงข้อสอบโดยใช้วิธีการสุ่มตัวอย่างแบบชั้นภูมิ (Stratified Random Sampling) คือ จำนวนข้อสอบที่สุ่มมาจากกลุ่มเนื้อหาในแต่ละชั้นที่มีค่าอำนาจจำแนกของข้อสอบสูงด้วยวิธีการสุ่มตัวอย่างแบบชั้นภูมิ ควรสุ่มข้อสอบจำนวนเท่าๆ กันจึงเหมาะสม โดยในช่วงเริ่มต้นการพัฒนา ได้ทดลองกำหนดจำนวนข้อสอบในการสุ่มไว้ที่ 10% ของจำนวนข้อสอบในแต่ละกลุ่มสาระการเรียนรู้ที่มีค่าอำนาจจำแนกของข้อสอบปานกลางถึงสูงทั้งหมดที่มีในคลังข้อสอบในแต่ละชั้น ปรากฏว่า เมื่อกำหนดจำนวนข้อสอบสุ่มมากขึ้น มีผลทำให้ข้อสอบที่มีอัตราการใช้ข้อสอบสูงกว่า 0.2 และจำนวนข้อสอบที่มีการใช้น้อยเกินไป อัตราการทับซ้อนของข้อสอบ และการแจกแจงอัตราการแสดงข้อสอบ มีแนวโน้มลดลง (บ่งชี้ว่า จำนวนข้อสอบถูกใช้ซ้ำบ่อยลดลง ซึ่งเป็นข้อดี) แต่ในขณะเดียวกันค่า RMSE และจำนวนข้อสอบที่ใช้ในการทดสอบกับมีแนวโน้มเพิ่มขึ้น (บ่งชี้ว่า ประสิทธิภาพด้านการประมาณค่าความสามารถของผู้สอบลดลง และใช้ข้อสอบในการทดสอบเพิ่มขึ้น ซึ่งเป็นข้อเสีย) ซึ่งเป็นผลมาจากเมื่อกำหนดจำนวนข้อสอบที่สุ่มไปใช้ในขั้นตอนการคัดเลือกข้อสอบข้อถัดไปให้มีจำนวนมากขึ้น ข้อสอบที่ไม่เหมาะสมกับผู้สอบ เช่น ข้อสอบที่มีค่าความยากของข้อสอบต่างจากค่าความสามารถของผู้สอบมาก จึงมีโอกาสถูกคัดเลือกไปให้ผู้สอบเพิ่มมากขึ้น ดังนั้น การพัฒนาวิธีการควบคุมการแสดงข้อสอบและการจัดสมดุลเนื้อหา นี้ จึงกำหนดจำนวนข้อสอบสุ่ม 4 ข้อ จาก 4 กลุ่มสาระการเรียนรู้ที่ออกข้อสอบ โดยอาศัยแนวคิดจากวิธีการ Randomesqus (Kingsbury & Zara, 1989)

2. การเปรียบเทียบประสิทธิภาพของวิธีการจัดคลังข้อสอบ

ผลการเปรียบเทียบการประมาณค่าความสามารถของผู้สอบ ได้แก่ ค่าความลำเอียงเฉลี่ยและค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย ปรากฏว่า วิธี IACB มีค่าน้อย วิธี CWA (8:0) แสดงให้เห็นว่า วิธี IACB มีประสิทธิภาพการประมาณค่าความสามารถของผู้สอบสูงกว่า วิธี CWA

เป็นไปตามสมมติฐานข้อที่ 1 เมื่อพิจารณารายข้อ ปรากฏว่า 1) ค่าความลำเอียงเฉลี่ย ปรากฏว่า วิธี IACB มีค่าน้อย 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 และค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย ปรากฏว่า วิธี IACB มีค่าน้อย วิธี CWA (4:0) แสดงให้เห็นว่า วิธี IACB มีประสิทธิภาพการประมาณค่าความสามารถของผู้สอบสูงกว่าสูงกว่า วิธี CWA ในคลังข้อสอบชั้นที่ 1 และ 3 อย่างมีนัยสำคัญทางสถิติที่ .05 และในคลังข้อสอบชั้นที่ 2 และ 4 อย่างมีนัยสำคัญทางสถิติที่ .01 เป็นไปตามสมมติฐานข้อที่ 1 เมื่อพิจารณา วิธีการจัดคลังข้อสอบวิธี IACB มีประสิทธิภาพด้านการประมาณค่าความสามารถของผู้สอบสูงกว่าวิธี CWA ที่พิจารณาเพียงค่าสารสนเทศที่ใกล้เคียงกับความสามารถของผู้สอบ ณ ขณะนั้น (Taylor, 2009, pp. 538-539; Cheng et al., 2009; Finch, 2010; Zhou, 2012) ซึ่งให้ความคลาดเคลื่อนกำลังสองเฉลี่ยและความลำเอียงเฉลี่ยมีแนวโน้มลดลง เหมือนกันเมื่อความยาวแบบทดสอบเพิ่มขึ้น แต่ความคลาดเคลื่อนกำลังสองเฉลี่ยในวิธี IACB ต่ำกว่า วิธี CWA

ผลการเปรียบเทียบประสิทธิภาพของการใช้ข้อสอบ 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 1) ข้อสอบที่มีการแสดงมากเกินไป ปรากฏว่า วิธี IACB มีค่าน้อย วิธี CWA (4:0) แสดงให้เห็นว่า วิธี IACB มีประสิทธิภาพด้านการใช้ข้อสอบที่มีการแสดงมากเกินไปสูงกว่า วิธี CWA ในทุกชั้นของคลังข้อสอบอย่างมีนัยสำคัญทางสถิติที่ .01 เป็นไปตามสมมติฐานข้อที่ 2 2) ข้อสอบที่มีการใช้น้อยเกินไป ปรากฏว่า วิธี CWA มีค่าน้อย วิธี IACB (4:0) แสดงให้เห็นว่า วิธี CWA มีประสิทธิภาพด้านการใช้จำนวนข้อสอบที่มีใช้น้อยเกินไปสูงกว่า วิธี IACB ในทุกชั้นของคลังข้อสอบอย่างมีนัยสำคัญทางสถิติที่ .01 ไม่เป็นไปตามสมมติฐานข้อที่ 2 3) อัตราการทับซ้อนของข้อสอบ ปรากฏว่า วิธี IACB มีค่าน้อย วิธี CWA (3:1) แสดงให้เห็นว่า วิธี IACB ประสิทธิภาพด้านจำนวนข้อสอบที่มีอัตราการทับซ้อนข้อสอบสูงกว่า วิธี CWA ในบางชั้นของคลังข้อสอบ ไม่แตกต่างกัน เป็นไปตามสมมติฐานข้อที่ 2 และ 4) การแจกแจงอัตราการแสดงข้อสอบ วิธี IACB มีค่าน้อย วิธี CWA (3:1) แสดงให้เห็นว่า วิธี IACB มีประสิทธิภาพด้านจำนวนข้อสอบที่มีการแจกแจงอัตราการแสดงข้อสอบสูงกว่า วิธี CWA ในบางชั้นของคลังข้อสอบ ไม่แตกต่างกัน เป็นไปตามสมมติฐานข้อที่ 2 ซึ่งสามารถควบคุมประสิทธิภาพของการใช้ข้อสอบ ได้แก่ ข้อสอบที่มีการแสดงมากเกินไป ข้อสอบที่มีการใช้น้อย อัตราการทับซ้อนของข้อสอบ และการแจกแจงอัตราการแสดงข้อสอบ ให้มีจำนวนลดลงได้ โดยส่วนใหญ่เป็นข้อสอบที่มีค่าอำนาจำแนกของข้อสอบระดับปานกลางถึงสูง วิธีการนี้ทำให้ข้อสอบที่มีความค่าอำนาจจำแนกของข้อสอบปานกลางถึงสูงมีโอกาสถูกเลือกใช้ในการทดสอบสูงขึ้น โดยการสุ่มข้อสอบที่มีค่าอำนาจจำแนกของข้อสอบปานกลางถึงสูง ด้วยวิธีการสุ่มตัวอย่างแบบชั้นภูมิ (Stratified Random Sampling) มาควบคุมการแสดงข้อสอบ โดยใช้ข้อสอบที่นำไปใช้ในขั้นตอนการคัดเลือกข้อสอบข้อถัดไป คือ ข้อสอบที่มีค่าอำนาจจำแนกของข้อสอบต่ำถึงปานกลาง และข้อสอบที่มีค่าอำนาจจำแนกของข้อสอบสูง ซึ่งถูกสุ่มมาจากกลุ่มสาระการเรียนรู้ในแต่ละชั้นที่มีการจัดสมดุลเนื้อหาตามกลุ่มสาระ

การเรียนรู้ที่ออกข้อสอบ O-NET เป็น 4 กลุ่มสาระการเรียนรู้ที่ออกข้อสอบ ได้แก่ 1) จำนวนและการดำเนินการ 2) การวัด3) พีชคณิต และ 4) การวิเคราะห์ข้อมูลและความน่าจะเป็น ด้วยวิธีการสุ่มตัวอย่างแบบชั้นภูมิ โดยให้กลุ่มสาระการเรียนรู้เที่ออกข้อสอบเป็นชั้นภูมิ และให้ค่าความยากของข้อสอบ (b) ที่มีลักษณะเหมือนกัน (Homogenious) เป็นหน่วยการสุ่ม ด้วยกระบวนการนี้ ผู้สอบที่มีความสามารถเท่ากันหรือใกล้เคียงกัน จึงมีโอกาสได้รับข้อสอบข้อเดียวกันน้อยลง เนื่องจากข้อสอบที่มีค่าอำนาจจำแนกของข้อสอบสูงถูกหมุนเวียน (จากการสุ่ม) นำไปใช้ในขั้นตอนการคัดเลือกข้อสอบข้อถัดไป วิธีการนี้จึงสามารถลดจำนวนข้อสอบที่มีการแสดงมากเกินไป จำนวนข้อสอบที่มีการใช้น้อย อัตราการทับซ้อนของข้อสอบ และการแจกแจงอัตราการแสดงข้อสอบลงได้ สอดคล้องกับงานวิจัยของ Ozturk and Dogan (2015) ที่ได้เปรียบเทียบประสิทธิภาพของวิธีการควบคุมการแสดงข้อสอบ ปรากฏว่า วิธีการ Randomesque เป็นวิธีการควบคุมการแสดงข้อสอบด้วยวิธีการสุ่ม สามารถลดจำนวนข้อสอบที่มีการแสดงมากเกินไป จำนวนข้อสอบที่มีการใช้น้อย อัตราการทับซ้อนของข้อสอบ และ การแจกแจงอัตราการแสดงข้อสอบให้น้อยลงได้

3. การพัฒนาโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ

การพัฒนาโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติในงานวิจัยนี้ เป็นโปรแกรมสำหรับการทดสอบ O-NET วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 6 มีลักษณะเป็น Web Application จึงสามารถใช้ผ่านอุปกรณ์ทั่วไปที่เชื่อมต่อกับอินเทอร์เน็ตได้ เช่น คอมพิวเตอร์ โน๊ตบุ๊คหรือสมาร์ทโฟน โปรแกรมการทดสอบนี้ แบ่งออกเป็น 3 ส่วน คือ 1) ข้อมูลทั่วไปเกี่ยวกับโปรแกรม 2) การจัดการของผู้ดูแลระบบ และ 3) การจัดการสอบ เมื่อผู้สอบทดสอบเสร็จสิ้น จะรายงานผลการทดสอบให้ผู้สอบทราบทันที

โปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติที่พัฒนาขึ้นนี้ แต่ละขั้นตอนมี จุดเด่น ดังนี้ ขั้นตอนที่ 1 การสร้างคลังข้อสอบ โปรแกรมนี้รองรับข้อสอบทีมีลักษณะการตรวจให้คะแนน 2 ค่า คือ ถูก-ผิด หรือ 0-1 และมีรายการคำตอบ 4 หรือ 5 รายการ และสร้างข้อสอบตามแนวคิดของ Bloom et al. (1956); Anderson et al. (2001, pp. 67-68) และ Krathwohl, 2002, pp. 213-215) ซึ่งได้เพิ่มโครงสร้างจากมิติเดียวเป็นสองมิติ ทำให้โครงสร้างใหม่มีลักษณะเป็นสองมิติที่ประกอบด้วยมิติด้านความรู้ และมิติด้านกระบวนการพุทธิปัญญา โดยมิติด้านความรู้ 3 ส่วน คือ ความรู้เกี่ยวกับข้อเท็จจริง มโนทัศน์ และวิธีดำเนินการ ตามลำดับ และมิติด้านกระบวนการพุทธิปัญญา 4 กระบวนการ คือ จำ เข้าใจ การประยุกต์ และวิเคราะห์ ตามลำดับ แล้วนำมาวิเคราะห์หาคุณภาพของข้อสอบโดยใช้ทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ (MIRT) แบบ 3 พารามิเตอร์ ประกอบด้วย ค่าความยากของข้อสอบตั้งแต่ -2.50 ถึง 2.50 ค่าอำนาจจำแนกของข้อสอบตั้งแต่ 0.50 ถึง 2.50 และค่าโอกาสในการเดาของข้อสอบไม่เกิน 0.30 (Urry, 1977) และการพัฒนาวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา ผู้วิจัยประยุกต์จากวิธีจัดคลังข้อสอบตามระดับชั้นของค่าอำนาจจำแนกของข้อสอบ (a-Stratified Method: a-STR) ที่พัฒนาโดย Chang and Ying (1996) ซึ่งแบ่งคลังข้อสอบออกเป็น 4 ชั้น ตามช่วงของค่าอำนาจจำแนกของข้อสอบ (a) 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 ร่วมกับการจัดสมดุลเนื้อหา (Content Balancing) ภายในชั้นของค่าอำนาจจำแนก ได้แบ่งเป็น 4 กลุ่มสาระการเรียนรู้ที่การ

ออกข้อสอบ O-NET วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 6 ได้แก่ 1) จำนวนและการดำเนินการ 2) การวัด 3) พีชคณิต และ 4) การวิเคราะห์และความน่าจะเป็น ดังนั้นวิธีการควบคุมการแสดงข้อสอบ (Controlling Item Exposure) จึงใช้วิธีการสุ่มตัวอย่างแบบชั้นภูมิ (Stratified Random Sampling) โดยให้กลุ่มสาระการเรียนรู้เนื้อหาการออกข้อสอบ เป็นชั้นภูมิ และให้ค่าความยากของข้อสอบ (b) ที่มีลักษณะเหมือนกัน (Homogenious) เป็นหน่วยการสุ่ม เป็นวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา สำหรับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ ขั้นตอนที่ 2 การคัดเลือกข้อสอบข้อแรก คัดเลือกโดยสุ่มข้อสอบในคลังข้อสอบที่มีค่าความยากของข้อสอบระหว่าง -1.0 ถึง 1.0 ซึ่งวิธีการนี้มีประสิทธิภาพสูงกว่าวิธีการเลือกข้อสอบข้อใดข้อหนึ่ง ที่มีค่าความยากของข้อสอบปานกลาง มาเป็นข้อสอบข้อแรก เพราะสามารถลดปัญหาข้อสอบข้อแรกถูกเลือกไปใช้บ่อยครั้งเกินไปได้ เมื่อเทียบกับการเลือกข้อสอบที่มีความยากของข้อสอบระดับปานกลางข้อใดข้อหนึ่ง ขั้นตอนที่ 3 การคัดเลือกข้อสอบข้อถัดไปพิจารณาจากค่าความยากของข้อสอบ (b) ข้อก่อนหน้า โดยใช้วิธีควบคุมการแสดงข้อสอบด้วยวิธีการสุ่มตัวอย่างแบบชั้นภูมิ ที่คำนึงถึงการจัดสมดุลเนื้อหาข้อสอบ โดยให้กลุ่มสาระการเรียนรู้ที่ออกข้อสอบเป็นชั้นภูมิ และให้ค่าความยากของข้อสอบ (b) ที่มีลักษณะเหมือนกัน (Homogenious) เป็นหน่วยการสุ่ม โดยการบังคับกลุ่มสาระการเรียนรู้ที่ถูกเลือกไปให้ผู้สอบแล้วไม่ให้ถูกเลือกขึ้นมาซ้ำ ขั้นตอนที่ 4 การประมาณค่าความสามารถของผู้สอบ เลือกใช้วิธีการประมาณค่าแบบเบส์ (Bayesian Estimation) ซึ่งเป็นวิธีที่ได้รับความนิยมนำไปใช้อย่างแพร่หลายและเป็นวิธีที่สามารถคำนวณประมาณค่าความสามารถของผู้สอบได้ ในกรณีที่ผู้สอบตอบถูกทุกข้อหรือตอบผิดทุกข้อ ซึ่งวิธีการอื่นไม่สามารถคำนวณค่าได้ และ ขั้นตอนที่ 5 เกณฑ์การยุติการทดสอบในแต่ละชั้นและเลื่อนชั้นการทดสอบ ผู้สอบต้องทดสอบในชั้นนั้นๆให้ครบ 4 ข้อ ตามเนื้อหาสาระการเรียนรู้ที่ออกข้อสอบ O-NET วิชาคณิตศาสตร์ หรือ ค่าความคลาดเคลื่อนมาตรฐาน น้อยกว่าหรือเท่ากับ 0.30 (n=4, หรือ SE ≤ 0.30) ซึ่งเป็นเกณฑ์ที่มีความเชื่อถือและลดจำนวนข้อสอบที่ใช้ในการทดสอบได้มาก (Thompson & Weiss, 2011; Wainer et al., 2001, pp. 105-113)

การจัดคลังข้อสอบที่ใช้ในโปรแกรมนี้ นอกจากต้องคำนึงถึงการควบคุมการแสดงข้อสอบ ยังต้องคำนึงถึงการควบคุมการจัดสมดุลเนื้อหาของข้อสอบ (Content Balancing) เพื่อป้องกันไม่ให้ผู้สอบได้รับข้อสอบที่มีเนื้อหา (สาระการเรียนรู้ที่ออกข้อสอบ) ด้านใดด้านหนึ่งเพียงด้านเดียว โดยเฉพาะอย่างยิ่งการวัดผลการเรียนรู้ของผู้เรียน ซึ่งมีเนื้อหาในการเรียนที่หลากหลาย ในโปรแกรมนี้ควบคุมอัตราการแสดงข้อสอบด้วยวิธีการสุ่มตัวอย่างแบบชั้นภูมิ (Stratified Random Sampling) ดังนั้น โปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติที่พัฒนาขึ้นนี้ จึงมีประสิทธิภาพสูงเมื่อนำไปใช้จริง ปรากฏว่า การทดสอบด้วยโปรแกรมนี้มีความยาวของแบบทดสอบค่อนข้างน้อย (16 ข้อ) ผู้สอบจะได้รับการทดสอบด้วยข้อสอบในแต่ละชั้นของคลังข้อสอบ ชั้นละ 4 ข้อ ประกอบด้วย 4 สาระการเรียนรู้ที่ออกข้อสอบ ได้แก่ 1) จำนวนและการดำเนินการ 2) การวัด 3) พีชคณิต และ 4) การวิเคราะห์ข้อมูลและความน่าจะเป็น ผลการประมาณค่าความสามารถของผู้สอบมีความถูกต้องแม่นยำสูง รวมทั้งผลการประมาณค่าความสามารถของผู้สอบมีความคงเส้นคงวา สอดคล้องกับ Petersen et al., (2009, pp. 315-329) ที่กล่าวว่า มีความเป็นไปได้ที่จะใช้ข้อสอบจำนวน 5-7 ข้อสอบ การประมาณค่า

ความสามารถของผู้สอบได้พร้อมกันในสามมิติและสอดคล้องกับ Frey and Seitz (2009, p. 93) ที่กล่าวว่า การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ที่ใช้คลังข้อสอบที่พัฒนามาจากทฤษฎีการตอบสนองข้อสอบแบบพหุมิติเป็นการทดสอบที่มีประสิทธิภาพสูง สามารถลดจำนวนข้อสอบลงได้มากกว่าการทดสอบด้วยการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ที่ใช้คลังข้อสอบที่พัฒนามาจากทฤษฎีการตอบสนองข้อสอบแบบเอกมิติประมาณ 30-50 % และลดจำนวนข้อสอบได้มากกว่าการทดสอบแบบดั้งเดิมประมาณ 70% โดยไม่สูญเสียความแม่นยำในการทดสอบ

4. การเปรียบเทียบความสามารถของผู้สอบแต่ละระดับชั้นของค่าอำนาจจำแนกของข้อสอบ สำหรับการทดสอบ O-NET วิชาคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 จำแนกตามเพศและความสามารถด้านการเรียนวิชาคณิตศาสตร์

ผลการเปรียบเทียบความสามารถของผู้สอบแต่ละระดับชั้นของค่าอำนาจจำแนกของข้อสอบ โดยใช้โปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ สำหรับการทดสอบ O-NET วิชาคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 จำแนกตามเพศและความสามารถด้านการเรียนวิชาคณิตศาสตร์ ปรากฏว่า 1) เพศมีผลต่อความสามารถของผู้สอบไม่แตกต่างกัน 2) ผลร่วมกันระหว่างเพศกับความสามารถด้านการเรียนวิชาคณิตศาสตร์ของนักเรียนมีผลต่อความสามารถของผู้สอบไม่แตกต่างกัน และ 3) นักเรียนที่มีความสามารถด้านการเรียนวิชาคณิตศาสตร์กลุ่มสูง มีความสามารถของผู้สอบสูงกว่า นักเรียนที่มีความสามารถด้านการเรียนวิชาคณิตศาสตร์กลุ่มต่ำ อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ในทุกระดับชั้นของค่าอำนาจจำแนกของข้อสอบ ซึ่งตัวแปรที่สอดคล้องกับสมมติฐานคือความสามารถด้านการเรียนวิชาคณิตศาสตร์ แต่ตัวแปรเพศไม่สอดคล้องกับสมมติฐานที่ตั้งไว้ ตัวแปรที่สอดคล้องกับสมมติฐานคือความสามารถด้านการเรียนวิชาคณิตศาสตร์ แต่ตัวแปรเพศไม่สอดคล้องกับสมมติฐานที่ตั้งไว้ ซึ่งสอดคล้องกับผลการศึกษาของ OECD (2016, pp. 14-18) ที่พบว่า ใน PISA 2015 สำหรับประเทศไทย นักเรียนชายและนักเรียนหญิงมีคะแนนวิชาคณิตศาสตร์ใกล้เคียงกัน โดยนักเรียนหญิงมีคะแนนสูงกว่านักเรียนชาย 3 คะแนน (ชาย 414 คะแนน และหญิง 417 คะแนน) แต่ความแตกต่างนี้ไม่มีนัยสำคัญทางสถิติ สอดคล้องกับงานวิจัยของ วรรณศิรินทร์ ภูฉายา และคณะ (2561) ที่ได้ศึกษาระดับความสามารถในการแก้ปัญหาทางคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 1-3 ปรากฏว่า ปัจจัยด้านเพศไม่มีความสัมพันธ์ต่อความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ ส่วนนักเรียนที่มีความสามารถด้านการเรียนวิชาคณิตศาสตร์กลุ่มสูง มีความสามารถในการแก้ปัญหาสูงกว่า นักเรียนที่มีความสามารถด้านการเรียนวิชาคณิตศาสตร์กลุ่มต่ำ อย่างมีนัยสำคัญทางสถิติที่ระดับ.01 นอกจากนี้ พบว่าปัจจัยที่มีความสัมพันธ์ต่อความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ คือ เกรดวิชาคณิตศาสตร์ บรรยากาศในชั้นเรียน เจตคติต่อวิชาคณิตศาสตร์ และพฤติกรรมการสอนของครูที่สอนคณิตศาสตร์

ข้อเสนอแนะในการนำผลการวิจัยไปใช้

1. ผู้สนใจพัฒนาวิธีการจัดคลังข้อสอบในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ สามารถนำวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา ไปประยุกต์ใช้กับวิชาอื่น ๆ ได้ ซึ่งวิธีการนี้มีความเหมาะสมกับข้อสอบที่มีค่าอำนาจจำแนกของข้อสอบอยู่ในระดับปานกลาง และระดับสูง

2. ผู้ที่สนใจวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา ควรวิเคราะห์คุณภาพข้อสอบโดยใช้ทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ (MIRT) และคัดเลือกข้อสอบที่ผ่านเกณฑ์วิเคราะห์คุณภาพข้อสอบที่มีค่าอำนาจจำแนกของข้อสอบอยู่ในระดับปานกลางถึงระดับสูง และมีค่าความยากของข้อสอบกระจายตั้งแต่ -2.50 ถึง 2.50 ซึ่งเป็นข้อสอบที่เหมาะสมกับวิธีการจัดคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหา

3. ผู้สนใจพัฒนาวิธีจัดคลังข้อสอบควรคำนึงถึงวิธีควบคุมการแสดงข้อสอบ วิธีการจัดสมดุลเนื้อหา และการพัฒนาโปรแกรมในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ สามารถนำวิธีการพัฒนาวิธีการแสดงข้อสอบโดยใช้วิธีการสุ่มตัวอย่างแบบแบ่งชั้นภูมิ (Stratified Random Sampling) ของการวิจัยนี้ ไปใช้เป็นแนวทางในการพัฒนาวิธีการควบคุมแสดงข้อสอบวิธีการใหม่ที่ใช้วิธีการสุ่มตัวอย่างวิธีการอื่น ๆ ได้ เช่น วิธีการสุ่มเป็นหลายขั้นตอน (Multistage Stage Sampling)

4. ครู อาจารย์ ศึกษานิเทศก์ หรือผู้ที่เกี่ยวข้องสามารถนำข้อสอบที่ผ่านการวิเคราะห์คุณภาพข้อสอบตามทฤษฎีการตอบสนองของสอบแบบพหุมิติ มาจัดเก็บในคลังข้อสอบโดยจัดเก็บตามค่าอำนาจจำแนกของข้อสอบที่เหมาะสมและสามารถนำไปใช้กับโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ สำหรับการทดสอบ O-NET วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 6 ที่ได้จากการวิจัยนี้ ไปใช้วัดความสามารถในการเรียนด้านอื่นๆ หรือนำไปประยุกต์กับการวัดผลการศึกษาในรายวิชาและในระดับอื่น ๆ ได้

ข้อเสนอแนะสำหรับการวิจัยต่อไป

1. การศึกษาในครั้งนี้ เป็นการพัฒนาคลังข้อสอบในรายวิชาคณิตศาสตร์ โดยใช้แนวคิดของ Bloom et al. (1956); Anderson et al. (2001, pp. 67-68) และ Krathwohl, 2002, pp. 213-215) ซึ่งได้เพิ่มโครงสร้างจากมิติเดียวเป็นสองมิติ ทำให้โครงสร้างใหม่มีลักษณะเป็นสองมิติที่ประกอบด้วยมิติด้านกระบวนการพุทธิปัญญา และมิติด้านความรู้ โดยมิติด้านความรู้ 4 ส่วน คือ ความรู้เกี่ยวกับข้อเท็จจริง มโนทัศน์ วิธีดำเนินการ และอภิปัญญา และมิติด้านกระบวนการพุทธิปัญญา 6 กระบวนการ คือ จำ เข้าใจ ประยุกต์ วิเคราะห์ ประเมินค่า และสร้างสรรค์ ในการวิจัยในครั้งนี้ ผู้วิจัยดำเนินการสร้างคลังข้อสอบโดยนำมิติด้านความรู้ 3 ส่วน คือ ความรู้เกี่ยวกับข้อเท็จจริง มโนทัศน์ และวิธีดำเนินการ ส่วนมิติด้านกระบวนการพุทธิปัญญา 4 กระบวนการ คือ จำ เข้าใจ ประยุกต์ และวิเคราะห์ ดังนั้น ในการวิจัยครั้งต่อไปควรดำเนินการวิจัยโดยการสร้างคลังข้อสอบในเนื้อหาอื่น ๆ ที่ครอบคลุมมิติความรู้ 4 ส่วน และมิติกระบวนการทั้ง 6 กระบวนการ

2. ควรมีการสร้างคลังข้อสอบที่มีการแบ่งมิติตามทักษะและกระบวนการทางคณิตศาสตร์ตามสาระและมาตรฐานการเรียนรู้หลักสูตรแกนกลางการศึกษาขั้นพื้นฐานของกระทรวงศึกษาธิการ เช่น มิติความสามารถในการแก้ปัญหา มิติความสามารถในการให้เหตุผล มิติความสามารถในการสื่อสาร การสื่อความหมายทางคณิตศาสตร์และการนำเสนอ มิติความสามารถในการเชื่อมโยงความรู้ต่างๆ ทางคณิตศาสตร์ และเชื่อมโยงคณิตศาสตร์กับศาสตร์อื่น ๆ และมิติการมีความคิดริเริ่มสร้างสรรค์ เป็นต้น

3. วิธีการควบคุมการแสดงข้อสอบที่พัฒนาขึ้นในการวิจัยนี้ มุ่งเน้นเข้าควบคุมการแสดงข้อสอบที่มีอัตราการใช้ข้อสอบสูง (Over-Expoure Item) ให้ลดลงได้ แต่ไม่ได้ควบคุมข้อสอบที่ถูกเลือกใช้น้อยหรือไม่ถูกเลือกใช้เลย (Under-Utilized Item) ให้มีอัตราการใช้ข้อสอบเพิ่มขึ้น จึงควรพัฒนาวิธีการควบคุมการแสดงข้อสอบ ให้ครอบคลุมกรณีข้อสอบที่ถูกเลือกใช้น้อยหรือไม่ถูกเลือกใช้เลย

4. โปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติที่พัฒนาขึ้นในการวิจัยนี้ใช้กับข้อสอบที่มีลักษณะการตรวจให้คะแนน 2 ค่า คือ ถูก-ผิด หรือ 0-1 และมีรายการคำตอบอย่างน้อย 4 ตัวเลือก จึงควรพัฒนาโปรแกรมให้รองรับรูปแบบการทดสอบที่หลากหลายมากกว่านี้

5. โปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติที่พัฒนาขึ้นในการวิจัยนี้เหมาะสำหรับข้อสอบที่ผ่านการวิเคราะห์ข้อสอบตามหลักทฤษฎีการตอบสนองข้อสอบแบบพหุมิติข้อสอบมาแล้ว นั่นคือ เป็นข้อสอบที่ทราบค่าอำนาจจำแนกของข้อสอบ ค่าความยากของข้อสอบและค่าการเดาของข้อสอบดังนั้นเพื่อความสะดวกในการใช้งาน จึงควรพัฒนาโปรแกรมให้สามารถวิเคราะห์ค่าอำนาจจำแนกของข้อสอบ ค่าความยากของข้อสอบและค่าการเดาของข้อสอบได้

บรรณานุกรม

กระทรวงศึกษาธิการ. (2551). *หลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551*. กรุงเทพฯ: โรงพิมพ์ชุมนุมสหกรณ์การเกษตรแห่งประเทศไทย จำกัด

กระทรวงศึกษาธิการ, (2560). *แผนพัฒนาการศึกษาของกระทรวงศึกษาธิการ ฉบับที่ 12 (2560-2564)*. ค้นเมื่อวันที่ 20 กรกฎาคม 2560 http://www.moe.go.th/moe/th/news/detail.php?NewsID=47194&Key=news20

กระทรวงศึกษาธิการ, (2561). *แผนพัฒนาการศึกษาของชาติ (2560-2579).* ค้นเมื่อวันที่ 6 ม.ค. 61 http://www.onec.go.th/index.php/book/BookView/1540

วรรณศิรินทร์ ภูฉายา, ณัฐกรานต์ สุระภา, ชนิษฐา คลังใหญ่, นิตยา จันตะคุณ, และอรุณี จันทร์ศิลา (2561). การศึกษาระดับความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 1-3 โดยใช้หลักการตรรกศาสตร์คลุมเครือ. *วารสารวิชาการเซาธ์อีสท์บางกอก (สาขา มนุษยศาสตร์ และ สังคมศาสตร์)*, *7*(1), 98-114.

สถาบันทดสอบทางการศึกษาแห่งชาติ (องค์กรมหาชน). (2557). *รายงานประจำปี 2559*. ค้นเมื่อ 16 มกราคม 2561, จาก http://www.niets.or.th/

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

Ackerman, T. A. (1994). Using multidimensional item response theory to understand what items and tests are measuring. *Applied Measurement in Education*, *7*(4), 255-278.

Ackerman, T. A., Gierl, M. J., & Walker, C. M. (2003). Using multidimensional item response theory to evaluate educational and psychological tests. *Educational Measurement*: *Issues and Practice*, *22*(3), 37-51.

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*. New York: Longman.

Babcock, B., & Weiss, D. J. (2012). Termination criteria in computerized adaptive tests: Do variable-length CATs provide efficient and effective measurement?. *Journal of Computerized Adaptive Testing*, *1*(1), 1-18.

Baghaei, P. (2012). The Application of multidimensional Rasch models in large scale assessment and validation: An empirical example. *Electronic Journal od Research in Educational Psychology*, *10*(1), 233-252.

Barrada, J. R., Olea, J., Ponsoda, V., & Abad, F. J. (2008). Incorporating randomness in the Fisher information for improving item-exposure control in CATs. *British Journal of Mathematical and Statistical Psychology*, *61*(2), 493-513.

Becker, J., Fliege, H., Kocalevent, R.D., Bjorner, J. B., Rose, M., Walter, O, B., & Klapp, B. F.(2008). Functioning and validity of A Computerized Adaptive test to measure anxiety (A-CAT). *Depression and Anxiety*, *25*(12), E182-E192

Belov, D. I., Armstrong, R. D., & Weissman, A. (2008). A Monte Carlo approach for adaptive testing with content constraints. *Applied Psychological Measurement*, *32*(6), 431-446.

Bergstrom, B. A., & Lunz, M .E. (1999). CAT for certification and licensure. In F. Drasgow & J. Olson-Buchanan (Eds.), Innovations in computerized assessment (pp.67-91). Mahwah, New Jersey: Lawrence Erlbaum.

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, US: D. Mckay.

Bock, R. D., Gibbons, R., & Muraki, E. (1988). Full-information item factor analysis. *Applied Psychological Measurement*, *12*(3), 261-280.

Bock, R. D., & Schilling, S. G. (2003). IRT Based Item Factor Analysis. In M. du Toit (Ed.) *IRT from SSI: BILOG-MG, MULTILOG, PARSCALE, TESTFACT* (pp. 584-591). Lincolnwood, Illinois: Scientific Software International.

Boyd, A. M. (2003). *Strategies for controlling testlet exposure rates in computerized adaptive testing systems. (doctoral dissertation),* The University of Texas at Austin

Briggs, D. C., & Wilson, M. (2003). An introduction to multidimensional measurement using Rasch models. *Journal of applied measurement*, *4*(1), 87-100.

Bunderson, C. V., Inouye, D. K., & Olsen, J. B. (1989). The four generations of computerized educational measurement. ETS Research Report Series, 1988(1), i-148. In R. L. Linn (Ed.), *Educational Measurement* (3rd ed., pp. 367-407). New York: Macmillan.

Chalhoub-Deville, M., Alcaya, C., & Lozier, V. M. (1996). *An operational framework for constructing a computer-adaptive test of L2 reading ability*: Theoretical and practical issues. Center for Advanced Research in Language Acquisition.

Chen, S. Y., Ankenmann, R. D., & Spray, J. A. (1999). *Exploring the relationship between item exposure rate and test overlap rate in computerized adaptive testing. paper presented at the annual meeting of the National Council on Measurement in Education*, Montreal.

Chang, H. H. (2012). *Making computerized adaptive testing diagnostic tools for schools.* In R. W. Lissitz & H. Jiao (Eds.), Computers and their impact on state assessments: Recent history and predictions for the future (pp. 195–226). Charlotte, North Carolina: Information Age.

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

Chang, H. H., & Ying, Z. (1999). A-stratified multistage computerized adaptive testing. *Applied Psychological Measurement*, *23*(3), 211-222.

Chang, H. H., Qian, J., & Ying, Z. (2001). a-Stratified multistage computerized adaptive testing with b blocking. *Applied Psychological Measurement*, *25*(4), 333-341.

Chang, H. H., & van der Linden, W. J. (2003). Optimal stratification of item pools in α-stratified computerized adaptive testing. *Applied Psychological Measurement*, *27*(4), 262-274.

Cheng, Y. Y., Wang, W. C., & Ho, Y. H. (2009). Multidimensional Rasch analysis of a psychological test with multiple subtests: a statistical solution for the bandwidth-fidelity dilemma. *Educational and Psychological Measurement*, *69*(3), 369-388.

Chen, P. H. (2009). Comparison of adaptive Bayesian estimation and weighted Bayesian estimation in multidimensional computerized adaptive testing. In D. J. Weiss (Ed.), Proceedings of the 2009 GMAC Conference on Computerized Adaptive Testing. Retrieved [date] from www.psych.umn.edu/psylabs/CATCentral/

Cheng, Y., & Chang, H. H. (2009). The maximum priority index method for severely constrained item selection in computerized adaptive testing. *British Journal of Mathematical and Statistical Psychology*, *62*(2), 369-383.

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.

Culbertson, M. J. (2014). Balancing information with imbalanced item banks in multidimensional CAT. paper presented at the annual meeting of the National Council on Measurement in Education, Philadelphia, PA, April 4–6, 2014

Davis, L. L., & Dodd, B. G. (2003). Item exposure constraints for testlets in the verbal reasoning section of the MCAT. *Applied Psychological Measurement*, *27*(5), 335-356.

De Ayala, R. J. (2009). *The theory and practice of item response theory*. New York: The Guilford Press.

Deng, H., Ansley, T., & Chang, H. H. (2010). Stratified and maximum information item selection procedures in computer adaptive testing. *Journal of Educational Measuremen*t, 47(2), 202-226.

Diao, Q., & Reckase, M. (2009). *Comparison of ability estimation and item selection methods in multidimensional computerized adaptive testing*. Michigan State University. Department of Measurement and Quantitative Methods.

Edmonds, W. A., & Kennedy, T. D. (2017). *An applied guide to research designs: Quantitative, qualitative, and mixed methods*. London: Sage.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, New Jersey: Lawrence Erlbaum Associates Publications

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.

Frey, A., & Seitz, N. N. (2008). Assessing the attainment of educational standards: Gains in measurement efficiency by multidimensional adaptive testing. *International Journal of Psychology*, *43*(3), 81.

Frey, A., & Seitz, N. N. (2009). Multidimensional adaptive testing in educational and psychological measurement: Current state and future challenges. *Studies in Educational Evaluation*, *35*(2), 89-94.

Green, B. F., Bock, R. D., Humphreys, L. G., Linn, R. L., & Reckase, M. D. (1984). Technical guidelines for assessing computerized adaptive tests. *Journal of Educational Measurement*, *21*(4), 347-360.

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.

Goff, M., & Ackerman, P. L. (1992). Personality-intelligence relations: Assessment of typical intellectual engagement. *Journal of Educational Psychology*, *84*(4), 537-553.

Gu, L., & Reckase, M. D. (2007). Designing optimal item pools for computerized adaptive tests with Sympson-Hetter exposure control. In Proceedings of the 2007 GMAC Conference on Computerized Adaptive Testing. Retrieved [date] from www. psych. umn. edu/psylabs/CATCentral.

Han, K. T. (2009). A gradual maximum information ratio approach to item selection in computerized adaptive testing. In Proceedings of the 2009 GMAC Conference on Computerized Adaptive Testing. Retrieved [date] from www. psych. umn. edu/psylabs/CATCentral.

Han, K. T. (2012). An efficiency balanced information criterion for item selection in computerized adaptive testing. *Journal of Educational Measurement*, *49*(3), 225-246.

Hau, K. T., & Chang, H. H. (2001). Item selection in computerized adaptive testing: Should more discriminating items be used first?. *Journal of Educational Measurement*, *38*(3), 249-266.

He, W., & Reckase, M. D. (2014). Item pool design for an operational variable-length computerized adaptive test. *Educational and Psychological Measurement*, *74*(3), 473-494.

Hetter, R. D., & Sympson, J. B. (1997). Item exposure-control in the CAT-ASVAB. In W.A. Sands, B. K. Waters, & J.R. McBride (Eds.), Computerized adaptive testing: From Inquiry to operation (pp. 141-1440). Washington, District of Columbia: American Psychological Association.

Huebner, A. R., Wang, C., Quinlan, K., & Seubert, L. (2016). Item exposure control for multidimensional computer adaptive testing under maximum likelihood and expected a posteriori estimation. *Behavior Research Methods*, *48*(4), 1443-1453.

Kelderman, H., & Rijkes, C. P. (1994). Loglinear multidimensional IRT models for polytomously scored items. Psychometrika, 59(2), 149-176.

Kingsbury, G. G., & Zara, A. R. (1989). Procedures for selecting items for computerized adaptive tests. *Applied Measurement in Education*, 2(4), 359-375.

Kirisci, L., Tarter, R., Reynolds, M., Ridenour, T., Stone, C., & Vanyukov, M. (2012). Computer adaptive testing of liability to addiction: Identifying individuals at risk. *Drug & Alcohol Dependence*, *123*(1), S79-S86.

Kocalevent, R. D., Rose, M., Becker, J., Walter, O. B., Fliege, H., Bjorner, J. B., ... & Klapp, B. F. (2009). An evaluation of patient-reported outcomes found computerized adaptive testing was efficient in assessing stress perception. *Journal of Clinical Epidemiology*, *62*(3), 278-287.

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

Kuo, B. C., Daud, M., & Yang, C. W. (2015). Multidimensional Computerized Adaptive Testing for Indonesia Junior High School Biology. *Eurasia Journal of Mathematics, Science and Technology Education*, *11*(5), 1105-1118.

Latu, E., & Chapman, E. (2002). Computerised Adaptive Testing. *British Journal of Educational Technology*, *33*(5), 619-622.

Lazarsfeld, P. F. (1950). The Obligations of the 1950 Pollster to the 1984 Historian. *Public Opinion Quarterly*, *14*(4), 617-638.

Lee, Y. H., Ip, E. H., & Fuh, C. D. (2008). A strategy for controlling item exposure in multidimensional computerized adaptive testing. *Educational and Psychological Measurement*, *68*(2), 215-232.

Leroux, A. J., Lopez, M., Hembry, I., & Dodd, B. G. (2013). A comparison of exposure control procedures in CATs using the 3PL model. *Educational and Psychological Measurement*, *73*(5), 857-874.

Leung, C. K., Chang, H. H., & Hau, K. T. (2002). Item selection in computerized adaptive testing: Improving the a-stratified design with the Sympson-Hetter algorithm. *Applied Psychological Measurement*, *26*(4), 376-392.

Leung, C. K., Chang, H. H., & Hau, K. T. (2003). Computerized adaptive testing: A comparison of three content balancing methods. *The Journal of Technology, Learning and Assessment*, *2*(5), 1-16.

Li, Y .H. & Schafer, W.D. (2004). The context effects of multidimensional CAT on the accuracy of multidimensional abilities and the item exposure rates. *The paper presented at the 2004 annual meeting of the America Educational Research Association,* San Diego California, United States of America.

Lord, F. M. (1971). A theoretical study of two-stage testing. *Psychometrika*, *36*(3), 227-242.

Lord, F. M. (1980). Applications of item response theory to practical testing problems. Hillsdale, New Jersry: Erlbaum.

Mao, L. (2014). Designing p-Optimal Item Pools for Multidimensional Computerized Adaptive Testing. Unpublished Doctoral Dissertation. Michigan State University.

Mao, X., Ozdemir, B., Wang, Y., & Xiu, T. (2016). A comparison study of item exposure control strategies in MCAT. *Online Submission*.

Meijer, R. R., & Nering, M. L. (1999). Computerized adaptive testing: Overview and introduction. *Applied Psychological Measurement*, *23*(3), 187-194

.

Martin-Fernandez, M., & Revuelta, J. (2017). Bayesian Estimation of Multidimensional Item Response Models. A Comparison of Analytic and Simulation Algorithms. *Psicologica: International Journal of Methodology and Experimental Psychology*, *38*(1), 25-55.

Mislevy, R. J., & Verhelst, N. (1990). Modeling item responses when different subjects employ different solution strategies. *Psychometrika*, *55*(2), 195-215.

Molina, J. G., Pareja, I., & Sanmartín, J. (2008). Modeling item banking: Analysis and design of a computerized system. *REMA*, *13*(2), 1-14.

Mulder, J., & van der Linden, W. J. (2009). Multidimensional adaptive testing with Kullback– Leibler information item selection. In W. J. van der Linden, & C. A. W. Glas (Eds.), Elements of adaptive testing (pp. 77–101). New York: Springer.

Miller, T. R., & Hirsch, T. M. (1992). Cluster analysis of angular data in applications of multidimensional item-response theory. *Applied Measurement in Education*, *5*(3), 193-211.

Murphy, D. L., Dodd, B. G., & Vaughn, B. K. (2010). A comparison of item selection techniques for testlets. *Applied Psychological Measurement*, *34*(6), 424-437.

OECD. (2016). Modular structure of the PISA 2015 context assessment design. In *PISA 2015 Assessment and Analytical Framework: Science, Reading Mathematics and Financial Literacy*, Paris: OECD. Publishing. http://dx.doi.org/10.1787/9789264255425-en

Olino, T. M., Yu, L., Klein, D. N., Rohde, P., Seeley, J. R., Pilkonis, P. A., & Lewinsohn, P. M. (2012). Measuring depression using item response theory: An examination of three measures of depressive symptomatology. *International Journal of Methods in Psychiatric Research*, *21*(1), 76-85.

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

Parshall, C. G., Spray, J. A., Kalohn, J. C., & Davey, T. (2002). *Issues in innovative item types*. In C. G. Parshall, J. A. Spray, J. C. Kalohn, & T. Davey (Eds.), *Practical considerations in computer-based testing* (pp. 70–91). New York: Springer.

Pilkonis, P. A., Choi, S. W., Reise, S. P., Stover, A. M., Riley, W. T., Cella, D., & PROMIS Cooperative Group. (2011). Item banks for measuring emotional distress from the Patient-Reported Outcomes Measurement Information System (PROMIS®): depression, anxiety, and anger. *Assessment*, *18*(3), 263-283.

Reckase, M. D. (1997). The past and future of multidimensional item response theory. *Applied Psychological Measurement*, *21*(1), 25–36.

Reckase, M. D. (1985). The difficulty of test items that measure more than one ability. *Applied Psychological Measurement*, *9*(4), 401-412.

Reckase, M. D. (1997). The past and future of multidimensional item response theory. *Applied Psychological Measurement*, *21*(1), 25-36.

Reckase, M. D. (2009). *Multidimensional item response theory*. New York: Springer.

Revuelta, J., & Ponsoda, V. (1998). A comparison of item exposure control methods in computerized adaptive testing. *Journal of Educational Measurement*, 35(4), 311-327.

Rudner, L. M. (1998). Item banking. *Practical Assessment, Research, and Evaluation, 6*(1), 41.

Sari, H. İ., & Huggins-Manley, A. C. (2017). Examining content control in adaptive tests: computerized adaptive testing vs. computerized adaptive multistage testing. *Educational Sciences: Theory & Practice*, *17*(5), 1759–1781.

Segall, D. O. (1996). Multidimensional adaptive testing. Psychometrika, 61(2), 331-354.

Segall, D. O. (2002). Principles of Multidimensional Adaptive Testing. In W. J. Van Der Linden and C.A.W. Glas (Eds.). C*omputerized Adaptive Testing : Theory and Practice.* p.53–74. Netherlands : Kluwer.

Segall, D. O. (2004). A sharing item response theory model for computerized adaptive testing. Journal of Educational and Behavioral Statistics, 29(4), 439-460.

Segall, D. O. (2005). *Computerized adaptive testing.* in K. Kempf-Leonard (Ed.), Encyclopedia of Social Measurement. p.429-438. New York : Academic Press.

Segall, D. O. (2010). Principles of multidimensional adaptive testing," in Wim J. Van Der Linden and Cees AW Glas (Eds.). Elements of Adaptive Testing. New York: Dordrecht Heidelberg London.

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

Seo, D. G., & Weiss, D. J. (2015). Best design for multidimensional computerized adaptive testing with the bifactor model. *Educational and Psychological Measurement*, *75*(6), 954-978.

Spearman, C. (1927). *The abilities of man: Their nature and measurement* (Vol. 8). New York: Macmillan.

Stocking, M. L., & Swanson, L. (1993). A method for severely constrained item selection in adaptive testing. *Applied Psychological Measurement*, *17*(3), 277-292.

Stocking, M. L., & Lewis, C. (1995). A new method of controlling item exposure in computerized adaptive testing. *ETS Research Report Series*, 1995(2), i-29.

Stocking, M. L., & Swanson, L. (1998). Optimal design of item banks for computerized adaptive tests. *Applied Psychological Measurement*, *22*(3), 271-279.

Stocking, M. L., & Lewis, C. (1998). Controlling item exposure conditional on ability in computerized adaptive testing. *Journal of Educational and Behavioral Statistics*, *23*(1), 57-75.

Stocking, M. L., & Lewis, C. (2000). Methods of controlling the exposure of items in CAT In WJ van der Linden & CAW Glas (Eds.), Computerized adaptive testing: Theory and practice (pp. 163-182).

Sympson, J. B., & Hetter, R. D. (1985). Controlling item-exposure rates in computerized adaptive testing. *In Proceedings of the 27th Annual Meeting of the Military Testing Association* (pp. 973-977).

Taylor, B. W. (2009). *Introduction to Management Science* (11 th 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; 1990. pp. 161-186.

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

Thurstone, L. L. (1947). *Multiple-factor analysis*. Chicago: The University of Chicago Press.

Tseng, W. T. (2016). Measuring English Vocabulary Size via Computerized Adaptive Testing. *Computers & Education, 97*(1), 69-85.

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

Van der Linden, W. J. (1998). Optimal assembly of psychological and educational tests. *Applied Psychological Measurement*, *22*(3), 195-211.

Van der Linden, W. J. (1999). Multidimensional adaptive testing with a minimum error-variance criterion. *Journal of Educational and Behavioral Statistics*, *24*(4), 398-412.

Van der Linden, W. J., & Veldkamp, B. P. (2007). Conditional item-exposure control in adaptive testing using item-ineligibility probabilities. *Journal of Educational and Behavioral Statistics, 32*(4), 398-418.

Veldkamp, B. P., & Matteucci, M. (2013). Bayesian computerized adaptive testing. *Ensaio: Avaliação e Políticas Públicas em Educação*, *21*(78), 57-82.

Yao, L., & Boughton, K. A. (2007). A multidimensional item response modeling approach for improving subscale proficiency estimation and classification. *Applied Psychological Measurement*, *31*(2), 83-105.

Veldkamp, B. P., & van der Linden, W. J. (2002). Multidimensional adaptive testing with constraints on test content. *Psychometrika*, *67*(4), 575-588.

Wainer, H. (1989). The future of item analysis. *ETS Research Report Series*, *1988*(2).

Wainer, H., Dorans, N., Eignor, D., Flaugher, R., Green, B. F., Mislevy, R. J., & Steinberg, L. (2001). Computerized adaptive testing: A primer. *Qual Life Res*, *10*(8), 733-734.

Walter, O. B., Becker, J., Bjorner, J. B., Fliege, H., Klapp, B. F., & Rose, M. (2007). Development and evaluation of a computer adaptive test for 'Anxiety'(Anxiety-CAT). *Quality of Life Research*, *1*(16), 143-155.

Wang, W. C., & Chen, P. H. (2004). Implementation and measurement efficiency of multidimensional computerized adaptive testing. *Applied Psychological Measurement*, *28*(5), 295-316.

Wang, C., & Chang, H. (2009). Kullback-Leibler information in multidimensional *Conference on Computerized Adaptive Testing. Retrieved [date] from www. psych. umn. edu/psylabs/CATCentral*.

Wang, C., Chang, H. H., & Boughton, K. A. (2011). Kullback–Leibler information and its applications in multi-dimensional adaptive testing. *Psychometrika*,*76*(1), 13-39.

Wang, C., Chang, H. H., & Huebner, A. (2011). Restrictive stochastic item selection methods in cognitive diagnostic computerized adaptive testing. *Journal of Educational Measurement, 48*(3), 255-273.

Wang, T., & Kolen, M. J. (2001). Evaluating comparability in computerized adaptive testing: Issues, criteria and an example. *Journal of Educational Measurement*, *38*(1), 19-49.

Wang, H. P., Kuo, B. C., Chao, R. C., & Tsai, Y. H. (2012). The Development and Evaluation of a Computerized Adaptive Testing System for Chinese Proficiency – Base on CEFR. *Procedia-Social and Behavioral Sciences, 64*, 34-42.

Weiss, D. J. (1974). Strategies of adaptive ability measurement (Research Report 74-5). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, Computerized Adaptive Testing Laboratory

Weiss, D. J., & Kingsbury, G. (1985). Application of computerized adaptive testing to educational problems. *Journal of Educational Measurement*, *21*(4), 361-375.

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.

Whitely, S. E. (1980). Multicomponent latent trait models for ability tests. *Psychometrika*, *45*(4), 479-494.

Xing, D., & Hambleton, R. K. (2004). Impact of test design, item quality, and item bank size on the psychometric properties of computer-based credentialing examinations. *Educational and Psychological Measurement*, *64*(1), 5-21.

Yan, D., von Davier, A. A., & Lewis, C. (Eds.). (2016). *Computerized multistage testing: heory and Applications*. CRC Press.

Yao, L. (2011). Multidimensional linking for domain scores and overall scores for nonequivalent groups. *Applied Psychological Measurement*, *35*(1), 48-66.

Yao, L. (2016). Multidimensional CAT item selection methods for domain scores and composite scores with item exposure control and content constraints. *Journal of Educational Measurement*, *51*(1), 18-38.

Zhang, J. (2012). Calibration of response data using MIRT models with simple and mixed structures. *Applied Psychological Measurement*, *36*(5), 375-398.

Zhang, J., Chang, H. H., & Yi, Q. (2012). Comparing single-pool and multiple-pool designs regarding test security in computerized testing. *Behavior Research Methods*, *44*(3), 742-752.

Zhou, X. (2012). *Designing p-optimal item pools in computerized adaptive tests with polytomous items* (pp. 1-146). Michigan State University.

ภาคผนวก

ภาคผนวก ก
รายชื่อผู้เชี่ยวชาญ

รายชื่อผู้เชี่ยวชาญ

1. รายชื่อผู้เชี่ยวชาญสำหรับการประเมินความสอดคล้องตามกระบวนการพุทธิปัญญาของ Bloom el at.(1956) แบบปรับใหม่ มีดังนี้

1.1) รายชื่อผู้เชี่ยวชาญตรวจสอบความสอดคล้องด้านเนื้อหาวิชาคณิตศาสตร์

1) นายปศาสตร์ ต้นทอง ครู คศ. 2 โรงเรียนทวีธาภิเศก บางขุนเทียน กรุงเทพมหานคร

2) นางสาวจริยา ทศพร ครู คศ. 3 โรงเรียนทวีธาภิเศก บางขุนเทียน กรุงเทพมหานคร

3) นายธเนตร โชควิเศษชัยสิทธิ์ รองผู้อำนวยการโรงเรียนทวีธาภิเศก บางขุนเทียน กรุงเทพมหานคร

1.2) รายชื่อผู้เชี่ยวชาญตรวจสอบความสอดคล้องด้านเนื้อหาวิชาคณิตศาสตร์กับกระบวนการพุทธิปัญญาของ Bloom el at. (1956) แบบปรับใหม่

1) อาจารย์ ดร.กนก พานทอง อาจารย์ประจำสาขาวิชาการวัดและเทคโนโลยีทางวิทยาการปัญญา วิทยาลัยวิทยาการวิจัยและวิทยาการปัญญา มหาวิทยาลัยบูรพา

2) อาจารย์ ดร.หนึ่งฤทัย เมฆวทัต อาจารย์ประจำสาขาวิชาวิจัยทางการศึกษา คณะครุศาสตร์ มหาวิทยาลัยราชภัฏราชนครินทร์

3) อาจารย์ปานวาสน์ มหาลวเลิศ อาจารย์ประจำสำนักทดสอบทางการศึกษาและจิตวิทยา มหาวิทยาลัยศรีนครินทรวิโรฒ

2. รายชื่อผู้เชี่ยวชาญสำหรับการประเมินความเหมาะสมของโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ มีดังนี้

1) อาจารย์ ดร.มิ่ง เทพครเมือง อาจารย์ประจำสาระการเรียนรู้วิชาคณิตศาสตร์ โรงเรียนสาธิตมหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร มหาวิทยาลัยศรีนครินทรวิโรฒ

2) อาจารย์ ดร.ปิยะ ถิรพันธุ์เมธี อาจารย์ประจำสาขาวิชาคอมพิวเตอร์และเทคโนโลยีสารสนเทศ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยเทคโนโลยีราชมงคลกรุงเทพ

3) อาจารย์รัชกฤช ธนพัฒนดล อาจารย์ประจำสาขาวิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยอีสเทิร์นเอเชีย

ภาคผนวก ข
ผลการวิเคราะห์คุณภาพข้อสอบ

การวิเคราะห์คุณภาพของข้อสอบตามหลักการของทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ แบบ 3 พารามิเตอร์ โดยใช้โปรแกรม Multilog โดยการคัดเลือกข้อสอบตามเกณฑ์ (Urry,1977) ค่าอำนาจจำแนกของข้อสอบ (a) มีค่าตั้งแต่ 0.50 ถึง 2.50 ค่าความยากของข้อสอบ (b) มีค่าตั้งแต่ -2.50 ถึง 2.50 และค่าโอกาสการเดาของข้อสอบ (c) มีค่าไม่เกิน 0.30 มีจำนวนข้อสอบที่ผ่านเกณฑ์ 193 ข้อจากข้อสอบทั้งหมด 256 คิดเป็นร้อยละ 75.39

การนำเสนอค่าพารามิเตอร์ของข้อสอบตามชั้นของค่าอำนาจจำแนกของข้อสอบ จำแนกตามสาระการเรียนรู้ที่ออกข้อสอบและจุดประสงค์ทางการศึกษาด้านพุทธิปัญญา ดังตารางที่ ข-1

ตารางที่ ข-1 ผลการวิเคราะห์คุณภาพข้อสอบที่ผ่านเกณฑ์ โดยภาพรวม

ชั้นที่	จุดประสงค์ทางการศึกษาด้านพุทธิปัญญาที่มุ่งวัด	สาระการออกข้อสอบ				รวม
		จำนวนและการดำเนินการ	การวัด	พีชคณิต	การวิเคราะห์ข้อมูลและความน่าจะเป็น	
1	วิเคราะห์เกี่ยวกับข้อเท็จจริง	-	-	-	-	-
	เข้าใจเกี่ยวกับมโนทัศน์	-	-	-	1	1
	ประยุกต์เกี่ยวกับมโนทัศน์	1	-	1	-	2
	วิเคราะห์เกี่ยวกับมโนทัศน์	1	-	3	1	5
	ประยุกต์เกี่ยวกับวิธีดำเนินการ	1	1	5	3	10
	วิเคราะห์เกี่ยวกับวิธีดำเนินการ	-	-	-	-	-
	รวม	3	1	9	5	18
2	วิเคราะห์เกี่ยวกับข้อเท็จจริง	1	-	2	-	3
	เข้าใจเกี่ยวกับมโนทัศน์	-	-	-	5	5
	ประยุกต์เกี่ยวกับมโนทัศน์	-	-	1	2	3
	วิเคราะห์เกี่ยวกับมโนทัศน์	1	-	6	3	10
	ประยุกต์เกี่ยวกับวิธีดำเนินการ	3	2	10	10	25
	วิเคราะห์เกี่ยวกับวิธีดำเนินการ	-	-	3	-	3
	รวม	5	2	22	20	49
3	วิเคราะห์เกี่ยวกับข้อเท็จจริง	2	-	1	-	3
	เข้าใจเกี่ยวกับมโนทัศน์	1	-	1	2	4
	ประยุกต์เกี่ยวกับมโนทัศน์	-	-	2	2	4
	วิเคราะห์เกี่ยวกับมโนทัศน์	4	-	12	2	18
	ประยุกต์เกี่ยวกับวิธีดำเนินการ	9	4	7	14	34
	วิเคราะห์เกี่ยวกับวิธีดำเนินการ	-	-	1	-	1
	รวม	16	4	24	20	64

ตารางที่ ข-1 (ต่อ)

ชั้นที่	จุดประสงค์ทางการศึกษาด้านพุทธิปัญญาที่มุ่งวัด	สาระการออกข้อสอบ				รวม
		จำนวนและการดำเนินการ	การวัด	พีชคณิต	การวิเคราะห์ข้อมูลและความน่าจะเป็น	
4	วิเคราะห์เกี่ยวกับข้อเท็จจริง	2	-	-	-	2
	เข้าใจเกี่ยวกับมโนทัศน์	-	1	-	1	2
	ประยุกต์เกี่ยวกับมโนทัศน์	-	-	-	-	-
	วิเคราะห์เกี่ยวกับมโนทัศน์	3	1	11	1	16
	ประยุกต์เกี่ยวกับวิธีดำเนินการ	6	6	21	8	41
	วิเคราะห์เกี่ยวกับวิธีดำเนินการ	-	-	1	-	1
	รวม	11	8	33	10	62

จากตารางที่ ข-1 ผลการวิเคราะห์คุณภาพข้อสอบ O-NET วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 6 ระหว่างปีการศึกษา 2551–2553 และ ปีการศึกษา 2555–2559 ที่ผ่านเกณฑ์การคัดเลือกข้อสอบตามเกณฑ์ของ Urry (1977) จำนวน 193 ข้อ จำแนกข้อสอบตามชั้นของค่าอำนาจจำแนกของข้อสอบ เป็น 4 ชั้น ได้แก่ ชั้นที่ 1 ข้อสอบที่มีค่าอำนาจจำแนกของข้อสอบ มีค่าอยู่ในช่วงระหว่าง 0.50-0.99 มีข้อสอบที่ผ่านเกณฑ์ 18 ข้อ ชั้นที่ 2 ข้อสอบที่มีค่าอำนาจจำแนกของข้อสอบ มีค่าอยู่ในช่วงระหว่าง 1.00-1.49 มีข้อสอบที่ผ่านเกณฑ์ 49 ข้อ ชั้นที่ 3 ข้อสอบที่มีค่าอำนาจจำแนกของข้อสอบ มีค่าอยู่ในช่วงระหว่าง 1.50-1.99 มีข้อสอบที่ผ่านเกณฑ์ 64 ข้อ และชั้นที่ 4 ข้อสอบที่มีค่าอำนาจจำแนกของข้อสอบ มีค่าอยู่ในช่วงระหว่าง 2.00-2.50 มีข้อสอบที่ผ่านเกณฑ์ 62 ข้อ

ตารางที่ ข-2 ผลการวิเคราะห์คุณภาพข้อสอบที่ผ่านเกณฑ์ จำแนกตามชั้นที่ 1 ของค่าอำนาจจำแนก

ข้อที่		ชั้น	สาระ	จุดประสงค์การศึกษาด้านพุทธิปัญญาที่มุ่งวัด	การวิเคราะห์คุณภาพข้อสอบ			
					ค่า a	ค่า b	ค่า c	ระดับความยาก
1	Q_2551_1	1	1	ประยุกต์เกี่ยวกับมโนทัศน์	0.90	1.63	0.06	ยาก
2	Q_2551_35	1	5	วิเคราะห์เกี่ยวกับมโนทัศน์	0.81	1.90	0.20	ยาก
3	Q_2552_4	1	1	วิเคราะห์เกี่ยวกับมโนทัศน์	0.92	1.10	0.14	ยาก
4	Q_2553_10	1	4	วิเคราะห์เกี่ยวกับมโนทัศน์	0.82	1.53	0.15	ยาก
5	Q_2555_9	1	1	ประยุกต์เกี่ยวกับวิธีดำเนินการ	0.97	1.27	0.07	ยาก
6	Q_2555_17	1	4	ประยุกต์เกี่ยวกับวิธีดำเนินการ	0.80	1.40	0.11	ยาก
7	Q_2555_19	1	4	ประยุกต์เกี่ยวกับวิธีดำเนินการ	0.83	1.56	0.13	ยาก
8	Q_2556_28	1	5	เข้าใจเกี่ยวกับมโนทัศน์	0.63	0.99	0.08	ค่อนข้างยาก
9	Q_2556_29	1	5	ประยุกต์เกี่ยวกับวิธีดำเนินการ	0.94	1.30	0.15	ยาก

ตารางที่ ข-2 (ต่อ)

ข้อที่		ชั้น	สาระ	จุดประสงค์การศึกษาด้านพุทธิปัญญาที่มุ่งวัด	การวิเคราะห์คุณภาพข้อสอบ			
					ค่า a	ค่า b	ค่า c	ระดับความยาก
10	Q_2557_1	1	4	วิเคราะห์เกี่ยวกับมโนทัศน์	0.89	1.26	0.27	ยาก
11	Q_2557_27	1	5	ประยุกต์เกี่ยวกับวิธีดำเนินการ	0.51	0.44	0.00	ปานกลาง
12	Q_2557_32	1	5	ประยุกต์เกี่ยวกับวิธีดำเนินการ	0.77	1.17	0.12	ยาก
13	Q_2558_16	1	4	ประยุกต์เกี่ยวกับวิธีดำเนินการ	0.94	1.63	0.20	ยาก
14	Q_2558_22	1	4	ประยุกต์เกี่ยวกับวิธีดำเนินการ	0.77	-0.34	0.00	ปานกลาง
15	Q_2559_10	1	2	ประยุกต์เกี่ยวกับวิธีดำเนินการ	0.61	-0.36	0.00	ปานกลาง
16	Q_2559_14	1	4	ประยุกต์เกี่ยวกับมโนทัศน์	0.98	0.75	0.19	ค่อนข้างยาก
17	Q_2559_16	1	4	ประยุกต์เกี่ยวกับวิธีดำเนินการ	0.55	-0.61	0.00	ค่อนข้างง่าย
18	Q_2559_17	1	4	วิเคราะห์เกี่ยวกับมโนทัศน์	0.85	0.66	0.16	ค่อนข้างยาก

จากตารางที่ ข-2 ผลการวิเคราะห์คุณภาพข้อสอบ O-NET วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 6 ระหว่างปีการศึกษา 2551–2553 และ ปีการศึกษา 2555-2559 ปรากฏว่า ชั้นที่ 1 มีข้อสอบที่ผ่านเกณฑ์ 18 ข้อ มีค่าอำนาจจำแนกของข้อสอบ (a) มีค่าอยู่ในช่วงระหว่าง 0.50–0.99 ค่าความยากของข้อสอบ (b) มีค่าอยู่ในช่วงระหว่าง -0.61–1.90 และค่าโอกาสการเดา (c) มีค่าอยู่ในช่วงระหว่าง 0.00–0.27 ซึ่งข้อสอบอยู่ในสาระการเรียนรู้จำนวนและการดำเนินการ จำนวน 3 ข้อ สาระการเรียนรู้การวัด จำนวน 1 ข้อ สาระการเรียนรู้พีชคณิต จำนวน 9 ข้อ และสาระการเรียนรู้การวิเคราะห์ข้อมูลและความน่าจะเป็น จำนวน 5 ข้อ และข้อสอบมุ่งวัดจุดประสงค์ทางการศึกษาด้านพุทธิปัญญาใน 6 ด้าน ได้แก่ 1) วิเคราะห์เกี่ยวกับข้อเท็จจริง ไม่มีข้อสอบ 2) เข้าใจเกี่ยวกับมโนทัศน์ จำนวน 1 ข้อ 3) ประยุกต์เกี่ยวกับมโนทัศน์ จำนวน 2 ข้อ 4) วิเคราะห์เกี่ยวกับมโนทัศน์ จำนวน 5 ข้อ 5) ประยุกต์เกี่ยวกับวิธีการดำเนินการ จำนวน 10 ข้อ และ 6) วิเคราะห์เกี่ยวกับวิธีดำเนินการ ไม่มีข้อสอบ ตามลำดับ

ตารางที่ ข-3 ผลการวิเคราะห์คุณภาพข้อสอบที่ผ่านเกณฑ์ จำแนกตามชั้นที่ 2 ของค่าอำนาจจำแนก

ข้อที่		ชั้น	สาระ	จุดประสงค์การศึกษาด้านพุทธิปัญญาที่มุ่งวัด	การวิเคราะห์คุณภาพข้อสอบ			
					ค่า a	ค่า b	ค่า c	ระดับความยาก
19	Q_2551_2	2	1	ประยุกต์เกี่ยวกับวิธีดำเนินการ	1.05	1.07	0.30	ยาก
20	Q_2551_11	2	4	วิเคราะห์เกี่ยวกับวิธีดำเนินการ	1.31	-0.09	0.13	ปานกลาง
21	Q_2551_20	2	4	วิเคราะห์เกี่ยวกับมโนทัศน์	1.16	-0.18	0.11	ปานกลาง
22	Q_2551_22	2	4	วิเคราะห์เกี่ยวกับมโนทัศน์	1.39	1.94	0.22	ยาก
23	Q_2551_24	2	4	ประยุกต์เกี่ยวกับวิธีดำเนินการ	1.07	0.14	0.12	ปานกลาง
24	Q_2551_25	2	5	วิเคราะห์เกี่ยวกับมโนทัศน์	1.28	1.07	0.21	ยาก
25	Q_2551_32	2	5	เข้าใจเกี่ยวกับมโนทัศน์	1.20	1.50	0.21	ยาก
26	Q_2551_37	2	5	ประยุกต์เกี่ยวกับวิธีดำเนินการ	1.40	0.73	0.17	ค่อนข้างยาก

ตารางที่ ข-3 (ต่อ)

ข้อที่		ชั้น	สาระ	จุดประสงค์การศึกษาด้านพุทธิปัญญาที่มุ่งวัด	การวิเคราะห์คุณภาพข้อสอบ			
					ค่า a	ค่า b	ค่า c	ระดับความยาก
27	Q_2551_38	2	5	ประยุกต์เกี่ยวกับวิธีดำเนินการ	1.19	1.28	0.22	ยาก
28	Q_2552_2	2	4	วิเคราะห์เกี่ยวกับมโนทัศน์	1.40	0.67	0.29	ค่อนข้างยาก
29	Q_2552_11	2	4	วิเคราะห์เกี่ยวกับมโนทัศน์	1.31	0.87	0.16	ค่อนข้างยาก
30	Q_2552_12	2	4	ประยุกต์เกี่ยวกับมโนทัศน์	1.45	1.60	0.16	ยาก
31	Q_2552_15	2	4	ประยุกต์เกี่ยวกับวิธีดำเนินการ	1.33	2.09	0.16	ยากมาก
32	Q_2552_16	2	2	ประยุกต์เกี่ยวกับวิธีดำเนินการ	1.45	1.21	0.17	ยาก
33	Q_2552_29	2	5	วิเคราะห์เกี่ยวกับมโนทัศน์	1.26	0.63	0.25	ค่อนข้างยาก
34	Q_2552_30	2	5	เข้าใจเกี่ยวกับมโนทัศน์	1.25	0.70	0.13	ค่อนข้างยาก
35	Q_2552_31	2	5	เข้าใจเกี่ยวกับมโนทัศน์	1.27	1.97	0.24	ยาก
36	Q_2552_32	2	5	วิเคราะห์เกี่ยวกับมโนทัศน์	1.25	1.31	0.22	ยาก
37	Q_2552_36	2	5	เข้าใจเกี่ยวกับมโนทัศน์	1.18	1.10	0.12	ปานกลาง
38	Q_2553_2	2	4	วิเคราะห์เกี่ยวกับมโนทัศน์	1.14	0.67	0.18	ค่อนข้าง่าย
39	Q_2553_9	2	4	ประยุกต์เกี่ยวกับวิธีดำเนินการ	1.37	1.97	0.12	ยาก
40	Q_2553_16	2	4	ประยุกต์เกี่ยวกับวิธีดำเนินการ	1.24	2.50	0.07	ยากมาก
41	Q_2555_2	2	4	วิเคราะห์เกี่ยวกับข้อเท็จจริง	1.46	1.80	0.11	ยาก
42	Q_2555_3	2	1	ประยุกต์เกี่ยวกับวิธีดำเนินการ	1.36	1.41	0.16	ยาก
43	Q_2555_5	2	1	วิเคราะห์เกี่ยวกับข้อเท็จจริง	1.46	1.01	0.19	ยาก
44	Q_2555_12	2	4	วิเคราะห์เกี่ยวกับมโนทัศน์	1.41	1.02	0.26	ยาก
45	Q_2555_16	2	4	ประยุกต์เกี่ยวกับวิธีดำเนินการ	1.00	1.01	0.16	ยาก
46	Q_2555_27	2	5	ประยุกต์เกี่ยวกับวิธีดำเนินการ	1.32	1.71	0.21	ยาก
47	Q_2556_3	2	1	วิเคราะห์เกี่ยวกับมโนทัศน์	1.47	1.44	0.15	ยาก
48	Q_2556_26	2	5	ประยุกต์เกี่ยวกับวิธีดำเนินการ	1.05	1.84	0.18	ยาก
49	Q_2556_30	2	5	ประยุกต์เกี่ยวกับมโนทัศน์	1.32	0.58	0.13	ค่อนข้างยาก
50	Q_2556_32	2	5	เข้าใจเกี่ยวกับมโนทัศน์	1.28	1.38	0.18	ยาก
51	Q_2557_10	2	4	ประยุกต์เกี่ยวกับวิธีดำเนินการ	1.42	1.70	0.20	ยาก
52	Q_2557_21	2	4	วิเคราะห์เกี่ยวกับข้อเท็จจริง	1.35	1.59	0.21	ยาก
53	Q_2557_22	2	4	ประยุกต์เกี่ยวกับวิธีดำเนินการ	1.36	0.88	0.17	ค่อนข้างยาก
54	Q_2557_24	2	4	วิเคราะห์เกี่ยวกับวิธีดำเนินการ	1.37	1.06	0.22	ยาก
55	Q_2557_25	2	4	ประยุกต์เกี่ยวกับวิธีดำเนินการ	1.40	1.08	0.20	ยาก
56	Q_2557_26	2	5	ประยุกต์เกี่ยวกับวิธีดำเนินการ	1.03	1.01	0.18	ยาก
57	Q_2557_30	2	5	ประยุกต์เกี่ยวกับมโนทัศน์	1.37	1.42	0.16	ยาก
58	Q_2557_31	2	5	ประยุกต์เกี่ยวกับวิธีดำเนินการ	1.04	0.97	0.15	ค่อนข้างยาก
59	Q_2558_12	2	1	ประยุกต์เกี่ยวกับวิธีดำเนินการ	1.02	1.65	0.28	ยาก
60	Q_2558_15	2	2	ประยุกต์เกี่ยวกับวิธีดำเนินการ	1.49	1.90	0.22	ยาก
61	Q_2558_21	2	4	ประยุกต์เกี่ยวกับวิธีดำเนินการ	1.19	0.65	0.29	ค่อนข้างยาก
62	Q_2558_24	2	4	วิเคราะห์เกี่ยวกับวิธีดำเนินการ	1.41	1.55	0.18	ยาก

ตารางที่ ข-3 (ต่อ)

ข้อที่		ชั้น	สาระ	จุดประสงค์การศึกษาด้านพุทธิปัญญาที่มุ่งวัด	การวิเคราะห์คุณภาพข้อสอบ			
					ค่า a	ค่า b	ค่า c	ระดับความยาก
63	Q_2558_26	2	4	ประยุกต์เกี่ยวกับวิธีดำเนินการ	1.35	1.64	0.14	ยาก
64	Q_2558_27	2	5	ประยุกต์เกี่ยวกับวิธีดำเนินการ	1.01	-0.35	000.	ปานกลาง
65	Q_2558_32	2	5	ประยุกต์เกี่ยวกับวิธีดำเนินการ	1.44	0.83	0.21	ค่อนข้างยาก
66	Q_2559_31	2	5	ประยุกต์เกี่ยวกับวิธีดำเนินการ	1.18	1.20	0.12	ปานกลาง
67	Q_2559_32	2	5	ประยุกต์เกี่ยวกับวิธีดำเนินการ	1.18	1.30	0.12	ปานกลาง

จากตารางที่ ข-3 ผลการวิเคราะห์คุณภาพข้อสอบ O-NET วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 6 ระหว่างปีการศึกษา 2551–2553 และ ปีการศึกษา 2555-2559 ปรากฏว่า ชั้นที่ 2 มีข้อสอบที่ผ่านเกณฑ์ 49 ข้อ มีค่าอำนาจจำแนกของข้อสอบ (a) มีค่าอยู่ในช่วงระหว่าง 1.01–1.49 ค่าความยากของข้อสอบ (b) มีค่าอยู่ในช่วงระหว่าง -0.35–2.50 และค่าโอกาสการเดา (c) มีค่าอยู่ในช่วงระหว่าง 0.00–0.30 ซึ่งข้อสอบอยู่ในสาระการเรียนรู้จำนวนและการดำเนินการ จำนวน 5 ข้อ สาระการเรียนรู้การวัด จำนวน 2 ข้อ สาระการเรียนรู้พีชคณิต จำนวน 22 ข้อ และสาระการเรียนรู้การวิเคราะห์ข้อมูลและความน่าจะเป็น จำนวน 20 ข้อ และข้อสอบมุ่งวัดจุดประสงค์ทางการศึกษาด้านพุทธิปัญญาใน 6 ด้าน ได้แก่ 1) วิเคราะห์เกี่ยวกับข้อเท็จจริง จำนวน 3 ข้อ 2) เข้าใจเกี่ยวกับมโนทัศน์ จำนวน 5 ข้อ 3) ประยุกต์เกี่ยวกับมโนทัศน์ จำนวน 3 ข้อ 4) วิเคราะห์เกี่ยวกับมโนทัศน์ จำนวน 10 ข้อ 5) ประยุกต์เกี่ยวกับวิธีการดำเนินการ จำนวน 25 ข้อ และ 6) วิเคราะห์เกี่ยวกับวิธีดำเนินการ จำนวน 3 ข้อ ตามลำดับ

ตารางที่ ข-4 ผลการวิเคราะห์คุณภาพข้อสอบที่ผ่านเกณฑ์ จำแนกตามชั้นที่ 3 ของค่าอำนาจจำแนก

ข้อที่		ชั้น	สาระ	จุดประสงค์การศึกษาด้านพุทธิปัญญาที่มุ่งวัด	การวิเคราะห์คุณภาพข้อสอบ			
					ค่า a	ค่า b	ค่า c	ระดับความยาก
68	Q_2551_29	3	5	เข้าใจเกี่ยวกับมโนทัศน์	1.76	1.46	0.20	ยาก
69	Q_2551_30	3	5	เข้าใจเกี่ยวกับมโนทัศน์	1.50	1.34	0.13	ยาก
70	Q_2551_31	3	5	ประยุกต์เกี่ยวกับวิธีดำเนินการ	1.85	1.16	0.27	ยาก
71	Q_2551_40	3	5	ประยุกต์เกี่ยวกับวิธีดำเนินการ	1.53	1.26	0.23	ยาก
72	Q_2552_3	3	1	วิเคราะห์เกี่ยวกับมโนทัศน์	1.79	1.61	0.24	ยาก
73	Q_2552_7	3	1	เข้าใจเกี่ยวกับมโนทัศน์	1.80	1.60	0.04	ยาก
74	Q_2552_10	3	4	วิเคราะห์เกี่ยวกับมโนทัศน์	1.80	1.25	0.28	ยาก
75	Q_2552_23	3	5	ประยุกต์เกี่ยวกับวิธีดำเนินการ	1.69	1.33	0.23	ยาก
76	Q_2552_24	3	5	ประยุกต์เกี่ยวกับวิธีดำเนินการ	1.72	1.37	0.25	ยาก
77	Q_2552_25	3	5	วิเคราะห์เกี่ยวกับมโนทัศน์	1.92	2.49	0.21	ยากมาก
78	Q_2552_34	3	5	ประยุกต์เกี่ยวกับวิธีดำเนินการ	1.69	1.09	0.09	ยาก

ตารางที่ ข-4 (ต่อ)

ข้อที่		ชั้น	สาระ	จุดประสงค์การศึกษาด้านพุทธิปัญญาที่มุ่งวัด	การวิเคราะห์คุณภาพข้อสอบ			
					ค่า a	ค่า b	ค่า c	ระดับความยาก
79	Q_2553_1	3	4	วิเคราะห์เกี่ยวกับมโนทัศน์	1.77	0.66	0.05	ค่อนข้างง่าย
80	Q_2553_3	3	4	วิเคราะห์เกี่ยวกับมโนทัศน์	1.60	1.21	0.27	ยาก
81	Q_2553_11	3	4	ประยุกต์เกี่ยวกับวิธีดำเนินการ	1.82	1.40	0.23	ยาก
82	Q_2553_12	3	4	วิเคราะห์เกี่ยวกับมโนทัศน์	1.62	1.65	0.13	ยาก
83	Q_2555_1	3	4	วิเคราะห์เกี่ยวกับมโนทัศน์	1.51	1.40	0.12	ยาก
84	Q_2555_6	3	1	ประยุกต์เกี่ยวกับวิธีดำเนินการ	1.50	1.27	0.21	ยาก
85	Q_2555_11	3	4	ประยุกต์เกี่ยวกับวิธีดำเนินการ	1.62	1.72	0.17	ยาก
86	Q_2555_15	3	1	วิเคราะห์เกี่ยวกับมโนทัศน์	1.62	1.96	0.17	ยาก
87	Q_2555_22	3	4	ประยุกต์เกี่ยวกับวิธีดำเนินการ	1.91	1.29	0.22	ยาก
88	Q_2555_32	3	5	ประยุกต์เกี่ยวกับวิธีดำเนินการ	1.82	1.70	0.17	ยาก
89	Q_2556_2	3	4	วิเคราะห์เกี่ยวกับมโนทัศน์	1.84	1.63	0.14	ยาก
90	Q_2556_5	3	1	วิเคราะห์เกี่ยวกับมโนทัศน์	1.55	1.60	0.18	ยาก
91	Q_2556_11	3	4	ประยุกต์เกี่ยวกับวิธีดำเนินการ	1.76	1.72	0.25	ยาก
92	Q_2556_19	3	4	ประยุกต์เกี่ยวกับวิธีดำเนินการ	1.93	0.92	0.08	ค่อนข้างยาก
93	Q_2556_22	3	4	ประยุกต์เกี่ยวกับวิธีดำเนินการ	1.74	1.19	0.14	ยาก
94	Q_2556_25	3	4	วิเคราะห์เกี่ยวกับมโนทัศน์	1.80	1.37	0.16	ยาก
95	Q_2556_31	3	5	ประยุกต์เกี่ยวกับวิธีดำเนินการ	1.57	1.11	0.18	ยาก
96	Q_2557_2	3	4	วิเคราะห์เกี่ยวกับมโนทัศน์	1.60	1.03	0.15	ยาก
97	Q_2557_3	3	1	วิเคราะห์เกี่ยวกับมโนทัศน์	1.55	1.06	0.13	ยาก
98	Q_2557_4	3	1	วิเคราะห์เกี่ยวกับข้อเท็จจริง	1.76	1.32	0.10	ยาก
99	Q_2557_6	3	1	ประยุกต์เกี่ยวกับวิธีดำเนินการ	1.77	1.21	0.08	ยาก
100	Q_2557_7	3	1	ประยุกต์เกี่ยวกับวิธีดำเนินการ	1.98	1.41	0.09	ยาก
101	Q_2557_8	3	1	ประยุกต์เกี่ยวกับวิธีดำเนินการ	1.86	1.37	0.15	ยาก
102	Q_2557_11	3	4	วิเคราะห์เกี่ยวกับมโนทัศน์	1.78	1.68	0.20	ยาก
103	Q_2557_23	3	4	วิเคราะห์เกี่ยวกับมโนทัศน์	1.55	0.76	0.17	ค่อนข้างยาก
104	Q_2557_28	3	5	ประยุกต์เกี่ยวกับวิธีดำเนินการ	1.98	1.41	0.12	ยาก
105	Q_2557_29	3	5	วิเคราะห์เกี่ยวกับมโนทัศน์	1.64	1.47	0.13	ยาก
106	Q_2558_3	3	1	ประยุกต์เกี่ยวกับวิธีดำเนินการ	1.65	1.28	0.16	ยาก
107	Q_2558_5	3	1	ประยุกต์เกี่ยวกับวิธีดำเนินการ	1.63	1.49	0.16	ยาก
108	Q_2558_7	3	4	ประยุกต์เกี่ยวกับมโนทัศน์	1.75	1.24	0.13	ยาก
109	Q_2558_9	3	4	เข้าใจเกี่ยวกับมโนทัศน์	1.60	1.23	0.15	ยาก
110	Q_2558_13	3	2	ประยุกต์เกี่ยวกับวิธีดำเนินการ	1.72	1.07	0.14	ยาก
111	Q_2558_14	3	2	ประยุกต์เกี่ยวกับวิธีดำเนินการ	1.90	1.30	0.16	ยาก
112	Q_2558_19	3	4	วิเคราะห์เกี่ยวกับข้อเท็จจริง	1.76	1.48	0.25	ยาก
113	Q_2558_28	3	5	ประยุกต์เกี่ยวกับวิธีดำเนินการ	1.60	1.34	0.21	ยาก
114	Q_2558_29	3	5	ประยุกต์เกี่ยวกับมโนทัศน์	1.66	1.42	0.20	ยาก

ตารางที่ ข-4 (ต่อ)

ข้อที่		ชั้น	สาระ	จุดประสงค์การศึกษาด้านพุทธิปัญญาที่มุ่งวัด	การวิเคราะห์คุณภาพข้อสอบ			
					ค่า a	ค่า b	ค่า c	ระดับความยาก
115	Q_2558_30	3	5	ประยุกต์เกี่ยวกับวิธีดำเนินการ	1.58	1.52	0.19	ยาก
116	Q_2559_1	3	1	วิเคราะห์เกี่ยวกับข้อเท็จจริง	1.62	0.97	0.17	ค่อนข้างยาก
117	Q_2559_2	3	1	ประยุกต์เกี่ยวกับวิธีดำเนินการ	1.64	1.13	0.23	ยาก
118	Q_2559_4	3	1	ประยุกต์เกี่ยวกับวิธีดำเนินการ	1.94	1.23	0.25	ยาก
119	Q_2559_5	3	1	ประยุกต์เกี่ยวกับวิธีดำเนินการ	1.65	0.89	0.21	ค่อนข้างยาก
120	Q_2559_8	3	2	ประยุกต์เกี่ยวกับวิธีดำเนินการ	1.67	0.78	0.15	ค่อนข้างยาก
121	Q_2559_9	3	2	ประยุกต์เกี่ยวกับวิธีดำเนินการ	1.63	0.86	0.30	ค่อนข้างยาก
122	Q_2559_13	3	4	วิเคราะห์เกี่ยวกับมโนทัศน์	1.90	1.12	0.27	ยาก
123	Q_2559_18	3	4	วิเคราะห์เกี่ยวกับมโนทัศน์	1.67	0.93	0.20	ค่อนข้างยาก
124	Q_2559_23	3	4	ประยุกต์เกี่ยวกับวิธีดำเนินการ	1.84	1.13	0.21	ยาก
125	Q_2559_24	3	4	วิเคราะห์เกี่ยวกับวิธีดำเนินการ	1.71	0.69	0.19	ค่อนข้างยาก
126	Q_2559_25	3	4	ประยุกต์เกี่ยวกับมโนทัศน์	1.91	1.07	0.23	ยาก
127	Q_2559_26	3	5	ประยุกต์เกี่ยวกับวิธีดำเนินการ	1.81	0.98	0.18	ค่อนข้างยาก
128	Q_2559_27	3	5	ประยุกต์เกี่ยวกับวิธีดำเนินการ	1.89	0.96	0.16	ค่อนข้างยาก
129	Q_2559_28	3	5	ประยุกต์เกี่ยวกับวิธีดำเนินการ	1.84	0.91	0.16	ค่อนข้างยาก
130	Q_2559_29	3	5	ประยุกต์เกี่ยวกับวิธีดำเนินการ	1.68	0.77	0.17	ค่อนข้างยาก
131	Q_2559_30	3	5	ประยุกต์เกี่ยวกับมโนทัศน์	1.71	0.72	0.16	ค่อนข้างยาก

จากตารางที่ ข-4 ผลการวิเคราะห์คุณภาพข้อสอบ O-NET วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 6 ระหว่างปีการศึกษา 2551–2553 และ ปีการศึกษา 2555-2559 ปรากฏว่า ชั้นที่ 3 มีข้อสอบที่ผ่านเกณฑ์ 64 ข้อ มีค่าอำนาจจำแนกของข้อสอบ (a) มีค่าอยู่ในช่วงระหว่าง 1.50–1.98 ค่าความยากของข้อสอบ (b) มีค่าอยู่ในช่วงระหว่าง 0.66–2.49 และค่าโอกาสการเดา (c) มีค่าอยู่ในช่วงระหว่าง 0.04–0.30 ซึ่งข้อสอบอยู่ในสาระการเรียนรู้จำนวนและการดำเนินการ จำนวน 16 ข้อ สาระการเรียนรู้การวัด จำนวน 4 ข้อ สาระการเรียนรู้พีชคณิต จำนวน 24 ข้อ และสาระการเรียนรู้การวิเคราะห์ข้อมูลและความน่าจะเป็น จำนวน 20 ข้อ และข้อสอบมุ่งวัดจุดประสงค์ทางการศึกษาด้านพุทธิปัญญาใน 6 ด้าน ได้แก่ 1) วิเคราะห์เกี่ยวกับข้อเท็จจริง จำนวน 3 ข้อ 2) เข้าใจเกี่ยวกับมโนทัศน์ จำนวน 4 ข้อ 3) ประยุกต์เกี่ยวกับมโนทัศน์ จำนวน 4 ข้อ 4) วิเคราะห์เกี่ยวกับมโนทัศน์ จำนวน 18 ข้อ 5) ประยุกต์เกี่ยวกับวิธีการดำเนินการ จำนวน 34 ข้อ และ 6) วิเคราะห์เกี่ยวกับวิธีดำเนินการ จำนวน 1 ข้อ ตามลำดับ

ตารางที่ ข-5 ผลการวิเคราะห์คุณภาพข้อสอบที่ผ่านเกณฑ์ จำแนกตามชั้นที่ 4 ของค่าอำนาจจำแนก

ข้อที่		ชั้น	สาระ	จุดประสงค์การศึกษาด้านพุทธิปัญญาที่มุ่งวัด	การวิเคราะห์คุณภาพข้อสอบ			
					ค่า a	ค่า b	ค่า c	ระดับความยาก
132	Q_2551_8	4	2	เข้าใจเกี่ยวกับมโนทัศน์	2.12	0.80	0.28	ค่อนข้างยาก
133	Q_2551_9	4	2	วิเคราะห์เกี่ยวกับมโนทัศน์	2.03	1.31	0.20	ยาก
134	Q_2551_14	4	4	วิเคราะห์เกี่ยวกับมโนทัศน์	2.09	1.10	0.20	ยาก
135	Q_2551_16	4	4	ประยุกต์เกี่ยวกับวิธีดำเนินการ	2.43	0.64	0.20	ค่อนข้างยาก
136	Q_2551_19	4	4	ประยุกต์เกี่ยวกับวิธีดำเนินการ	2.05	0.95	0.25	ค่อนข้างยาก
137	Q_2551_23	4	4	วิเคราะห์เกี่ยวกับมโนทัศน์	2.19	1.30	0.15	ยาก
138	Q_2551_26	4	5	ประยุกต์เกี่ยวกับวิธีดำเนินการ	2.35	0.53	0.13	ค่อนข้างยาก
139	Q_2551_27	4	5	เข้าใจเกี่ยวกับมโนทัศน์	2.26	0.48	0.27	ปานกลาง
140	Q_2551_28	4	5	วิเคราะห์เกี่ยวกับมโนทัศน์	2.05	0.39	0.26	ปานกลาง
141	Q_2551_36	4	5	ประยุกต์เกี่ยวกับวิธีดำเนินการ	2.41	1.48	0.19	ยาก
142	Q_2551_39	4	5	ประยุกต์เกี่ยวกับวิธีดำเนินการ	2.46	0.58	0.22	ค่อนข้างยาก
143	Q_2552_13	4	4	ประยุกต์เกี่ยวกับวิธีดำเนินการ	2.30	1.25	0.18	ยาก
144	Q_2552_17	4	2	ประยุกต์เกี่ยวกับวิธีดำเนินการ	2.18	1.37	0.25	ยาก
145	Q_2552_20	4	4	ประยุกต์เกี่ยวกับวิธีดำเนินการ	2.36	0.76	0.26	ค่อนข้างยาก
146	Q_2552_22	4	4	ประยุกต์เกี่ยวกับวิธีดำเนินการ	2.26	1.10	0.25	ยาก
147	Q_2552_28	4	5	ประยุกต์เกี่ยวกับวิธีดำเนินการ	2.24	0.69	0.23	ค่อนข้างยาก
148	Q_2552_33	4	5	ประยุกต์เกี่ยวกับวิธีดำเนินการ	2.20	0.93	0.18	ค่อนข้างยาก
149	Q_2553_4	4	4	วิเคราะห์เกี่ยวกับมโนทัศน์	2.06	2.14	0.05	ยากมาก
150	Q_2553_6	4	1	วิเคราะห์เกี่ยวกับมโนทัศน์	2.32	1.55	0.20	ยาก
151	Q_2553_17	4	4	ประยุกต์เกี่ยวกับวิธีดำเนินการ	2.19	1.09	0.10	ยาก
152	Q_2553_19	4	4	ประยุกต์เกี่ยวกับวิธีดำเนินการ	2.31	1.23	0.22	ยาก
153	Q_2555_7	4	1	ประยุกต์เกี่ยวกับวิธีดำเนินการ	2.28	1.22	0.20	ยาก
154	Q_2555_8	4	4	ประยุกต์เกี่ยวกับวิธีดำเนินการ	2.35	1.40	0.23	ยาก
155	Q_2555_13	4	4	วิเคราะห์เกี่ยวกับมโนทัศน์	2.28	1.87	0.14	ยาก
156	Q_2555_18	4	4	ประยุกต์เกี่ยวกับวิธีดำเนินการ	2.26	1.43	0.23	ยาก
157	Q_2555_21	4	4	ประยุกต์เกี่ยวกับวิธีดำเนินการ	2.33	1.06	0.23	ยาก
158	Q_2555_23	4	2	ประยุกต์เกี่ยวกับวิธีดำเนินการ	2.38	1.23	0.23	ยาก
159	Q_2555_25	4	2	ประยุกต์เกี่ยวกับวิธีดำเนินการ	2.26	1.51	0.21	ยาก
160	Q_2555_26	4	2	ประยุกต์เกี่ยวกับวิธีดำเนินการ	2.26	1.58	0.19	ยาก
161	Q_2555_28	4	5	วัประยุกต์เกี่ยวกับวิธีดำเนินการ	2.08	1.57	0.23	ยาก
162	Q_2555_31	4	5	ประยุกต์เกี่ยวกับวิธีดำเนินการ	2.16	1.71	0.18	ยาก
163	Q_2556_1	4	4	วิเคราะห์เกี่ยวกับมโนทัศน์	2.09	1.45	0.11	ยาก
164	Q_2556_4	4	1	วิเคราะห์เกี่ยวกับข้อเท็จจริง	2.07	1.40	0.18	ยาก
165	Q_2556_6	4	1	ประยุกต์เกี่ยวกับวิธีดำเนินการ	2.20	1.49	0.13	ยาก
166	Q_2556_7	4	4	ประยุกต์เกี่ยวกับวิธีดำเนินการ	2.45	1.06	0.18	ยาก
167	Q_2556_9	4	4	ประยุกต์เกี่ยวกับวิธีดำเนินการ	2.22	1.27	0.21	ยาก

ตารางที่ ข-5 (ต่อ)

ข้อที่		ชั้น	สาระ	จุดประสงค์การศึกษาด้านพุทธิปัญญาที่มุ่งวัด	การวิเคราะห์คุณภาพข้อสอบ			
					ค่า a	ค่า b	ค่า c	ระดับความยาก
168	Q_2556_10	4	4	ประยุกต์เกี่ยวกับวิธีดำเนินการ	2.02	1.19	0.20	ยาก
169	Q_2556_13	4	4	วิเคราะห์เกี่ยวกับมโนทัศน์	2.13	1.57	0.13	ยาก
170	Q_2556_18	4	2	ประยุกต์เกี่ยวกับวิธีดำเนินการ	2.41	1.54	0.18	ยาก
171	Q_2556_20	4	4	วิเคราะห์เกี่ยวกับวิธีดำเนินการ	2.03	1.17	0.17	ยาก
172	Q_2556_23	4	4	ประยุกต์เกี่ยวกับวิธีดำเนินการ	2.04	1.33	0.17	ยาก
173	Q_2556_24	4	4	ประยุกต์เกี่ยวกับวิธีดำเนินการ	2.02	1.06	0.16	ยาก
174	Q_2556_27	4	5	ประยุกต์เกี่ยวกับวิธีดำเนินการ	2.38	1.65	0.16	ยาก
175	Q_2557_5	4	1	วิเคราะห์เกี่ยวกับข้อเท็จจริง	2.10	1.28	0.11	ยาก
176	Q_2557_9	4	4	วิเคราะห์เกี่ยวกับมโนทัศน์	2.23	1.60	0.15	ยาก
177	Q_2557_12	4	4	วิเคราะห์เกี่ยวกับมโนทัศน์	2.34	1.64	0.21	ยาก
178	Q_2557_13	4	4	วิเคราะห์เกี่ยวกับมโนทัศน์	2.26	1.50	0.16	ยาก
179	Q_2557_14	4	4	ประยุกต์เกี่ยวกับวิธีดำเนินการ	2.37	1.42	0.18	ยาก
180	Q_2557_15	4	2	ประยุกต์เกี่ยวกับวิธีดำเนินการ	2.09	1.28	0.23	ยาก
181	Q_2558_2	4	1	วิเคราะห์เกี่ยวกับมโนทัศน์	2.50	1.37	0.27	ยาก
182	Q_2558_4	4	1	ประยุกต์เกี่ยวกับวิธีดำเนินการ	2.43	1.47	0.20	ยาก
183	Q_2558_6	4	1	ประยุกต์เกี่ยวกับวิธีดำเนินการ	2.44	1.31	0.18	ยาก
184	Q_2558_8	4	4	วิเคราะห์เกี่ยวกับมโนทัศน์	2.00	1.65	0.22	ยาก
185	Q_2558_11	4	4	ประยุกต์เกี่ยวกับวิธีดำเนินการ	2.23	1.75	0.18	ยาก
186	Q_2558_17	4	4	ประยุกต์เกี่ยวกับวิธีดำเนินการ	2.01	1.36	0.20	ยาก
187	Q_2558_18	4	4	ประยุกต์เกี่ยวกับวิธีดำเนินการ	2.40	1.62	0.17	ยาก
188	Q_2559_3	4	1	ประยุกต์เกี่ยวกับวิธีดำเนินการ	2.36	1.38	0.21	ยาก
189	Q_2559_6	4	1	วิเคราะห์เกี่ยวกับมโนทัศน์	2.42	0.64	0.14	ค่อนข้างยาก
190	Q_2559_7	4	1	ประยุกต์เกี่ยวกับวิธีดำเนินการ	2.41	0.66	0.12	ค่อนข้างยาก
191	Q_2559_11	4	4	วิเคราะห์เกี่ยวกับมโนทัศน์	2.21	1.39	0.21	ยาก
192	Q_2559_20	4	4	ประยุกต์เกี่ยวกับวิธีดำเนินการ	2.16	1.14	0.16	ยาก
193	Q_2559_22	4	4	ประยุกต์เกี่ยวกับวิธีดำเนินการ	2.49	1.30	0.21	ยาก

จากตารางที่ ข-5 ผลการวิเคราะห์คุณภาพข้อสอบ O-NET วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 6 ระหว่างปีการศึกษา 2551–2553 และ ปีการศึกษา 2555-2559 ปรากฏว่า ชั้นที่ 4มีข้อสอบที่ผ่านเกณฑ์ 62 ข้อ มีค่าอำนาจจำแนกของข้อสอบ (a) มีค่าอยู่ในช่วงระหว่าง 2.01–2.50 ค่าความยากของข้อสอบ (b) มีค่าอยู่ในช่วงระหว่าง 0.39–2.14 และค่าโอกาสการเดา (c) มีค่าอยู่ในช่วงระหว่าง 0.05–0.27 ซึ่งข้อสอบอยู่ในสาระการเรียนรู้จำนวนและการดำเนินการ จำนวน 11 ข้อ สาระการเรียนรู้การวัด จำนวน 8 ข้อ สาระการเรียนรู้พีชคณิต จำนวน 33 ข้อ และสาระการเรียนรู้การวิเคราะห์ข้อมูลและความน่าจะเป็น จำนวน 10 ข้อ และข้อสอบมุ่งวัดจุดประสงค์ทางการศึกษาด้านพุทธิปัญญาใน 6 ด้าน ได้แก่ 1) วิเคราะห์เกี่ยวกับข้อเท็จจริง จำนวน 2 ข้อ 2) เข้าใจ

เกี่ยวกับมโนทัศน์ จำนวน 2 ข้อ 3) ประยุกต์เกี่ยวกับมโนทัศน์ ไม่มีข้อสอบ 4) วิเคราะห์เกี่ยวกับมโนทัศน์ จำนวน 16 ข้อ 5) ประยุกต์เกี่ยวกับวิธีการดำเนินการ จำนวน 41 ข้อ และ 6) วิเคราะห์เกี่ยวกับวิธีดำเนินการ จำนวน 1 ข้อ ตามลำดับ

ภาคผนวก ค
ผลการศึกษาในสถานการณ์จำลอง ด้านการประมาณความสามารถของผู้สอบ

ค1. ค่าความสามารถที่แท้จริงของผู้สอบชั้นที่ 1 ค่าประมาณความสามารถของผู้สอบและจำนวนข้อสอบที่ใช้ในการทดสอบของการศึกษาในสถานการณ์จำลองการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ

ค2. การคำนวณค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย

ค3. การคำนวนค่าความลำเอียงเฉลี่ย

ผลการศึกษาในสถานการณ์จำลองการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ

ตารางที่ ค1 ค่าความสามารถที่แท้จริงของผู้สอบชั้นที่ 1 ค่าประมาณความสามารถของผู้สอบและจำนวนข้อสอบที่ใช้ในการทดสอบของการศึกษาในสถานการณ์จำลองการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ

ที่	วิธี IACB		วิธี CWA	
	ค่า θ	ค่าประมาณ $\hat{\theta}$	ค่า θ	ค่าประมาณ $\hat{\theta}$
1	0.687	1.041	0.804	-0.008
2	1.436	1.242	0.243	-0.717
3	0.702	0.979	0.403	-0.924
4	1.096	0.819	1.282	-0.903
5	0.931	0.983	1.137	-0.807
6	0.280	1.242	1.979	-0.035
7	1.061	1.727	1.037	-0.990
8	1.875	1.036	0.109	-0.651
9	1.546	1.060	1.063	0.290
10	0.820	1.552	0.919	0.082
11	1.097	1.762	0.863	-0.640
12	0.935	1.101	1.063	-0.825
13	0.880	1.151	1.850	0.575
14	2.192	0.825	0.781	-0.182
15	1.207	1.242	1.349	0.155
16	0.674	1.205	-0.164	0.680
17	0.300	1.121	0.480	0.082
18	0.895	1.113	1.490	-0.877
19	1.422	0.850	2.092	0.575
20	1.228	1.066	1.351	0.014
21	1.392	1.208	-0.015	-0.733
22	0.678	1.650	0.802	0.575
23	0.826	0.964	1.233	0.291
24	0.966	1.179	0.463	-0.562
25	1.752	1.007	0.760	-0.947
26	0.745	1.131	1.390	0.154
27	1.174	1.749	0.985	-0.860
28	1.103	0.963	1.823	0.317
29	1.071	0.887	0.527	-0.659
30	1.456	0.964	0.686	0.820
31	1.380	0.856	0.706	-0.136
32	0.474	1.716	0.430	-0.878
33	0.796	1.263	1.312	0.092

ตารางที่ ค1 (ต่อ)

ที่	วิธี IACB		วิธี CWA	
	ค่า θ	ค่าประมาณ $\hat{\theta}$	ค่า θ	ค่าประมาณ $\hat{\theta}$
34	0.933	1.049	1.322	-0.827
35	1.645	0.935	0.905	-0.649
36	1.909	1.861	1.497	-0.020
37	1.415	1.075	1.095	0.575
38	1.236	0.866	1.141	-0.690
39	1.020	1.285	1.497	-0.641
40	1.052	0.947	0.819	-0.873
41	1.708	1.014	0.969	-0.834
42	0.593	0.788	0.851	-0.717
43	1.263	0.974	2.654	-0.904
44	-0.256	1.084	1.823	-0.896
45	2.060	1.034	0.935	-0.823
46	1.551	1.752	0.794	0.052
47	1.539	0.761	0.631	-0.182
48	0.441	1.797	1.378	0.046
49	0.729	0.920	0.546	-0.732
50	0.446	0.812	0.695	0.957
51	1.311	1.173	0.082	0.034
52	1.553	1.244	1.209	-0.741
53	1.669	1.216	-0.271	-0.734
54	0.752	1.134	0.666	-0.606
55	0.787	2.257	0.977	0.082
56	0.243	0.780	1.211	-0.839
57	0.297	1.103	0.823	-0.824
58	1.077	1.049	0.649	-0.640
59	0.883	1.123	1.651	-0.893
60	0.869	0.997	1.186	-0.639
61	0.469	1.105	0.367	-0.158
62	1.772	1.280	0.868	0.193
63	1.747	0.831	0.000	-0.809
64	1.066	1.702	0.506	-0.704
65	1.591	0.950	1.904	-0.828
66	1.000	0.969	0.552	-0.748
67	0.443	2.261	0.132	-0.741
68	1.302	1.007	1.510	0.575
69	0.732	0.827	1.569	-0.088
70	0.337	1.576	0.339	0.957

ตารางที่ ค1 (ต่อ)

ที่	วิธี IACB		วิธี CWA	
	ค่า θ	ค่าประมาณ $\hat{\theta}$	ค่า θ	ค่าประมาณ $\hat{\theta}$
71	0.826	1.085	0.612	-0.730
72	0.585	0.872	1.653	-0.898
73	0.834	1.292	1.152	0.244
74	0.807	0.855	1.395	-0.188
75	1.076	0.797	1.401	1.140
76	0.950	1.069	0.565	0.291
77	0.996	0.874	0.322	-0.725
78	0.392	0.750	0.236	-0.794
79	1.029	1.151	1.277	0.738
80	1.238	0.993	0.567	-0.900
81	1.261	0.805	1.707	-0.615
82	1.011	0.968	1.208	-0.872
83	0.602	1.330	0.198	-0.871
84	0.825	0.911	0.480	-0.885
85	-0.016	1.077	0.719	0.173
86	1.949	1.226	0.630	0.170
87	1.699	0.968	1.165	0.062
88	1.562	1.250	1.355	-1.000
89	0.859	1.042	1.783	-0.835
90	1.188	1.169	1.104	-0.643
91	0.957	1.647	-0.289	-0.008
92	0.769	1.702	1.616	-0.744
93	0.345	1.267	1.506	0.820
94	0.501	1.268	1.482	-0.682
95	0.498	1.698	-1.018	0.159
96	0.806	0.917	1.783	0.376
97	1.483	1.727	0.420	0.291
98	0.404	1.212	0.953	-0.717
99	1.227	1.275	1.044	0.045
100	0.974	0.780	1.558	-0.881
101	0.994	1.112	0.832	-0.797
102	2.281	1.261	0.198	1.120
103	-0.083	0.964	0.534	-0.621
104	1.244	1.117	1.110	-0.935
105	0.234	1.815	0.616	0.820
106	0.324	0.791	1.747	-0.947
107	0.517	1.211	1.347	-0.113

ตารางที่ ค1 (ต่อ)

ที่	วิธี IACB		วิธี CWA	
	ค่า θ	ค่าประมาณ $\hat{\theta}$	ค่า θ	ค่าประมาณ $\hat{\theta}$
108	0.960	0.894	1.084	-0.641
109	0.877	1.800	0.991	-0.593
110	0.955	1.015	0.833	0.291
111	0.959	1.593	0.506	-0.640
112	1.661	1.184	0.335	0.150
113	0.852	1.127	1.311	-0.797
114	1.147	1.602	0.953	-0.885
115	1.053	1.185	0.866	-0.076
116	0.788	1.765	1.688	-0.799
117	1.656	1.229	2.147	-0.717
118	0.725	1.204	0.763	-0.912
119	0.754	2.261	1.326	-0.605
120	1.371	1.016	0.829	0.396
121	-0.072	0.867	0.602	-0.809
122	0.932	1.120	1.697	-0.176
123	1.271	0.874	0.543	-0.702
124	1.107	0.814	1.072	-0.560
125	1.098	1.592	0.467	-0.807
126	0.508	0.963	0.265	-0.642
127	1.289	0.855	0.632	-0.605
128	1.519	0.853	0.631	0.082
129	0.960	1.208	0.909	-0.578
130	0.476	1.158	0.474	0.026
131	0.987	0.887	0.501	-0.993
132	1.695	1.069	0.507	0.073
133	0.869	1.167	1.249	-0.741
134	1.872	1.662	0.718	0.301
135	0.885	1.267	0.626	-0.950
136	0.657	1.037	1.819	0.154
137	1.048	1.755	0.442	-0.622
138	0.745	0.969	1.204	-0.795
139	1.071	1.040	0.659	-0.740
140	0.667	0.900	-0.097	0.701
141	1.110	1.287	0.649	0.094
142	2.294	0.881	1.423	0.037
143	1.264	1.576	1.112	0.047
144	1.239	0.768	1.681	-0.613

ตารางที่ ค1 (ต่อ)

ที่	วิธี IACB		วิธี CWA	
	ค่า θ	ค่าประมาณ $\hat{\theta}$	ค่า θ	ค่าประมาณ $\hat{\theta}$
145	1.111	1.782	0.986	-0.877
146	1.389	1.134	0.506	-0.173
147	0.969	1.105	0.894	-0.641
148	1.103	0.923	1.528	-0.632
149	1.336	1.040	1.235	0.052
150	1.342	1.269	1.044	-0.600
151	1.185	1.011	0.238	-0.724
152	1.373	0.964	0.056	-0.643
153	1.604	1.037	0.718	-0.907
154	0.869	1.077	1.392	-0.020
155	0.596	0.889	1.633	-0.073
156	1.126	1.248	0.888	-0.943
157	0.629	1.019	1.652	0.000
158	0.578	1.000	1.288	-0.842
159	1.376	0.838	0.262	0.778
160	1.233	0.697	1.175	-0.090
161	0.230	1.727	0.625	-0.613
162	0.718	2.027	1.045	0.701
163	1.795	0.781	0.817	-0.820
164	1.104	0.820	1.398	0.282
165	1.301	1.725	1.221	-0.741
166	0.514	1.104	0.786	-0.025
167	0.440	1.752	1.327	0.913
168	0.012	0.776	1.281	-0.099
169	0.960	2.318	0.302	0.575
170	1.838	0.754	1.079	0.062
171	0.833	0.865	0.865	0.191
172	1.487	0.970	0.792	-0.812
173	0.862	1.153	1.373	0.820
174	0.436	1.105	1.536	1.133
175	0.650	1.196	1.059	-0.187
176	1.447	0.897	1.188	-0.469
177	0.362	1.029	1.332	0.820
178	1.553	0.920	0.747	-0.609
179	0.852	1.322	0.999	0.262
180	0.984	1.304	0.607	-0.087
181	1.435	1.056	1.485	-0.958

ตารางที่ ค1 (ต่อ)

ที่	วิธี IACB		วิธี CWA	
	ค่า θ	ค่าประมาณ $\hat{\theta}$	ค่า θ	ค่าประมาณ $\hat{\theta}$
182	0.614	0.998	1.131	0.255
183	0.505	0.931	0.176	0.820
184	1.171	1.260	1.178	-0.976
185	0.567	0.968	0.827	-0.061
186	1.796	1.061	0.471	-0.640
187	0.817	1.727	0.622	0.957
188	0.284	1.212	1.970	0.049
189	0.473	1.047	1.147	-0.643
190	0.780	0.845	1.385	-0.226
191	0.926	0.856	0.669	-0.640
192	1.564	0.901	1.231	0.172
193	0.734	1.751	1.110	-0.790
194	0.746	0.996	1.553	0.820
195	0.971	1.116	0.885	-0.866
196	0.969	1.712	0.416	0.019
197	0.396	1.158	0.506	0.247
198	1.198	1.124	1.078	0.154
199	1.193	1.212	0.652	-0.725
200	1.042	1.650	1.161	0.207
201	1.176	1.016	0.095	0.045
202	1.654	1.121	1.315	0.820
203	0.943	0.828	0.730	0.154
204	-0.084	1.594	1.284	-0.828
205	1.557	1.330	1.241	1.039
206	0.665	1.148	0.802	-0.022
207	1.267	1.156	1.611	-0.174
208	1.131	1.292	0.762	0.005
209	0.689	0.983	1.224	0.575
210	0.381	0.995	2.131	0.820
211	1.739	0.766	2.118	-0.651
212	0.979	1.236	1.022	0.045
213	0.757	1.208	1.042	-0.639
214	1.278	1.464	1.033	-0.885
215	0.759	1.256	1.915	-0.942
216	1.067	0.931	0.499	0.925
217	1.072	0.973	0.837	0.575
218	1.314	1.137	1.477	0.777

ตารางที่ ค1 (ต่อ)

ที่	วิธี IACB		วิธี CWA	
	ค่า θ	ค่าประมาณ $\hat{\theta}$	ค่า θ	ค่าประมาณ $\hat{\theta}$
219	1.411	1.714	0.769	-0.966
220	1.321	1.083	0.659	-0.950
221	0.794	0.993	1.058	-0.173
222	1.034	1.581	0.769	-0.605
223	0.433	1.113	0.710	-0.061
224	0.970	1.151	0.789	0.630
225	1.041	1.031	0.338	-0.651
226	0.978	0.984	0.253	0.193
227	0.311	1.582	0.566	-0.941
228	0.362	0.889	0.614	1.140
229	1.515	1.040	0.297	-0.697
230	1.598	0.952	1.380	-0.832
231	1.017	1.022	0.877	-0.605
232	0.647	1.632	1.936	0.575
233	1.792	0.866	1.282	0.330
234	0.316	1.042	1.225	-0.704
235	0.805	1.196	0.869	-0.668
236	0.788	0.836	1.160	-0.642
237	0.996	1.036	1.466	-0.643
238	1.093	1.765	0.323	-0.616
239	1.689	1.019	1.249	0.957
240	0.723	1.253	0.681	0.575
241	0.990	1.022	0.536	0.158
242	1.308	1.133	1.150	-0.986
243	1.252	1.149	0.744	-0.834
244	1.170	1.174	0.666	-0.904
245	1.847	0.997	0.670	-0.682
246	0.789	1.250	1.289	0.113
247	0.754	1.212	0.618	0.109
248	1.638	1.122	1.069	0.291
249	1.255	1.044	0.615	1.133
250	0.935	1.800	0.865	-0.816
251	1.569	1.260	1.018	-0.733
252	1.278	0.788	1.014	-0.002
253	1.188	0.872	0.588	-0.951
254	1.382	1.123	0.737	-0.157
255	1.279	1.675	1.093	0.153

ตารางที่ ค1 (ต่อ)

ที่	วิธี IACB		วิธี CWA	
	ค่า θ	ค่าประมาณ $\hat{\theta}$	ค่า θ	ค่าประมาณ $\hat{\theta}$
256	1.110	1.179	1.214	-0.659
257	0.911	1.174	0.886	-0.740
258	0.655	0.914	1.352	-0.642
259	0.318	1.196	1.569	-0.625
260	1.613	1.216	0.937	0.675
261	1.556	1.237	0.258	0.064
262	0.309	1.126	1.136	0.258
263	0.971	1.304	0.555	-0.853
264	0.546	1.490	0.149	-0.173
265	0.676	0.840	0.142	-0.739
266	1.440	0.891	0.921	0.317
267	1.157	0.951	1.378	-0.934
268	1.141	1.628	1.589	-0.835
269	0.928	0.827	0.766	-0.816
270	1.197	0.891	1.598	-0.725
271	1.546	0.881	0.772	0.112
272	1.708	1.288	1.390	-0.659
273	0.934	1.752	0.934	0.317
274	0.088	0.766	0.210	-0.658
275	1.005	1.174	0.999	-0.154
276	1.794	1.815	1.187	-0.097
277	0.629	0.948	1.038	-0.828
278	0.383	0.999	1.216	0.005
279	0.466	0.980	0.864	-0.947
280	1.345	1.108	1.200	-0.615
281	0.835	1.487	0.495	0.285
282	1.051	1.493	0.192	0.076
283	0.853	0.838	0.854	-0.173
284	0.571	1.059	1.339	-0.108
285	1.138	1.797	0.710	-0.808
286	0.069	1.036	1.038	-0.895
287	1.268	1.069	0.850	-0.741
288	1.266	0.698	1.812	-0.642
289	1.820	0.775	0.718	-0.623
290	1.532	0.781	2.053	0.154
291	1.090	1.249	0.951	0.005
292	0.613	1.061	1.116	-0.342

ตารางที่ ค1 (ต่อ)

ที่	วิธี IACB		วิธี CWA	
	ค่า θ	ค่าประมาณ $\hat{\theta}$	ค่า θ	ค่าประมาณ $\hat{\theta}$
293	0.932	1.062	0.646	0.049
294	1.579	0.697	1.016	-0.632
295	1.784	0.821	1.928	-0.767
296	0.694	1.648	1.237	0.154
297	0.161	2.261	1.096	-0.749
298	1.135	1.300	0.887	1.140
299	1.090	0.957	1.650	-0.891
300	0.564	1.280	1.274	-0.808
301	0.741	0.871	1.519	-0.959
302	1.173	0.887	0.347	0.427
303	1.594	1.144	1.173	-0.704
304	0.948	0.916	0.558	-0.747
305	0.902	0.860	1.421	0.575
306	0.397	0.886	1.007	-0.106
307	1.411	0.984	1.285	-0.615
308	0.823	0.815	0.532	-0.257
309	0.668	0.898	0.899	-0.812
310	0.491	1.105	1.517	-0.833
311	1.018	1.128	1.024	-0.816
312	1.107	0.776	1.478	0.488
313	0.782	1.762	1.220	0.820
314	0.796	1.300	1.034	-0.805
315	-0.046	1.665	0.278	-0.739
316	0.384	1.169	1.602	-0.692
317	1.199	1.152	1.033	-0.257
318	0.682	0.867	0.771	0.026
319	1.406	1.178	0.639	1.277
320	1.089	2.232	0.823	0.494
321	1.454	1.137	1.342	-0.877
322	0.532	1.170	0.571	-0.733
323	1.464	1.311	0.828	-0.826
324	0.937	0.955	1.077	-0.641
325	0.949	1.038	1.420	-0.098
326	0.882	0.754	0.401	-0.735
327	1.118	0.706	1.641	0.625
328	1.085	1.861	0.631	-0.921
329	1.613	1.103	2.059	-0.908

ตารางที่ ค1 (ต่อ)

ที่	วิธี IACB		วิธี CWA	
	ค่า θ	ค่าประมาณ $\hat{\theta}$	ค่า θ	ค่าประมาณ $\hat{\theta}$
330	0.300	1.235	0.620	-0.747
331	0.597	1.602	-0.076	-0.651
332	1.354	1.020	0.761	0.052
333	1.711	0.894	0.506	0.037
334	1.214	0.969	1.097	-0.315
335	1.392	0.860	1.116	0.588
336	0.919	1.116	2.213	-0.830
337	0.771	0.698	0.935	0.146
338	0.303	1.042	1.290	-0.606
339	1.014	1.120	0.802	-0.156
340	0.773	0.918	1.543	-0.951
341	0.405	0.891	0.343	-0.011
342	0.990	1.103	1.155	-0.320
343	0.523	1.117	1.059	0.736
344	0.749	1.257	1.027	-0.138
345	1.484	1.802	1.325	-0.885
346	0.676	0.914	0.047	-0.986
347	0.878	1.869	0.632	-0.827
348	1.903	1.240	0.801	-0.649
349	1.118	0.841	1.663	0.127
350	0.654	0.994	1.319	-0.693
351	1.189	1.212	0.456	0.957
352	1.259	1.044	1.130	-0.606
353	1.303	0.855	1.097	0.593
354	0.763	1.151	1.091	0.067
355	1.060	0.697	1.735	0.957
356	0.467	0.757	1.043	0.632
357	0.557	0.878	1.144	-0.090
358	0.285	1.295	0.829	-0.824
359	1.245	0.997	1.058	0.154
360	0.459	0.997	0.666	-0.109
361	0.290	0.863	0.649	0.052
362	1.048	0.908	0.623	-0.617
363	0.765	1.267	0.831	0.701
364	1.433	1.438	0.883	0.063
365	0.626	1.257	0.665	-0.804
366	2.222	1.124	0.710	-0.087

ตารางที่ ค1 (ต่อ)

ที่	วิธี IACB		วิธี CWA	
	ค่า θ	ค่าประมาณ $\hat{\theta}$	ค่า θ	ค่าประมาณ $\hat{\theta}$
367	1.092	1.116	0.963	-0.639
368	0.463	1.151	0.369	-0.993
369	1.531	0.846	1.039	-0.857
370	1.026	0.839	1.616	0.288
371	0.494	1.784	0.249	-0.919
372	1.106	0.966	1.278	-0.935
373	-0.017	0.827	0.870	0.575
374	0.941	0.965	1.741	-0.990
375	0.771	1.037	0.635	-0.822
376	1.166	0.875	0.930	0.127
377	0.531	0.786	1.340	-0.739
378	1.669	1.063	0.999	0.076
379	1.245	1.242	-0.152	0.095
380	0.355	1.158	0.043	-0.262
381	0.366	0.761	0.933	-0.903
382	1.355	1.011	1.073	-0.903
383	0.562	1.254	1.167	-0.149
384	0.461	1.034	1.330	0.039
385	1.345	1.859	0.206	0.005
386	0.555	1.122	0.763	-0.109
387	1.513	1.236	0.855	-0.790
388	1.067	1.647	1.232	0.376
389	0.852	1.683	0.714	0.024
390	0.731	1.299	1.791	0.052
391	1.332	1.698	1.303	0.140
392	1.118	1.114	1.383	-0.993
393	0.898	0.957	1.107	-0.823
394	0.991	1.716	0.926	-0.889
395	1.718	2.257	0.578	-0.907
396	1.400	0.869	0.406	-0.659
397	1.238	0.983	0.942	-0.591
398	0.296	1.163	0.593	-0.717
399	1.069	1.260	0.628	0.225
400	0.958	1.002	0.681	-0.092
401	1.035	1.244	1.293	-0.820
402	0.445	1.109	0.743	0.225
403	0.835	0.952	0.898	-0.958

ตารางที่ ค1 (ต่อ)

ที่	วิธี IACB		วิธี CWA	
	ค่า θ	ค่าประมาณ $\hat{\theta}$	ค่า θ	ค่าประมาณ $\hat{\theta}$
404	1.538	1.081	0.787	-0.967
405	0.746	1.300	1.672	0.026
406	0.858	1.147	1.602	-0.930
407	1.918	1.567	0.057	-0.818
408	0.606	1.232	1.181	-0.730
409	0.585	0.978	1.370	-0.633
410	0.551	1.605	1.589	-0.651
411	0.310	0.918	0.828	0.317
412	0.337	1.735	0.368	-0.877
413	1.470	1.638	1.329	-0.016
414	1.593	1.111	1.506	-0.695
415	0.970	1.583	1.720	-0.717
416	0.854	0.794	1.346	0.040
417	0.807	1.205	1.684	-0.595
418	-0.072	1.206	1.431	0.153
419	0.780	0.886	1.666	-0.175
420	0.809	0.663	1.045	-0.893
421	1.448	1.582	1.464	-0.649
422	1.347	1.647	1.596	0.023
423	1.171	1.016	0.913	0.154
424	0.302	0.881	1.212	-0.895
425	0.374	1.689	1.524	0.018
426	0.937	1.041	0.766	-0.958
427	1.439	1.028	1.498	-0.169
428	1.038	1.714	1.342	0.925
429	0.535	0.964	0.574	-0.887
430	1.562	0.999	1.013	-0.248
431	0.895	1.727	1.103	-0.613
432	0.940	1.185	0.045	-0.809
433	0.575	1.224	0.647	-0.643
434	1.953	1.267	1.412	-0.749
435	-0.016	2.241	0.493	-1.000
436	1.254	0.790	0.102	-0.877
437	0.425	1.060	1.313	0.023
438	0.850	1.037	1.085	0.211
439	0.850	0.974	1.011	0.127
440	0.788	0.882	1.047	-0.900

ตารางที่ ค1 (ต่อ)

ที่	วิธี IACB		วิธี CWA	
	ค่า θ	ค่าประมาณ $\hat{\theta}$	ค่า θ	ค่าประมาณ $\hat{\theta}$
441	0.721	1.346	1.219	-0.110
442	1.037	1.388	1.050	-0.826
443	0.755	1.066	0.588	-0.963
444	1.206	0.952	0.941	-0.518
445	0.615	0.718	1.455	0.376
446	1.004	1.083	1.519	-0.912
447	0.660	1.621	0.117	0.291
448	0.736	0.862	0.804	-0.790
449	1.086	1.093	0.764	-0.825
450	1.289	0.645	0.733	-0.888
451	1.393	1.623	1.021	-0.075
452	1.129	1.124	1.252	-0.650
453	0.754	1.175	0.468	-0.905
454	1.223	1.125	0.687	-0.735
455	0.605	0.804	1.677	0.820
456	1.023	1.775	1.009	0.153
457	1.088	0.846	1.674	-0.823
458	0.087	1.007	0.815	0.108
459	1.702	0.860	0.995	0.132
460	1.197	0.964	1.991	-0.824
461	0.823	1.191	1.285	-0.816
462	1.714	0.954	1.215	-0.943
463	0.709	1.221	1.099	-0.741
464	1.220	1.229	0.762	0.701
465	1.518	1.185	2.013	-0.812
466	0.376	0.935	1.558	0.192
467	1.533	0.905	1.588	0.021
468	0.417	0.894	0.822	0.131
469	0.736	0.920	1.159	-0.113
470	1.098	0.964	1.311	-0.877
471	1.677	1.800	0.645	-0.801
472	1.557	0.935	0.011	-0.830
473	2.252	0.993	0.663	-0.733
474	1.652	1.304	0.870	-0.893
475	0.231	1.442	0.977	-0.902
476	1.612	1.144	1.520	-0.731
477	1.273	1.392	0.503	-0.908

ตารางที่ ค1 (ต่อ)

ที่	วิธี IACB		วิธี CWA	
	ค่า θ	ค่าประมาณ $\hat{\theta}$	ค่า θ	ค่าประมาณ $\hat{\theta}$
478	0.391	1.059	0.309	-0.262
479	0.891	1.121	0.854	-0.738
480	0.778	1.034	0.624	-0.905
481	1.296	1.010	1.481	1.133
482	0.496	1.038	0.626	0.097
483	2.286	0.878	1.003	0.341
484	0.251	0.651	0.616	0.285
485	2.063	1.251	0.588	0.235
486	1.880	1.185	1.216	0.722
487	1.860	0.861	1.032	-0.651
488	0.935	1.131	0.718	1.140
489	1.376	2.260	0.720	-0.152
490	1.077	1.669	0.776	0.274
491	0.439	0.849	-0.007	-0.769
492	1.349	1.169	0.360	0.291
493	1.148	1.193	1.107	-0.852
494	1.249	0.878	1.338	-0.792
495	1.198	1.227	1.445	-0.703
496	1.004	0.901	1.240	0.925
497	0.638	0.777	0.793	-0.846
498	0.459	0.974	1.247	-0.720
499	0.335	1.082	1.379	0.234
500	1.362	1.039	1.213	0.211
Mean	997.	1.158	978.	289.-
SD	0.46	0.32	0.48	0.58
Average Bias	-	0.001	-	0.002-
RMSE	-	0.030	-	0.041

จากตารางที่ ค1 แสดงค่าความสามารถที่แท้จริงของผู้สอบ (θ) ของวิธีการจัดคลังข้อสอบวิธี IACB และวิธี CWA ในคลังข้อสอบชั้นที่ 1 ซึ่งใช้เป็นตัวอย่างของการศึกษาในสถานการณ์จำลอง การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ดำเนินการศึกษาซ้ำทั้งหมด 10 รอบ ชั้นที่ 1 ใช้คลังข้อสอบจริงขนาด 18 ข้อ ตามวิธีการจัดคลังข้อสอบวิธี IACB และวิธี CWA ใช้ค่าพารามิเตอร์ของข้อสอบไปประมาณค่าความสามารถที่แท้จริงของผู้สอบที่ละชั้น ชั้นละ 10 รอบ รอบละ 500 ค่า รวมทั้งหมดในชั้นที่ 1 จำนวน 5000 ค่า ผลที่ได้จากการศึกษาแต่ละครั้ง จะได้ค่าประมาณความสามารถของผู้สอบ $(\hat{\theta})$ ดังนั้นในแต่ละวิธีจะได้รับค่าประมาณความสามารถของผู้สอบในชั้นที่ 1 อย่างละ 20000 ค่าเนื่องด้วยข้อจำกัดในการแสดงผล จึงของแสดงผลเฉพาะชั้นที่ 1 รอบที่ 1 เท่านั้น (ดัง

ตารางที่ ข-1) ค่าประมาณความสามารถของผู้สอบนี้ นำไปคำนวณหาค่าความลำเอียงเฉลี่ยและ ค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย เพื่อเปรียบเทียบประสิทธิภาพด้านการประมาณค่าความสามารถของผู้สอบ (แสดงตัวอย่างการคำนวณในหัวข้อถัดไป)

การคำนวณค่าความลำเอียงเฉลี่ย

การคำนวณค่าความลำเอียงเฉลี่ย (Average Bias) ดำเนินการโดยนำค่าความสามารถที่แท้จริงของผู้สอบ (θ) และค่าประมาณความสามารถของผู้สอบ $(\hat{\theta})$ ที่ได้รับจากการศึกษาในสถานการณ์จำลอง (แสดงในตารางที่ ข1) มาแทนค่าในสมการที่ 58 หน้า 68 (Yan, von Davier, & Lewis, 2016, p. 115)

แสดงตัวอย่างการคำนวณค่า Average Bias และ ค่า RMSE ของวิธีการจัดคลังข้อสอบวิธี IACB รอบที่ 1 คลังข้อสอบชั้นที่ 1 ได้ดังนี้

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

เมื่อ n แทน จำนวนผู้สอบทั้งหมด

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

θ_i แทน ค่าความสามารถที่แท้จริงของผู้สอบคนที่ i

แทนค่า

$$\begin{aligned} AverageBias &= \frac{1}{500}\sum_{i=1}^{500}\left(\hat{\theta}_i - \theta_i\right) \\ &= \frac{1}{500}\left[\left(\hat{\theta}_1 - \theta_1\right) + \left(\hat{\theta}_2 - \theta_2\right) + \left(\hat{\theta}_3 - \theta_3\right) + ... + \left(\hat{\theta}_{500} - \theta_{500}\right)\right] \\ &= \frac{1}{500}\left[\left(0.687 - 1.041\right) + \left(1.436 - 1.242\right) + \left(0.702 - 0.979\right) + ... + \left(1.362 - 1.039\right)\right] \\ &= 0.000 \end{aligned}$$

การคำนวณค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย

การคำนวณค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Square Error: RMSE) ดำเนินการโดยนำค่าความสามารถที่แท้จริงของผู้สอบ (θ) และค่าประมาณความสามารถของผู้สอบ $(\hat{\theta})$ ที่ได้รับจากการศึกษาในสถานการณ์จำลอง (แสดงในตารางที่ ข1) มาแทนค่าในสมการที่ 57 หน้า 68 (Yan, von Davier, & Lewis, 2016, p. 115)

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{\theta}_i - \theta_i\right)^2}$$

เมื่อ n แทน จำนวนผู้สอบทั้งหมด

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

θ_i แทน ค่าความสามารถที่แท้จริงของผู้สอบคนที่ i

แทนค่า

$$RMSE = \sqrt{\frac{1}{500}\sum_{i=1}^{500}\left(\hat{\theta}_i - \theta_i\right)^2}$$

$$= \sqrt{\frac{1}{500}\left[\left(\hat{\theta}_1 - \theta_1\right)^2 + \left(\hat{\theta}_2 - \theta_2\right)^2 + \left(\hat{\theta}_3 - \theta_3\right)^3 + ... + \left(\hat{\theta}_{500} - \theta_{500}\right)^2\right]}$$

$$= \sqrt{\frac{1}{500}\left[\left(0.687 - 1.041\right)^2 + \left(1.436 - 1.242\right)^2 + \left(0.702 - 0.979\right)^2 + ... + \left(1.362 - 1.039\right)^2\right]}$$

=0.000

ส่วนการคำนวณค่าความลำเอียงเฉลี่ย และค่า RMSE ของวิธี IACB และวิธี CWA ในชั้นอื่นๆ สามารถคำนวณได้ในทำนองเดียวกัน

ภาคผนวก ง
ผลการศึกษาในสถานการณ์จำลอง ด้านประสิทธิภาพของการใช้ข้อสอบ

ค่าพารามิเตอร์ของข้อสอบในคลังข้อสอบจำลอง จำนวนการใช้ข้อสอบ ข้อสอบที่มีการแสดงมากเกินไป ข้อสอบที่มีการใช้น้อยเกินไป อัตราการทับซ้อนของข้อสอบ และการแจกแจงอัตราการแสดงข้อสอบ จากการศึกษาในสถานการณ์การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ

ง1. การคำนวณข้อสอบที่มีการแสดงมากเกินไป และการคำนวณข้อสอบที่มีการใช้น้อยเกินไป

ง2. การคำนวณอัตราการทับซ้อนของข้อสอบ

ง3. การคำนวณการแจกแจงอัตราการแสดงข้อสอบ

คลังข้อสอบจำลองที่ใช้ในสถานการณ์จำลองการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ

ตารางที่ ง1 แสดงค่าค่าพารามิเตอร์ของข้อสอบในคลังข้อสอบจำลอง จำนวนการใช้ข้อสอบ ข้อสอบที่มีการแสดงมากเกินไป ข้อสอบที่มีการใช้น้อยเกินไป

ข้อที่	ค่าพารามิเตอร์ของข้อสอบ			จำนวนการใช้ข้อสอบ (ครั้ง)		ข้อสอบที่มีการแสดงมากเกินไป (>0.2)		ข้อสอบที่มีการใช้น้อยเกินไป(<0.2)	
	a	b	c	วิธี IACB	วิธี CWA	วิธี IACB	วิธี CWA	วิธี IACB	วิธี CWA
1	0.90	1.63	0.06	315	0	0.63	0	0.63	0
2	0.81	1.90	0.20	243	0	0.486	0	0.486	0
3	0.92	1.10	0.14	63	126	0.126	0.252	0.126	0.252
4	0.82	1.53	0.15	129	34	0.258	0.068	0.258	0.068
5	0.97	1.27	0.07	122	73	0.244	0.146	0.244	0.146
6	0.80	1.40	0.11	99	142	0.198	0.284	0.198	0.284
7	0.83	1.56	0.13	64	13	0.128	0.026	0.128	0.026
8	0.63	0.99	0.08	47	157	0.094	0.314	0.094	0.314
9	0.94	1.30	0.15	132	89	0.264	0.178	0.264	0.178
10	0.89	1.26	0.27	68	101	0.136	0.202	0.136	0.202
11	0.51	0.44	0.00	0	132	0	0.264	0	0.264
12	0.77	1.17	0.12	78	138	0.156	0.276	0.156	0.276
13	0.94	1.63	0.20	83	0	0.166	0	0.166	0
14	0.77	-0.34	0.00	0	105	0	0.21	0	0.21
15	0.61	-0.36	0.00	500	500	1	1	1	1
16	0.98	0.75	0.19	57	144	0.114	0.288	0.114	0.288
17	0.55	-0.61	0.00	0	106	0	0.212	0	0.212
18	0.85	0.66	0.16	0	140	0	0.28	0	0.28
19	1.05	1.07	0.30	26	0	0.052	0	0.052	0
20	1.31	-0.09	0.13	0	30	0	0.06	0	0.06
21	1.16	-0.18	0.11	41	137	0.082	0.274	0.082	0.274
22	1.39	1.94	0.22	0	32	0	0.064	0	0.064
23	1.07	0.14	0.12	0	29	0	0.058	0	0.058
24	1.28	1.07	0.21	111	7	0.222	0.014	0.222	0.014
25	1.20	1.50	0.21	0	44	0	0.088	0	0.088
26	1.40	0.73	0.17	0	34	0	0.068	0	0.068
27	1.19	1.28	0.22	11	0	0.022	0	0.022	0
28	1.40	0.67	0.29	0	69	0	0.138	0	0.138
29	1.31	0.87	0.16	37	13	0.074	0.026	0.074	0.026
30	1.45	1.60	0.16	26	2	0.052	0.004	0.052	0.004
31	1.33	2.09	0.16	31	97	0.062	0.194	0.062	0.194

ตารางที่ ง1 (ต่อ)

ข้อที่	ค่าพารามิเตอร์ของข้อสอบ			จำนวนการใช้ข้อสอบ (ครั้ง)		ข้อสอบที่มีการแสดงมากเกินไป (>0.2)		ข้อสอบที่มีการใช้น้อยเกินไป(<0.2)	
	a	b	c	วิธี IACB	วิธี CWA	วิธี IACB	วิธี CWA	วิธี IACB	วิธี CWA
32	1.45	1.21	0.17	34	4	0.068	0.008	0.068	0.008
33	1.26	0.63	0.25	35	177	0.07	0.354	0.07	0.354
34	1.25	0.70	0.13	39	400	0.078	0.8	0.078	0.8
35	1.27	1.97	0.24	37	0	0.074	0	0.074	0
36	1.25	1.31	0.22	15	0	0.03	0	0.03	0
37	1.18	0.00	0.12	99	18	0.198	0.036	0.198	0.036
38	1.14	0.67	0.18	0	84	0	0.168	0	0.168
39	1.37	1.97	0.12	33	113	0.066	0.226	0.066	0.226
40	1.24	2.50	0.07	39	0	0.078	0	0.078	0
41	1.46	1.80	0.11	43	3	0.086	0.006	0.086	0.006
42	1.36	1.41	0.16	0	70	0	0.14	0	0.14
43	1.46	1.01	0.19	34	153	0.068	0.306	0.068	0.306
44	1.41	1.02	0.26	7	0	0.014	0	0.014	0
45	1.00	1.01	0.16	9	0	0.018	0	0.018	0
46	1.32	1.71	0.21	22	0	0.044	0	0.044	0
47	1.47	1.44	0.15	126	2	0.252	0.004	0.252	0.004
48	1.05	1.84	0.18	139	36	0.278	0.072	0.278	0.072
49	1.32	0.58	0.13	42	8	0.084	0.016	0.084	0.016
50	1.28	1.38	0.18	45	21	0.09	0.042	0.09	0.042
51	1.42	1.70	0.20	26	0	0.052	0	0.052	0
52	1.35	1.59	0.21	124	2	0.248	0.004	0.248	0.004
53	1.36	0.88	0.17	37	0	0.074	0	0.074	0
54	1.37	1.06	0.22	0	72	0	0.144	0	0.144
55	1.40	1.08	0.20	35	3	0.07	0.006	0.07	0.006
56	1.03	1.01	0.18	24	0	0.048	0	0.048	0
57	1.37	1.42	0.16	39	0	0.078	0	0.078	0
58	1.04	0.97	0.15	42	60	0.084	0.12	0.084	0.12
59	1.02	1.65	0.28	25	10	0.05	0.02	0.05	0.02
60	1.49	1.90	0.22	23	9	0.046	0.018	0.046	0.018
61	1.19	0.65	0.29	46	11	0.092	0.022	0.092	0.022
62	1.41	1.55	0.18	39	0	0.078	0	0.078	0
63	1.35	1.64	0.14	33	33	0.066	0.066	0.066	0.066
64	1.01	-0.35	0.00	0	0	0	0	0	0
65	1.44	0.83	0.21	401	0	0.802	0	0.802	0
66	1.18	0.00	0.12	0	58	0	0.116	0	0.116
67	1.18	0.00	0.12	25	0	0.05	0	0.05	0

ตารางที่ ง1 (ต่อ)

ข้อที่	ค่าพารามิเตอร์ของข้อสอบ			จำนวนการใช้ข้อสอบ (ครั้ง)		ข้อสอบที่มีการแสดงมากเกินไป (>0.2)		ข้อสอบที่มีการใช้น้อยเกินไป(<0.2)	
	a	b	c	วิธี IACB	วิธี CWA	วิธี IACB	วิธี CWA	วิธี IACB	วิธี CWA
68	1.76	1.46	0.20	35	0	0.07	0	0.07	0
69	1.50	1.34	0.13	31	0	0.062	0	0.062	0
70	1.85	1.16	0.27	20	35	0.04	0.07	0.04	0.07
71	1.53	1.26	0.23	27	0	0.054	0	0.054	0
72	1.79	1.61	0.24	61	0	0.122	0	0.122	0
73	1.80	1.60	0.04	55	0	0.11	0	0.11	0
74	1.80	1.25	0.28	28	0	0.056	0	0.056	0
75	1.69	1.33	0.23	37	0	0.074	0	0.074	0
76	1.72	1.37	0.25	25	0	0.05	0	0.05	0
77	1.92	2.49	0.21	41	0	0.082	0	0.082	0
78	1.69	1.09	0.09	28	68	0.056	0.136	0.056	0.136
79	1.77	0.66	0.05	0	28	0	0.056	0	0.056
80	1.60	1.21	0.27	26	1	0.052	0.002	0.052	0.002
81	1.82	1.40	0.23	26	0	0.052	0	0.052	0
82	1.62	1.65	0.13	24	0	0.048	0	0.048	0
83	1.51	1.40	0.12	36	0	0.072	0	0.072	0
84	1.50	1.27	0.21	26	0	0.052	0	0.052	0
85	1.62	1.72	0.17	23	0	0.046	0	0.046	0
86	1.62	1.96	0.17	30	0	0.06	0	0.06	0
87	1.91	1.29	0.22	29	0	0.058	0	0.058	0
88	1.82	1.70	0.17	49	0	0.098	0	0.098	0
89	1.84	1.63	0.14	50	0	0.1	0	0.1	0
90	1.55	1.60	0.18	58	0	0.116	0	0.116	0
91	1.76	1.72	0.25	18	0	0.036	0	0.036	0
92	1.93	0.92	0.08	0	63	0	0.126	0	0.126
93	1.74	1.19	0.14	21	8	0.042	0.016	0.042	0.016
94	1.80	1.37	0.16	35	0	0.07	0	0.07	0
95	1.57	1.11	0.18	25	82	0.05	0.164	0.05	0.164
96	1.60	1.03	0.15	22	89	0.044	0.178	0.044	0.178
97	1.55	1.06	0.13	17	90	0.034	0.18	0.034	0.18
98	1.76	1.32	0.10	24	0	0.048	0	0.048	0
99	1.77	1.21	0.08	28	1	0.056	0.002	0.056	0.002
100	1.98	1.41	0.09	35	0	0.07	0	0.07	0
101	1.86	1.37	0.15	38	0	0.076	0	0.076	0
102	1.78	1.68	0.20	22	0	0.044	0	0.044	0
103	1.55	0.76	0.17	0	27	0	0.054	0	0.054

ตารางที่ ง1 (ต่อ)

ข้อที่	ค่าพารามิเตอร์ของข้อสอบ			จำนวนการใช้ข้อสอบ (ครั้ง)		ข้อสอบที่มีการแสดงมากเกินไป (>0.2)		ข้อสอบที่มีการใช้น้อยเกินไป(<0.2)	
	a	b	c	วิธี IACB	วิธี CWA	วิธี IACB	วิธี CWA	วิธี IACB	วิธี CWA
104	1.98	1.41	0.12	27	0	0.054	0	0.054	0
105	1.64	1.47	0.13	44	0	0.088	0	0.088	0
106	1.65	1.28	0.16	32	0	0.064	0	0.064	0
107	1.63	1.49	0.16	43	0	0.086	0	0.086	0
108	1.75	1.24	0.13	24	0	0.048	0	0.048	0
109	1.60	1.23	0.15	29	0	0.058	0	0.058	0
110	1.72	1.07	0.14	0	110	0	0.22	0	0.22
111	1.90	1.30	0.16	500	0	1	0	1	0
112	1.76	1.48	0.25	24	0	0.048	0	0.048	0
113	1.60	1.34	0.21	31	0	0.062	0	0.062	0
114	1.66	1.42	0.20	36	0	0.072	0	0.072	0
115	1.58	1.52	0.19	44	0	0.088	0	0.088	0
116	1.62	0.97	0.17	0	52	0	0.104	0	0.104
117	1.64	1.13	0.23	21	499	0.042	0.998	0.042	0.998
118	1.94	1.23	0.25	32	0	0.064	0	0.064	0
119	1.65	0.89	0.21	0	43	0	0.086	0	0.086
120	1.67	0.78	0.15	0	19	0	0.038	0	0.038
121	1.63	0.86	0.30	0	18	0	0.036	0	0.036
122	1.90	1.12	0.27	20	71	0.04	0.142	0.04	0.142
123	1.67	0.93	0.20	0	50	0	0.1	0	0.1
124	1.84	1.13	0.21	22	189	0.044	0.378	0.044	0.378
125	1.71	0.69	0.19	0	24	0	0.048	0	0.048
126	1.91	1.07	0.23	21	108	0.042	0.216	0.042	0.216
127	1.81	0.98	0.18	0	44	0	0.088	0	0.088
128	1.89	0.96	0.16	0	37	0	0.074	0	0.074
129	1.84	0.91	0.16	0	47	0	0.094	0	0.094
130	1.68	0.77	0.17	0	15	0	0.03	0	0.03
131	1.71	0.72	0.16	0	23	0	0.046	0	0.046
132	2.12	0.80	0.28	0	68	0	0.136	0	0.136
133	2.03	1.31	0.20	57	0	0.114	0	0.114	0
134	2.09	1.10	0.20	24	118	0.048	0.236	0.048	0.236
135	2.43	0.64	0.20	0	24	0	0.048	0	0.048
136	2.05	0.95	0.25	0	50	0	0.1	0	0.1
137	2.19	1.30	0.15	21	0	0.042	0	0.042	0
138	2.35	0.53	0.13	0	25	0	0.05	0	0.05
139	2.26	0.48	0.27	0	13	0	0.026	0	0.026

ตารางที่ ง1 (ต่อ)

ข้อที่	ค่าพารามิเตอร์ของข้อสอบ			จำนวนการใช้ข้อสอบ (ครั้ง)		ข้อสอบที่มีการแสดงมากเกินไป (>0.2)		ข้อสอบที่มีการใช้น้อยเกินไป(<0.2)	
	a	b	c	วิธี IACB	วิธี CWA	วิธี IACB	วิธี CWA	วิธี IACB	วิธี CWA
140	2.05	0.39	0.26	0	16	0	0.032	0	0.032
141	2.41	1.48	0.19	109	0	0.218	0	0.218	0
142	2.46	0.58	0.22	0	13	0	0.026	0	0.026
143	2.30	1.25	0.18	24	0	0.048	0	0.048	0
144	2.18	1.37	0.25	47	0	0.094	0	0.094	0
145	2.36	0.76	0.26	0	67	0	0.134	0	0.134
146	2.26	1.10	0.25	23	98	0.046	0.196	0.046	0.196
147	2.24	0.69	0.23	0	60	0	0.12	0	0.12
148	2.20	0.93	0.18	0	64	0	0.128	0	0.128
149	2.06	2.14	0.05	2	0	0.004	0	0.004	0
150	2.32	1.55	0.20	0	0	0	0	0	0
151	2.19	1.09	0.10	31	67	0.062	0.134	0.062	0.134
152	2.31	1.23	0.22	18	20	0.036	0.04	0.036	0.04
153	2.28	1.22	0.20	89	129	0.178	0.258	0.178	0.258
154	2.35	1.40	0.23	12	0	0.024	0	0.024	0
155	2.28	1.87	0.14	0	0	0	0	0	0
156	2.26	1.43	0.23	13	0	0.026	0	0.026	0
157	2.33	1.06	0.23	33	126	0.066	0.252	0.066	0.252
158	2.38	1.23	0.23	42	4	0.084	0.008	0.084	0.008
159	2.26	1.51	0.21	78	0	0.156	0	0.156	0
160	2.26	1.58	0.19	132	0	0.264	0	0.264	0
161	2.08	1.57	0.23	148	0	0.296	0	0.296	0
162	2.16	1.71	0.18	123	0	0.246	0	0.246	0
163	2.09	1.45	0.11	13	0	0.026	0	0.026	0
164	2.07	1.40	0.18	84	0	0.168	0	0.168	0
165	2.20	1.49	0.13	0	0	0	0	0	0
166	2.45	1.06	0.18	20	109	0.04	0.218	0.04	0.218
167	2.22	1.27	0.21	21	0	0.042	0	0.042	0
168	2.02	1.19	0.20	31	500	0.062	1	0.062	1
169	2.13	1.57	0.13	9	0	0.018	0	0.018	0
170	2.41	1.54	0.18	91	0	0.182	0	0.182	0
171	2.03	1.17	0.17	25	52	0.05	0.104	0.05	0.104
172	2.04	1.33	0.17	17	0	0.034	0	0.034	0
173	2.02	1.06	0.16	17	98	0.034	0.196	0.034	0.196
174	2.38	1.65	0.16	120	0	0.24	0	0.24	0
175	2.10	1.28	0.11	89	0	0.178	0	0.178	0

ตารางที่ ง1 (ต่อ)

ข้อที่	ค่าพารามิเตอร์ของข้อสอบ			จำนวนการใช้ข้อสอบ (ครั้ง)		ข้อสอบที่มีการแสดงมากเกินไป (>0.2)		ข้อสอบที่มีการใช้น้อยเกินไป(<0.2)	
	a	b	c	วิธี IACB	วิธี CWA	วิธี IACB	วิธี CWA	วิธี IACB	วิธี CWA
176	2.23	1.60	0.15	9	0	0.018	0	0.018	0
177	2.34	1.64	0.21	0	0	0	0	0	0
178	2.26	1.50	0.16	13	0	0.026	0	0.026	0
179	2.37	1.42	0.18	18	0	0.036	0	0.036	0
180	2.09	1.28	0.23	53	0	0.106	0	0.106	0
181	2.50	1.37	0.27	78	0	0.156	0	0.156	0
182	2.43	1.47	0.20	0	0	0	0	0	0
183	2.44	1.31	0.18	67	0	0.134	0	0.134	0
184	2.00	1.65	0.22	5	0	0.01	0	0.01	0
185	2.23	1.75	0.18	0	0	0	0	0	0
186	2.01	1.36	0.20	14	0	0.028	0	0.028	0
187	2.40	1.62	0.17	19	0	0.038	0	0.038	0
188	2.36	1.38	0.21	93	0	0.186	0	0.186	0
189	2.42	0.64	0.14	0	12	0	0.024	0	0.024
190	2.41	0.66	0.12	0	16	0	0.032	0	0.032
191	2.21	1.39	0.21	19	0	0.038	0	0.038	0
192	2.16	1.14	0.16	28	69	0.056	0.138	0.056	0.138
193	2.49	1.30	0.21	21	0	0.042	0	0.042	0
Min	0.51	-0.61	0.00	0	0	0	0	0	0
Max	2.50	2.50	0.30	500	500	1.000	1.00	1.000	1.00
Mean	1.69	1.20	0.18	41.45	38.86	0.083	0.078	0.083	0.078
SD	0.48	.49	0.06	68.27	77.76	0.14	0.16	0.14	0.16
				ชั้นที่ 1 (1-18)		14	15	8 (6)	4 (11)
				ชั้นที่ 2 (19-67)		38	33	33(5)	28(5)
				ชั้นที่ 3 (68-131)		49	27	48(1)	24(3)
				ชั้นที่ 4 (132-193)		44	24	43(1)	18(6)
				รวม (193)		145	99	132(13)	74(25)
				ข้อที่ไม่ได้ถูกนำมาใช้		48	94	48	94

การคำนวณอัตราการแสดงข้อสอบ

การคำนวณข้อสอบที่มีการแสดงมากเกินไป (Overexposure Item) หมายถึง ข้อสอบใดๆ ที่มีอัตราการแสดงสูงกว่า 0.2 เมื่ออัตราการแสดงข้อสอบ หมายถึง จำนวนครั้งของข้อสอบที่จัดให้แก่ผู้สอบต่อจำนวนผู้สอบทั้งหมดที่ทำการทดสอบ และข้อสอบที่มีการใช้น้อยเกินไป (Underutilized Item) หมายถึง ข้อสอบใดๆ ที่มีอัตราการแสดงน้อยกว่า 0.2 ถ้าข้อสอบที่มีการใช้น้อยมีจำนวนมาก

แสดงถึงข้อสอบในคลังข้อสอบมีโอกาสนำไปใช้น้อยขาด ด้านประสิทธิภาพของการใช้ข้อสอบ สะท้อนถึงประสิทธิภาพวิธีเลือกข้อสอบไม่เหมาะสม และขาดประสิทธิภาพในการบริหารต้นทุนในการพัฒนาข้อสอบ ซึ่งคำนวณได้ดังสมการ 59 หน้า 68 ได้ดังนี้

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

ตัวอย่างการคำนวณอัตราการแสดงข้อสอบ

1) กรณีข้อสอบที่มีการแสดงมากเกินไป ข้อสอบข้อที่ 1 วิธี IACB ในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ ซึ่งข้อสอบข้อที่ 1 ถูกนำมาใช้ในการทดสอบทั้งหมด 315 ครั้ง จากผู้สอบทั้งหมด 500 คน ดังนั้นอัตราการแสดงข้อสอบของข้อสอบข้อที่ 1 วิธี IACB คือ $\frac{315}{500} = 0.63$ แสดงให้เห็นว่าข้อสอบข้อที่ 1 วิธี IACB เป็นข้อสอบที่มีการแสดงมากเกินไป $\left(0.63 > 0.2\right)$

2) กรณีข้อสอบที่มีการใช้น้อยเกินไป ข้อสอบข้อที่ 3 วิธี IACB ในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ ซึ่งข้อสอบข้อที่ 3 ถูกนำมาใช้ในการทดสอบทั้งหมด 63 ครั้ง จากผู้สอบทั้งหมด 500 คน ดังนั้นอัตราการแสดงข้อสอบของข้อสอบข้อที่ 1 วิธี IACB คือ $\frac{63}{500} = 0.126$ แสดงให้เห็นว่าข้อสอบข้อที่ 3 วิธี IACB เป็นข้อสอบที่มีการใช้น้อยเกินไป $\left(0.126 < 0.2\right)$

ตารางที่ ง2 แสดงผลการศึกษาในสถานการณ์การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ ด้านอัตราการทับซ้อนของข้อสอบ และการแจกแจงอัตราการแสดงข้อสอบของวิธี IACB

ชุดที่	รายการข้อสอบ	จำนวนผู้สอบ (คน)	Item Overlap rate	tem Exposue Rat Distribution
1	Q_2559_10.pngQ_2552_4.pngQ_2553_10.pngQ_2551_35.png	1	0	-
2	Q_2559_10.pngQ_2552_4.pngQ_2555_17.pngQ_2551_35.png	1	0	-
3	Q_2559_10.pngQ_2552_4.pngQ_2555_19.pngQ_2551_35.png	1	0	-
4	Q_2559_10.pngQ_2552_4.pngQ_2556_29.pngQ_2553_10.png	1	0	-
5	Q_2559_10.pngQ_2552_4.pngQ_2556_29.pngQ_2555_17.png	1	0	-
6	Q_2559_10.pngQ_2552_4.pngQ_2556_29.pngQ_2555_19.png	1	0	-
7	Q_2559_10.pngQ_2552_4.pngQ_2556_29.pngQ_2558_16.png	2	0.4	0.142222
8	Q_2559_10.pngQ_2552_4.pngQ_2557_1.pngQ_2551_35.png	2	0.4	0.142222
9	Q_2559_10.pngQ_2552_4.pngQ_2557_32.pngQ_2553_10.png	2	0.4	0.142222
10	Q_2559_10.pngQ_2552_4.pngQ_2557_32.pngQ_2558_16.png	2	0.4	0.142222
11	Q_2559_10.pngQ_2552_4.pngQ_2558_16.pngQ_2551_35.png	3	0.6	0.642222
12	Q_2559_10.pngQ_2553_10.pngQ_2551_1.pngQ_2551_35.png	3	0.6	0.642222

ตารางที่ ง2 (ต่อ)

ชุดที่	รายการข้อสอบ	จำนวนผู้สอบ (คน)	Item Overlap rate	tem Exposue Rat Distribution
13	Q_2559_10.pngQ_2555_17.pngQ_2551_1.pngQ_2551_35.png	3	0.6	0.642222
14	Q_2559_10.pngQ_2555_9.pngQ_2553_10.pngQ_2551_35.png	3	0.6	0.642222
15	Q_2559_10.pngQ_2555_9.pngQ_2555_17.pngQ_2551_35.png	3	0.6	0.642222
16	Q_2559_10.pngQ_2555_9.pngQ_2555_19.pngQ_2551_35.png	3	0.6	0.642222
17	Q_2559_10.pngQ_2555_9.pngQ_2556_29.pngQ_2553_10.png	3	0.6	0.642222
18	Q_2559_10.pngQ_2555_9.pngQ_2556_29.pngQ_2555_17.png	4	0.8	1.502222
19	Q_2559_10.pngQ_2555_9.pngQ_2556_29.pngQ_2555_19.png	4	0.8	1.502222
20	Q_2559_10.pngQ_2555_9.pngQ_2556_29.pngQ_2558_16.png	4	0.8	1.502222
21	Q_2559_10.pngQ_2555_9.pngQ_2558_16.pngQ_2551_35.png	4	0.8	1.502222
22	Q_2559_10.pngQ_2556_28.pngQ_2551_1.pngQ_2558_16.png	4	0.8	1.502222
23	Q_2559_10.pngQ_2556_28.pngQ_2552_4.pngQ_2555_17.png	4	0.8	1.502222
24	Q_2559_10.pngQ_2556_28.pngQ_2553_10.pngQ_2551_1.png	5	1	2.722222
25	Q_2559_10.pngQ_2556_28.pngQ_2555_17.pngQ_2551_1.png	5	1	2.722222
26	Q_2559_10.pngQ_2556_28.pngQ_2555_19.pngQ_2551_1.png	5	1	2.722222
27	Q_2559_10.pngQ_2556_28.pngQ_2555_9.pngQ_2553_10.png	5	1	2.722222
28	Q_2559_10.pngQ_2556_28.pngQ_2555_9.pngQ_2555_19.png	5	1	2.722222
29	Q_2559_10.pngQ_2556_28.pngQ_2557_1.pngQ_2551_1.png	5	1	2.722222
30	Q_2559_10.pngQ_2556_28.pngQ_2557_1.pngQ_2555_9.png	6	1.2	4.302222
31	Q_2559_10.pngQ_2556_28.pngQ_2558_16.pngQ_2551_1.png	6	1.2	4.302222
32	Q_2559_10.pngQ_2556_29.pngQ_2551_1.pngQ_2558_16.png	7	1.4	6.242222
33	Q_2559_10.pngQ_2556_29.pngQ_2553_10.pngQ_2551_1.png	7	1.4	6.242222
34	Q_2559_10.pngQ_2556_29.pngQ_2555_17.pngQ_2551_1.png	7	1.4	6.242222
35	Q_2559_10.pngQ_2556_29.pngQ_2555_19.pngQ_2551_1.png	7	1.4	6.242222
36	Q_2559_10.pngQ_2556_29.pngQ_2558_16.pngQ_2551_1.png	8	1.6	8.542222
37	Q_2559_10.pngQ_2557_1.pngQ_2551_1.pngQ_2551_35.png	8	1.6	8.542222
38	Q_2559_10.pngQ_2557_1.pngQ_2555_9.pngQ_2551_35.png	8	1.6	8.542222
39	Q_2559_10.pngQ_2557_1.pngQ_2555_9.pngQ_2556_29.png	8	1.6	8.542222
40	Q_2559_10.pngQ_2557_1.pngQ_2556_29.pngQ_2551_1.png	9	1.8	11.20222
41	Q_2559_10.pngQ_2557_32.pngQ_2551_1.pngQ_2558_16.png	9	1.8	11.20222
42	Q_2559_10.pngQ_2557_32.pngQ_2553_10.pngQ_2551_1.png	10	2	14.22222
43	Q_2559_10.pngQ_2557_32.pngQ_2555_17.pngQ_2551_1.png	10	2	14.22222
44	Q_2559_10.pngQ_2557_32.pngQ_2555_19.pngQ_2551_1.png	10	2	14.22222
45	Q_2559_10.pngQ_2557_32.pngQ_2555_9.pngQ_2555_17.png	10	2	14.22222
46	Q_2559_10.pngQ_2557_32.pngQ_2555_9.pngQ_2555_19.png	10	2	14.22222
47	Q_2559_10.pngQ_2557_32.pngQ_2555_9.pngQ_2558_16.png	10	2	14.22222
48	Q_2559_10.pngQ_2557_32.pngQ_2557_1.pngQ_2551_1.png	11	2.2	17.60222

ตารางที่ ง2 (ต่อ)

ชุดที่	รายการข้อสอบ	จำนวนผู้สอบ (คน)	Item Overlap rate	tem Exposue Rat Distribution
49	Q_2559_10.pngQ_2557_32.pngQ_2557_1.pngQ_2555_9.png	11	2.2	17.60222
50	Q_2559_10.pngQ_2557_32.pngQ_2558_16.pngQ_2551_1.png	11	2.2	17.60222
51	Q_2559_10.pngQ_2559_14.pngQ_2551_1.pngQ_2551_35.png	12	2.4	21.34222
52	Q_2559_10.pngQ_2559_14.pngQ_2552_4.pngQ_2556_29.png	13	2.6	25.44222
53	Q_2559_10.pngQ_2559_14.pngQ_2555_9.pngQ_2551_35.png	14	2.8	29.90222
54	Q_2559_10.pngQ_2559_14.pngQ_2555_9.pngQ_2556_29.png	14	2.8	29.90222
55	Q_2559_10.pngQ_2559_14.pngQ_2556_28.pngQ_2551_1.png	14	2.8	29.90222
56	Q_2559_10.pngQ_2559_14.pngQ_2556_28.pngQ_2552_4.png	14	2.8	29.90222
57	Q_2559_10.pngQ_2559_14.pngQ_2556_28.pngQ_2555_9.png	14	2.8	29.90222
58	Q_2559_10.pngQ_2559_14.pngQ_2556_29.pngQ_2551_1.png	19	3.8	57.60222
59	Q_2559_10.pngQ_2559_14.pngQ_2557_32.pngQ_2551_1.png	48	9.6	395.7422
60	Q_2559_10.pngQ_2559_14.pngQ_2557_32.pngQ_2555_9.png	66	13.2	757.9022
	ผลรวม	500	98.8	1666.240

การคำนวณอัตราการทับซ้อนของข้อสอบ

อัตราการทับซ้อนของข้อสอบ (Item Overlap Rate) หมายถึง สัดส่วนของข้อสอบที่ใช้ร่วมกันหารจำนวนผู้เข้าสอบทั้งแล้วคูณด้วย100 เมื่อกำหนดให้ความยาวของแบบทดสอบคือ L ข้อ และจำนวนผู้สอบคือ P คน อัตราการทับซ้อนของข้อสอบคำนวณได้จากสมการ 60 หน้า 69

$$\text{อัตราการทับซ้อนของข้อสอบ} = \frac{\text{จำนวนผู้สอบที่ใช้ข้อสอบร่วมกัน}}{\text{จำนวนผู้สอบทั้งหมด}} \quad X \quad 100$$

อัตราการทับซ้อนของข้อสอบเป็นดัชนีสรุปสำหรับการควบคุมการแสดงข้อสอบ ใช้กำหนดขนาดและส่วนประกอบของคลังข้อสอบในการทดสอบแบบปรับเหมาะ ให้ภาพรวมว่าชุดข้อสอบถูกจัดให้บ่อยครั้งเพียงไรและคาดว่าจำนวนข้อสอบใช้ร่วมกันระหว่างคู่ของผู้สอบต้องมีจำนวนน้อยที่สุด ถ้าอัตราการทับซ้อนของข้อสอบสูงจะคุกคามความปลอดภัยของข้อสอบ

ตัวอย่างอัตราการทับซ้อนของข้อสอบ

กรณีอัตราการทับซ้อนของข้อสอบ วิธี IACB ในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ สามารถจัดชุดข้อสอบได้ 60 ชุดข้อสอบ ซึ่งอัตราการทับซ้อนของข้อสอบ จะคำนวณจากชุดข้อสอบที่มีผู้สอบได้รับชุดข้อสอบร่วมกัน (2 คนขึ้นไป) โดยนำชุดข้อสอบชุดที่ 7 – 60 มาคำนวณ เช่น ข้อสอบชุดที่ 7 มีจำนวนผู้ใช้ข้อสอบร่วมกัน 2 คน จากผู้สอบทั้งหมด 500 คน คือ $\left(\frac{2}{500}\right)*100 = 0.4$ และข้อสอบชุดที่ 60 มีจำนวนผู้ใช้ข้อสอบร่วมกัน 66 คน จากผู้สอบทั้งหมด 500

คน คือ $\left(\frac{66}{500}\right)*100=13.2$ ถ้าจะหาอัตราทับซ้อนของข้อสอบทั้งชุดคือผลรวมของอัตราทับซ้อนข้อสอบทั้งหมด

ผลรวมอัตราทับซ้อนข้อสอบ $=(0.4+0.4+0.4+...+13.2)$

= 98.8

อัตราการทับซ้อนของข้อสอบเป็นดัชนีสรุปสำหรับการควบคุมการแสดงข้อสอบ ใช้กำหนดขนาดและส่วนประกอบของคลังข้อสอบในการทดสอบแบบปรับเหมาะ ให้ภาพรวมว่าชุดข้อสอบถูกจัดให้บ่อยครั้งเพียงไรและคาดว่าจำนวนข้อสอบใช้ร่วมกันระหว่างคู่ของผู้สอบต้องมีจำนวนน้อยที่สุด ถ้าอัตราการทับซ้อนของข้อสอบสูงจะคุกคามความปลอดภัยของข้อสอบ

ตารางที่ ง3 แสดงผลการศึกษาในสถานการณ์การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ ด้านอัตราการทับซ้อนของข้อสอบ และการแจกแจงอัตราการแสดงข้อสอบของวิธี CWA

ชุดที่	รายการข้อสอบ	จำนวนผู้สอบ (คน)	Item Overlap rate	tem Exposue Rat Distribution
1	Q_2559_10.pngQ_2552_4.pngQ_2556_28.pngQ_2558_22.png	1	0	-
2	Q_2559_10.pngQ_2552_4.pngQ_2556_28.pngQ_2559_17.png	1	0	-
3	Q_2559_10.pngQ_2552_4.pngQ_2557_27.pngQ_2558_22.png	1	0	-
4	Q_2559_10.pngQ_2552_4.pngQ_2559_14.pngQ_2559_17.png	1	0	-
5	Q_2559_10.pngQ_2552_4.pngQ_2559_17.pngQ_2558_22.png	1	0	-
6	Q_2559_10.pngQ_2555_17.pngQ_2552_4.pngQ_2559_14.png	1	0	-
7	Q_2559_10.pngQ_2555_17.pngQ_2552_4.pngQ_2559_16.png	1	0	-
8	Q_2559_10.pngQ_2555_17.pngQ_2553_10.pngQ_2552_4.png	1	0	-
9	Q_2559_10.pngQ_2555_17.pngQ_2553_10.pngQ_2557_27.png	1	0	-
10	Q_2559_10.pngQ_2555_17.pngQ_2553_10.pngQ_2557_32.png	1	0	-
11	Q_2559_10.pngQ_2555_17.pngQ_2553_10.pngQ_2558_22.png	1	0	-
12	Q_2559_10.pngQ_2555_17.pngQ_2553_10.pngQ_2559_16.png	1	0	-
13	Q_2559_10.pngQ_2555_17.pngQ_2555_9.pngQ_2558_22.png	1	0	-
14	Q_2559_10.pngQ_2555_17.pngQ_2555_9.pngQ_2559_17.png	1	0	-
15	Q_2559_10.pngQ_2555_17.pngQ_2556_28.pngQ_2559_17.png	1	0	-
16	Q_2559_10.pngQ_2555_17.pngQ_2556_29.pngQ_2552_4.png	1	0	-
17	Q_2559_10.pngQ_2555_17.pngQ_2556_29.pngQ_2556_28.png	1	0	-
18	Q_2559_10.pngQ_2555_17.pngQ_2557_1.pngQ_2556_28.png	1	0	-
19	Q_2559_10.pngQ_2555_17.pngQ_2557_1.pngQ_2559_16.png	1	0	-
20	Q_2559_10.pngQ_2555_17.pngQ_2557_1.pngQ_2559_17.png	1	0	-
21	Q_2559_10.pngQ_2555_17.pngQ_2557_32.pngQ_2557_27.png	1	0	-
22	Q_2559_10.pngQ_2555_17.pngQ_2559_14.pngQ_2559_17.png	1	0	-
23	Q_2559_10.pngQ_2555_9.pngQ_2552_4.pngQ_2559_17.png	1	0	-

ตารางที่ ง3 (ต่อ)

ชุดที่	รายการข้อสอบ	จำนวนผู้สอบ (คน)	Item Overlap rate	tem Exposue Rat Distribution
24	Q_2559_10.pngQ_2555_9.pngQ_2556_28.pngQ_2552_4.png	1	0	-
25	Q_2559_10.pngQ_2555_9.pngQ_2556_28.pngQ_2558_22.png	1	0	-
26	Q_2559_10.pngQ_2555_9.pngQ_2556_28.pngQ_2559_16.png	1	0	-
27	Q_2559_10.pngQ_2555_9.pngQ_2556_28.pngQ_2559_17.png	1	0	-
28	Q_2559_10.pngQ_2555_9.pngQ_2556_29.pngQ_2556_28.png	1	0	-
29	Q_2559_10.pngQ_2555_9.pngQ_2556_29.pngQ_2558_22.png	1	0	-
30	Q_2559_10.pngQ_2555_9.pngQ_2556_29.pngQ_2559_16.png	1	0	-
31	Q_2559_10.pngQ_2555_9.pngQ_2557_1.pngQ_2552_4.png	1	0	-
32	Q_2559_10.pngQ_2555_9.pngQ_2557_1.pngQ_2557_27.png	1	0	-
33	Q_2559_10.pngQ_2555_9.pngQ_2557_1.pngQ_2557_32.png	1	0	-
34	Q_2559_10.pngQ_2555_9.pngQ_2557_1.pngQ_2558_22.png	1	0	-
35	Q_2559_10.pngQ_2555_9.pngQ_2557_27.pngQ_2559_16.png	1	0	-
36	Q_2559_10.pngQ_2555_9.pngQ_2557_32.pngQ_2557_1.png	1	0	-
37	Q_2559_10.pngQ_2555_9.pngQ_2557_32.pngQ_2559_16.png	1	0	-
38	Q_2559_10.pngQ_2555_9.pngQ_2557_32.pngQ_2559_17.png	1	0	-
39	Q_2559_10.pngQ_2555_9.pngQ_2559_14.pngQ_2557_27.png	1	0	-
40	Q_2559_10.pngQ_2555_9.pngQ_2559_17.pngQ_2559_16.png	1	0	-
41	Q_2559_10.pngQ_2556_28.pngQ_2552_4.pngQ_2559_17.png	1	0	-
42	Q_2559_10.pngQ_2556_28.pngQ_2559_14.pngQ_2559_16.png	1	0	-
43	Q_2559_10.pngQ_2556_29.pngQ_2552_4.pngQ_2559_17.png	1	0	-
44	Q_2559_10.pngQ_2556_29.pngQ_2555_17.pngQ_2556_28.png	1	0	-
45	Q_2559_10.pngQ_2556_29.pngQ_2555_9.pngQ_2552_4.png	1	0	-
46	Q_2559_10.pngQ_2556_29.pngQ_2555_9.pngQ_2555_17.png	1	0	-
47	Q_2559_10.pngQ_2556_29.pngQ_2556_28.pngQ_2557_27.png	1	0	-
48	Q_2559_10.pngQ_2556_29.pngQ_2556_28.pngQ_2558_22.png	1	0	-
49	Q_2559_10.pngQ_2556_29.pngQ_2557_1.pngQ_2557_27.png	1	0	-
50	Q_2559_10.pngQ_2556_29.pngQ_2557_1.pngQ_2558_22.png	1	0	-
51	Q_2559_10.pngQ_2556_29.pngQ_2557_1.pngQ_2559_16.png	1	0	-
52	Q_2559_10.pngQ_2556_29.pngQ_2557_32.pngQ_2556_28.png	1	0	-
53	Q_2559_10.pngQ_2556_29.pngQ_2557_32.pngQ_2559_14.png	1	0	-
54	Q_2559_10.pngQ_2556_29.pngQ_2557_32.pngQ_2559_16.png	1	0	-
55	Q_2559_10.pngQ_2556_29.pngQ_2557_32.pngQ_2559_17.png	1	0	-
56	Q_2559_10.pngQ_2556_29.pngQ_2559_14.pngQ_2556_28.png	1	0	-
57	Q_2559_10.pngQ_2556_29.pngQ_2559_14.pngQ_2558_22.png	1	0	-
58	Q_2559_10.pngQ_2556_29.pngQ_2559_14.pngQ_2559_17.png	1	0	-
59	Q_2559_10.pngQ_2556_29.pngQ_2559_17.pngQ_2558_22.png	1	0	-
60	Q_2559_10.pngQ_2557_1.pngQ_2552_4.pngQ_2557_27.png	1	0	-

ตารางที่ ง3 (ต่อ)

ชุดที่	รายการข้อสอบ	จำนวนผู้สอบ (คน)	Item Overlap rate	tem Exposue Rat Distribution
61	Q_2559_10.pngQ_2557_1.pngQ_2556_28.pngQ_2557_27.png	1	0	-
62	Q_2559_10.pngQ_2557_1.pngQ_2556_28.pngQ_2558_22.png	1	0	-
63	Q_2559_10.pngQ_2557_1.pngQ_2557_27.pngQ_2559_16.png	1	0	-
64	Q_2559_10.pngQ_2557_1.pngQ_2557_32.pngQ_2559_14.png	1	0	-
65	Q_2559_10.pngQ_2557_1.pngQ_2557_32.pngQ_2559_17.png	1	0	-
66	Q_2559_10.pngQ_2557_1.pngQ_2559_14.pngQ_2559_16.png	1	0	-
67	Q_2559_10.pngQ_2557_1.pngQ_2559_17.pngQ_2558_22.png	1	0	-
68	Q_2559_10.pngQ_2557_1.pngQ_2559_17.pngQ_2559_16.png	1	0	-
69	Q_2559_10.pngQ_2557_32.pngQ_2552_4.pngQ_2558_22.png	1	0	-
70	Q_2559_10.pngQ_2557_32.pngQ_2552_4.pngQ_2559_14.png	1	0	-
71	Q_2559_10.pngQ_2557_32.pngQ_2552_4.pngQ_2559_16.png	1	0	-
72	Q_2559_10.pngQ_2557_32.pngQ_2552_4.pngQ_2559_17.png	1	0	-
73	Q_2559_10.pngQ_2557_32.pngQ_2556_28.pngQ_2552_4.png	1	0	-
74	Q_2559_10.pngQ_2557_32.pngQ_2556_28.pngQ_2557_27.png	1	0	-
75	Q_2559_10.pngQ_2557_32.pngQ_2556_28.pngQ_2558_22.png	1	0	-
76	Q_2559_10.pngQ_2557_32.pngQ_2556_28.pngQ_2559_16.png	1	0	-
77	Q_2559_10.pngQ_2557_32.pngQ_2557_1.pngQ_2559_17.png	1	0	-
78	Q_2559_10.pngQ_2557_32.pngQ_2559_14.pngQ_2559_16.png	1	0	-
79	Q_2559_10.pngQ_2557_32.pngQ_2559_14.pngQ_2559_17.png	1	0	-
80	Q_2559_10.pngQ_2552_4.pngQ_2556_28.pngQ_2557_32.png	2	0.4	0.142222
81	Q_2559_10.pngQ_2552_4.pngQ_2556_28.pngQ_2559_14.png	2	0.4	0.142222
82	Q_2559_10.pngQ_2552_4.pngQ_2556_28.pngQ_2559_16.png	2	0.4	0.142222
83	Q_2559_10.pngQ_2552_4.pngQ_2557_32.pngQ_2556_28.png	2	0.4	0.142222
84	Q_2559_10.pngQ_2552_4.pngQ_2557_32.pngQ_2559_14.png	2	0.4	0.142222
85	Q_2559_10.pngQ_2552_4.pngQ_2559_14.pngQ_2556_28.png	2	0.4	0.142222
86	Q_2559_10.pngQ_2552_4.pngQ_2559_17.pngQ_2559_16.png	2	0.4	0.142222
87	Q_2559_10.pngQ_2555_17.pngQ_2552_4.pngQ_2556_28.png	2	0.4	0.142222
88	Q_2559_10.pngQ_2555_17.pngQ_2552_4.pngQ_2557_27.png	2	0.4	0.142222
89	Q_2559_10.pngQ_2555_17.pngQ_2552_4.pngQ_2559_17.png	2	0.4	0.142222
90	Q_2559_10.pngQ_2555_17.pngQ_2553_10.pngQ_2556_28.png	2	0.4	0.142222
91	Q_2559_10.pngQ_2555_17.pngQ_2556_28.pngQ_2558_22.png	2	0.4	0.142222
92	Q_2559_10.pngQ_2555_17.pngQ_2556_29.pngQ_2557_32.png	2	0.4	0.142222
93	Q_2559_10.pngQ_2555_17.pngQ_2556_29.pngQ_2559_17.png	2	0.4	0.142222
94	Q_2559_10.pngQ_2555_17.pngQ_2557_1.pngQ_2552_4.png	2	0.4	0.142222
95	Q_2559_10.pngQ_2555_17.pngQ_2557_1.pngQ_2557_32.png	2	0.4	0.142222
96	Q_2559_10.pngQ_2555_17.pngQ_2557_32.pngQ_2556_28.png	2	0.4	0.142222
97	Q_2559_10.pngQ_2555_17.pngQ_2557_32.pngQ_2558_22.png	2	0.4	0.142222

ตารางที่ ง3 (ต่อ)

ชุดที่	รายการข้อสอบ	จำนวนผู้สอบ (คน)	Item Overlap rate	tem Exposue Rat Distribution
98	Q_2559_10.pngQ_2555_17.pngQ_2559_14.pngQ_2556_28.png	2	0.4	0.142222
99	Q_2559_10.pngQ_2555_17.pngQ_2559_14.pngQ_2557_27.png	2	0.4	0.142222
100	Q_2559_10.pngQ_2555_17.pngQ_2559_14.pngQ_2559_16.png	2	0.4	0.142222
101	Q_2559_10.pngQ_2555_9.pngQ_2556_29.pngQ_2557_1.png	2	0.4	0.142222
102	Q_2559_10.pngQ_2555_9.pngQ_2556_29.pngQ_2557_27.png	2	0.4	0.142222
103	Q_2559_10.pngQ_2555_9.pngQ_2557_1.pngQ_2559_16.png	2	0.4	0.142222
104	Q_2559_10.pngQ_2555_9.pngQ_2557_32.pngQ_2559_14.png	2	0.4	0.142222
105	Q_2559_10.pngQ_2555_9.pngQ_2559_14.pngQ_2556_28.png	2	0.4	0.142222
106	Q_2559_10.pngQ_2555_9.pngQ_2559_14.pngQ_2558_22.png	2	0.4	0.142222
107	Q_2559_10.pngQ_2555_9.pngQ_2559_17.pngQ_2558_22.png	2	0.4	0.142222
108	Q_2559_10.pngQ_2556_28.pngQ_2552_4.pngQ_2558_22.png	2	0.4	0.142222
109	Q_2559_10.pngQ_2556_28.pngQ_2552_4.pngQ_2559_16.png	2	0.4	0.142222
110	Q_2559_10.pngQ_2556_29.pngQ_2552_4.pngQ_2557_32.png	2	0.4	0.142222
111	Q_2559_10.pngQ_2556_29.pngQ_2556_28.pngQ_2552_4.png	2	0.4	0.142222
112	Q_2559_10.pngQ_2556_29.pngQ_2556_28.pngQ_2559_17.png	2	0.4	0.142222
113	Q_2559_10.pngQ_2556_29.pngQ_2557_32.pngQ_2557_1.png	2	0.4	0.142222
114	Q_2559_10.pngQ_2556_29.pngQ_2559_17.pngQ_2557_27.png	2	0.4	0.142222
115	Q_2559_10.pngQ_2557_1.pngQ_2552_4.pngQ_2559_17.png	2	0.4	0.142222
116	Q_2559_10.pngQ_2557_1.pngQ_2556_28.pngQ_2559_17.png	2	0.4	0.142222
117	Q_2559_10.pngQ_2557_1.pngQ_2557_32.pngQ_2552_4.png	2	0.4	0.142222
118	Q_2559_10.pngQ_2557_1.pngQ_2557_32.pngQ_2556_28.png	2	0.4	0.142222
119	Q_2559_10.pngQ_2557_1.pngQ_2557_32.pngQ_2559_16.png	2	0.4	0.142222
120	Q_2559_10.pngQ_2557_1.pngQ_2559_14.pngQ_2558_22.png	2	0.4	0.142222
121	Q_2559_10.pngQ_2557_1.pngQ_2559_17.pngQ_2557_27.png	2	0.4	0.142222
122	Q_2559_10.pngQ_2557_32.pngQ_2556_28.pngQ_2559_14.png	2	0.4	0.142222
123	Q_2559_10.pngQ_2557_32.pngQ_2556_28.pngQ_2559_17.png	2	0.4	0.142222
124	Q_2559_10.pngQ_2557_32.pngQ_2557_1.pngQ_2556_28.png	2	0.4	0.142222
125	Q_2559_10.pngQ_2557_32.pngQ_2557_27.pngQ_2558_22.png	2	0.4	0.142222
126	Q_2559_10.pngQ_2557_32.pngQ_2557_27.pngQ_2559_16.png	2	0.4	0.142222
127	Q_2559_10.pngQ_2557_32.pngQ_2559_14.pngQ_2557_27.png	2	0.4	0.142222
128	Q_2559_10.pngQ_2557_32.pngQ_2559_14.pngQ_2558_22.png	2	0.4	0.142222
129	Q_2559_10.pngQ_2552_4.pngQ_2556_28.pngQ_2557_27.png	3	0.6	0.642222
130	Q_2559_10.pngQ_2552_4.pngQ_2557_27.pngQ_2559_16.png	3	0.6	0.642222
131	Q_2559_10.pngQ_2552_4.pngQ_2557_32.pngQ_2557_27.png	3	0.6	0.642222
132	Q_2559_10.pngQ_2552_4.pngQ_2557_32.pngQ_2558_22.png	3	0.6	0.642222
133	Q_2559_10.pngQ_2552_4.pngQ_2559_17.pngQ_2557_27.png	3	0.6	0.642222
134	Q_2559_10.pngQ_2555_17.pngQ_2553_10.pngQ_2557_1.png	3	0.6	0.642222

ตารางที่ ง3 (ต่อ)

ชุดที่	รายการข้อสอบ	จำนวนผู้สอบ (คน)	Item Overlap rate	tem Exposue Rat Distribution
135	Q_2559_10.pngQ_2555_17.pngQ_2553_10.pngQ_2559_17.png	3	0.6	0.642222
136	Q_2559_10.pngQ_2555_17.pngQ_2556_28.pngQ_2552_4.png	3	0.6	0.642222
137	Q_2559_10.pngQ_2555_17.pngQ_2556_28.pngQ_2557_27.png	3	0.6	0.642222
138	Q_2559_10.pngQ_2555_17.pngQ_2556_29.pngQ_2553_10.png	3	0.6	0.642222
139	Q_2559_10.pngQ_2555_17.pngQ_2556_29.pngQ_2557_1.png	3	0.6	0.642222
140	Q_2559_10.pngQ_2555_17.pngQ_2557_27.pngQ_2559_16.png	3	0.6	0.642222
141	Q_2559_10.pngQ_2555_17.pngQ_2557_32.pngQ_2557_1.png	3	0.6	0.642222
142	Q_2559_10.pngQ_2555_17.pngQ_2557_32.pngQ_2559_16.png	3	0.6	0.642222
143	Q_2559_10.pngQ_2555_17.pngQ_2557_32.pngQ_2559_17.png	3	0.6	0.642222
144	Q_2559_10.pngQ_2555_17.pngQ_2559_14.pngQ_2558_22.png	3	0.6	0.642222
145	Q_2559_10.pngQ_2555_9.pngQ_2552_4.pngQ_2557_32.png	3	0.6	0.642222
146	Q_2559_10.pngQ_2555_9.pngQ_2556_29.pngQ_2559_14.png	3	0.6	0.642222
147	Q_2559_10.pngQ_2555_9.pngQ_2559_17.pngQ_2557_27.png	3	0.6	0.642222
148	Q_2559_10.pngQ_2556_28.pngQ_2552_4.pngQ_2557_27.png	3	0.6	0.642222
149	Q_2559_10.pngQ_2556_28.pngQ_2552_4.pngQ_2557_32.png	3	0.6	0.642222
150	Q_2559_10.pngQ_2556_28.pngQ_2552_4.pngQ_2559_14.png	3	0.6	0.642222
151	Q_2559_10.pngQ_2556_28.pngQ_2559_14.pngQ_2557_27.png	3	0.6	0.642222
152	Q_2559_10.pngQ_2556_28.pngQ_2559_14.pngQ_2558_22.png	3	0.6	0.642222
153	Q_2559_10.pngQ_2556_29.pngQ_2555_17.pngQ_2557_27.png	3	0.6	0.642222
154	Q_2559_10.pngQ_2556_29.pngQ_2555_17.pngQ_2559_17.png	3	0.6	0.642222
155	Q_2559_10.pngQ_2556_29.pngQ_2555_9.pngQ_2559_14.png	3	0.6	0.642222
156	Q_2559_10.pngQ_2557_1.pngQ_2552_4.pngQ_2559_16.png	3	0.6	0.642222
157	Q_2559_10.pngQ_2557_1.pngQ_2559_14.pngQ_2556_28.png	3	0.6	0.642222
158	Q_2559_10.pngQ_2557_32.pngQ_2552_4.pngQ_2556_28.png	3	0.6	0.642222
159	Q_2559_10.pngQ_2557_32.pngQ_2557_1.pngQ_2558_22.png	3	0.6	0.642222
160	Q_2559_10.pngQ_2557_32.pngQ_2557_1.pngQ_2559_16.png	3	0.6	0.642222
161	Q_2559_10.pngQ_2557_32.pngQ_2559_17.pngQ_2558_22.png	3	0.6	0.642222
162	Q_2559_10.pngQ_2559_14.pngQ_2556_28.pngQ_2557_27.png	3	0.6	0.642222
163	Q_2559_10.pngQ_2559_14.pngQ_2556_28.pngQ_2559_17.png	3	0.6	0.642222
164	Q_2559_10.pngQ_2552_4.pngQ_2559_14.pngQ_2559_16.png	4	0.8	1.502222
165	Q_2559_10.pngQ_2555_17.pngQ_2555_9.pngQ_2556_29.png	4	0.8	1.502222
166	Q_2559_10.pngQ_2555_17.pngQ_2557_27.pngQ_2558_22.png	4	0.8	1.502222
167	Q_2559_10.pngQ_2555_17.pngQ_2559_17.pngQ_2559_16.png	4	0.8	1.502222
168	Q_2559_10.pngQ_2556_28.pngQ_2559_14.pngQ_2552_4.png	4	0.8	1.502222
169	Q_2559_10.pngQ_2556_29.pngQ_2555_17.pngQ_2559_14.png	4	0.8	1.502222
170	Q_2559_10.pngQ_2556_29.pngQ_2557_27.pngQ_2558_22.png	4	0.8	1.502222
171	Q_2559_10.pngQ_2556_29.pngQ_2559_17.pngQ_2559_16.png	4	0.8	1.502222

ตารางที่ ง3 (ต่อ)

ชุดที่	รายการข้อสอบ	จำนวนผู้สอบ (คน)	Item Overlap rate	tem Exposue Rat Distribution
172	Q_2559_10.pngQ_2557_32.pngQ_2557_1.pngQ_2552_4.png	4	0.8	1.502222
173	Q_2559_10.pngQ_2557_32.pngQ_2559_17.pngQ_2559_16.png	4	0.8	1.502222
174	Q_2559_10.pngQ_2559_14.pngQ_2556_28.pngQ_2552_4.png	4	0.8	1.502222
175	Q_2559_10.pngQ_2559_14.pngQ_2556_28.pngQ_2558_22.png	4	0.8	1.502222
176	Q_2559_10.pngQ_2552_4.pngQ_2557_32.pngQ_2559_16.png	5	1	2.722222
177	Q_2559_10.pngQ_2555_17.pngQ_2559_17.pngQ_2557_27.png	5	1	2.722222
178	Q_2559_10.pngQ_2556_28.pngQ_2557_27.pngQ_2559_16.png	5	1	2.722222
179	Q_2559_10.pngQ_2556_28.pngQ_2559_17.pngQ_2557_27.png	5	1	2.722222
180	Q_2559_10.pngQ_2556_28.pngQ_2559_17.pngQ_2558_22.png	5	1	2.722222
181	Q_2559_10.pngQ_2556_29.pngQ_2555_17.pngQ_2553_10.png	5	1	2.722222
182	Q_2559_10.pngQ_2557_1.pngQ_2557_27.pngQ_2558_22.png	5	1	2.722222
183	Q_2559_10.pngQ_2557_1.pngQ_2557_32.pngQ_2555_9.png	5	1	2.722222
184	Q_2559_10.pngQ_2557_32.pngQ_2557_1.pngQ_2559_14.png	5	1	2.722222
185	Q_2559_10.pngQ_2557_32.pngQ_2559_17.pngQ_2557_27.png	5	1	2.722222
186	Q_2559_10.pngQ_2552_4.pngQ_2557_32.pngQ_2557_1.png	6	1.2	4.302222
187	Q_2559_10.pngQ_2555_9.pngQ_2557_27.pngQ_2558_22.png	6	1.2	4.302222
188	Q_2559_10.pngQ_2559_14.pngQ_2556_28.pngQ_2559_16.png	6	1.2	4.302222
189	Q_2559_10.pngQ_2559_14.pngQ_2557_27.pngQ_2558_22.png	6	1.2	4.302222
190	Q_2559_10.pngQ_2559_14.pngQ_2559_17.pngQ_2557_27.png	6	1.2	4.302222
191	Q_2559_10.pngQ_2559_14.pngQ_2559_17.pngQ_2559_16.png	6	1.2	4.302222
192	Q_2559_10.pngQ_2555_17.pngQ_2559_17.pngQ_2558_22.png	7	1.4	6.242222
193	Q_2559_10.pngQ_2556_28.pngQ_2559_14.pngQ_2559_17.png	7	1.4	6.242222
194	Q_2559_10.pngQ_2556_28.pngQ_2557_27.pngQ_2558_22.png	8	1.6	8.542222
195	Q_2559_10.pngQ_2557_1.pngQ_2552_4.pngQ_2557_32.png	8	1.6	8.542222
196	Q_2559_10.pngQ_2559_14.pngQ_2559_17.pngQ_2558_22.png	9	1.8	11.20222
197	Q_2559_10.pngQ_2555_9.pngQ_2556_29.pngQ_2555_17.png	10	2	14.22222
198	Q_2559_10.pngQ_2556_28.pngQ_2559_17.pngQ_2559_16.png	10	2	14.22222
199	Q_2559_10.pngQ_2559_14.pngQ_2557_27.pngQ_2559_16.png	12	2.4	21.34222
200	Q_2559_10.pngQ_2555_17.pngQ_2553_10.pngQ_2555_19.png	13	2.6	25.44222
	ผลรวมทั้งหมด	500	84.2	216.5089

การคำนวณการแจกแจงอัตราการแสดงข้อสอบ

การแจกแจงอัตราการแสดงข้อสอบ (Item Exposure Rate Distribution) หมายถึง ลักษณะการกระจายอัตราการแสดงของข้อสอบในคลังข้อสอบบ่งบอกถึงประสิทธิภาพของการใช้ข้อสอบในคลังข้อสอบโดยรวม พิจารณาจากสถิติไค-สแควร์ ซึ่งคำนวณได้ดังสมการ 61 หน้า 69

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

เมื่อ er_j แทน อัตราการแสดงข้อสอบสำหรับข้อสอบที่ j

L แทน ความยาวแบบทดสอบ

N แทน ขนาดคลังข้อสอบ

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

er_j เป็นค่าสังเกต ขณะที่ $\frac{L}{N}$ เป็นค่าคาดหวังว่าข้อสอบทุกข้อมีอัตราการแสดงข้อสอบเท่าเทียมกัน ดังนั้น สมการแสดงถึงความแตกต่างระหว่างค่าสังเกตและค่าคาดหวังของอัตราการแสดงข้อสอบ ถ้าค่า χ^2 เล็ก แสดงถึง การใช้ข้อสอบในคลังข้อสอบมีความสมดุลกัน

ตัวอย่างการคำนวณการแจกแจงอัตราการแสดงข้อสอบ

กรณีการแจกแจงอัตราการแสดงข้อสอบ วิธี CWA ในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ สามารถจัดชุดข้อสอบได้ 200 ชุดข้อสอบ ซึ่งการแจกแจงอัตราการแสดงข้อสอบ จะคำนวณจากชุดข้อสอบที่มีผู้สอบได้รับชุดข้อสอบร่วมกัน (2 คนขึ้นไป) โดยนำชุดข้อสอบชุดที่ 80 – 200 มาคำนวณ เช่น ข้อสอบชุดที่ 80 มีจำนวนผู้ใช้ข้อสอบร่วมกัน 2 คน จากผู้สอบทั้งหมด 500 คน คือ $\left(\frac{\left(\left(\frac{2}{500} \right) * 100 \right) - \frac{4}{18}}{\frac{4}{18}} \right) = 0.1422$ และข้อสอบชุดที่ 60 มีจำนวนผู้ใช้ข้อสอบร่วมกัน 13 คน จากผู้สอบทั้งหมด 500 คน คือ $\left(\frac{\left(\left(\frac{13}{500} \right) * 100 \right) - \frac{4}{18}}{\frac{4}{18}} \right) = 25.4422$ ถ้าจะหาการแจกแจงอัตราการแสดงข้อสอบ ทั้งชุดคือผลรวมของการแจกแจงอัตราการแสดงข้อสอบ ทั้งหมด

$$\chi^2 = \left(0.1422 + 0.1422 + 0.1422 + ... + 25.4422 \right)$$

$$\chi^2 = 216.509$$

ดังนั้น สมการแสดงถึงความแตกต่างระหว่างค่าสังเกตและค่าคาดหวังของอัตราการแสดงข้อสอบ ถ้าค่า χ^2 เล็ก แสดงถึง การใช้ข้อสอบในคลังข้อสอบมีความสมดุลกัน

การเปรียบเทียบการแจกแจงอัตราการแสดงข้อสอบ 2 วิธี ได้แก่ วิธีที่ 1 และวิธีที่ 2 โดยใช้อัตราส่วน F (F - ratio) จะพิจารณาจากอัตราส่วนของค่า χ^2 วิธีที่ 1 หารด้วย χ^2 วิธีที่ 2 นั่นคือ ซึ่งคำนวณได้ดังสมการ 62 หน้า 70

$$F_{1.2} = \frac{\chi_1^2}{\chi_2^2}$$

เมื่อ $F_{1.2}$ แทน อัตราส่วนการแจกแจงอัตราการแสดงข้อสอบ2วิธีหรือวิธีที่1หารด้วยวิธีที่ 2

χ_1^2 แทน การแจกแจงอัตราการแสดงข้อสอบวิธีที่ 1

χ_2^2 แทน การแจกแจงอัตราการแสดงข้อสอบวิธีที่ 2

ถ้า $F < 1$ แสดงว่า วิธีที่ 1 ให้สมดุลของอัตราการแสดงข้อสอบโดยรวมดีกว่าวิธีที่ 2

ตัวอย่างการเปรียบเทียบการแจกแจงอัตราการแสดงข้อสอบ 2 วิธี

จากการคำนวนแจกแจงอัตราการแสดงข้อสอบ 2 วิธี จากตารางที่ ค3 แสดงค่า $\chi^2_{NBCA} = 1666.240$ และจากตาราง ค4 แสดงค่า $\chi^2_{CWA} = 216.509$ โดยใช้อัตราส่วน F (F - ratio) จะพิจารณาจากอัตราส่วนของค่า χ^2 วิธีที่ 1 หารด้วย χ^2 วิธีที่ 2 นั่นคือ ซึ่งคำนวณได้ดังสมการ

$$F_{1.2} = \frac{\chi_1^2}{\chi_2^2}$$

$$F_{1.2} = \frac{1666.240}{216.509}$$

$$F_{1.2} = 7.696$$

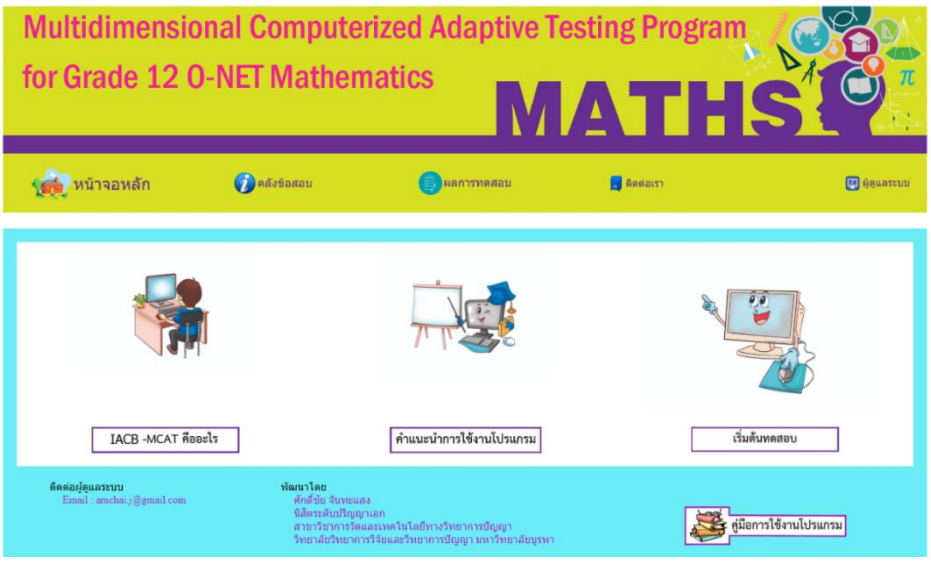
ปรากฏว่า ค่า $F_{ratio} > 1$ แสดงว่า วิธีที่ CWA ให้สมดุลของอัตราการแสดงข้อสอบโดยรวมสำหรับคลังข้อสอบในชั้นที่ 1 ดีกว่าวิธีที่ IACB

ภาคผนวก จ

คู่มือการใช้โปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ สำหรับการทดสอบ O-NET วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 6

คู่มือการใช้โปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ สำหรับการทดสอบ O-NET วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 6

User Guide for Multidimensional Computerized Adaptive testing Program for Grade 12 O-NET Mathematics



สาขาวิชาการวัดและเทคโนโลยีทางวิทยาการปัญญา
วิทยาลัยวิทยาการวิจัยและวิทยาการปัญญา มหาวิทยาลัยบูรพา
กรกฎาคม 2561
ลิขสิทธิ์เป็นของมหาวิทยาลัยบูรพา

คำนำ

คู่มือการใช้โปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ สำหรับการทดสอบ O-NET วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 6 เล่มนี้ แบ่งออกเป็น 3 ส่วน คือ 1) ข้อมูลทั่วไปเกี่ยวกับโปรแกรม เป็นที่อธิบายส่วนประกอบต่างๆของโปรแกรม 2) การจัดการของผู้ดูแลระบบ เป็นส่วนที่ครู อาจารย์ หรือผู้ที่เกี่ยวข้องในการจัดการข้อสอบ นำข้อสอบที่ผ่านการวิเคราะห์คุณภาพข้อสอบตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ (Multidimensional Item Response Theory: MIRT) และจัดข้อสอบตามชั้นของค่าอำนาจจำแนกของข้อสอบและตามกลุ่มเนื้อหาวิชา มาใส่ในโปรแกรมเพื่อใช้เป็นคลังข้อสอบ รวมทั้งการลบ และการแก้ไขข้อสอบเดิมที่มีอยู่ในคลังข้อสอบ นอกจากนี้ยังมีรายงานผลการทดสอบของผู้สอบทั้งหมด เพื่อให้ครู อาจารย์ และผู้เกี่ยวข้องนำผลการทดสอบไปใช้ในการประเมินผลการเรียนรู้ต่อไป 3) การจัดการสอบ เป็นส่วนที่ให้ผู้สอบ คือนักเรียนใช้โปรแกรมนี้เพื่อทดสอบความสามารถทางวิชาคณิตศาสตร์ ซึ่งเมื่อผู้สอบทำการทดสอบเสร็จสิ้นรายงานผลการทดสอบให้ผู้สอบทราบทันที

คู่มือการใช้โปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ สำหรับทดสอบ O-NET วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 6 เล่มนี้ นอกจากจะใช้ประกอบการใช้โปรแกรมแล้ว ยังเหมาะกับผู้สนใจพัฒนาโปรแกรมเดี่ยวกับการวัดผลอีกด้วย และหากคู่มือการใช้โปรแกรมเล่มนี้มีข้อผิดพลาดประการใด ผู้วิจัยขออภัยมา ณ ที่นี้ด้วย

ศักดิ์ชัย จันทะแสง

สารบัญ

เรื่อง		หน้า
	คำนำ	ก
	สารบัญ	ข
	สารบัญภาพ	ค-ง
1	คู่มือการใช้โปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ	1
	วัตถุประสงค์ของโปรแกรม	2
	ประโยชน์ที่คาดว่าจะได้รับ	3
2	วิธีการใช้โปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ	4
	ส่วนที่ 1 ข้อมูลทั่วไปเกี่ยวกับโปรแกรม	4
	หน้าจอหลักของโปรแกรม	5
	หน้าจอคลังข้อสอบ	8
	หน้าจอเข้าสู่ระบบค้นหาผลการทดสอบ	10
	หน้าจอติดต่อเรา	12
	หน้าจอเริ่มต้นผู้ดูแลระบบ	12
	ส่วนที่ 2 การจัดการของผู้ดูแลระบบ	13
	การเพิ่มข้อสอบ	16
	การแก้ไขข้อสอบ	17
	การลบข้อสอบ	18
	ส่วนที่ 3 การจัดการสอบ	19
	เริ่มต้นการทดสอบ	19
	การดำเนินการสอบ	21
	รายงานผลการทดสอบ	21
	แบบประเมินความเหมาะสมของโปรแกรม	24

สารบัญภาพ

ภาพที่		หน้า
1	หน้าจอหลักของโปรแกรมการทดสอบ.....	4
2	หน้าจอ IACB -MCAT คืออะไร.....	5
3	หน้าจอคำแนะนำการใช้งานโปรแกรม.....	5
4	หน้าจอเริ่มต้นทดสอบ.....	6
5	หน้าจอลงทะเบียนเพื่อเข้าใช้งานระบบ.....	6
6	หน้าจอคลังข้อสอบ.....	7
7	หน้าจอความหมายของคลังข้อสอบ.....	8
8	หน้าจอรูปแบบการจัดคลังข้อสอบ.....	8
9	หน้าจอเกณฑ์การประเมินความสามารถของผู้สอ.....	9
10	หน้าจอเข้าสู่ระบบค้นหาผลการทดสอบ.....	9
11	หน้าจอรายงานผลการทดสอบ.....	10
12	หน้าจอติดต่อเรา.....	11
13	หน้าจอเริ่มต้นผู้ดูแลระบบ.....	11
14	หน้าจอหลักของโปรแกรมการทดสอใ.....	13
15	หน้าจอเริ่มต้นผู้ดูแลระบบ.....	13
16	หน้าจอค้นหาข้อมู.....	14
17	หน้าจอแสดงผลการค้นห.....	14
18	หน้าจอการเพิ่มข้อสอ.....	15
19	หน้าจอแก้ไขข้อสอบ.....	16
20	หน้าเริ่มต้นเข้าสู่ระบบทดสอบ.....	18
21	หน้าจอลืมรหัสผ่าน.....	19
22	หน้าจอลงทะเบียนเพื่อเข้าใช้งานในระบ.....	19
23	หน้าจอข้อสอบข้อแรก.....	20
24	หน้าจอรายงานผลการทดสอบ.....	21
25	หน้าจอรายงานผลการตอบข้อสอบเป็นรายข้อจำแนกตามชั้นของค่าอำนาจจำแนกของข้อสอ.....	22
26	หน้าจอรายงานผลการตอบจำแนกตามจุดประสงค์ทางการศึกษาด้านพุทธิปัญญา	22
27	หน้าจอรายงานผลการวิเคราะห์ความสามารถของผู้สอบจำแนกตามชั้นของค่าอำนาจจำแนกข้อสอบ.....	23
28	หน้าจอพิมพ์รายงาน.....	24
29	แบบประเมินความเหมาะสมของโปรแกรม.....	24
30	แบบประเมินความเหมาะสมของโปรแกรม (สำหรับผู้เชี่ยวชาญ).....	25
31	แบบประเมินความเหมาะสมของโปรแกรม (สำหรับผู้ใช้งาน).....	26

คู่มือการใช้โปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ

(User Guide for Multidimensional Computerized Adaptive Testing)

คู่มือการใช้โปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ สำหรับทดสอบ O-NET วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 6 เล่มนี้ แบ่งออกเป็น 3 ส่วน คือ 1) ข้อมูลทั่วไปเกี่ยวกับโปรแกรม เป็นที่อธิบายส่วนประกอบต่างๆของโปรแกรม 2) การจัดการของผู้ดูแลระบบ เป็นส่วนที่ครู อาจารย์ หรือผู้ที่เกี่ยวข้องในการจัดการข้อสอบ นำข้อสอบที่ผ่านการวิเคราะห์คุณภาพข้อสอบตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ (Multidimensional Item Response Theory : MIRT) และจัดข้อสอบตามชั้นของค่าอำนาจจำแนกของข้อสอบและตามกลุ่มเนื้อหาวิชา มาใส่ในโปรแกรมเพื่อใช้เป็นคลังข้อสอบ รวมทั้งการลบ และการแก้ไขข้อสอบเดิมที่มีอยู่ในคลังข้อสอบ นอกจากนี้ยังมีรายงานผลการทดสอบของผู้สอบทั้งหมด เพื่อให้ครู อาจารย์ และผู้เกี่ยวข้องนำผลการทดสอบไปใช้ในการประเมินผลการเรียนรู้ต่อไป 3) การจัดการสอบ เป็นส่วนที่ให้ผู้สอบ คือนักเรียนใช้โปรแกรมนี้เพื่อทดสอบความสามารถทางวิชาคณิตศาสตร์ ซึ่งเมื่อผู้สอบทำการทดสอบเสร็จสิ้น จะรายงานผลการทดสอบให้ผู้สอบทราบทันที

โปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ สำหรับทดสอบ O-NET วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 6 เป็นโปรแกรมที่พัฒนาขึ้น โดยมีวัตถุประสงค์เพื่อใช้เป็นเครื่องมือในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ สำหรับทดสอบ O-NET วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 6 ตามสาระการเรียนรู้เนื้อหาการออกข้อสอบ O-NET วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 6 ได้แก่ 1) จำนวนและการดำเนินการ 2) การวัด 3) พีชคณิต และ 4) การวิเคราะห์ข้อมูลและความน่าจะเป็น โดยกระบวนการพุทธิปัญญาเป็นการค้นคว้าเกี่ยวกับกระบวนการคิด การให้เหตุผลหรือกิจกรรมทางสติปัญญาของนักเรียน แนวคิดเกี่ยวกับพุทธิปัญญาที่ได้รับการยอมรับและนำมาใช้ในการเรียนการสอนในปัจจุบันคือแนวคิดของ Bloom et al. (1956); Anderson et al. (2001, p. XXI); Krathwohl (2002, p. 213) โดยการปรับลำดับขั้นและคำศัพท์ที่ใช้ในกระบวนการพุทธิปัญญา (The Cognitive Process Dimension) มี 6 กระบวนการเหมือนเดิม แต่ 3 กระบวนการแรกเปลี่ยนชื่อเป็น จำ (Remember) เข้าใจ (Understand) และประยุกต์ (Apply) ส่วน 3 กระบวนการหลังเปลี่ยนชื่อที่มีลักษณะเป็นคำนามไปเป็นคำกริยา และสลับกระบวนการที่ 5 กับ 6 และสร้างสรรค์ (Create) เปลี่ยนชื่อมาจาก การสังเคราะห์ (Synthesis) (Anderson et al, 2001, p. XXI; Krathwohl, 2002, p. 213) และเปลี่ยนโครงสร้างจากมิติเดียวเป็นสองมิติ โดย (Anderson et al. (2001, p. XXI) และ Krathwohl (2002, p. 213) ได้เพิ่มโครงสร้างในมิติด้านความรู้ (Knowledge Dimension) เข้ามาในโครงสร้างของจุดประสงค์ทางการศึกษา ด้านพุทธิปัญญา ทำให้โครงสร้างใหม่มีลักษณะเป็นสองมิติประกอบด้วยมิติด้านกระบวนการพุทธิปัญญาและมิติด้านความรู้ และอธิบายความหมายของมิติด้านความรู้ที่เพิ่มเติมเข้ามาได้ (Krathwohl, 2002, p. 213) โดยมิติด้านความรู้ แบ่งเป็น 4 ส่วน คือ ความรู้เกี่ยวกับข้อเท็จจริง (Factual Knowledge) ความรู้เกี่ยวกับมโนทัศน์ (Conceptual Knowledge) ความรู้เกี่ยวกับวิธีดำเนินการ (Procedural Knowledge) และความรู้เกี่ยวกับอภิปัญญา (Metacognitive Knowledge) และมิติด้านกระบวนการพุทธิปัญญา แบ่งเป็น 6 กระบวนการ คือ จำ (Remember)

เข้าใจ (Understand) ประยุกต์ (Apply) วิเคราะห์ (Analyze) ประเมินค่า (Evaluate) และสร้างสรรค์ (Create) จำแนกตามจุดประสงค์ทางการศึกษาด้านพุทธิปัญญาที่มุ่งวัดและสาระการออกข้อสอบ ได้ดังนี้ 1) วิเคราะห์เกี่ยวกับข้อเท็จจริง จำนวน 11 ข้อ จำแนกเป็น สาระการเรียนรู้จำนวนและการดำเนินการ 7 ข้อ และสาระการเรียนรู้พีชคณิต 4 ข้อ 2) เข้าใจเกี่ยวกับมโนทัศน์ จำนวน 17 ข้อ จำแนกเป็น สาระการเรียนรู้จำนวนและการดำเนินการ 1 ข้อ การวัด 1 ข้อ พีชคณิต 2 ข้อ และการวิเคราะห์ข้อมูลและความน่าจะเป็น 13 ข้อ 3) ประยุกต์เกี่ยวกับมโนทัศน์ จำนวน 10 ข้อ จำแนกเป็น สาระการเรียนรู้จำนวนและการดำเนินการ 1 ข้อ การวัด 1 ข้อ พีชคณิต 4 ข้อ และการวิเคราะห์ข้อมูลและความน่าจะเป็น 4 ข้อ 4) วิเคราะห์เกี่ยวกับมโนทัศน์ จำนวน 65 ข้อ จำแนกเป็น สาระการเรียนรู้จำนวนและการดำเนินการ 14 ข้อ การวัด 1 ข้อ พีชคณิต 41 ข้อ และการวิเคราะห์ข้อมูลและความน่าจะเป็น 9 ข้อ 5) ประยุกต์เกี่ยวกับวิธีดำเนินการ จำนวน 148 ข้อ จำแนกเป็น สาระการเรียนรู้จำนวนและการดำเนินการ 22 ข้อ การวัด 28 ข้อ พีชคณิต 59 ข้อ และ การวิเคราะห์ข้อมูลและความน่าจะเป็น 39 ข้อ และ 6) วิเคราะห์เกี่ยวกับวิธีดำเนินการ จำนวน 5ข้อ จำแนกเป็น สาระการเรียนรู้พีชคณิต 5 ข้อ

ผลการวิเคราะห์คุณภาพข้อสอบ O-NET วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 6 ระหว่างปีการศึกษา 2551–2553 และปีการศึกษา 2555–2559 ที่ผ่านเกณฑ์การคัดเลือกข้อสอบตามเกณฑ์ของ Urry (1977) จำนวน 193 ข้อ จำแนกข้อสอบตามชั้นของค่าอำนาจจำแนกของข้อสอบ 4 ชั้น ได้แก่ 1) ชั้นที่ 1 0.50 - 0.99 มีข้อสอบ 18 ข้อ 2) ชั้นที่ 2 1.00 - 1.49 มีข้อสอบ 49 ข้อ 3) ชั้นที่ 3 1.50 - 1.99 มีข้อสอบ 64 ข้อ และ 4) ชั้นที่ 4 2.00 - 2.50 มีข้อสอบที่ 62 ข้อ

แนวคิดการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ มีหลักการทำงานของโปรแกรม โดยโปรแกรมจะคัดเลือกข้อสอบขั้นแรกในคลังข้อสอบชั้นที่หนึ่งให้กับผู้สอบ เริ่มต้นจากข้อสอบที่มีค่าความยากปานกลางก่อน ซึ่งผู้สอบแต่ละคนจะได้รับข้อสอบต่างกัน ขึ้นอยู่กับความสามารถของผู้สอบ หากผู้สอบตอบถูก ให้จัดข้อสอบข้อถัดไปที่มีค่าความยากสูงขึ้น แต่หากผู้สอบตอบผิด ให้จัดข้อสอบข้อถัดไปที่มีค่าความยากน้อยลง ทำเช่นนี้จนกว่าจะสิ้นสุดการทดสอบ

วัตถุประสงค์ของโปรแกรม

1. เพื่อใช้จัดทำคลังข้อสอบโดยแบ่งคลังข้อสอบออกเป็นชั้นตามค่าอำนาจจำแนกของข้อสอบแบบเป็นช่วงของแต่ละระดับชั้นร่วมกับการจัดสมดุลเนื้อหาข้อสอบ O-NET วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 6
2. เพื่อศึกษาระดับความสามารถของผู้สอบจำแนกตามชั้นของค่าอำนาจจำแนกของข้อสอบ O-NET วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 6
3. เพื่อใช้ทดสอบในรายวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 6

ประโยชน์ที่คาดว่าจะได้รับ

1. สามารถนำโปรแกรมไปใช้งานตามวัตถุประสงค์ได้อย่างมีประสิทธิภาพ
2. สามารถนำไปเป็นแนวทางในการพัฒนาโปรแกรมการทดสอบผู้เรียนในรายวิชาอื่นๆ ได้
3. ครูอาจารย์สามารถนำไปใช้ตรวจสอบความสามารถของนักเรียนในการทบทวนก่อนการทดสอบ O-NET วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 6 ได้

วิธีการใช้โปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ
สำหรับการทดสอบ O-NET วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 6

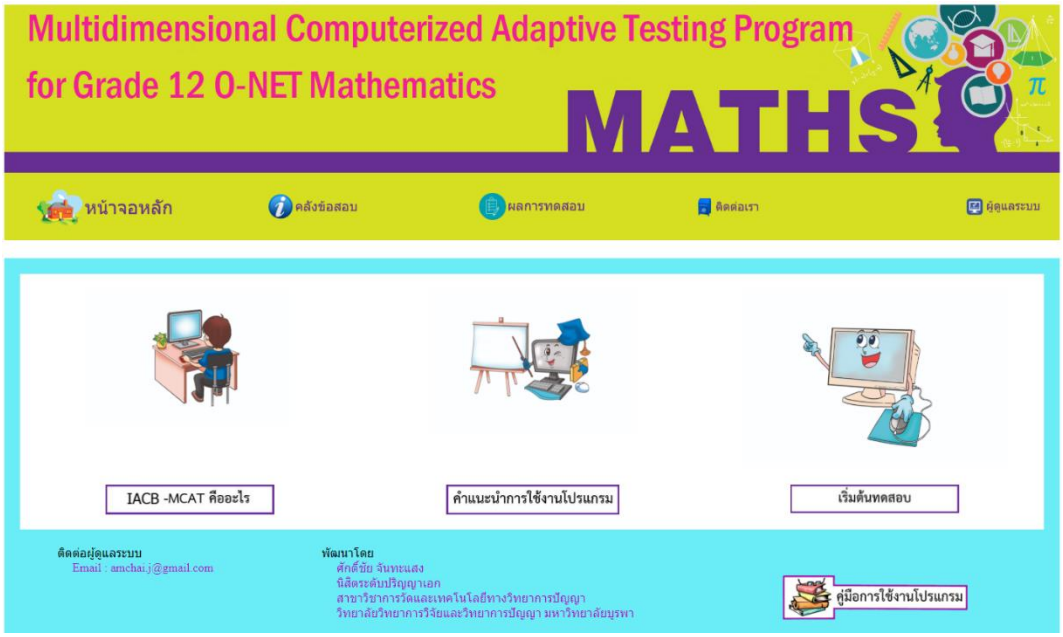
โปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์นี้ เป็น Web Application ที่พัฒนาขึ้นบนโปรแกรม Notepad++ ร่วมกับการใช้ Tools ของ Cloud9 โดยพัฒนาให้สามารถใช้งานบนอุปกรณ์ที่เชื่อมต่อกับอินเทอร์เน็ตได้ เช่น คอมพิวเตอร์ โน๊ตบุ๊คหรือสมาร์ทโฟน โปรแกรมการทดสอบนี้ แบ่งออกเป็น 3 ส่วน คือ 1) ข้อมูลทั่วไปเกี่ยวกับโปรแกรม เป็นที่อธิบายส่วนประกอบต่างๆ ของโปรแกรม 2) การจัดการของผู้ดูแลระบบ เป็นส่วนที่ครู อาจารย์ หรือผู้ที่เกี่ยวข้องในการจัดการข้อสอบ นำข้อสอบที่ผ่านการวิเคราะห์คุณภาพข้อสอบตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ (Multidimensional Item Response Theory: MIRT) และจัดข้อสอบตามชั้นของค่าอำนาจจำแนกของข้อสอบและตามกลุ่มเนื้อหาวิชา มาใส่ในโปรแกรมเพื่อใช้เป็นคลังข้อสอบ รวมทั้งการลบ และการแก้ไขข้อสอบเดิมที่มีอยู่ในคลังข้อสอบ นอกจากนี้ยังมีรายงานผลการทดสอบของผู้สอบทั้งหมด เพื่อให้ครู อาจารย์ และผู้เกี่ยวข้องนำผลการทดสอบไปใช้ในการประเมินผลการเรียนรู้ต่อไป 3) การจัดการสอบ เป็นส่วนที่ให้ผู้สอบ คือนักเรียน ใช้โปรแกรมนี้เพื่อทดสอบความสามารถทางวิชาคณิตศาสตร์ ซึ่งเมื่อผู้สอบทำการทดสอบเสร็จสิ้น จะรายงานผลการทดสอบให้ผู้สอบทราบทันที

ส่วนที่ 1 ข้อมูลทั่วไปเกี่ยวกับโปรแกรม

การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ (Multidimensional Computerized Adaptive testing: MCAT) เป็นการทดสอบด้วยคอมพิวเตอร์ ซึ่งผู้สอบแต่ละคนจะได้รับแบบทดสอบต่างชุดกัน ขึ้นอยู่กับความสามารถของผู้สอบ เริ่มต้นจากข้อสอบที่มีค่าความยากปานกลางก่อน หากผู้สอบตอบถูก ให้จัดข้อสอบข้อถัดไปที่มีค่าความยากสูงขึ้น แต่หากผู้สอบตอบผิดให้จัดข้อสอบข้อถัดไปที่มีค่าความยากน้องลง ทำเช่นนี้จนกว่าจะสิ้นสุดการทดสอบ

โปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ สำหรับทดสอบ O-NET วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 6 นี้ การพัฒนาคลังข้อสอบเริ่มจากแบ่งชั้นคลังข้อสอบเป็น 4 ชั้นตามระดับค่าอำนาจจำแนกข้อสอบ 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

การพัฒนาคลังข้อสอบเริ่มจากแบ่งชั้นคลังข้อสอบเป็น 4 ชั้นตามระดับค่าอำนาจจำแนกข้อสอบร่วมกับการจัดสมดุลเนื้อหา ในแต่ละชั้น โดยในแต่ละชั้นได้จัดกลุ่มสาระการเรียนรู้ที่เนื้อหาที่ออกข้อสอบ O-NET วิชาคณิตศาสตร์ โดยจำแนกเป็น 4 กลุ่มสาระการเรียนรู้ ได้แก่ 1) จำนวนและการดำเนินการ 2) การวัด 3) พีชคณิต และ 4) การวิเคราะห์ข้อมูลและความน่าจะเป็น ผู้ใช้งานจะเข้าใช้โปรแกรมได้จากhttp://mcatonetmath.ddns.net:8080/MCATWebApp/ สามารถแสดงหน้าเข้าใช้โปรแกรมได้ ดังภาพที่ 1



ภาพที่ 1 หน้าจอหลักของโปรแกรมการทดสอบ

จากภาพที่ 1 แสดงหน้าจอหลักของโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ ซึ่งในหน้าหลักของโปรแกรม สามารถแบ่งการใช้งานตามรูปต่างๆได้ดังนี้

หน้าจอหลัก เมื่อกดปุ่มนี้จะกลับมาที่หน้าเริ่มต้น

คลังข้อสอบ เมื่อกดปุ่มนี้จะปรากฏลิงค์ไปยังหน้า อธิบายความหมายของคลังข้อสอบ

ผลการทดสอบ เมื่อกดปุ่มนี้จะปรากฏลิงค์ไปยังหน้า เริ่มต้นเข้าสู่ระบบ แสดงผลการทดสอบ

ติดต่อเรา เมื่อกดปุ่มนี้จะปรากฏลิงค์ไปยังหน้า แสดงช่องทางการติดต่อผู้จัดทำ

ผู้ดูแลระบบ เมื่อกดปุ่มนี้จะปรากฏลิงค์ไปยังหน้า เริ่มต้นเข้าสู่ระบบ สำหรับผู้ดูแลระบบ

คู่มือการใช้งานโปรแกรม เมื่อกดปุ่มนี้จะปรากฏลิงค์ไปยังหน้า คู่มือการใช้งานโปรแกรม

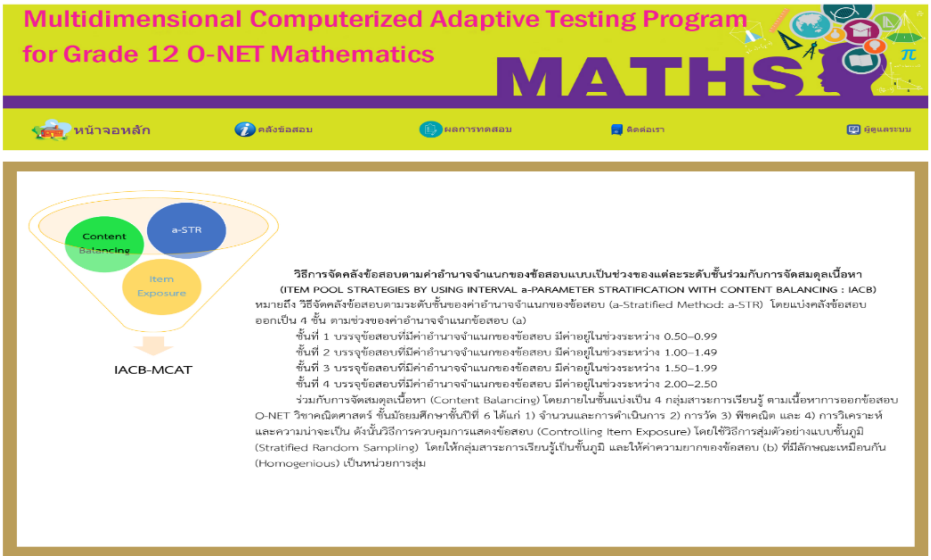
หน้าจอหลักของโปรแกรม เมื่อกดปุ่มนี้ “ หน้าจอหลัก ” จะกลับมาที่หน้าเริ่มต้น และจะมีลิงค์ไปยังหน้าต่างๆ 3 ส่วน ได้ดังนี้

IACB -MCAT คืออะไร เมื่อกดปุ่มนี้จะปรากฏลิงค์ไปยังหน้า แสดงความหมายของ IACB-MCAT

คำแนะนำการใช้งานโปรแกรม เมื่อกดปุ่มนี้จะปรากฏลิงค์ไปยังหน้า แสดงคำแนะนำการใช้งานโปรแกรม

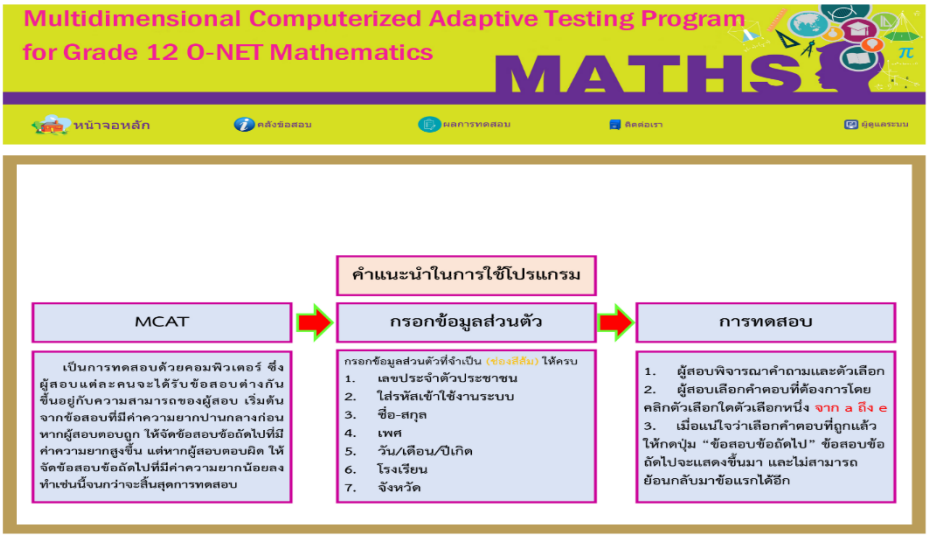
เริ่มต้นทดสอบ เมื่อกดปุ่มนี้จะปรากฏลิงค์ไปยังหน้า เริ่มต้นเข้าสู่ระบบทดสอบ

1.1) หน้าจอ IACB-MCAT คืออะไร เมื่อกดปุ่มนี้ ” IACB -MCAT คืออะไร ” จะปรากฏลิงค์ไปยังหน้า แสดงความหมายของ IACB-MCAT คืออะไร ดังภาพที่ 2



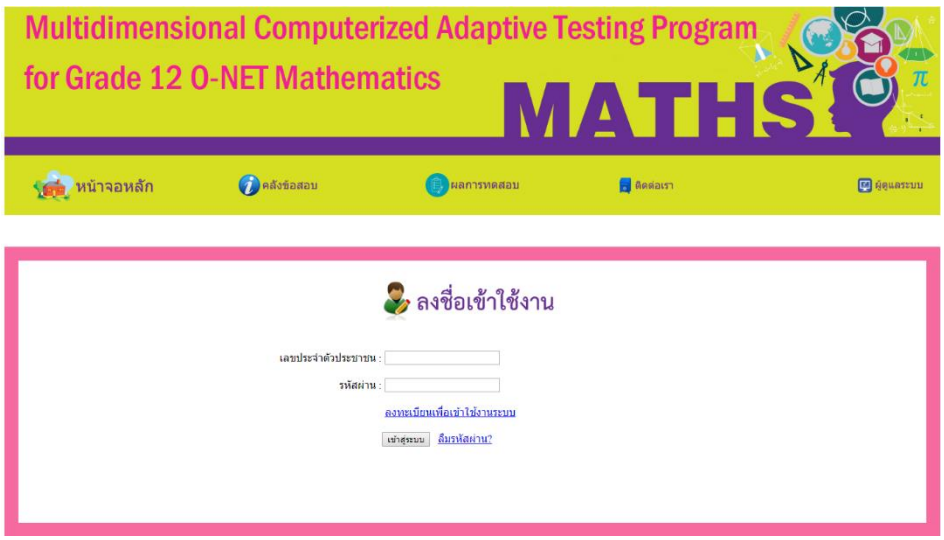
ภาพที่ 2 หน้าจอ IACB-MCAT คืออะไร

1.2) หน้าจอคำแนะนำการใช้งานโปรแกรม เมื่อกดปุ่มนี้ “ คำแนะนำการใช้งานโปรแกรม ” จะปรากฏลิงค์ไปยังหน้า แสดงคำแนะนำการใช้งานโปรแกรม ดังภาพที่ 3



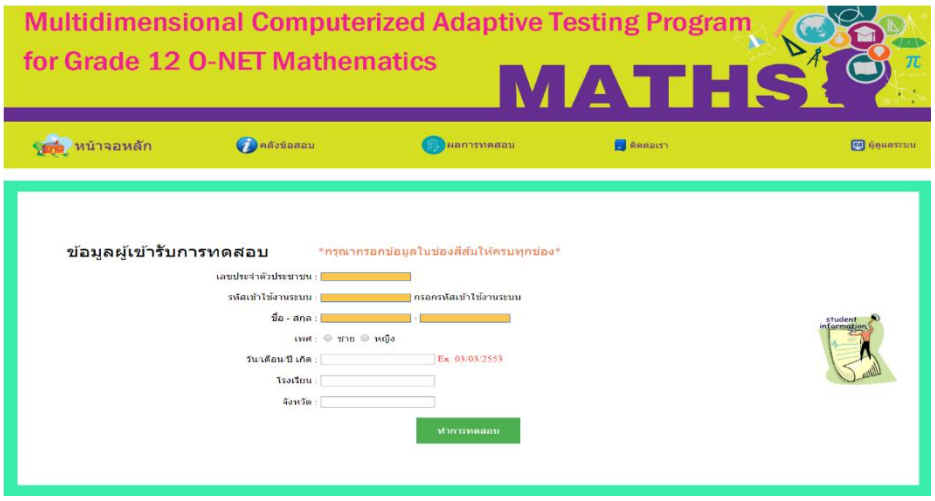
ภาพที่ 3 หน้าจอคำแนะนำการใช้งานโปรแกรม

1.3) หน้าจอเริ่มต้นทดสอบ เมื่อกดปุ่มนี้ “ เริ่มต้นทดสอบ ” จะปรากฏลิงค์ไปยังหน้า เริ่มต้นเข้าสู่ระบบทดสอบ ดังภาพที่ 4



ภาพที่ 4 หน้าจอเริ่มต้นทดสอบ

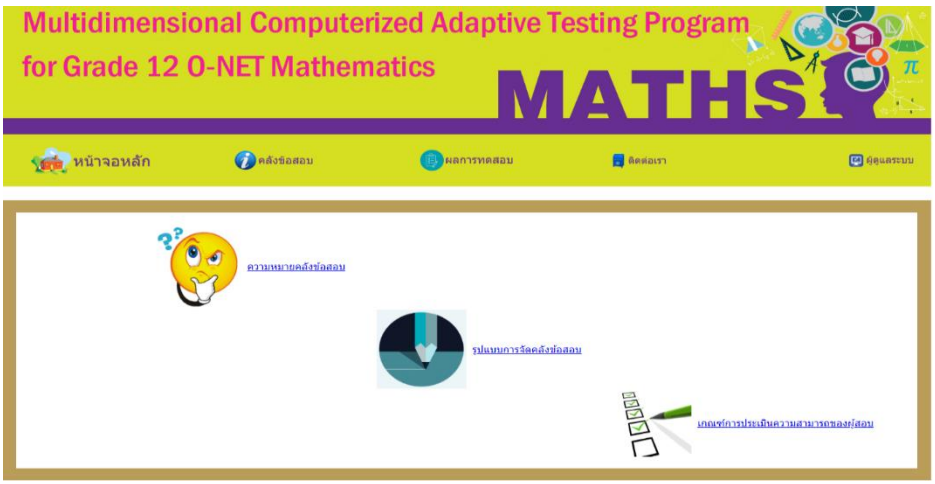
จากภาพที่ 4 หน้าจอเริ่มต้นการทดสอบ สำหรับผู้เริ่มต้นใหม่ ให้กดที่ข้อความ “ลงทะเบียนเพื่อเข้าใช้งานระบบ” หากใครเคยลงทะเบียนแล้ว สามารถใส่ เลขประจำตัวประชาชน กับ รหัสผ่านเพื่อเข้าสู่หน้าทำการทดสอบ ถ้าผู้ใช้งานลืมรหัสผ่านสามารถกดปุ่มเพื่อตั้งรหัสผ่านใหม่ได้ และหน้าจอลงทะเบียนเพื่อเข้าใช้งานในระบบ ดังภาพที่ 5



ภาพที่ 5 หน้าจอลงทะเบียนเพื่อเข้าใช้งานระบบ

จากภาพที่ 5 หน้าจอลงทะเบียนเพื่อเข้าใช้งานระบบ กรอกข้อมูลให้ครบทุกช่อง แล้วกดปุ่มทำการทดสอบ เพื่อไปยังหน้าทำการทดสอบ

หน้าจอคลังข้อสอบ เมื่อกดปุ่มนี้ “ คลังข้อสอบ ” จะปรากฏลิงค์ไปยังหน้าอธิบายความหมายของคลังข้อสอบ สามารถแบ่งการใช้งานตามรูปต่างๆได้ ดังภาพที่ 6



ภาพที่ 6 หน้าจอคลังข้อสอบ

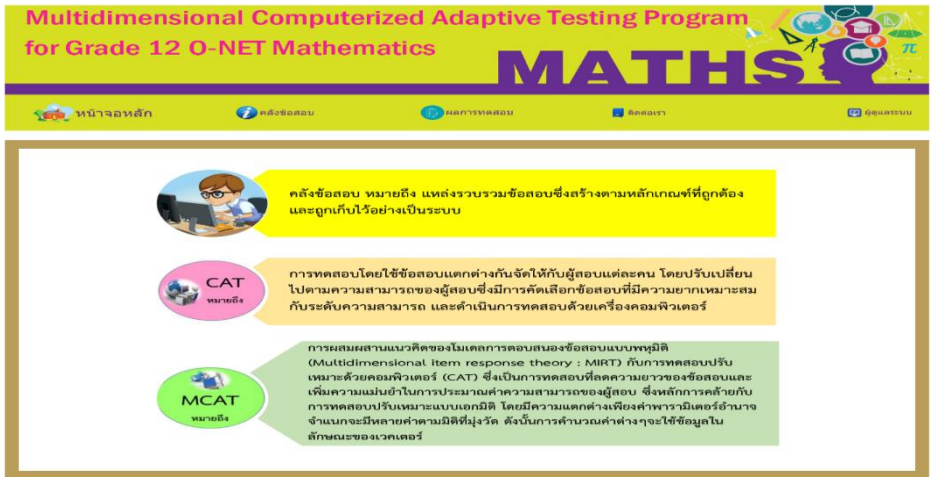
จากภาพที่ 6 หน้าจอคลังข้อสอบจะมีลิงค์ไปยังหน้าต่างๆ ได้ดังนี้

เมื่อกดปุ่ม จะปรากฏลิงค์ไปยังหน้า สดงความหมายของคลังข้อสอบ

เมื่อกดปุ่ม จะปรากฏลิงค์ไปยังหน้า แสดงความหมายของรูปแบบการจัดคลังข้อสอบ

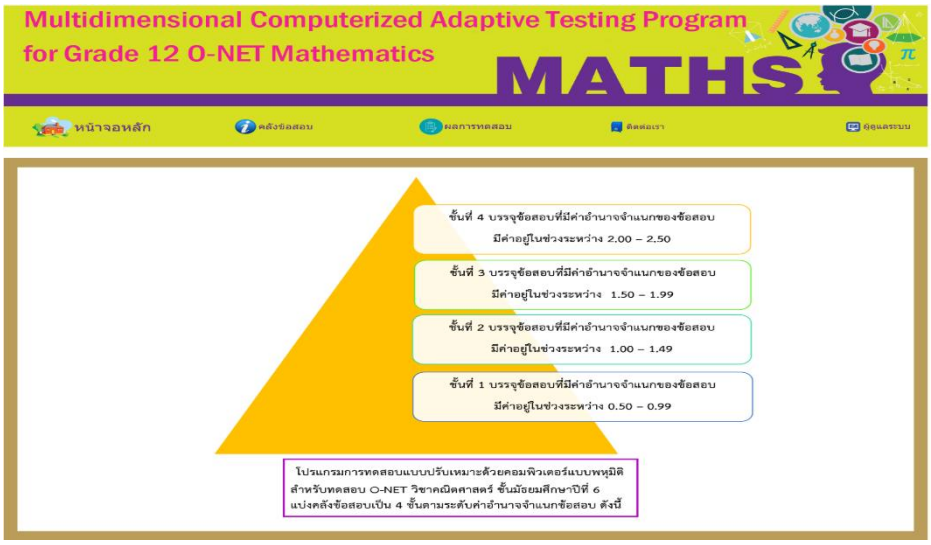
เมื่อกดปุ่ม จะปรากฏลิงค์ไปยังหน้า แสดงเกณฑ์การประเมินความสามารถของผู้สอบ

2.1) หน้าจอความหมายของคลังข้อสอบ เมื่อกดลิงค์นี้ “ ” จะไปยังหน้าแสดงความหมายของคลังข้อสอบ ดังภาพที่ 7



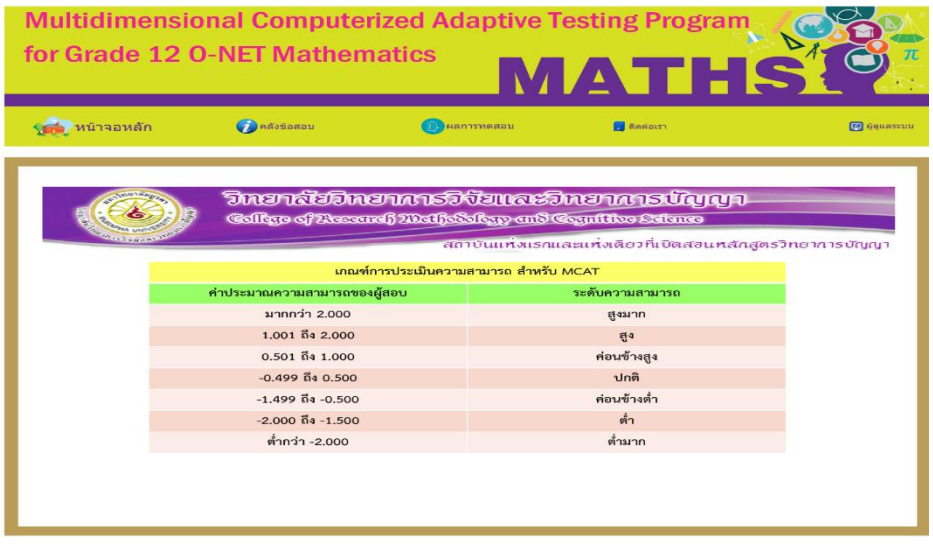
ภาพที่ 7 หน้าจอความหมายของคลังข้อสอบ

2.2) หน้าจอรูปแบบการจัดคลังเมื่อกดลิงค์นี้ “ “จะไปยังหน้า แสดงความหมายของรูปแบบการจัดคลังข้อสอบ ดังภาพที่ 8



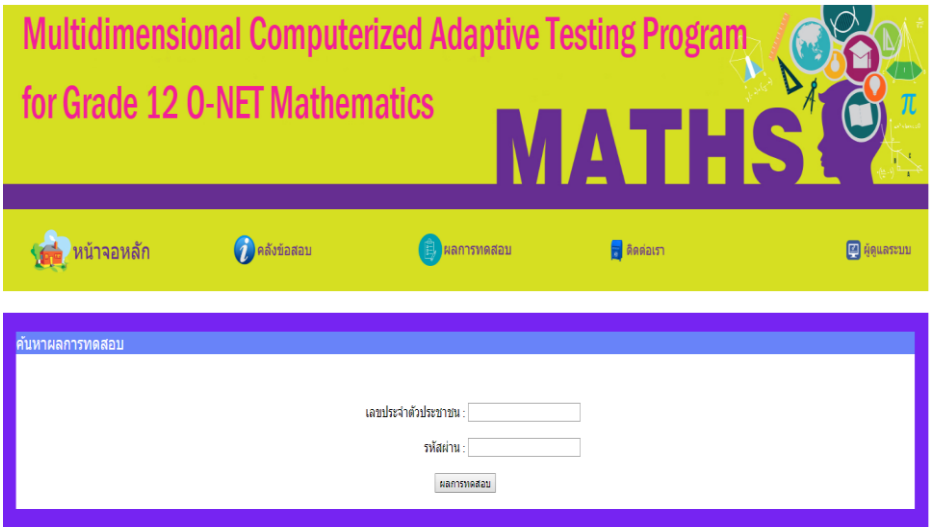
ภาพที่ 8 หน้าจอรูปแบบการจัดคลังข้อสอบ

2.3) หน้าจอเกณฑ์การประเมินความสามารถของผู้สอบ เมื่อกดลิงค์นี้ “ ” จะไปยังหน้า แสดงเกณฑ์การประเมินความสามารถของผู้สอบ ดังภาพที่ 9



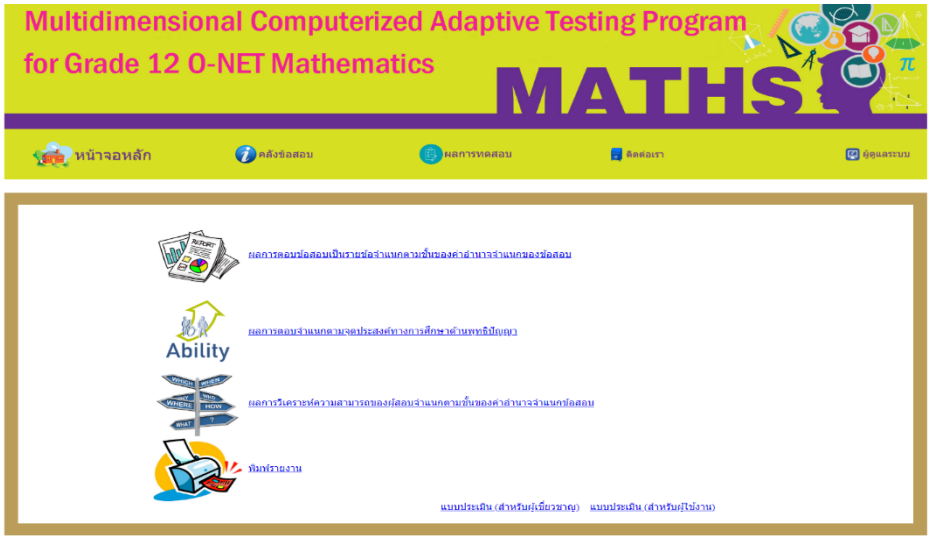
ภาพที่ 9 หน้าจอเกณฑ์การประเมินความสามารถของผู้สอบ

หน้าจอเข้าสู่ระบบค้นหาผลการทดสอบ เมื่อกดปุ่มนี้ “ ผลการทดสอบ” จะปรากฏลิงค์ไปยังหน้า เริ่มต้นเข้าสู่ระบบ ได้ ดังภาพที่ 10



ภาพที่ 10 หน้าจอเข้าสู่ระบบค้นหาผลการทดสอบ

จากภาพที่ 10 เมื่อกรอกเลขประจำตัวประชาชน กับ รหัสผ่าน แล้วกดปุ่มผลการทดสอบ จะไปยังหน้าจอรายงานผลการทดสอบ ดังภาพที่ 11



ภาพที่ 11 หน้าจอรายงานผลการทดสอบ

จากภาพที่ 11 หน้าจอรายงานผลการทดสอบ จะมีลิงค์ไปยังหน้าจอรายงานผลการทดสอบ 4 ส่วน ดังนี้

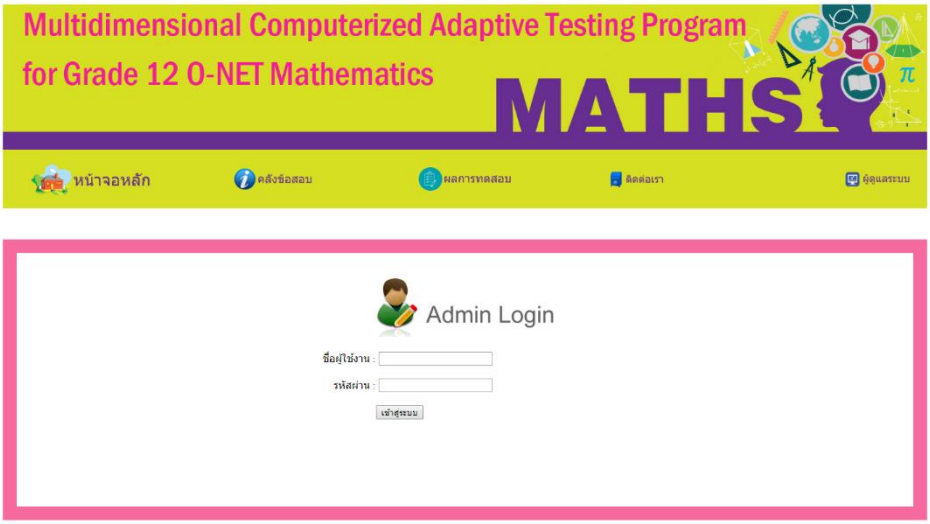
เมื่อกดปุ่มนี้จะแสดงรายงาน ผลการตอบข้อสอบเป็นรายข้อจำแนกตามชั้นของค่าอำนาจจำแนกของข้อสอบ

เมื่อกดปุ่มนี้จะแสดงรายงาน ผลการตอบจำแนกตามจุดประสงค์ทางการศึกษาด้านพุทธิปัญญา
เมื่อกดปุ่มนี้จะแสดงรายงาน ผลการวิเคราะห์ความสามารถของผู้สอบจำแนกตามชั้นของค่าอำนาจจำแนกข้อสอบ
เมื่อกดปุ่มนี้จะแสดงรายงานผล ในรูปแบบการพิมพ์ทางเครื่องพิมพ์

หน้าจอติดต่อเรา เมื่อกดปุ่มนี้ ” ติดต่อเรา ” จะปรากฏลิงค์ไปยังหน้า แสดงช่องทางการติดต่อผู้จัดทำ ดังภาพที่ 12

ภาพที่ 12 หน้าจอติดต่อเรา

หน้าจอเริ่มต้นผู้ดูแลระบบ เมื่อกดปุ่มนี้ “ ผู้ดูแลระบบ ” จะปรากฏลิงค์ไปยังหน้าเริ่มต้นเข้าสู่ระบบ สำหรับผู้ดูแลระบบดังภาพที่ 13



ภาพที่ 13 หน้าจอเริ่มต้นผู้ดูแลระบบ

จากภาพที่ 13 เมื่อกรอกชื่อผู้ใช้งาน กับรหัสผ่านถูกต้องแล้ว กดปุ่มเข้าสู่ระบบจะไปยังหน้า ค้นหาข้อสอบ

ส่วนที่ 2 การจัดการของผู้ดูแลระบบ

การจัดการข้อสอบเป็นส่วนที่ให้ครู อาจารย์ หรือผู้เกี่ยวของในการจัดการของผู้ดูแลระบบใส่ข้อสอบที่ผ่านการคัดเลือกเป็นไปตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ (Multidimensional Item Response Theory: MIRT) และคลังข้อสอบที่ใช้ในโปรแกรมนี้เป็นข้อสอบในการทดสอบ O-NET วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 6 ซึ่งได้รับความอนุเคราะห์ข้อสอบและผลการตอบจากสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) (สมศ.) โดยนำมาวิเคราะห์ข้อสอบตามแนวคิดของ Bloom et al. (1956); Anderson et al. (2001, p. XXI) ; Krathwohl (2002, p. 213) โดยการปรับลำดับขั้นและคำศัพท์ที่ใช้ในกระบวนการพุทธิปัญญา (The Cognitive Process Dimension) มี 6 กระบวนการเหมือนเดิม แต่ 3 กระบวนการแรกเปลี่ยนชื่อเป็น จำ (Remember) เข้าใจ (Understand) และประยุกต์ (Apply) ส่วน 3 กระบวนการหลังเปลี่ยนชื่อที่มีลักษณะเป็นคำนามไปเป็นคำกริยา และสลับกระบวนการที่ 5 กับ 6 และสร้างสรรค์ (Create) เปลี่ยนชื่อมาจากการสังเคราะห์ (Synthesis) (Anderson et al, 2001, p. XXI; Krathwohl, 2002, p. 213) และเปลี่ยนโครงสร้างจากมิติเดียวเป็นสองมิติ โดย (Anderson et al. (2001, p. XXI) และ Krathwohl (2002, p. 213) ได้เพิ่มโครงสร้างในมิติด้านความรู้ (Knowledge Dimension) เข้ามาในโครงสร้างของจุดประสงค์ทางการศึกษา ด้านพุทธิปัญญา ทำให้โครงสร้างใหม่มีลักษณะเป็นสองมิติประกอบด้วยมิติด้านกระบวนการพุทธิปัญญาและมิติด้านความรู้ และอธิบายความหมายของมิติด้านความรู้ที่เพิ่มเติมเข้ามาได้ (Krathwohl, 2002, p. 213) โดยมิติด้านความรู้ แบ่งเป็น 4 ส่วน คือ ความรู้เกี่ยวกับข้อเท็จจริง (Factual Knowledge) ความรู้เกี่ยวกับมโนทัศน์ (Conceptual Knowledge) ความรู้เกี่ยวกับวิธีดำเนินการ (Procedural Knowledge) และความรู้เกี่ยวกับอภิปัญญา (Metacognitive Knowledge) และมิติด้านกระบวนการพุทธิปัญญา แบ่งเป็น 6 กระบวนการ คือ จำ (Remember) เข้าใจ (Understand) ประยุกต์ (Apply) วิเคราะห์ (Analyze) ประเมินค่า (Evaluate) และสร้างสรรค์ (Create)การวิเคราะห์ค่าพารามิเตอร์ของข้อสอบตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ (MIRT) โดยใช้โปรแกรม Multilog ซึ่งมีเกณฑ์การคัดเลือกข้อสอบ (Urry, 1977) ดังนี้ 1) ค่าอำนาจจำแนกของข้อสอบ (a) มีค่าตั้งแต่ 0.50 ถึง 2.50 2) ค่าความยากของข้อสอบ (b) มีค่าตั้งแต่ -2.50 ถึง 2.50 และ 3) ค่าโอกาสในการเดาของผู้สอบ (c) มีค่าไม่เกิน 0.30 ซึ่งมีข้อสอบที่ผ่านเกณฑ์ 193 ข้อ ประกอบด้วยโจทย์ รายการคำตอบแบบ 4 ตัวเลือกและ 5 ตัวเลือก เริ่มจากแบ่งชั้นคลังข้อสอบเป็น 4 ชั้นตามระดับค่าอำนาจจำแนกของข้อสอบ ดังนี้

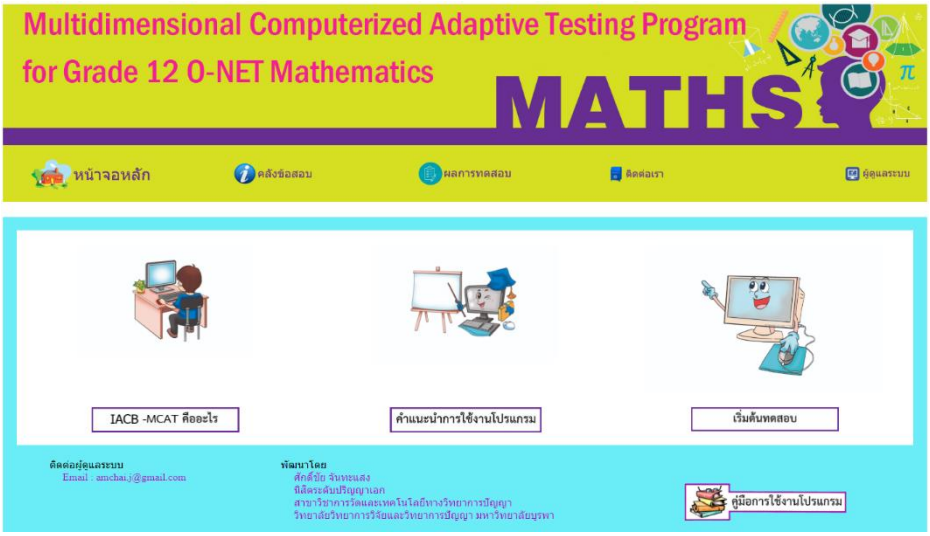
ชั้นที่ 1 0.50 – 0.99 มีข้อสอบ จำนวน 18 ข้อ ได้แก่ 1) สาระการเรียนรู้จำนวนและการดำเนินการ 3 ข้อ 2) สาระการเรียนรู้การวัด 1 ข้อ 3) สาระการเรียนรู้พีชคณิต 9 ข้อ และ 4) สาระการเรียนรู้การวิเคราะห์ข้อมูลและความน่าจะเป็น 5 ข้อ

ชั้นที่ 2 1.00 – 1.49 มีข้อสอบ จำนวน 49 ข้อ ได้แก่ 1) สาระการเรียนรู้จำนวนและการดำเนินการ 5 ข้อ 2) สาระการเรียนรู้การวัด 2 ข้อ 3) สาระการเรียนรู้พีชคณิต 22 ข้อ และ 4) สาระการเรียนรู้การวิเคราะห์ข้อมูลและความน่าจะเป็น 20 ข้อ

ชั้นที่ 3 1.50 – 1.99 มีข้อสอบ จำนวน 64 ข้อ ได้แก่ 1) สาระการเรียนรู้จำนวนและการดำเนินการ 16 ข้อ 2) สาระการเรียนรู้การวัด 4 ข้อ 3) สาระการเรียนรู้พีชคณิต 24 ข้อ และ 4) สาระการเรียนรู้การวิเคราะห์ข้อมูลและความน่าจะเป็น 20 ข้อ

ชั้นที่ 4 2.00 – 2.50 มี ข้อสอบ จำนวน 62 ข้อ ได้แก่ 1) สาระการเรียนรู้จำนวนและการดำเนินการ 11 ข้อ 2) สาระการเรียนรู้การวัด 8 ข้อ 3) พีชคสาระการเรียนรู้ณิต 33 ข้อ และ 4) สาระการเรียนรู้การวิเคราะห์ข้อมูลและความน่าจะเป็น 10 ข้อ

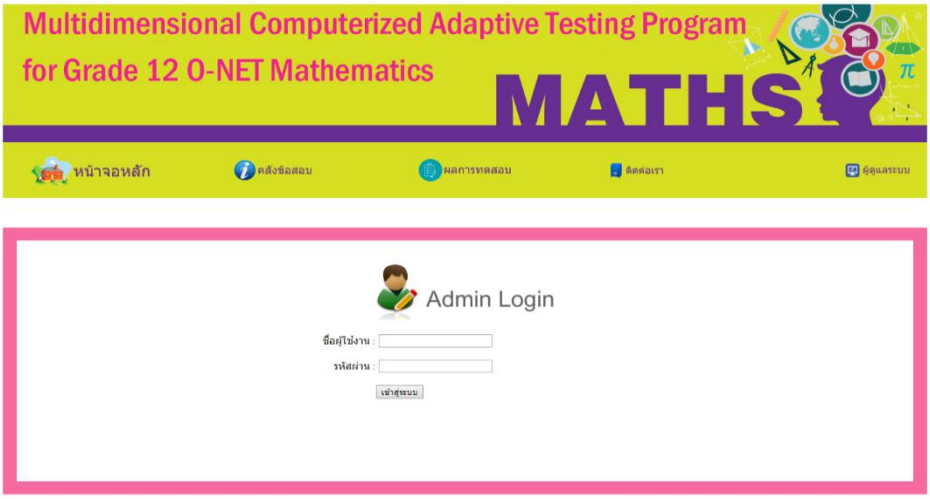
ในส่วนของการจัดการคลังข้อสอบ เพื่อทำการเพิ่มหรือทำการลบ และแก้ไขข้อสอบเดิมที่มีอยู่ในคลังข้อสอบ หลังจากผู้ใช้เลือก “เข้าสู่ระบบ” ในหน้าหลักโปรแกรมจะเข้าสู่หน้าจอหลักของโปรแกรมของส่วนการจัดการของผู้ดูแลระบบบ แสดงได้ดังภาพที่ 14



ภาพที่ 14 หน้าจอหลักของโปรแกรมการทดสอบ

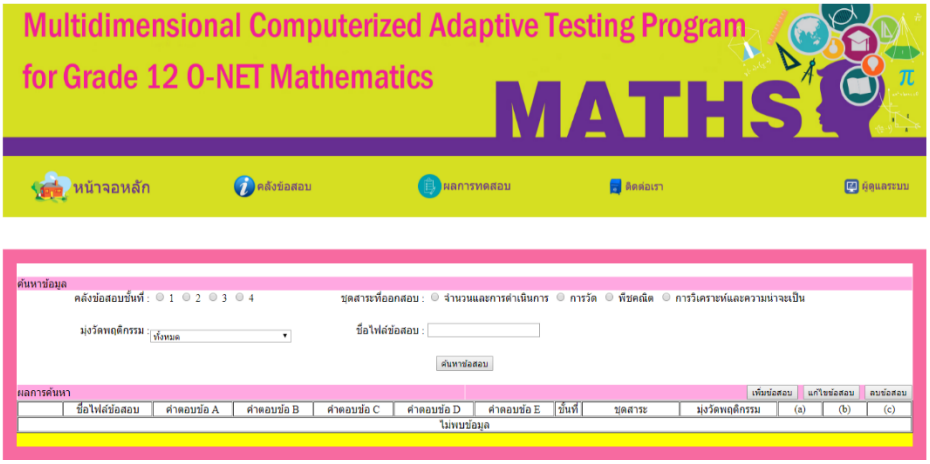
จากภาพที่ 14 หน้าจอหลักของโปแกรมการทดสอบ เมื่อกดปุ่มนี้ “

” จะปรากฏลิงค์ไปยังหน้า เริ่มต้นเข้าสู่ระบบสำหรับผู้ดูแลระบบ ดังภาพที่ 15



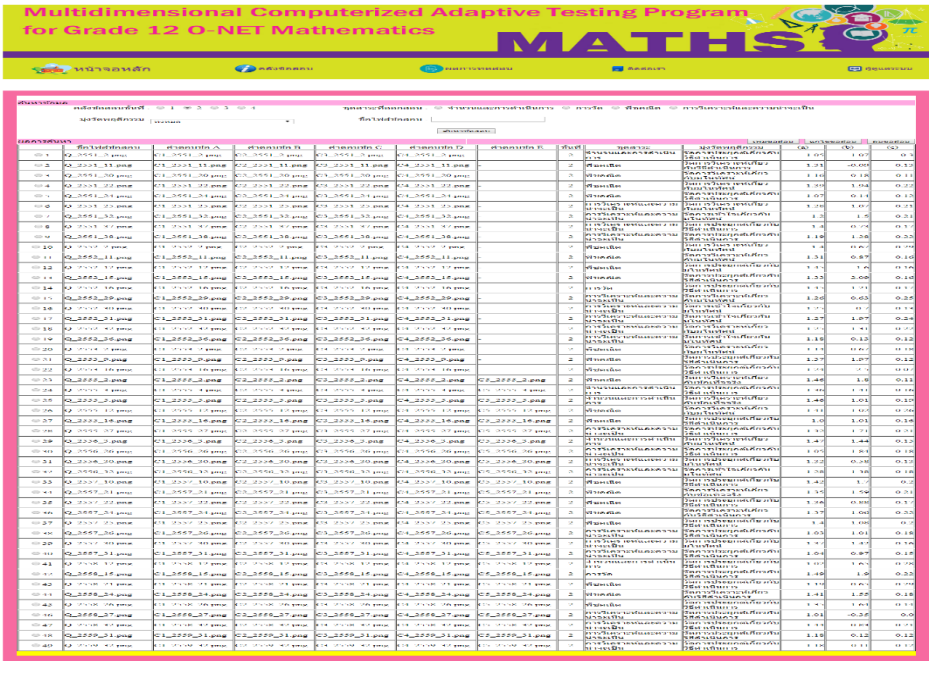
ภาพที่ 15 หน้าจอเริ่มต้นผู้ดูแลระบบ

จากภาพที่ 15 แสดงหน้าจอ Admin Login เข้าส่วนการจัดการข้อสอบ เมื่อกรอกชื่อผู้ใช้งาน กับ รหัสผ่านถูกต้องแล้ว กดปุ่ม “เข้าสู่ระบบ” จะไปยังหน้า ค้นหาข้อมูล ดังภาพที่ 16



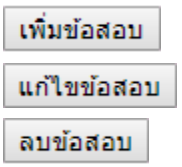
ภาพที่ 16 หน้าจอค้นหาข้อมูล

จากภาพที่ 16 หน้าจอค้นหาข้อมูล นี้จะประกอบด้วย 2 ส่วน คือ ส่วนค้นหาข้อมูล กับส่วนแสดงผลการค้นหา เมื่อใส่ข้อมูลการค้นหา แล้วกดปุ่ม”ค้นหาข้อสอบ” จะแสดงรายชื่อข้อสอบในระบบตามที่ตัวเลือกการค้นหา ในส่วนแสดงผลการค้นหา ดังภาพที่ 17



ภาพที่ 17 หน้าจอแสดงผลการค้นหา

จากภาพที่ 17 หน้าจอแสดงผลการค้นหา จะมีเมนูย่อยหลักๆ 3 ส่วน ประกอบด้วย 1) เพิ่มข้อสอบ 2) แก้ไขข้อสอบ และ 3) ลบข้อสอบ ดังนี้



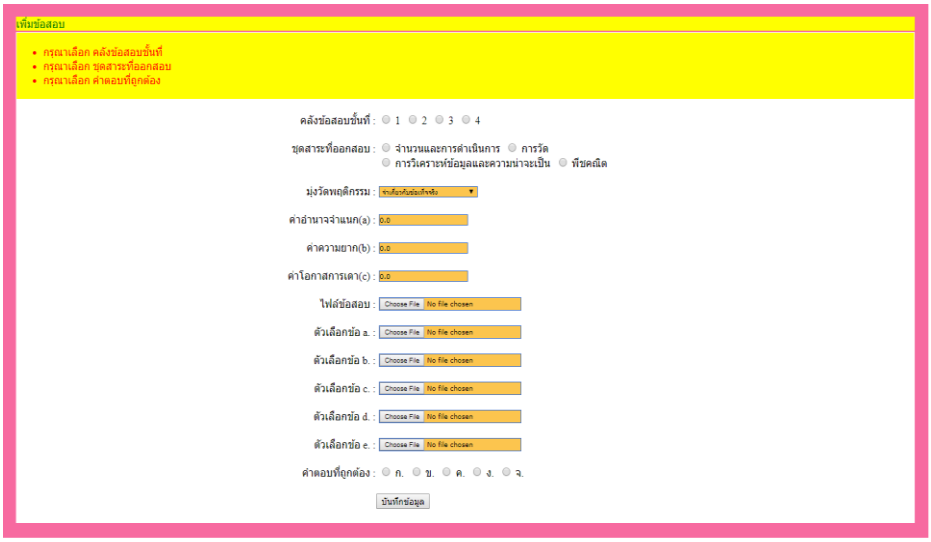
เมื่อกดปุ่มนี้จะแสดงหน้าจอ “เพิ่มข้อสอบ”

เมื่อกดปุ่มนี้จะแสดงหน้าจอ “แก้ไขข้อสอบ”

เมื่อกดปุ่มนี้แสดงหน้าจอ “การลบข้อสอบ”

การเพิ่มข้อสอบ

การเพิ่มข้อสอบลงในโปรแกรม โดยให้ผู้ใช้กดปุ่ม “เพิ่มข้อสอบ” แล้วโปรแกรมจะเข้าสู่หน้าจอ ดังภาพที่ 18



ภาพที่ 18 หน้าจอการเพิ่มข้อสอบ

จากภาพที่ 18 แสดงหน้าจอการเพิ่มข้อสอบลงในโปรแกรม โดยข้อสอบที่ใช้ในโปรแกรมนี้ต้องเป็นข้อสอบที่มีรายการคำตอบอย่างน้อย 4 ตัวเลือก และผ่านการวิเคราะห์ข้อสอบตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ จึงมีค่าพารามิเตอร์ของข้อสอบ ได้แก่ ค่าอำนาจจำแนกของข้อสอบ ค่าความยากของข้อสอบ และค่าโอกาสการเดาของข้อสอบ (ผู้สนใจสามารถคำนวณค่าพารามิเตอร์ของข้อสอบได้จาก โปรแกรม Multilog มีขั้นตอนการเพิ่มข้อสอบดังนี้

ขั้นที่ 1 เลือกข้อสอบที่ต้องการเพิ่มข้อสอบเข้าไป

ขั้นที่ 2 เลือกคลังข้อสอบชั้นที่ 1, 2, 3 และ 4 ตามค่าอำนาจจำแนกของข้อสอบ

ขั้นที่ 3 เลือกชุดสาระการเรียนรู้ที่ออกข้อสอบ ได้แก่ 1) จำนวนและการดำเนินการ 2) การวัด 3) พีชคณิต และ 4) การวิเคราะห์ข้อมูลและความน่าจะเป็น

ขั้นที่ 4 เลือกกระบวนการพุทธิปัญญาที่มุ่งวัด ได้แก่ 1) วิเคราะห์เกี่ยวกับข้อเท็จจริง 2) เข้าใจเกี่ยวกับมโนทัศน์ 3) ประยุกต์เกี่ยวกับมโนทัศน์ 4) วิเคราะห์เกี่ยวกับมโนทัศน์ 5) ประยุกต์เกี่ยวกับวิธีการดำเนินการ และ 6) วิเคราะห์เกี่ยวกับวิธีดำเนินการ

ขั้นที่ 5 กรอกรายละเอียดค่าอำนาจจำแนกของข้อสอบ (Discrimination Power Parameter: a_parameter) ต้องมีค่าระหว่าง -2.50 ถึง 2.50

ขั้นที่ 6 กรอกรายละเอียดค่าความยากของข้อสอบ (Difficulty Parameter: b_parameter) ต้องมีค่าระหว่าง 0.50 ถึง 2.50

ขั้นที่ 7 กรอกรายละเอียดค่าพารามิเตอร์ของข้อสอบ ค่าโอกาสการเดาของข้อสอบ (Guess Parameter: c_parameter) ต้องมีค่าไม่เกิน 0.30

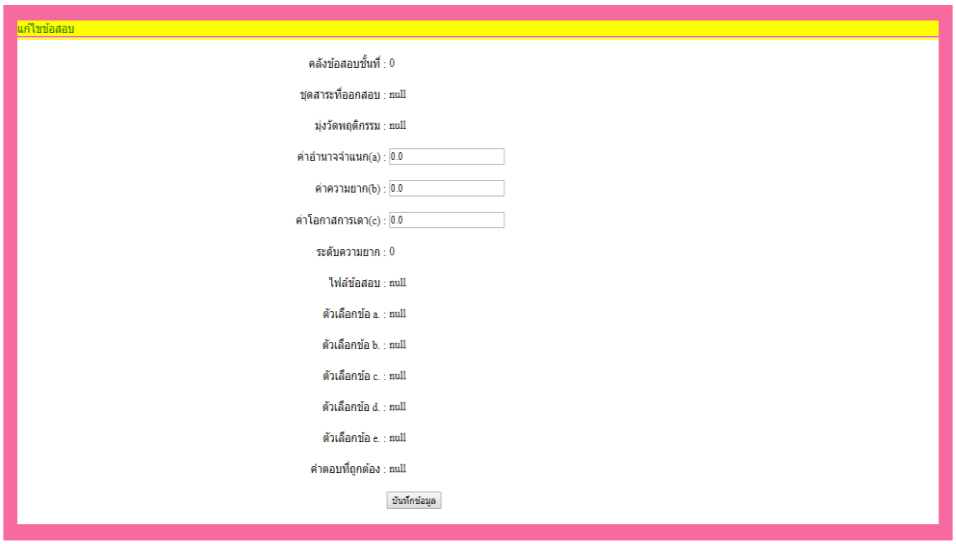
ขั้นที่ 8 เลือกไฟล์ข้อสอบที่บันทึกเป็นไฟล์รูปภาพ (.png)

ขั้นที่ 9 เลือกไฟล์ตัวเลือกข้อ a, b, c, d และ e ที่บันทึกเป็นไฟล์รูปภาพ (.png)

ขั้นที่ 10 เลือกคำตอบที่ถูกต้อง เสร็จกดปุ่ม “ บันทึกข้อมูล ”

การแก้ไขข้อสอบ

การแก้ไขข้อสอบเดิมในคลังข้อสอบ ดำเนินการโดยไปหน้าจอค้นหาข้อมูลข้อสอบ ดังภาพที่ 16 แล้วทำการกดปุ่ม “แก้ไข” ซึ่งปรากฏอยู่ด้านหน้าของข้อสอบที่ต้องการแก้ไข แล้วจะได้หน้าจอดังภาพที่ 19



ภาพที่ 19 หน้าจอแก้ไขข้อสอบ

จากภาพที่ 19 แสดงหน้าจอแก้ไขข้อสอบ เมื่อต้องการแก้ไขข้อสอบ มีขั้นตอนการเพิ่มข้อสอบดังนี้

ขั้นที่ 1 เลือกค้นหาข้อสอบที่ต้องการจะแก้ไข

ขั้นที่ 2 เลือกคลังข้อสอบชั้นที่ 1, 2, 3 และ 4 ตามค่าอำนาจจำแนกของข้อสอบ

ขั้นที่ 3 เลือกชุดสาระการเรียนรู้ที่ออกข้อสอบ ได้แก่ 1) จำนวนและการดำเนินการ 2) การวัด 3) พีชคณิต และ 4) การวิเคราะห์ข้อมูลและความน่าจะเป็น

ขั้นที่ 4 เลือกกระบวนการพุทธิปัญญาที่มุ่งวัด ได้แก่ 1) วิเคราะห์เกี่ยวกับข้อเท็จจริง 2) เข้าใจเกี่ยวกับมโนทัศน์ 3) ประยุกต์เกี่ยวกับมโนทัศน์ 4) วิเคราะห์เกี่ยวกับมโนทัศน์ 5) ประยุกต์เกี่ยวกับวิธีการดำเนินการ และ 6) วิเคราะห์เกี่ยวกับวิธีดำเนินการ

ขั้นที่ 5 กรอกรายละเอียดค่าอำนาจจำแนกของข้อสอบ (Discrimination Power Parameter: a_parameter) ต้องมีค่าระหว่าง -2.50 ถึง 2.50

ขั้นที่ 6 กรอกรายละเอียดค่าความยากของข้อสอบ (Difficulty Parameter: b_parameter) ต้องมีค่าระหว่าง 0.50 ถึง 2.50

ขั้นที่ 7 กรอกรายละเอียดค่าพารามิเตอร์ของข้อสอบ ค่าโอกาสการเดาของข้อสอบ (Guess Parameter: c_parameter) ต้องมีค่าไม่เกิน 0.30

ขั้นที่ 8 เลือกไฟล์ข้อสอบที่บันทึกเป็นไฟล์รูปภาพ (.png)

ขั้นที่ 9 เลือกไฟล์ตัวเลือกข้อ a, b, c, d และ e ที่บันทึกเป็นไฟล์รูปภาพ (.png)

ขั้นที่ 10 เลือกคำตอบที่ถูกต้อง เสร็จกดปุ่ม “ ”

การลบข้อสอบ

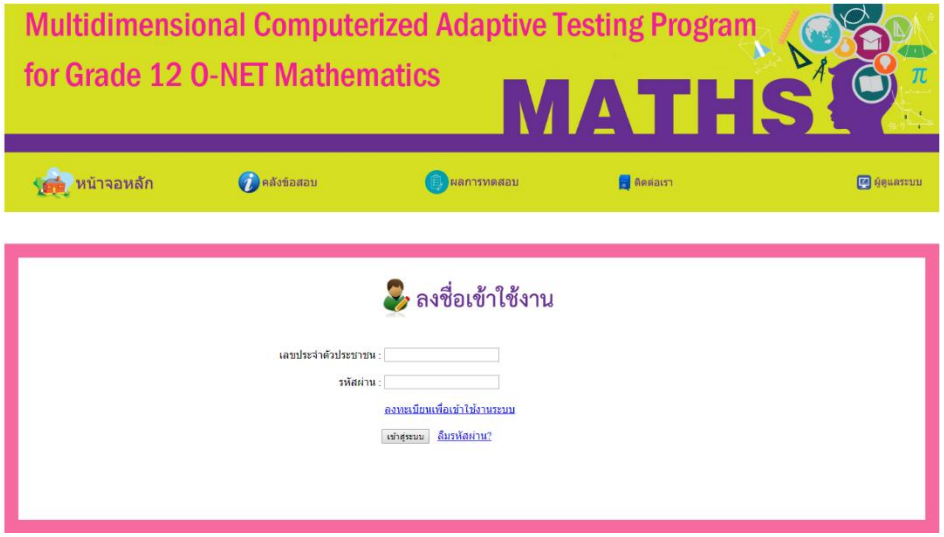
การลบข้อสอบออกจากคลังข้อสอบที่อยู่ในโปรแกรม ดำเนินการโดยไปหน้าจอค้นหาข้อมูลข้อสอบ ดังภาพที่ 16 จากนั้นให้เลือกชุดของข้อสอบที่ต้องการลบข้อสอบออกแล้วทำการกดปุ่ม “ลบข้อสอบ” ซึ่งปรากฏอยู่ด้านท้ายของข้อสอบที่ต้องการลบ

ส่วนที่ 3 การจัดการสอบ

การจัดการสอบนี้ เป็นส่วนที่ให้นักเรียน เข้าใช้โปรแกรมเพื่อทดสอบข้อสอบ O-NET วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 6 ซึ่งการทดสอบแบบรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ มีขั้นตอนการทดสอบดังต่อไปนี้

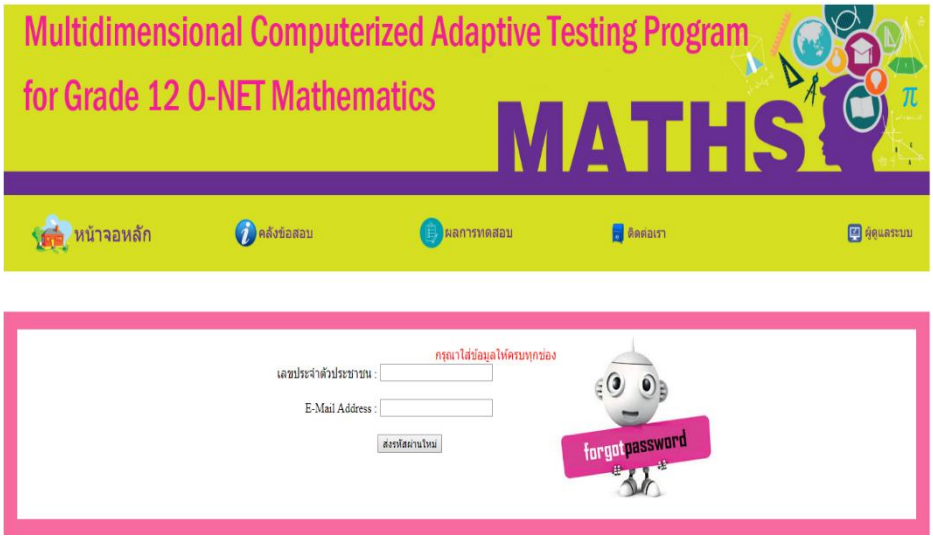
3.1 เริ่มต้นการทดสอบ

เมื่อผู้สอบต้องการเริ่มต้นการทดสอบ ให้กดปุ่มนี้ “ ” จะปรากฏลิงค์ไปยังหน้า เริ่มต้นเข้าสู่ระบบทดสอบ ดังภาพที่ 20



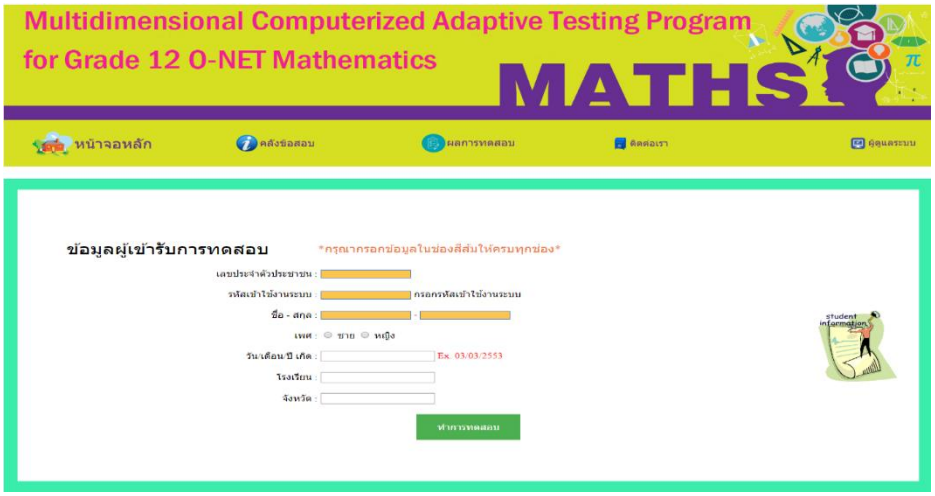
ภาพที่ 20 หน้าเริ่มต้นเข้าสู่ระบบทดสอบ

จากภาพที่ 20 สำหรับผู้เริ่มต้นใหม่ ให้กดข้อความ “ลงทะเบียนเพื่อเข้าใช้งานระบบ” ถ้าใครเคยลงทะเบียนแล้ว สามารถใส่ เลขประจำตัวประชาชนกับและรหัสผ่าน เพื่อเข้าสู่หน้าทำการทดสอบได้เลย และกรณีจำรหัสผ่านไม่ได้สามารถกดปุ่มลืมรหัส ให้กดข้อความ “ ลืมรหัสผ่าน? ” จะปรากฏลิงค์ไปยังหน้า หน้าจอลงทะเบียนเพื่อเข้าใช้งานในระบบ ดังภาพที่ 21 แต่สำหรับคนที่ยังไม่ลงทะเบียน ให้กดปุ่ม “ ลงทะเบียนเพื่อเข้าใช้งานระบบ ”จะปรากฏลิงค์ไปยังหน้า หน้าจอลงทะเบียนเพื่อเข้าใช้งานในระบบ ดังภาพที่ 22



ภาพที่ 21 หน้าจอลืมรหัสผ่าน

จากภาพที่ 21 กรอกข้อมูลให้ครบทุกช่อง เลขประจำตัวประชาชน และ E-Mail Address หลักจากนั้นให้กดปุ่ม “ ส่งรหัสผ่านใหม่ ” ระบบจะทำการส่งรหัสผ่านใหม่ไปที่ E-Mail Address ที่ผู้ใช้งานกรอกข้อมูลไว้

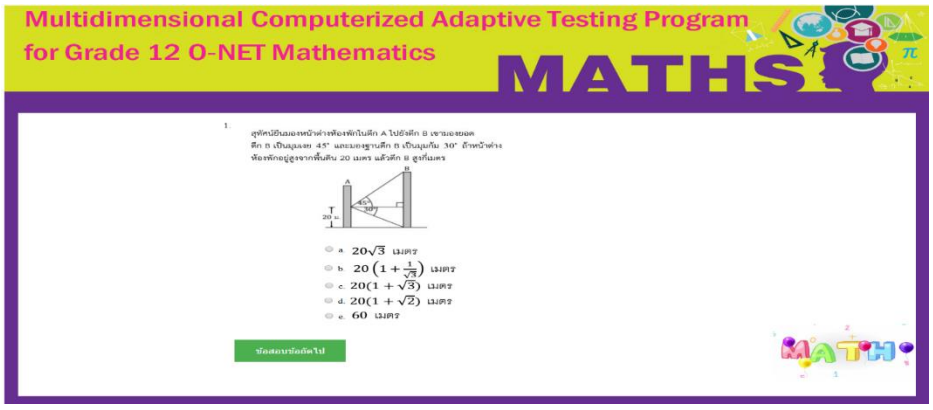


ภาพที่ 22 หน้าจอลงทะเบียนเพื่อเข้าใช้งานในระบบ

จากภาพที่ 22 กรอกข้อมูลให้ครบทุกช่อง เลขประจำตัวประชาชน รหัสเข้าใช้งานระบบ ชื่อ-สกุล เพศ วัน/เดือน/ปีเกิด โรงเรียน จังหวัด หลักจากนั้นให้กดปุ่ม “ทำการทดสอบ” เพื่อดำเนินการทดสอบต่อไป

3.2 การดำเนินการสอบ

หลักจากนั้นให้กดปุ่ม “ทำการทดสอบ” เพื่อดำเนินการทดสอบเรียบร้อยแล้ว จะเข้าสู่การดำเนินการทดสอบ โดยจะปรากฏหน้าจอดังภาพที่ 23

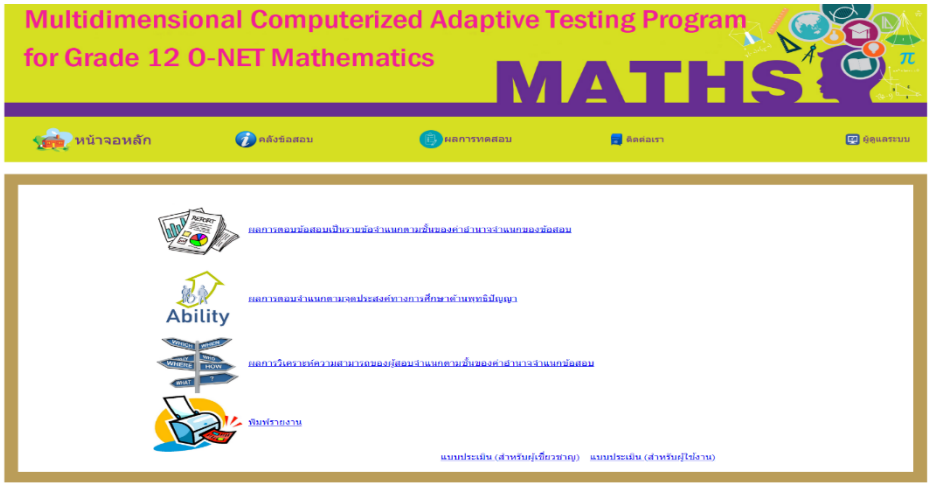


ภาพที่ 23 หน้าจอข้อสอบข้อแรก

จากภาพที่ 23 หน้าจอข้อสอบข้อแรก โดยโปรแกรมจะทำการสุ่มเลือกข้อสอบขึ้นมาให้ขึ้นอยู่กับความสามารถของผู้สอบโดยเรียงตามชั้นของค่าอำนาจจำแนกข้อสอบ หลังจากนั้นให้ผู้สอบเลือกคำตอบ แล้วกดปุ่ม “ข้อสอบข้อถัดไป” เพื่อรับข้อสอบข้อถัดไป ซึ่งการดำเนินการทดสอบจะดำเนินต่อไป จนสิ้นสุดการทดสอบ

3.3. รายงานผลารทดสอบ

หลังจากผู้สอบดำเนินการทดสอบเสร็จสิ้นแล้ว โปรแกรมจะให้กดปุ่ม “สรุปผลการทดสอบ” จะมีลิงค์ไปยังหน้าจอรายงานผลการทดสอบ ดังภาพที่ 24



ภาพที่ 24 หน้าจอรายงานผลการทดสอบ

จากภาพที่ 24 หน้าจอรายงานผลการทดสอบ จะมีลิงค์ไปยังหน้าจอแสดงผลการทดสอบ 4 ส่วน ดังนี้

ผลการตอบข้อสอบเป็นรายข้อจำแนกตามชั้นของค่าอำนาจจำแนกของข้อสอบ

เมื่อกดปุ่มนี้จะแสดงรายงาน ผลการตอบข้อสอบเป็นรายข้อจำแนกตามชั้นของค่าอำนาจจำแนกของข้อสอบ

ผลการตอบจำแนกตามจุดประสงค์ทางการศึกษาด้านพุทธิปัญญา

เมื่อกดปุ่มนี้จะแสดงรายงาน ผลการตอบจำแนกตามจุดประสงค์ทางการศึกษาด้านพุทธิปัญญา

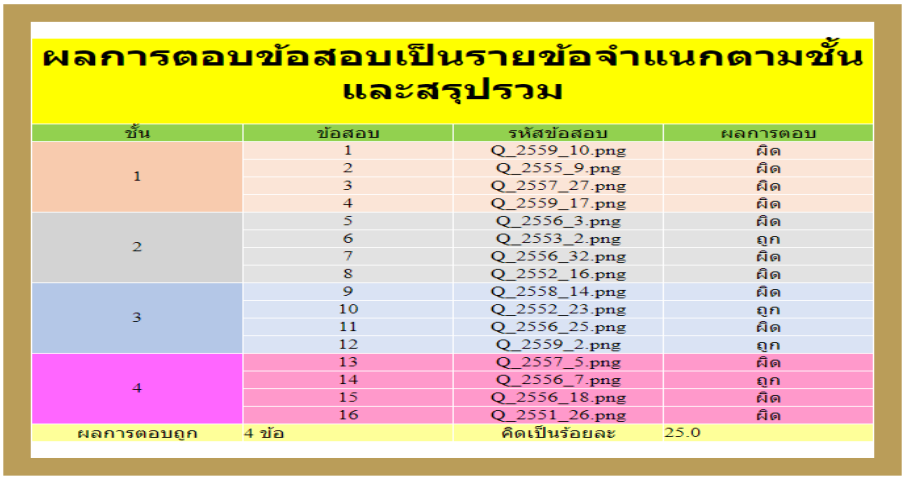
ผลการวิเคราะห์ความสามารถของผู้สอบจำแนกตามชั้นของค่าอำนาจจำแนกข้อสอบ

เมื่อกดปุ่มนี้จะแสดงรายงาน ผลการวิเคราะห์ความสามารถของผู้สอบจำแนกตามชั้นของค่าอำนาจจำแนกข้อสอบ

พิมพ์รายงาน

เมื่อกดปุ่มนี้จะแสดงรายงาน ผลในรูปแบบการพิมพ์ทางเครื่องพิมพ์

3.3.1) หน้าจอรายงานผลการตอบข้อสอบเป็นรายข้อจำแนกตามชั้นของค่าอำนาจจำแนกของข้อสอบ เมื่อกดปุ่มนี้ “ ผลการตอบข้อสอบเป็นรายข้อจำแนกตามชั้นของค่าอำนาจจำแนกของข้อสอบ ” จะปรากฏลิงค์ไปยังหน้า หน้าจอรายงานผลการตอบข้อสอบเป็นรายข้อจำแนกตามชั้นของค่าอำนาจจำแนกของข้อสอบ ดังภาพที่ 25



ผลการตอบข้อสอบเป็นรายข้อจำแนกตามชั้นและสรุปรวม

ชั้น	ข้อสอบ	รหัสข้อสอบ	ผลการตอบ
1	1	Q_2559_10.png	ผิด
	2	Q_2555_9.png	ผิด
	3	Q_2557_27.png	ผิด
	4	Q_2559_17.png	ผิด
2	5	Q_2556_3.png	ผิด
	6	Q_2553_2.png	ถูก
	7	Q_2556_32.png	ผิด
	8	Q_2552_16.png	ผิด
3	9	Q_2558_14.png	ผิด
	10	Q_2552_23.png	ถูก
	11	Q_2556_25.png	ผิด
	12	Q_2559_2.png	ถูก
4	13	Q_2557_5.png	ผิด
	14	Q_2556_7.png	ถูก
	15	Q_2556_18.png	ผิด
	16	Q_2551_26.png	ผิด
ผลการตอบถูก	4 ข้อ	คิดเป็นร้อยละ	25.0

ภาพที่ 25 หน้าจอรายงานผลการตอบข้อสอบเป็นรายข้อจำแนกตามชั้นของค่าอำนาจจำแนกของข้อสอบ

จากภาพที่ 25 แสดงหน้าจอรายงานผลการตอบข้อสอบเป็นรายข้อจำแนกตามชั้นของค่าอำนาจจำแนกของข้อสอบ โดยแสดงรายละเอียดเกี่ยวกับจำนวนข้อสอบในแต่ละชั้น รหัสข้อสอบ ผลการตอบ จำนวนข้อสอบที่ตอบถูกและร้อยละของผลการตอบถูก

3.3.2) หน้าจอรายงานผลการตอบจำแนกตามจุดประสงค์ทางการศึกษาด้านพุทธิปัญญา เมื่อกดปุ่มนี้ “ Ability ผลการตอบจำแนกตามจุดประสงค์ทางการศึกษาด้านพุทธิปัญญา ” จะปรากฏลิงค์ไปยังหน้าจอรายงานผลการตอบจำแนกตามจุดประสงค์ทางการศึกษาด้านพุทธิปัญญา ดังภาพที่ 26

ผลการวิเคราะห์ชั้นของคลังข้อสอบกับกระบวนการพุทธิปัญญาภาพรวม

ชั้น	ข้อสอบ	กระบวนการพุทธิปัญญา						ผลการตอบ
		การวิเคราะห์เกี่ยวกับข้อเท็จจริง	การเข้าใจเกี่ยวกับมโนทัศน์	การประยุกต์เกี่ยวกับมโนทัศน์	การวิเคราะห์เกี่ยวกับมโนทัศน์	การประยุกต์เกี่ยวกับวิธีดำเนินการ	การวิเคราะห์เกี่ยวกับวิธีดำเนินการ	
1	1					✓		ผิด
	2					✓		ผิด
	3					✓		ผิด
	4				✓			ผิด
2	5				✓			ผิด
	6				✓			ถูก
	7		✓					ผิด
	8					✓		ผิด
3	9					✓		ผิด
	10					✓		ถูก
	11				✓			ผิด
	12					✓		ถูก
4	13	✓						ผิด
	14					✓		ถูก
	15					✓		ผิด
	16					✓		ผิด
รวม		1	1	0	4	10	0	

ภาพที่ 26 หน้าจอรายงานผลการตอบจำแนกตามจุดประสงค์ทางการศึกษาด้านพุทธิปัญญา

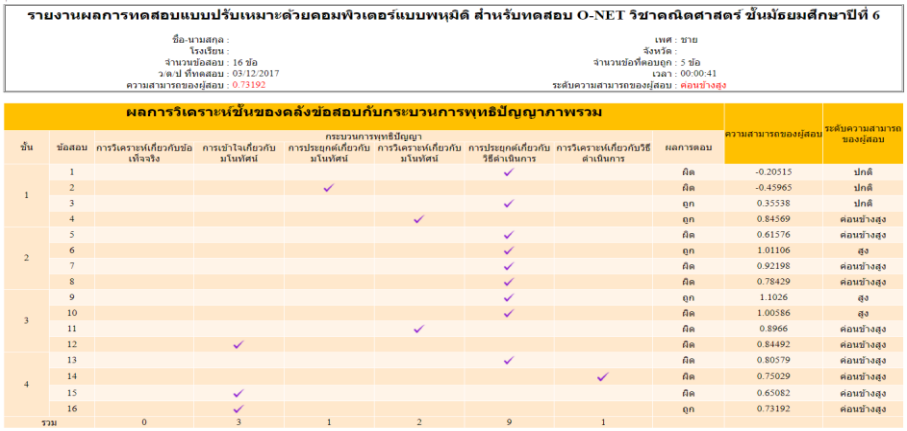
จากภาพที่ 26 แสดงหน้าจอรายงานผลการตอบจำแนกตามจุดประสงค์ทางการศึกษาด้านพุทธิปัญญา โดยแสดงรายละเอียดเกี่ยวกับจำนวนข้อสอบในแต่ละชั้น กระบวนการพุทธิปัญญา และผลการตอบ

3.3.3) หน้าจอรายงานผลการวิเคราะห์ความสามารถของผู้สอบจำแนกตามชั้นของค่าอำนาจจำแนกข้อสอบ เมื่อกดปุ่มนี้ “ ” จะปรากฏลิงค์ไปยังหน้าหน้าจอรายงานผลการวิเคราะห์ความสามารถของผู้สอบจำแนกตามชั้นของค่าอำนาจจำแนกข้อสอบ ดังภาพที่ 27

ภาพที่ 27 หน้าจอรายงานผลการวิเคราะห์ความสามารถของผู้สอบจำแนกตามชั้นของค่าอำนาจจำแนกข้อสอบ

จากภาพที่ 27 แสดงหน้าจอรายงานผลการวิเคราะห์ความสามารถของผู้สอบจำแนกตามชั้นของค่าอำนาจจำแนกข้อสอบ โดยแสดงรายละเอียดเกี่ยวกับชั้นของข้อสอบ จำนวนข้อทั้งหมด ข้อสอบที่ตอบถูก ร้อยละของการตอบถูกในแต่ละชั้น ค่าความสามารถของผู้สอบ และระดับความสามารถของผู้สอบ

3.3.4) หน้าจอพิมพ์รายงาน เมื่อกดปุ่มนี้ “ ” จะปรากฏลิงค์ไปยัง หน้าจอพิมพ์รายงาน ดังภาพที่ 28



รายงานผลการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ สำหรับทดสอบ O-NET วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 6

ชื่อ-นามสกุล :
โรงเรียน :
จำนวนข้อสอบ : 16 ข้อ
ว/ด/ป ที่ทดสอบ : 03/12/2017
ความสามารถของผู้สอบ : 0.73192

เพศ : ชาย
จังหวัด :
จำนวนข้อที่ตอบถูก : 5 ข้อ
เวลา : 00:00:41
ระดับความสามารถของผู้สอบ : ค่อนข้างสูง

ผลการวิเคราะห์ขั้นของคลังข้อสอบกับกระบวนการพุทธิปัญญาภาพรวม

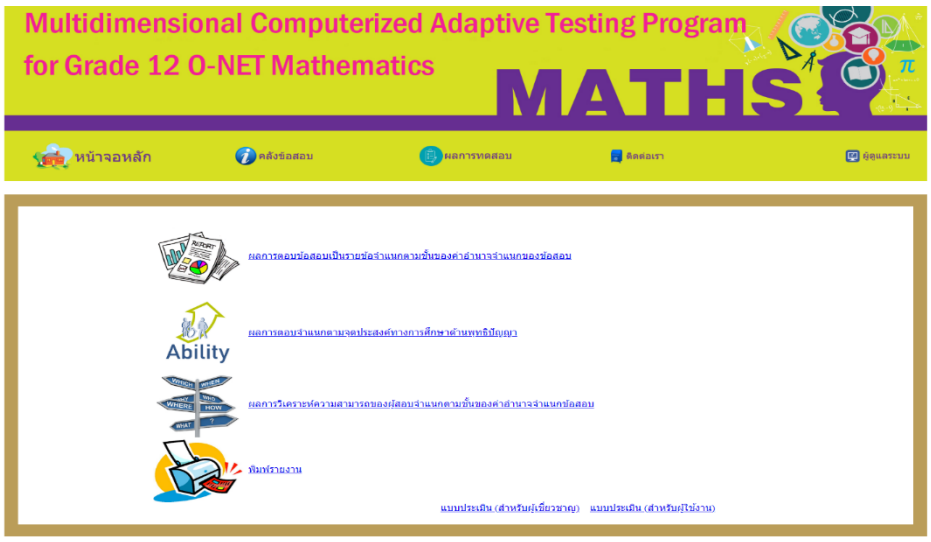
ขั้น	ข้อสอบ	การวิเคราะห์เกี่ยวกับข้อเท็จจริง	การเข้าใจเกี่ยวกับมโนทัศน์	กระบวนการพุทธิปัญญา: การประยุกต์เกี่ยวกับมโนทัศน์	การวิเคราะห์เกี่ยวกับมโนทัศน์	การประยุกต์เกี่ยวกับวิธีดำเนินการ	การวิเคราะห์เกี่ยวกับวิธีดำเนินการ	ผลการตอบ	ความสามารถของผู้สอบ	ระดับความสามารถของผู้สอบ
1	1					✓		ผิด	-0.20515	ปกติ
	2			✓				ผิด	-0.45965	ปกติ
	3					✓		ถูก	0.35538	ปกติ
	4				✓			ถูก	0.84569	ค่อนข้างสูง
2	5					✓		ผิด	0.61576	ค่อนข้างสูง
	6					✓		ถูก	1.01106	สูง
	7					✓		ผิด	0.92198	ค่อนข้างสูง
	8					✓		ผิด	0.78429	ค่อนข้างสูง
3	9					✓		ถูก	1.1026	สูง
	10					✓		ผิด	1.00586	สูง
	11				✓			ผิด	0.8966	ค่อนข้างสูง
	12		✓					ผิด	0.84492	ค่อนข้างสูง
4	13					✓		ผิด	0.80579	ค่อนข้างสูง
	14						✓	ผิด	0.75029	ค่อนข้างสูง
	15		✓					ผิด	0.65082	ค่อนข้างสูง
	16		✓					ถูก	0.73192	ค่อนข้างสูง
รวม		0	3	1	2	9	1			

ภาพที่ 28 หน้าจอพิมพ์รายงาน

จากภาพที่ 28 แสดงหน้าจอพิมพ์รายงาน หน้าจอพิมพ์รายงานการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ โดยแสดงรายละเอียดผู้สอบ ได้แก่ ชื่อ-สกุล เพศ โรงเรียน จังหวัด จำนวนข้อสอบทั้งหมด จำนวนข้อสอบที่ตอบถูก วัน/เดือน/ปี เวลาที่ใช้ในการทดสอบ ค่าประมาณความสามารถของผู้สอบ (Ability Estimate) และระดับความสามารถของผู้สอบ (Ability Level) ทั้งนี้มีเกณฑ์การประเมินระดับความสามารถของผู้สอบมีรายละเอียดดังภาพที่ 9

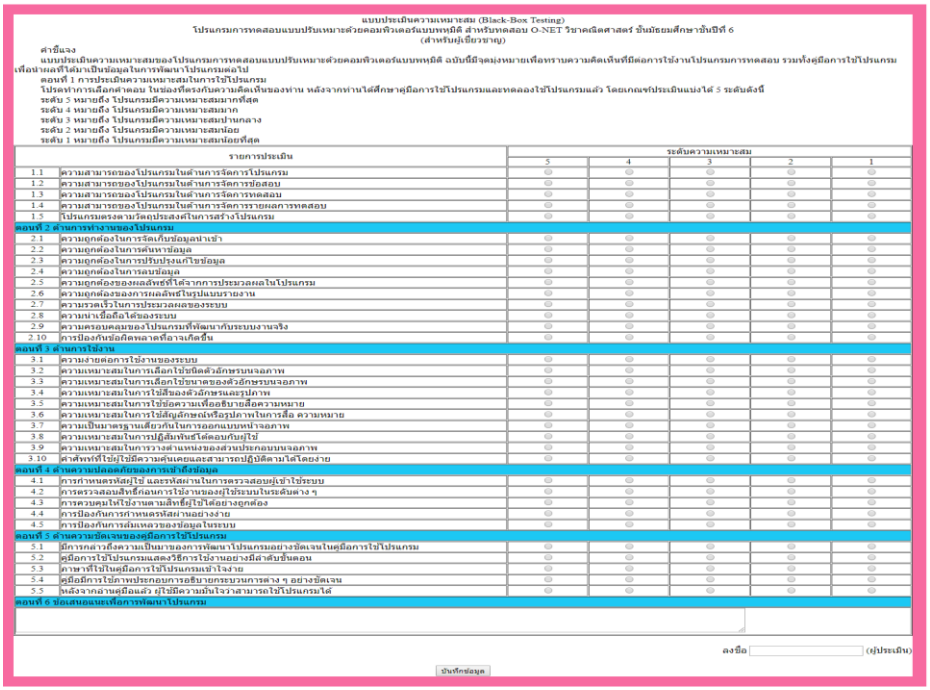
3.4 แบบประเมินความเหมาะสมของโปรแกรม

แบบประเมินความเหมาะสมของโปรแกรมจะอยู่ด้านล่างของหน้าจอรายงานผลการทดสอบ ประกอบด้วย 1) แบบประเมิน (สำหรับผู้เชี่ยวชาญ) และ 2) แบบประเมิน (สำหรับผู้ใช้งาน) ดังภาพที่ 29



ภาพที่ 29 แบบประเมินความเหมาะสมของโปรแกรม

แบบประเมินความเหมาะสมของโปรแกรม (สำหรับผู้เชี่ยวชาญ) เมื่อกดปุ่มนี้ “ แบบประเมิน (สำหรับผู้เชี่ยวชาญ) ” จะปรากฏลิงค์ไปยัง แบบประเมินความเหมาะสมของโปรแกรม (สำหรับผู้เชี่ยวชาญ) ดังภาพที่ 30



แบบประเมินความเหมาะสม (Black-Box Testing)
โปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบทฤษฎี สำหรับทดสอบ O-NET วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาชั้นปีที่ 6
(สำหรับผู้เชี่ยวชาญ)

คำชี้แจง
แบบประเมินความเหมาะสมของโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบทฤษฎี ฉบับนี้มีจุดมุ่งหมายเพื่อทราบความคิดเห็นที่มีต่อการใช้งานโปรแกรมการทดสอบ รวมทั้งคู่มือการใช้โปรแกรม เพื่อนำผลที่ได้มาเป็นข้อมูลในการพัฒนาโปรแกรมต่อไป
ตอนที่ 1 การประเมินความเหมาะสมในการใช้โปรแกรม
โปรดทำการเลือกคำตอบ ให้ตรงกับความคิดเห็นของท่าน หลังจากท่านได้ศึกษาคู่มือการใช้โปรแกรมและทดลองใช้โปรแกรมแล้ว โดยเกณฑ์ประเมินแบ่งได้ 5 ระดับดังนี้
ระดับ 5 หมายถึง โปรแกรมมีความเหมาะสมมากที่สุด
ระดับ 4 หมายถึง โปรแกรมมีความเหมาะสมมาก
ระดับ 3 หมายถึง โปรแกรมมีความเหมาะสมปานกลาง
ระดับ 2 หมายถึง โปรแกรมมีความเหมาะสมน้อย
ระดับ 1 หมายถึง โปรแกรมมีความเหมาะสมน้อยที่สุด

รายการประเมิน		ระดับความเหมาะสม				
		5	4	3	2	1
1.1	ความสามารถของโปรแกรมในด้านการจัดการโปรแกรม					
1.2	ความสามารถของโปรแกรมในด้านการจัดการข้อสอบ					
1.3	ความสามารถของโปรแกรมในด้านการจัดการทดสอบ					
1.4	ความสามารถของโปรแกรมในด้านการจัดการรายผลการทดสอบ					
1.5	โปรแกรมตรงตามวัตถุประสงค์ในการสร้างโปรแกรม					
ตอนที่ 2 ด้านการทำงานของโปรแกรม						
2.1	ความถูกต้องในการจัดเก็บข้อมูลนำเข้า					
2.2	ความถูกต้องในการค้นหาข้อมูล					
2.3	ความถูกต้องในการปรับปรุงแก้ไขข้อมูล					
2.4	ความถูกต้องในการลบข้อมูล					
2.5	ความถูกต้องของผลลัพธ์ที่ได้จากการประมวลผลในโปรแกรม					
2.6	ความถูกต้องของการแสดงผลรายงานในรูปแบบรายงาน					
2.7	ความรวดเร็วในการประมวลผลของระบบ					
2.8	ความน่าเชื่อถือได้ของระบบ					
2.9	ความครบคลุมของโปรแกรมที่พัฒนากับระบบงานจริง					
2.10	การป้องกันข้อผิดพลาดที่อาจเกิดขึ้น					
ตอนที่ 3 ด้านการใช้งาน						
3.1	ความง่ายต่อการใช้งานของระบบ					
3.2	ความเหมาะสมในการเลือกใช้ชนิดตัวอักษรบนจอภาพ					
3.3	ความเหมาะสมในการเลือกใช้ขนาดของตัวอักษรบนจอภาพ					
3.4	ความเหมาะสมในการใช้สีของตัวอักษรและรูปภาพ					
3.5	ความเหมาะสมในการใช้ข้อความเพื่ออธิบายสื่อความหมาย					
3.6	ความเหมาะสมในการใช้สัญลักษณ์หรือรูปภาพในการสื่อ ความหมาย					
3.7	ความเป็นมาตรฐานเดียวกันในการออกแบบหน้าจอภาพ					
3.8	ความเหมาะสมในการปฏิสัมพันธ์โต้ตอบกับผู้ใช้					
3.9	ความเหมาะสมในการวางตำแหน่งของส่วนประกอบบนจอภาพ					
3.10	คำศัพท์ที่ใช้ผู้ใช้มีความคุ้นเคยและสามารถปฏิบัติตามได้โดยง่าย					
ตอนที่ 4 ด้านความปลอดภัยของการเข้าถึงข้อมูล						
4.1	การกำหนดรหัสผู้ใช้ และรหัสผ่านในการตรวจสอบผู้เข้าใช้ระบบ					
4.2	การตรวจสอบสิทธิ์ก่อนการใช้งานของผู้ใช้ระบบในระดับต่าง ๆ					
4.3	การควบคุมให้ใช้งานตามสิทธิ์ซึ่งผู้ใช้ได้อย่างถูกต้อง					
4.4	การป้องกันการกำหนดรหัสผ่านอย่างง่าย					
4.5	การป้องกันการดัดแปลงของข้อมูลในระบบ					
ตอนที่ 5 ด้านความชัดเจนของคู่มือการใช้โปรแกรม						
5.1	มีการกล่าวถึงความเป็นมาของการพัฒนาโปรแกรมอย่างชัดเจนในคู่มือการใช้โปรแกรม					
5.2	คู่มือการใช้โปรแกรมแสดงวิธีการใช้งานอย่างมีลำดับขั้นตอน					
5.3	ภาษาที่ใช้ในคู่มือการใช้โปรแกรมเข้าใจง่าย					
5.4	คู่มือมีการใช้ภาพประกอบการอธิบายกระบวนการต่าง ๆ อย่างชัดเจน					
5.5	หลังจากอ่านคู่มือแล้ว ผู้ใช้มีความมั่นใจว่าสามารถใช้โปรแกรมได้					
ตอนที่ 6 ข้อเสนอแนะเพื่อการพัฒนาโปรแกรม						

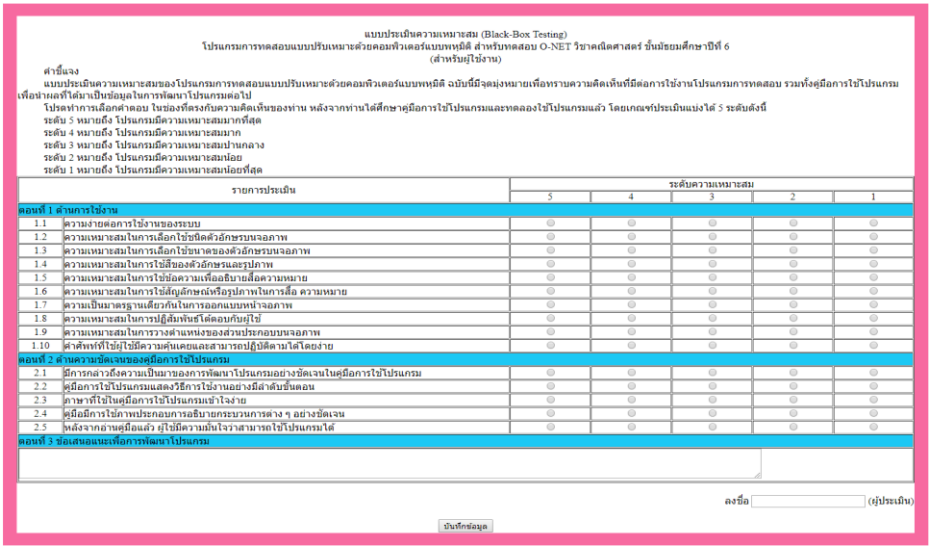
ลงชื่อ _____ (ผู้ประเมิน)

บันทึกข้อมูล

ภาพที่ 30 แบบประเมินความเหมาะสมของโปรแกรม (สำหรับผู้เชี่ยวชาญ)

จากภาพที่ 30 แสดงแบบประเมินความเหมาะสมของโปรแกรมทั้ง 6 ด้าน ได้แก่ 1) ด้านความต้องการของผู้ใช้โปรแกรม 2) ด้านความต้องการของผู้ใช้โปรแกรม 3) ด้านการใช้งาน 4) ด้านการรักษาความปลอดภัยของการเข้าถึงข้อมูล 5) ด้านความชัดเจนของคู่มือการใช้งาน และ 6) ข้อเสนอแนะเพื่อการพัฒนาโปรแกรม โดยใช้แบบประเมินด้วยวิธี Black-Box Testing ลักษณะมาตรประมาณค่า (Rating Scale) 5ระดับ

แบบประเมินความเหมาะสมของโปรแกรม (สำหรับผู้ใช้งาน) เมื่อกดปุ่มนี้ “ แบบประเมิน (สำหรับนักเรียน) ” จะปรากฏลิงค์ไปยัง แบบประเมินความเหมาะสมของโปรแกรม (สำหรับผู้ใช้งาน) ดังภาพที่ 31



แบบประเมินความเหมาะสม (Black-Box Testing)
โปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ สำหรับทดสอบ O-NET วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 6
(สำหรับผู้ใช้งาน)

คำชี้แจง
แบบประเมินความเหมาะสมของโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ ฉบับนี้มีจุดมุ่งหมายเพื่อทราบความคิดเห็นที่มีต่อการใช้งานโปรแกรมการทดสอบ รวมทั้งคู่มือการใช้โปรแกรม เพื่อนำผลที่ได้มาเป็นข้อมูลในการพัฒนาโปรแกรมต่อไป
โปรดทำการเลือกคำตอบ ในช่องที่ตรงกับความคิดเห็นของท่าน หลังจากท่านได้ศึกษาคู่มือการใช้โปรแกรมและทดลองใช้โปรแกรมแล้ว โดยเกณฑ์ประเมินแบ่งได้ 5 ระดับดังนี้
ระดับ 5 หมายถึง โปรแกรมมีความเหมาะสมมากที่สุด
ระดับ 4 หมายถึง โปรแกรมมีความเหมาะสมมาก
ระดับ 3 หมายถึง โปรแกรมมีความเหมาะสมปานกลาง
ระดับ 2 หมายถึง โปรแกรมมีความเหมาะสมน้อย
ระดับ 1 หมายถึง โปรแกรมมีความเหมาะสมน้อยที่สุด

รายการประเมิน		ระดับความเหมาะสม				
		5	4	3	2	1
ตอนที่ 1 ด้านการใช้งาน						
1.1	ความง่ายต่อการใช้งานของระบบ					
1.2	ความเหมาะสมในการเลือกใช้ชนิดตัวอักษรบนจอภาพ					
1.3	ความเหมาะสมในการเลือกใช้ขนาดของตัวอักษรบนจอภาพ					
1.4	ความเหมาะสมในการใช้สีของตัวอักษรและรูปภาพ					
1.5	ความเหมาะสมในการใช้ข้อความเพื่ออธิบายสื่อความหมาย					
1.6	ความเหมาะสมในการใช้สัญลักษณ์หรือรูปภาพในการสื่อ ความหมาย					
1.7	ความเป็นมาตรฐานเดียวกันในการออกแบบหน้าจอภาพ					
1.8	ความเหมาะสมในการปฏิสัมพันธ์โต้ตอบกับผู้ใช้					
1.9	ความเหมาะสมในการวางตำแหน่งของส่วนประกอบบนจอภาพ					
1.10	คำศัพท์ที่ใช้ผู้ใช้มีความคุ้นเคยและสามารถปฏิบัติตามได้โดยง่าย					
ตอนที่ 2 ด้านความชัดเจนของคู่มือการใช้โปรแกรม						
2.1	มีการกล่าวถึงความเป็นมาของการพัฒนาโปรแกรมอย่างชัดเจนในคู่มือการใช้โปรแกรม					
2.2	คู่มือการใช้โปรแกรมแสดงวิธีการใช้งานอย่างมีลำดับขั้นตอน					
2.3	ภาษาที่ใช้ในคู่มือการใช้โปรแกรมเข้าใจง่าย					
2.4	คู่มือมีการใช้ภาพประกอบการอธิบายกระบวนการต่าง ๆ อย่างชัดเจน					
2.5	หลังจากอ่านคู่มือแล้ว ผู้ใช้มีความมั่นใจว่าสามารถใช้โปรแกรมได้					
ตอนที่ 3 ข้อเสนอแนะเพื่อการพัฒนาโปรแกรม						

ลงชื่อ (ผู้ประเมิน)

บันทึกข้อมูล

ภาพที่ 31 แบบประเมินความเหมาะสมของโปรแกรม (สำหรับผู้ใช้งาน)

จากภาพที่ 31 แสดงแบบประเมินความเหมาะสมของโปรแกรมทั้ง 3 ด้าน ได้แก่ 1) ด้านการใช้งาน 2) ด้านความชัดเจนของคู่มือการใช้งาน และ 3) ข้อเสนอแนะเพื่อการพัฒนาโปรแกรม โดยใช้แบบประเมินด้วยวิธี Black-Box Testing ลักษณะมาตรประมาณค่า (Rating Scale) 5ระดับ

ภาคผนวก ฉ
แบบประเมินความเหมาะสมของโปรแกรม
การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ สำหรับผู้เชี่ยวชาญ

แบบประเมินความเหมาะสม (Black-Box Testing)
โปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ
สำหรับทดสอบ O-NET วิชาคณิตศาสตร์ มัธยมศึกษาชั้นปีที่ 6
(สำหรับผู้เชี่ยวชาญ)

คำชี้แจง

แบบประเมินความเหมาะสมของโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ ฉบับนี้มีจุดมุ่งหมายเพื่อทราบความคิดเห็นที่มีต้อการใช้งานโปรแกรมการทดสอบ รวมทั้งคู่มือการใช้โปรแกรม เพื่อนำผลที่ได้มาเป็นข้อมูลในการพัฒนาโปรแกรมต่อไป

ตอนที่ 1 การประเมินความเหมาะสมในการใช้โปรแกรม

โปรดทำกดปุ่ม ⦿ ในช่องที่ตรงกับความคิดเห็นของท่านหลังจากท่านได้ศึกษาคู่มือการใช้โปรแกรมและทดลองใช้โปรแกรมแล้ว โดยเกณฑ์ประเมินแบ่งได้ 5 ระดับดังนี้

ระดับ 5 หมายถึง โปรแกรมมีความเหมาะสมมากที่สุด
ระดับ 4 หมายถึง โปรแกรมมีความเหมาะสมมาก
ระดับ 3 หมายถึง โปรแกรมมีความเหมาะสมปานกลาง
ระดับ 2 หมายถึง โปรแกรมมีความเหมาะสมน้อย
ระดับ 1 หมายถึง โปรแกรมมีความเหมาะสมน้อยที่สุด

รายการประเมิน		ระดับความเหมาะสม				
		5	4	3	2	1
1.ด้านความต้องการของผู้ใช้โปรแกรม (Functional Requirement)						
1.1	ความสามารถของโปรแกรมในด้านการจัดการโปรแกรม					
1.2	ความสามารถของโปรแกรมในด้านการจัดการข้อสอบ					
1.3	ความสามารถของโปรแกรมในด้านการจัดการทดสอบ					
1.4	ความสามารถของโปรแกรมในด้านการจัดการรายผลการทดสอบ					
1.5	โปรแกรมตรงตามวัตถุประสงค์ในการสร้างโปรแกรม					
2. ด้านการทำงานของโปรแกรม (Functional)						
2.1	ความถูกต้องในการจัดเก็บข้อมูลนำเข้า					
2.2	ความถูกต้องในการค้นหาข้อมูล					
2.3	ความถูกต้องในการปรับปรุงแก้ไขข้อมูล					
2.4	ความถูกต้องในการลบข้อมูล					
2.5	ความถูกต้องของผลลัพธ์ที่ได้จากการประมวลผลในโปรแกรม					
2.6	ความถูกต้องของการผลลัพธ์ในรูปแบบรายงาน					
2.7	ความรวดเร็วในการประมวลผลของระบบ					
2.8	ความน่าเชื่อถือได้ของระบบ					
2.9	ความครอบคลุมของโปรแกรมที่พัฒนากับระบบงานจริง					
2.10	การป้องกันข้อผิดพลาดที่อาจเกิดขึ้น					

รายการประเมิน		ระดับความเหมาะสม				
		5	4	3	2	1
3. ด้านการใช้งาน (Usability)						
3.1	ความง่ายต่อการใช้งานของระบบ					
3.2	ความเหมาะสมในการเลือกใช้ชนิดตัวอักษรบนจอภาพ					
3.3	ความเหมาะสมในการเลือกใช้ขนาดของตัวอักษรบนจอภาพ					
3.4	ความเหมาะสมในการใช้สีของตัวอักษรและรูปภาพ					
3.5	ความเหมาะสมในการใช้ข้อความเพื่ออธิบายสื่อความหมาย					
3.6	ความเหมาะสมในการใช้สัญลักษณ์หรือรูปภาพในการสื่อความหมาย					
3.7	ความเป็นมาตรฐานเดียวกันในการออกแบบหน้าจอภาพ					
3.8	ความเหมาะสมในการปฏิสัมพันธ์โต้ตอบกับผู้ใช้					
3.9	ความเหมาะสมในการวางตำแหน่งของส่วนประกอบบนจอภาพ					
3.10	คำศัพท์ที่ใช้ผู้ใช้มีความคุ้นเคยและสามารถปฏิบัติตามได้โดยง่าย					
4. ด้านการรักษาความปลอดภัย (Security)						
4.1	การกำหนดรหัสผู้ใช้ และรหัสผ่านในการตรวจสอบผู้เข้าใช้ระบบ					
4.2	การตรวจสอบสิทธิ์ก่อนการใช้งานของผู้ใช้ระบบในระดับต่าง ๆ					
4.3	การควบคุมให้ใช้งานตามสิทธิ์ผู้ใช้ได้อย่างถูกต้อง					
4.4	การป้องกันการกำหนดรหัสผ่านอย่างง่าย					
4.5	การป้องกันการล้มเหลวของข้อมูลในระบบ					
5.ด้านความชัดเจนของคู่มือการใช้โปรแกรม (Program Manual)						
5.1	มีการกล่าวถึงความเป็นมาของการพัฒนาโปรแกรมอย่างชัดเจนในคู่มือการใช้โปรแกรม					
5.2	คู่มือการใช้โปรแกรมแสดงวิธีการใช้งานอย่างมีลำดับขั้นตอน					
5.3	ภาษาที่ใช้ในคู่มือการใช้โปรแกรมเข้าใจง่าย					
5.4	คู่มือมีการใช้ภาพประกอบการอธิบายกระบวนการต่าง ๆ อย่างชัดเจน					
5.5	หลังจากอ่านคู่มือแล้ว ผู้ใช้มีความมั่นใจว่าสามารถใช้โปรแกรมได้					

ตอนที่ 2 ข้อเสนอแนะเพื่อการพัฒนาโปรแกรม

.....

.....

.....

.....

ภาคผนวก ช
แบบประเมินความเหมาะสมของโปรแกรม
การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ สำหรับผู้ใช้งาน

แบบประเมินความเหมาะสม (Black-Box Testing)
โปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ
สำหรับทดสอบ O-NET วิชาคณิตศาสตร์ มัธยมศึกษาชั้นปีที่ 6
(สำหรับผู้ใช้งาน)

คำชี้แจง

แบบประเมินความเหมาะสมของโปรแกรมการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ ฉบับนี้มีจุดมุ่งหมายเพื่อทราบความคิดเห็นที่มีต้อการใช้งานโปรแกรมการทดสอบ รวมทั้งคู่มือการใช้โปรแกรม เพื่อนำผลที่ได้มาเป็นข้อมูลในการพัฒนาโปรแกรมต่อไป

ตอนที่ 1 การประเมินความเหมาะสมในการใช้โปรแกรม

โปรดทำกดปุ่ม ◉ ในช่องที่ตรงกับความคิดเห็นของท่านหลังจากท่านได้ศึกษาคู่มือการใช้โปรแกรมและทดลองใช้โปรแกรมแล้ว โดยเกณฑ์ประเมินแบ่งได้ 5 ระดับ ดังนี้

ระดับ 5 หมายถึง โปรแกรมมีความเหมาะสมมากที่สุด
ระดับ 4 หมายถึง โปรแกรมมีความเหมาะสมมาก
ระดับ 3 หมายถึง โปรแกรมมีความเหมาะสมปานกลาง
ระดับ 2 หมายถึง โปรแกรมมีความเหมาะสมน้อย
ระดับ 1 หมายถึง โปรแกรมมีความเหมาะสมน้อยที่สุด

รายการประเมิน		ระดับความเหมาะสม				
		5	4	3	2	1
1. ด้านการใช้งาน (Usability)						
1.1	ความง่ายต่อการใช้งานของระบบ					
1.2	ความเหมาะสมในการเลือกใช้ชนิดตัวอักษรบนจอภาพ					
1.3	ความเหมาะสมในการเลือกใช้ขนาดของตัวอักษรบนจอภาพ					
1.4	ความเหมาะสมในการใช้สีของตัวอักษรและรูปภาพ					
1.5	ความเหมาะสมในการใช้ข้อความเพื่ออธิบายสื่อความหมาย					
1.6	ความเหมาะสมในการใช้สัญลักษณ์หรือรูปภาพในการสื่อความหมาย					
1.7	ความเป็นมาตรฐานเดียวกันในการออกแบบหน้าจอภาพ					
1.8	ความเหมาะสมในการปฏิสัมพันธ์โต้ตอบกับผู้ใช้					
1.9	ความเหมาะสมในการวางตำแหน่งของส่วนประกอบบนจอภาพ					
1.10	คำศัพท์ที่ใช้ผู้ใช้มีความคุ้นเคยและสามารถปฏิบัติตามได้โดยง่าย					
2.ด้านความชัดเจนของคู่มือการใช้โปรแกรม (Program Manual)						
2.1	มีการกล่าวถึงความเป็นมาของการพัฒนาโปรแกรมอย่างชัดเจนในคู่มือการใช้โปรแกรม					
2.2	คู่มือการใช้โปรแกรมแสดงวิธีการใช้งานอย่างมีลำดับขั้นตอน					

รายการประเมิน		ระดับความเหมาะสม				
		5	4	3	2	1
2.3	ภาษาที่ใช้ในคู่มือการใช้โปรแกรมเข้าใจง่าย					
2.4	คู่มือมีการใช้ภาพประกอบการอธิบายกระบวนการต่าง ๆ อย่างชัดเจน					
2.5	หลังจากอ่านคู่มือแล้ว ผู้ใช้มีความมั่นใจว่าสามารถใช้โปรแกรมได้					

ตอนที่ 2 ข้อเสนอแนะเพื่อการพัฒนาโปรแกรม

.....

.....

.....

.....