

ประสิทธิภาพของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
ภายใต้เงื่อนไขต่างกันโดยใช้วิธีการจำลองข้อมูล

ปริญญานิพนธ์
ของ
สุดารัตน์ หวลมุกดา

เสนอต่อบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาการศึกษาคุณวุฒิปบัณฑิต สาขาวิชาการทดสอบและวัดผลการศึกษา
พฤษภาคม 2550

ประสิทธิภาพของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของคำอ่านจำแนก
ภายใต้เงื่อนไขต่างกันโดยใช้วิธีการจำลองข้อมูล

บทคัดย่อ
ของ
สุดาร์ตน์ หวลมุกดา

เสนอต่อบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาการศึกษาคุณวุฒิปบัณฑิต สาขาวิชาการทดสอบและวัดผลการศึกษา
พฤษภาคม 2550

สุดารัตน์ หวลมุกดา. (2550). *ประสิทธิภาพของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกภายใต้เงื่อนไขต่างกันโดยใช้วิธีการจำลองข้อมูล*. ปรินญาณพนธ์ กศ.ด. (การทดสอบและวัดผลการศึกษา). กรุงเทพฯ: บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. คณะกรรมการควบคุม : รองศาสตราจารย์ ดร.บุญเชิด ภิญโญนนตพงษ์, รองศาสตราจารย์ ดร.สมสรร วงศ์อยู่น้อย, ผู้ช่วยศาสตราจารย์ ดร.องอาจ นัยพัฒน์.

การวิจัยครั้งนี้มีจุดมุ่งหมายเพื่อศึกษาประสิทธิภาพการประมาณค่าความสามารถของผู้สอบ และประสิทธิภาพสมมูลการใช้ข้อสอบในคลังข้อสอบของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก เมื่อใช้จำนวนข้อสอบในชั้นคลังข้อสอบ ขนาดคลังข้อสอบ ความยาวแบบทดสอบ และการควบคุมการแสดงผลข้อสอบ ต่างกัน ด้วยวิธีจำลองข้อมูล ค่าความสามารถจริงของกลุ่มตัวอย่าง 5,000 ค่าที่สุ่มจากการแจกแจงโค้งปกติ และนำมาทดสอบตามขั้นตอนการทดสอบแบบปรับเหมาะตามระดับชั้นของค่าอำนาจจำแนกด้วยโปรแกรมคอมพิวเตอร์ที่ผู้วิจัยพัฒนาขึ้น ภายใต้ปัจจัยที่แปรเปลี่ยน 4 เงื่อนไข คือ 1. จำนวนข้อสอบในชั้นคลังข้อสอบต่างกัน คือ 1.1 อัตราส่วนจำนวนข้อสอบเท่ากันทุกชั้น (1:1:1:1) 1.2 อัตราส่วนจำนวนข้อสอบชั้นแรกมากกว่าชั้นอื่น (2:0.5:0.5:1) 1.3 อัตราส่วนจำนวนข้อสอบชั้นสุดท้ายมากกว่าชั้นอื่น (1:0.5:0.5:2) 1.4 อัตราส่วนจำนวนข้อสอบชั้นแรกและชั้นสุดท้ายมากกว่าชั้นอื่น (1.5:0.5:0.5:1.5) 2. ขนาดคลังข้อสอบที่ต่างกันสี่ระดับ คือ จำนวน 200 400 600 และ 800 ข้อ 3. ความยาวแบบทดสอบต่างกันสามระดับ คือ มีจำนวน 20 32 และ 44 ข้อ 4. มีการควบคุมและไม่ควบคุมการแสดงผลข้อสอบ ว่ามีประสิทธิภาพการประมาณค่าความสามารถของผู้สอบและประสิทธิภาพสมมูลการใช้ข้อสอบในคลังข้อสอบอย่างไร โดยพิจารณาความคลาดเคลื่อนกำลังสองเฉลี่ยและความลำเอียงเฉลี่ย อัตราการทับซ้อนข้อสอบ ข้อสอบที่มีการแสดงมากเกินไป ข้อสอบที่มีการใช้ต่ำ และการแจกแจงอัตราการแสดงผลข้อสอบ

ผลการวิจัยสรุปได้ดังนี้

การจัดจำนวนข้อสอบเพิ่มขึ้นในชั้นคลังข้อสอบโดยเฉพาะการใช้ข้อสอบเพิ่มขึ้นในชั้นที่มีค่าอำนาจจำแนกสูงและการเพิ่มขนาดคลังข้อสอบ สามารถลดความคลาดเคลื่อนกำลังสองเฉลี่ยและความลำเอียงเฉลี่ย รวมทั้งสามารถลดอัตราการทับซ้อนของข้อสอบ ลดจำนวนข้อสอบที่มีการแสดงมากเกินไป และเพิ่มจำนวนข้อสอบที่มีการใช้ต่ำ การเพิ่มความยาวแบบทดสอบสามารถลดความคลาดเคลื่อนกำลังสองเฉลี่ย ลดความลำเอียงเฉลี่ย และลดข้อสอบที่มีการใช้ต่ำ แต่ทำให้อัตราการทับซ้อนข้อสอบสูงขึ้นและเพิ่มข้อสอบที่มีการแสดงมากเกินไป และการควบคุมการแสดงผลข้อสอบสามารถลดความคลาดเคลื่อนกำลังสองเฉลี่ย ลดความลำเอียงเฉลี่ย และลดข้อสอบที่มีการใช้ต่ำ แต่เพิ่มอัตราการทับซ้อนและข้อสอบที่มีการแสดงมากเกินไป ในขณะเดียวกันก็สามารถให้สมมูลการแจกแจงอัตราการแสดงผลข้อสอบได้

EFFICIENCY OF A-STRATIFIED COMPUTERIZED ADAPTIVE TESTING UNDER
DIFFERENT CONDITIONS USING SIMULATIVE DATA

AN ABSTRACT

BY

SUDARATANA HUALMUKDA

Presented in Partial Fulfillment of the Requirements for the
Doctor of Education Degree in Testing and Measurement
at Srinakharinwirot University

May 2007

Sudaratana Hualmukda. (2007). *Efficiency of a-Stratified Computerized Adaptive Testing under Different Conditions Using Simulative Data*. Dissertation, Ed.D. (Testing and Measurement) Bangkok : Graduate School, Srinakharinwirot University. Advisor Committee : Assoc. Prof. Dr. Boonchird Pinyoanuntopong, Assoc. Prof. Dr. Somsan Wongyounoi, Assist. Prof. Dr. Ongart Naiyapatana.

The purpose of this research was to investigate the estimative efficiency and item utilizable efficiency of a-stratified computerized adaptive testing with different number of items in the item pool, item pool sizes, test lengths, and item exposure control. It was done by the use of simulative data of true abilities of 5,000 samples selected by normal curve distribution random sampling. The data were then tested according to the procedures of a-stratified computerized adaptive testing developed by the researcher under four conditions of independent variables. 1) The number of items in the item pool was different in four ratios: 1.1 the ratio of item number in every stratum was equal (1:1:1:1), 1.2 the ratio of item number in first stratum was more than other strata (2:0.5:0.5:1), 1.3 the ratio of item number in last stratum was more than other strata (1:0.5:0.5:2), 1.4 the ratio of item number in first and last strata was more than other strata (1.5:0.5:0.5:1.5). 2) The item pool was in four different sizes of 200, 400, 600 and 800 items. 3) The test length was in three different levels of 20, 32 and 44 items. 4) The item exposure was controlled and uncontrolled. The estimative efficiency and item utilizable efficiency were examined by using mean square error and average bias, item overlap rate, overexposed items, underutilized items, and item exposure rate distribution.

The findings of research were as follows.

The increased number of items in item pool stratum, especially the use of increased items in the stratum with high discrimination, and the increased item pool could help reduce the mean square error and average bias, the item overlap rate, and overexposed items. They however increased the underutilized items. The arrangement of equal number of items in item pool stratum could provide the balance of item exposure rate distribution better than other methods. The increased item length could reduce the mean square error and average bias, and underutilized items. It increased item overlap rate and overexposed items. The item exposure control could reduce mean square error and average bias, and underutilized items. It increased item overlap rate, and overexposed items. At the same time, it provided the balance of item exposure rate distribution.

งานวิจัยนี้ได้รับทุนอุดหนุนการวิจัย
จาก
สำนักงานคณะกรรมการการอุดมศึกษา

ประกาศคุณูปการ

ปริญญานิพนธ์ฉบับนี้สำเร็จลุล่วงได้ด้วยดี จากความร่วมมือของบุคคลหลายฝ่าย ผู้วิจัยขอกราบขอบพระคุณ รองศาสตราจารย์ ดร.บุญเชิด ภิญโญนนตพงษ์ ประธานกรรมการควบคุมปริญญานิพนธ์ รองศาสตราจารย์ ดร. สมสรร วงษ์อยู่น้อย และ ผู้ช่วยศาสตราจารย์ ดร. งามอาจ นัยพัฒน์ กรรมการควบคุมปริญญานิพนธ์ ซึ่งได้เสียสละเวลาให้คำปรึกษาและคอยแนะนำตลอดระยะเวลาทำปริญญานิพนธ์

ขอกราบขอบพระคุณ อาจารย์ ดร. สุวพร เข้มแสง กรรมการแต่งตั้งเพิ่มเติม ที่กรุณาให้ข้อเสนอแนะที่มีคุณค่าทำให้ปริญญานิพนธ์ฉบับนี้มีความสมบูรณ์ยิ่งขึ้น

ขอกราบขอบพระคุณ ผู้ช่วยศาสตราจารย์ ดร. ปานใจ ธารทัศนวงศ์ กรรมการแต่งตั้งเพิ่มเติม ที่ให้ความช่วยเหลือและสละเวลาอันมีค่าในการพัฒนาโปรแกรมคอมพิวเตอร์สำหรับการวิจัยครั้งนี้ รวมทั้งบุคคลในครอบครัวของท่านที่ร่วมเป็นกำลังใจและเอื้ออาทรผู้วิจัยด้วยดีเสมอมา

ขอกราบขอบพระคุณ รองศาสตราจารย์ ดร.บุญเชิด ภิญโญนนตพงษ์ และอาจารย์ ดร. สุวพร เข้มแสง อาจารย์ที่ปรึกษานิพนธ์ ซึ่งคอยให้ความดูแลด้วยความห่วงใย และให้คำแนะนำอย่างดียิ่งตลอดระยะเวลาที่ศึกษาในมหาวิทยาลัย

ขอกราบขอบพระคุณ Prof. David J Weiss ภาควิชาจิตวิทยา มหาวิทยาลัยมินเนโซต้า ประเทศสหรัฐอเมริกา ที่กรุณาอนุเคราะห์จัดส่งเอกสารที่มีคุณค่านำมาใช้เป็นพื้นฐานความรู้สำหรับการทำวิจัยครั้งนี้

ขอกราบขอบพระคุณ ดร. ละเอียด แจ่มจันทร์ ผู้อำนวยการวิทยาลัยพยาบาลบรมราชชนนีนพรัตน์วชิระ ที่สนับสนุนให้ผู้วิจัยมีโอกาสศึกษาต่อในระดับปริญญาเอก และพี่น้องอาจารย์ทุกท่านที่ปฏิบัติหน้าที่แทนผู้วิจัยตลอดช่วงเวลาของการลาศึกษาต่อ

ขอกราบขอบพระคุณ บิดา มารดา ที่เป็นกำลังใจให้ผู้วิจัยมีจิตใจที่มุ่งมั่นต่อการเผชิญอุปสรรคต่างๆ

สุดท้ายนี้ ขอกราบขอบพระคุณอาจารย์ทุกท่านที่เคยอบรมสั่งสอนในช่วงเวลาที่ผ่านมามาจนผู้วิจัยสามารถประสบความสำเร็จด้านการศึกษาในวันนี้

สุดารัตน์ หวลมุกดา

สารบัญ

บทที่	หน้า
1 บทนำ.....	1
ภูมิหลัง	1
ความมุ่งหมายของการวิจัย	7
ความสำคัญของการวิจัย	7
ขอบเขตการวิจัย	8
กรอบแนวคิดในการวิจัย	10
นิยามศัพท์เฉพาะ	12
สมมุติฐานในการวิจัย	14
2 เอกสารและงานวิจัยที่เกี่ยวข้อง	15
ทฤษฎีการตอบข้อสอบ	16
แนวคิดทฤษฎีการตอบข้อสอบ	17
ข้อตกลงเบื้องต้นของทฤษฎีการตอบข้อสอบ	17
รูปแบบการตอบข้อสอบ	18
สารสนเทศข้อสอบและสารสนเทศแบบทดสอบ.....	20
การประมาณค่าความสามารถและพารามิเตอร์ข้อสอบ	21
การทดสอบแบบปรับเหมาะ	26
ความเป็นมาของการทดสอบแบบปรับเหมาะ	28
ความแตกต่างระหว่างแบบทดสอบแบบปรับเหมาะและแบบทดสอบ ประเพณีนิยม.....	31
กลยุทธ์ของการทดสอบแบบปรับเหมาะ	33
ความก้าวหน้าของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์.....	36
การกำหนดรูปแบบการตอบข้อสอบ	36
คลังข้อสอบ	38
ความยาวแบบทดสอบ	40
จุดเริ่มต้นการทดสอบ	41

สารบัญ (ต่อ)

บทที่		หน้า
2(ต่อ)	การเลือกข้อสอบ	43
	การประมาณค่าความสามารถ	45
	การแสดงข้อสอบ	47
	การจัดสมมูลเนื้อหา	51
	คุณภาพแบบทดสอบ	53
	กฎการยุติ	55
	การศึกษาในสถานการณ์จำลอง	57
	การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับขั้นของค่าอำนาจ จำแนก.....	59
	งานวิจัยที่เกี่ยวข้อง	72
3	วิธีดำเนินงานวิจัย	81
	กลุ่มตัวอย่าง	81
	ตัวแปรศึกษา	81
	เครื่องมือใช้ในการวิจัย	84
	การรวบรวมข้อมูล	85
	การวิเคราะห์ข้อมูล	85
	รูปแบบของการวิจัย	85
4	ผลการวิเคราะห์ข้อมูล.....	98
	ประสิทธิภาพการประมาณค่าความสามารถของผู้สอบ.....	100
	ประสิทธิภาพสมมูลการใช้ข้อสอบในคลังข้อสอบ.....	118
	สรุปวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับขั้นของค่า อำนาจจำแนกที่ไม่สามารถให้ผลการวิเคราะห์ได้	133
5	สรุป อภิปรายผล และข้อเสนอแนะ.....	138
	สรุปผลการวิจัย.....	141
	อภิปรายผล.....	145

สารบัญ (ต่อ)

บทที่		หน้า
5(ต่อ)	ประสิทธิภาพการประมาณค่าความสามารถของผู้สอบ.....	141
	ประสิทธิภาพสมมูลการใช้ข้อสอบในคลังข้อสอบ.....	143
	ข้อเสนอแนะ.....	152
	บรรณานุกรม	155
	ภาคผนวก	164
	ประวัติย่อผู้วิจัย.....	224

บัญชีตาราง

ตาราง	หน้า
1 โปรแกรมการประมาณค่าพารามิเตอร์รูปแบบต่างๆ	25
2 จำนวนข้อสอบในแต่ละชั้นคลังข้อสอบของวิธีการทดสอบแบบปรับเหมาะด้วย คอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก 4 วิธี.....	87
3 ความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) และความลำเอียงเฉลี่ย (Bias) ของ วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจ จำแนกวิธี ที่ 1 (a-STR1).....	101
ความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) และความลำเอียงเฉลี่ย (Bias) ของ วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจ จำแนกวิธี ที่ 2 (a-STR2).....	103
5 ความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) และความลำเอียงเฉลี่ย(Bias) ของ วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจ จำแนกวิธี ที่ 3 (a-STR3).....	105
6 ความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) และความลำเอียงเฉลี่ย (Bias) ของ วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจ จำแนกวิธี ที่ 4 (a-STR4).....	107
7 เปรียบเทียบความคลาดเคลื่อนกำลังสองเฉลี่ยและความลำเอียงเฉลี่ยของวิธีการ ทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก 4วิธี.....	109
8 เปรียบเทียบความคลาดเคลื่อนกำลังสองเฉลี่ยและความลำเอียงเฉลี่ยของวิธีการ ทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก 4วิธี ขนาดคลังข้อสอบ 200 ข้อ ความยาวแบบทดสอบ 32 ข้อ.....	111
9 เปรียบเทียบความคลาดเคลื่อนกำลังสองเฉลี่ยและความลำเอียงเฉลี่ยของวิธีการ ทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก 4วิธี ที่ขนาดคลังข้อสอบ 600 ข้อ ความยาวแบบทดสอบ 32 ข้อ.....	112
10 เปรียบเทียบความคลาดเคลื่อนกำลังสองเฉลี่ยและความลำเอียงเฉลี่ยของวิธีการ ทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก ร่วมกับการควบคุมการแสดงข้อสอบ 4 วิธี.....	114

บัญชีตาราง (ต่อ)

ตาราง		หน้า
11	เปรียบเทียบความคลาดเคลื่อนกำลังสองเฉลี่ยของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก 4 วิธี ร่วมกับการควบคุมและไม่ควบคุมการแสดงผลข้อสอบ ขนาดคลังข้อสอบ 800 ข้อ ความยาวแบบทดสอบ 32 ข้อ.....	116
12	เปรียบเทียบความลำเอียงเฉลี่ยของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก 4 วิธี ร่วมกับการควบคุมและไม่ควบคุมการแสดงผลข้อสอบ ขนาดคลังข้อสอบ 800 ข้อ ความยาวแบบทดสอบ 32 ข้อ.....	116
13	อัตราการทับซ้อนข้อสอบ ข้อสอบที่มีการแสดงมากเกินไป และข้อสอบที่มีการใช้ต่ำของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นค่าอำนาจจำแนกวิธีที่ 1 (a-STR1).....	119
14	อัตราการทับซ้อนข้อสอบ ข้อสอบที่มีการแสดงมากเกินไป ข้อสอบที่มีการใช้ต่ำของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นค่าอำนาจจำแนกวิธีที่ 2 (a-STR2)	121
15	อัตราการทับซ้อนข้อสอบ ข้อสอบที่มีการแสดงมากเกินไป ข้อสอบที่มีการใช้ต่ำของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นค่าอำนาจจำแนกวิธีที่ 3 (a-STR3).....	123
16	อัตราการทับซ้อนข้อสอบ ข้อสอบที่มีการแสดงมากเกินไป ข้อสอบที่มีการใช้ต่ำของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นค่าอำนาจจำแนกวิธีที่ 4 (a-STR4).....	125
17	เปรียบเทียบอัตราการทับซ้อนข้อสอบและข้อสอบที่มีการแสดงมากเกินไปของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก 4 วิธี.....	127
18	เปรียบเทียบข้อสอบที่มีการใช้ต่ำและการแจกแจงอัตราการแสดงผลข้อสอบของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก 4 วิธี.....	128
19	เปรียบเทียบอัตราการทับซ้อนข้อสอบและข้อสอบที่มีการแสดงมากเกินไปของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกทั้ง 4 วิธีร่วมกับการควบคุมการแสดงผลข้อสอบ.....	129

บัญชีตาราง (ต่อ)

ตาราง		หน้า
20	เปรียบเทียบข้อสอบที่มีการใช้คำและการแจกแจงอัตราการแสดงข้อสอบของ วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของคำอ่าน จำแนกทั้ง 4 วิธีร่วมกับการควบคุมการแสดงผลข้อสอบ.....	130
21	เปรียบเทียบอัตราการทั้บซ้อนข้อสอบของวิธีการทดสอบแบบปรับเหมาะด้วย คอมพิวเตอร์ตามระดับชั้นของคำอ่านจำแนก 4 วิธี ร่วมกับการควบคุมและ ไม่ควบคุมการแสดงผลข้อสอบ ขนาดคลังข้อสอบ 800 ข้อ ความยาว แบบทดสอบ 32 ข้อ.....	131
22	เปรียบเทียบข้อสอบที่มีการแสดงมากเกินไปของวิธีการทดสอบแบบปรับเหมาะ ด้วยคอมพิวเตอร์ตามระดับชั้นของคำอ่านจำแนก 4 วิธี ร่วมกับการควบคุม และไม่ควบคุมการแสดงผลข้อสอบ ขนาดคลังข้อสอบ 800 ข้อ ความยาว แบบทดสอบ 32 ข้อ.....	132
23	เปรียบเทียบข้อสอบที่มีการใช้คำของวิธีการทดสอบแบบปรับเหมาะด้วย คอมพิวเตอร์ตามระดับชั้นของคำอ่านจำแนก 4 วิธี ร่วมกับการควบคุมและ ไม่ควบคุมการแสดงผลข้อสอบ ขนาดคลังข้อสอบ 800 ข้อ ความยาว แบบทดสอบ 32ข้อ.....	132
24	สรุปวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นคำอ่าน จำแนก 4 วิธีไม่สามารถให้ผลการวิเคราะห์ได้.....	134
25	การทดสอบของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้น ของคำอ่านจำแนก ขนาดคลังข้อสอบ 200 ข้อ ความยาวแบบทดสอบ 20ข้อ	165
26	การทดสอบของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้น ของคำอ่านจำแนกร่วมกับการควบคุมการแสดงผลข้อสอบ ขนาดคลังข้อสอบ 200 ข้อ ความยาวแบบทดสอบ 20 ข้อ.....	166
27	การทดสอบของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้น ของคำอ่านจำแนก ขนาดคลังข้อสอบ 200 ข้อ ความยาวแบบทดสอบ 32ข้อ	167
28	การทดสอบของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้น ของคำอ่านจำแนก ขนาดคลังข้อสอบ 200 ข้อ ความยาวแบบทดสอบ 44ข้อ	168
29	การทดสอบของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้น ของคำอ่านจำแนก ขนาดคลังข้อสอบ 400 ข้อ ความยาวแบบทดสอบ 20ข้อ	169

บัญชีตาราง (ต่อ)

[illegible]

บัญชีตาราง (ต่อ)

ตาราง	หน้า
41 การทดสอบของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้น ของค่าอำนาจจำแนก ขนาดคลังข้อสอบ 800 ข้อ ความยาวแบบทดสอบ 20 ข้อ	181
42 การทดสอบของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้น ของค่าอำนาจจำแนกร่วมกับการควบคุมการแสดงผลข้อสอบ ขนาดคลัง ข้อสอบ 800 ข้อ ความยาวแบบทดสอบ 20 ข้อ.....	182
43 การทดสอบของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้น ของค่าอำนาจจำแนก ขนาดคลังข้อสอบ 800 ข้อ ความยาวแบบทดสอบ 32 ข้อ	183
44 การทดสอบของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้น ของค่าอำนาจจำแนกร่วมกับการควบคุมการแสดงผลข้อสอบ ขนาดคลัง ข้อสอบ 800 ข้อ ความยาวแบบทดสอบ 32 ข้อ.....	184
45 การทดสอบของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้น ของค่าอำนาจจำแนก ขนาดคลังข้อสอบ 800 ข้อ ความยาวแบบทดสอบ 44 ข้อ	185
46 การทดสอบของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้น ของค่าอำนาจจำแนกร่วมกับการควบคุมการแสดงผลข้อสอบ ขนาดคลัง ข้อสอบ 800 ข้อ ความยาวแบบทดสอบ 44 ข้อ.....	186
47 ความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Square Error : MSE) เมื่อสิ้นสุดการ ทดสอบของวิธีกำหนด.....	187
48 ความลำเอียงเฉลี่ย (Average Bias) เมื่อสิ้นสุดการทดสอบของวิธีกำหนด.....	188
49 จำนวนข้อสอบที่มีการแสดงมากเกินไป (Overexpose Item) เมื่อสิ้นสุดการ ทดสอบของวิธีกำหนด.....	189
50 จำนวนข้อสอบที่มีการใช้ต่ำ (Underutilized Item) เมื่อสิ้นสุดการทดสอบของ วิธีกำหนด.....	190
51 อัตราการทับซ้อนข้อสอบ (Item Overlap Rate) * เมื่อสิ้นสุดการทดสอบของวิธี กำหนด.....	191
52 การแจกแจงอัตราการแสดงผลข้อสอบ (Item Exposure Rate Distribution) ด้วย ไคสแควร์ เมื่อสิ้นสุดการทดสอบของวิธีกำหนด.....	192

บัญชีภาพประกอบ

ภาพประกอบ	หน้า
1 กรอบแนวคิดการวิจัย.....	11
2 ขั้นตอนการทดสอบแบบปรับเหมาะ.....	27
3 สารสนเทศข้อสอบต่อค่าอำนาจจำแนกของข้อสอบ 5 กรณี.....	60
4 ขั้นตอนการกำหนดจุดเริ่มต้นการทดสอบด้วยข้อสอบเทียม 3 ข้อ.....	89
5 ขั้นตอนวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของ ค่าอำนาจจำแนก.....	92
6 ขั้นตอนการเตรียมพารามิเตอร์การแสดงผลข้อสอบของวิธีการทดสอบแบบ ปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกร่วมกับ การควบคุมการแสดงผลข้อสอบ.....	94
7 ขั้นตอนวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของ ค่าอำนาจจำแนกร่วมกับการควบคุมการแสดงผลข้อสอบ.....	96
8 การแจกแจงโค้งปกติมาตรฐานของค่าความยากของข้อสอบ.....	141

บทที่ 1

บทนำ

ภูมิหลัง

การวัดผลทางการศึกษานั้นอยู่บนพื้นฐานแนวคิดที่ว่า พฤติกรรมแสดงออกกำหนดมาจากลักษณะภายในตัวบุคคลซึ่งสังเกตไม่ได้โดยตรง ต้องอาศัยสิ่งเร้ามากระตุ้นให้บุคคลแสดงพฤติกรรมตอบสนอง เครื่องมือการวัดผลสัมฤทธิ์ทางการศึกษาชนิดหนึ่งคือแบบทดสอบ (Test) ใช้เป็นตัวแทนเนื้อหาการเรียนรู้ของผู้เรียน เปรียบเสมือนสิ่งเร้านำไปกระตุ้นผู้เรียนให้แสดงพฤติกรรมตอบสนอง ผลการตอบแบบทดสอบคือสารสนเทศเกี่ยวกับผลสัมฤทธิ์การศึกษาตามหลักสูตรกำหนด หรือความสามารถทางการเรียนรู้ด้านต่างๆ ที่ต้องการให้เกิดขึ้นภายในตัวผู้เรียน ถ้าได้รับสารสนเทศมากเพียงพอระดับความสามารถภายในตัวผู้เรียนจะประมาณได้อย่างแม่นยำและลดการตัดสินใจผิดพลาด

แบบทดสอบที่นิยมใช้กันโดยทั่วไป ได้แก่ แบบทดสอบประเพณีนิยม (Conventional Test หรือ Traditional Test) หรือเรียกว่า แบบทดสอบใช้กระดาษ-ดินสอ (Paper-Pencil Test : P&P Test) อาศัยหลักทฤษฎีการทดสอบแบบมาตรฐานเดิม (Classical Test Theory : CTT) มีลักษณะดังนี้ คือ ผู้สอบกลุ่มเป้าหมายทุกคนจะได้รับแบบทดสอบชุดเดียวกัน จำนวนข้อสอบเท่ากัน แบบทดสอบส่วนใหญ่ออกแบบให้มีค่าความยากปานกลาง ทำให้มีประสิทธิภาพสูงเหมาะกับการวัดผู้สอบกลุ่มระดับความสามารถปานกลาง แต่จำนวนข้อสอบเหมาะสมกับผู้สอบระดับความสามารถสูงและความสามารถต่ำในแบบทดสอบนั้นมีจำนวนน้อย ดังนั้นประสิทธิภาพในการวัดผู้สอบระดับความสามารถสูงและความสามารถต่ำจะลดลง ผู้สอบระดับความสามารถสูงเสียเวลาในการทำข้อสอบง่ายขึ้นแต่ได้รับสารสนเทศเพียงเล็กน้อย ในขณะที่ผู้สอบระดับความสามารถต่ำต้องทำข้อสอบยากเกินไปมักเกิดความท้อแท้และได้รับสารสนเทศไม่เพียงพอหรือไม่ได้เลย ข้อจำกัดอื่นๆ คือ ทฤษฎีการทดสอบแบบมาตรฐานเดิมไม่สามารถบอกประสิทธิภาพของข้อสอบรายข้อ สำหรับการวัดผู้สอบระดับความสามารถต่างๆ ได้ จึงไม่สามารถทำนายคุณสมบัติการวัดทางจิตวิทยาของแบบทดสอบได้ชัดเจนเมื่อจัดให้กับผู้สอบกลุ่มเป้าหมาย (Lord, 1980 : 3-10) นอกจากนี้ การวัดความคลาดเคลื่อนสำหรับผู้สอบทุกคนที่ตอบแบบทดสอบฉบับนั้นถือว่ามีความเท่ากันซึ่งเป็นไปไม่ได้ในสถานการณ์จริง จากข้อจำกัดของทฤษฎีการทดสอบแบบมาตรฐานเดิสดังกล่าว เป็นเหตุผลักดันไปสู่การค้นคว้าเพื่อหาแนวทางแก้ไข พัฒนาเป็นทฤษฎีการตอบข้อสอบ (Item Response Theory : IRT) หรือเรียกอีกชื่อหนึ่งว่า ทฤษฎีลักษณะแฝง (Latent Trait Theory) ในช่วงเวลาต่อมา

ทฤษฎีการตอบข้อสอบใช้แนวทางคณิตศาสตร์อธิบายสิ่งที่เกิดขึ้นเมื่อผู้สอบเผชิญกับข้อสอบ จากแนวคิดที่ว่า ความน่าจะเป็นของการตอบข้อสอบถูกต้องอธิบายได้ด้วยฟังก์ชันตำแหน่งบุคคลบน ลักษณะแฝง (Latent Trait) ควบคู่ไปกับพารามิเตอร์ลักษณะของข้อสอบ เรียกว่า ฟังก์ชันการตอบ ข้อสอบ (Item Response Function : IRF) และ แสดงเป็นกราฟ เรียกว่า โค้งลักษณะข้อสอบ (Item Characteristic Curve : ICC) ทฤษฎีการตอบข้อสอบมุ่งเน้นการวัดความเป็นมิตินี้ของความรู้อหรือ ลักษณะที่ต้องการ ตำแหน่งของข้อสอบแต่ละข้อบนมิตินี้อยู่ในรูปของความยากของข้อสอบ (b) และตำแหน่งของผู้สอบแต่ละคนบนมิตินี้อยู่ในรูปของค่าความสามารถของผู้สอบ (θ) ความน่าจะเป็นของการตอบข้อสอบถูกต้องแสดงในรูปของปฏิสัมพันธ์ระหว่างค่าความยากของข้อสอบและค่า ความสามารถของผู้สอบ

หลักการของทฤษฎีการตอบข้อสอบ (Item Response Theory : IRT) นำไปพัฒนาเป็นการ ทดสอบแบบปรับเหมาะ (Adaptive Testing) มีลักษณะเป็นการทดสอบที่ใช้ข้อสอบแตกต่างกันจัด ให้กับผู้สอบแต่ละคนโดยพิจารณาตามความสามารถภายในของผู้สอบ หรือผู้สอบแต่ละคนได้รับ ข้อสอบเหมาะสมกับความสามารถของตน การทดสอบเริ่มต้นจากผู้สอบได้รับข้อสอบข้อแรกที่มีความ ยากปานกลาง ถ้าตอบข้อสอบถูกต้องข้อสอบข้อถัดไปจะมีความยากเพิ่มขึ้น ในทางกลับกัน ถ้าตอบ ข้อสอบผิดข้อสอบข้อถัดไปจะมีความยากลดลง การทดสอบดำเนินต่อไปจนกระทั่งผลการทดสอบ สามารถบอกค่าความสามารถของผู้สอบใกล้เคียงความเป็นจริงมากที่สุด การทดสอบจึงยุติ ดังนั้น วัตถุประสงค์ของการทดสอบแบบปรับเหมาะ คือ เพื่อประมาณค่าความสามารถหรือลักษณะของ ผู้สอบโดยการตอบชุดข้อสอบที่เหมาะสมกับความสามารถของผู้สอบคนนั้น เมื่อนำคอมพิวเตอร์เข้ามา ช่วยจัดการกับการคำนวณสถิติที่ซับซ้อน เรียกโดยทั่วไปว่า การทดสอบแบบปรับเหมาะด้วย คอมพิวเตอร์ (Computerized Adaptive Testing : CAT)

การเลือกข้อสอบนับว่าเป็นหัวใจสำคัญของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ (Weiss; & Kingsbury. 1984) วิธีการเลือกข้อสอบที่ดีต้องพิจารณาประสิทธิภาพสูงสุดในการประมาณ ค่าความสามารถให้ใกล้เคียงค่าความสามารถจริงของผู้สอบ รวมทั้งพิจารณาการจัดสมดุลของเนื้อหา และการควบคุมการแสดงข้อสอบไปพร้อมๆ กันด้วย การควบคุมสมดุลเนื้อหาเพื่อต้องการให้แน่ใจว่า แบบทดสอบแตกต่างกันที่จัดให้ผู้สอบแต่ละคนครอบคลุมสัดส่วนเนื้อหาเหมือนกัน สามารถอ้างอิงไป ยังการวัดผู้สอบคนอื่นๆ ในลักษณะเดียวกัน การควบคุมการแสดงผลข้อสอบเพื่อไม่ต้องการให้ข้อสอบ บางข้อนำไปใช้บ่อยเกินไปจนรู้จักกันทั่วไปในกลุ่มผู้สอบ เป็นสาเหตุทำให้ความเที่ยงตรงในการวัด ลดลงได้ การเลือกข้อสอบโดยทั่วไปใช้วิธีสารสนเทศสูงสุดของฟิชเชอร์ (Fisher Maximum Information) ซึ่งเลือกข้อสอบที่ให้สารสนเทศสูงสุด ณ ตำแหน่งค่าประมาณความสามารถของผู้สอบ คนนั้น วิธีนี้ถึงแม้จะให้ประสิทธิภาพสูงโดยใช้ข้อสอบจำนวนน้อย แต่ข้อสอบสารสนเทศสูงสุดมัก เป็นข้อสอบที่มีค่าอำนาจจำแนกสูงด้วย ข้อสอบลักษณะนี้ใช้บ่อยครั้งเกินไปจนเป็นที่รู้จักทั่วไปในกลุ่ม

ผู้สอบ นำไปสู่ปัญหาความไม่ปลอดภัยของข้อสอบ ในขณะที่เดียวกัน ข้อสอบอำนาจจำแนกต่ำใช้น้อยหรือไม่ได้ใช้เลยและมีจำนวนมาก เพื่อให้เกิดความสมดุลของการใช้ข้อสอบนี้ นักทดสอบหลายท่านได้พยายามคิดค้นและพัฒนาวิธีการควบคุมความถี่ในการแสดงข้อสอบ วิธีซึ่งได้รับความนิยมมากคือวิธีการควบคุมการแสดงผลข้อสอบซิมสัน-เฮตเตอร์ (Simpson-Hetter Procedure : SH) (Stocking; & Lewis. 2000 : 163-182 ; citing Simpson; & Hetter. 1985) เมื่อใช้การเลือกข้อสอบวิธีสารสนเทศสูงสุดของฟิชเชอร์ร่วมกับวิธีการควบคุมการแสดงผลข้อสอบซิมสัน-เฮตเตอร์ พบว่า สามารถควบคุมการแสดงผลข้อสอบไม่ให้สูงเกินอัตรากำหนด แต่ไ้การแจกแจงอัตราการแสดงผลข้อสอบโดยรวมพบว่ามีแนวโน้ม (Chang; & Ying. 1999) โดยเฉพาะข้อสอบอำนาจจำแนกต่ำมีอัตราการแสดงผลข้อสอบเป็นศูนย์และข้อสอบอำนาจจำแนกสูงยังพบว่ามีผลการแสดงผลข้อสอบสูงในระยะเริ่มต้นของการทดสอบแบบปรับเหมาะเป็นเช่นนี้เพราะว่า วิธีการควบคุมการแสดงผลข้อสอบซิมสัน-เฮตเตอร์ควบคุมความน่าจะเป็นตามเงื่อนไขการจัดข้อสอบให้ผู้สอบภายหลังจากการเลือกข้อสอบแล้ว แต่ไม่ได้ควบคุมความน่าจะเป็นต่อการเลือกข้อสอบโดยตรง ดังนั้น ข้อสอบที่มีความน่าจะเป็นในการเลือกต่ำก็ยังคงมีการแสดงผลข้อสอบต่ำอยู่ (Deng; & Ansley. 2003)

ซางและยিং (Chang; & Ying. 1999) นำเสนอวิธีการเลือกข้อสอบสำหรับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ เรียกทั่วไปว่า การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก (a-Stratified Computerized Adaptive Testing : a-STR) เป็นการทดสอบโดยใช้ข้อสอบแตกต่างกันจัดให้กับผู้สอบแต่ละคน พิจารณาเลือกใช้ข้อสอบตามความสามารถภายในของผู้สอบที่ต้องการวัดและดำเนินการทดสอบตามระดับชั้นของค่าอำนาจจำแนกของคลังข้อสอบด้วยเครื่องคอมพิวเตอร์ วิธีนี้ประยุกต์มาจากกลยุทธ์รูปแบบปรับระดับชั้นของไวส์ (Weiss. 1973) เพื่อแก้ปัญหาความไม่สมดุลการใช้ข้อสอบในคลังข้อสอบ จากการเลือกข้อสอบวิธีสารสนเทศสูงสุดของฟิชเชอร์

ขั้นตอนวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกมีดังนี้คือ แบ่งชั้นคลังข้อสอบตามค่าอำนาจจำแนกของข้อสอบ ชั้นแรกบรรจุข้อสอบอำนาจจำแนกต่ำ ชั้นถัดไปบรรจุข้อสอบอำนาจจำแนกสูงขึ้น จนกระทั่งถึงชั้นสุดท้ายบรรจุข้อสอบอำนาจจำแนกสูงสุด ข้อสอบในแต่ละชั้นคลังข้อสอบมีจำนวนเท่ากัน แบ่งการทดสอบเป็นขั้นในแต่ละชั้นของคลังข้อสอบเลือกข้อสอบตามค่าความยากของข้อสอบใกล้เคียงกับค่าประมาณความสามารถของผู้สอบครั้งหลังสุด เริ่มทดสอบจากชั้นแรกไปจนถึงชั้นสุดท้าย การทดสอบถึงยุติ

วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกมีข้อดี คือ ให้ประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบและให้สมดุลการใช้ข้อสอบในคลังข้อสอบไปพร้อมกันโดยใช้ข้อสอบอำนาจจำแนกต่ำระยะเริ่มต้นของการทดสอบ และใช้ข้อสอบอำนาจจำแนกสูงระยะท้ายของการทดสอบ ด้วยเหตุผลว่าในระยะเริ่มต้นการทดสอบ เมื่อค่าประมาณความสามารถยังห่างไกลจากค่าความสามารถจริง ข้อสอบอำนาจจำแนกสูงยังไม่ให้สารสนเทศสูงสุด ทำให้สูญเสีย

ประสิทธิภาพที่ควรได้รับ ดังนั้น ควรใช้ข้อสอบอำนาจจำแนกต่ำแทนในระยะเริ่มต้นการทดสอบเมื่อค่าประมาณความสามารถเข้าใกล้ค่าความสามารถจริงระยะท้ายๆ ของการทดสอบ การใช้ข้อสอบอำนาจจำแนกสูงจะให้สารสนเทศสูงสุด ช่วยให้การประมาณค่าความสามารถของผู้สอบเป็นไปอย่างมีประสิทธิภาพ (Chang; & Ying. 1996)

การศึกษาประสิทธิภาพการเลือกข้อสอบวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก เมื่อเปรียบเทียบกับวิธีสารสนเทศสูงสุดของฟิชเชอร์ร่วมกับวิธีการควบคุมการแสดงผลข้อสอบของซิมสัน-เฮดเทอร์ พบว่า วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก มีผลดีในด้านให้ความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำ เพิ่มสมมูลการใช้ข้อสอบในคลังข้อสอบโดยกระจายการใช้ข้อสอบไปยังข้อสอบอื่นในคลังข้อสอบ ลดความเบ้ของโค้งการแจกแจงอัตราการแสดงผลข้อสอบ และให้อัตราการทับซ้อนข้อสอบต่ำ (Chang; & Ying. 1999) เมื่อประยุกต์วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก ร่วมกับวิธีการควบคุมการแสดงผลข้อสอบของซิมสัน-เฮดเทอร์ หรือมีชื่อเรียกเฉพาะว่า วิธีขยายช่วงระดับชั้น (Enhanced Stratified Method) ผลการศึกษาพบว่า วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกร่วมกับการควบคุมการแสดงผลข้อสอบวิธีซิมสัน-เฮดเทอร์ ให้สมมูลการใช้ข้อสอบได้ดีกว่าการใช้วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกแบบเดิมเพียงอย่างเดียว (Leung; Chang; & Hau. 2002)

การศึกษาประสิทธิภาพการเลือกข้อสอบวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกได้รับความสนใจและมีงานวิจัยเผยแพร่อย่างต่อเนื่อง (Chang; & Ying. 1999) ; (Parshall; Harnes; & Kromrey. 2000) ; (Hau; & Chang. 2001) ; (Chang; Qian; & Ying. 2001) ; (Leung; Chang; & Hau. 2002) ; (Pastor; Dodd; & Chang. 2002) ; (Leung; Chang; & Hau. 2003) ; (Chang; & van der Linden. 2003) ; (Deng; & Ansley. 2003) ; (van der Linden; & Chang. 2003) การพัฒนาวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกมีแนวโน้มเพื่อเพิ่มประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบให้สูงขึ้น โดยใช้ข้อสอบจำนวนน้อยที่สุดเท่าที่เป็นไปได้ ในขณะเดียวกัน จำเป็นต้องจัดสมมูลการใช้ข้อสอบในคลังข้อสอบให้มีความถี่เท่าเทียมกัน และอยู่ในอัตราการแสดงผลข้อสอบกำหนดเพื่อรักษาความปลอดภัยของข้อสอบ นอกจากนี้ การเพิ่มประสิทธิภาพการทดสอบจะสัมพันธ์กับเงื่อนไขการทดสอบแบบปรับเหมาะต่างๆ เช่น การกำหนดรูปแบบการตอบข้อสอบ ลักษณะคลังข้อสอบ จุดเริ่มต้นการทดสอบ วิธีการเลือกข้อสอบ วิธีการประมาณค่าความสามารถ กฎการยุติการทดสอบ เป็นต้น กล่าวได้ว่าการเลือกข้อสอบวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกเพื่อให้ได้ประสิทธิภาพสูงนั้นนอกจากคุณสมบัติภายในตัวมันเองแล้ว ยังขึ้นอยู่กับ การวางเงื่อนไขต่างๆ ในการ

ทดสอบ เมื่อเงื่อนไขข้อใดข้อหนึ่งเปลี่ยนแปลงไปจำเป็นต้องทำการทดสอบเบื้องต้นเพื่อศึกษาผลจากการเปลี่ยนแปลงที่อาจกระทบต่อประสิทธิภาพการทดสอบ

วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับขั้นของค่าอำนาจจำแนก เมื่อวางเงื่อนไขการใช้ข้อสอบจำนวนไม่เท่ากันในแต่ละชั้นของคลังข้อสอบ มีความเป็นไปได้ที่จะเพิ่มประสิทธิภาพการประมาณค่าความสามารถของผู้สอบ และประสิทธิภาพสมดุลการใช้ข้อสอบในคลังข้อสอบให้สูงขึ้นกว่าการใช้ข้อสอบจำนวนเท่ากันในแต่ละชั้นของคลังข้อสอบ จากเหตุผลบางประการดังนี้

การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับขั้นของค่าอำนาจจำแนกสำหรับประชากรผู้สอบจำนวนมาก ค่าความสามารถของประชากรผู้สอบมีแนวโน้มจะกระจายอยู่รอบๆ ค่ากลางของโค้งปกติ การเลือกข้อสอบมักเลือกข้อสอบความยากปานกลางจัดให้ประชากรผู้สอบ ทำให้ข้อสอบความยากปานกลางอาจมีจำนวนไม่เพียงพอต่อการประมาณค่าความสามารถของผู้สอบให้เป็นอย่างมีประสิทธิภาพ และการแสดงข้อสอบความยากปานกลางอาจมีความถี่สูงจนเสี่ยงต่อความพลอดภัยของข้อสอบ ดังนั้น ถ้าจัดข้อสอบให้มีจำนวนมากเพียงพอรอบๆ ระดับความสามารถปานกลางทั้งระยะเริ่มต้นและระยะท้ายของการทดสอบ คาดว่า จะช่วยให้การประมาณค่าความสามารถของผู้สอบมีประสิทธิภาพมากขึ้นและการแสดงข้อสอบไม่สูงเกินอัตรากำหนด

ระยะเริ่มต้นของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับขั้นของค่าอำนาจจำแนก การใช้ข้อสอบอำนาจจำแนกต่ำโดยเลือกข้อสอบความยากใกล้เคียงค่าประมาณความสามารถของผู้สอบสามารถให้สารสนเทศสูงได้ (Leung; Chang; & Hau. 2001) ช่วยลดความแตกต่างระหว่างค่าประมาณความสามารถและค่าความสามารถจริงของผู้สอบได้ดีในระดับหนึ่ง ก่อนใช้ข้อสอบอำนาจจำแนกสูงในการทดสอบลำดับต่อไป ดังนั้น ถ้าจัดข้อสอบอำนาจจำแนกต่ำให้มีจำนวนมากเพียงพอในระยะเริ่มต้นของการทดสอบ คาดว่า จะช่วยเพิ่มประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบ และกระจายการใช้ข้อสอบไปยังข้อสอบอำนาจจำแนกต่ำ สร้างสมดุลการใช้ข้อสอบได้มากขึ้น

ระยะท้ายของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับขั้นของค่าอำนาจจำแนก เมื่อค่าประมาณความสามารถใกล้เคียงค่าความสามารถจริงของผู้สอบ การใช้ข้อสอบอำนาจจำแนกสูงจะให้ประสิทธิภาพสูงสุดในการทดสอบ ด้วยเหตุนี้ ข้อสอบอำนาจจำแนกสูงควรมีจำนวนมากเพียงพอระยะท้ายๆ ของการทดสอบ ผลที่คาดว่าจะได้รับ คือ เพิ่มประสิทธิภาพในการประมาณค่าความสามารถให้ใกล้เคียงค่าความสามารถจริงของผู้สอบมากที่สุด พร้อมทั้งกระจายการแสดงผลข้อสอบไปยังข้อสอบอำนาจจำแนกสูงข้ออื่นๆ ไม่ให้มีอัตราการแสดงสูงเกินเกณฑ์กำหนด

การศึกษาประสิทธิภาพในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับขั้นของค่าอำนาจจำแนก มีความเกี่ยวข้องกับการวางเงื่อนไขต่างๆ ในการทดสอบ เงื่อนไขสำคัญควรศึกษา ได้แก่ การควบคุมการแสดงผลข้อสอบ ขนาดคลังข้อสอบ และความยาวแบบทดสอบ ด้วยเหตุผลดังนี้

การควบคุมการแสดงข้อสอบ ถึงแม้ไม่มีผลการวิจัยสรุปชัดเจนแต่มีความเป็นไปได้ว่า การควบคุมการแสดงข้อสอบอาจลดประสิทธิภาพการประมาณค่าความสามารถของผู้สอบ เมื่อข้อสอบความยากใกล้เคียงค่าประมาณความสามารถของผู้สอบมากที่สุดไม่ได้จัดให้ เนื่องจากข้อสอบข้อนั้นมีอัตราการแสดงข้อสอบสูงเกินเกณฑ์กำหนด ทำให้แบบทดสอบต้องเลือกข้อสอบความยากอันดับรองลงมาหรือใกล้เคียงจัดให้แทน การลดประสิทธิภาพดังกล่าวอาจไม่เกิดขึ้นถ้าคลังข้อสอบมีขนาดใหญ่พอและความยาวแบบทดสอบสั้นลง นอกจากนี้ การกำหนดอัตราการแสดงข้อสอบมีผลต่อจำนวนข้อสอบในคลังข้อสอบ ถ้ากำหนดอัตราการแสดงข้อสอบต่ำจะต้องใช้ข้อสอบในคลังข้อสอบจำนวนเพิ่มขึ้น (Eggen, 2001 : Online)

ขนาดคลังข้อสอบ คลังข้อสอบขนาดใหญ่ช่วยให้แบบทดสอบมีโอกาสเลือกข้อสอบที่เหมาะสมได้ดีกว่าคลังข้อสอบขนาดเล็ก ให้สารสนเทศแบบทดสอบเพิ่มขึ้น และเพิ่มประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบ แต่การสร้างคลังข้อสอบขนาดใหญ่ต้องใช้ข้อสอบจำนวนมาก ข้อสอบอาจมีจำนวนไม่เพียงพอที่จะครอบคลุมทุกระดับความสามารถของประชากรผู้สอบ หรือทุกระดับชั้นของค่าอำนาจจำแนก รวมทั้งการสร้างข้อสอบคุณภาพดีทำได้ยากโดยเฉพาะข้อสอบอำนาจจำแนกสูง เสียเวลามาก ค่าใช้จ่ายในการสร้างสูง และการบำรุงรักษาค้างข้อสอบทำได้ยาก

ความยาวแบบทดสอบ ความคลาดเคลื่อนของการประมาณค่าความสามารถจะค่อยๆ ลดลงเมื่อความยาวแบบทดสอบค่อยๆ เพิ่มขึ้น จนได้ค่าประมาณความสามารถเข้าใกล้ค่าความสามารถจริงของผู้สอบมากที่สุด แต่ถ้าความยาวแบบทดสอบมากเกินไปพบว่า การแสดงข้อสอบจะสูงเกินอัตราการแสดงข้อสอบกำหนด (Leung; Chang; & Hau, 2003 : Online) และอาจให้สารสนเทศแบบทดสอบมีลักษณะเป็นโค้งแบน (Thissen, 1990 : 169) ปัญหานี้สามารถแก้ไขได้โดยสร้างคลังข้อสอบให้มีขนาดใหญ่เพียงพอ

เห็นได้ว่าเงื่อนไข การควบคุมการแสดงข้อสอบ ขนาดคลังข้อสอบ ความยาวแบบทดสอบ ให้ผลสัมพันธ์ซึ่งกันและกัน การจัดเงื่อนไขต่างๆ ให้เหมาะสมกับลักษณะวิธีกำหนด (Algorithm) สำหรับการทดสอบ จะรักษาประสิทธิภาพสูงสุดในการประมาณค่าความสามารถของผู้สอบและสมคุณการใช้ข้อสอบในคลังข้อสอบของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ได้

ด้วยเหตุผลข้างต้น นำไปสู่ประเด็นสำคัญสนับสนุนการวิจัยประสิทธิภาพของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกภายใต้เงื่อนไขต่างกัน โดยใช้วิธีจำลองข้อมูล

ความมุ่งหมายของการวิจัย

การวิจัยครั้งนี้มีจุดมุ่งหมายหลักเพื่อ ศึกษาประสิทธิภาพการประมาณค่าความสามารถของผู้สอบและประสิทธิภาพสมมูลการใช้ข้อสอบในคลังข้อสอบของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับขั้นของค่าอำนาจจำแนกภายใต้เงื่อนไขต่างกัน โดยใช้วิธีจำลองข้อมูล

โดยมีวัตถุประสงค์เฉพาะ ดังนี้

1. เพื่อศึกษาประสิทธิภาพการประมาณค่าความสามารถของผู้สอบของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับขั้นของค่าอำนาจจำแนก เมื่อใช้ จำนวนข้อสอบในแต่ละชั้นคลังข้อสอบ การควบคุมการแสดงผลข้อสอบ ขนาดคลังข้อสอบ ความยาวแบบทดสอบ ต่างกัน โดยใช้วิธีจำลองข้อมูล
2. เพื่อศึกษาประสิทธิภาพสมมูลการใช้ข้อสอบในคลังข้อสอบของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับขั้นของค่าอำนาจจำแนก เมื่อใช้ จำนวนข้อสอบในแต่ละชั้นคลังข้อสอบการควบคุมการแสดงผลข้อสอบ ขนาดคลังข้อสอบ ความยาวแบบทดสอบ ต่างกัน โดยใช้วิธีจำลองข้อมูล

ความสำคัญของการวิจัย

การศึกษาประสิทธิภาพการประมาณค่าความสามารถของผู้สอบ และประสิทธิภาพสมมูลการใช้ข้อสอบในคลังข้อสอบของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับขั้นของค่าอำนาจจำแนก คาดว่า วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับขั้นของค่าอำนาจจำแนกเมื่อใช้ จำนวนข้อสอบในแต่ละชั้นคลังข้อสอบ การควบคุมการแสดงผลข้อสอบ ขนาดคลังข้อสอบ และความยาวแบบทดสอบ ในระดับที่เหมาะสม สามารถเพิ่มประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบ และประสิทธิภาพสมมูลการใช้ข้อสอบในคลังข้อสอบให้มากยิ่งขึ้น ผลการศึกษาที่ได้นำไปใช้เป็นพื้นฐานสำหรับการประยุกต์ลักษณะคลังข้อสอบและวิธีกำหนดที่เหมาะสม ขณะเดียวกันยังคงประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบและสมมูลการใช้ข้อสอบในคลังข้อสอบไว้ได้ ได้แก่ การกำหนดจำนวนข้อสอบในชั้นของคลังข้อสอบตามค่าอำนาจจำแนก การจัดคลังข้อสอบให้มีขนาดเล็กลง ช่วยในการลดค่าใช้จ่ายในการสร้างข้อสอบและการบำรุงรักษาคลังข้อสอบทำได้ง่าย ความยาวแบบทดสอบอาจสามารถปรับให้มีขนาดสั้นลง ทำให้ผู้สอบลดเวลาทำการทดสอบและประหยัดข้อสอบในการทดสอบ การรักษาความปลอดภัยของข้อสอบยังคงอยู่ด้วยคุณสมบัติของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับขั้นของค่าอำนาจจำแนก ความถี่การแสดงผลข้อสอบถูกควบคุมไม่ให้สูงเกินเกณฑ์กำหนด สามารถคาดการณ์ วินิจฉัย คัดเลือก ลักษณะคลังข้อสอบและวิธีกำหนดรูปแบบต่างๆ โดยยังรักษาประสิทธิภาพสูงสุดในการประมาณค่าความสามารถของผู้สอบและสมมูลการใช้ข้อสอบในคลังข้อสอบ เพื่อใช้ในการจัดการทดสอบแบบ

ปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกต่อไป ข้อสังเกตที่ได้รับระหว่างการทดสอบจะนำไปใช้ศึกษาเพื่อเชื่อมโยงความรู้เข้าด้วยกัน สร้างเป็นชุดความรู้ใหม่สำหรับการเลือกข้อสอบวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก ภายใต้เงื่อนไขจำนวนข้อสอบในแต่ละชั้นคลังข้อสอบ การควบคุมการแสดงข้อสอบ ขนาดคลังข้อสอบ ความยาวแบบทดสอบ ต่างกัน สามารถนำไปประยุกต์ใช้ในทางปฏิบัติสถานการณ์จริง และการศึกษาวิจัยนี้ สามารถใช้เป็นแนวทางในการศึกษาการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกภายใต้เงื่อนไขต่างกัน ได้แก่ จำนวนข้อสอบในแต่ละชั้นคลังข้อสอบ ขนาดคลังข้อสอบ ความยาวแบบทดสอบ และการควบคุมการแสดงข้อสอบ สำหรับงานวิจัยที่จะเกิดขึ้นในอนาคตต่อไป

ขอบเขตของการวิจัย

การวิจัยครั้งนี้ศึกษาโดยใช้วิธีการจำลองข้อมูล อาศัยแนวทางการทดสอบตามหลักการทฤษฎีการตอบข้อสอบ (Item Response Theory : IRT) ใช้รูปแบบโลจิสติก 3 พารามิเตอร์ (3 Parameter Logistic Model : 3 PLM) ผลการตอบข้อสอบมี 2 ค่า คือ ตอบถูกต้องได้คะแนนเท่ากับ 1 ตอบผิดได้คะแนนเท่ากับ 0 (Dichotomous Scoring) กำหนดให้คลังข้อสอบมีความเป็นมิติเดียว (Unidimension) สำหรับ กลุ่มตัวอย่าง พารามิเตอร์ของข้อสอบ และตัวแปรศึกษา กำหนดไว้ดังนี้

กลุ่มตัวอย่าง

ค่าความสามารถจริง (θ) ของกลุ่มตัวอย่างผู้สอบ สร้างขึ้นโดยสุ่มจากการแจกแจงโค้งปกติมาตรฐาน มีค่าเฉลี่ยเท่ากับ 0 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1 [$\theta \sim N(0,1)$] การแจกแจงค่าความสามารถจริงมีค่าตั้งแต่ -3.0 ถึง 3.0 จำนวนค่าความสามารถจริงของกลุ่มตัวอย่างผู้สอบในการทดสอบแต่ละวิธีกำหนดเท่ากับ 5,000 ค่า เมื่อเปลี่ยนวิธีกำหนดจะทำการสุ่มค่าความสามารถจริงของกลุ่มตัวอย่างผู้สอบใหม่ทุกครั้ง

พารามิเตอร์ข้อสอบ

พารามิเตอร์ข้อสอบ สร้างขึ้นโดยกำหนดค่าการเดา (c) คงที่เท่ากับ 0 ค่าอำนาจจำแนก (a) 4 ค่า คือ 0.5 1.0 1.5 และ 2.0 ตามจำนวนการแบ่งชั้นคลังข้อสอบ 4 ชั้น โดยให้ข้อสอบที่บรรจุในชั้นคลังข้อสอบเดียวกันมีค่าอำนาจจำแนกเท่ากัน ค่าความยาก (b) สร้างโดยสุ่มจากการแจกแจงโค้งปกติมาตรฐานมีค่าเฉลี่ยเท่ากับ 0 และมีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1 [$b \sim N(0,1)$] บนมาตรฐานวัดเดียวกันกับเส้นต่อเนื่องค่าความสามารถจริงของกลุ่มตัวอย่างผู้สอบ การแจกแจงค่าความยากมีค่าตั้งแต่ -3.0 ถึง 3.0

ตัวแปรศึกษา

1. ตัวแปรต้น มี 4 ตัวแปร ดังนี้

1.1 การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก

(a-Stratified Computerized Adaptive Testing : a-STR) 4 วิธี ได้แก่

วิธีที่ 1 หรือวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก กำหนดอัตราส่วนจำนวนข้อสอบเท่ากันทุกชั้น (1 : 1 : 1 : 1)

(a-STR1)

วิธีที่ 2 หรือวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก กำหนดอัตราส่วนจำนวนข้อสอบชั้นแรกมากกว่าชั้นอื่น

(2 : 0.5 : 0.5 : 1) (a-STR2)

วิธีที่ 3 หรือวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก กำหนดอัตราส่วนจำนวนข้อสอบชั้นสุดท้ายมากกว่าชั้นอื่น

(1 : 0.5 : 0.5 : 2) (a-STR3)

วิธีที่ 4 หรือวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก กำหนดอัตราส่วนจำนวนข้อสอบชั้นแรกและชั้นสุดท้ายมากกว่าชั้นอื่น

(1.5 : 0.5 : 0.5 : 1.5) (a-STR4)

1.2 วิธีควบคุมการแสดงข้อสอบ 2 ระดับ ได้แก่ ไม่ควบคุมการแสดงข้อสอบ และ ควบคุมการแสดงข้อสอบด้วยวิธีซิมสัน-สเตเทอร์ หรือ วิธีขยายช่วงระดับชั้น

(Enhanced a-Stratified Method)

1.3 ขนาดคลังข้อสอบ 3 ระดับ ได้แก่ ขนาดคลังข้อสอบ 200 400 และ 600 ข้อ

1.4 ความยาวแบบทดสอบ 3 ระดับ ได้แก่ ความยาวแบบทดสอบ 20 32 และ 44 ข้อ

2. ตัวแปรตาม มี 2 ตัวแปร ดังนี้

2.1 ประสิทธิภาพการประมาณค่าความสามารถของผู้สอบ พิจารณาจาก

ความคลาดเคลื่อนกำลังสองเฉลี่ย

ความลำเอียงเฉลี่ย

2.2 ประสิทธิภาพสมมูลการใช้ข้อสอบในคลังข้อสอบ พิจารณาจาก

ข้อสอบที่มีการแสดงมากเกินไป

ข้อสอบที่มีการใช้ต่ำ

อัตราการทับซ้อนข้อสอบ

การแจกแจงอัตราการแสดงข้อสอบ

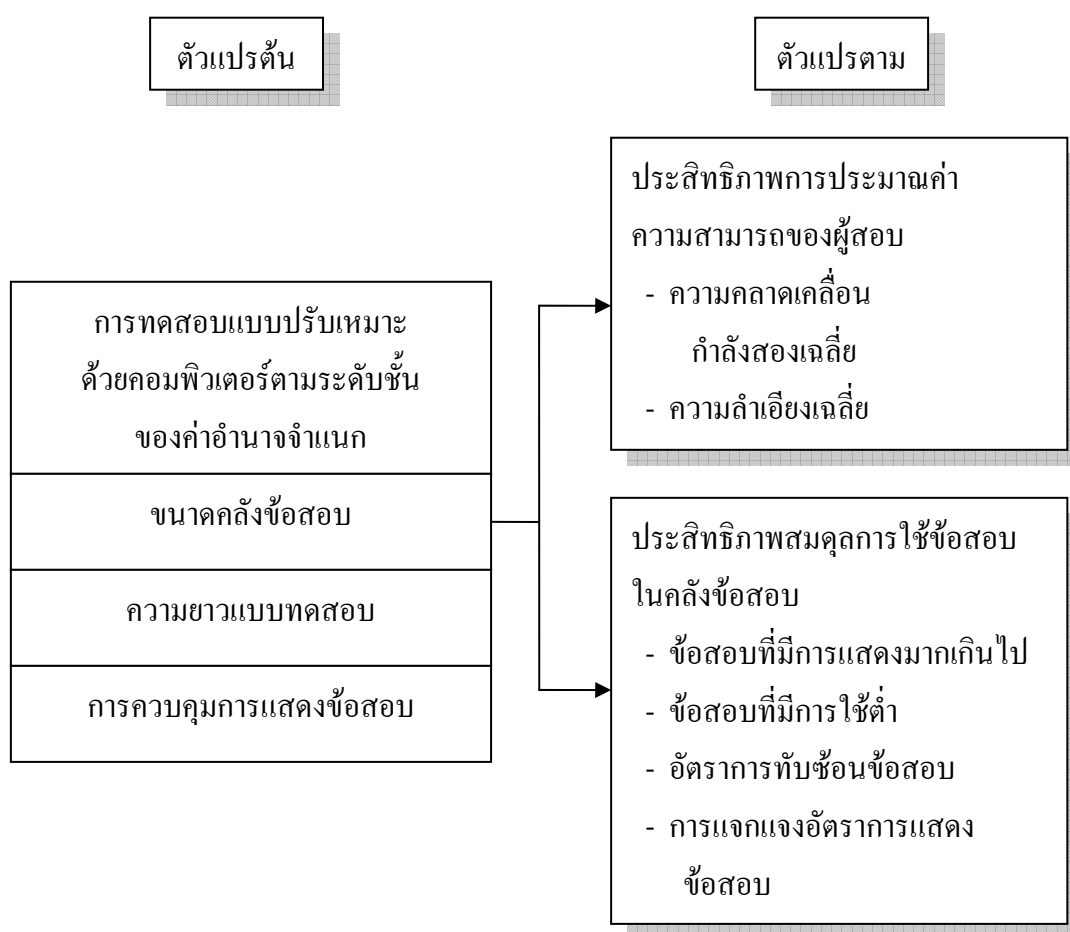
กรอบแนวคิดในการวิจัย

การศึกษาประสิทธิภาพการประมาณค่าความสามารถของผู้สอบ และประสิทธิภาพสมมูลการใช้ข้อสอบในคลังข้อสอบของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก (a-Stratified Computerized Adaptive Testing : a-STR) ตัวแปรที่สนใจได้แก่ การจัดจำนวนข้อสอบในแต่ละชั้นคลังข้อสอบเท่ากัน (a-STR1) และไม่เท่ากัน (a-STR2 a-STR3 และ a-STR4) วิธีการจัดจำนวนข้อสอบในแต่ละชั้นคลังข้อสอบเท่ากันเป็นวิธีที่มีลักษณะเทียบได้กับวิธีที่เคยศึกษามา (Chang; & Ying, 1999) พบว่า สามารถให้ประสิทธิภาพการประมาณค่าความสามารถของผู้สอบ โดยให้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยและค่าความลำเอียงเฉลี่ยต่ำ ขณะเดียวกันให้ประสิทธิภาพสมมูลการใช้ข้อสอบในคลังข้อสอบ โดยกระจายการใช้ข้อสอบไปยังข้อสอบข้ออื่นๆ ทำให้ข้อสอบมีอัตราการแสดงข้อสอบไม่เกินเกณฑ์กำหนด ลดความเบ้ของโค้งการแจกแจงอัตราการแสดงข้อสอบ และอัตราการทับซ้อนต่ำ ประสิทธิภาพการประมาณค่าความสามารถของผู้สอบและประสิทธิภาพสมมูลการใช้ข้อสอบในคลังข้อสอบจะเพิ่มขึ้นเมื่อเงื่อนไขข้ออื่นๆ เหมาะสม ได้แก่ คลังข้อสอบขนาดใหญ่และความยาวแบบทดสอบมากเพียงพอ แต่จำเป็นต้องระวังถ้าใช้คลังข้อสอบขนาดใหญ่เกินไป จะเป็นการสิ้นเปลืองโดยไม่จำเป็นและยากในการบำรุงรักษา และถ้าความยาวแบบทดสอบมากเกินไป อาจให้การแสดงข้อสอบสูงเกินอัตราการแสดงข้อสอบกำหนด ลดประสิทธิภาพสมมูลการใช้ข้อสอบในคลังข้อสอบได้ การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก กำหนดอัตราส่วนจำนวนข้อสอบในแต่ละชั้นคลังข้อสอบเท่ากัน (a-STR1) เมื่อใช้ร่วมกับการควบคุมการแสดงผลข้อสอบจะให้สมมูลการใช้ข้อสอบได้ดีกว่าขณะที่ไม่ควบคุมการแสดงผลข้อสอบ แต่ต้องติดตามการลดประสิทธิภาพการประมาณค่าความสามารถของผู้สอบ ซึ่งอาจเกิดจากแบบทดสอบจัดข้อสอบความยากใกล้เคียงรองลงมาจัดให้แทน เนื่องจากข้อสอบความยากใกล้เคียงค่าประมาณความสามารถของผู้สอบมากที่สุดมีอัตราการแสดงข้อสอบสูงเกินเกณฑ์กำหนด ผลการศึกษาการทดสอบแบบปรับเหมาะ ด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก วิธีจัดจำนวนข้อสอบในแต่ละชั้นคลังข้อสอบเท่ากัน (a-STR1) เป็นไปได้ว่าผลนี้สามารถเกิดได้กับวิธีจัดจำนวนข้อสอบในแต่ละชั้นคลังข้อสอบไม่เท่ากัน (a-STR2 a-STR3 และ a-STR4) ที่สนใจศึกษา แต่ผลอาจจะแตกต่างกันไปตามลักษณะการจัดวางจำนวนข้อสอบในแต่ละชั้นคลังข้อสอบของแต่ละวิธี

ดังนั้น แนวทางการศึกษาให้ความสนใจ ลักษณะการจัดจำนวนข้อสอบในแต่ละชั้นคลังข้อสอบ ขนาดคลังข้อสอบ ความยาวแบบทดสอบ และการควบคุมการแสดงผลข้อสอบ สามารถส่งผลกระทบซึ่งกันและกัน เมื่อลักษณะเงื่อนไขข้อใดอย่างหนึ่งเปลี่ยนแปลง ย่อมส่งผลกระทบต่อประสิทธิภาพการประมาณค่าความสามารถของผู้สอบ โดยพิจารณาได้จาก ความคลาดเคลื่อนกำลังสองเฉลี่ยและความลำเอียงเฉลี่ย และประสิทธิภาพสมมูลการใช้ข้อสอบในคลังข้อสอบ โดยพิจารณาได้จาก ข้อสอบที่มีการแสดงมากเกินไป ข้อสอบที่มีการใช้ต่ำ อัตราการทับซ้อนข้อสอบ และการ

แจกแจงอัตราการแสดงข้อสอบ การวิจัยนี้มีจุดมุ่งหมายที่จะศึกษาลักษณะการจัดจำนวนข้อสอบในแต่ละชั้นคลังข้อสอบ ขนาดคลังข้อสอบ ความยาวแบบทดสอบ และการควบคุมการแสดงข้อสอบว่าควรมีลักษณะอย่างไร จึงจะให้ประสิทธิภาพการประมาณค่าความสามารถของผู้สอบและประสิทธิภาพสมมูลการใช้ข้อสอบในคลังข้อสอบของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก

แสดงเป็นกรอบแนวคิดการวิจัยดังภาพประกอบ 1 ดังนี้



ภาพประกอบ 1 กรอบแนวคิดการวิจัย

ภาพประกอบ 1 แสดงกรอบแนวคิดในการวิจัยโดยสรุป อธิบายเชื่อมโยงไปสู่แนวทางวิจัยได้ว่า ค่าความสามารถจริงของกลุ่มตัวอย่างผู้สอบ 5,000 ค่า เมื่อนำไปผ่านขั้นตอนการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกตามวิธีกำหนด ได้แก่ วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก (a-STR) 4 วิธี แตกต่างกันที่บรรจุข้อสอบจำนวนไม่เท่ากันในแต่ละชั้นของคลังข้อสอบ การควบคุมการแสดงข้อสอบ 2 ระดับ ได้แก่ ไม่ควบคุมและควบคุมการแสดงข้อสอบ ขนาดคลังข้อสอบ 3 ระดับ ได้แก่ 200 400 และ 600 ข้อ และความยาวแบบทดสอบ 3 ระดับ ได้แก่ 20 32 และ 44 ข้อ จะได้วิธีกำหนดทั้งสิ้น $4 \times 2 \times 3 \times 3$ เท่ากับ 72 วิธี เก็บข้อมูลเพื่อศึกษาประสิทธิภาพการประมาณค่าความสามารถของผู้สอบโดยพิจารณาจากความคลาดเคลื่อนกำลังสองเฉลี่ย ความลำเอียงเฉลี่ย และศึกษาประสิทธิภาพสมมูลการใช้ข้อสอบในคลังข้อสอบ โดยพิจารณาจาก ข้อสอบที่มีการแสดงมากเกินไป ข้อสอบที่มีการใช้ต่ำ อัตราการทับซ้อนข้อสอบ และการแจกแจงอัตราการแสดงข้อสอบ

นิยามศัพท์เฉพาะ

1. การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ (Computerized Adaptive Testing : CAT) หมายถึง การทดสอบโดยใช้ข้อสอบแตกต่างกันจัดให้กับผู้สอบแต่ละคน พิจารณาเลือกข้อสอบจากความสามารถภายในของผู้สอบที่ต้องการวัด และดำเนินการทดสอบด้วยเครื่องคอมพิวเตอร์

2. การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก (a-Stratified Computerized Adaptive Testing : a-STR) หมายถึง การทดสอบโดยใช้ข้อสอบแตกต่างกันจัดให้กับผู้สอบแต่ละคน พิจารณาเลือกข้อสอบตามความสามารถภายในของผู้สอบที่ต้องการวัด และดำเนินการทดสอบตามระดับชั้นของค่าอำนาจจำแนกของคลังข้อสอบด้วยเครื่องคอมพิวเตอร์ แบ่งเป็น 4 วิธี

2.1 วิธีที่ 1 หรือ a-STR1 หมายถึง การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก กำหนดอัตราส่วนจำนวนข้อสอบในแต่ละชั้นคลังข้อสอบเท่ากัน (1:1:1:1)

2.2 วิธีที่ 2 หรือ a-STR2 หมายถึง การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก กำหนดอัตราส่วนจำนวนข้อสอบชั้นแรกมากกว่าชั้นอื่น (2:0.5:0.5:1)

2.3 วิธีที่ 3 หรือ a-STR3 หมายถึง การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก กำหนดอัตราส่วนจำนวนข้อสอบชั้นสุดท้ายมากกว่าชั้นอื่น (1:0.5:0.5:2)

2.4 วิธีที่ 4 หรือ a-STR4 หมายถึง การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก กำหนดอัตราส่วนจำนวนข้อสอบชั้นแรกและชั้นสุดท้ายมากกว่าชั้นอื่น (1.5:0.5:0.5:1.5)

3. **ขนาดคลังข้อสอบ (Item Pool Size)** หมายถึง จำนวนข้อสอบทั้งหมดบรรจุในคลังข้อสอบซึ่งเป็นแหล่งรวบรวมข้อสอบหรือคำถามแบบทดสอบ

4. **ความยาวแบบทดสอบ (Test Length)** หมายถึง จำนวนข้อสอบในแบบทดสอบที่ต้องการจัดให้แก่ผู้สอบ

5. **กฎการยุติ (Stopping Rule)** หมายถึง ข้อกำหนดเพื่อใช้ในการสิ้นสุดการทดสอบ

6. **การเลือกข้อสอบ (Item Selection)** หมายถึง ขั้นตอนการทดสอบเพื่อค้นหาข้อสอบที่มีค่าความยากเท่ากับหรือใกล้เคียงค่าประมาณความสามารถของผู้สอบครั้งสุดท้าย

7. **การแสดงผลข้อสอบ (Item Exposure)** หมายถึง การจัดข้อสอบที่ผ่านการเลือกแล้วให้แก่ผู้สอบ

8. **อัตราการแสดงผลข้อสอบ (Item Exposure Rate)** หมายถึง จำนวนครั้งของข้อสอบที่จัดให้แก่ผู้สอบต่อจำนวนผู้สอบทั้งหมดที่ทำการทดสอบ

9. **การควบคุมการแสดงผลข้อสอบ (Item Exposure Control)** หมายถึง การควบคุมจำนวนครั้งของข้อสอบที่จัดให้แก่ผู้สอบตามอัตราการแสดงผลข้อสอบกำหนด

10. **ประสิทธิภาพการประมาณค่าความสามารถของผู้สอบ (Estimative Efficiency)** หมายถึง ความสามารถของวิธีกำหนดในการประมาณค่าความสามารถของผู้สอบให้เท่ากับหรือใกล้เคียงค่าความสามารถจริงของผู้สอบมากที่สุด พิจารณาจาก ความคลาดเคลื่อนกำลังสองเฉลี่ย และความลำเอียงเฉลี่ย

10.1 **ความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Square Error : MSE)** หมายถึง ค่าบอกความถูกต้องของค่าประมาณความสามารถของผู้สอบ แสดงเป็นผลต่างกำลังสองเฉลี่ยระหว่างค่าประมาณความสามารถและค่าความสามารถจริงของผู้สอบ หรือมีค่าเท่ากับความแปรปรวนของค่าประมาณความสามารถรวมกับความลำเอียงกำลังสอง

10.2 **ความลำเอียงเฉลี่ย (Average Bias)** หมายถึง ค่าบอกความเที่ยงตรงของค่าประมาณความสามารถของผู้สอบ แสดงเป็นผลต่างเฉลี่ยระหว่างค่าประมาณความสามารถและค่าความสามารถจริงของผู้สอบ บอกทิศทางการประมาณค่าว่าให้ผลสูงหรือต่ำกว่าค่าความสามารถจริงของผู้สอบ ถ้าความลำเอียงกำลังสองเท่ากับ 0 ความคลาดเคลื่อนกำลังสองเฉลี่ยจะเท่ากับความแปรปรวนของค่าประมาณความสามารถของผู้สอบ

11. **ประสิทธิภาพสมมูลการใช้ข้อสอบ (Item Utilizable Efficiency)** หมายถึง ความสามารถของวิธีกำหนดที่ทำให้ข้อสอบทุกข้อในคลังข้อสอบมีโอกาสนำไปใช้ในการทดสอบ พิจารณาจากข้อสอบที่มีการแสดงมากเกินไป ข้อสอบที่นำไปใช้ต่ำ อัตราการทับซ้อนข้อสอบ และการแจกแจงอัตราการแสดงผลข้อสอบ

11.1 ข้อสอบที่มีการแสดงมากเกินไป (Overexposed Item) หมายถึง ข้อสอบใดๆ ที่มีอัตราการแสดงข้อสอบสูงกว่า 0.2

11.2 ข้อสอบที่มีการใช้ต่ำ (Underutilized Item) หมายถึง ข้อสอบใดๆ ที่มีอัตราการแสดงข้อสอบเท่ากับหรือน้อยกว่า 0.2

11.3 อัตราการทับซ้อนข้อสอบ (Item Overlap Rate) หมายถึง สัดส่วนของข้อสอบที่ใช้ร่วมกันโดยคู่ของผู้สอบต่อข้อสอบที่ใช้โดยคู่ของผู้สอบทั้งหมด

11.4 การแจกแจงอัตราการแสดงข้อสอบ (Item Exposure Rate Distribution) หมายถึง ลักษณะการกระจายอัตราการแสดงของข้อสอบในคลังข้อสอบ บ่งบอกถึงประสิทธิภาพการใช้ข้อสอบในคลังข้อสอบโดยรวม พิจารณาจากการประยุกต์ใช้ค่าไคสแควร์

สมมติฐานในการวิจัย

1. เงื่อนไขการทดสอบ ได้แก่ จำนวนข้อสอบในแต่ละชั้นคลังข้อสอบ การควบคุมการแสดงผลข้อสอบ ขนาดคลังข้อสอบ ความยาวแบบทดสอบ ที่เหมาะสม สามารถให้ประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก

2. เงื่อนไขการทดสอบ ได้แก่ จำนวนข้อสอบในแต่ละชั้นคลังข้อสอบ การควบคุมการแสดงผลข้อสอบ ขนาดคลังข้อสอบ ความยาวแบบทดสอบ ที่เหมาะสม สามารถให้ประสิทธิภาพสมดุลการใช้ข้อสอบในคลังข้อสอบของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก

บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

การวิจัยครั้งนี้มีความมุ่งหมายเพื่อศึกษาประสิทธิภาพการประมาณค่าความสามารถของผู้สอบ และประสิทธิภาพสมมูลการใช้ข้อสอบในคลังข้อสอบของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ตามระดับชั้นของค่าอำนาจจำแนก เมื่อใช้ จำนวนข้อสอบในแต่ละชั้นคลังข้อสอบ การควบคุมการ แสดงข้อสอบ ขนาดคลังข้อสอบ ความยาวแบบทดสอบ ต่างกัน โดยใช้วิธีจำลองข้อมูล ผู้วิจัยได้ศึกษา เอกสารกรอบแนวคิด ทฤษฎี หลักการ และงานวิจัยที่เกี่ยวข้อง เพื่อนำองค์ความรู้ที่ได้รับจากการศึกษา มาประยุกต์ใช้ในการวิจัยครั้งนี้ นำเสนอเป็นกรอบแนวทางการศึกษา 6 ประเด็นดังนี้

1. ทฤษฎีการตอบข้อสอบ
 - 1.1 แนวคิดทฤษฎีการตอบข้อสอบ
 - 1.2 ข้อตกลงเบื้องต้นของทฤษฎีการตอบข้อสอบ
 - 1.3 รูปแบบการตอบข้อสอบ
 - 1.4 สารสนเทศข้อสอบและแบบทดสอบ
 - 1.5 การประมาณค่าความสามารถและพารามิเตอร์ข้อสอบ
2. การทดสอบแบบปรับเหมาะ
 - 2.1 ความเป็นมาของการทดสอบแบบปรับเหมาะ
 - 2.2 ความแตกต่างระหว่างแบบทดสอบแบบปรับเหมาะและแบบทดสอบ ประเพณีนิยม
 - 2.3 กลยุทธ์ของการทดสอบแบบปรับเหมาะ
3. ความก้าวหน้าของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์
 - 3.1 การกำหนดรูปแบบการตอบข้อสอบ
 - 3.2 คลังข้อสอบ
 - 3.3 ความยาวแบบทดสอบ
 - 3.4 จุดเริ่มต้นการทดสอบ
 - 3.5 การเลือกข้อสอบ
 - 3.6 การประมาณค่าความสามารถ
 - 3.7 การแสดงข้อสอบ
 - 3.8 การจัดสมมูลเนื้อหา

3.9 คุณภาพแบบทดสอบ

3.10 กฎการยุติ

4. การศึกษาในสถานการณ์จำลอง

5. การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของคำอ่านจำแนก

6. งานวิจัยที่เกี่ยวข้อง

1. ทฤษฎีการตอบข้อสอบ (Item Response Theory : IRT)

แบบทดสอบตามหลักการจิตวิทยา มุ่งวัดลักษณะภายในตัวบุคคล เช่น ความสามารถทางด้านเชาวน์ปัญญา เจตคติ ค่านิยม เป็นต้น การหาตัวอย่างพฤติกรรมบ่งบอกลักษณะภายในว่ามีมากน้อยเพียงใดสามารถทำได้ แต่จะไม่มีพฤติกรรมชัดเจนครอบคลุมลักษณะภายในที่มุ่งศึกษานั้นทั้งหมดจึงเป็นเหตุผลให้ลักษณะภายในแต่ละลักษณะเรียกอีกชื่อหนึ่งว่า ลักษณะแฝง (Latent Trait) การจัดการทดสอบโดยเสนอสิ่งเร้าซึ่งหมายถึงข้อสอบหรือแบบทดสอบให้กับผู้สอบ เพื่อต้องการให้ผู้สอบแสดงพฤติกรรมการตอบแบบทดสอบ ผลการตอบแบบทดสอบจะชี้บอกตำแหน่งผู้สอบบนเส้นต่อเนื่องของลักษณะภายใน และใช้วิธีการทางสถิติช่วยพยากรณ์ลักษณะภายในของผู้สอบนั้น (Fischer; & Molenaar. 1995 : 3-14) ; (Lord; & Novick. 1968 : 358-359)

หลักการทดสอบตามทฤษฎีการตอบข้อสอบ มุ่งประมาณลักษณะภายใน (Trait) ของผู้สอบ โดยอาศัยการตอบข้อสอบและคุณสมบัติของข้อสอบ ความน่าจะเป็นในการตอบข้อสอบนำไปอธิบายตำแหน่งผู้สอบบนเส้นต่อเนื่องของลักษณะภายในแสดงเป็นฟังก์ชันคณิตศาสตร์ เรียกว่า โคง์ลักษณะข้อสอบ (Item Characteristic Curve : ICC) หรือ ฟังก์ชันการตอบข้อสอบ (Item Response Function : IRF) มีผู้เสนอแนวคิดเกี่ยวกับทฤษฎีการตอบข้อสอบในระยะแรกๆ และเริ่มปรากฏชัดเจนขึ้นโดย ลอร์ดเลย์ในปี ค.ศ. 1943-1944 ซึ่งสร้างวิธีการประมาณค่าพารามิเตอร์แต่ยังพบข้อจำกัดบางประการ ปี ค.ศ. 1952 ลอร์ดอธิบายรูปแบบนอร์มัลโอโจส 2 พารามิเตอร์และรูปแบบประยุกต์อื่นๆ ปี ค.ศ. 1968 ลอร์ดและโนวิก (Lord; & Novick. 1968 : 358-387) แนะนำทฤษฎีการตอบข้อสอบ (Item Response Theory : IRT) หรือรู้จักอีกชื่อหนึ่งคือ ทฤษฎีลักษณะแฝง (Latent Trait Theory) ในหนังสือชื่อ " Statistic Theory of Mental Test Score " และเอกสารต่างๆ มากมาย นับเป็นพื้นฐานการวัดทางจิตวิทยาแนวใหม่ แต่ยังติดขัดด้านการใช้สถิติทดสอบและขาดโปรแกรมคอมพิวเตอร์ที่มีประสิทธิภาพเพียงพอในการคำนวณ ทำให้การพัฒนาในระยะนั้นเป็นไปอย่างเชื่องช้า การพัฒนาอีกทางหนึ่งในปี ค.ศ. 1960 ราสช์ (Wainer; & Mislevy. 1990 ; citing Rasch. 1901-1980) พัฒนารูปแบบของ IRT โดยใช้พารามิเตอร์เดียวคือ พารามิเตอร์ความยาก (b) เรียกูปแบบนี้ว่า ราสช์โมเดล (Rasch Model) ความน่าจะเป็นของการตอบข้อสอบในรูปพารามิเตอร์เดียวนี้นี้จะหมายถึง ความน่าจะเป็น

เป็นสำหรับผู้สอบระดับความสามารถที่ 0 ตอบข้อสอบค่าความยากที่ 6 ได้ถูกต้อง รูปแบบนี้เมื่อค่าความยากของข้อสอบแตกต่างกัน โ้คงลักษณะข้อสอบมีลักษณะคู่ขนานกัน ต่อมาเบิร์นบอมได้พัฒนาฟังก์ชันนอร์มัลโอโจส์รูปแบบ 2 พารามิเตอร์ คือ มีพารามิเตอร์ค่าความยากและค่าอำนาจจำแนกซึ่งลอร์ดพัฒนาไว้สร้างเป็นฟังก์ชันโลจิสติก นำไปสู่การพัฒนาในรูปแบบอื่นๆ เช่น รูปแบบโลจิสติก 3 และ 4 พารามิเตอร์ ความก้าวหน้าของทฤษฎีการตอบข้อสอบช่วยให้นักวิจัยประเมินคุณภาพแบบทดสอบว่า ข้อสอบนั้นมีความเหมาะสมเพียงไร และใช้เป็นเครื่องมือมาตรฐานสำหรับวัดบุคคลได้อย่างมีประสิทธิภาพ (Fischer; & Molenaar. 1995 : 3-14) ; (Embretson; & Reise. 2000 : 3-9) ; (Baker. 1992 : 1-23)

1.1 แนวคิดของทฤษฎีการตอบข้อสอบ

ทฤษฎีการตอบข้อสอบตั้งอยู่บนแนวคิดสำคัญ 2 ประการ ได้แก่ 1. พฤติกรรมในการตอบข้อสอบของผู้สอบนำไปใช้ทำนายความสามารถ (Ability) หรือลักษณะภายใน (Trait) ของผู้สอบ และ 2. พฤติกรรมการตอบข้อสอบของผู้สอบสัมพันธ์กับความสามารถ (Ability) หรือลักษณะภายใน (Trait) ของผู้สอบ (Hambleton; Swaminathan; & Rogers. 1991 : 7)

แนวคิดนี้อยู่บนพื้นฐานความเชื่อว่า ลักษณะภายในตัวบุคคลซึ่งสังเกตไม่ได้เป็นตัวกำหนดพฤติกรรมหรือผลการตอบข้อสอบซึ่งสังเกตได้ ความสัมพันธ์ดังกล่าวสามารถอธิบายได้ด้วยฟังก์ชันทางคณิตศาสตร์เรียกว่า ฟังก์ชันการตอบข้อสอบ (Item Response Function) มี 2 ลักษณะคือ ฟังก์ชันโลจิสติก (Logistic Function) และฟังก์ชันปกติสะสม (Normal Ogive Function) ฟังก์ชันการตอบข้อสอบนี้เมื่อแสดงด้วยกราฟจะได้โค้งลักษณะข้อสอบ (Item Characteristic Curve : ICC) ลักษณะข้อสอบประกอบด้วยพารามิเตอร์ข้อสอบต่างๆ ได้แก่ พารามิเตอร์ความยาก (Difficulty Parameter) พารามิเตอร์อำนาจจำแนก (Discrimination Parameter) พารามิเตอร์การเดา (Guessing Parameter) และพารามิเตอร์ความสะเพร่า (Careless Parameter) นำมาใช้กำหนดลักษณะรูปแบบ (Model) ต่างๆ ได้แก่ รูปแบบ 1 พารามิเตอร์ (One-Parameter Model) รูปแบบ 2 พารามิเตอร์ (Two-Parameter Model) รูปแบบ 3 พารามิเตอร์ (Three-Parameter Model) และรูปแบบ 4 พารามิเตอร์ (Four-Parameter Model) (Hambleton; Swaminathan; & Rogers. 1991 : 7-31)

1.2 ข้อตกลงเบื้องต้นของทฤษฎีการตอบข้อสอบ

ข้อตกลงเบื้องต้นตามทฤษฎีการตอบข้อสอบกล่าวถึงความเป็นมิติเดียว ความเป็นอิสระในการตอบข้อสอบและโค้งลักษณะข้อสอบ มีรายละเอียดดังนี้

1.2.1 ความเป็นมิติเดียว (Unidimension) ผลการตอบข้อสอบของผู้สอบสามารถอธิบายความสามารถหรือคุณลักษณะภายในเพียงด้านใดด้านหนึ่งของผู้สอบ คุณลักษณะภายในด้าน

เดียนีมีความหมายเหมือนเป็นมิติเดียว ในสภาพความเป็นจริง ข้อตกลงนี้พบว่า ยังมีคุณลักษณะภายในหลายประการที่ส่งผลร่วมต่อการตอบข้อสอบเข้ามาเกี่ยวข้อง ดังนั้นจึงได้ผ่อนปรนจากความเป็นมิติเดียวให้เป็นลักษณะเด่น (Dominant) หรือลักษณะหลัก ซึ่งสามารถนำไปอธิบายผลการตอบข้อสอบของผู้สอบได้ (Hambleton; Swaminathan; & Rogers. 1991: 9) ; (Green; & et al. 1984)

1.2.2 ความเป็นอิสระในการตอบข้อสอบ (Local Independence) เมื่อค่าความสามารถของผู้สอบเป็นค่าแน่นอน การตอบข้อสอบแต่ละข้อของผู้สอบคนหนึ่งจะมีความเป็นอิสระจากกันหรืออาจกล่าวได้ว่าการตอบข้อสอบข้อใดๆ ของผู้สอบจะไม่มีผลต่อข้อสอบข้ออื่นๆ สิ่งที่ส่งผลต่อการตอบข้อสอบแต่ละข้อเป็นผลมาจากความสามารถของผู้สอบเท่านั้น (Hambleton; Swaminathan; & Rogers. 1991: 10) ความเป็นอิสระในการตอบข้อสอบทำให้พารามิเตอร์ข้อสอบยังเป็นค่าคงที่ ไม่ว่าข้อสอบข้อนั้นอยู่ตำแหน่งใดๆ บนแบบทดสอบ การละเมิดความเป็นอิสระในการตอบข้อสอบเกิดขึ้นได้ เช่น การตอบข้อสอบข้อใดข้อหนึ่งถูกต้องจำเป็นต้องใช้ข้อมูลจากข้อสอบอื่นช่วย หรือกรณีเนื้อหาของข้อสอบหนึ่งมีความสัมพันธ์กับการตอบข้อสอบข้ออื่น หรือมีปฏิสัมพันธ์ระหว่างเนื้อหาข้อสอบ เป็นต้น ถ้ามีการละเมิดข้อตกลงเกิดขึ้น รูปแบบความน่าจะเป็นของการตอบข้อสอบจะเปลี่ยนแปลงไม่คงที่กับรูปแบบของทฤษฎีการตอบข้อสอบที่ได้กำหนดไว้ (Embretson; & Reise. 2000 : 187-189)

1.2.3 โค้งลักษณะข้อสอบ (Item Characteristic Curve) หรือ ฟังก์ชันการตอบข้อสอบ (Item Response Function : IRF) จะเป็นฟังก์ชันทางคณิตศาสตร์แสดงความสัมพันธ์ระหว่างความน่าจะเป็นของการตอบข้อสอบถูกต้องกับระดับความสามารถของผู้สอบ (Hambleton; Swaminathan; & Rogers. 1991: 10)

1.3 รูปแบบการตอบข้อสอบ (Item Response Model)

ในระยะเริ่มแรกของทฤษฎีการตอบข้อสอบ มีการพัฒนารูปแบบนอร์มัลโอจีฟ (Normal Ogive Model) ขึ้นแต่ยังพบข้อจำกัดในการใช้ จนกระทั่งเบิร์นบอมพัฒนารูปแบบโลจิสติก (Logistic Model) แทนรูปแบบนอร์มัลโอจีฟเพื่อให้ส่วนปลายโค้งแอสซิมโทส (Asymptotes) ของรูปแบบโลจิสติกเปลี่ยนแปลงได้น้อยกว่าแอสซิมโทสของรูปแบบนอร์มัลโอจีฟ เมื่อรูปแบบการตอบข้อสอบเป็นรูปสนิท (Fit) กับข้อมูลแล้ว จะให้ความไม่แปรเปลี่ยน (Invariance) ไปตามกลุ่มผู้สอบ มีผลให้การประมาณค่าความสามารถของผู้สอบถูกต้องยิ่งขึ้น (Hambleton; & Swaminathan. 1985 : 1-13)

รูปแบบโลจิสติกเมื่อการตอบข้อสอบประเภทให้ผลสองค่า (Dichotomous Scoring) คือตอบข้อสอบถูกต้องได้ 1 คะแนน หรือตอบข้อสอบผิดได้ 0 คะแนน ได้แก่

รูปแบบโลจิสติก 1 พารามิเตอร์

$$P_i(\theta) = \frac{e^{(\theta - b_i)}}{1 + e^{(\theta - b_i)}} \quad i = 1, 2, \dots, n$$

รูปแบบโลจิสติก 2 พารามิเตอร์

$$P_i(\theta) = \frac{e^{Da_i (\theta - b_i)}}{1 + e^{Da_i (\theta - b_i)}} \quad i = 1, 2, \dots, n$$

รูปแบบโลจิสติก 3 พารามิเตอร์

$$P_i(\theta) = c_i + (1 - c_i) \frac{e^{Da_i (\theta - b_i)}}{1 + e^{Da_i (\theta - b_i)}} \quad i = 1, 2, \dots, n$$

รูปแบบโลจิสติก 4 พารามิเตอร์

$$P_i(\theta) = c_i + (\gamma_i - c_i) \frac{e^{Da_i (\theta - b_i)}}{1 + e^{Da_i (\theta - b_i)}} \quad i = 1, 2, \dots, n$$

เมื่อ P_i แทน ความน่าจะเป็นที่ผู้สอบระดับความสามารถที่ θ ตอบข้อสอบข้อที่ i ได้ถูกต้อง
 θ แทน ระดับความสามารถของผู้สอบ มีค่าอยู่ระหว่าง $-\infty$ ถึง $+\infty$ ในทางปฏิบัติมีค่าระหว่าง -3.0 ถึง $+3.0$

a_i แทน ค่าอำนาจจำแนกของข้อสอบข้อที่ i หาได้จากความชัน (slope) ของโค้งลักษณะข้อสอบที่ตำแหน่ง b_i ค่านี้บอกว่า ข้อสอบจำแนกผู้สอบความสามารถสูงออกจากผู้สอบความสามารถต่ำได้ดีเพียงไร ในทางปฏิบัติมีค่าเป็นบวกอยู่ระหว่าง 0 ถึง $+2.5$

b_i แทน ค่าความยากของข้อสอบข้อที่ i เป็นค่าบนสเกลความสามารถ ณ ตำแหน่งโค้งลักษณะข้อสอบมีความน่าจะเป็นในการตอบข้อสอบถูกต้องเท่ากับ 0.5 ถ้าเป็นรูปแบบที่มีพารามิเตอร์การเดา ค่าความยากจะได้จากตำแหน่งโค้งลักษณะข้อสอบมีความน่าจะเป็นในการตอบข้อสอบถูกต้องเท่ากับ $(1 + c_i)/2$ ค่าความยากมีค่าอยู่ระหว่าง $-\infty$ ถึง $+\infty$ ในทางปฏิบัติมีค่าอยู่ระหว่าง -2.5 ถึง $+2.5$

c_i แทน ค่าการเดาของข้อสอบข้อที่ i หรือค่าความน่าจะเป็นที่ผู้สอบความสามารถต่ำจะทำข้อสอบข้อที่ i ได้ถูกต้อง โดยปกติมีค่าอยู่ระหว่าง 0 ถึง 1 ในทางปฏิบัติใช้ค่าการเดาไม่เกิน 0.3

γ_i แทน ค่าความสะเพร่า เกิดจากผู้สอบความสามารถสูงจะตอบข้อสอบข้อที่ i ผิด

n แทน จำนวนข้อสอบในแบบทดสอบ

D แทน ค่าคงที่ใช้ปรับสเกลให้ฟังก์ชันโลจิสติกและฟังก์ชันนอร์มัลโอโจไฟท์ให้ใกล้เคียงกันมากที่สุด มีค่าเท่ากับ 1.7

e แทน ค่าคงที่ของลอการิทึมฐานธรรมชาติ (Natural Log) มีค่าเท่ากับ 2.718

1.4 สารสนเทศข้อสอบและสารสนเทศแบบทดสอบ (Item and Test Information)

สารสนเทศข้อสอบเป็นเครื่องบ่งชี้คุณภาพข้อสอบ สามารถให้ค่าสูงสุดเมื่อข้อสอบมีค่าความยากใกล้เคียงหรือเท่ากับความสามารถของผู้สอบ ค่าอำนาจจำแนกสูง และค่าการเดาน้อยมากจนเข้าใกล้ศูนย์ เมื่อใช้รูปแบบการตอบข้อสอบประเภทให้ผลสองค่า (Dichotomous Scoring) สารสนเทศข้อสอบแสดงเป็นสมการดังนี้

$$I_i(\theta) = \frac{[P'_i(\theta)]}{P_i(\theta)Q_i(\theta)} \quad i = 1, 2, 3, \dots, k$$

เมื่อ $I_i(\theta)$ แทน สารสนเทศได้จากข้อสอบข้อที่ i สำหรับผู้สอบความสามารถ θ

$P_i(\theta)$ แทน ความน่าจะเป็นที่ผู้สอบความสามารถ θ จะตอบข้อสอบที่ i ได้ถูกต้อง

$P'_i(\theta)$ แทน อนุพันธ์ของ $P_i(\theta)$ เมื่อพิจารณาจากตำแหน่งความสามารถ θ

$Q_i(\theta)$ แทน ความน่าจะเป็นที่ผู้สอบความสามารถ θ จะตอบข้อสอบที่ i ผิดหรือมีค่าเท่ากับ $1 - P_i(\theta)$

สารสนเทศข้อสอบแต่ละข้อของแบบทดสอบทั้งฉบับเมื่อนำมารวมเข้าด้วยกันได้สารสนเทศแบบทดสอบบนตำแหน่งความสามารถเดียวกัน ให้ความแม่นยำในการประมาณค่าความสามารถของผู้สอบว่า เมื่อทำแบบทดสอบทั้งฉบับ ผลการทดสอบบอกถึงผู้สอบมีความสามารถอยู่ ณ ตำแหน่งใด

$$T(\theta) = \sum_{i=1}^n I_i(\theta) \quad i = 1, 2, 3, \dots, n$$

เมื่อ $T(\theta)$ แทน สารสนเทศแบบทดสอบสำหรับผู้สอบที่มีความสามารถ θ

$I_i(\theta)$ แทน สารสนเทศข้อสอบข้อที่ i สำหรับผู้สอบที่มีความสามารถ θ

เมื่อนำสารสนเทศแบบทดสอบไปหาค่าความคลาดเคลื่อนมาตรฐานของการประมาณค่าความสามารถของผู้สอบ (Standard Error of Estimation : SE) ได้ดังนี้

$$SE(\hat{\theta}) = \frac{1}{T(\theta)}$$

ค่าความคลาดเคลื่อนมาตรฐานของการประมาณค่าความสามารถของผู้สอบสำหรับทฤษฎีการตอบข้อสอบมีบทบาทเหมือนความคลาดเคลื่อนมาตรฐานของการวัด (Standard Error of Measurement : SEM) ในทฤษฎีการทดสอบแบบมาตรฐานเดิม แต่ความคลาดเคลื่อนมาตรฐานของการประมาณค่าความสามารถในทฤษฎีการตอบข้อสอบสามารถเปลี่ยนแปลงไปตามค่าความสามารถของผู้สอบได้ (Hambleton; Swaminathan; & Rogers. 1991 : 91-96)

1.5 การประมาณค่าความสามารถและการประมาณค่าพารามิเตอร์ข้อสอบ

การประมาณค่าความสามารถและการประมาณค่าพารามิเตอร์ข้อสอบตามทฤษฎีการตอบข้อสอบ มีแนวทางดังนี้

1.5.1 การประมาณค่าความสามารถ (Estimation of Ability) แบ่งเป็นวิธีดังนี้

1.5.1.1 การประมาณค่าความเป็นไปได้สูงสุด(Maximum Likelihood Estimation : MLE)

การประมาณค่าความสามารถกระทำโดยให้สุ่มเลือกผู้สอบทำข้อสอบ n ข้อ ได้รูปแบบการตอบข้อสอบ คือ $U_1, U_2, \dots, U_j, \dots, U_n$ U_j เท่ากับ 1 เมื่อตอบข้อสอบข้อที่ j ถูกต้อง หรือ เท่ากับ 0 เมื่อตอบข้อสอบข้อที่ j ผิด ตามข้อตกลงของความเป็นอิสระในการตอบข้อสอบ (Local Independence) ความน่าจะเป็นร่วมของรูปแบบการตอบข้อสอบจะเท่ากับผลคูณความน่าจะเป็นของการตอบข้อสอบแต่ละข้อ นั่นคือ

$$P(U_1 \dots U_2 \dots U_3 \dots U_n | \theta) = P(U_1 | \theta) P(U_2 | \theta) \dots P(U_j | \theta) \dots P(U_n | \theta)$$

หรือ

$$P(U_1 \dots U_2 \dots U_n | \theta) = \prod_{j=1}^n P(U_j | \theta)$$

ฟังก์ชันความเป็นไปได้ (Likelihood Function) เมื่อ U_j เท่ากับ 1 หรือ 0 แสดงได้ดังนี้

$$P(U_1 \dots U_2 \dots U_n | \theta) = \prod_{j=1}^n P(U_j | \theta)^{U_j} [1 - P(U_j | \theta)]^{1-U_j}$$

$$P(U_1 \dots U_2 \dots U_n | \theta) = \prod_{j=1}^n P_j^{u_j} Q_j^{1-u_j}$$

เมื่อ $P_j = P(U_j | \theta)$ และ $Q_j = 1 - P(U_j | \theta)$

ฟังก์ชันความเป็นไปได้ (Likelihood Function) คือ

$$L(u_1 \dots u_2 \dots u_n | \theta) = \prod_{j=1}^n P_j^{u_j} Q_j^{1-u_j}$$

เมื่อ $U_j = u_j$ (u_j คือการตอบข้อสอบข้อที่ j)

เมื่อนำฟังก์ชันความเป็นไปได้ใส่ลอการิทึม (Logarithms) จะได้

$$\ln L(u | \theta) = \sum_{j=1}^n [u_j \ln P_j + (1 - u_j) \ln(1 - P_j)]$$

ความสามารถในการตอบข้อสอบที่ทำให้ฟังก์ชันความเป็นไปได้นี้มีค่าสูงสุด เป็นการประมาณค่าความเป็นไปได้สูงสุด (Maximum Likelihood Estimation : MLE) ของความสามารถในการตอบข้อสอบสำหรับผู้สอบคนนั้น (Hambleton; Swaminathan; & Rogers. 1991 : 32-50)

1.5.1.2 การประมาณค่าเบย์เซียน (Bayesian Estimation) เป็นวิธีการประมาณค่าความสามารถซึ่งมีแนวคิดพื้นฐานประยุกต์มาจากฟังก์ชันความเป็นไปได้ (Likelihood Function) รวมเข้ากับสารสนเทศเบื้องต้น (Prior Information) ของผู้สอบ ทฤษฎีของเบย์ (Bayes Theory) กล่าวถึงความน่าจะเป็นของเหตุการณ์ A ทำให้เกิดเหตุการณ์ B ว่า

$$P(A|B) \propto P(B|A)P(A)$$

เมื่อ $P(A)$ แทน ความน่าจะเป็นในการเกิดเหตุการณ์ A ก่อน จะเป็นจริงเมื่อ A คือความสามารถของผู้สอบและ B คือรูปแบบการตอบข้อสอบ ทฤษฎีของเบย์เซียนเป็นสมการดังนี้

$$f(\theta|u) \propto f(u|\theta)f(\theta)$$

$f(u|\theta)$ ก็คือ ฟังก์ชันความเป็นไปได้ (likelihood function) ดังนั้น

$$f(\theta|u) \propto L(u|\theta)f(\theta)$$

การปรับฟังก์ชันความเป็นไปได้เรียกว่า การรวมเข้าภายหลัง (Posterior Density)

ถ้าการแจกแจงความสามารถเบื้องต้นเป็นรูปแบบเดียวกัน $f(\theta)$ จะเป็นค่าคงที่ ดังนั้น

$$f(\theta|u) \propto L(u|\theta)$$

การแจกแจงค่าการรวมเข้าภายหลัง (Posterior Density) มีความหมายเหมือนค่าเฉลี่ยของการ

แจกแจงคำนวณโดยการประมาณการแจกแจง k ค่า ความถี่ที่จุด θ_j ($j = 1, 2, \dots, k$) ก็คือ $f(\theta_j|u)$ ค่าเฉลี่ยที่ได้คือ

$$U(\theta|u) = \frac{\sum_{j=1}^k \theta_j f(\theta_j|u)}{\sum_{j=1}^k f(\theta_j|u)}$$

หรือเรียกว่า การประมาณค่าภายหลังคาดหวัง [Expected A Posteriori (EAP) Estimation]

(Hambleton; Swaminathan; & Rogers. 1991 : 39 ; citing Bock & Mislevy. 1982)

1.5.2 การประมาณค่าพารามิเตอร์ข้อสอบ (Estimation of Item Parameter)

การประมาณค่าพารามิเตอร์ข้อสอบต้องทราบค่าความสามารถของผู้สอบแต่ละคน การตอบข้อสอบข้อหนึ่งของผู้สอบ N คน จะได้ฟังก์ชันความเป็นไปได้ (Likelihood Function) ดังนี้

$$L(u_1, u_2, \dots, u_n | \theta, a, b, c) = \prod_{j=1}^n P_j^{u_j} Q_j^{1-u_j}$$

ความแตกต่างระหว่างฟังก์ชันความเป็นไปได้สำหรับผู้สอบคนหนึ่งและข้อสอบข้อหนึ่งกล่าวได้ว่า สำหรับข้อสอบข้อหนึ่งจะไม่อ้างถึงความเป็นอิสระในการตอบข้อสอบเพราะว่า ผู้สอบ N คน ตอบข้อสอบหนึ่งข้อมีความเป็นอิสระอยู่แล้ว แต่จะให้ความสนใจความเป็นอิสระในการตอบข้อสอบเมื่อผู้สอบหนึ่งคนตอบข้อสอบหลายข้อ อีกประการหนึ่ง ฟังก์ชันความเป็นไปได้สำหรับผู้สอบเป็นพหุคูณสำหรับพารามิเตอร์ข้อสอบ กรณีรูปแบบ 3 พารามิเตอร์ คือ ค่าความยาก (a) ค่าอำนาจจำแนก

(b) และค่าการเดา (c) ดังนั้น การประมาณค่าความเป็นไปได้สูงสุดของ a b และ c ต้องหาค่าของ a b และ c ที่สัมพันธ์กับค่าสูงสุดทั้งสามมิติ ทำได้โดยหาอนุพันธ์แรกของฟังก์ชันความเป็นไปได้ของพารามิเตอร์แต่ละตัว จัดอนุพันธ์เหล่านี้เป็นศูนย์และแก้ปัญหาร่วมๆ กัน (Hambleton; Swaminathan; & Rogers. 1991 : 39-40)

การประมาณค่าความสามารถของผู้สอบและพารามิเตอร์ข้อสอบ สามารถประมาณค่าได้โดยเลือกใช้โปรแกรมคอมพิวเตอร์ที่เหมาะสมกับรูปแบบ สรุปไว้ในตาราง 1 ดังนี้คือ

ตาราง 1 โปรแกรมการประมาณค่าพารามิเตอร์รูปแบบต่างๆ

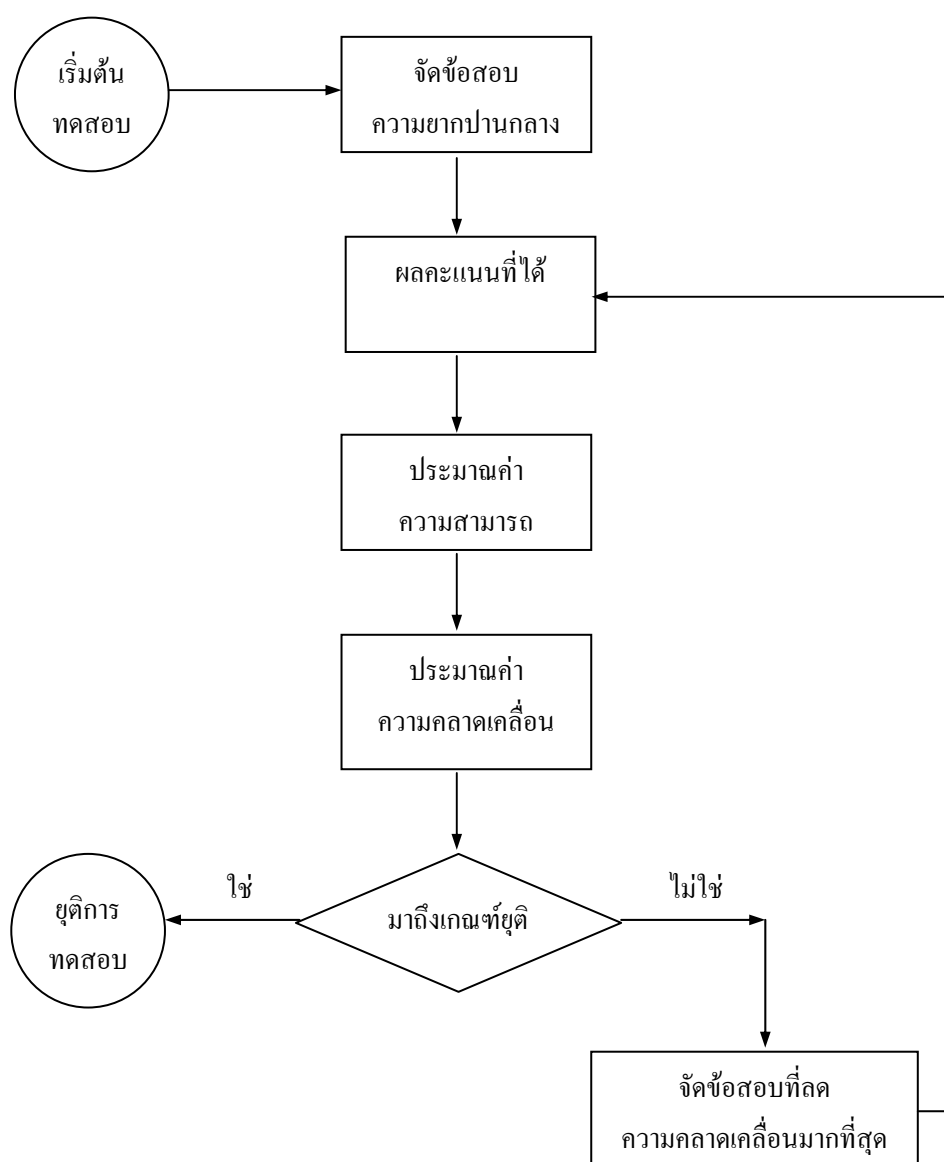
โปรแกรม	รูปแบบ	วิธีการประมาณค่า	ผู้พัฒนา
BICAL (replaced by BIG- SCALE)	1 PL	Unconditional Maximum Likelihood	Wright; & et al (1979) ; Wright; & et al (1989)
MICROSCALE	1 PL Multicategory	Unconditional Maximum Likelihood	Mediix Interactive Technologies (1986)
PML	1 PL	Conditional Maximum Likelihood	Gustafsson (1980a)
RASCAL	1 PL	Unconditional Maximum Likelihood	Assessment Systems Corp. (1988)
RIDA	1 PL	Conditional or Marginal Maximum Likelihood	Glas (1990)
ANCILLES	3 PL	Heuristic	Urry (1974, 1978)
ASCAL	1 PL, 2 PL, 3 PL	Modified Bayesian	Assessment Systems Corp. (1988)
LOGIST	1 PL, 2 PL, 3 PL	Unconditional Maximum Likelihood	Wingersky (1983) Wingersky; & et al.(1982)
BILOG	1 PL, 2 PL, 3 PL	Marginal Maximum Likelihood	Mislevy; & Bock (1984)
NOHARM	1 PL, 2 PL, 3 PL	Least Squares	Fraser; & McDonald (1988)
MULTILOG	Multicategory		Thissen (1986)
MIRTE	1 PL, 2 PL, 3 PL	Unconditional Maximum Likelihood	Carlson (1987)

ที่มา: Hambleton Ronald K.; Swaminathan H.; & Rogers H. Jane. (1991). Ability and Item Parameter Estimation. in *Fundamentals of Item Response Theory*. p. 48-50.

2. การทดสอบแบบปรับเหมาะ (Adaptive testing)

แบบทดสอบแบบปรับเหมาะ (Adaptive Test) เป็นแบบทดสอบที่ใช้ข้อสอบแตกต่างกันจัดให้กับผู้สอบแต่ละคน พิจารณาเลือกข้อสอบจากความสามารถภายในของผู้สอบที่ต้องการวัด (Weiss. 1985) ดังนั้น การทดสอบแบบปรับเหมาะ (Adaptive Testing : AT) กล่าวได้ว่า เป็นการทดสอบด้วยแบบทดสอบที่ใช้ข้อสอบแตกต่างกันจัดให้กับผู้สอบ หรือผู้สอบแต่ละคนจะได้รับข้อสอบแตกต่างกัน ขึ้นอยู่กับระดับความสามารถของผู้สอบแต่ละคนนั่นเอง

ในการทดสอบแบบปรับเหมาะนั้น อาศัยแนวคิดตามทฤษฎีการตอบข้อสอบ (Item Response Theory : IRT) โดยทั่วไปนิยมใช้รูปแบบ 3 พารามิเตอร์ ข้อสอบต้องทราบค่าพารามิเตอร์และรวบรวมข้อสอบทั้งหมดไว้ในคลังข้อสอบ ขั้นตอนการทดสอบแบบปรับเหมาะสรุปดังนี้



ภาพประกอบ 2 ขั้นตอนการทดสอบแบบปรับเหมาะ

ที่มา: Vispoel, Waite P. (1999). Creating Computerized Adaptive Tests of Music Aptitude : Problems, Solutions, and Future Directions. in Innovations in Computerized Assessment. p. 151-176.

ภาพประกอบ 2 สรุปขั้นตอนการทดสอบแบบปรับเหมาะ เริ่มต้นการทดสอบทำโดยเลือกข้อสอบจากคลังข้อสอบกำหนดความยากของข้อสอบที่ระดับปานกลางหรือความยากของข้อสอบใกล้เคียงกับความสามารถของผู้สอบตามสารสนเทศเดิม ผลคะแนนที่ได้นำไปคำนวณหาค่าประมาณความสามารถของผู้สอบ เพื่อใช้ค่าประมาณความสามารถนี้ทำการเลือกข้อสอบข้อที่ให้ความคลาดเคลื่อนน้อยที่สุดและจัดให้ผู้สอบ นั่นหมายถึง ถ้าตอบข้อสอบถูกต้อง ข้อสอบข้อต่อไปมีระดับความยากเพิ่มขึ้น ถ้าตอบข้อสอบผิด ข้อสอบข้อต่อไปมีระดับความยากลดลง การทดสอบดำเนินต่อเนื่องไปเรื่อยๆ และสิ้นสุดเมื่อความคลาดเคลื่อนของการประมาณค่าความสามารถไม่เกินเกณฑ์กำหนด หรือค่าประมาณความสามารถอยู่ในระดับยอมรับได้ หรือเมื่อใช้ข้อสอบตามจำนวนต้องการ หรือ ดังนั้น ข้อสอบที่จัดให้แก่ผู้สอบกล่าวได้ว่าไม่ยากเกินไปหรือยากเกินไปกว่าความสามารถของผู้สอบ เหมาะกับการวัดระดับความสามารถของผู้สอบแต่ละคน

2.1 ความเป็นมาของการทดสอบแบบปรับเหมาะ

การทดสอบตามทฤษฎีการตอบข้อสอบ แบบทดสอบจะให้ความแม่นยำในการวัดทุกระดับความสามารถของผู้สอบเมื่อค่าความยากของข้อสอบใกล้เคียงความสามารถของผู้สอบ จากเหตุผลนี้นำไปสู่แนวคิดสำหรับการสร้างแบบทดสอบแบบปรับเหมาะในเวลาต่อมา

การทดสอบแบบปรับเหมาะ (Adaptive Testing) หรือเรียกชื่ออื่นๆ เช่น Tailored Testing, Response-Contingent Testing, Programmed Testing, Individualized Testing, Branched Testing และ Sequential Testing เป็นต้น (Weiss, 1985) เริ่มใช้ทดสอบระยะแรก ปี ค.ศ. 1908 โดยอัลเฟรด บิเน็ต (Alfred Binet) และผู้ร่วมงาน ประยุกต์ใช้แบบทดสอบแบบปรับเหมาะสำหรับการทดสอบเชาว์ปัญญา เรียกว่า แบบทดสอบเชาว์ปัญญาบิเน็ต (Binet Intelligence Test) ลักษณะของแบบทดสอบเชาว์ปัญญาบิเน็ต คือ ข้อสอบตามอายุจะถูกจัดให้กับผู้สอบแต่ละคน เริ่มต้นด้วยข้อสอบตามอายุพื้นฐาน ถ้าผู้สอบตอบข้อสอบได้ถูกต้อง ข้อสอบข้อต่อไปมีความยากเพิ่มขึ้น (ข้อสอบระดับอายุสูงกว่าอายุพื้นฐาน) ถ้าผู้สอบตอบข้อสอบผิด ข้อสอบข้อต่อไปมีความยากลดลง (ข้อสอบระดับอายุต่ำกว่าอายุพื้นฐาน)

ประมาณปี ค.ศ. 1900-1950 มีการทดลองใช้แบบทดสอบแบบปรับเหมาะสำหรับศึกษากลยุทธ์ต่างๆ เพื่อกำหนดหนทางการทดสอบผ่านคลังข้อสอบ ในระยะแรกใช้กลยุทธ์สองขั้นตอน รูปแบบทางแยกคงที่แบบง่ายๆ โดยไม่เน้นจุดเริ่มต้นหรือจุดยุติการทดสอบ ระยะต่อมาขยายเป็นกลยุทธ์สามหรือสี่ขั้นตอนมีความซับซ้อนเพิ่มขึ้น รูปแบบทางแยกคงที่รูปแบบหนึ่งคือ รูปแบบปิรามิด (Pyramid Test) มีกฎทางแยกคือ จัดข้อสอบข้อต่อไปที่มีความยากมากขึ้นเมื่อตอบข้อสอบถูก และจัดข้อสอบที่มีความยากลดลงเมื่อตอบข้อสอบผิด รูปแบบปิรามิดนี้จำเป็นต้องเตรียมข้อสอบจำนวนมาก จึงได้มีการพัฒนารูปแบบที่ใช้ข้อสอบน้อยลง เรียกว่า รูปแบบเฟล็กซิเลเวล (Flexilevel Test) และพัฒนาต่อมาเป็นรูปแบบปรับระดับชั้น (Stratified Adaptive Test หรือ Stradaptive Test) รูปแบบนี้

แบ่งข้อสอบในคลังข้อสอบตามระดับค่าความยาก เริ่มต้นการทดสอบด้วยข้อสอบความยากตามที่กำหนด ถ้าตอบข้อสอบถูกต้องข้อสอบข้อต่อไปเป็นข้อสอบความยากสูงขึ้น ถ้าตอบข้อสอบผิดข้อสอบข้อต่อไปเป็นข้อสอบความยากลดลง ทั้งรูปแบบเฟล็กซิเบิลและรูปแบบปรับระดับขั้นยังคงเป็นรูปแบบทางแยกคงที่ ซึ่งยังมีข้อจำกัดได้แก่ 1. รูปแบบทางแยกคงที่ที่ใช้กับการทดสอบด้วยกระดาษคำตอบ เทคนิคง่ายๆ ในการคำนวณ แต่ใช้ยากกับข้อมูลที่มีกระบวนการซับซ้อน 2. รูปแบบทางแยกคงที่ให้ความสำคัญกับค่าความยากในการเลือกข้อสอบในคลังข้อสอบ โดยไม่พิจารณาค่าอำนาจจำแนกและค่าการเดา (ยกเว้นรูปแบบปรับระดับขั้นที่ใช้ค่าอำนาจจำแนกช่วยเรียงข้อสอบในแต่ละชั้น) 3. วิธีการให้คะแนนไม่ชัดเจน และ 4. กลยุทธ์ออกแบบมาให้ความยาวแบบทดสอบเท่ากัน (ยกเว้นรูปแบบปรับระดับขั้น) จากข้อจำกัดดังกล่าวมา นำไปสู่การพัฒนาแบบทางแยกแปรผัน (Variable-Branching Model) เพื่อแก้ปัญหาของรูปแบบทางแยกคงที่และให้กำลัง (Power) สูงกว่า (Weiss, 1985) ต่อมาในปี ค.ศ. 1960 ลอร์ดมีความเห็นว่า แบบทดสอบที่มีความยาวเท่ากันจะไม่ได้ประสิทธิภาพการทดสอบแก่ผู้สอบทุกคน โดยเฉพาะผู้สอบความสามารถระดับต่ำและระดับสูง จึงได้ปรับความยาวแบบทดสอบให้ลดลงโดยยังรักษาสารสนเทศความสามารถของผู้สอบสูงสุดไว้ และผู้สอบแต่ละคนจะได้รับแบบทดสอบลักษณะ เฉพาะแต่ละคน (Hambleton; Swaminathan; & Rogers. 1991 : 145)

การทดสอบแบบปรับเหมาะสามารถนำไปปฏิบัติได้มากขึ้นเมื่อใช้ร่วมกับคอมพิวเตอร์ จากความสามารถในการคำนวณการประมาณค่าความสามารถของผู้สอบ การเลือกข้อสอบ การแสดงคำถาม การให้คะแนนการตอบแบบทดสอบ และการออกแบบคลังข้อสอบ (De Ayala; Dodd; & Koch. 1990) นับได้ว่าเป็นการรวมกันของทฤษฎีการตอบข้อสอบ การทดสอบแบบปรับเหมาะ และใช้การดำเนินการด้วยคอมพิวเตอร์รวมเข้าไว้ด้วยกัน ตั้งแต่ปี ค.ศ. 1980 เป็นต้นมา ผลงานการศึกษาการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์มีการเผยแพร่อย่างต่อเนื่อง เช่น การประชุมการวัดผลการศึกษาในระดับชาติ (The National Council on Measurement in Education) ประจำปี ค.ศ. 1999 ของสหรัฐอเมริกา พบว่า 25 เปอร์เซ็นต์ของงานวิจัยด้านการวัดผลการศึกษาทั้งหมดเกี่ยวข้องกับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ (Meijer; & Nering. 1999) ปัจจุบันการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์นำไปประยุกต์ใช้ในการทดสอบมาตรฐานต่างๆ อย่างกว้างขวาง [เช่น The Graduate Record Examination (GRE) The Graduate Management Admission Test (GMAT) The Test of English as a Foreign Language (TOEFL) เป็นต้น]

สต็อกกิงและสแวนสัน (Stocking; & Swanson. 1993) ได้ศึกษาผลงานวิจัยต่างๆ ที่เกี่ยวข้องกับการทดสอบแบบปรับเหมาะ สรุปเป็นพื้นฐานของการทดสอบแบบปรับเหมาะภายใต้ทฤษฎีการตอบข้อสอบ ดังนี้ 1. การเลือกข้อสอบขึ้นอยู่กับพื้นฐานข้อมูลของผู้สอบและจัดให้กับผู้สอบ 2. การตอบข้อสอบจะนำไปใช้ปรับค่าประมาณความสามารถของผู้สอบ 3. การเลือกข้อสอบข้อต่อไปขึ้นอยู่กับ

ค่าประมาณความสามารถครั้งสุดท้าย และดำเนินการต่อไปจนถึงเกณฑ์ยุติ และ 4. คะแนนสุดท้ายของผู้สอบได้แก่ค่าประมาณความสามารถของผู้สอบ ภายหลังทำข้อสอบทุกข้อที่จัดให้แล้ว

นอกจากนี้ สต็อกกิงและสแวนสัน (Stocking; & Swanson. 1998 ; citing Davey; & Parshall. 1995) กล่าวถึงวัตถุประสงค์ของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ไว้ 4 ประการ ได้แก่ 1. เพื่อเพิ่มประสิทธิภาพของแบบทดสอบโดยเลือกข้อสอบที่เหมาะสมที่สุดให้กับผู้สอบ 2. เพื่อให้แบบทดสอบวัดลักษณะ (Trait) เดียวกันสำหรับผู้สอบแต่ละคน โดยใช้การควบคุมแบบไม่ใช้สถิติ (Non-Statistical Constraint) 3. เพื่อป้องกันความปลอดภัยของคลังข้อสอบโดยควบคุมอัตราการแสดงข้อสอบ และ 4. เพื่อเพิ่มการใช้ข้อสอบในคลังข้อสอบให้มากที่สุดโดยไม่เสียความปลอดภัยของข้อสอบ

ไวเนอร์ (Wainer. 1990 : 11 ; citing Green. 1983) สรุปข้อดีของแบบทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ได้แก่ แบบทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ให้ความปลอดภัยในการทดสอบมากกว่าแบบทดสอบใช้กระดาษคำตอบ ข้อสอบทั้งหมดจะเก็บในคลังข้อสอบในคอมพิวเตอร์ ผู้สอบได้รับข้อสอบโดยผ่านขั้นตอนการเลือกข้อสอบให้เหมาะสมกับความสามารถของตนโดยไม่สามารถทราบล่วงหน้าว่าจะได้รับข้อสอบข้อใด ถ้าผู้สอบล่วงรู้ข้อสอบจำนวนมากในคลังข้อสอบก็ถือว่าเป็นสิ่งดีเพราะถือว่าผู้สอบมีความรอบรู้ในเนื้อหาวิชานั้น ความสามารถของผู้สอบจะเพิ่มขึ้นและคะแนนสูงตามไปด้วย ผู้สอบสามารถกำหนดการทดสอบด้วยตนเองใช้เวลาการทำข้อสอบโดยอิสระ ผู้สอบความสามารถสูงจะทำข้อสอบจำนวนน้อย ขณะที่ผู้สอบความสามารถต่ำต้องทำข้อสอบจำนวนมากกว่า ขึ้นอยู่กับระดับความสามารถและความไวในการตอบของแต่ละคน ความสนใจและตั้งใจของผู้สอบจะจ่อกับการทำข้อสอบตรงหน้าจากการได้รับข้อสอบไม่เหมือนกัน เพื่อให้ได้ผลการทดสอบซึ่งเป็นผลมาจากความสามารถของตนเองในการทำข้อสอบ แบบทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ไม่ใช่กระดาษคำตอบทำให้ลดปัญหาจากการใช้กระดาษคำตอบที่อาจส่งผลกระทบต่อคะแนนสอบ เมื่อพบความบกพร่องของข้อสอบสามารถลบข้อสอบนั้นออกได้ทันที รูปแบบการแสดงผลข้อสอบทำได้หลากหลาย เช่น การใช้ใช้มัลติมีเดียช่วยในการออกแบบการทดสอบ และให้ผู้สอบทำข้อสอบตามรูปแบบที่จัดไว้ การตรวจให้คะแนนทำได้ง่ายสามารถแจ้งผลกลับไปยังผู้สอบได้ทันที

อย่างไรก็ตามข้อจำกัดของแบบทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ (Jacobs; & Chase. 1992) มีดังนี้คือ การสร้างคลังข้อสอบต้องใช้ข้อสอบจำนวนมากเพียงพอและต้องครอบคลุมเนื้อหาวิชาที่ต้องการทดสอบ ข้อสอบระยะแรกต้องทดสอบกับผู้สอบจำนวนมากเพื่อให้ได้ค่าพารามิเตอร์ข้อสอบมีความถูกต้องแม่นยำ ค่าใช้จ่ายสูงในการดำเนินการจัดซื้ออุปกรณ์และพัฒนาระบบ ผู้สอบจะต้องมีความรู้การใช้คอมพิวเตอร์เพียงพอ และการแสดงผลการสอบและแปลความหมายควรมีการออกแบบอย่างดีเพราะผู้สอบส่วนมากเคยชินกับการประเมินผลเป็นระดับเกรดที่ใช้กันทั่วไป

สรุป การทดสอบแบบปรับเหมาะเป็นการทดสอบโดยใช้ข้อสอบแตกต่างกันจัดให้แก่ผู้สอบ การพิจารณาเลือกข้อสอบขึ้นอยู่กับค่าประมาณความสามารถของผู้สอบ ถ้าตอบข้อสอบถูกต้อง ข้อสอบข้อต่อไปจะมีความยากเพิ่มขึ้น ถ้าตอบข้อสอบผิดข้อสอบข้อต่อไปจะมีระดับความยากลดลง ผลการตอบข้อสอบถูกนำไปประมาณค่าความสามารถใหม่สำหรับใช้เลือกข้อสอบข้อต่อไป การทดสอบจะดำเนินต่อไปเรื่อยๆ จนถึงกฎการยุติถึงสิ้นสุดการทดสอบ การทดสอบแบบปรับเหมาะในระยะแรกเริ่มจากการใช้กลยุทธ์ง่ายๆ พัฒนาจนมีความซับซ้อนมากขึ้น เพื่อแก้ปัญหาต่างๆ และเพิ่มประสิทธิภาพในการทดสอบ ปัจจุบันการทดสอบแบบปรับเหมาะเป็นที่นิยมแพร่หลายสังเกตได้จากการใช้การทดสอบแบบปรับเหมาะในการทดสอบมาตรฐานสำคัญต่างๆ และผลงานวิจัยตีพิมพ์เผยแพร่อย่างต่อเนื่อง

2.2 ความแตกต่างระหว่างแบบทดสอบแบบปรับเหมาะ (Adaptive Test) และแบบทดสอบประเพณีนิยม (Conventional Test)

แบบทดสอบประเพณีนิยมอาศัยหลักการตามทฤษฎีการทดสอบแบบมาตรฐานเดิม (Classical Test Theory : CTT) คุณภาพของข้อสอบได้มากี่ต่อเมื่อผู้สอบทุกคนทำข้อสอบทุกข้อในแบบทดสอบชุดเดียวกัน ผลการทดสอบจะได้ค่าความสามารถของผู้สอบกระจายเป็นช่วงกว้างบนเส้นต่อเนื่องของค่าความสามารถตามจุดประสงค์การวัดของแต่ละแบบทดสอบนั้น

ข้อสอบในแบบทดสอบประเพณีนิยมมักกำหนดค่าความยากปานกลางประมาณ 0.5 ผู้สอบระดับความสามารถบริเวณช่วงค่าความยาก 0.5 จะได้รับการวัดผลอย่างแม่นยำ แบบทดสอบประเพณีนิยมจึงใช้ได้ดีกับผู้สอบระดับความสามารถปานกลางเท่านั้น อีกประการหนึ่ง การให้ผู้สอบทุกคนต้องทำข้อสอบจำนวนเท่าๆ กัน ทำให้มีข้อสอบจำนวนน้อยที่เหมาะสมกับผู้สอบระดับความสามารถสูงและผู้สอบระดับความสามารถต่ำ ความแม่นยำในการวัดสำหรับผู้สอบเหล่านี้จึงลดลงอย่างรวดเร็ว เมื่อข้อสอบมีความยากเกินไปสำหรับผู้สอบความสามารถต่ำหรือง่ายเกินไปสำหรับผู้สอบความสามารถสูง สารสนเทศของผู้สอบจึงน้อยลงตามไปด้วย เพราะฉะนั้น ข้อสอบยากเกินไปหรือง่ายเกินไปเมื่อจัดให้ผู้สอบจะไม่วัดความสามารถของผู้สอบได้ถูกต้อง (Weiss. 1974 ; 1985) มีผลต่อความเชื่อมั่นในภาพรวมและให้ความเที่ยงตรงลดลง

การจำกัดเวลาในการทดสอบเพื่อความสะดวกของผู้คุมสอบ อาจนำมาซึ่งความคลาดเคลื่อนของคะแนนแบบทดสอบได้ เนื่องจากผู้สอบบางคนตอบข้อสอบได้ในเวลาจำกัดในขณะที่บางคนทำไม่ได้ ผลคือ การจำกัดเวลาจะส่งผลกระทบต่อผู้สอบแตกต่างกัน การประมาณขนาดความคลาดเคลื่อนของคะแนนแบบทดสอบทำได้ไม่แน่นอน (Weiss. 1974)

การทดสอบโดยทั่วไปถึงแม้ว่าถูกออกแบบเพื่อจัดผลของผู้คุมสอบต่อผู้สอบ แต่ปฏิสัมพันธ์ระหว่างผู้คุมสอบกับผู้สอบยังพบได้ภายใต้เงื่อนไขการจัดการสอบ ดังนั้น ความแตกต่างของผู้คุมสอบ

สามารถส่งผลต่อความคลาดเคลื่อนของคะแนนของผู้สอบได้ในการทดสอบประเพณีนิยม (Weiss. 1974 ; citing Weiss; & Betz. 1973)

จากเหตุผลดังกล่าว มีการศึกษาวิธีการทดสอบอื่นๆ นอกเหนือจากการทดสอบประเพณีนิยม เพื่อเพิ่มความเชื่อมั่นและความเที่ยงตรงของคะแนนการทดสอบ วิธีที่ได้รับความนิยมคือ การทดสอบแบบปรับเหมาะตามหลักการตามทฤษฎีการตอบข้อสอบ (Item Response Theory : IRT) จากแนวคิดที่ว่า ข้อสอบที่จัดให้แก่ผู้สอบแต่ละคนมีค่าความยากเหมาะสมกับความสามารถของผู้สอบคนนั้น ทำให้ความน่าจะเป็นของการตอบถูกต้องหรือตอบผิดมีค่าเท่ากับ 0.5 ดังนั้น ข้อสอบที่จัดให้แก่ผู้สอบจะวัดระดับความสามารถของผู้สอบคนนั้นอย่างแม่นยำ มีความเฉพาะเจาะจงมากกว่าการทดสอบประเพณีนิยมซึ่งให้เพียงระดับความสามารถโดยเฉลี่ยของกลุ่มผู้สอบ

ภายใต้กลยุทธ์ของการทดสอบแบบปรับเหมาะ การใช้ข้อสอบจำนวนมากหรือน้อยนั้นขึ้นอยู่กับระดับความสามารถของผู้สอบ ข้อสอบความยากเหมาะสมกับความสามารถของผู้สอบจะลดผลที่เป็นอุปสรรคและการเดา รวมทั้งจำนวนข้อสอบต่ำกว่าหรือสูงกว่าระดับความสามารถของผู้สอบจะใช้น้อยลง ลดความเบื่อหน่ายของผู้สอบในการทำแบบทดสอบ โดยทั่วไป การทดสอบแบบปรับเหมาะจะไม่จำกัดเวลา อนุญาตให้ผู้สอบได้ใช้เวลาตามต้องการ การจัดจำนวนข้อสอบให้มีความเฉพาะในแต่ละบุคคล ทำให้ผู้สอบบางคนลดเวลาในการทำแบบทดสอบ นอกจากนี้ เวลาในการตอบข้อสอบสามารถใช้เป็นข้อมูลในการวัดทางจิตวิทยาได้ ไวส์ (Weiss. 1974 ; citing Weiss; & Betz. 1973) กล่าวว่า การทดสอบแบบปรับเหมาะสามารถลดเวลาโดยไม่ลดความเชื่อมั่นและความเที่ยงตรง และอาจจะเพิ่มความเชื่อมั่นและความเที่ยงตรงได้

การทดสอบแบบปรับเหมาะประยุกต์ใช้กับคอมพิวเตอร์มีข้อดีเหนือกว่าการทดสอบประเพณีนิยมหรือการทดสอบที่ใช้กระดาษและดินสอ ให้ความเชื่อมั่นและความเที่ยงตรงได้เท่าหรือมากกว่าการทดสอบที่ใช้กระดาษและดินสอ อย่างไรก็ตาม สิ่งที่ต้องระวังในการใช้การทดสอบแบบปรับเหมาะด้วยเครื่องคอมพิวเตอร์ ได้แก่ 1. การแสดงข้อสอบ (Item Exposure) ข้อสอบคุณภาพสูงมักนำไปแสดงให้แก่ผู้สอบมากเกินไป กระทบต่อความปลอดภัยของข้อสอบ 2. ความเที่ยงตรงของข้อสอบ เนื่องจากสัดส่วนข้อสอบต้องครอบคลุมเนื้อหาวิชา แบบทดสอบของผู้สอบแต่ละคนจึงเป็นเหมือนแบบทดสอบคู่ขนานกัน และสามารถเปรียบเทียบได้กับแบบทดสอบประเพณีนิยม ดังนั้น การเลือกข้อสอบสำหรับการทดสอบแบบปรับเหมาะด้วยเครื่องคอมพิวเตอร์ นับเป็นปัจจัยสำคัญที่ควรคำนึงถึง (Leung; Chang; & Hau. 2001 : On Line)

สรุป แบบทดสอบแบบปรับเหมาะมีข้อได้เปรียบมากกว่าแบบทดสอบประเพณีนิยม เพราะการเลือกข้อสอบในแบบทดสอบแบบปรับเหมาะ จะเลือกข้อสอบความยากเหมาะสมกับความสามารถของผู้สอบแต่ละคน การทดสอบจึงให้ความแม่นยำที่ทุกระดับความสามารถของผู้สอบ แตกต่างจาก

แบบทดสอบประเพณีนิยม ซึ่งให้ความแม่นยำในการทดสอบเฉพาะผู้สอบความสามารถระดับปานกลางเท่านั้น นอกจากนี้ การทดสอบแบบปรับเหมาะยอมให้ผู้สอบใช้เวลาทำข้อสอบได้ตามต้องการ และจัดผลกระทบของผู้คุมสอบที่มีต่อผลการทดสอบ ด้วยเหตุผลดังกล่าว การทดสอบแบบปรับเหมาะจึงสามารถให้ความเชื่อมั่นและความเที่ยงตรงเป็นที่ยอมรับได้สำหรับใช้ในการทดสอบ

2.3 กลยุทธ์ของการทดสอบแบบปรับเหมาะ (Strategies of Adaptive Testing)

การทดสอบแบบปรับเหมาะมีกลยุทธ์ต่างๆ ซึ่งควบคุมโดยคลังข้อสอบ แต่ละกลยุทธ์แตกต่างกันด้วยทิศทางในการทดสอบผ่านคลังข้อสอบตามการจัดลำดับในแต่ละกลยุทธ์ (Weiss, 1973) ซึ่งเป็นผลมาจากการตอบข้อสอบตามระดับความสามารถของผู้สอบแต่ละคน กลยุทธ์ต่างๆ มีดังนี้คือ

2.3.1 กลยุทธ์สองขั้นตอน (Two-Stage Strategies)

กลยุทธ์สองขั้นตอนเป็นรูปแบบง่ายที่สุดในกลยุทธ์การทดสอบแบบปรับเหมาะ ประกอบด้วย 2 ขั้นตอนและใช้แบบทดสอบ 2 ฉบับ ขั้นตอนแรก ผู้สอบต้องทำแบบทดสอบ เรียกว่าแบบทดสอบแยกทิศทาง (Routing Test) ประกอบด้วยข้อสอบจำนวนไม่มากนัก ข้อสอบอาจมีระดับความยากปานกลางหรือมีความยากแตกต่างกันจากง่ายไปยาก ผู้สอบต้องทำข้อสอบครบทุกข้อ ผลคะแนนที่ได้นำไปประมาณค่าความสามารถของผู้สอบเริ่มแรก ก่อนนำไปใช้ในขั้นตอนที่สองเพื่อเลือกแบบทดสอบให้เหมาะสมกับความสามารถของผู้สอบเป็นรายบุคคล แบบทดสอบขั้นตอนที่สอง เรียกว่าแบบทดสอบวัด (Main หรือ Measurement Test) มีหลายฉบับครอบคลุมระดับความยากจากง่ายที่สุดถึงยากที่สุด แบบทดสอบวัดจึงให้สารสนเทศได้มากกว่าแบบทดสอบแยกทิศทาง ถ้าผู้สอบตอบข้อสอบในขั้นตอนแรกได้ถูกต้องมากข้อจะได้รับแบบทดสอบวัดฉบับความยากสูง แต่ถ้าผู้สอบตอบข้อสอบในขั้นตอนแรกได้ถูกต้องน้อยข้อ จะได้รับแบบทดสอบวัดฉบับความยากต่ำลงมา

2.3.2 กลยุทธ์หลายขั้นตอน (Multi-Stage Strategies)

กลยุทธ์นี้แตกต่างจากกลยุทธ์สองขั้นตอน เนื่องจาก กลยุทธ์สองขั้นตอนตัดสินทางแยกครั้งเดียว จากแบบทดสอบแยกทิศทางไปยังแบบทดสอบวัด แต่กลยุทธ์หลายขั้นตอนตัดสินทางแยกหลายลำดับขั้นหลังจากผู้สอบตอบคำถามแต่ละข้อ ประกอบด้วยรูปแบบต่างๆ ดังนี้คือ

รูปแบบทางแยกคงที่ (Fixed Branching Model) รูปแบบนี้มีผังการจัดเรียงข้อสอบและเส้นทาง แยกในการทำข้อสอบขั้นต่อไปแน่นอน การทดสอบเคลื่อนผ่านผังโครงสร้างไปในทิศทางต่างกันหลังจากตอบข้อสอบแต่ละข้อ ผลการตอบข้อสอบนำไปหาค่าพารามิเตอร์ของข้อสอบและค่าประมาณความสามารถของผู้สอบเพื่อกำหนดทางแยกในขั้นต่อไป แบ่งเป็นรูปแบบต่างๆ ได้แก่ รูปแบบปิรามิดต่างๆ รูปแบบเฟล็กซิเบิล และรูปแบบปรับระดับขั้น

รูปแบบทางแยกแปรผัน (Variable Branching Model) รูปแบบนี้กฎทางแยกไม่เป็นไปตามโครงสร้างคลังข้อสอบที่ใช้ผลการตอบถูกหรือผิด แต่คลังข้อสอบจะปฏิบัติการคำนวณโดย

พิจารณาค่าความยากและค่าอำนาจจำแนกของข้อสอบ และเลือกข้อสอบเหมาะสมที่สุดจัดให้แก่ผู้สอบ รูปแบบทางแยกแปรผันจำแนกเป็น การประมาณค่าเบย์เซียน (Bayesian Estimation) และการประมาณค่าความเป็นไปได้สูงสุด (Maximum likelihood Estimation)

ในการวิจัยครั้งนี้ สนใจ ศึกษาการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก (a-Stratified Model) ซึ่งพัฒนามาจากกลยุทธ์หลายขั้นตอนรูปแบบปรับระดับชั้นของไวส์ (Weiss. 1973 ; 1974) เพื่อทำความเข้าใจแนวคิดและหลักการเบื้องต้น ขอบิยารายละเอียดของรูปแบบปรับระดับชั้น ดังนี้

รูปแบบปรับระดับชั้น (Stradaptive Model) การทดสอบรูปแบบนี้เริ่มจากการเตรียมคลังข้อสอบโดย แบ่งแบบทดสอบเป็นกลุ่มข้อสอบย่อยตามค่าความยากและจัดเข้าแต่ละชั้น (Stratum) ของคลังข้อสอบ กลุ่มข้อสอบย่อยในแต่ละชั้นกระจายอยู่รอบๆ ระดับค่าความยากแสดงเป็นยอดโค้ง (Peak) ส่วนสูงของยอดโค้งแสดงถึงความหนาแน่นของจำนวนข้อสอบในระดับความยากตามที่กำหนด ค่าความยากเฉลี่ยของข้อสอบชั้นถัดไปจะสูงกว่าค่าความยากเฉลี่ยของข้อสอบชั้นก่อน นั่นคือ ชั้นที่ 1 บรรจุด้วยกลุ่มข้อสอบมีค่าความยากน้อยที่สุด ชั้นที่ 2 บรรจุด้วยกลุ่มข้อสอบมีค่าความยากเฉลี่ยสูงกว่าชั้นที่ 1 ชั้นอื่นๆ บรรจุด้วยกลุ่มข้อสอบมีค่าความยากเฉลี่ยสูงขึ้นเรื่อยๆ จนถึงชั้นสูงสุด

การทดสอบแบบปรับเหมาะรูปแบบปรับระดับชั้นเริ่มด้วยการประมาณค่าระดับความสามารถของผู้สอบ พิจารณาจากสารสนเทศเบื้องต้นหรือข้อมูลส่วนตัวนำไปกำหนดจุดนำเข้า (Entry Point) ผู้สอบความสามารถต่ำจะเริ่มต้นการทดสอบด้วยการทำข้อสอบง่าย และผู้สอบความสามารถสูงจะเริ่มต้นการทดสอบด้วยข้อสอบยาก ผลการตอบข้อสอบข้อแรกนำไปประมาณค่าความสามารถและกำหนดทางแยกตามกฎที่เลือกใช้ กรณีกฎทางแยกขึ้นหนึ่ง-ลงหนึ่งเมื่อตอบข้อสอบถูกต้อง จะนำไปสู่ข้อสอบในชั้นต่อไปที่มีระดับความยากสูงสุดในชั้นนั้น ถ้าตอบข้อสอบผิด ข้อสอบข้อต่อไปจะอยู่ในชั้นต่ำลงมาทำให้ค่าความยากโดยเฉลี่ย ดังนั้น การออกแบบทางแยกในรูปแบบปรับระดับชั้นเป็นไปเพื่อหาค่าความยากของข้อสอบให้เหมาะสมกับผู้สอบ ชั้นคลังข้อสอบที่ผู้สอบตอบข้อสอบทั้งหมดถูกต้องอ้างเป็นชั้นพื้นฐาน (Basal Stratum) และชั้นคลังข้อสอบที่ผู้สอบตอบข้อสอบผิดทั้งหมดอ้างเป็นชั้นเพดาน (Ceiling Stratum) ความสามารถจริงของผู้สอบตกอยู่ระหว่าง 2 ชั้นนี้ โดยถือว่า ผู้สอบสามารถตอบข้อสอบได้ถูกต้องประมาณ 50 เปอร์เซ็นต์ของข้อสอบที่ทำ

ไวส์ (Weiss. 1973) อธิบายข้อดีของการทดสอบแบบปรับระดับชั้น (Stradaptive Testing) ว่าเหนือกว่าการทดสอบในกลยุทธ์สองขั้นตอนและกลยุทธ์หลายขั้นตอนในรูปแบบอื่นๆ ดังนี้

การทดสอบแบบปรับเหมาะรูปแบบปรับระดับชั้นสามารถย้อนกลับตำแหน่งการทดสอบเดิมได้ (Complete Recovery) การย้อนกลับนี้เป็นไปไม่ได้ในรูปแบบปิรามิดและเฟล็กซิเลเวล สำหรับกลยุทธ์สองขั้นตอนพบว่ามักเกิดความคลาดเคลื่อนบ่อยๆ คลังข้อสอบรูปแบบนี้สามารถให้สารสนเทศสูงสุดแก่ผู้สอบ ผู้สอบส่วนมากได้รับสัดส่วนการตอบถูกต้องทั้งหมดประมาณ .50 ของการทดสอบ

และมีกฎการยุติถูกออกแบบไว้ชัดเจนสำหรับอธิบายการเคา ผู้สอบทั้งหมดไม่จำเป็นต้องเริ่มทดสอบด้วยข้อสอบเดียวกัน และการทดสอบโดยใช้สารสนเทศเดิมของผู้สอบสามารถลดจำนวนการใช้ข้อสอบในผู้สอบแต่ละคนได้ วิธีนี้ยอมให้จำนวนข้อสอบเปลี่ยนแปลงตามผู้สอบ การทดสอบดำเนินอย่างต่อเนื่องจนได้ระดับความแม่นยำตามต้องการ ถ้าคลังข้อสอบมีจำนวนข้อสอบเพียงพอ การทดสอบจะให้ความแม่นยำเท่าเทียมกันทุกระดับความสามารถของผู้สอบ ผลการทดสอบแบบปรับเหมาะรูปแบบปรับระดับชั้นให้คะแนนคงเส้นคงวา เมื่อคะแนนมีความแน่นอนสามารถใช้ในการทำนายครั้งต่อไปได้ถูกต้อง การทดสอบแบบปรับเหมาะรูปแบบปรับระดับชั้นออกแบบมาเพื่อใช้กับเครื่องคอมพิวเตอร์ แต่ยังสามารถใช้เป็นเครื่องมือการทดสอบเฉพาะตามจุดประสงค์ได้

การพัฒนากล้องข้อสอบสำหรับการทดสอบรูปแบบปรับระดับชั้น (Weiss, 1973) มีแนวทางว่าข้อสอบต้องมีจำนวนมากเพียงพอและครอบคลุมระดับความสามารถของผู้สอบ ขนาดคลังข้อสอบขึ้นกับคุณภาพข้อสอบ และขนาดคลังข้อสอบเมื่อแบ่งชั้นแล้วจะใช้ได้ผลดีเมื่อมีข้อสอบมากกว่า 200 ข้อขึ้นไป ข้อสอบจะนำไปหาค่าความยากและค่าอำนาจจำแนกของข้อสอบจากกลุ่มผู้สอบขนาดใหญ่ ข้อสอบอำนาจจำแนกต่ำจะขจัดออกไป ต่อจากนั้น จัดคลังข้อสอบโดยแบ่งชั้นตามระดับค่าความยากในแต่ละชั้นจะได้ยอดโค้งของข้อสอบโดยไม่มีการทับซ้อนค่าความยากของข้อสอบระหว่างชั้น จำนวนชั้นคลังข้อสอบหรือจำนวนยอดโค้งข้อสอบขึ้นอยู่กับขนาดคลังข้อสอบครั้งแรก จำนวนชั้นยิ่งมากจะให้ความแม่นยำสำหรับกลุ่มผู้สอบระดับความสามารถต่างๆ อย่างเท่าเทียมกัน จำนวนชั้นควรมีประมาณ 9-10 ชั้น ครอบคลุมระดับความสามารถโดยไม่จำเป็นต้องใช้คลังข้อสอบขนาดใหญ่นัก แต่ละชั้นบรรจุข้อสอบ 10-15 ข้อ และจำนวนข้อสอบไม่จำเป็นต้องเท่ากันทุกชั้น ชั้นความยากต่ำปานกลาง ต้องการข้อสอบมากกว่าชั้นความยากสูง ข้อสอบในแต่ละชั้นถูกจัดลำดับค่าอำนาจจำแนกข้อสอบจากมากไปน้อย ระยะแรกของการทดสอบเมื่อใช้ข้อสอบจำนวนน้อย การจัดค่าอำนาจจำแนกข้อสอบเป็นช่วงกว้างจะเหมาะสมกว่า ในทางตรงข้าม ระยะท้ายของการทดสอบ เมื่อค่าประมาณความสามารถของผู้สอบใกล้เคียงความจริง ควรจัดค่าอำนาจจำแนกข้อสอบเป็นช่วงแคบ การทดสอบจึงจะได้ประสิทธิภาพดี

สรุป กลยุทธ์ของการทดสอบแบบปรับเหมาะจะแสดงทิศทางการทดสอบผ่านคลังข้อสอบขึ้นอยู่กับการจัดลำดับของแต่ละกลยุทธ์ ซึ่งมีลักษณะแตกต่างกันประกอบด้วย กลยุทธ์สองขั้นตอน กลยุทธ์หลายขั้นตอน ได้แก่ รูปแบบทางแยกคงที่และรูปแบบทางแยกแปรผัน รูปแบบปรับระดับชั้นเป็นรูปแบบหนึ่งในรูปแบบทางแยกคงที่ โดยคลังข้อสอบถูกแบ่งชั้นและแบบทดสอบถูกแบ่งเป็นกลุ่มข้อสอบย่อยจัดเข้าในแต่ละชั้นของคลังข้อสอบสอดคล้องกับค่าความยาก ชั้นแรกบรรจุด้วยข้อสอบค่าความยากต่ำ ชั้นถัดไปบรรจุด้วยข้อสอบความยากเพิ่มขึ้นเรื่อยๆ จนถึงชั้นสุดท้าย การทดสอบเริ่มต้นด้วยนำค่าประมาณความสามารถของผู้สอบจากสารสนเทศเบื้องต้นหรือข้อมูลส่วนตัว นำไปกำหนดจุด

นำเข้า ผู้สอบค่าประมาณความสามารถต่ำจะเริ่มต้นทดสอบด้วยข้อสอบความยากต่ำ ผู้สอบค่าประมาณความสามารถสูงจะเริ่มทดสอบด้วยข้อสอบความยากสูง ผลการตอบข้อสอบนำไปประมาณค่าความสามารถและกำหนดทางแยกตามกฎที่เลือกใช้ต่อไป เมื่อมาถึงกฎการยุติจึงสิ้นสุดการทดสอบ

3. ความก้าวหน้าของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์

(Advances of Computerized Adaptive Testing)

การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ (Computerized Adaptive Testing : CAT) เป็นการทดสอบที่ใช้ข้อสอบแตกต่างกันจัดให้กับผู้สอบแต่ละคน โดยพิจารณาจากความสามารถภายในของผู้สอบที่ต้องการวัดและดำเนินการทดสอบด้วยเครื่องคอมพิวเตอร์ ดังนั้น วัตถุประสงค์ของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์จึงเป็นการสร้างแบบทดสอบให้เหมาะสมกับผู้สอบแต่ละคน (Meijer; & Nering. 1999) และอาศัยคอมพิวเตอร์ช่วยดำเนินขั้นตอนการทดสอบต่างๆ ปัจจุบัน การศึกษาด้านการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์มีความก้าวหน้าอย่างต่อเนื่อง องค์ประกอบสำคัญและความก้าวหน้าของการศึกษาการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์มีรายละเอียดดังนี้

3.1 การกำหนดรูปแบบการตอบข้อสอบ (Item Response Model)

การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ใช้หลักการตามทฤษฎีการตอบข้อสอบ รูปแบบ 1 2 และ 3 พารามิเตอร์เป็นไปตามข้อตกลงของความเป็นมิติเดียว (Unidimension) ความเป็นอิสระในการตอบข้อสอบ (Local Independence) และโค้งลักษณะข้อสอบ (Item Characteristic Curve) ก่อนพิจารณากำหนดรูปแบบการตอบข้อสอบ ความเป็นมิติเดียวของแบบทดสอบใดๆ มีความหมายเหมือนจำนวนของลักษณะแฝง (Latent Trait) หรือ มิติ (Dimension) ของลักษณะภายในหรือความสามารถที่ต้องการทดสอบ ในขณะเดียวกัน ต้องการความเป็นอิสระในการตอบข้อสอบเพื่อให้การตอบข้อสอบข้อหนึ่งไม่มีความเกี่ยวข้องกับข้อสอบข้ออื่น รักษาพารามิเตอร์ของข้อสอบให้คงที่ไม่ว่าข้อสอบนั้นอยู่ตำแหน่งใดบนแบบทดสอบ ผลที่ได้รับคือ การประมาณค่าความสามารถของผู้สอบเป็นอิสระจากชุดข้อสอบ แม้ว่าผู้สอบแต่ละคนจะได้รับชุดข้อสอบแตกต่างกัน ความยากต่างกัน แต่สามารถเปรียบเทียบค่าประมาณความสามารถของผู้สอบเหล่านี้ได้ (Chalhoub-Deville; Alcaya; & Lozier. 1996 : Online) และถ้าความเป็นอิสระในการตอบข้อสอบอยู่บนเงื่อนไขบนความเป็นมิติเดียว จะสามารถอธิบายความแตกต่างของรูปแบบการตอบข้อสอบระหว่างผู้สอบได้ (Lord; & Novick. 1968 : 358-394)

การกำหนดรูปแบบการตอบข้อสอบว่าจะเลือกใช้รูปแบบโลจิสติก 1 2 และ 3 พารามิเตอร์ใดนั้น มีคำแนะนำดังนี้

กรีนและคณะ (Green; & et al. 1984) แนะนำว่า การใช้รูปแบบโลจิสติก 3 พารามิเตอร์ถือว่ามีความเหมาะสมกับการทดสอบแบบปรับเหมาะมากกว่าใช้รูปแบบโลจิสติก 1 หรือ 2 พารามิเตอร์ เพราะรูปแบบจะไม่สั่นไหว (Fit) กับข้อสอบหลายตัวเลือกถ้าการเดาเท่ากับศูนย์ ไวนอร์และมีสวาลี (Wainer; & Mislevy. 1990 : 72) กล่าวว่า ในทางปฏิบัติถึงแม้ว่าการเดาอาจไม่จำเป็นเท่าไรนักในระหว่างการดำเนินการทดสอบ เพราะเป็นไปได้ยากที่ผู้สอบจะได้รับข้อสอบความยากไม่เหมาะสมกับความสามารถของตนเอง แต่การใช้ค่าการเดาร่วมด้วยเป็นสิ่งจำเป็นด้วยเหตุผลว่า ในระยะเริ่มแรกของการทดสอบ สารสนเทศของผู้สอบอาจไม่เพียงพอที่จะทำให้การประมาณค่าความสามารถถูกต้องได้ โอกาสเกิดการเดาจากการจัดข้อสอบไม่เหมาะสมในระยะเริ่มแรกจึงเป็นไปได้ และค่าการเดาจำเป็นต้องใช้เพื่อประสิทธิภาพในการคำนวณ อย่างไรก็ตาม การใช้รูปแบบโลจิสติก 3 พารามิเตอร์ก็พบปัญหาได้คือพารามิเตอร์ค่าการเดาประมาณค่าได้ยาก เพื่อแก้ปัญหานี้ ชาฮูบเดวีล อลคายา และโลซีเออร์ (Chalhoub-Deville; Alcaya; & Lozier. 1996 : Online) แนะนำให้กำหนดค่าการเดาคงที่เท่ากับ 1 และหารด้วยจำนวนตัวเลือก เช่น ข้อสอบมีตัวเลือก 5 ตัวเลือก กำหนดค่าการเดาจะถูกกำหนดไว้ที่ .20 อย่างไรก็ตาม วิธีนี้ช่วยแก้ปัญหาคงที่หน้าได้ในระดับหนึ่ง แต่ยังไม่สามารถตอบปัญหาในการระบุค่าการเดาแน่นอนได้

การเลือกใช้รูปแบบโลจิสติก 2 พารามิเตอร์ ชาฮูบเดวีล อลคายา และโลซีเออร์ (Chalhoub-Deville; Alcaya; & Lozier. 1996 : Online) ให้ความเห็นว่า การใช้รูปแบบโลจิสติก 2 พารามิเตอร์ซึ่งยอมให้ค่าอำนาจจำแนกเปลี่ยนแปลงไปมาได้ สามารถให้สารสนเทศมากพอในการเลือกใช้ข้อสอบข้อต่อไป

การเลือกใช้รูปแบบโลจิสติก 1 พารามิเตอร์ มีข้อดีในเชิงแนวคิดและปฏิบัติมากกว่าการเลือกใช้รูปแบบโลจิสติก 2 และ 3 พารามิเตอร์ รวมทั้งเพิ่มความถูกต้องในการประมาณค่าความสามารถลดค่าใช้จ่าย ลดขนาดกลุ่มตัวอย่าง และง่ายในการแปลผล (Chalhoub-Deville; Alcaya; & Lozier. 1996 : Online)

เมื่อพิจารณาผลงานวิจัยการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ด้านต่างๆ พบว่า มีการเลือกใช้ทั้งรูปแบบโลจิสติก 3 พารามิเตอร์ [(Gialluca; & Weiss. 1979) ; (Stocking. 1997) ; (Nicewander; & Thomasson. 1999) ; (Ban; & et al. 2001) ; (Vispoel; Hendrickson; & Bleiler. 2000) ; (Zwick; & Thayer. 2002)] รูปแบบโลจิสติก 2 พารามิเตอร์ [(van der Linden. 1998, 1999) ; (van Rijn; & et al. 2002) ; (Ackerman. 1991) ; (Harwell; & Janosky. 1991) ; (Reise; & Yu. 1990)] และรูปแบบโลจิสติก 1 พารามิเตอร์ [(Lunz; Bergstrom; & Wright. 1992) ; (Lunz; & Bergstrom. 1994) ; (Blais; & Raiche. 2002 : Online) ; (Veerkamp; & Glas. 2000)] ข้อสังเกตดังกล่าวให้ความสอดคล้องกับข้อสรุปของแฮมเบตันและสวามินาธาน (Hambleton; & Swaminathan. 1985 : 307-309)

กล่าวว่า การเลือกใช้รูปแบบการตอบข้อสอบรูปแบบใดนั้น ความสำคัญอยู่ที่การพิจารณาลักษณะของข้อมูลและภาวะสนิทรูป (Fit) ระหว่างข้อมูลการตอบข้อสอบและรูปแบบที่เลือกใช้เป็นหลักสำคัญ

สรุป การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามทฤษฎีการตอบข้อสอบ รูปแบบการตอบข้อสอบ 1 2 และ 3 พารามิเตอร์จะต้องเป็นไปตามข้อตกลง ความเป็นมิติเดียว ความเป็นอิสระในการตอบข้อสอบและไค้ลักษณะข้อสอบก่อน ส่วนการเลือกใช้รูปแบบโลจิสติก 1 2 และ 3 พารามิเตอร์รูปแบบใดนั้น ควรพิจารณาลักษณะของข้อมูลและภาวะสนิทรูป (Fit) ระหว่างข้อมูลการตอบข้อสอบและรูปแบบที่เลือกใช้

3.2 คลังข้อสอบ (Item Pool)

คลังข้อสอบ (Item Pool หรือ Question Banks, Item Collection, Item Reservoirs, Test Item Libraries) เป็นแหล่งรวบรวมข้อสอบหรือคำถามแบบทดสอบ การทดสอบด้วยแบบทดสอบประเพณีนิยม ผู้สอบทั้งหมดจะได้รับข้อสอบชุดเดียวกัน แต่สำหรับการทดสอบแบบปรับเหมาะผู้สอบจะได้รับข้อสอบเหมาะสมกับความสามารถของผู้สอบแต่ละคน คลังข้อสอบจึงเป็นเสมือนแหล่งรวบรวมชุดแบบทดสอบจำนวนมากและมีความเป็นคู่ขนานกัน ข้อสอบบรรจุในคลังข้อสอบมีลักษณะ ค่าความยากกระจายเต็มระดับความสามารถของประชากรผู้สอบ ค่าอำนาจจำแนกสูงเพื่อให้ประสิทธิภาพการทดสอบสูงสุด และค่าการเดาเข้าใกล้หรือเท่ากับศูนย์ ข้อสอบควรมีจำนวนมากเพียงพอและกระจายทุกระดับความสามารถของผู้สอบ ดังนั้น ต้องใช้ข้อสอบจำนวนมากสำหรับสร้างคลังข้อสอบ ทำให้คลังข้อสอบสำหรับการทดสอบแบบปรับเหมาะมีขนาดใหญ่กว่าการทดสอบประเพณีนิยม (Flaugher, 1990 : 41-62)

การสร้างข้อสอบจำนวนมากใช้สำหรับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ขั้นตอนดังนี้ ได้แก่ 1. สร้างข้อสอบใหม่ให้มีจำนวนเพียงพอในแต่ละเนื้อหาเน้นที่ครอบคลุมเนื้อหาหลักและกฎเบื้องต้นในการเขียนข้อสอบ 2. ตรวจสอบคุณภาพของข้อสอบเพื่อให้ข้อสอบมีคุณภาพสูงสุดเท่าที่เป็นไปได้ 3. ทำการทดสอบเบื้องต้นเพื่อตรวจสอบข้อบกพร่องของข้อสอบใหม่ 4. ทำการเลือกชุดย่อยของข้อสอบใหม่ ตอบข้อสอบ และวิเคราะห์ผลสอบตามทฤษฎีการทดสอบแบบมาตรฐานเดิมและทฤษฎีการตอบข้อสอบ เพื่อจัดข้อสอบไม่เหมาะสมออกไปหรือทำการปรับปรุงข้อสอบใหม่ 5. เปรียบเทียบสมมูลเนื้อหาของคลังข้อสอบด้วยแบบทดสอบตามทฤษฎีการทดสอบแบบมาตรฐานเดิม เพื่ออ้างอิงสมมูลเนื้อหาของคลังข้อสอบในการทดสอบแบบปรับเหมาะ ประเมินระบบการทดสอบด้วยการจำลองพฤติกรรมกรรมการตอบข้อสอบของผู้สอบระดับความสามารถต่างๆ กัน คะแนนทั้ง 2 การทดสอบนี้ต้องเปรียบเทียบกันได้ และ 6. นำแสดงข้อสอบตามแนวทางการจัดการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ต่อไป (Flaugher, 1990 : 43)

วิธีเลือกข้อสอบของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ให้สามารถเปรียบเทียบได้กับการทดสอบใช้กระดาษและดินสอ มีจำเป็นต้องจัดสมดุลเนื้อหาเพื่อใช้เป็นตัวแทนของแบบทดสอบ ถ้าการทดสอบใช้กระดาษและดินสอเป็นพหุมติ แต่การทดสอบแบบปรับเหมาะข้อตกลงกำหนดว่า คลังข้อสอบต้องเป็นมิติเดียว การแก้ปัญหานี้อาจใช้วิธีการแบ่งคลังข้อสอบใหญ่เป็นคลังข้อสอบย่อย คลังข้อสอบย่อยแต่ละคลังแทนแต่ละมิติ คะแนนแต่ละส่วนมาจากแต่ละมิติหรือแต่ละเนื้อหา ผลการทดสอบหาโดยคำนวณรวมน้ำหนักของคะแนนต่างๆ ในแต่ละมิติ แต่วิธีนี้อาจยุ่งยากในการทดสอบและการรายงานผลตามมา (Flaugher, 1990 : 60) กรีนและคณะ (Green; & et al. 1984) ให้เหตุผลสำหรับใช้คลังข้อสอบมิติเดียวว่า รูปแบบการทดสอบแบบปรับเหมาะที่เหมาะสมเกิดได้ก็ต่อเมื่อข้อสอบมีมิติเด่นมิติเดียว (Dominant Dimension) ข้อสอบมิติเด่น 2 มิติหรือข้อสอบพหุมติมีแนวโน้มทำให้ค่าอำนาจจำแนกต่ำ และการเลือกข้อสอบอำนาจจำแนกสูงมีแนวโน้มให้ความเป็นมิติเดียวโดยตัวของมันเองอยู่แล้ว

การสร้างคลังข้อสอบมากกว่าหนึ่งคลังข้อสอบหรือแบ่งคลังข้อสอบใหญ่เป็นคลังข้อสอบย่อยสามารถสุ่มเลือกคลังข้อสอบและหมุนเวียนการใช้ชุดแบบทดสอบไปตามคลังข้อสอบต่างๆ ได้ (Wang; & Kolen, 2001 ; citing Stocking, 1988 ; 1994) การสร้างคลังข้อสอบจำนวนมากต้องทำการศึกษาเปรียบเทียบระหว่างคลังข้อสอบก่อนนำไปใช้ พิจารณาปัจจัยจากต่างๆ ดังนี้ 1. ขนาดของคลังข้อสอบ 2. จำนวนคลังข้อสอบเมื่อใช้พร้อมๆ กัน 3. การทับซ้อนของคลังข้อสอบ 4. วิธีเข้าสู่แต่ละคลังข้อสอบ 5. การหมุนเวียนใช้คลังข้อสอบและเกณฑ์การออกจากคลังข้อสอบ 6. กฎการนำข้อสอบกลับมาใช้ใหม่ 7. การทดสอบข้อสอบก่อนนำมาใช้ วิธีการคำนวณข้อสอบ และวิธีการควบคุมช่วงความสามารถบนเส้นต่อเนื่อง และ 8. วิธีการเลือกข้อสอบเพื่อให้ครอบคลุมแต่ละเนื้อหาวิชา ควบคุมการแสดงข้อสอบ อัตราการทับซ้อนและอื่นๆ (Wang; & Kolen, 2001)

ปฏิบัติการรวมแบบทดสอบในคลังข้อสอบย่อยเข้าเป็นแบบทดสอบใหญ่ อาจประยุกต์จากผลการศึกษาของเวนเดอลินเดน (van der Linden, 1998) ศึกษาเปรียบเทียบวิธีรวมข้อสอบเข้าแบบทดสอบ 4 วิธี (1. Heuristic-Base Test 2. 0-1 Linear Programming 3. Network-Float Programming และ 4. Optimal Design Approach) ขณะที่สต็อกกิงและสเวนสัน (Stocking; & Swanson, 1998) ประยุกต์วิธีการออกแบบเหมาะสม (Optimal Design) เพื่อใช้สำหรับรวบรวมแบบทดสอบแบบปรับเหมาะเข้าไว้ด้วยกัน การศึกษาเหล่านี้อยู่ในระยะเริ่มต้นอาจนำไปใช้กับการรวมคลังข้อสอบสำหรับการทดสอบแบบปรับเหมาะในโอกาสต่อไป

ขนาดคลังข้อสอบมีส่วนช่วยเพิ่มประสิทธิภาพการทดสอบ ไวส์ (Weiss, 1985) แนะนำว่าการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ จะให้ประสิทธิภาพเมื่อคลังข้อสอบมีจำนวนข้อสอบมากเพียงพอสำหรับจำแนกผู้สอบ การศึกษาของซิงและแฮมเบิลตัน (Xing; & Hambleton, 2004) พบว่าคลังข้อสอบขนาดใหญ่ทำให้สารสนเทศแบบทดสอบเพิ่มขึ้นตามไปด้วย สอดคล้องกับความคิดเห็นของ

ลอร์ด (Lord, 1980 : 150-161) ซึ่งอ้างถึงการศึกษาในสถานการณ์จำลอง เมื่อใช้ความยาวแบบทดสอบ 25 ข้อ ขนาดคลังข้อสอบ 363 ข้อ และ 183 ข้อ พบว่า ขนาดคลังข้อสอบ 363 ข้อสามารถให้สารสนเทศได้มากกว่า เป็นเช่นนี้เพราะคลังข้อสอบขนาดใหญ่มีโอกาสเลือกข้อสอบดีที่สุดจัดให้ผู้สอบได้มากกว่าคลังข้อสอบขนาดเล็ก

การกำหนดขนาดคลังข้อสอบพิจารณาปัจจัยหลายอย่าง เชาฮูบคิวิลี อัลคายา และโลซีเออร์ (Chalhoub-Deville; Alcaya; & Lozier, 1996 : Online ; citing Stocking, 1994) กล่าวถึงปัจจัยในการออกแบบคลังข้อสอบที่ส่งผลต่อขนาดคลังข้อสอบ เช่น วิธีการเลือกข้อสอบ การควบคุมเนื้อหา มาตรฐานทางจิตวิทยา กฎการยุติ การทับซ้อนข้อสอบ การให้คะแนนแบบทดสอบ ความเป็นคู่ขนานกับแบบทดสอบใช้กระดาษและดินสอ เป็นต้น เบิร์กสตรอมและลูนซ์ (Bergstrom; & Lunz, 1999 : 67-91 ; citing Way, 1997) อ้างถึงลักษณะแบบทดสอบสามารถนำไปประมาณขนาดคลังข้อสอบได้ว่า 1. ความยาวแบบทดสอบต่อขนาดคลังข้อสอบเท่ากับ 1 ต่อ 6 - 8 2. ใช้ข้อสอบเพียง 10-15 เปอร์เซนต์ของประชากรผู้สอบ 3. อัตราการทับซ้อนข้อสอบที่จัดให้แก่ผู้สอบแต่ละคู่ต้องไม่มากเกินไป 15 - 20 เปอร์เซนต์ และ 4. เปอร์เซนต์การทับซ้อนข้อสอบสูงสุดของแบบทดสอบแบบปรับเหมาะ 2 ชุดไม่เกิน 40 เปอร์เซนต์ ระหว่างผู้สอบความสามารถระดับเดียวกัน

เบิร์กสตรอมและลูนซ์อ้างถึงจำนวนผู้สอบถ้าน้อยกว่า 1,000 คน ขนาดคลังข้อสอบอย่างน้อยที่สุดควรประมาณ 500 ข้อ ขนาดคลังข้อสอบโดยทั่วไปประมาณ 600 ถึง 800 ข้อ และต้องกระจายครอบคลุมเนื้อหาข้อสอบ (Bergstrom; & Lunz, 1999 : 67-91 ; citing Stahl; & Lunz, 1993) ซีเกล (Segall, 2003? : Online) เสนอว่า ขนาดคลังข้อสอบควรมีขนาดประมาณ 6 - 8 เท่าของความยาวแบบทดสอบ

สรุป คลังข้อสอบเป็นแหล่งรวบรวมข้อสอบ จำนวนข้อสอบต้องมีจำนวนมากเพียงพอกับทุกระดับความสามารถของผู้สอบ ข้อสอบในคลังข้อสอบควรมีค่าความยากกระจายเต็มระดับความสามารถของประชากรผู้สอบ ค่าอำนาจจำแนกสูง และค่าการเดาเข้าใกล้หรือเท่ากับศูนย์ แบบทดสอบของการทดสอบแบบปรับเหมาะสำหรับผู้สอบแต่ละคนต้องมีความเป็นคู่ขนานเปรียบเทียบกันได้ คลังข้อสอบควรมีความเป็นมิติเดียว ถ้าคลังข้อสอบมีหลายมิติอาจแบ่งคลังข้อสอบใหญ่เป็นคลังข้อสอบย่อยหลายๆ คลังข้อสอบแทนแต่ละมิติ คลังข้อสอบขนาดใหญ่จะให้สารสนเทศแบบทดสอบได้ดีกว่าคลังข้อสอบขนาดเล็ก การกำหนดขนาดคลังข้อสอบควรพิจารณาปัจจัยต่างๆ นำมาพิจารณาออกแบบคลังข้อสอบ

3.3 ความยาวแบบทดสอบ (Test Length)

ทฤษฎีการตอบข้อสอบให้ความสำคัญกับสารสนเทศแบบทดสอบ สารสนเทศแบบทดสอบบนระดับความสามารถหนึ่ง ได้รับมาจากผลรวมสารสนเทศของข้อสอบทุกข้อในแบบทดสอบบนระดับ

ความสามารถนั้นๆ ดังนั้น เมื่อความยาวแบบทดสอบเพิ่มขึ้นสารสนเทศแบบทดสอบจะเพิ่มขึ้นตามไปด้วย หลังจากการตอบข้อสอบแต่ละข้อ ความคลาดเคลื่อนของการประมาณค่าความสามารถของผู้สอบจะลดลง จนค่าประมาณความสามารถเข้าใกล้ค่าความสามารถจริงของผู้สอบมากที่สุด (Hambleton; & Swaminathan, 1985 : 101-124) ไวน์เนอร์และมิสวาลี (Wainer; & Mislevy, 1990 : 77) เสนอว่า ในทางปฏิบัติการทดสอบแบบปรับเหมาะ ความถูกต้องของการประมาณค่าความสามารถของผู้สอบทำได้โดยการเพิ่มความยาวแบบทดสอบ และเพิ่มข้อสอบความยากเหมาะสมเข้าไป ข้อเสนอแนะนี้ตรงกับการศึกษาของ ลุง ชาง และ ฮัว (Leung; Chang; & Hau, 2003 : Online) พบว่า เมื่อความยาวแบบทดสอบเพิ่มขึ้นผลตามมาก็คือ ความสัมพันธ์ของค่าประมาณความสามารถและค่าความสามารถจริงของผู้สอบสูงขึ้นตามกัน ความคลาดเคลื่อนของการประมาณค่าความสามารถของผู้สอบค่อยๆ ลดลง ดังนั้น การเพิ่มความยาวแบบทดสอบจะเพิ่มประสิทธิภาพการประมาณค่าความสามารถของผู้สอบสูงตามไปด้วย การศึกษาของ ลุง ชาง และฮัว (Leung; Chang; & Hau, 2003 : Online) รายงานผลข้างเคียงของการเพิ่มความยาวแบบทดสอบว่า ทำให้การแสดงข้อสอบมากเกินไปอัตราการแสดงข้อสอบกำหนด และให้ค่าความลำเอียงเพิ่มขึ้นแต่ไม่มากนัก ในขณะที่ ทิสเซน (Thissen, 1990 : 169) รายงานผลความยาวแบบทดสอบที่ไม่เหมาะสมอาจให้สารสนเทศแบบทดสอบมีลักษณะเป็นโค้งแบน ความยาวแบบทดสอบจึงเป็นสิ่งสำคัญมีผลต่อประสิทธิภาพการประมาณค่าความสามารถของผู้สอบ และสมมูลการใช้ข้อสอบในคลังข้อสอบ การออกแบบความยาวแบบทดสอบควรพิจารณาปัจจัยเกี่ยวข้องต่างๆ ในการทดสอบ อาจกำหนดความยาวแบบทดสอบเบื้องต้นจากข้อเสนอของซีเกลให้คลังข้อสอบมีขนาดประมาณ 6 - 8 เท่าของความยาวแบบทดสอบ (Segall, 2003? : Online)

สรุป การเพิ่มความยาวแบบทดสอบสามารถเพิ่มประสิทธิภาพการประมาณค่าความสามารถของผู้สอบ แต่ความยาวแบบทดสอบมากเกินไปอาจทำให้สารสนเทศแบบทดสอบลดลง และการแสดงข้อสอบมากเกินไปอัตราการแสดงข้อสอบกำหนด ขาดสมมูลการใช้ข้อสอบในคลังข้อสอบ การกำหนดความยาวแบบทดสอบจึงเป็นสิ่งสำคัญ ต้องพิจารณาถึงปัจจัยประกอบต่างๆ ที่เกี่ยวข้องในการออกแบบการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ร่วมไปด้วย

3.4 จุดเริ่มต้นการทดสอบ (Entry Point)

การกำหนดจุดเริ่มต้นการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์มีหลายวิธี วิธีโดยทั่วไปจะใช้ข้อสอบข้อแรกมีค่าความยากปานกลางจัดให้กับผู้สอบแต่ละคน เมื่อไม่ทราบค่าสารสนเทศเบื้องต้นหรือระดับความสามารถของผู้สอบ (Green; & et al. 1984) แสมเบลตัน สวามินาธาน และโรเจอร์ (Hambleton; Swaminathan; & Rogers, 1991) และ ฮูลิน คราสโกลว และพาร์สัน (Hulin; Drasgow; & Parsons, 1983) เสนอทางเลือกการกำหนดจุดเริ่มต้นการทดสอบโดยใช้สารสนเทศเบื้องต้นเกี่ยวกับ

ระดับความสามารถของผู้สอบ เช่น ข้อมูลการศึกษาที่ผ่านมานำมาใช้กำหนดจุดเริ่มต้นการทดสอบ วิธีนี้ประยุกต์ในการศึกษาของ ชาฮูบดิวีลลี อัลคายา และโลซีเออร์ (Chalhoub-Deville; Alcaya; & Lozier. 1996) เริ่มจากให้ผู้สอบประเมินความสามารถของตนเองก่อนกำหนดจุดเริ่มต้นการทดสอบ วิธีนี้ต้องเตรียมข้อสอบความยากระดับต่างๆ กระจายครอบคลุมค่าความสามารถของประชากรผู้สอบ และผู้สอบต้องตอบข้อสอบอย่างน้อยที่สุดคือ ตอบข้อสอบถูก 1 ข้อ และตอบข้อสอบผิด 1 ข้อ เพื่อนำไปใช้คำนวณค่าประมาณความสามารถของผู้สอบ ฟานและซุ (Fan; & Zhu. 2001 : Online ; citing Zhu; & Fan. 1999) เสนอการใช้สารสนเทศเกี่ยวกับความสามารถของผู้สอบสำหรับการเลือกและปรับจุดเริ่มต้นของผู้สอบ สามารถลดการแสดงข้อสอบรอบๆ ระดับความยากปานกลาง และเสนอรายงานว่าการทดสอบเบื้องต้นก่อนการทดสอบจริงและนำผลมากำหนดจุดเริ่มต้นในการทดสอบแบบปรับเหมาะ มีผลทำให้การใช้ความยาวแบบทดสอบสั้นลง และใช้ข้อสอบจำนวนน้อยกว่าการเริ่มต้นการทดสอบด้วยข้อสอบความยากปานกลางหรือสารสนเทศจากผลการเรียนรายวิชาและผลการเรียนเฉลี่ย (Fan; & Zhu. 2001)

ชาง และ ยิง (Chang; & Ying. 1999) เสนอการเริ่มต้นการทดสอบแบบปรับเหมาะด้วยการใช้ข้อสอบเทียม 3 ข้อ ข้อสอบข้อแรกถูกเลือกโดยมี ค่าอำนาจจำแนก (a) เท่ากับ 1 ค่าความยาก (b) สุ่มเลือกจากการแจกแจงโค้งปกติ $N(0,1)$ และค่าการเดา (c) เท่ากับ .2 ถ้าตอบข้อสอบเทียมข้อแรกถูกต้อง ข้อสอบข้อที่ 2 มีค่าความยากมากขึ้น ($b_2 = b_1 + 2$) แต่ถ้าตอบข้อสอบเทียมข้อแรกผิด ข้อสอบข้อที่ 2 มีค่าความยากลดลง ($b_2 = b_1 - 2$) ข้อสอบข้อที่ 3 ปฏิบัติคล้ายกัน เมื่อดำเนินการทดสอบครบ 3 ข้อ จะได้สารสนเทศของผู้สอบเบื้องต้นสำหรับการเริ่มต้นการทดสอบต่อไป ค่าอำนาจจำแนกและค่าความยากจะไม่เปลี่ยนแปลงระหว่างการทดสอบ ผลการตอบข้อสอบเทียมใช้ประมาณค่าความสามารถของผู้สอบโดยวิธีการประมาณค่าความเป็นไปได้สูงสุด (Maximum Likelihood Estimation : MLE) และสามารถใช่วิธีการประมาณค่าความเป็นไปได้สูงสุดตั้งแต่เริ่มต้นจนยุติการทดสอบ ถ้าผลการตอบข้อสอบเทียมถูกหมดหรือผิดหมด ต้องเริ่มทดสอบด้วยข้อสอบเทียมใหม่ตามขั้นตอนที่กล่าวมา วิธีนี้พบได้ในงานวิจัยต่างๆ (Leung; Chang; & Hua. 2002) ; (Leung; Chang; & Hua. 2003) ; (Chang; Qian; & Ying. 2001)

สรุป การกำหนดจุดเริ่มต้นการทดสอบทำโดย 1. ใช้ข้อสอบค่าความยากปานกลางจัดให้กับผู้สอบแต่ละคนเมื่อไม่ทราบค่าสารสนเทศเบื้องต้นของผู้สอบ 2. ใช้สารสนเทศเบื้องต้นเกี่ยวกับระดับความสามารถของผู้สอบหรือข้อมูลการศึกษาที่ผ่านมาใช้กำหนดจุดเริ่มต้นการทดสอบ 3. ทำการทดสอบเบื้องต้นก่อนการทดสอบจริงและนำผลมากำหนดจุดเริ่มต้นการทดสอบ 4. ใช้ข้อสอบเทียม 3 ข้อหาสารสนเทศของผู้สอบเบื้องต้นนำไปกำหนดจุดเริ่มต้นการทดสอบ วิธีการกำหนดจุดเริ่มต้นการ

ทดสอบดังกล่าวมีความเหมาะสมแตกต่างกัน การนำไปใช้ขึ้นอยู่กับกรอบแบบและเงื่อนไขการทดสอบแบบปรับเหมาะตามที่กำหนด

3.5 การเลือกข้อสอบ (Item Selection)

ประสิทธิภาพของแบบทดสอบแบบปรับเหมาะพิจารณาจาก แบบทดสอบแบบปรับเหมาะที่ให้สารสนเทศเกี่ยวกับความสามารถของผู้สอบมากที่สุดโดยใช้ข้อสอบจำนวนน้อย (Lunz; Bergstrom; & Wright, 1992) ดังนั้น การเลือกข้อสอบแต่ละข้อจึงมีความสำคัญต่อการทดสอบ ข้อสอบจะให้สารสนเทศสูงสุดเมื่อมีค่าความยาก (b) ใกล้เคียงค่าความสามารถของผู้สอบมากที่สุด ค่าอำนาจจำแนก (a) สูง ค่าการเดา (c) เข้าใกล้ศูนย์ และเลือกข้อสอบคุณสมบัติดังกล่าวจัดให้แก่ผู้สอบ ในขณะที่เดียวกัน การเลือกข้อสอบต้องคำนึงถึงการป้องกันการใช้ข้อสอบข้อใดข้อหนึ่งมากเกินไปและเพิ่มการใช้ข้อสอบที่ใช้น้อยหรือไม่ได้ใช้เลย รักษาสมดุลเนื้อหา ร่วมกับการให้ประสิทธิภาพการประมาณค่าความสามารถของผู้สอบไปพร้อมๆ กัน

เดวิส และดอร์ค (Davis; & Dodd, 2003 citing Parshall; & et al. 1998) สรุปจุดมุ่งหมายการเลือกข้อสอบว่า เพื่อให้ความแม่นยำในการประมาณค่าความสามารถของผู้สอบมากที่สุด การรักษาความปลอดภัยของข้อสอบ และรักษาสมดุลเนื้อหาแบบทดสอบพร้อมกันด้วย

วิธีการเลือกข้อสอบแต่ละวิธีให้คุณลักษณะแตกต่างกัน ได้แก่

การเลือกข้อสอบโดยใช้วิธีการสุ่มเลือกข้อสอบ (Random Item Selection) เนื่องจากวิธีการสุ่มเลือกเปิดโอกาสให้ข้อสอบทุกข้อมีโอกาสถูกเลือกเท่าเทียมกัน วิธีนี้จึงช่วยกระจายการใช้ข้อสอบในคลังข้อสอบ แต่ให้ผลลดประสิทธิภาพการประมาณค่าความสามารถของผู้สอบและไม่ให้สมดุลเนื้อหา (Eggen, 2001 : Online)

การเลือกข้อสอบโดยพิจารณาจากฟังก์ชันสารสนเทศ (Information Function) วิธีนี้สามารถให้ประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบสูง แต่ข้อสอบที่ให้สารสนเทศสูงมักนำมาใช้บ่อยเกินไป ในขณะที่ข้อสอบบางข้อเลือกใช้น้อยหรือไม่ได้เลือกเลย ข้อสอบใช้มากเกินไปทำให้ผู้สอบเกิดการเรียนรู้หรือข้อสอบเป็นที่รู้จักกันทั่วไปในกลุ่มผู้สอบ เสี่ยงต่อความปลอดภัยของข้อสอบ และสูญเสียคุณสมบัติทางจิตวิทยาการทดสอบจากความเที่ยงตรงของข้อสอบลดลง

วิธีการเลือกข้อสอบที่นิยมแพร่หลาย คือการเลือกข้อสอบวิธีสารสนเทศสูงสุด (Maximum Information Item Selection) การเลือกข้อสอบวิธีเบย์เซียน (Bayesian Item Selection) มีผู้ทำการศึกษา ดังนี้ คือ

3.5.1 การเลือกข้อสอบวิธีสารสนเทศสูงสุด (Maximum Information Item Selection)

วิธีสารสนเทศสูงสุดจะเลือกข้อสอบที่ให้สารสนเทศสูงสุด ตามค่าประมาณความสามารถของผู้สอบครั้งหลังสุด เมื่อสารสนเทศแบบทดสอบสูงขึ้น ความคลาดเคลื่อนมาตรฐานของการประมาณค่า

ความสามารถจะลดลง ทำให้ค่าประมาณความสามารถเข้าใกล้ค่าความสามารถจริงของผู้สอบมากที่สุด การเลือกข้อสอบสารสนเทศสูงสุดในการทดสอบแบบปรับเหมาะมีความแม่นยำเทียบได้กับการทดสอบที่ใช้กระดาษและดินสอแต่ใช้จำนวนข้อสอบน้อยกว่า 60 เปอร์เซ็นต์ เปรียบเทียบระหว่างวิธีสารสนเทศสูงสุดกับวิธีการสุ่มเลือกข้อสอบ พบว่า วิธีการสุ่มเลือกข้อสอบใช้ข้อสอบมากกว่า 20 เปอร์เซ็นต์ และ ยังให้ความถูกต้องในการประมาณค่าความสามารถของผู้สอบน้อยกว่าวิธีสารสนเทศสูงสุด (Eggen; & Straetmans. 2000)

ปัจจุบันการเลือกข้อสอบวิธีสารสนเทศสูงสุดประยุกต์ใช้กับการเลือกข้อสอบวิธีอื่นๆ เพื่อเพิ่มประสิทธิภาพในการเลือกข้อสอบ ได้แก่ การเลือกข้อสอบโดยใช้ค่าสารสนเทศโดยรวมเฉลี่ย (Kullback -Leibler Information : KLI) [(Chang; & Ying. 1996) ; (Eggen. 1999 ; citing Cover; & Thomas. 1991)] การเลือกข้อสอบด้วยวิธีเกณฑ์ข้างเคียงใกล้ที่สุด (Nearest-Neighbors Criterion) (Cheng; & Liou. 2003) วิธีเกณฑ์ข้างเคียงใกล้ที่สุดเป็นการรวมวิธีสารสนเทศสูงสุดของฟิชเชอร์และวิธีค่าความยากข้อสอบเหมาะสม (Optimal Item Difficulty) เข้าไว้ด้วยกัน นอกจากนี้ การศึกษาวิธีเลือกข้อสอบจากสารสนเทศเฉลี่ยสูงสุด นำไปใช้ผ่อนคลายค่าประมาณความสามารถระหว่างการเลือกข้อสอบ เช่น เกณฑ์สารสนเทศตามน้ำหนักทั่วไป (General Weight Information Criterion) เกณฑ์สารสนเทศช่วงชั้น (Interval Information Criterion) และ เกณฑ์สารสนเทศตามน้ำหนักเป็นไปได้ (Likelihood Weight Information Criterion) (Veerkamp; & Berger. 1997)

3.5.2 การเลือกข้อสอบวิธีเบย์เซียน (Bayesian Item Selection)

วิธีนี้เลือกข้อสอบที่ให้ความแปรปรวนของการประมาณค่าความสามารถของผู้สอบน้อยที่สุด การประยุกต์วิธีเบย์เซียนสำหรับใช้ในการเลือกข้อสอบ พบได้ในงานวิจัยของ เวน เดอ ลินเดน (van der Linden. 1998) [ได้แก่ Maximum Expected (Posterior Weighted) Information]

การเลือกข้อสอบทั้งวิธีสารสนเทศสูงสุดและวิธีการเลือกข้อสอบวิธีเบย์เซียน สามารถลดความคลาดเคลื่อนกำลังสองเฉลี่ยได้ ทั้ง 2 วิธีให้ผลการเลือกข้อสอบใกล้เคียงกันเมื่อความยาวแบบทดสอบมากเพียงพอ การทดสอบแบบปรับเหมาะสามารถใช้วิธีสารสนเทศสูงสุดและวิธีเบย์เซียนในการทดสอบเดียวกัน พบในผลงานวิจัยของ เวน เดอ ลินเดน (van der Linden. 1999) นอกจากนี้ การศึกษาเปรียบเทียบวิธีการเลือกข้อสอบยังสามารถพบได้ในงานวิจัยต่างๆ (Thissen; & Mislevy. 1990 : 111-114) ; (Schnipke; & Green. 1995) ; (Chen; Ankenmann; & Chang. 2000)

วิธีการเลือกข้อสอบได้รับความสนใจไม่มานานมานี้ ได้แก่ วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก (a-Stratified Method : a-STR) เสนอโดย ชาง และ ยิง (Chang; & Ying. 1999) การทดสอบเริ่มจากแบ่งคลังข้อสอบออกเป็นชั้นตามค่าอำนาจจำแนกและแบ่งแบบทดสอบเข้าในแต่ละชั้นคลังข้อสอบ ข้อสอบจะได้รับเลือกเมื่อมีค่าความยากใกล้เคียงค่าประมาณ

ความสามารถของผู้สอบครั้งหลังสุด วิธีนี้ให้ประสิทธิภาพการประมาณค่าความสามารถของผู้สอบและประสิทธิภาพสมมูลการใช้ข้อสอบในคลังข้อสอบไปพร้อมๆ กัน รายละเอียดวิธีนี้จะนำเสนอต่อไป

สรุป การเลือกข้อสอบเป็นขั้นตอนสำคัญของการทดสอบแบบปรับเหมาะ มีเป้าหมายเพื่อให้ประสิทธิภาพสูงสุดในการประมาณค่าความสามารถของผู้สอบ รักษาความปลอดภัยของข้อสอบ และสมมูลเนื้อหาของแบบทดสอบ แบบทดสอบจะทำการเลือกข้อสอบความยากใกล้เคียงกับค่าประมาณความสามารถของผู้สอบครั้งหลังสุดและจัดให้ผู้สอบ วิธีการเลือกข้อสอบนิยมใช้โดยทั่วไป ได้แก่ วิธีการเสนอแนะสูงสุด วิธีเบเซียน และวิธีการทดสอบแบบปรับเหมาะตามระดับขั้นของค่าอำนาจจำแนกซึ่งเป็นวิธีที่เริ่มศึกษาไม่นานมานี้ ทั้ง 3 วิธีมีการประยุกต์เป็นรูปแบบต่างๆ การเลือกใช้วิธีการเลือกข้อสอบวิธีใดนั้นมีทั้งข้อดีและข้อจำกัด ซึ่งผู้ใช้ควรพิจารณาเงื่อนไขต่างๆ ของการทดสอบร่วมด้วย

3.6 การประมาณค่าความสามารถ (Ability Estimation)

การปฏิบัติการทดสอบแบบปรับเหมาะตามทฤษฎีการตอบข้อสอบ ผลการตอบข้อสอบและพารามิเตอร์ข้อสอบจะนำไปคำนวณหาค่าประมาณความสามารถครั้งหลังสุดเพื่อนำไปใช้ในการเลือกข้อสอบข้อต่อไป วิธีการประมาณค่าความสามารถแบ่งเป็น 2 วิธี คือ

3.6.1 วิธีประมาณค่าความเป็นไปได้สูงสุด (Maximum Likelihood Estimation : MLE)

วิธีประมาณค่าความเป็นไปได้สูงสุด สามารถให้พารามิเตอร์ข้อสอบค่อนข้างเป็นรูปสันทัดกับรูปแบบตามทฤษฎีการตอบข้อสอบ และให้ค่าประมาณความสามารถใกล้เคียงค่าความสามารถจริงของผู้สอบมากที่สุด วิธีประมาณค่าความเป็นไปได้สูงสุดมีความมั่นคงและประสิทธิภาพ (Hambleton; & Swaminathan. 1985 : 76-91) แต่สามารถพบข้อจำกัดบางประการ (Meijer; & Nering. 1999 ; citing Lord. 1983) ได้แก่ ขาดโอกาสของฟังก์ชันความเป็นไปได้ (Likelihood) จะไม่เป็นจริงเมื่อรูปแบบคะแนนเป็นการตอบถูกทั้งหมดหรือตอบผิดทั้งหมด และค่าประมาณความสามารถจะมากเกินไปเมื่อค่าความสามารถจริงเป็นค่าทางบวก และค่าประมาณความสามารถต่ำเกินไปเมื่อค่าความสามารถจริงเป็นค่าทางลบ

การศึกษาเกี่ยวกับการประยุกต์วิธีประมาณค่าความเป็นไปได้สูงสุดแนวทางอื่นๆ สามารถพบได้ในงานวิจัย เช่น การประมาณค่าความเป็นไปได้สูงสุดตามน้ำหนัก (Weighted Maximum Likelihood Estimation : WLE) ให้ค่าประมาณความสามารถเป็นฟังก์ชันความเป็นไปได้ตามน้ำหนักสูงสุด วิธีนี้ให้ความลำเอียงน้อยกว่าวิธีประมาณค่าความเป็นไปได้สูงสุด และใช้วิธีนี้เป็นทางเลือกหนึ่งเมื่อรูปแบบคะแนนนั้นไม่สามารถใช้ได้กับวิธีประมาณค่าความเป็นไปได้สูงสุด (Meijer; & Nering. 1999 ; citing Warm. 1989)

3.6.2 วิธีประมาณค่าเบเซียน (Bayesian Estimation)

วิธีประมาณค่าเบเซียนสำหรับใช้ในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ได้แก่ การประมาณค่าภายหลังคาดหวัง (Expected A Posteriori : EAP) และ รูปแบบเบส์ (Bayes Model : BM) หรือการประมาณค่าภายหลังสูงสุด (Maximum A Posteriori : MAP) หลักการวิธีประมาณค่าเบเซียนใช้สารสนเทศจากรูปแบบการตอบข้อสอบหรือฟังก์ชันความเป็นไปได้ (Likelihood Function) และใช้สารสนเทศเดิมของผู้สอบ (Prior Information) นำไปคำนวณหาค่าประมาณความสามารถของผู้สอบ ถ้าพบว่าการประมาณค่าความเป็นไปได้ และค่าเฉลี่ยการแจกแจงค่าความสามารถเริ่มต้นมีความต่างกันมาก ค่าประมาณความสามารถจะถอยกลับไปยังค่าเฉลี่ยเบื้องต้น (Meijer; & Nering, 1999)

ในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ การประมาณค่าความสามารถของผู้สอบเพื่อเริ่มการเลือกข้อสอบ ถ้าใช้วิธีการประมาณค่าความเป็นไปได้สูงสุด (Maximum Likelihood Estimation : MLE) จะให้การประมาณค่าความสามารถไม่แน่นอนถ้าตอบข้อสอบถูกต้องหมดหรือผิดหมด จึงไม่ใช้วิธีนี้ในการเริ่มต้นการทดสอบแต่จะเลือกใช้วิธีอื่นที่ให้ค่าประมาณความสามารถแน่นอนกว่า การใช้วิธีเบเซียน เช่น การประมาณค่าภายหลังคาดหวังจำเป็นต้องใช้ข้อมูลการตอบข้อสอบที่มีอยู่ก่อนหน้านี้หรือข้อมูลพื้นฐานของผู้สอบ นำไปอ้างเป็นค่าประมาณความสามารถเริ่มต้น ซึ่งแต่ละวิธีมีข้อจำกัดบางประการทั้งสิ้น งานวิจัยบางเรื่องจึงใช้การประมาณค่าความสามารถด้วยวิธีเบเซียนในระยะเริ่มต้นหรือระหว่างการทดสอบ และใช้วิธีการประมาณค่าความเป็นไปได้สูงสุดในช่วงท้ายของการทดสอบเดียวกัน (Leung; Chang; & Hua, 2003. : Online) ; (Parshall; Harnes; & Kromrey, 2000) ; (Deng; & Ansley, 2003 : Online) ; (Ban; & et al, 2002) นอกจากนี้ ชาง และยิง (Chang; & Ying, 1999) เสนอการเริ่มต้นการทดสอบโดยใช้ข้อสอบเทียม 3 ข้อ ควบคุมผลการตอบข้อสอบไม่ให้ตอบถูกต้องหมดหรือผิดหมด ประมาณค่าความสามารถโดยใช้วิธีการประมาณค่าความเป็นไปได้สูงสุดและสามารถใช้วิธีนี้ได้ตลอดการทดสอบ

สรุป ฟังก์ชันความสามารถของผู้สอบหาได้จากผลการตอบข้อสอบและพารามิเตอร์ข้อสอบ การประมาณค่าความสามารถของผู้สอบโดยทั่วไป นิยมใช้ทั้งวิธีการประมาณค่าความเป็นไปได้สูงสุด และวิธีประมาณค่าเบเซียน ทั้ง 2 วิธีนี้ให้ประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบแต่ก็มีข้อจำกัดบางประการ ได้แก่ วิธีการประมาณค่าความเป็นไปได้สูงสุดใช้ได้จำกัดเมื่อการตอบข้อสอบถูกต้องหมดหรือผิดหมด จึงไม่ใช้วิธีนี้สำหรับการเริ่มต้นการทดสอบ วิธีประมาณค่าเบเซียนอาจให้การประมาณค่าความสามารถของผู้สอบลดลง เมื่อค่าประมาณความเป็นไปได้และค่าเฉลี่ยการแจกแจงความสามารถเบื้องต้นมีความแตกต่างกันมาก การทดสอบสามารถรวม 2 วิธีนี้เข้าไว้ใน การทดสอบเดียวกันเพื่อลดข้อจำกัดแต่ละวิธี หรืออาจใช้ข้อสอบเทียม 3 ข้อ แก้ปัญหาข้อจำกัดของวิธีการประมาณค่าความเป็นไปได้สูงสุด ให้สามารถใช้วิธีการประมาณค่าความเป็นไปได้สูงสุดตลอดการทดสอบ

3.7 การแสดงข้อสอบ (Item Exposure)

การแสดงข้อสอบเป็นการจัดข้อสอบที่ผ่านการเลือกแล้วจัดให้แก่ผู้สอบ อ้างได้ถึงวิธีการเลือกข้อสอบนั้นว่ามีประสิทธิภาพหรือไม่ การแสดงข้อสอบจำแนกเป็น 2 ประเภทใหญ่ ได้แก่

3.7.1 การไม่ควบคุมการแสดงข้อสอบ (Uncontrolling Item Exposure)

การไม่ควบคุมการแสดงข้อสอบพบได้ทั่วไปในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์เมื่อต้องการเลือกข้อสอบสารสนเทศสูงสุดเพื่อความแม่นยำในการประมาณค่าความสามารถของผู้สอบ ทำให้ข้อสอบสารสนเทศสูงบางข้อนำไปใช้บ่อยครั้งเกินไป จนเป็นรู้จักแพร่หลายในกลุ่มผู้สอบ ทำให้ผู้สอบสามารถทำคะแนนได้มากเกินความสามารถจริงของตนเอง สูญเสียคุณสมบัติทางจิตวิทยาการทดสอบและลดความเที่ยงตรงของข้อสอบ ถ้าปัญหาดังกล่าวไม่ได้รับการแก้ไข จะส่งผลกระทบต่อความปลอดภัยของข้อสอบและประสิทธิภาพการใช้ข้อสอบในคลังข้อสอบตามมา ไมเจอร์และเนอริง (Meijer; & Nering. 1999 ; citing Segall; & Moreno. 1997) เสนอแนวทางป้องกันการรู้ข้อสอบล่วงหน้า ได้แก่ การเพิ่มความยาวแบบทดสอบ การเพิ่มจำนวนชุดของแบบทดสอบหรือจำนวนคลังข้อสอบ การเพิ่มจำนวนข้อสอบในคลังข้อสอบ และการลดอัตราการแสดงข้อสอบ ซึ่งการลดอัตราการแสดงข้อสอบนี้ทำได้โดยใช้วิธีการควบคุมการแสดงข้อสอบเข้าร่วมในการทดสอบด้วย

3.7.2 การควบคุมการแสดงข้อสอบ (Controlling Item Exposure)

การควบคุมการแสดงข้อสอบเพื่อลดการใช้ข้อสอบโดยควบคุมความถี่ของข้อสอบที่จัดให้กับผู้สอบ บอยด์ (Boyd. 2003 : Online ; citing Way; & et al. 1998) รายงานถึงวิธีการควบคุมการแสดงข้อสอบ 2 ประเภท ได้แก่ วิธีการเลือกอย่างสุ่ม (Randomization Procedures) และวิธีการเลือกอย่างมีเงื่อนไข (Conditional Selection Procedures) ขณะที่ ชางและยิง (Chang; & Ying. 1996) เสนอวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับขั้นของค่าอำนาจจำแนก (a-Stratified Method : a-STR) ซึ่งเป็นวิธีเลือกข้อสอบและสามารถควบคุมการแสดงข้อสอบโดยอัตโนมัติ รายละเอียดดังนี้

วิธีการเลือกอย่างสุ่ม (Randomization Procedures)

วิธีนี้การเลือกกลุ่มข้อสอบที่เหมาะสมที่สุดจากชุดข้อสอบ แทนการเลือกข้อสอบสารสนเทศสูงสุดเพียงข้อเดียว และสุ่มเลือกข้อสอบข้อหนึ่งจากกลุ่มข้อสอบจัดให้กับผู้สอบ วิธีการเลือกอย่างสุ่มไม่รับรองว่าการแสดงข้อสอบจะควบคุมได้ตามเกณฑ์ที่ต้องการ ตัวอย่างวิธีนี้ เช่น

วิธี 5-4-3-2-1 (5-4-3-2-1 Procedure) เสนอโดย เมคบริดและมาร์ติน (Davis; & Dodd. 2003 ; citing McBride; & Martin. 1983 ; Hetter; & Sympton. 1997) วิธีนี้เน้นการเลือกข้อสอบในระยะเริ่มต้นของการทดสอบ เมื่อผู้สอบมีแนวโน้มว่าเริ่มต้นการทดสอบที่ระดับความสามารถเท่าเทียมกัน (Stocking; & Lewis. 2000 : 163-182) เริ่มจากการเลือกข้อสอบสารสนเทศสูงสุดมา 5 ข้อ ข้อสอบข้อแรกที่จัดให้สุ่มเลือกจากข้อสอบสารสนเทศสูงสุด 5 อันดับ ข้อสอบข้อที่ 2 สุ่มเลือกจาก

ข้อสอบสารสนเทศสูงสุดที่เหลืออีก 4 อันดับ และดำเนินการไปเรื่อยๆ จนถึงข้อที่ 5 ข้อดีของวิธีนี้คือทำได้ง่าย แต่มีข้อจำกัดบางประการ คือ การแสดงข้อสอบสูงโดยเฉพาะกลุ่มข้อสอบสารสนเทศสูงและไม่รับรองว่าการแสดงข้อสอบจะควบคุมได้ตามเกณฑ์ที่ต้องการ

วิธีสุ่มเลือกจากกลุ่มข้อสอบ (Randomesque Procedure) เสนอโดย คิงบูรีและซารา (Boyd. 2003 : Online ; citing Kingsbury; & Zara. 1989) วิธีนี้จะสุ่มเลือกข้อสอบเพียงข้อเดียวจากกลุ่มข้อสอบสารสนเทศสูงสุดจำนวนเท่าๆ กัน

วิธีสุ่มเลือกข้อสอบหนึ่งในสาม (Choose One of Three) เสนอโดย โรมัสสัน (Davis; & Dodd. 2003 ; citing Thomasson. 1998) ใช้วิธีสุ่มเลือกข้อสอบจากข้อสอบสารสนเทศสูงสุด 3 ข้อ ข้อสอบอีก 2 ข้อที่ไม่ได้รับเลือกจะนำกลับไปจัดชุดใหม่เพื่อทำการเลือกครั้งต่อไป

วิธีสุ่มเลือกโลจิตส์ (Within .10 Logits Procedure) เสนอโดย ลูนซ์และสเทอัล (Davis; & Dodd. 2003 ; citing Lunz; & Stahl. 1998) วิธีนี้ข้อสอบได้รับเลือกจากการสุ่มในกลุ่มข้อสอบความยากเดียวกันและจับคู่ระหว่างค่าความสามารถของผู้สอบกับค่าความยากของข้อสอบ ถ้าไม่ได้ข้อสอบความยากตามที่ต้องการ จะสุ่มเลือกข้อสอบความยากใกล้เคียงค่าความยากเป้าหมายแทน

วิธีสุ่มเลือกแบบก้าวหน้า (Progressive Procedure) เสนอโดย รีโวลตาและพอนโซดา (Revuelta; & Ponsoda. 1998) เป็นวิธีรวมการสุ่มเลือกและการเลือกจากสารสนเทศสูงสุดเข้าไว้ด้วยกัน ในระยะแรกใช้วิธีการสุ่มเลือกมากกว่าการเลือกจากสารสนเทศสูงสุด แต่เมื่อการทดสอบดำเนินต่อไป สารสนเทศจะมีผลต่อการเลือกข้อสอบมากกว่าวิธีการสุ่มเลือก เอกเกน (Eggen. 2001 : Online) ใช้วิธีนี้นำไปประยุกต์เพื่อใช้ควบคุมการแสดงผลข้อสอบที่มีการแสดงดำเนินไป

วิธีการเลือกอย่างมีเงื่อนไข (Conditional Selection Procedures)

วิธีการเลือกอย่างมีเงื่อนไขเป็นวิธีควบคุมความน่าจะเป็นของข้อสอบที่ได้รับการเลือกแล้วให้มีการแสดงผลข้อสอบเป็นไปตามกำหนด วิธีนี้รับรองว่าอัตราการแสดงผลข้อสอบสูงสุดเป็นไปตามกำหนด แต่ขั้นตอนมีความซับซ้อน และต้องศึกษาในสถานการณ์จำลองในด้านจำนวนครั้งสำหรับการใช้ปฏิบัติ เมื่อองค์ประกอบต่างๆในการทดสอบเปลี่ยนแปลงไป จำเป็นต้องศึกษาในสถานการณ์จำลองก่อนใช้ในสถานการณ์จริง มีวิธีต่างๆ ได้แก่

วิธีซิมสันและเฮตเตอร์ (Sympson-Hetter Procedure : SH) วิธีนี้เป็นที่รู้จักกันแพร่หลาย เสนอโดย ซิมสันและเฮตเตอร์ (Stocking; & Lewis. 2000 166-170 ; citing Sympson; & Hetter. 1985) ใช้ความถี่ของข้อสอบที่จัดให้กับกลุ่มผู้สอบจำนวนมากในสถานการณ์จำลอง เปรียบเทียบกับอัตราการแสดงผลข้อสอบเป้าหมายเพื่อหาพารามิเตอร์การแสดงผลข้อสอบ ขบวนการนี้ทำซ้ำๆ กันจนพารามิเตอร์การแสดงผลข้อสอบแต่ละข้อมีค่าระหว่าง 0 ถึง 1 พารามิเตอร์นี้ใช้สำหรับการเลือกข้อสอบในการทดสอบ สถานการณ์จริงโดยสร้างตัวเลขสุ่มจากการแจกแจงเหมือนกัน (0 ถึง 1) เปรียบเทียบกับพารามิเตอร์การแสดงผลข้อสอบ ถ้าพารามิเตอร์การแสดงผลข้อสอบมากกว่าตัวเลขสุ่มจะจัดข้อสอบให้ แต่ถ้าพารามิเตอร์

การแสดงข้อสอบน้อยกว่าตัวเลขผู้สอบจะทำการเลือกข้อสอบข้อใหม่แทน ข้อดีของวิธีนี้คือ ยอมให้มีการกำหนดอัตราการแสดงข้อสอบล่วงหน้า เพื่อยืนยันว่าการแจกแจงอัตราการแสดงข้อสอบควรมีลักษณะเดียวกัน วิธีนี้ยังเป็นวิธีพื้นฐานประยุกต์ไปสู่วิธีการเลือกอย่างมีเงื่อนไขอื่นๆ อีกด้วย

ข้อควรระวังสำหรับการใช้วิธีควบคุมการแสดงผลข้อสอบซิมสัน-เฮตเตอร์ ได้แก่ 1. ถ้าการแจกแจงค่าความสามารถของผู้สอบไม่เหมาะสม นำไปสู่การเคร่งครัดหรือผ่อนปรนพารามิเตอร์การแสดงผลข้อสอบมากเกินไป ดังนั้น เมื่อมีการแจกแจงค่าความสามารถของผู้สอบใหม่ จะต้องคำนวณพารามิเตอร์การแสดงผลข้อสอบใหม่ซ้ำทุกครั้ง 2. พารามิเตอร์การแสดงผลข้อสอบวิธีซิมสัน-เฮตเตอร์เป็นค่าโดยรวมของการควบคุมการแสดงผลข้อสอบ จะไม่เหมาะสมสำหรับค่าความสามารถของผู้สอบบริเวณส่วนปลายของโค้งการแจกแจงค่าความสามารถของผู้สอบ 3. วิธีซิมสัน-เฮตเตอร์ให้ผลในการลดข้อสอบที่มีอัตราการแสดงผลข้อสอบสูงเกินไป แต่ไม่มีผลต่อข้อสอบที่มีอัตราการแสดงผลข้อสอบต่ำ (Boyd, 2003 : Online ; citing Parshall; Davey; & Nering, 1998) เอกเกน (Eggen, 2001 : Online) เสนอทางแก้ไขโดยใช้วิธีซิมสัน-เฮตเตอร์ร่วมกับวิธีสุ่มเลือกแบบก้าวหน้า (Progressive Procedure) ควบคุมการแสดงผลข้อสอบไม่ให้สูงหรือต่ำเกินไป และแนะนำว่า การกำหนดอัตราการแสดงผลข้อสอบไว้ต่ำจำเป็นต้องเพิ่มจำนวนข้อสอบให้มากขึ้น

วิธีซิมสัน-เฮตเตอร์แบบมีเงื่อนไข (The Conditional Symptom-Hetter Procedure) วิธีนี้พัฒนามาจากวิธีซิมสัน-เฮตเตอร์ เสนอโดย สต็อกกิงและเลวิส (Boyd, 2003 : Online ; citing Stocking; & Lewis, 1995) พารามิเตอร์การแสดงผลข้อสอบได้จากการประมาณระดับค่าความสามารถแทนการแจกแจงค่าความสามารถของผู้สอบ ระหว่างขั้นตอนสถานการณ์จำลอง จะมีการสร้าง $n \times m$ เมทริกซ์ เมื่อแถวบน n คือจำนวนข้อสอบในคลังข้อสอบ และแถวตั้ง m คือ จำนวนจุดค่าความสามารถที่กระจายบนช่วงการแจกแจงค่าความสามารถ ดังนั้น ข้อสอบแต่ละข้อจะมีพารามิเตอร์การแสดงผลข้อสอบที่แต่ละจุดค่าความสามารถของผู้สอบ

วิธีดาเวย์และพาร์เชลล์ (Davey-Parshall Procedure) เสนอโดย ดาเวย์และพาร์เชลล์ (Boyd, 2003 : Online ; citing Davey; & Parshall, 1995) วิธีนี้เน้นการลดการทับซ้อนข้อสอบระหว่างกลุ่มผู้สอบ โดยสร้างพารามิเตอร์การแสดงผลข้อสอบจาก $n \times n$ เมทริกซ์ เมื่อ n คือ จำนวนข้อสอบในคลังข้อสอบ เงื่อนไขพารามิเตอร์ในไดอะโกนอลเมทริกซ์ (Diagonal Matrix) ควบคุมความถี่ของข้อสอบจากการเปรียบเทียบรายคู่ (Pairwise Comparisons) ในสถานการณ์จำลองเพื่อควบคุมการทับซ้อนข้อสอบ ไดอะโกนอลเมทริกซ์มีพารามิเตอร์การแสดงผลข้อสอบปฏิบัติคล้ายพารามิเตอร์การแสดงผลข้อสอบในวิธี ซิมสัน-เฮตเตอร์

วิธีโดยรวมเงื่อนไข (Triconditional Procedure) เสนอโดย พาร์เชล โฮเกอร์ตี้ และโครมเรย์ (Boyd, 2003 : Online ; citing Parshall; Hogerty; & Kromrey, 1999) วิธีนี้รวมวิธีซิมสัน-เฮตเตอร์

(Simpson-Hetter Technique) วิธีชิมสัน-เฮตเทอร์แบบมีเงื่อนไข (The Conditional Simpson-Hetter Procedure) และ วิธีดาเวย์และพาร์เชลล์ (Davey-Parshall Procedure) เข้าไว้ด้วยกัน

วิธีมัลติโนเมียลแบบไม่มีเงื่อนไข (Unconditional Multinomial Procedure) เสนอโดย สต็อกกิงและเลวิส (Boyd. 2003 : Online ; citing Stocking; & Lewis. 1995) วิธีนี้พารามิเตอร์การ แสดงข้อสอบได้จากวิธีชิมสัน-เฮตเทอร์ในช่วงระยะแรก ระยะที่ 2 ใช้รูปแบบมัลติโนเมียลในการเลือก ข้อสอบ วิธีนี้ประยุกต์ไปสู่การควบคุมการแสดงผลข้อสอบมัลติโนเมียลแบบมีเงื่อนไข (Conditional Multinomial Procedure) ลักษณะต่างๆ (Stocking; & Lewis. 1998)

วิธีจำกัดสารสนเทศสูงสุด (Restricted Maximum Information Procedure) เสนอโดย รีโวลตาและพอนโซดา (Revuelta; & Ponsoda. 1998) วิธีนี้กำหนดอัตราการแสดงผลข้อสอบสูงสุดไว้ล่วงหน้าและจำกัดจำนวนครั้งการจัดข้อสอบให้ผู้สอบในชุดแบบทดสอบทั้งหมด การเลือกข้อสอบ พิจารณาจากสารสนเทศสูงสุด ข้อสอบที่มีอัตราการแสดงผลข้อสอบครั้งหลังสุดต่ำกว่าอัตราการแสดงผล ข้อสอบสูงสุดจะนำไปรวมไว้ในคลังข้อสอบ ถ้าข้อสอบมีอัตราการแสดงผลข้อสอบมากกว่าอัตราการ ผลลัพธ์ข้อสอบสูงสุด ข้อสอบจะถูกขับออกจากคลังข้อสอบและไม่นำมาพิจารณาในการเลือกข้อสอบ

วิธีจำกัดความก้าวหน้า (The Progressive Restricted Procedure) เสนอโดย รีโวลตาและ พอนโซดา (Revuelta; & Ponsoda. 1998) โดยรวมวิธีจำกัดสารสนเทศสูงสุด (Restricted Maximum Information Procedure) และวิธีสุ่มเลือกแบบก้าวหน้า (Progressive Procedure) ไว้ด้วยกัน วิธีจำกัด สารสนเทศสูงสุดนำมาใช้ เพื่อไม่ให้ข้อสอบมีอัตราการแสดงผลข้อสอบเกินอัตราการแสดงผลข้อสอบสูงสุด ขณะที่วิธีสุ่มเลือกแบบก้าวหน้าใช้สำหรับเลือกข้อสอบให้ผู้สอบ

3.7.2.3 วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจ จำแนก (a-Stratified Method : a-STR) วิธีนี้เสนอโดย ชางและยิง (Chang; & Ying. 1996) ประยุกต์ มาจากรูปแบบปรับระดับชั้นของไวส์ (Weiss. 1973) เป็นวิธีการเลือกข้อสอบและให้ประสิทธิภาพใน การควบคุมการแสดงผลข้อสอบโดยอัตโนมัติ เริ่มจากการแบ่งชั้นคลังข้อสอบตามค่าอำนาจจำแนกของ ข้อสอบ และแบ่งแบบทดสอบเข้าในแต่ละชั้นคลังข้อสอบ ชั้นแรกบรรจุข้อสอบอำนาจจำแนกต่ำสุด ชั้นต่อไปบรรจุข้อสอบอำนาจจำแนกสูงกว่าชั้นแรก ทำเช่นนี้เรื่อยๆ จนกระทั่งชั้นสุดท้ายบรรจุข้อสอบ ค่าอำนาจจำแนกสูงสุด การทดสอบจะดำเนินจากชั้นแรกจนถึงชั้นสุดท้าย วิธีนี้กระจายการใช้ข้อสอบ ในคลังข้อสอบให้เท่าเทียมกัน รายละเอียดวิธีนี้จะนำเสนอต่อไป

วิธีขยายช่วงระดับชั้น (Enhanced Stratified Procedure) เสนอโดย ลุง ชาง และฮัว (Leung; Chang; & Hua. 2002) เป็นการประยุกต์รวมวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ตามระดับ ชั้นของค่าอำนาจจำแนกและวิธีควบคุมการแสดงผลข้อสอบชิมสัน-เฮตเทอร์เข้าไว้ด้วยกัน รายละเอียดวิธีนี้จะนำเสนอต่อไป

การแก้ปัญหาด้านความปลอดภัยของข้อสอบ และประสิทธิภาพการใช้ข้อสอบในคลังข้อสอบ นอกจากวิธีการควบคุมการแสดงผลข้อสอบแล้ว ยังสามารถใช้วิธีอื่นๆ ได้แก่ การจัดคลังข้อสอบหลายๆ คลังข้อสอบและหมุนเวียนการใช้คลังข้อสอบ (Stocking; & Swanson. 1993) ; (van der Linden. 1998) การตรวจสอบการรู้ข้อสอบล่วงหน้าทำได้โดย ศึกษารูปแบบการตอบข้อสอบว่าจะแนบเป็นรูปสันทกับ รูปแบบการตอบข้อสอบหรือไม่ (Mcleod; Lewis; & Thissen. 2003) ; (van Krimpen-Stoop; & Meijer. 2001) ; (Mcleod; & Lewis. 1999)

สรุป การแสดงผลข้อสอบเป็นการจัดข้อสอบที่ผ่านการเลือกแล้วให้แก่ผู้สอบ กรณีไม่ควบคุมการแสดงผลข้อสอบ ข้อสอบสารสนเทศสูงสุดมักได้รับเลือกและจัดให้ผู้สอบบ่อยครั้งจนรู้จักแพร่หลายในกลุ่มผู้สอบ เสี่ยงต่อความปลอดภัยของข้อสอบ ปัญหานี้แก้ไขโดยควบคุมการแสดงผลข้อสอบไม่ให้ใช้ข้อสอบบางข้อมากเกินไป และเพิ่มการใช้ข้อสอบที่มีการแสดงผลข้อสอบต่ำ วิธีการควบคุมการแสดงผลข้อสอบแบ่งเป็น วิธีการเลือกอย่างสุ่มเป็นการเลือกกลุ่มข้อสอบที่เหมาะสมที่สุดจากชุดข้อสอบและสุ่มให้ผู้สอบ วิธีนี้จะไม่รับรองการแสดงผลข้อสอบว่าเป็นไปตามอัตราการแสดงผลข้อสอบกำหนด และวิธีการเลือกอย่างมีเงื่อนไขซึ่งให้การแสดงผลข้อสอบเป็นไปตามอัตราการแสดงผลข้อสอบกำหนด วิธีอื่นๆ เช่น วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก เป็นวิธีการเลือกข้อสอบและให้ประสิทธิภาพในการควบคุมการแสดงผลข้อสอบโดยอัตโนมัติ การป้องกันความปลอดภัยของข้อสอบอาจทำได้หลายวิธี เช่น การจัดการคลังข้อสอบหลายๆ คลังข้อสอบและหมุนเวียนการใช้คลังข้อสอบ การตรวจสอบการรู้ข้อสอบล่วงหน้าของผู้สอบ เป็นต้น

3.8.1 การจัดสมดุลเนื้อหา (Content Balancing)

การจัดสมดุลเนื้อหาเป็นปัจจัยสำคัญหนึ่งซึ่งบอกถึงประสิทธิภาพการเลือกข้อสอบ โดยทั่วไปสำหรับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ถึงแม้ผู้สอบจะได้รับชุดแบบทดสอบแตกต่างกัน แต่จำเป็นต้องครอบคลุมขอบเขตเนื้อหาตามกำหนด เพื่อให้แน่ใจว่าผู้สอบแต่ละคนได้รับการประเมินด้วยแบบทดสอบคู่ขนานกัน และผลการทดสอบสามารถนำไปเปรียบเทียบกันได้ การเลือกข้อสอบมักเลือกข้อสอบสารสนเทศสูงสุดใกล้เคียงค่าประมาณความสามารถของผู้สอบ การเลือกข้อสอบวิธีนี้จะไม่ได้รับสมดุลเนื้อหาตามกำหนด ส่งผลตามาคือ ผู้สอบความสามารถน้อยหรือไม่มีความรู้ในการตอบข้อสอบตรงส่วนเนื้อหาที่ขาดไป อาจได้รับการประมาณค่าความสามารถสูงเกินความสามารถจริง ในขณะที่ผู้สอบความสามารถสูง หรือมีความรู้ในการตอบข้อสอบตรงส่วนเนื้อหาที่ขาดไปจะได้รับการประมาณค่าความสามารถต่ำกว่าความสามารถจริง ดังนั้น การเลือกข้อสอบในการทดสอบแบบปรับเหมาะนอกจากจะต้องรักษาประสิทธิภาพการประมาณค่าความสามารถของผู้สอบและควบคุมการแสดงผลข้อสอบแล้ว จำเป็นต้องทำการควบคุมสมดุลเนื้อหาไปพร้อมๆ กันด้วย การจัดสมดุลเนื้อหากล่าวได้ว่า

ไม่ใช่การวัดทางจิตวิทยาจึงไม่เกี่ยวข้องกับรูปแบบการวัดต่างๆ ถือว่า เป็นการควบคุมไม่ใช่สถิติ (Non-Statistical Constraint) หรือเรียกว่า การควบคุมเนื้อหา (Content Constraint) แนวคิดเกี่ยวกับการจัดสมดุลเนื้อหานี้ สต็อกกิงและ สเวนสัน (Stocking; & Swanson. 1993) อธิบายลักษณะของข้อสอบไม่ใช่สถิติสำหรับการจำแนกและเลือกข้อสอบ พิจารณาจาก 1. คุณสมบัติภายในข้อสอบ (Intrinsic Item Properties) กล่าวถึงลักษณะของข้อสอบด้านขอบเขตเนื้อหา 2. การควบคุมการทับซ้อน (Overlap Constraints) กล่าวถึงข้อสอบมุ่งวัดสิ่งอื่นนอกเหนือจากลักษณะที่ต้องการวัด 3. การควบคุมชุดข้อสอบ (Item Set Constraints) กล่าวถึงการยอมให้กลุ่มข้อสอบแบ่งสิ่งเร้าหรือจัดสิ่งมุ่งหมายร่วมกัน การศึกษาเทคนิคการจัดสมดุลเนื้อหาจากงานวิจัยต่างๆ พบได้ไม่มากนัก ได้แก่

การทดสอบแบบปรับเหมาะแบบควบคุมสมดุลเนื้อหา (Constrained Computerized Adaptive Testing : CCAT) นำเสนอโดย กิงบูรี และ ชาร่า (Leung; Chang; & Hau. 2003 : Online ; citing Kingsbury; & Zara. 1989) วิธีนี้เลือกข้อสอบที่เหมาะสมที่สุดจากขอบเขตเนื้อหา โดยพิจารณาอัตรา การแสดงข้อสอบครั้งสุดท้าย และห่างจากเปอร์เซ็นต์ของการจัดข้อสอบตามกำหนดมากที่สุด

การทดสอบแบบปรับเหมาะแบบควบคุมสมดุลเนื้อหาประยุกต์ (Modified Constrained Computerized Adaptive Testing : MCCAT) เสนอโดย ลุง ชาง และฮัว (Leung; Chang; & Hau. 2003 : Online ; citing Leung; Chang; & Hua. 2000) วิธีนี้ประยุกต์มาจากการทดสอบแบบปรับเหมาะแบบควบคุมสมดุลเนื้อหา เพื่อลดการคาดเดาการเรียงเนื้อหาและให้การควบคุมสมดุลเนื้อหาให้ดียิ่งขึ้น ข้อสอบเลือกมาจากขอบเขตเนื้อหาทั้งหมดที่ยังไม่ถูกใช้เพื่อจัดผลจากการเรียงลำดับ

รูปแบบมัลติโนเมียลประยุกต์ (Modified Multinomial Model : MMM) เสนอโดย เซน เอนเคนแมน และสเปย์ (Leung; Chang; & Hau. 2003 : Online ; citing Chen; Ankenmann; & Spray. 1999) วิธีการทดสอบแบบปรับเหมาะแบบควบคุมสมดุลเนื้อหาสามารถคาดเดาอันดับการเรียงเนื้อหา ดังนั้น จึงได้พัฒนารูปแบบมัลติโนเมียลประยุกต์เพื่อต้องการให้เนื้อหามีความสมดุลกัน เมื่อขอบเขตเนื้อหาถึงเปอร์เซ็นต์กำหนด จะมีการสร้างการแจกแจงมัลติโนเมียลเพื่อปรับเปอร์เซ็นต์ที่ยังไม่เต็มของขอบเขตเนื้อหาที่เหลืออยู่

การควบคุมสมดุลเนื้อหาเป็นการควบคุมไม่ใช่สถิติ (Non-Statistical Constraint) ศึกษาการเปรียบเทียบวิธีควบคุมสมดุลเนื้อหา 3 วิธี ได้แก่ การทดสอบแบบปรับเหมาะแบบควบคุมสมดุลเนื้อหา การทดสอบแบบปรับเหมาะแบบควบคุมสมดุลเนื้อหาประยุกต์ รวมทั้งรูปแบบมัลติโนเมียลประยุกต์ ภายใต้เงื่อนไขการเลือกข้อสอบวิธีสารสนเทศสูงสุด ร่วมกับการควบคุมการแสดงผลข้อสอบวิธีซิมสัน-เฮดเทอร์ พบว่า วิธีควบคุมสมดุลเนื้อหาทั้ง 3 วิธีไม่มีผลเชิงระบบต่อประสิทธิภาพของการประมาณค่าความสามารถของผู้สอบและการใช้ข้อสอบในคลังข้อสอบ (Leung; Chang; & Hau. 2003) : (Leung; Chang; & Hau. 2003 : Online)

สรุป การจัดสมมูลเนื้อหาในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ เพื่อต้องการให้กลุ่มผู้สอบได้รับชุดแบบทดสอบแตกต่างกัน แต่ได้รับเนื้อหาการทดสอบเดียวกันหรือเป็นแบบทดสอบคู่ขนานกัน สามารถเปรียบเทียบผลการทดสอบของกลุ่มผู้สอบกันได้ การจัดสมมูลเนื้อหาไม่เกี่ยวข้องกับการวัดอื่นๆ ถือว่าเป็นการควบคุมไม่ใช่สถิติ (Non-Statistical Constraint) การเลือกข้อสอบควรพิจารณาการจัดสมมูลเนื้อหาควบคู่ไปกับ การควบคุมการแสดงข้อสอบ และประสิทธิภาพการประมาณค่าความสามารถของผู้สอบด้วย

3.9 คุณภาพแบบทดสอบ (Test Quality)

แบบทดสอบตามทฤษฎีการตอบข้อสอบตรวจสอบคุณภาพได้เหมือนแบบทดสอบตามทฤษฎีการทดสอบแบบมาตรฐานเดิม โดยพิจารณาจากความเชื่อมั่นและความเที่ยงตรงของแบบทดสอบ แต่มีความแตกต่างในรายละเอียดบางประการตามแนวทางทฤษฎีการตอบข้อสอบ ดังนี้คือ

3.9.1 ความเชื่อมั่น (Reliability)

ความเชื่อมั่น (Reliability) เป็นหลักบอกคุณภาพของแบบทดสอบ มีความหมายเกี่ยวข้องกับคะแนนแบบทดสอบที่ปราศจากความคลาดเคลื่อนของการวัด (American Psychological Association, 1985) หรือความหมายตามการวัดทางจิตวิทยาว่า การทดสอบให้ผลเหมือนกันไม่ว่าจะทดสอบกี่ครั้งก็ตาม อธิบายได้ว่า เมื่อผู้สอบตอบแบบทดสอบ 2 ชุดซึ่งวัดลักษณะเดียวกัน คะแนนของผู้สอบที่ตอบแบบทดสอบทั้ง 2 ชุดนั้นจะมีความสัมพันธ์กันอย่างสมบูรณ์ แต่ถ้าพบว่าไม่สัมพันธ์กัน แสดงว่าเป็นผลมาจากมีความคลาดเคลื่อนของการวัด สัมประสิทธิ์ความเชื่อมั่นเป็นสัมประสิทธิ์ความสัมพันธ์ระหว่างแบบทดสอบ 2 ชุด ซึ่งบอกความสัมพันธ์ของขนาดความคลาดเคลื่อนของการวัดหรือการวัดความเชื่อมั่นของแบบทดสอบนั่นเอง (Thissen, 1990 : 162 ; citing Otis, 1925)

การทดสอบตามทฤษฎีการตอบข้อสอบมีข้อได้เปรียบกว่าการทดสอบตามทฤษฎีการทดสอบแบบมาตรฐานเดิม ทฤษฎีการทดสอบแบบมาตรฐานเดิมมีข้อดกลงเบื้องต้นว่าความคลาดเคลื่อนของการวัดเป็นค่าคงที่ทุกระดับค่าความสามารถของผู้สอบ สิ่งนี้ไม่เป็นจริงในสภาพทั่วไป ทิศเช่นกล่าวว่า ผู้สอบความสามารถต่ำควรมีค่าความคลาดเคลื่อนของการวัดสูง ตรงข้ามกับผู้สอบความสามารถสูงควรมีค่าความคลาดเคลื่อนต่ำ (Thissen, 1990 : 168-170) เมื่อนำทฤษฎีการตอบข้อสอบมาใช้ในการทดสอบพบว่า สามารถให้ความแม่นยำในการวัดไม่ว่าผู้สอบจะมีความสามารถระดับใด (Meijer; & Nering, 1999 ; citing Moreno; & Segall, 1997) ดังนั้น ผลการทดสอบตามทฤษฎีการตอบข้อสอบอธิบายความสามารถของผู้สอบเป็นรายบุคคลได้อย่างแม่นยำ

การทดสอบตามทฤษฎีการตอบข้อสอบให้ความแม่นยำในการวัดที่ระดับความสามารถต่างๆ ของผู้สอบ เปรียบเทียบความแม่นยำในการวัดได้กับทฤษฎีการทดสอบแบบมาตรฐานเดิม โดย

ทฤษฎีการตอบข้อสอบรายงานผลเป็นความคลาดเคลื่อนมาตรฐานของการประมาณค่าความสามารถ ในขณะที่ทฤษฎีการทดสอบแบบมาตรฐานเดิมรายงานผลเป็นความเชื่อมั่น (Thissen, 1990 : 164) เนื่องจากมีความสัมพันธ์ระหว่าง ความเชื่อมั่น (ρ) ความแปรปรวน (σ_e^2) และความคลาดเคลื่อนมาตรฐาน (σ_e) ยกตัวอย่างเช่น ถ้า $\rho = 0.9$ ผลตามมามี $\sigma_e^2 = 0.1$ และ $\sigma_e \sim 0.3$ ทั้งหมดนี้ให้ความหมายถึงความแม่นยำในการวัดและสามารถนำมาคิดกลับไปได้ (Thissen, 1990 : 164 ; citing Lord; & Novick, 1968)

ความคลาดเคลื่อนมาตรฐานเป็นหลักบอกความแม่นยำ หรือความเชื่อมั่นในการวัดตามทฤษฎีการตอบข้อสอบ ความแม่นยำของการวัดจะเปลี่ยนแปลงไปตามฟังก์ชันค่าความสามารถของผู้สอบ แสดงเป็นความคลาดเคลื่อนมาตรฐานหรือความแปรปรวนความคลาดเคลื่อน (Wainer, 1990 : 15) ; (Thissen, 1990 : 163 ; citing Samejima, 1977)

สรุป ความแม่นยำของการวัดตามทฤษฎีการตอบข้อสอบ แสดงผลเป็นความคลาดเคลื่อนมาตรฐานหรือความแปรปรวนความคลาดเคลื่อน ขณะที่ทฤษฎีการทดสอบแบบมาตรฐานเดิมความแม่นยำในการวัดแสดงผลเป็นค่าความเชื่อมั่น แต่สามารถนำไปคิดเปรียบเทียบกันได้ นอกจากนี้การทดสอบตามทฤษฎีการตอบข้อสอบให้ความแม่นยำของการวัดที่ระดับความสามารถต่างๆ ของผู้สอบ และแปรเปลี่ยนไปตามฟังก์ชันค่าความสามารถของผู้สอบ

3.9.2 ความเที่ยงตรง (Validity)

ความเที่ยงตรงมีความสำคัญในการประเมินคุณภาพแบบทดสอบ แนวคิดของความเที่ยงตรงอ้างถึงการให้คุณค่า การให้ความหมาย และการใช้ประโยชน์จากคะแนนแบบทดสอบ ความเที่ยงตรงของการทดสอบจึงเป็นกระบวนการรวบรวมหลักฐานที่จะสนับสนุนข้อสรุปบางประการ (American Psychological Association, 1985) จัดเป็นกลุ่มได้ 3 กลุ่ม ดังนี้

1. ความเที่ยงตรงเชิงเนื้อหา (Content-Related Evidence of Validity) เป็นสิ่งแสดงขนาดความสัมพันธ์ระหว่างขอบเขตเนื้อหาของแบบทดสอบกับจุดประสงค์กำหนด (American Psychological Association, 1985) ตามบริบทของการทดสอบแบบปรับเหมาะ ความเที่ยงตรงเชิงเนื้อหากระทำโดยจัดคลังข้อสอบและการเลือกข้อสอบให้ได้ขอบเขตเนื้อหาตามต้องการ การเลือกข้อสอบจะต้องไม่เปลี่ยนความเฉพาะเจาะจงของเนื้อหา แต่จะต้องให้สมดุลเนื้อหาไปพร้อมกัน เพื่อผู้สอบแต่ละคนได้รับข้อสอบเป็นเหมือนตัวแทนขอบเขตเนื้อหาเท่าเทียมกัน (Steinberg; Thissen & Wainer, 1990 : 189)

2. ความเที่ยงตรงตามเกณฑ์ (Criterion-Related Evidence of Validity) เป็นสิ่งแสดงขนาดความสัมพันธ์ระหว่างคะแนนแบบทดสอบกับเกณฑ์กำหนด (American Psychological

Association, 1985) การเลือกแบบทดสอบใดก็ตามเมื่อคาดการณ์ได้ว่า ผู้สอบที่ได้คะแนนสูงจะตอบข้อสอบได้มาก และผู้สอบที่ได้คะแนนต่ำจะตอบข้อสอบได้น้อย (Steinberg; Thissen & Wainer, 1990 : 197)

3. ความเที่ยงตรงเชิงโครงสร้าง (Construct-Related Evidence of Validity) เป็นสิ่งสนับสนุนการแปลผลคะแนนแบบทดสอบที่สร้างตามจุดประสงค์มาจากโครงสร้างสัมพันธ์กับทฤษฎี (American Psychological Association, 1985) การประเมินความเที่ยงตรงเชิงโครงสร้างของการทดสอบแบบปรับเหมาะ พิจารณาจาก การวิเคราะห์ข้อสอบในคลังข้อสอบว่าเป็นมิตินี้เดียว หรือ วัดเพียงแฟกเตอร์เดียว และการวิเคราะห์แบบทดสอบว่าความสามารถของผู้สอบสอดคล้องกับผลการตอบข้อสอบ (Steinberg; Thissen & Wainer, 1990 : 190-191)

ความเที่ยงตรงของการทดสอบแบบปรับเหมาะเปรียบเทียบกับกับการทดสอบที่ใช้กระดาษและดินสอตามทฤษฎีการทดสอบแบบมาตรฐานเดิม กรีนและคณะ (Green; & et al. 1984) กล่าวว่า เมื่อใช้ข้อสอบที่มีค่าอำนาจจำแนกสูง การทดสอบแบบปรับเหมาะจะมีความบริสุทธิ์ (จัดสิ่งที่จะทำให้ข้อสอบเปลี่ยนแปลง) และให้ความแม่นยำสูง ผลที่ตามมาคือทำให้ความเที่ยงตรงสูงโดยตัวมันเอง สิ่งกระทบต่อความเที่ยงตรงของการทดสอบแบบปรับเหมาะมาจาก 1. ถ้าข้อสอบมีความบกพร่องมากผลคะแนนอาจปราศจากความหมาย ไม่สามารถพยากรณ์ได้ 2. การทดสอบโดยคอมพิวเตอร์อาจเปลี่ยนขบวนการทางจิตวิทยาต่อการตอบข้อสอบ ความเที่ยงตรงจากการสรุปผลคะแนนเหล่านี้อาจเปลี่ยนไป 3. การเลือกข้อสอบไม่ครอบคลุมเนื้อหาอาจกระทบต่อความเที่ยงตรงจากการสรุปผลคะแนน (Wainer, 1990 : 16)

สรุป ความเที่ยงตรงบอกถึงคุณภาพของแบบทดสอบแบบปรับเหมาะ ความเที่ยงตรงเชิงเนื้อหาได้รับเมื่อผู้สอบความสามารถต่างๆ ทำข้อสอบที่มีขอบเขตเนื้อหาเท่าเทียมกัน ทำได้โดยจัดคลังข้อสอบและการเลือกข้อสอบพร้อมกับการควบคุมสมดุลเนื้อหาไปด้วย ความเที่ยงตรงตามเกณฑ์ที่ได้รับเมื่อการทำแบบทดสอบคาดการณ์ได้ว่า ผู้สอบคะแนนสูงจะตอบข้อสอบได้มากและผู้สอบคะแนนต่ำจะตอบข้อสอบได้น้อย ความเที่ยงตรงเชิงโครงสร้างได้รับเมื่อผลการตอบข้อสอบสอดคล้องกับความสามารถของผู้สอบ หรือคลังข้อสอบเป็นมิตินี้เดียวและแบบทดสอบมีความเป็นคู่ขนานกัน นอกจากนี้การทดสอบที่ใช้ข้อสอบอำนาจจำแนกสูงจะให้ความเที่ยงตรงสูงโดยตัวมันเอง

3.11 กฎการยุติ (Stopping Rule)

การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ดำเนินการทดสอบต่อเนื่อง จนมาถึงกฎการยุติจึงสิ้นสุดการทดสอบ การกำหนดกฎการยุติขึ้นอยู่กับ วัตถุประสงค์ในการทดสอบ ลักษณะคลังข้อสอบ

และการควบคุมการปฏิบัติต่างๆ (Segall. 2003? : Online) กฎการยุติโดยทั่วไปแบ่งเป็น 2 ประเภท ได้แก่

3.11.1 กฎความยาวคงที่ (Fixed-Length) เป็นการกำหนดจำนวนข้อสอบเท่าๆ กันและจัดให้กับผู้สอบแต่ละคน เมื่อการทดสอบครบตามจำนวนข้อสอบกำหนด การทดสอบจึงยุติ การศึกษาโดยใช้กฎความยาวคงที่ มักสนใจความคลาดเคลื่อนมาตรฐานของค่าประมาณความสามารถของผู้สอบเมื่อสิ้นสุดการทดสอบว่ามากน้อยเท่าไร (Thissen; & Mislevy. 1990 : 114) กฎนี้ง่ายในการใช้ สามารถทำนายอัตราการใช้ข้อสอบได้ แต่ต้องระวังการใช้กฎความยาวคงที่อาจให้ฟังก์ชันสารสนเทศเป็นโค้งลักษณะแบน ดังนั้น กฎความยาวคงที่อาจไม่ให้ความแม่นยำเท่าเทียมกันที่ระดับความสามารถต่างๆ (Thissen. 1990 : 169)

3.11.2 กฎความยาวแปรผัน (Variable-Length) เป็นการจัดข้อสอบให้อย่างต่อเนื่องจนระดับความแม่นยำของค่าประมาณความสามารถของแต่ละบุคคล มาถึงระดับความแม่นยำตามกำหนด การทดสอบจึงยุติ การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ใช้กฎความยาวแปรผันเพื่อกำหนดความยาวแบบทดสอบว่า จำนวนข้อสอบเท่าไรควรจัดให้แก่ผู้สอบระดับความสามารถต่างๆ (Thissen; & Mislevy. 1990 : 114) การใช้กฎการยุติจากค่าความคลาดเคลื่อนมาตรฐานจะให้ความแม่นยำในการวัดผู้สอบทั้งระดับความสามารถสูงและต่ำ (Meijer; & Nering. 1999) โดยทั่วไปจะใช้กฎการยุติจากค่าความคลาดเคลื่อนมาตรฐานกำหนดไว้ที่ 0.3 (Thissen. 1990 : 168)

นอกจากนี้ มีความพยายามประยุกต์ใช้กฎความยาวแปรผันในงานวิจัยต่างๆ ได้แก่ บลายสและไรคี (Blais; & Raiche. 2002 : Online) ทำการศึกษาโดยใช้รูปแบบราส์ส (Rasch Model) และวิธีประมาณค่าความสามารถภายหลังคาดหวัง (EAP) พบว่า ถ้าการแจกแจงการประมาณค่าความสามารถเป็นโค้งปกติ กฎการยุติจะสอดคล้องกับจำนวนข้อสอบอย่างน้อย 13 ข้อ ยี แวงและแบน (Yi; Wang; & Ban. 2001, Fall) ศึกษาการใช้สารสนเทศเป้าหมาย (Target Information) เป็นกฎการยุติสำหรับวิธีประมาณค่าความสามารถต่างๆ ชางและยิง (Chang; & Ying. 2001 : Online) ใช้ทฤษฎีกลุ่มตัวอย่างขนาดใหญ่ (Large Sample Theory) เพื่อกำหนดกฎความยาวแปรผัน ขณะนี้อยู่ในระหว่างการศึกษา

สรุป การกำหนดกฎการยุติขึ้นอยู่กับ วัตถุประสงค์ในการทดสอบ ลักษณะคลังข้อสอบ และการควบคุมการทดสอบ การใช้กฎความยาวคงที่โดยกำหนดจำนวนข้อสอบเท่าๆ กันจัดให้ผู้สอบมักสนใจศึกษาขนาดความคลาดเคลื่อนมาตรฐานเมื่อสิ้นสุดการทดสอบ การใช้กฎความยาวแปรผันโดยจัดข้อสอบให้อย่างต่อเนื่องจนกระทั่งระดับความแม่นยำของแต่ละผู้สอบมาถึงเกณฑ์กำหนด มักสนใจศึกษาจำนวนข้อสอบที่ใช้ไปสำหรับผู้สอบระดับความสามารถต่างๆ เนื่องจากกฎความยาวแปรผันให้ความแม่นยำในการวัดผู้สอบระดับความสามารถต่างๆ จึงมีการศึกษาและพัฒนากฎความยาวแปรผันอย่างกว้างขวางให้มีประสิทธิภาพมากขึ้น

4. การศึกษาในสถานการณ์จำลอง (Simulation Study)

การศึกษาในสถานการณ์จำลอง (Simulation Study) หรือการศึกษามอนติคาร์โล (Monte Carlo Study) เป็นการออกแบบการทดลองโดยจำลองสถานการณ์เสมือนจริงและดำเนินการทดลองเป็นระบบภายใต้เงื่อนไขกำหนด ผลการศึกษานำไปอ้างอิงได้กับการทดลองเชิงประจักษ์ จะพิจารณาใช้เมื่อไม่สามารถได้ข้อมูลจากการวิจัยทางอื่นหรือยากจะได้รับ

ฮาร์เวลและคณะ (Harwell; & et al. 1996) กล่าวถึงการศึกษาในสถานการณ์จำลองกับงานวิจัยที่ใช้ทฤษฎีการตอบข้อสอบมีลักษณะเฉพาะดังนี้ คือ

1. ใช้คำถามการวิจัย 1 คำถามหรือมากกว่า นำไปกำหนดจุดประสงค์การศึกษา
2. สร้างเงื่อนไขการศึกษาแสดงเป็นตัวแปรอิสระ
3. ออกแบบการทดลองอย่างเหมาะสม
4. สร้างข้อมูลการตอบข้อสอบตามรูปแบบพื้นฐานทางทฤษฎี
5. ประเมินค่าพารามิเตอร์และใช้การตอบข้อสอบจำลอง
6. เปรียบเทียบผลการศึกษาที่ได้
7. ขบวนการจะถูกทำซ้ำหลายครั้งในแต่ละเซลล์ (Cell) ของการออกแบบ
8. ผลการทำซ้ำหลายครั้งจะนำไปวิเคราะห์ด้วยวิธีเชิงพรรณนา (Descriptive Method)

และวิธีเชิงอ้างอิง (Inferential Method) นำผลสรุปการทดลองไปตอบคำถามการวิจัย

การศึกษาในสถานการณ์จำลองร่วมกับงานวิจัยตามทฤษฎีการตอบข้อสอบ สามารถแก้ไข ปัญหาที่ปฏิบัติไม่ได้ในการทดลองเชิงประจักษ์จากเหตุผลความซับซ้อนของการปฏิบัติ ประเมินสิ่งคุกคามความเที่ยงตรงภายในและภายนอก คำถามการวิจัยสามารถแก้ไขได้ตามแนวทางอุดมคติ นอกจากนี้ การศึกษาในสถานการณ์จำลองสามารถศึกษาปัจจัยต่างๆ ในเวลาเดียวกันหรือแยกศึกษาแต่ละปัจจัยได้ เปรียบเทียบระหว่างวิธีปฏิบัติต่างๆ และประหยัดค่าใช้จ่ายเมื่อต้องทำการศึกษาเกี่ยวข้องกับมนุษย์ ข้อจำกัดบางประการซึ่งอาจพบได้ ได้แก่ ผลการศึกษาในสถานการณ์จำลองนำไปอ้างอิงได้น้อย ถ้าสิ่งกำหนดไม่เป็นไปในสถานการณ์จริง นอกจากนี้ ผลการศึกษาอาจเปลี่ยนแปลงได้ขึ้นกับจำนวนการทำซ้ำและความแม่นยำจากการคำนวณด้วยคอมพิวเตอร์

ขั้นตอนการศึกษาในสถานการณ์จำลองของการทดสอบตามทฤษฎีการตอบข้อสอบ (Harwell; & et al. 1996) มีดังนี้คือ

1. กำหนดปัญหา ผู้วิจัยจะเป็นผู้กำหนดปัญหาที่สนใจ ตั้งคำถาม สมมุติฐานการทดสอบ และสิ่งที่ต้องการวัด โดยทั่วไปการกำหนดปัญหาได้จากการทบทวนวรรณกรรมที่ผ่านมา
2. ออกแบบการศึกษา ทั้งการวิจัยเชิงประจักษ์และการศึกษาในสถานการณ์จำลองให้ความสำคัญกับการออกแบบการทดลองในลักษณะเหมือนกัน เช่น การเลือกวิธีการออกแบบการ

ทดลองที่เหมาะสม การกำหนดตัวแปรต้นจากคำถามการวิจัย กำหนดตัวแปรตามที่เกี่ยวข้องตัวแปรต้น การเลือกจำนวนการทำซ้ำ (Replication) ซึ่งหมายถึงขนาดกลุ่มตัวอย่างเพื่อต้องการลดความแปรปรวนของกลุ่มตัวอย่าง สร้างความแน่นอนและเพิ่มความเชื่อมั่นของผลการศึกษาให้สูงขึ้น

3. การใช้โปรแกรมคอมพิวเตอร์และ/หรือการเขียนคำสั่งโปรแกรมใหม่ แต่ละส่วนของคำสั่งโปรแกรมจะต้องผ่านการประเมินความถูกต้องและความเหมาะสมในการใช้ เช่น การสร้างผลการตอบข้อสอบ การเลือกค่าผลลัพธ์ การสร้างตัวเลขสุ่ม การประมาณค่าพารามิเตอร์รูปแบบการตอบข้อสอบ เป็นต้น

4. การวิเคราะห์ผลการศึกษา การใช้วิธีวิเคราะห์เชิงพรรณนาเพียงอย่างเดียว เช่น การสรุปตารางกราฟสถิติซึ่งเป็นการแสดงผลโดยรวม จะไม่ให้ความชัดเจนในการตรวจสอบผลปฏิสัมพันธ์สำหรับประมาณขนาดและอธิบายความคลาดเคลื่อนจากข้อมูลที่สร้างขึ้นในสถานการณ์จำลอง ดังนั้นจึงควรใช้วิธีวิเคราะห์เชิงพรรณนาร่วมกับการวิเคราะห์เชิงอ้างอิงที่เหมาะสม (Harwell, 1997) เช่น การถดถอย (Regression) การวิเคราะห์ความแปรปรวน (Analysis of Variance : ANOVA) เป็นต้น

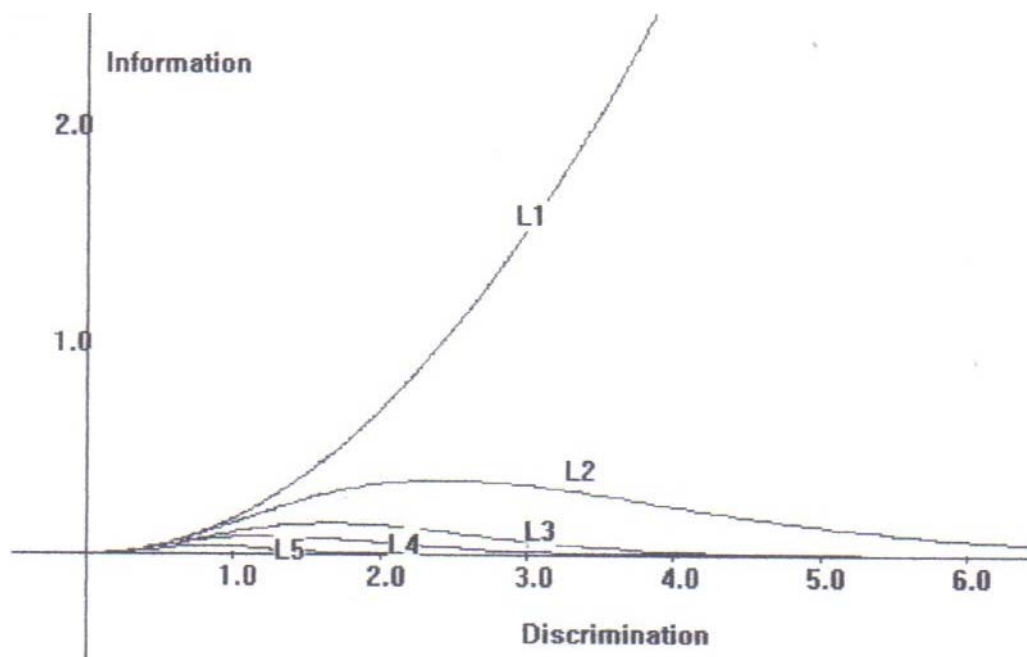
ประสิทธิภาพการทดสอบแบบปรับเหมาะขึ้นอยู่กับปัจจัยหลัก 2 ประการ คือ 1. ลักษณะคลังข้อสอบ (Item Pool Characteristics) ได้แก่ ขนาดคลังข้อสอบ การกระจายพารามิเตอร์ข้อสอบ สมดุลเนื้อหา ฯลฯ และ 2. ลักษณะวิธีกำหนด (Algorithm Characteristics) ได้แก่ กฎการหยุด การควบคุมเนื้อหา การควบคุมการแสดงผลข้อสอบ ฯลฯ ทั้งลักษณะคลังข้อสอบและวิธีกำหนดเป็นปฏิสัมพันธ์เชิงซ้อน ผลที่เกิดขึ้นจะทำให้ประสิทธิภาพในการทดสอบ นักพัฒนาแบบทดสอบจึงต้องใช้ในการศึกษาในสถานการณ์จำลองเพื่อปรับวิธีกำหนดและทดสอบคลังข้อสอบก่อนทำการทดสอบในสถานการณ์จริงโดยกำหนดค่าประมาณพารามิเตอร์ข้อสอบ บรรจุข้อสอบในคลังข้อสอบ ทำการทดสอบกับกลุ่มผู้สอบจำลองระดับความสามารถต่างๆ ผลการศึกษาจะให้ค่าเฉลี่ยและความแปรปรวนของการประมาณค่าความสามารถของผู้สอบอย่างมีเงื่อนไข ดอด คือยาลา และน็อซ (Dodd; De Ayala; & Koch, 1995) กล่าวถึงข้อดีของการใช้คลังข้อสอบจำลองว่า การกำหนดค่าพารามิเตอร์สำหรับการจัดการออกแบบระบบทำให้สามารถติดตามตัวแปรพื้นฐานที่สนใจได้ การศึกษาในสถานการณ์จำลองนี้จะยอมให้ผลการทดสอบมาจากการเปลี่ยนแปลงต่างๆ เช่น การควบคุมเนื้อหา ขนาดคลังข้อสอบ กฎการหยุด อัตราการแสดงผลข้อสอบ เป็นต้น ซึ่งจะนำมาทดสอบและเปรียบเทียบ ผลการศึกษาใช้เป็นพื้นฐานสำหรับการกำหนดลักษณะคลังข้อสอบและวิธีกำหนดที่เหมาะสมต่อไป (Segall, 2003? : Online) การศึกษาเปรียบเทียบการทดสอบแบบปรับเหมาะในสถานการณ์จำลองและในสถานการณ์การทดสอบปฏิบัติจริงเมื่อใช้รูปแบบตามทฤษฎีการตอบข้อสอบพบว่า การทดสอบทั้งสองให้ผลสอดคล้องกันไม่แตกต่างกันมากนัก (Weiss, 1985 ; citing Johnson; & Weiss, 1980) เช่นการศึกษาการเลือกข้อสอบงานวิจัยวิธีตามระดับขั้นของค่าอำนาจจำแนก (Chang; & Ying, 1999) ; (Leung; Chang; & Hau, 2002)

สรุป การศึกษาในสถานการณ์จำลองจะเป็นการออกแบบการทดลองเพื่อจำลองสถานการณ์เสมือนจริง เมื่อไม่สามารถได้ข้อมูลจากการวิจัยอื่นหรือยากจะได้รับ การทดลองดำเนินการเป็นระบบภายใต้เงื่อนไขกำหนด ผลการศึกษานำไปอ้างอิงได้กับผลการทดลองเชิงประจักษ์ ขั้นตอนเริ่มด้วยการกำหนดปัญหาการทดสอบ การออกแบบการศึกษา การใช้โปรแกรมคอมพิวเตอร์และ/หรือการเขียนคำสั่งใหม่ และการวิเคราะห์ผลการศึกษา การทดสอบแบบปรับเหมาะเมื่อศึกษาในสถานการณ์จำลองจะให้ความสำคัญกับ ลักษณะคลังข้อสอบ วิธีการกำหนด และผลจากปฏิสัมพันธ์เชิงซ้อนของการทดสอบซึ่งส่งผลต่อประสิทธิภาพการทดสอบ เมื่อเงื่อนไขอย่างใดอย่างหนึ่งเปลี่ยนไป จะต้องทำการศึกษาในสถานการณ์จำลองก่อนเพื่อใช้เป็นข้อมูลพื้นฐานสำหรับการกำหนดลักษณะคลังข้อสอบและวิธีการกำหนดอย่างเหมาะสมในการทดสอบจริงต่อไป

5. การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก

(a-Stratified Computerized Adaptive Testing : a-STR)

การทดสอบแบบปรับเหมาะสร้างบนพื้นฐานความรู้ตามทฤษฎีการตอบข้อสอบ นิยมใช้รูปแบบโลจิสติก 3 พารามิเตอร์ คือ ค่าอำนาจจำแนก (a) ค่าความยาก (b) และค่าการเดา (c) และใช้สัญลักษณ์ θ แทนค่าความสามารถของผู้สอบ ความน่าจะเป็นเมื่อผู้สอบความสามารถระดับ θ ตอบข้อสอบถูกต้องนำไปใช้กำหนดสารสนเทศข้อสอบ ลุง ชาง และฮัว (Leung; Chang; & Hau. 2002) ทำการศึกษาสารสนเทศข้อสอบต่อค่าอำนาจจำแนกข้อสอบ โดยกำหนดให้ค่า c คงที่ และค่า $\theta - b$ แตกต่างกัน แสดงเป็นโค้งสารสนเทศข้อสอบตามภาพประกอบ 3 ดังนี้ คือ



ภาพประกอบ 3 สารสนเทศข้อสอบต่อค่าอำนาจจำแนกของข้อสอบ 5 กรณี

ที่มา: Leung, Chi-Keung; Chang, Hua-Hua; & Hau, Kit-Tai. (2001, March). Making a-Stratified Computerized Adaptive Testing Design More Practical : Imposing Non-Statistical Constraints. (On Line). Available: <http://www.ied.edu.hk/scast/staff/math/mathslungck.html>. Retrieved January 10, 2004.

จากภาพประกอบ 3 เมื่อ c คงที่เท่ากับ 0.2 ความแตกต่างระหว่างค่าความสามารถของผู้สอบและค่าความยากข้อสอบเท่ากับ L_1 คือ $\theta - b = 0$, L_2 คือ $\theta - b = 1$, L_3 คือ $\theta - b = -1$, L_4 คือ $\theta - b = 2$, L_5 คือ $\theta - b = -2$ ภาพประกอบแสดงให้เห็นว่า เมื่อเลือกข้อสอบความยากเท่ากับค่าความสามารถของผู้สอบ สารสนเทศข้อสอบจะเพิ่มขึ้นจนสูงสุดตามการเพิ่มขึ้นของค่าอำนาจจำแนกข้อสอบ

พิจารณาตามเส้น L_1 ($\theta - b = 0$) เมื่อเลือกข้อสอบความยากเท่ากับค่าความสามารถของผู้สอบ สารสนเทศข้อสอบเพิ่มขึ้นจนสูงสุดตามการเพิ่มขึ้นของค่าอำนาจจำแนกข้อสอบ ดังนั้น เมื่อ

ทราบค่าความสามารถจริงของผู้สอบและเลือกข้อสอบค่าความยากใกล้เคียงหรือเท่ากับค่าความสามารถจริงของผู้สอบ ถ้าข้อสอบข้อนั้นมีค่าอำนาจจำแนกสูงด้วย จะได้รับสารสนเทศข้อสอบสูงสุด

อย่างไรก็ตาม ระยะเริ่มต้นการทดสอบ เมื่อค่าประมาณความสามารถและค่าความสามารถจริงมีความแตกต่างกันมากหรือเมื่อยังไม่ทราบค่าความสามารถจริงของผู้สอบ การเลือกข้อสอบความยากใกล้เคียงค่าประมาณความสามารถผู้สอบจะให้สารสนเทศซึ่งห่างไกลจากความเป็นจริง ถึงแม้ว่าจะใช้ข้อสอบอำนาจจำแนกสูงก็ตาม ตัวอย่าง ได้แก่ เส้น L_3 ($\theta - b = -1$) ตำแหน่งค่าอำนาจจำแนกเท่ากับ 3.0 จะให้สารสนเทศข้อสอบน้อยกว่าตำแหน่งค่าอำนาจจำแนกเท่ากับ 1.0 ความสัมพันธ์นี้บอกถึงวิธีการเลือกข้อสอบโดยใช้ข้อสอบอำนาจจำแนกสูงไม่ได้ให้สารสนเทศสูงสุดเสมอไป ดังนั้น ระยะเริ่มต้นการทดสอบไม่จำเป็นต้องใช้ข้อสอบค่าอำนาจจำแนกสูง เมื่อค่าประมาณความสามารถยังห่างไกลจากค่าความสามารถจริงของผู้สอบมาก

การทดสอบแบบปรับเหมาะโดยทั่วไป การเลือกข้อสอบจะใช้วิธีสารสนเทศสูงสุดของฟิชเชอร์ (Fisher Maximum Information) ทำการเลือกข้อสอบที่ให้สารสนเทศสูงสุดบนตำแหน่งค่าประมาณความสามารถของผู้สอบ วิธีการเลือกข้อสอบนี้ให้ประสิทธิภาพสูงในการประมาณค่าความสามารถของผู้สอบด้วยข้อสอบจำนวนน้อย แต่ปัญหาตามมาก็คือ ข้อสอบที่ให้สารสนเทศสูงสุดมักเป็นข้อสอบอำนาจจำแนกสูง ข้อสอบลักษณะนี้ผ่านการเลือกใช้บ่อยเกินไปจนเป็นที่รู้จักทั่วไปในกลุ่มผู้สอบ ส่งผลกระทบต่อความปลอดภัยของข้อสอบรวมทั้งข้อสอบอำนาจจำแนกสูงสร้างได้ยากและต้องการจำนวนมาก เพิ่มค่าใช้จ่ายสำหรับสร้างข้อสอบและบำรุงรักษาค้างข้อสอบ ในขณะเดียวกันข้อสอบอำนาจจำแนกต่ำพบว่านำไปใช้น้อยหรือไม่ได้ใช้เลย ทำให้เสียสมดุลการใช้ข้อสอบในคลังข้อสอบ เมื่อใช้การเลือกข้อสอบวิธีสารสนเทศสูงสุดของฟิชเชอร์ร่วมกับวิธีควบคุมการแสดงผลข้อสอบซิมสัน-เฮตเตอร์ (Simpson-Hetter Method) เพื่อแก้ไขปัญหาการแสดงผลข้อสอบมากเกินไป ผลพบว่าสามารถควบคุมการแสดงผลข้อสอบไม่ให้สูงเกินอัตราการแสดงผลข้อสอบกำหนด แต่ข้อสอบที่มีการใช้ต่ำไม่สามารถเพิ่มการแสดงผลข้อสอบได้ ทำให้เกิดการแจกแจงอัตราการแสดงผลข้อสอบโดยรวมยังเบี่ยงเหตุผลเนื่องจาก วิธีควบคุมการแสดงผลข้อสอบซิมสัน-เฮตเตอร์ควบคุมความถี่การแสดงผลข้อสอบหลังจากผ่านการเลือกข้อสอบ แต่ไม่ได้ควบคุมการเลือกข้อสอบโดยตรง ข้อสอบที่มีการใช้ต่ำมีโอกาสนในการเลือกใช้น้อย ดังนั้น การจัดสมดุลการใช้ข้อสอบในคลังข้อสอบจึงต้องได้รับการปรับปรุงแก้ไข นอกจากนี้ยังพบว่า ประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบด้วยวิธีสารสนเทศสูงสุดของฟิชเชอร์ร่วมกับวิธีควบคุมการแสดงผลข้อสอบซิมสัน-เฮตเตอร์จะน้อยกว่าการใช้วิธีสารสนเทศสูงสุดของฟิชเชอร์เพียงอย่างเดียวเมื่อความยาวแบบทดสอบค่อยๆ เพิ่มขึ้น (Leung; Chang; & Hau. 2002)

จากเหตุผลดังกล่าว ชางและยิง (Chang; & Ying. 1999) เสนอการเลือกข้อสอบวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับขั้นของค่าอำนาจจำแนก (a-Stratified Method : a-STR) ซึ่งประยุกต์มาจากกลยุทธ์รูปแบบปรับระดับขั้นของไวส์ (Weiss. 1973) จุดประสงค์เพื่อเพิ่ม

ประสิทธิภาพสมมูลการใช้ข้อสอบในคลังข้อสอบโดยยังรักษาประสิทธิภาพการประมาณค่าความสามารถของผู้สอบไว้ได้ นำมาใช้เป็นทางเลือกหนึ่งสำหรับวิธีการเลือกข้อสอบ ด้วยเหตุผล 2 ประการ คือ

1. ในระยะเริ่มต้นของการทดสอบ ค่าประมาณความสามารถยังห่างไกลจากค่าความสามารถจริงของผู้สอบ การใช้ข้อสอบอำนาจจำแนกสูงจะไม่ให้ประสิทธิภาพการประมาณค่าความสามารถของผู้สอบสูงสุดเท่าที่ควรจะเป็น การเลือกใช้ข้อสอบอำนาจจำแนกต่ำสามารถให้ประสิทธิภาพการประมาณค่าความสามารถของผู้สอบได้ดีในระดับหนึ่ง ดังนั้น ควรเลือกใช้ข้อสอบอำนาจจำแนกต่ำในระยะเริ่มต้นของการทดสอบ เมื่อค่าประมาณความสามารถใกล้เคียงค่าความสามารถจริงของผู้สอบในระยะท้ายๆ ของการทดสอบ จึงใช้ข้อสอบอำนาจจำแนกสูงเพื่อให้ได้ประสิทธิภาพการประมาณค่าความสามารถของผู้สอบสูงสุด (Chang; & Ying. 1996)

2. การเลือกข้อสอบจากสารสนเทศสูงสุดมักให้การแสดงข้อสอบมากเกินไป (Item Overexposure) ซึ่งพบได้เมื่อใช้ข้อสอบอำนาจจำแนกสูง การเลือกข้อสอบวิธีตามระดับขั้นของค่าอำนาจจำแนกสามารถลดการแสดงข้อสอบสำหรับข้อสอบอำนาจจำแนกสูงและเพิ่มการแสดงข้อสอบสำหรับข้อสอบอำนาจจำแนกต่ำ เป็นการกระจายการใช้ข้อสอบอำนาจจำแนกระดับต่างๆ สร้างสมมูลการใช้ข้อสอบในคลังข้อสอบ

วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับขั้นของค่าอำนาจจำแนก (a-Stratified Method : a-STR) ดำเนินการตามขั้นตอนดังนี้คือ

1. แบ่งคลังข้อสอบเป็นชั้นตามค่าอำนาจจำแนกข้อสอบ (พารามิเตอร์ a) ชั้นแรกบรรจุข้อสอบอำนาจจำแนกต่ำสุด ชั้น 2 บรรจุข้อสอบอำนาจจำแนกสูงกว่าชั้นแรก และเพิ่มขึ้นเรื่อยๆ ตามลำดับชั้น จนกระทั่งชั้นสุดท้ายบรรจุข้อสอบอำนาจจำแนกสูงสุด
2. แบ่งชั้นการทดสอบหรือความยาวแบบทดสอบตามการแบ่งชั้นของคลังข้อสอบ
3. ที่ตำแหน่งค่าความสามารถของผู้สอบบนจุดเริ่มต้นการทดสอบ เลือกข้อสอบความยากใกล้เคียงกับค่าความสามารถของผู้สอบมากที่สุดและจัดให้ผู้สอบ ความน่าจะเป็นในการตอบข้อสอบถูกต้องและผลการตอบข้อสอบนำไปคำนวณหาค่าประมาณความสามารถของผู้สอบก่อนนำไปใช้เลือกข้อสอบข้อต่อไป เมื่อทำการทดสอบจนครบตามความยาวแบบทดสอบย่อยในชั้นนั้น จึงเลื่อนการทดสอบขึ้นไปยังชั้นและชั้นคลังข้อสอบต่อไป
4. ทำการทดสอบจนครบทุกชั้นการทดสอบและทุกชั้นคลังข้อสอบ การทดสอบจึงยุติ

วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับขั้นของค่าอำนาจจำแนก มีแนวโน้มให้การใช้ข้อสอบอำนาจจำแนกระดับต่างๆ มีความทัดเทียมกันมากขึ้น เพิ่มความปลอดภัยข้อสอบและรักษาความมั่นคงของคลังข้อสอบไว้ด้วย การศึกษาในสถานการณ์จำลองของ ชางและยัง

(Chang; & Ying. 1999) แสดงผลให้เห็นว่า วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกเพิ่มการใช้ข้อสอบอำนาจจำแนกต่ำ ให้ประสิทธิภาพในการลดความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Square Error : MSE) และลดความลำเอียงเฉลี่ย (Average Bias) ของค่าประมาณความสามารถของผู้สอบ พร้อมกับให้สมดุลการใช้ข้อสอบในคลังข้อสอบ ฮัวและชาง (Hua; & Chang. 2001) แสดงผลการศึกษาว่า วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกลดอัตราการใช้ข้อสอบ เพิ่มการใช้ข้อสอบในคลังข้อสอบ ให้ประสิทธิภาพการประมาณค่าความสามารถของผู้สอบเปรียบเทียบกับวิธีสารสนเทศสูงสุดของฟิชเชอร์และให้สมดุลการใช้ข้อสอบ รวมทั้งโครงสร้างคลังข้อสอบมีความมั่นคงต่อการทดสอบด้วยการเติมข้อสอบอย่างต่อเนื่อง

สรุป วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกให้ผลดีเมื่อเปรียบเทียบกับวิธีสารสนเทศสูงสุดของฟิชเชอร์ (Hua; & Chang. 2001) อย่างน้อย 3 ประการ คือ

ประการที่ 1 ให้ประสิทธิภาพการประมาณค่าความสามารถของผู้สอบ เปรียบเทียบได้กับวิธีสารสนเทศสูงสุดของฟิชเชอร์

ประการที่ 2 วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกให้การควบคุมการแสดงผลข้อสอบโดยอัตโนมัติ วิธีสารสนเทศสูงสุดของฟิชเชอร์มักพบปัญหาด้านการแจกแจงการแสดงผลข้อสอบและความปลอดภัยของข้อสอบเมื่อเลือกใช้ข้อสอบอำนาจจำแนกสูงมากเกินไป แต่ข้อสอบอำนาจจำแนกปานกลางและต่ำมีการใช้ต่ำ เมื่อใช้วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก พบว่า การแสดงผลข้อสอบมีการแจกแจงมากขึ้น เนื่องจากข้อสอบอำนาจจำแนกระดับต่างๆ มีโอกาสได้รับการเลือกใช้ การออกแบบด้วยการแบ่งชั้นคลังข้อสอบตามค่าอำนาจจำแนก จึงเป็นเหมือนการควบคุมการแสดงผลข้อสอบโดยอัตโนมัติ

ประการที่ 3 วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกใช้ได้ง่ายเมื่อเปรียบเทียบกับวิธีสารสนเทศสูงสุดของฟิชเชอร์ร่วมกับวิธีควบคุมการแสดงผลข้อสอบซิมสัน-เฮดเทอร์ ซึ่งใช้เวลามากสำหรับการปฏิบัติและต้องใช้คอมพิวเตอร์ความไวสูง แตกต่างจากวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกซึ่งไม่ต้องคำนวณพารามิเตอร์ควบคุมการแสดงผลข้อสอบให้ยุ่งยาก

อย่างไรก็ตาม วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกอาจให้ข้อจำกัดบางประการ (Chang; & Ying. 1999) คือ 1. วิธีนี้ไม่รับรองการควบคุมการแสดงผลข้อสอบให้เป็นไปตามกำหนด และ 2. การออกแบบสำหรับวิธีควบคุมอื่นๆ ยังต้องพัฒนาต่อไป เช่น การควบคุมเนื้อหา หรือการควบคุมอื่นๆ ที่เกี่ยวข้อง

ลุง ชาง และฮัว (Leung; Chang; & Hua. 2002) ได้พิจารณาปัญหาวิธีการทดสอบแบบปรับ
 เหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกกรณีไม่รับรองการแสดงผลข้อสอบว่าจะเป็นไป
 ตามอัตราการแสดงผลข้อสอบกำหนด และเสนอแนวทางแก้ปัญหาโดยประยุกต์วิธีตามระดับชั้นของค่า
 อำนาจจำแนกร่วมกับวิธีควบคุมการแสดงผลข้อสอบซิมสัน-เฮดเทอร์ เรียกว่า วิธีขยายช่วงระดับชั้น
 (Enhanced Stratified Design) การรวม 2 วิธีเข้าด้วยกันจะทำให้การแจกแจงการแสดงผลข้อสอบมีความ
 สมดุลมากขึ้น และอยู่ในอัตราการแสดงผลข้อสอบกำหนด ขณะเดียวกัน ยังรักษาประสิทธิภาพการ
 ประเมินค่าความสามารถของผู้สอบไว้ได้

ขั้นตอนวิธีขยายช่วงระดับชั้น (Enhanced Stratified Design) มีดังนี้ คือ

1. กำหนดค่าอัตราการแสดงผลข้อสอบสูงสุดเป้าหมายเท่ากับ 0.2 เพื่อใช้เป็นจุดตัด
2. แบ่งขั้นการทดสอบเข้าในชั้นของคลังข้อสอบ ขั้นตอนการทดสอบทำตามอธิบายไว้ใน
 วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก

3. หาพารามิเตอร์การแสดงผลข้อสอบของข้อสอบแต่ละข้อจากสถานการณ์จำลอง เริ่มต้นจาก
 กำหนดให้พารามิเตอร์การแสดงผลของข้อสอบทุกข้อเท่ากับ 1 ดำเนินการทดสอบแบบปรับเหมาะจำลอง
 ตามขั้นตอนอธิบายไว้ในวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจ
 จำแนก นำไปทดสอบกับค่าความสามารถของกลุ่มตัวอย่างผู้สอบ อัตราการแสดงผลสำหรับข้อสอบแต่ละ
 ข้อนำไปเปรียบเทียบกับอัตราการแสดงผลข้อสอบสูงสุดเป้าหมาย ถ้าข้อสอบมีอัตราการแสดงผลข้อสอบ
 มากกว่าอัตราการแสดงผลข้อสอบสูงสุดเป้าหมาย พารามิเตอร์การแสดงผลข้อสอบต้องปรับให้ลดลงโดยคูณ
 ด้วยแฟกเตอร์น้อยกว่า 1 (เช่น 0.95) ถ้าข้อสอบมีอัตราการแสดงผลข้อสอบน้อยกว่าอัตราการแสดงผล
 ข้อสอบสูงสุดเป้าหมาย พารามิเตอร์การแสดงผลข้อสอบต้องปรับให้สูงขึ้นและค่าสูงสุดไม่เกิน 1 โดยคูณ
 ด้วยแฟกเตอร์มากกว่า 1 (เช่น 1.04) การปรับนี้จะดำเนินอย่างต่อเนื่องผ่านการทำซ้ำๆ จนกระทั่ง
 พารามิเตอร์การแสดงผลข้อสอบมีความคงที่

4. ที่แต่ละขั้นการทดสอบ ข้อสอบความยากใกล้เคียงค่าประมาณความสามารถของผู้สอบ
 ครั้งหลังสุดจะได้รับเลือก พารามิเตอร์ควบคุมการแสดงผลข้อสอบของข้อสอบข้อนั้นนำไปเปรียบเทียบกับ
 กับตัวเลขสุ่มที่สร้างจากการแจกแจงแบบเดียวกัน $U(0,1)$ ถ้าพารามิเตอร์ควบคุมการแสดงผลข้อสอบสูง
 กว่าตัวเลขสุ่ม จะจัดข้อสอบนั้นให้ แต่ถ้าพารามิเตอร์ควบคุมการแสดงผลข้อสอบต่ำกว่าตัวเลขสุ่ม จะไม่
 จัดข้อสอบนั้นให้ ทำการเลือกข้อสอบใหม่ที่มีค่าความยากใกล้เคียงค่าประมาณความสามารถของผู้สอบ
 อันตรรองลงมา และพิจารณาพารามิเตอร์ควบคุมการแสดงผลข้อสอบเปรียบเทียบกับตัวเลขสุ่มค่าใหม่
 ก่อนจัดข้อสอบให้ ขั้นตอนนี้ต่อเนื่องจนกระทั่งสามารถหาข้อสอบจัดให้แก่ผู้สอบ

5. ในกรณีที่ไม่สามารถหาข้อสอบจัดให้ได้และไม่มีข้อสอบอื่น ข้อที่ 4 จะทำซ้ำจนกระทั่งหา
 ข้อสอบข้อใดข้อหนึ่งจัดให้

6. ข้อที่ 4 ถึงข้อที่ 5 จะทำซ้ำๆ จนกระทั่งความยาวแบบทดสอบย่อยครบตามกำหนดที่ขึ้น การทดสอบนั้น การทดสอบจะต่อเนื่องไปยังขั้นต่อไป

7. ข้อที่ 4 ถึงข้อที่ 6 จะทำซ้ำๆ จนกระทั่งครบทุกขั้นการทดสอบและครบทุกชั้นคลัง ข้อสอบ การทดสอบจึงยุติ

การประเมินประสิทธิภาพการประมาณค่าความสามารถของผู้สอบและประสิทธิภาพสมมูล การใช้ข้อสอบในคลังข้อสอบของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับขั้นของค่าอำนาจจำแนกพบได้ในงานวิจัยต่างๆ [(Chang; & Ying. 1999) ; (Parshall; Harmes; & Kromrey. 2000) ; (Hua; & Chang. 2001) ; (Chang; Qian; & Ying. 2001) ; (Pastor; Dodd; & Chang. 2002) ; (Leung; Chang; & Hua. 2002) ; (Leung; Chang; & Hua. 2003) ; (Leung; Chang; & Hua. 2003 : Online) ; (Chang; & van der Linden. 2003) ; (Deng; & Ansley. 2003 : Online) ; (van der Linden; & Chang. 2003)] ดังนี้คือ

การประเมินประสิทธิภาพการประมาณค่าความสามารถของผู้สอบ

ประสิทธิภาพการประมาณค่าความสามารถของผู้สอบ หมายถึง ความสามารถของวิธีกำหนด ในการประมาณค่าความสามารถของผู้สอบให้ใกล้เคียงหรือเท่ากับค่าความสามารถจริงของผู้สอบมากที่สุด เมื่อเราสนใจศึกษาพารามิเตอร์ที่ไม่รู้ค่า (ค่าความสามารถจริง) สถิติที่นิยมใช้ทั่วไปเรียกว่า ตัวประมาณค่า (Estimator) การประมาณค่าความสามารถว่าใกล้เคียงหรือเท่ากับค่าความสามารถจริงของผู้สอบหรือไม่ พิจารณาจาก

1. ความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Square Error : MSE) หมายถึง ค่าบอกความถูกต้องของค่าประมาณความสามารถของผู้สอบ แสดงเป็นผลต่างกำลังสองเฉลี่ยระหว่างค่าประมาณความสามารถและค่าความสามารถจริงของผู้สอบ นอกจากนี้ MSE ยังมีค่าเท่ากับความแปรปรวนของค่าประมาณความสามารถรวมกับความลำเอียงกำลังสอง กล่าวได้ว่า MSE ให้ความหมายเหมือนคุณภาพของค่าประมาณความสามารถ MSE ขนาดเล็กบ่งชี้ถึงการประมาณค่าความสามารถมีความแตกต่างจากค่าความสามารถจริงของผู้สอบน้อยมาก สะท้อนให้เห็นถึงประสิทธิภาพของวิธีเลือกข้อสอบ

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

เมื่อ n แทน จำนวนผู้สอบทั้งหมด
 $\hat{\theta}_i$ แทน ค่าประมาณความสามารถของผู้สอบ
 θ_i แทน ค่าความสามารถจริงของผู้สอบ

2. ความลำเอียงเฉลี่ย (Average Bias) หมายถึง ค่าบอกความเที่ยงตรงของค่าประมาณความสามารถของผู้สอบ แสดงเป็นผลต่างเฉลี่ยระหว่างค่าประมาณความสามารถและค่าความสามารถจริงของผู้สอบ บอกทิศทางการประมาณค่าว่าให้ผลสูงหรือต่ำกว่าค่าความสามารถจริงของผู้สอบ ถ้าความลำเอียงกำลังสองเท่ากับ 0 ความคลาดเคลื่อนกำลังสองเฉลี่ยจะเท่ากับความแปรปรวนของค่าประมาณความสามารถของผู้สอบ ความลำเอียงเฉลี่ยนขนาดเล็กสะท้อนให้เห็นถึงประสิทธิภาพของวิธีเลือกข้อสอบ

$$Bias = \frac{1}{n} \sum_{i=1}^n (\hat{\theta}_i - \theta_i)^2$$

เมื่อ n แทน จำนวนผู้สอบทั้งหมด
 $\hat{\theta}_i$ แทน ค่าประมาณความสามารถของผู้สอบ
 θ_i แทน ค่าความสามารถจริงของผู้สอบ

การประเมินประสิทธิภาพสมมูลการใช้ข้อสอบในคลังข้อสอบ

ประสิทธิภาพสมมูลการใช้ข้อสอบในคลังข้อสอบ หมายถึง ความสามารถของวิธีกำหนดที่ทำให้ข้อสอบทุกข้อในคลังข้อสอบมีโอกาสนำไปใช้ในการทดสอบ พิจารณาจาก

1. ข้อสอบที่มีการแสดงมากเกินไป (Overexposure Item) หมายถึง ข้อสอบใดๆที่มีอัตราการแสดงสูงกว่า 0.2 เมื่ออัตราการแสดงข้อสอบ หมายถึง จำนวนครั้งของข้อสอบที่จัดให้แก่ผู้สอบต่อจำนวนผู้สอบทั้งหมดที่ทำการทดสอบ

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

ถ้าข้อสอบในคลังข้อสอบมีอัตราการแสดงสูงจำนวนมาก แสดงถึงแนวโน้มข้อสอบเหล่านั้นนำไปใช้ในการทดสอบมากเกินไป เป็นที่รู้จักแพร่หลายในกลุ่มผู้สอบ ส่งผลกระทบต่อความปลอดภัยของข้อสอบ ขาดสมมูลการใช้ข้อสอบในคลังข้อสอบ และสะท้อนถึงประสิทธิภาพวิธีเลือกข้อสอบไม่เหมาะสม

2. ข้อสอบที่มีการใช้ต่ำ (Underutilized Item) หมายถึง ข้อสอบใดๆ ที่มีอัตราการแสดงต่ำกว่า 0.2 ถ้าข้อสอบที่มีการใช้ต่ำมีจำนวนมากแสดงถึงข้อสอบในคลังข้อสอบมีโอกาสนำไปใช้น้อย ขาดสมดุลการใช้ข้อสอบในคลังข้อสอบ สะท้อนถึงประสิทธิภาพวิธีเลือกข้อสอบไม่เหมาะสม และขาดประสิทธิภาพในการบริหารต้นทุนในการพัฒนาข้อสอบ

3. อัตราการทับซ้อนข้อสอบ (Item Overlap Rate) หมายถึง สัดส่วนของข้อสอบที่ใช้ร่วมกันโดยคู่ของผู้สอบต่อข้อสอบที่ใช้ร่วมกันโดยคู่ของผู้สอบที่เป็นไปได้ทั้งหมด

เมื่อกำหนดให้ความยาวของแบบทดสอบคือ L ข้อ และจำนวนผู้สอบคือ P คน อัตราการทับซ้อนข้อสอบคำนวณโดย

1. นับจำนวนข้อสอบที่ใช้ร่วมกันแต่ละคู่ของผู้สอบ $P(P-1)/2$ คู่
2. บวกจำนวนข้อสอบที่ใช้ร่วมกัน $P(P-1)/2$ คู่
3. หารจำนวนทั้งหมดด้วย $L P(P-1)/2$

อัตราการทับซ้อนข้อสอบเป็นดัชนีสรุปสำหรับการควบคุมการแสดงผลข้อสอบ ใช้กำหนดขนาดและส่วนประกอบของคลังข้อสอบในการทดสอบแบบปรับเหมาะ ให้ภาพรวมว่าชุดข้อสอบถูกจัดให้บ่อยครั้งเพียงไรและคาดว่าจำนวนข้อสอบใช้ร่วมกันระหว่างคู่ของผู้สอบต้องมีจำนวนน้อยที่สุด ถ้าอัตราการทับซ้อนข้อสอบสูงจะคุกคามความปลอดภัยของข้อสอบ

4. การแจกแจงอัตราการแสดงผลข้อสอบ (Item Exposure Rate Distribution) หมายถึง ลักษณะการกระจายอัตราการแสดงผลของข้อสอบในคลังข้อสอบ บ่งบอกถึงประสิทธิภาพการใช้ข้อสอบในคลังข้อสอบโดยรวม พิจารณาจากสถิติไคสแควร์

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

er_j แทน อัตราการแสดงผลข้อสอบสำหรับข้อสอบที่ j

L แทน ความยาวแบบทดสอบ

N แทน ขนาดคลังข้อสอบ

L/N แทน อัตราการแสดงผลข้อสอบที่ต้องการ

er_j เป็นค่าสังเกต ขณะที่ L/N เป็นค่าคาดหวังว่าข้อสอบทุกข้อมีอัตราการแสดงผลข้อสอบเท่าเทียมกัน ดังนั้น สมการแสดงถึงความแตกต่างระหว่างค่าสังเกตและค่าคาดหวังของอัตราการแสดงผลข้อสอบ 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

$$F_{1.2} = \chi_1^2 / \chi_2^2$$

เมื่อ $F_{1.2}$ แทน อัตราส่วนการแจกแจงอัตราแสดงข้อสอบ 2 วิธี หรือ วิธีที่ 1 หารด้วยวิธีที่ 2
 χ_1^2 แทน การแจกแจงอัตราแสดงข้อสอบวิธีที่ 1
 χ_2^2 แทน การแจกแจงอัตราแสดงข้อสอบวิธีที่ 2
 ถ้า $F < 1$ แสดงว่า วิธีที่ 1 ให้สมดุลของอัตราแสดงข้อสอบโดยรวมดีกว่าวิธีที่ 2

วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก นับว่าเป็นวิธีการเลือกข้อสอบที่มีประสิทธิภาพในปัจจุบัน ให้ประสิทธิภาพการประมาณค่าความสามารถของผู้สอบเทียบได้กับวิธีสารสนเทศสูงสุดของฟิชเชอร์ มีคุณสมบัติควบคุมการแสดงข้อสอบโดยอัตโนมัติ สร้างสมดุลการใช้ข้อสอบอำนาจจำแนกระดับต่างๆ จากการศึกษาที่ผ่านมา วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกใช้รูปแบบการแบ่งชั้นคลังข้อสอบและแบ่งการทดสอบเข้าชั้นตามระดับค่าอำนาจจำแนกจากต่ำไปสูง โดยกำหนดจำนวนข้อสอบจำนวนเท่าๆ กันในแต่ละชั้นคลังข้อสอบ อย่างไรก็ตาม มีเหตุผลบางประการคาดว่า วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกสามารถให้ประสิทธิภาพการประมาณค่าความสามารถของผู้สอบและประสิทธิภาพสมดุลการใช้ข้อสอบในคลังข้อสอบเพิ่มขึ้น เมื่อใช้จำนวนข้อสอบไม่เท่ากันในแต่ละชั้นคลังข้อสอบ ด้วยเหตุผลสนับสนุนดังนี้

1. ระยะเริ่มต้นของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกสำหรับประชากรผู้สอบจำนวนมาก การแจกแจงค่าความสามารถของผู้สอบเหล่านี้มักจะกระจายอยู่รอบๆ ค่ากลางของโค้งการแจกแจงปกติ การเลือกข้อสอบในระยะเริ่มต้นจึงเลือกข้อสอบค่าความยากปานกลางจัดให้กับประชากรผู้สอบ ทำให้ข้อสอบความยากปานกลางอาจมีจำนวนไม่เพียงพอที่จะให้การประมาณค่าความสามารถของผู้สอบเป็นไปอย่างมีประสิทธิภาพ รวมทั้งการแสดงข้อสอบความยากรอบๆ ค่าความสามารถปานกลางมีความถี่สูง เสี่ยงต่อความปลอดภัยของข้อสอบ ดังนั้น การจัดข้อสอบให้มีจำนวนมากเพียงพอรอบๆ ระดับความสามารถปานกลางทั้งระยะเริ่มต้นการทดสอบและระยะท้ายการทดสอบ จะทำให้การเลือกข้อสอบมีตัวเลือกมากขึ้น และเติมเต็มข้อสอบความยากใกล้เคียงเมื่อข้อสอบเลือกใช้บ่อยมาถึงความถี่ตามเกณฑ์กำหนด ผลคาดว่า การประมาณค่าความสามารถของผู้สอบจะมีประสิทธิภาพมากขึ้นและการแสดงข้อสอบไม่สูงเกินอัตรากำหนด มีการกระจายการใช้ข้อสอบในคลังข้อสอบมากขึ้น ช่วยรักษาความปลอดภัยของข้อสอบ

2. ระยะเริ่มต้นของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก เมื่อค่าประมาณความสามารถยังห่างไกลจากค่าความสามารถจริงของผู้สอบ พบว่า ค่าความลำเอียงและค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสูง (Cheng; & Liou. 2000) และจากการศึกษาของ ลุง ชาง และฮัว (Leung; Chang; & Hau. 2001) พบว่า ถ้าเลือกข้อสอบความยากตรงกับความสามารถของผู้สอบ ถึงแม้ว่าเป็นข้อสอบอำนาจจำแนกต่ำก็สามารถให้สารสนเทศสูงได้ ในทางปฏิบัติข้อสอบอำนาจจำแนกสูงมีจำนวนไม่มากนัก การใช้ข้อสอบอำนาจจำแนกต่ำในระยะเริ่มต้นการทดสอบสามารถให้สารสนเทศได้ระดับหนึ่ง ช่วยลดความแตกต่างระหว่างค่าประมาณความสามารถและค่าความสามารถจริงของผู้สอบ ก่อนจะใช้ข้อสอบอำนาจจำแนกสูงในการทดสอบลำดับต่อไป ดังนั้น ถ้าจัดข้อสอบอำนาจจำแนกต่ำจำนวนมากในระยะเริ่มต้นการทดสอบ และรักษาข้อสอบอำนาจจำแนกสูงไว้ใช้ระยะท้ายการทดสอบ คาดว่าจะช่วยเพิ่มประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบ รวมทั้งยังกระจายการใช้ข้อสอบ โดยเฉพาะข้อสอบอำนาจจำแนกต่ำให้มีการใช้สูงขึ้น สร้างสมดุลการใช้ข้อสอบในคลังข้อสอบ

3. ระยะท้ายๆ ของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก เมื่อค่าประมาณความสามารถใกล้เคียงค่าความสามารถจริงของผู้สอบ การใช้ข้อสอบอำนาจจำแนกสูงจะให้ประสิทธิภาพสูงสุดในการทดสอบ (Chang; & Ying. 1996) ด้วยเหตุนี้ ข้อสอบอำนาจจำแนกสูงควรมีจำนวนมากเพียงพอระยะท้ายๆ ของการทดสอบ การศึกษาของ ชาง ควิน และยिंग (Chang; Qian; & Ying. 2001) และ ลุง ชาง และฮัว (Leung; Chang; & Hau. 2003) อ้างถึงการใช่วิธี a-STR ร่วมกับการจัดบล็อกค่าความยาก (b) (a-STR with b Blocking) เพื่อกระจายข้อสอบความยากระดับต่างๆ ตามระดับชั้นของค่าอำนาจจำแนก (a) สามารถลดความลำเอียงและความคลาดเคลื่อนกำลังสองเฉลี่ย การแสดงข้อสอบต่ำ ความเบ้ของการแจกแจงการแสดงข้อสอบลดลง ซึ่งได้ผลดีกว่าวิธี a-STR แบบเดิม แต่รูปแบบการจัดบล็อกค่าความยากเป็นรูปแบบในอุดมคติ โดยถือว่าการแจกแจงค่าความยากไม่มีผลต่อการแบ่งชั้น นั่นคือ ค่าอำนาจจำแนกและค่าความยากไม่มีความสัมพันธ์กัน แต่ในทางปฏิบัติข้อสรุปนี้เป็นไปได้ยาก ค่าอำนาจจำแนกและค่าความยากมักพบความสัมพันธ์ทางบวกบ่อยๆ กล่าวได้ว่า ข้อสอบอำนาจจำแนกสูงมีแนวโน้มที่จะมีค่าความยากสูงด้วย ชั้นค่าอำนาจจำแนกสูงของคลังข้อสอบจึงพบการขาดข้อสอบความยากต่ำ เป็นเหตุให้ข้อสอบความยากต่ำจำนวนน้อยเลือกใช้บ่อยเกินไปมีผลต่อความปลอดภัยของข้อสอบ สิ่งนี้ยืนยันจากการสังเกตของ ชาง ควิน และยिंग (Chang; Qian; & Ying. 2001) ว่า วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก ข้อสอบจำนวนมากมีอัตราการแสดงข้อสอบสูงโดยข้อสอบเหล่านี้มาจากชั้นที่ 3 และ 4 จากจำนวนทั้งหมด 4 ชั้น และถ้าค่าประมาณความสามารถของผู้สอบห่างไกลจากค่าความยากข้อสอบ ประสิทธิภาพการประมาณค่าความสามารถจะลดลงถึงแม้ว่าข้อสอบนั้นจะมีค่าอำนาจจำแนกสูงก็ตาม (Leung; Chang; & Hau. 2001) ดังนั้น วิธีที่คาดว่าจะให้ประสิทธิภาพได้ใกล้เคียงวิธี a-STR

ร่วมกับการจัดล่อค่าความยากที่เป็นไปได้มากในทางปฏิบัติคือ การเพิ่มจำนวนข้อสอบโดยเฉพาะข้อสอบอำนาจจำแนกสูงระยะท้ายๆ ของการทดสอบ เป็นการเติมเต็มข้อสอบเมื่อข้อสอบเลือกใช้บ่อยมาถึงอัตราการแสดงตามเกณฑ์กำหนด ก็จะมีข้อสอบข้อใหม่เพียงพอพร้อมได้รับเลือกต่อไป ผลคาดว่าจะได้รับคือ การเพิ่มจำนวนข้อสอบอำนาจจำแนกสูงระยะท้ายการทดสอบ จะช่วยเพิ่มประสิทธิภาพในการประมาณค่าความสามารถใกล้เคียงค่าความสามารถจริงของผู้สอบมากที่สุด พร้อมทั้งกระจายการใช้ข้อสอบไปยังข้อสอบอำนาจจำแนกสูงข้ออื่นๆ ระยะท้ายการทดสอบไม่ให้มีอัตราการแสดงสูงกว่าเกณฑ์กำหนด

จากเหตุผลดังกล่าว นำไปสู่ประเด็นสำคัญสนับสนุนการศึกษาประสิทธิภาพการประมาณค่าความสามารถของผู้สอบ และสมมูลการใช้ข้อสอบในคลังข้อสอบของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับขั้นของค่าอำนาจจำแนกโดยใช้ข้อสอบจำนวนไม่เท่ากันในแต่ละชั้น ผลคาดหวังว่า วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับขั้นของค่าอำนาจจำแนกโดยใช้ข้อสอบจำนวนไม่เท่ากันในแต่ละชั้นคลังข้อสอบ สามารถเพิ่มประสิทธิภาพการประมาณค่าความสามารถของผู้สอบ และประสิทธิภาพสมมูลการใช้ข้อสอบในคลังข้อสอบได้มากกว่าวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับขั้นของค่าอำนาจจำแนกที่ใช้ข้อสอบจำนวนเท่ากันในแต่ละชั้นคลังข้อสอบ

เนื่องจากการศึกษาประสิทธิภาพของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับขั้นของค่าอำนาจจำแนกมีความเกี่ยวข้องกับการวางเงื่อนไขต่างๆ ในการทดสอบ การวิจัยครั้งนี้สนใจทำการศึกษาเงื่อนไขสำคัญ คือ การควบคุมการแสดงผลข้อสอบ ขนาดคลังข้อสอบ และความยาวแบบทดสอบ ด้วยเหตุผลดังนี้คือ

การควบคุมการแสดงผลข้อสอบอาจลดประสิทธิภาพการประมาณค่าความสามารถของผู้สอบสังเกตจากผลการศึกษาว่า ประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบสำหรับการเลือกข้อสอบวิธีสารสนเทศสูงสุดร่วมกับวิธีควบคุมการแสดงผลข้อสอบชิมสัน-เฮดเทอร์ ลดน้อยกว่าวิธีสารสนเทศสูงสุดเพียงอย่างเดียวเมื่อความยาวแบบทดสอบเพิ่มขึ้น (Leung; Chang; & Hau. 2002) ดังนั้น มีความเป็นไปได้เมื่อข้อสอบความยากใกล้เคียงค่าประมาณความสามารถของผู้สอบมากที่สุด ไม่ได้จัดให้ เนื่องจากข้อสอบข้อนั้นมีการแสดงผลข้อสอบสูงเกินอัตราการแสดงผลข้อสอบกำหนด ทำให้แบบทดสอบต้องเลือกข้อสอบความยากอันดับรองลงมาจัดให้แทน ผลคือลดความคลาดเคลื่อนได้น้อยและช้ากว่า การลดประสิทธิภาพดังกล่าวอาจไม่เกิดขึ้นถ้าคลังข้อสอบมีขนาดใหญ่เพียงพอและความยาวแบบทดสอบสั้นลง

คลังข้อสอบขนาดใหญ่ช่วยให้แบบทดสอบมีโอกาสเลือกข้อสอบที่ดีที่สุดได้มากกว่าคลังข้อสอบขนาดเล็ก ให้สารสนเทศแบบทดสอบเพิ่มขึ้น เพิ่มประสิทธิภาพในการประมาณค่า

ความสามารถของผู้สอบ และกระจายการแสดงข้อสอบไปยังข้อสอบอื่นๆ ได้ดี แต่การสร้างคลังข้อสอบขนาดใหญ่ต้องเสียเวลาและค่าใช้จ่ายสูง รวมทั้งการสร้างข้อสอบคุณภาพดีทำได้ยากโดยเฉพาะข้อสอบค่าอำนาจจำแนกสูง

เมื่อความยาวแบบทดสอบค่อยๆ เพิ่มขึ้น ความคลาดเคลื่อนของการประมาณค่าความสามารถของผู้สอบมีแนวโน้มลดลง จนค่าประมาณความสามารถเข้าใกล้ค่าความสามารถจริงของผู้สอบ ให้ประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบ แต่ถ้าความยาวแบบทดสอบมากเกินไปพบว่า การแสดงข้อสอบจะสูงเกินอัตราการแสดงข้อสอบกำหนด (Leung; Chang; & Hau. 2003 : Online) และอาจให้สารสนเทศแบบทดสอบมีลักษณะเป็นโค้งแบน (Thissen. 1990 : 169) ปัญหานี้สามารถลดลงได้ถ้าคลังข้อสอบมีขนาดใหญ่เพียงพอ

จะเห็นได้ว่า เงื่อนไขทั้ง การควบคุมการแสดงข้อสอบ ขนาดคลังข้อสอบ และความยาวแบบทดสอบมีความสัมพันธ์ซึ่งกันและกัน การจัดเงื่อนไขให้มีความเหมาะสมกับลักษณะวิธีกำหนดในการทดสอบ จะรักษาประสิทธิภาพการประมาณค่าความสามารถของผู้สอบและสมดุลการใช้ข้อสอบในคลังข้อสอบไว้ได้ ดังนั้น จึงได้ทำการศึกษาเงื่อนไข การควบคุมการแสดงข้อสอบ ขนาดคลังข้อสอบ และความยาวแบบทดสอบ พร้อมกันกับการเลือกข้อสอบวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก เมื่อใช้จำนวนข้อสอบไม่เท่ากันในแต่ละชั้นคลังข้อสอบ ในการวิจัยครั้งนี้ด้วย

6. งานวิจัยที่เกี่ยวข้อง

ซางและยิง (Chang; & Ying. 1996) ตั้งข้อสังเกตว่าการเลือกใช้ข้อสอบที่ดีที่สุดในระยะเริ่มต้นของการทดสอบอาจไม่ให้ประสิทธิภาพเมื่อค่าประมาณความสามารถยังไม่เข้าใกล้ค่าความสามารถจริง ดังนั้นจึงได้ศึกษาประสิทธิภาพวิธีการเลือกข้อสอบ จากวิธีนิยามกันมากคือ การเลือกข้อสอบวิธีสารสนเทศสูงสุดของฟิชเชอร์ (Fisher Item Information) เปรียบเทียบกับวิธีสารสนเทศโดยรวมเฉลี่ย (Global Information) (หรือ Kullback-Leibler Information) ศึกษาในสถานการณ์จำลอง 2 สถานการณ์ คือ

1. เปรียบเทียบวิธีสารสนเทศโดยรวมเฉลี่ยและวิธีสารสนเทศสูงสุดของฟิชเชอร์ เมื่อใช้ข้อสอบจำนวนน้อยในระยะเริ่มต้นของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์

กำหนดขนาดคลังข้อสอบ 800 ข้อ ช่วงค่าพารามิเตอร์ข้อสอบต่างๆ คือ ค่าอำนาจจำแนก (.5 ถึง 2.5) ค่าความยาก (-3.6 ถึง 3.6) ค่าการเดา (0.0 ถึง .25) กำหนดช่วงความสามารถต่อเนื่อง 8 ค่า (-3.0 -2.0 -1.5 -1.0 0.0 1.0 2.0 3.0) ความยาวข้อสอบ 14 ข้อ เริ่มทำการทดสอบแบบปรับเหมาะโดยใช้การเลือกข้อสอบทั้งสองวิธี ทำซ้ำ 1000 ครั้ง ผลการศึกษาพบว่า การเลือกข้อสอบโดยวิธีสารสนเทศโดยรวมเฉลี่ยจะให้ความลำเอียง และความคลาดเคลื่อนกำลังสองเฉลี่ย น้อยกว่าวิธีสารสนเทศสูงสุดของฟิชเชอร์ในระยะเริ่มต้นเมื่อใช้ข้อสอบจำนวนน้อย แต่เมื่อเพิ่มจำนวนข้อสอบขึ้น ความแตกต่างจะค่อยๆ ลดลง

2. เปรียบเทียบวิธีสารสนเทศโดยรวมเฉลี่ยและวิธีสารสนเทศสูงสุดของฟิชเชอร์ เมื่อใช้ค่าการคำนวณมาจากข้อสอบมาตรฐาน 254 ข้อ

วิธีเหมือนการศึกษาที่ 1 แตกต่างกันที่กำหนดช่วงความสามารถต่อเนื่องเท่ากับ -2.0, -1.0, 1.0 และ 2.0 ค่าความยากเริ่มต้นที่ 0 ความยาวแบบทดสอบ 40 ข้อ ข้อสอบ 254 ข้อ แบ่งเป็น รูปแบบการตอบข้อสอบ 2 พารามิเตอร์จำนวน 122 ข้อ และรูปแบบการประมาณค่า 3 พารามิเตอร์จำนวน 132 ข้อ โดยพารามิเตอร์ข้อสอบเหล่านี้มีการแจกแจงแตกต่างกัน ผลการศึกษาพบว่า วิธีสารสนเทศโดยรวมเฉลี่ยลดความลำเอียงได้ดีกว่าวิธีสารสนเทศสูงสุดของฟิชเชอร์เมื่อช่วงความสามารถต่อเนื่องเท่ากับ -2.0 และ -1.0 แต่ไม่พบความแตกต่างเมื่อช่วงความสามารถต่อเนื่องเท่ากับ 1.0 และ 2.0 วิธีสารสนเทศโดยรวมเฉลี่ยยังให้ความคลาดเคลื่อนกำลังสองเฉลี่ยลดลงอย่างมีนัยสำคัญในช่วงความสามารถต่อเนื่องเท่ากับ -2.0 ส่วนช่วงความสามารถต่อเนื่องอื่นๆ ค่าความ คลาดเคลื่อนกำลังสองเฉลี่ยลดลงเมื่อข้อสอบมีจำนวนน้อย และไม่พบความแตกต่างทั้งสองวิธีเมื่อจำนวนข้อสอบเพิ่มขึ้นมากกว่า 30 ข้อ

ผลการศึกษาสรุปว่า วิธีสารสนเทศโดยรวมเฉลี่ยสามารถลดความลำเอียงและให้ความคลาดเคลื่อนกำลังสองเฉลี่ยขนาดเล็กกว่าในระยะเริ่มต้นของการทดสอบ แนวคิดของวิธีสารสนเทศ

โดยรวมเฉลี่ยนำไปประยุกต์ใช้ข้อสอบอำนาจจำแนกต่ำแทนการใช้ข้อสอบอำนาจจำแนกสูงในระยะเริ่มต้นของการทดสอบแบบปรับเหมาะ เมื่อยังไม่มีรู้ตำแหน่งค่าความสามารถจริงที่แน่นอน

ซางและยিং (Chang; & Ying. 1999) เสนอการเลือกข้อสอบวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับขั้นของค่าอำนาจจำแนก (a-Stratified Computerized Adaptive Testing : a-STR) ทำการศึกษาในสถานการณ์จำลองเพื่อเปรียบเทียบวิธี a-STR กับวิธีการสารสนเทศสูงสุดของฟิชเชอร์ร่วมกับการควบคุมการแสดงผลข้อสอบซิมสัน-เฮตเตอร์ (Sympson- Hetter Method) (FSH) และวิธีการเลือกข้อสอบแบบเขียน (EAP) ร่วมกับการควบคุมการแสดงผลข้อสอบซิมสัน-เฮตเตอร์ (BSH)

การศึกษาที่ 1

ใช้ขนาดคลังข้อสอบ 400 ข้อ แบ่ง 4 ระดับๆ ละ 100 ข้อ กำหนดข้อสอบอำนาจจำแนก (a) เท่ากัน สำหรับ 100 ข้อระดับแรกเท่ากับ .5 และเพิ่มขึ้นระดับละ .5 ค่าความยาก (b) สร้างโดยสุ่มจากการแจกแจงโค้งปกติมาตรฐาน $N(0,1)$ ค่าการเดา (c) เท่ากับ 0 ความยาวแบบทดสอบ 40 และ 60 ข้อ ทั้ง 2 ชุดจัดทดสอบด้วยค่าความสามารถ 3,000 ค่าสร้างจากการแจกแจงโค้งปกติมาตรฐาน $N(0,1)$ การเลือกข้อสอบ 3 วิธีคือ a-STR FSH และ BSH ทั้ง 3 วิธีใช้ข้อสอบเทียม 3 ข้อประมาณค่าความสามารถโดยวิธีการประมาณค่าความเป็นไปได้สูงสุด (MLE) เพื่อหาสารสนเทศค่าความสามารถเบื้องต้นเป็นจุดเริ่มต้นการทดสอบ ระหว่างการทดสอบใช้ MLE ในการประมาณค่าความสามารถ สำหรับวิธี a-STR และ FSH วิธี BSH ใช้การประมาณค่าความสามารถตามวิธี EAP ผลการศึกษาพิจารณา ความลำเอียงเฉลี่ย ความคลาดเคลื่อนกำลังสองเฉลี่ย และการแจกแจงอัตราการแสดงผลข้อสอบด้วยสถิติไคสแควร์

การศึกษาที่ 2

การศึกษาที่ 2 คล้ายการศึกษาที่ 1 แต่ใช้พารามิเตอร์ของข้อสอบปฏิบัติจริง (ได้แก่ The Reading Assessment 1992 National Assessment of Educational Progress) ข้อสอบทั้งหมด 254 ข้อ ความยาวแบบทดสอบ 20 ข้อ รูปแบบการตอบข้อสอบ 3 พารามิเตอร์ ข้อสอบแบ่งเป็นกลุ่มย่อย 4 กลุ่มสอดคล้องกับค่าอำนาจจำแนก

ผลการศึกษา

ผลการศึกษาที่ 1 สำหรับความยาวแบบทดสอบ 40 ข้อพบว่า วิธีการเลือกข้อสอบทั้ง 3 วิธีให้ความลำเอียงเฉลี่ยและความคลาดเคลื่อนกำลังสองเฉลี่ยใกล้เคียงกัน อัตราการทับซ้อนข้อสอบของวิธี a-STR น้อยกว่าประมาณครึ่งหนึ่งของวิธี FSH และวิธี BSH วิธี a-STR ยังเพิ่มการใช้ข้อสอบที่ใช้น้อยในวิธี FSH และวิธี BSH และให้การแสดงผลข้อสอบสมคูลมากกว่า เมื่อพิจารณาการแจกแจงอัตราการแสดงผลข้อสอบด้วยสถิติไคสแควร์พบว่าวิธี a-STR สัมพันธ์กับวิธี FSH และ BSH สำหรับความยาวแบบทดสอบ 60 ข้อ ทั้ง 3 วิธีให้ความลำเอียงเฉลี่ยต่ำกว่าเมื่อเทียบกับความยาวแบบทดสอบ 40 ข้อ

วิธี a-STR ให้อัตราการทับซ้อนข้อสอบน้อยกว่าวิธี FSH และวิธี BSH วิธี a-STR ให้ความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุดขณะที่วิธี BSH ให้มากที่สุด วิธี a-STR ลดความเบ้ได้มากกว่า 50 เปอร์เซนต์เมื่อเทียบกับวิธี FSH และวิธี BSH นอกจากนี้วิธี a-STR ให้สมคุณการแจกแจงอัตราการแสดงข้อสอบมากกว่าและยังเพิ่มการแสดงข้อสอบอำนาจจำแนกต่ำ

ผลการศึกษาที่ 2 แสดงถึงความคลาดเคลื่อนกำลังสองเฉลี่ยและความลำเอียงเฉลี่ยของทั้ง 3 วิธีไม่แตกต่างกัน แต่วิธี a-STR มีแนวโน้มให้ความคลาดเคลื่อนกำลังสองเฉลี่ยค่อยๆ เพิ่มขึ้น ลดอัตราการทับซ้อนข้อสอบ ลดจำนวนข้อสอบที่มีการแสดงข้อสอบน้อย และลดความเบ้ของการแจกแจงอัตราการแสดงข้อสอบเมื่อเปรียบเทียบกับวิธี FSH และวิธี BSH

ลุง ชาง และฮัว (Leung; Chang; & Hau. 2002) ศึกษาเปรียบเทียบวิธีการเลือกข้อสอบ 3 วิธี คือ วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก (a-Stratified Method : a-STR) ร่วมกับการควบคุมการแสดงผลข้อสอบวิธีซิมสัน-เฮตเตอร์ (a-Stratified Method with Simpson-Hetter Procedure : a-STR-SH) และวิธีสารสนเทศสูงสุดร่วมกับการควบคุมการแสดงผลข้อสอบวิธีซิมสัน-เฮตเตอร์ (Maximum Information with Simpson-Hetter procedure : Max-I-SH) ประเมินผลโดย ค่าความเชื่อถือ (Fedelity) ความลำเอียงเฉลี่ย ความคลาดเคลื่อนกำลังสองเฉลี่ย จำนวนข้อสอบที่มีการแสดงมากเกินไป จำนวนข้อสอบที่มีการใช้ต่ำ การแจกแจงอัตราการแสดงข้อสอบด้วยสถิติไคสแควร์ อัตราการทับซ้อนข้อสอบ เวลาในการคำนวณ (Computing Time) การศึกษาทำในสถานการณ์จำลอง 4 สถานการณ์ 2 สถานการณ์แรก การทดสอบจำลองในกลุ่มตัวอย่างค่าความสามารถ 3,000 ค่าซึ่งสุ่มจากโค้งการแจกแจงมาตรฐาน $N(0,1)$ 2 สถานการณ์หลังใช้ผู้สอบค่าความสามารถ 10,000 ค่า (จำแนกระดับความสามารถ -3.0, ..., 3.0)

การศึกษาที่ 1 ใช้ข้อสอบจำลองและความสามารถสุ่ม

เปรียบเทียบวิธี a-STR และวิธี a-STR-SH กำหนดพารามิเตอร์ข้อสอบ คือ ค่าการเดา (c) คงที่เท่ากับ 0 ค่าอำนาจจำแนก (a) กำหนดไว้ที่ 0.5, 1.0, 1.5 และ 2.0 ค่าความยาก (b) สร้างโดยสุ่มจากโค้งปกติมาตรฐาน $N(0,1)$ ความยาวแบบทดสอบ 40 และ 60 ข้อ ขนาดคลังข้อสอบ 400 ข้อ แบ่งเป็น 4 ชั้นๆ ละ 100 ข้อ ผู้สอบแต่ละคนได้รับแบบทดสอบ 6 กรณี คือ ความยาวแบบทดสอบ 40 และ 60 ข้อ ใช้วิธีเลือกข้อสอบ 3 วิธี ในแต่ละกรณี ทดสอบด้วยข้อสอบเทียม 3 ข้อเพื่อหาจุดเริ่มต้นการทดสอบ การศึกษาที่ 2 ใช้ข้อสอบปฏิบัติจริงและความสามารถสุ่ม

ใช้พารามิเตอร์จากข้อสอบปฏิบัติจริง 254 ข้อ แบ่งคลังข้อสอบเป็น 4 ชั้น ความยาวแบบทดสอบชั้นละ 6 ข้อ รวม 24 ข้อ วิธีทดสอบเหมือนการศึกษาที่ 1

การศึกษาที่ 3 ใช้ข้อสอบจำลองและความสามารถคงที่

ใช้คลังข้อสอบและความยาวแบบทดสอบเหมือนการศึกษาที่ 1 ทดสอบด้วยค่าความสามารถผู้สอบ 10,000 ค่า มาจากระดับความสามารถที่ -3.0 -2.6... 3.0 ตามลำดับ ศึกษาเปรียบเทียบความลำเอียง และ ความคลาดเคลื่อนกำลังสองเฉลี่ยทั้ง 3 วิธี

การศึกษาที่ 4 ใช้ข้อสอบปฏิบัติจริงและความสามารถคงที่

ใช้ข้อสอบปฏิบัติจริง ความยาวข้อสอบ 24 ข้อจากการศึกษาที่ 2 และความสามารถคงที่จากการศึกษาที่ 3

ผลการศึกษาที่ 1 พบว่า ทั้งความยาวแบบทดสอบ 40 และ 60 ข้อ สัมประสิทธิ์ความเชื่อถือ (Fidelity Coefficients) ทั้ง 3 วิธีมีความสัมพันธ์ของการประมาณค่าความสามารถสูง ความลำเอียงเฉลี่ยเข้าใกล้ 0 ทั้ง 3 วิธี ความคลาดเคลื่อนกำลังสองเฉลี่ยสำหรับความยาวแบบทดสอบ 40 ข้อของวิธี Max-I-SH น้อยกว่าวิธี a-STR และวิธี a-STR-SH แต่เมื่อความยาวแบบทดสอบ 60 ข้อ พบว่า วิธี a-STR ให้ผลน้อยกว่าวิธี a-STR-SH และวิธี Max-I-SH ข้อสอบที่มีการแสดงข้อสอบมากเกินไปของวิธี a-STR-SH มีจำนวนน้อยที่สุดเมื่อเทียบกับอีก 2 วิธี การทับซ้อนข้อสอบของวิธี Max-I-SH สูงกว่าอีก 2 วิธีเมื่อใช้ความยาวแบบทดสอบ 40 ข้อ แต่เมื่อจำนวนข้อสอบเพิ่มขึ้นความแตกต่างจะลดลง วิธี a-STR-SH จะให้ค่าการทับซ้อนข้อสอบเล็กที่สุด การสร้างพารามิเตอร์ควบคุมการแสดงผลข้อสอบทั้งความยาวแบบทดสอบ 40 และ 60 ข้อ พบว่า วิธี a-STR-SH ใช้เวลาน้อยกว่าวิธี Max-I-SH

ผลการศึกษาที่ 2 พบว่า สัมประสิทธิ์ความเชื่อถือทั้ง 3 วิธีค่อยๆ ลดลงเมื่อความยาวแบบทดสอบสั้นลงและขนาดคลังข้อสอบเล็กลงแต่ยังเปรียบเทียบกันได้ ความลำเอียงเฉลี่ยทั้ง 3 วิธีมีขนาดเล็กและเปรียบเทียบได้ ความคลาดเคลื่อนกำลังสองเฉลี่ยพบว่าวิธี Max-I-SH มีประสิทธิภาพมากขึ้น สมดุลการใช้ข้อสอบในคลังข้อสอบได้ผลดีในวิธี a-STR และวิธี a-STR-SH อัตราการทับซ้อนข้อสอบของวิธี Max-I-SH มีมากกว่าวิธี a-STR และวิธี a-STR-SH วิธี a-STR-SH ให้การควบคุมการแสดงผลข้อสอบดีที่สุด อัตราการทับซ้อนข้อสอบน้อยที่สุด ให้การแจกแจงอัตราการแสดงผลข้อสอบน้อยที่สุด และใช้เวลาในการคำนวณน้อยที่สุด

ผลการศึกษาที่ 3 และ 4 พบว่า ทั้ง 3 วิธีให้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยใกล้เคียงกัน วิธี a-STR และวิธี a-STR-SH ให้ผลบนมาตรฐานวัดความสามารถคล้ายอีก 2 การศึกษา วิธี Max-I-SH จะให้ประสิทธิภาพเมื่อความยาวแบบทดสอบสั้น ค่าความลำเอียงทั้ง 3 วิธีจะใกล้เคียงกัน

ซาง ควิน และยิง (Chang; Qian; & Ying. 2001) ศึกษาประสิทธิภาพการเลือกข้อสอบวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก (a-Stratified Method : a-STR) เปรียบเทียบกับวิธีตามระดับชั้นของค่าอำนาจจำแนกและค่าความยาก (a-Stratified Method with b Blocking : ab-STR) ในด้านประสิทธิภาพการประมาณค่าความสามารถของผู้สอบ การใช้ข้อสอบในคลังข้อสอบ และความปลอดภัยข้อสอบ

การศึกษาทำในสถานการณ์จำลองโดยใช้ข้อสอบ 360 ข้อ ความยาวแบบทดสอบ 40 ข้อ ค่าความสามารถ 3000 ค่าสร้างจากการแจกแจงโค้งปกติมาตรฐาน $N(0,1)$ ใช้รูปแบบการตอบข้อสอบ 3 พารามิเตอร์ เตรียมคลังข้อสอบโดยวิธี a-STR คลังข้อสอบถูกแบ่ง 4 ชั้นตามค่าอำนาจจำแนก (a) จากข้างล่างขึ้นข้างบน ชั้นแรกบรรจุข้อสอบค่า a ต่ำที่สุดและชั้นที่ 4 บรรจุข้อสอบค่า a สูงสุด สำหรับวิธี ab-STR คลังข้อสอบจะถูกแบ่ง 4 ระดับเหมือนกัน แต่จัดบล็อก 90 บล็อกด้วยกลุ่มค่าความยาก (b) เหมือนกัน แต่ละกลุ่มประกอบด้วยข้อสอบ 4 ข้อที่มีค่า b เหมือนกันมากที่สุด ดังนั้น กลุ่มที่ 1 ประกอบด้วยข้อสอบ 4 ข้อที่มีค่า b ต่ำที่สุดและกลุ่มที่ 90 บรรจุข้อสอบ 4 ข้อที่มีค่า b สูงสุด ข้อสอบค่า a ต่ำสุดมีอยู่ในทุกๆ กลุ่มระดับแรก ข้อสอบค่า a ต่ำเป็นอันดับ 2 มีอยู่ในทุกๆ กลุ่มระดับที่ 2 ค่า a อื่นๆ ถูกสร้างในทำนองเดียวกัน ใช้วิธีการประมาณค่าความเป็นไปได้สูงสุด (Maximum Likelihood Estimation : MLE) ประมาณค่าความสามารถทั้งสองวิธีนี้เริ่มทดสอบด้วยข้อสอบเทียม 3 ข้อแรกตามวิธีของซางและยिंग (Chang; & Ying. 1999)

ผลการศึกษาพบว่า วิธี ab-STR ให้ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างความสามารถจริงและค่าประมาณความสามารถสูงเมื่อเทียบกับวิธี a-STR วิธี ab-STR ดีกว่าวิธี a-STR ในด้านการลดความลำเอียง ความคลาดเคลื่อนกำลังสองเฉลี่ย ลดการทับซ้อนข้อสอบ และให้ประสิทธิภาพการใช้ข้อสอบในคลังข้อสอบ การแจกแจงอัตราการแสดงข้อสอบสำหรับวิธี ab-STR ประมาณ 1 ใน 5 ของวิธี a-STR ประมาณ 74 % ของความเป็นวิธี ab-STR ถูกทำให้ลดลงเมื่อเทียบกับวิธี a-STR

ลุง ซาง และฮัว (Leung; Chang; & Hua. 2003 : Online) เปรียบเทียบวิธีการเลือกข้อสอบ 3 วิธี คือ 1. วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก (a-Stratified Method : a-STR) 2. วิธีตามระดับชั้นของค่าอำนาจจำแนกและค่าความยาก (a-Stratified Method with b Blocking : ab-STR) และ 3. วิธีตามระดับชั้นของค่าอำนาจจำแนก ค่าความยาก และควบคุมเนื้อหา (a-Stratified Method with b Blocking and Content : abC-STR) โดยใช้การจัดสมดุลเนื้อหา 3 วิธี คือ 1. การทดสอบแบบปรับเหมาะแบบควบคุมสมดุลเนื้อหา (Constrained Computerized Adaptive Testing : CCAT) 2. รูปแบบมัลติโนเมียลประยุกต์ (Modified Multinomial Model : MMM) และ 3. การทดสอบแบบปรับเหมาะแบบควบคุมสมดุลเนื้อหาประยุกต์ (Modified Constrained Computerized Adaptive Testing : MCCAT) จำนวนเนื้อหา 4 เนื้อหา ขนาดคลังข้อสอบ 700 ข้อ ความยาวแบบทดสอบ 35 ข้อ ค่าความสามารถ 5,000 ค่า ประเมินผลพิจารณาความสัมพันธ์ของค่าประมาณความสามารถและค่าความสามารถจริง ความคลาดเคลื่อนกำลังสองเฉลี่ย การแจกแจงอัตราการแสดงข้อสอบด้วยสถิติไคสแควร์ จำนวนข้อสอบที่มีการใช้ต่ำ จำนวนข้อสอบที่มีการแสดงข้อสอบมากเกินไป และอัตราการทับซ้อนข้อสอบ ผลการศึกษาแสดงวิธีการเลือกข้อสอบทั้ง 3 วิธีให้ความสัมพันธ์ของค่าประมาณความสามารถและค่าความสามารถจริงเหมือนกัน ความถูกต้องและ

ประสิทธิภาพของการทดสอบสามารถเปรียบเทียบกันได้ โดยพิจารณาจากค่าความคลาดเคลื่อนกำลังสองเฉลี่ย แต่วิธี abC-STR ให้การใช้ข้อสอบในคลังข้อสอบดีที่สุด อาจเกิดจากการจัดข้อสอบให้เพียงพอสำหรับการจัดเนื้อหา วิธี CCAT ให้อัตราการทับซ้อนข้อสอบสูง เมื่อพิจารณาการใช้ข้อสอบในคลังข้อสอบต่อค่าไคสแควร์ วิธี MMM ให้สมดุลการใช้ข้อสอบได้ดีกับข้อสอบทั้งหมดขณะที่ CCAT ให้ผลน้อย ดังนั้นวิธี abC-STR ที่ใช้ MMM อาจเป็นวิธีการเลือกข้อสอบที่เหมาะสมในการควบคุมการทับซ้อนข้อสอบ การใช้ข้อสอบในคลังข้อสอบ และให้สมดุลเนื้อหา

ลุง ชาง และฮัว (Leung; Chang; & Hua. 2003) ศึกษาเปรียบเทียบการจัดสมดุลเนื้อหาด้วยวิธี 1. การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ร่วมกับการควบคุมสมดุลเนื้อหา (Constrained Computerized Adaptive Testing : CCAT) 2. รูปแบบมัลติโนเมียลประยุกต์ (Modified Multinomial Model : MMM) และ 3. การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ร่วมกับการควบคุมสมดุลเนื้อหาประยุกต์ (Modified Constrained Computerized Adaptive Testing : MCCAT) โดยใช้วิธีการเลือกข้อสอบจากสารสนเทศสูงสุดเมื่อความยาวแบบทดสอบต่างกัน และวิธีการควบคุมอัตราการแสดงข้อสอบซิมสัน-เฮตเทอร์ กำหนดอัตราการแสดงข้อสอบที่ระดับ .1 และ .2 ประเมินผลจากความสัมพันธ์ของค่าประมาณความสามารถและค่าความสามารถจริง ความลำเอียงเฉลี่ย ความคลาดเคลื่อนกำลังสองเฉลี่ย การแจกแจงอัตราการแสดงข้อสอบด้วยสถิติไคสแควร์ จำนวนข้อสอบที่มีการใช้ซ้ำ จำนวนข้อสอบที่มีการแสดงข้อสอบมากเกินไป ผลการศึกษาพบว่า การจัดสมดุลเนื้อหาทั้ง 3 วิธีให้ความถูกต้องในการประมาณค่าความสามารถของผู้สอบ ความคลาดเคลื่อนกำลังสองเฉลี่ยสัมประสิทธิ์สหสัมพันธ์ค่าประมาณความสามารถและค่าความสามารถจริงไม่แตกต่างกัน ความถูกต้องและความแม่นยำจะเพิ่มขึ้นเมื่อความยาวแบบทดสอบเพิ่มขึ้น และอัตราการแสดงข้อสอบเป้าหมายเพิ่มจาก 0.1 เป็น 0.2 ทั้ง 3 วิธีให้การแจกแจงอัตราการแสดงข้อสอบไม่แตกต่างกัน

เดน (Deng. 2001 : Online) ศึกษาวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก (a-STR) ร่วมกับวิธีควบคุมการแสดงผลข้อสอบซิมสัน-เฮตเทอร์ (SH Method) โดยกำหนดอัตราการแสดงข้อสอบในแต่ละชั้นคลังข้อสอบไม่เท่ากัน ชั้นข้อสอบอำนาจจำแนกสูงให้อัตราการแสดงผลข้อสอบกำหนดที่ 0.2 ชั้นข้อสอบอำนาจจำแนกต่ำให้อัตราการแสดงผลข้อสอบกำหนดที่ 0.1 ชั้นตรงกลางให้อัตราการแสดงผลข้อสอบกำหนดที่ 1.5 เพื่อต้องการเพิ่มประสิทธิภาพของแบบทดสอบ ผลการศึกษาพบว่า วิธี a-STR ร่วมกับวิธีควบคุมการแสดงผลข้อสอบซิมสัน-เฮตเทอร์ โดยกำหนดอัตราการแสดงข้อสอบในแต่ละชั้นคลังข้อสอบไม่เท่ากันให้ประสิทธิภาพดีกว่าวิธี a-STR แบบเดิมและวิธีสารสนเทศสูงสุดร่วมกับวิธีควบคุมการแสดงผลข้อสอบซิมสัน-เฮตเทอร์

เชน เอนเคนแมน และชาง (Chen; Ankenmann; & Chang. 2000) ทำการศึกษาเพื่อสนับสนุนการใช้ข้อสอบอำนาจจำแนกต่ำในระยะเริ่มต้นการทดสอบ จากเหตุผลที่ว่า เมื่อค่าประมาณความสามารถไกลจากค่าความสามารถจริงของผู้สอบไม่ควรใช้ข้อสอบอำนาจจำแนกสูงตามวิธีสารสนเทศสูงสุดของฟิชเชอร์ โดยศึกษาเปรียบเทียบวิธีการเลือกข้อสอบในระยะแรกของการทดสอบแบบปรับเหมาะ 5 วิธี คือ 1. วิธีสารสนเทศสูงสุดของฟิชเชอร์ (Fisher Information : FI) 2. วิธีสารสนเทศช่วงระยะของฟิชเชอร์ (Fisher Interval Information : FII) 3. วิธีสารสนเทศของฟิชเชอร์ร่วมกับการแจกแจงภายหลัง (Fisher Information with a Posterior Distribution : FIP) 4. วิธีสารสนเทศโดยรวมเฉลี่ย (Kullback-Leibler Information : KL) 5. วิธีสารสนเทศโดยรวมเฉลี่ยร่วมกับการแจกแจงภายหลัง (Kullback-Leibler Information with a Posterior Distribution : KLP) เพื่อศึกษาประสิทธิภาพและความแม่นยำของการประมาณค่าความสามารถของผู้สอบระยะเริ่มต้นของการทดสอบแบบปรับเหมาะ ประเมินผลด้วย ความลำเอียงเฉลี่ย รากกำลังสองของความคลาดเคลื่อนกำลังสองเฉลี่ย ความคลาดเคลื่อนกำลังสองเฉลี่ย ผลการศึกษาพบว่า ในระยะเริ่มต้นของการทดสอบแบบปรับเหมาะ วิธีการเลือกข้อสอบอื่นๆ ปฏิบัติได้ดีกว่าวิธีสารสนเทศสูงสุดของฟิชเชอร์เมื่อใช้ข้อสอบไม่มากนัก แต่เมื่อความยาวแบบทดสอบเพิ่มขึ้นมากกว่า 10 ข้อ ทุกวิธีการเลือกข้อสอบให้ผลไม่แตกต่างกัน

พาร์เชล ฮาร์มีส และโครมเลย์ (Parshall; Harnes; & Kromrey. 2000, Fall) สังเกตวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับขั้นของค่าอำนาจจำแนก (a-STR) ร่วมกับวิธีควบคุมการแสดงข้อสอบซิมสัน-เฮดเทอร์ (SH Method) แม้ว่าจะให้สมดุลการใช้ข้อสอบในคลังข้อสอบแต่ยังพบข้อสอบบางข้อมีการใช้มากเกินไป จึงได้เสนอวิธีการหยุดข้อสอบ (Item Freezing) ร่วมกับ a-STR และศึกษาผลของจำนวนขั้นต่อ a-STR เพื่อเปรียบเทียบระดับการควบคุมการแสดงข้อสอบ ข้อสอบอัตราการแสดงมากเกินไปจะถูกหยุดกะทันหันหรือไม่ถูกนำไปใช้ในการเลือก ผลการศึกษาพบว่า วิธี a-STR ร่วมกับวิธีการหยุดข้อสอบจะควบคุมอัตราการแสดงข้อสอบได้ดีตามกำหนดโดยไม่สูญเสียความแม่นยำของแบบทดสอบ จำนวนของขั้นต่างๆ (4 6 และ 8 ขั้น) ใน a-STR ไม่ส่งผลต่อการศึกษามากนัก

เวน ชาง และฮัว (Wen; Chang; & Hau. 2001 : Online) ศึกษาจำนวนขั้นที่เหมาะสมของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับขั้นของค่าอำนาจจำแนก (a-STR) จากข้อสังเกตว่า ถ้าจำนวนขั้นน้อยที่สุดเท่ากับ 1 การเลือกข้อสอบจากค่าความยาก (b) ให้ใกล้เคียงกับค่าประมาณความสามารถให้ประสิทธิภาพน้อยกว่าวิธีสารสนเทศสูงสุด แต่ถ้าจำนวนขั้นมากเกินไปอาจไม่มีจำนวนข้อสอบความยากระดับต่างๆ เพียงพอในแต่ละขั้น การศึกษาทำในคลังข้อสอบ 200

400 และ 800 ข้อ ความยาวแบบทดสอบ 48 ข้อ จำนวนชั้นคลังข้อสอบ คือ 1 2 3 4 6 8 12 16 24 และ 48 ชั้น ผลการศึกษาพบว่า ประสิทธิภาพการประมาณค่าความสามารถของผู้สอบและสมมูลการใช้ข้อสอบจะไม่เพิ่มขึ้นหรือลดลงตามจำนวนชั้น จำนวนชั้นที่ใช้ในแต่ละการประยุกต์วิธีขึ้นกับโครงสร้างคลังข้อสอบ ความยาวแบบทดสอบ และเงื่อนไขในการทดสอบอื่นๆ

ซางและยิง (Chang; & Ying. 2001, July : Online) เสนอว่า ความยาวแบบทดสอบควรขึ้นกับผู้สอบแต่ละคนในการทดสอบแบบปรับเหมาะ จึงได้พัฒนาการใช้ทฤษฎีกลุ่มตัวอย่างขนาดใหญ่ (Large Sample Theory) เพื่อกำหนดความยาวแบบทดสอบแปรผัน (Variable Length) และประมาณค่าเป็นจุดหรือช่วงความสามารถ ศึกษาในสถานการณ์จำลองซึ่งแสดงการประมาณกลุ่มตัวอย่างขนาดใหญ่ที่ใช้กับขนาดกลุ่มตัวอย่างในความเป็นจริง เนื่องจากการศึกษายังอยู่ในระยะเริ่มแรก ข้อมูลยังต้องได้รับการสนับสนุนจากงานวิจัยต่อไป

ฮัว และซาง (Hua; & Chang. 2001) ทดสอบการเลือกข้อสอบวิธี a-DSTR (Descending a-Stratified Method) และวิธี N-STR (Non Systematic Stratified Method) เปรียบเทียบกับวิธี การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก (a-STR) เมื่อ a-DSTR คือ วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก (a-STR) เรียงระดับค่าอำนาจจำแนก (a) หักกลับ คือ ชั้นคลังข้อสอบแรกจะบรรจุข้อสอบอำนาจจำแนกสูง และค่าอำนาจจำแนกลดลงตามชั้นจนถึงชั้นสุดท้ายบรรจุข้อสอบอำนาจจำแนกต่ำ N-STR คือ วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก (a-STR) ชั้นแรกบรรจุด้วยข้อสอบอำนาจจำแนกปานกลาง ชั้นท้ายๆ บรรจุด้วยข้อสอบอำนาจจำแนกสูงและต่ำ ผลการศึกษาพบว่า วิธี a-STR และ วิธี a-DSTR ให้ความคลาดเคลื่อนกำลังสองเฉลี่ยและความลำเอียงเฉลี่ยมีแนวโน้มลดลงเหมือนกันเมื่อความยาวแบบทดสอบเพิ่มขึ้น แต่ความคลาดเคลื่อนกำลังสองเฉลี่ยในวิธี a-STR เล็กกว่า วิธี a-DSTR ซึ่งสนับสนุนงานของซางและยิง (Chang; & Ying. 1999) ว่า ข้อสอบอำนาจจำแนกต่ำจะให้สารสนเทศโดยรวมได้มากกว่าในระยะเริ่มต้นของการทดสอบ เมื่อค่าประมาณความสามารถไกลจากค่าความสามารถจริง ความลำเอียงให้ผลด้านบวก ด้านลบ และเข้าใกล้ 0 เมื่อความยาวแบบทดสอบเพิ่มขึ้น วิธี a-STR ดีกว่าวิธี a-DSTR เพราะให้สมมูลการใช้ข้อสอบดีกว่า ส่วนวิธี N-STR ให้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยกึ่งกลางระหว่างวิธี a-STR และวิธี a-DSTR

เอจเจน (Eggen. 2001 : Online) ศึกษาในสถานการณ์จำลองถึงวิธีการแก้ปัญหาการแสดงข้อสอบมากเกินไปและปัญหาการแสดงข้อสอบต่ำเกินไปในการทดสอบแบบปรับเหมาะ โดยประยุกต์รวมการใช้วิธีซิมสัน-เฮตเตอร์ (Simpson-Hetter Method) เพื่อแก้ไขการแสดงข้อสอบมากเกินไปและ

วิธีสุ่มเลือกแบบก้าวหน้า (Progressive Method) เพื่อแก้ไขการแสดงข้อสอบต่ำเกินไป ผลการศึกษาสรุปว่า ปัญหาทั้ง 2 นี้ได้รับการแก้ไขอย่างมีประสิทธิภาพ

เซงและลีย (Cheng; & Liou. 2003) ศึกษาเปรียบเทียบความถูกต้องในการประมาณค่าความสามารถของผู้สอบและอัตราการแสดงข้อสอบของวิธีการเลือกข้อสอบในการทดสอบแบบปรับเหมาะ 6 วิธี คือ 1. วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก (a-Stratified Method : a-STR) 2. วิธีค่าความยากข้อสอบเหมาะสม (Optimal Item Difficulty : OID) 3. วิธีสารสนเทศสูงสุดของฟิชเชอร์ (Maximum Fisher Information : MFI) 4. วิธีสุ่มครึ่งหนึ่งของเมคบริดและมาร์ติน (McBride and Martin's Semi-Random Method : M&M) 5. วิธีสารสนเทศคาดหวังสูงสุด (Maximum Expected Information : MEI) และ 6. วิธีเกณฑ์ใกล้เคียงที่สุด (Nearest - Neighbors Criterion : NN) ซึ่งนำเสนอในการศึกษาครั้งนี้ประยุกต์แนวคิดมาจากการรวมวิธี MFI และ OID เข้าด้วยกัน การทดสอบใช้ข้อสอบมาตรฐานต่อวิธีการเลือกข้อสอบ 6 วิธี พิจารณา ความถูกต้องในการประมาณค่าความสามารถของผู้สอบด้วยความคลาดเคลื่อนกำลังสองเฉลี่ย อัตราการแสดงข้อสอบ การแจกแจงอัตราการแสดงข้อสอบด้วยสถิติไคสแควร์ ผลการศึกษาพบว่า ที่ระดับค่าความสามารถเฉลี่ยต่างๆ วิธี OID จะให้อัตราการแสดงข้อสอบที่ดีที่สุดแต่ให้ความคลาดเคลื่อนกำลังสองเฉลี่ยขนาดใหญ่ ผลนี้จะมีลักษณะใกล้เคียงกับวิธี NN และ a-STR

บทที่ 3

วิธีดำเนินการวิจัย

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อศึกษาประสิทธิภาพการประมาณค่าความสามารถของผู้สอบ และประสิทธิภาพสมมูลการใช้ข้อสอบในคลังข้อสอบของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ ตามระดับชั้นของค่าอำนาจจำแนกภายใต้เงื่อนไขต่างกัน ได้แก่ จำนวนข้อสอบในแต่ละชั้นคลังข้อสอบ การควบคุมการแสดงข้อสอบ ขนาดคลังข้อสอบ และความยาวแบบทดสอบ โดยใช้วิธีการจำลอง ข้อมูล โดยผู้วิจัยได้ดำเนินการวิจัยดังต่อไปนี้

กลุ่มตัวอย่าง

กลุ่มตัวอย่างสำหรับการศึกษาในครั้งนี้ ได้แก่ ค่าความสามารถจริง (θ) ของกลุ่มตัวอย่าง ผู้สอบสร้างขึ้นโดยสุ่มการแจกแจงโคงปกติมาตรฐาน ซึ่งมีค่าเฉลี่ยเท่ากับ 0 และส่วนเบี่ยงเบน มาตรฐานเท่ากับ 1 [$\theta \sim N(0,1)$] และกำหนดให้มีค่าตั้งแต่ -3.0 ถึง 3.0 ค่าความสามารถจริงของกลุ่ม ตัวอย่างผู้สอบสำหรับการทดสอบแต่ละวิธีกำหนดมีจำนวนเท่ากับ 5,000 ค่า เมื่อเปลี่ยนวิธีกำหนดจะทำการสุ่มค่าความสามารถจริงของกลุ่มตัวอย่างผู้สอบใหม่ทุกครั้ง

ตัวแปรศึกษา

ตัวแปรต้นที่ผู้วิจัยสนใจทำการศึกษากำหนดไว้ 4 กลุ่มดังนี้

กลุ่มที่ 1 ได้แก่ วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจ จำแนก (a-Stratified Computerized Adaptive Testing : a-STR) 4 วิธีคือ วิธีที่ 1 คือวิธีการทดสอบแบบ ปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก กำหนดอัตราส่วนจำนวนข้อสอบเท่ากัน ทุกชั้น (1 : 1 : 1 : 1) (a-STR1) วิธีที่ 2 คือวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้น ของค่าอำนาจจำแนก กำหนดอัตราส่วนจำนวนข้อสอบชั้นแรกมากกว่าชั้นอื่น (2 : 0.5 : 0.5 : 1) (a-STR2) วิธีที่ 3 คือวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก กำหนดอัตราส่วนจำนวนข้อสอบชั้นสุดท้ายมากกว่าชั้นอื่น (1 : 0.5 : 0.5 : 2) (a-STR3) และวิธีที่ 4 คือ วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก กำหนดอัตราส่วน จำนวนข้อสอบชั้นแรกและชั้นสุดท้ายมากกว่าชั้นอื่น (1.5 : 0.5 : 0.5 : 1.5) (a-STR4)

กลุ่มที่ 2 ได้แก่ การควบคุมการแสดงข้อสอบ 2 ระดับคือ ไม่ควบคุมการแสดงข้อสอบ และ ควบคุมการแสดงข้อสอบด้วยวิธีชิมสัน-สเตเทอร์ หรือเรียกว่า วิธีขยายช่วงระดับชั้น

กลุ่มที่ 3 ได้แก่ ขนาดคลังข้อสอบ 3 ระดับคือ 200 400 และ 600 ข้อ

กลุ่มที่ 4 ได้แก่ ความยาวแบบทดสอบ 3 ระดับคือ 20 32 และ 44 ข้อ

ตัวแปรตาม ที่ผู้วิจัยสนใจทำการศึกษาแบ่งเป็น 2 กลุ่มดังนี้

กลุ่มที่ 1 ประสิทธิภาพการประมาณค่าความสามารถของผู้สอบพิจารณาด้วย ความคลาดเคลื่อนกำลังสองเฉลี่ยและความลำเอียงเฉลี่ย วิเคราะห์ข้อมูลดังนี้

ความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Square Error : MSE) หมายถึง ค่าบอกความถูกต้องของค่าประมาณความสามารถของผู้สอบ แสดงเป็นผลต่างกำลังสองเฉลี่ยระหว่างค่าประมาณความสามารถและค่าความสามารถจริงของผู้สอบ หรือมีค่าเท่ากับความแปรปรวนของค่าประมาณความสามารถของผู้สอบรวมกับความลำเอียงกำลังสอง

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

เมื่อ n แทน จำนวนผู้สอบทั้งหมด

$\hat{\theta}_i$ แทน ค่าประมาณความสามารถของผู้สอบ

θ_i แทน ค่าความสามารถจริงของผู้สอบ

ความลำเอียงเฉลี่ย (Average Bias) หมายถึง ค่าบอกความเที่ยงตรงของค่าประมาณความสามารถของผู้สอบ แสดงเป็นผลต่างเฉลี่ยระหว่างค่าประมาณความสามารถและค่าความสามารถจริงของผู้สอบ บอกทิศทางการประมาณค่าว่าให้ผลสูงหรือต่ำกว่าค่าความสามารถจริงของผู้สอบ ถ้าความลำเอียงกำลังสองเท่ากับ 0 ความคลาดเคลื่อนกำลังสองเฉลี่ยจะเท่ากับความแปรปรวนของค่าประมาณความสามารถของผู้สอบ

$$Bias = \frac{1}{n} \sum_{i=1}^n (\hat{\theta}_i - \theta_i)^2$$

เมื่อ n แทน จำนวนผู้สอบทั้งหมด

$\hat{\theta}_i$ แทน ค่าประมาณความสามารถของผู้สอบ

θ_i แทน ค่าความสามารถจริงของผู้สอบ

กลุ่มที่ 2 ประสิทธิภาพสมมูลการใช้ข้อสอบในคลังข้อสอบพิจารณาด้วย ข้อสอบที่มีการแสดงมากเกินไป ข้อสอบที่มีการใช้ต่ำ อัตราการทับซ้อนข้อสอบ และการแจกแจงอัตราการแสดงข้อสอบ วิเคราะห์ข้อมูลดังนี้

ข้อสอบที่มีการแสดงมากเกินไป หมายถึง ข้อสอบใดๆ ที่มีอัตราการแสดงข้อสอบสูงกว่า 0.2 เมื่ออัตราการแสดงข้อสอบ หมายถึง จำนวนครั้งของข้อสอบที่จัดให้ผู้สอบต่อจำนวนผู้สอบทั้งหมดที่ทำการทดสอบ

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

ข้อสอบที่มีการใช้ต่ำ หมายถึง ข้อสอบใดๆ ที่มีอัตราการแสดงข้อสอบต่ำกว่า 0.2

อัตราการทับซ้อนข้อสอบ หมายถึง สัดส่วนของข้อสอบที่ใช้ร่วมกันโดยกลุ่มของผู้สอบต่อข้อสอบที่ใช้ร่วมกันโดยกลุ่มของผู้สอบที่เป็นไปได้ทั้งหมด

เมื่อกำหนดให้ความยาวของแบบทดสอบคือ L ข้อ และจำนวนผู้สอบคือ P คน อัตราการทับซ้อนข้อสอบคำนวณโดย 1. นับจำนวนข้อสอบที่ใช้ร่วมกันแต่ละกลุ่มของผู้สอบ $P(P-1)/2$ คู่ 2. บวกจำนวนข้อสอบที่ใช้ร่วมกัน $P(P-1)/2$ คู่ และ 3.หารจำนวนทั้งหมดด้วย $L P(P-1)/2$

การแจกแจงอัตราการแสดงข้อสอบ หมายถึง ลักษณะการกระจายอัตราการแสดงของข้อสอบในคลังข้อสอบ บ่งบอกถึงประสิทธิภาพสมมูลการใช้ข้อสอบในคลังข้อสอบโดยรวม พิจารณาจากสถิติไคสแควร์

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

er_j แทน อัตราการแสดงข้อสอบสำหรับข้อสอบที่ j

L แทน ความยาวแบบทดสอบ

N แทน ขนาดคลังข้อสอบ

L/N แทน อัตราการแสดงข้อสอบที่ต้องการ

สมการแสดงถึงความแตกต่างระหว่างค่าสังเกตและค่าคาดหวังของอัตราการแสดงข้อสอบ
การเปรียบเทียบการแจกแจงอัตราการแสดงข้อสอบ 2 วิธี ได้แก่ วิธีที่ 1 และวิธีที่ 2 โดยใช้
อัตราส่วน F (F ratio) จะพิจารณาจากอัตราส่วนของค่า χ^2 วิธีที่ 1 หารด้วย χ^2 วิธีที่ 2 นั่นคือ

$$F_{1,2} = \chi_1^2 / \chi_2^2$$

เมื่อ $F_{1,2}$ แทน อัตราส่วนการแจกแจงอัตราการแสดงข้อสอบ 2 วิธี หรือ วิธีที่ 1 หารด้วยวิธีที่ 2
 χ_1^2 แทน การแจกแจงอัตราการแสดงข้อสอบวิธีที่ 1
 χ_2^2 แทน การแจกแจงอัตราการแสดงข้อสอบวิธีที่ 2
 ถ้า $F < 1$ แสดงว่า วิธีที่ 1 ให้สมดุลของอัตราการแสดงข้อสอบโดยรวมดีกว่าวิธีที่ 2

เครื่องมือใช้ในการวิจัย

ผู้วิจัยได้ออกแบบโปรแกรมคอมพิวเตอร์สร้างเป็นเครื่องมือสำหรับการทดสอบแบบ
ปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก โดยดำเนินการตามขั้นตอนดังนี้

ขั้นตอนที่ 1 ศึกษาความมุ่งหมายการวิจัย นำไปวิเคราะห์สิ่งที่ต้องการศึกษาและรวบรวม
ข้อมูลที่เกี่ยวข้อง กำหนดความชัดเจนของข้อมูล พิจารณาความสัมพันธ์ระหว่างตัวแปรต้นและตัวแปร
ตามของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก

ขั้นตอนที่ 2 เขียนผังขั้นตอนการทดสอบ โดยกำหนดลักษณะหรือโครงร่างการทดสอบ
ครอบคลุมสิ่งที่ต้องการศึกษา แสดงความสัมพันธ์ของขั้นตอนการทดสอบตั้งแต่เริ่มต้นจนถึงสิ้นสุด
การทดสอบ คาดการณ์ความเป็นไปได้ในการทดสอบและวิธีเก็บรวบรวมผลการทดสอบ

ขั้นตอนที่ 3 ออกแบบโปรแกรมคอมพิวเตอร์ด้วยภาษาเบสิก (Basic Language หรือ
Beginner's All-Purpose Symbolic Instruction Code) และใช้ภาษาเบสิกชนิดคิวเบสิก (Q-Basic) ซึ่ง
เป็นภาษาที่เขียนได้ง่าย สามารถปฏิบัติการคำนวณขั้นสูงและใช้กับเครื่องคอมพิวเตอร์ระบบ
ไมโครคอมพิวเตอร์ได้ โปรแกรมออกแบบขึ้นสำหรับจัดกระทำข้อมูลที่ใช้ในการวิจัย โดยเขียนคำสั่ง
ต่างๆ ที่ละขั้นตอนการทำงานเป็นไปตามลำดับและมีความถูกต้องครบถ้วนตามผังขั้นตอนการทดสอบ

ขั้นตอนที่ 4 ทดสอบและปรับปรุงโปรแกรม การตรวจสอบทำตามขั้นตอนการทำงานของ
โปรแกรมโดยติดตามการทำงานทีละคำสั่งจนกว่าจะจบคำสั่ง ทดสอบโดยการควบคุมเครื่อง
คอมพิวเตอร์ให้ทำงานตามคำสั่งด้วยข้อมูลที่รู้ค่าผลลัพธ์แล้วและตรวจสอบว่าได้ผลตรงกันหรือไม่
ถ้าพบข้อบกพร่องจะทำการปรับปรุงแก้ไขและนำไปทดสอบซ้ำ เมื่อแน่ใจว่าการทำงานของโปรแกรม
ถูกต้องและผลการทดสอบแม่นยำเพียงพอสำหรับเก็บข้อมูลตัวแปรตามได้ครบถ้วน จึงนำไปใช้
ทดสอบในงานวิจัยต่อไป

การรวบรวมข้อมูล

การรวบรวมข้อมูลกระทำโดยโปรแกรมคอมพิวเตอร์ภาษาคิวเบสิก เพื่อรวบรวมผลที่เกิดขึ้นจากตัวแปรต้นและรายงานผลข้อมูลตัวแปรตาม เมื่อผ่านการทดสอบที่แต่ละชั้นของคลังข้อสอบและภายหลังการทดสอบเสร็จสิ้นในแต่ละวิธีกำหนด นำผลข้อมูลตัวแปรตามมาพิจารณาตรวจสอบความสมบูรณ์ครบถ้วนก่อนนำข้อมูลไปดำเนินการต่อไป

การวิเคราะห์ข้อมูล

การวิเคราะห์ข้อมูลทำการปฏิบัติการของโปรแกรมคอมพิวเตอร์ภาษาคิวเบสิก ผลการวิเคราะห์ข้อมูลได้รับเมื่อการทดสอบผ่านแต่ละชั้นคลังข้อสอบและเมื่อสิ้นสุดการทดสอบ

รูปแบบของการวิจัย

ในการวิจัยครั้งนี้ผู้วิจัยทำการศึกษาค้นคว้าประกอบสำคัญของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์และจุดมุ่งหมายการวิจัย นำมากำหนดรูปแบบการวิจัยมีรายละเอียดดังนี้

1. การกำหนดข้อตกลงเบื้องต้น

การวิจัยได้กำหนดข้อตกลงเบื้องต้น ได้แก่ การทดสอบอาศัยหลักการตามทฤษฎีการตอบข้อสอบ (Item Response Theory : IRT) โดยใช้รูปแบบโลจิสติก 3 พารามิเตอร์ (3 Parameter Logistic Model : 3 PLM) ผลการตอบข้อสอบมี 2 ค่า คือ ตอบถูกต้องได้คะแนนเท่ากับ 1 ตอบผิดได้คะแนนเท่ากับ 0 (Dichotomous Scoring) และกำหนดให้คลังข้อสอบมีความเป็นมิติเดียว (Unidimension) โดยการวิจัยศึกษาในสถานการณ์จำลอง

2. การสร้างคลังข้อสอบ

การสร้างคลังข้อสอบเริ่มจากแบ่งชั้นคลังข้อสอบเป็น 4 ชั้นตามระดับค่าอำนาจจำแนกข้อสอบ ชั้นแรกบรรจุด้วยข้อสอบอำนาจจำแนกต่ำสุดเท่ากับ 0.5 ชั้นที่ 2 บรรจุด้วยข้อสอบอำนาจจำแนกเท่ากับ 1.0 ชั้นที่ 3 บรรจุด้วยข้อสอบอำนาจจำแนกเท่ากับ 1.5 และชั้นสุดท้ายบรรจุด้วยข้อสอบอำนาจจำแนกสูงสุดเท่ากับ 2.0 จำนวนข้อสอบแตกต่างกันคิดเป็นอัตราส่วนจำนวนข้อสอบในแต่ละชั้นคลังข้อสอบเป็นไปตามวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นค่าอำนาจจำแนก (a-STR) 4 วิธี

การกำหนดขนาดคลังข้อสอบเริ่มต้นจาก 200 400 และ 600 ข้อ เมื่อทำการทดสอบในเบื้องต้นพบว่า สามารถให้ผลการทดสอบได้ครบเมื่อไม่ได้ใช้ร่วมกับการควบคุมการแสดงข้อสอบ แต่ผลการ

ทดสอบจะแตกต่างออกไปเมื่อวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกใช้ร่วมกับการควบคุมการแสดงผลข้อสอบ บางครั้งพบว่าไม่สามารถให้ผลการทดสอบได้ สาเหตุมาจากจำนวนข้อสอบในชั้นของค่าอำนาจจำแนกมีจำนวนไม่เพียงพอ ได้แก่ ชั้นที่ 2 และ 3 ของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกวิธีที่ 2 (a-STR2) วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกวิธีที่ 3 (a-STR3) และวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกวิธีที่ 4 (a-STR4) จำนวนข้อสอบน้อยเกินไปมีผลทำให้พารามิเตอร์การแสดงผลข้อสอบของข้อสอบแต่ละข้อมีค่าต่ำลงมาก เนื่องจากการใช้ข้อสอบมากเกินไป ผลทำให้มีข้อสอบให้เลือกใช้น้อยลง แบบทดสอบไม่สามารถเลือกข้อสอบความยากใกล้เคียงค่าประมาณความสามารถของผู้สอบสำหรับการทดสอบได้ เหตุการณ์นี้มักเกิดขึ้นร่วมกับการใช้ขนาดคลังข้อสอบเล็กเกินไป และความยาวข้อสอบมากเกินไป

จากเหตุผลดังกล่าว ผู้วิจัยได้เพิ่มการศึกษาที่ขนาดคลังข้อสอบ 800 ข้อ เพื่อต้องการศึกษาและเปรียบเทียบผลที่ได้รับของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกแต่ละวิธีร่วมกับการควบคุมการแสดงผลข้อสอบ การเพิ่มขนาดคลังข้อสอบ 800 ข้อคาดว่า ขนาดคลังข้อสอบ 800 ข้อ สามารถเพิ่มจำนวนข้อสอบในแต่ละชั้นคลังข้อสอบให้มีมากเพียงพอที่แต่ละระดับความสามารถบนเส้นต่อเนื่อง ให้พารามิเตอร์การแสดงผลข้อสอบของข้อสอบแต่ละข้อยังคงมีค่าใกล้เคียงหรือเท่ากับ 1 เพิ่มโอกาสที่แบบทดสอบจะสามารถเลือกข้อสอบความยากใกล้เคียงค่าประมาณความสามารถของผู้สอบและมีข้อสอบให้เลือกจำนวนมากเพียงพอสำหรับการทดสอบ

ตาราง 2 จำนวนข้อสอบในแต่ละชั้นคลังข้อสอบของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก 4 วิธี

วิธี	ชั้นค่าอำนาจ จำแนก (a)	ขนาดคลังข้อสอบ			
		200 ข้อ	400 ข้อ	600 ข้อ	800 ข้อ
a-STR1 1:1:1:1	a 0.5	50	100	150	200
	a 1.0	50	100	150	200
	a 1.5	50	100	150	200
	a 2.0	50	100	150	200
a-STR2 2:0.5:0.5:1.0	a 0.5	100	200	300	400
	a 1.0	25	50	75	100
	a 1.5	25	50	75	100
	a 2.0	50	100	150	200
a-STR3 1.0:0.5:0.5:2	a 0.5	50	100	150	200
	a 1.0	25	50	75	100
	a 1.5	25	50	75	100
	a 2.0	100	200	300	400
a-STR4 1.5:0.5:0.5:1.5	a 0.5	75	150	225	300
	a 1.0	25	50	75	100
	a 1.5	25	50	75	100
	a 2.0	75	150	225	300

จำนวนข้อสอบตามตาราง 2 สรุปดังนี้

จำนวนข้อสอบในแต่ละชั้นคลังข้อสอบมาจากจำนวนข้อสอบทั้งหมดในคลังข้อสอบแบ่งเป็น 4 กลุ่ม จำนวนข้อสอบแต่ละกลุ่มแตกต่างกันตามวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก 4 วิธี เช่น เมื่อใช้ขนาดคลังข้อสอบ 200 ข้อ วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกวิธีที่ 1(a-STR1) กำหนดอัตราส่วนจำนวนข้อสอบเท่ากันทุกชั้นเท่ากับ 1 : 1 : 1 : 1 จำนวนข้อสอบเท่ากับ 50 50 50 และ 50 ข้อตามระดับชั้นของค่าอำนาจจำแนก วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกวิธีที่ 2 (a-STR2) กำหนดอัตราส่วนจำนวนข้อสอบชั้นแรกมากกว่าชั้นอื่นเท่ากับ 2 : 0.5 : 0.5 : 1 จำนวน

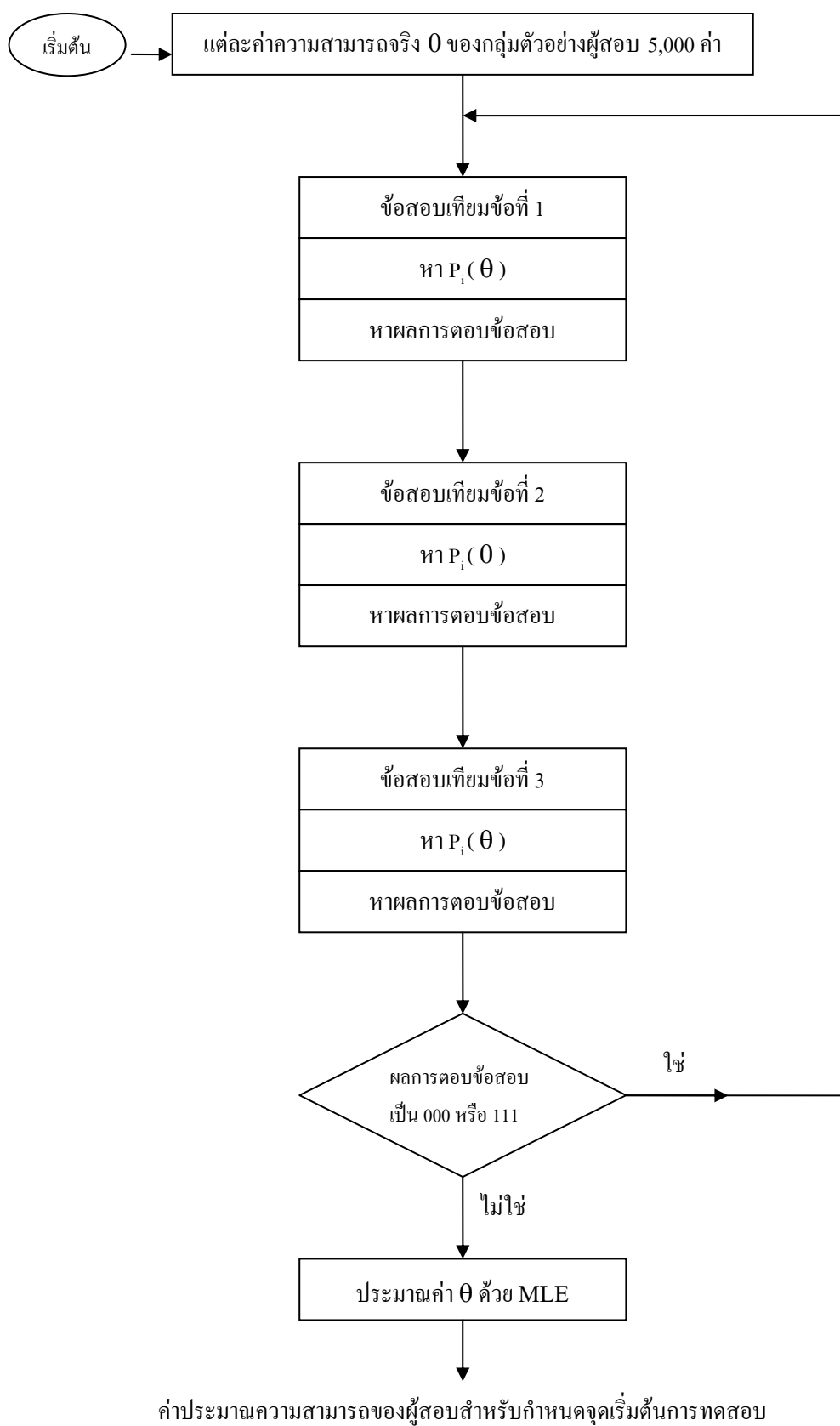
ข้อสอบเท่ากับ 100 25 25 และ 50 ข้อตามระดับชั้นของค่าอำนาจจำแนก วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกวิธีที่ 3 (a-STR3) กำหนดอัตราส่วนจำนวนข้อสอบชั้นสุดท้ายมากกว่าชั้นอื่นเท่ากับ $1 : 0.5 : 0.5 : 2$ จำนวนข้อสอบเท่ากับ 50 25 25 และ 100 ข้อตามระดับชั้นของค่าอำนาจจำแนก และวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกวิธีที่ 4 (a-STR4) กำหนดอัตราส่วนจำนวนข้อสอบชั้นแรกและชั้นสุดท้ายมากกว่าชั้นอื่นเท่ากับ $1.5 : 0.5 : 0.5 : 1.5$ จำนวนข้อสอบเท่ากับ 75 25 25 และ 75 ข้อ เป็นต้น

ข้อสอบแต่ละข้อในคลังข้อสอบกำหนดให้มีพารามิเตอร์ ดังนี้คือ ค่าการเดา (c) ลงที่เท่ากับ 0 ค่าอำนาจจำแนก (a) กำหนด 4 ค่า คือ 0.5 1.0 1.5 และ 2.0 ตามลำดับชั้นคลังข้อสอบ 4 ชั้น ข้อสอบบรรจุในชั้นคลังข้อสอบเดียวกันมีค่าอำนาจจำแนกเท่ากัน ค่าความยาก (b) สร้างโดยสุ่มจากการแจกแจงโค้งปกติมาตรฐาน มีค่าเฉลี่ยเท่ากับ 0 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1 บนมาตรวัดเดียวกันกับเส้นต่อเนื่องค่าความสามารถจริงของกลุ่มตัวอย่างผู้สอบ การแจกแจงค่าความยากตั้งแต่ -3.0 ถึง 3.0

3. การกำหนดจุดเริ่มต้นการทดสอบ (Entry Point)

การกำหนดจุดเริ่มต้นการทดสอบกระทำก่อนเริ่มต้นการทดสอบจริงโดยใช้ข้อสอบเทียม 3 ข้อ ประยุกต์มาจากการวิจัยของชางและยิง (Chang; & Ying, 1999) เริ่มจากการเตรียมข้อสอบเทียมโดยให้พารามิเตอร์ข้อสอบเทียมข้อแรกมีค่าอำนาจจำแนก (a) เท่ากับ 1 ค่าการเดา (c) เท่ากับ 0.2 และค่าความยาก (b) ได้จากการสุ่มจากการแจกแจงโค้งปกติมาตรฐานมีค่าเฉลี่ยเท่ากับ 0 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1 $[N(0,1)]$ บนมาตรวัดเดียวกันกับเส้นต่อเนื่องค่าความสามารถจริงของกลุ่มตัวอย่างผู้สอบ การแจกแจงค่าความยากตั้งแต่ -3.0 ถึง 3.0

แต่ละค่าความสามารถจริงของผู้สอบจะได้รับการทดสอบด้วยข้อสอบเทียม 3 ข้อ เพื่อหาความน่าจะเป็นในการตอบข้อสอบและผลการตอบข้อสอบ ก่อนนำไปคำนวณหาค่าประมาณความสามารถของผู้สอบเพื่อนำไปใช้สำหรับกำหนดจุดเริ่มต้นการทดสอบต่อไป



ภาพประกอบ 4 ขั้นตอนการกำหนดจุดเริ่มต้นการทดสอบด้วยข้อสอบเทียม 3 ข้อ

ขั้นตอนในภาพประกอบ 4 สรุปดังนี้

แต่ละค่าความสามารถจริงของผู้สอบจะได้รับข้อสอบเทียมข้อแรกมีค่าความยากใกล้เคียงค่าความสามารถจริงของผู้สอบมากที่สุด นำไปหาความน่าจะเป็นในการตอบข้อสอบถูกต้องและหาผลการตอบข้อสอบ ถ้าผลการตอบข้อสอบข้อแรกถูกต้อง ข้อสอบข้อที่ 2 มีความยากมากขึ้น คือ $b_2 = b_1 + 2$ แต่ถ้าผลการตอบข้อสอบข้อแรกผิด ข้อสอบข้อที่ 2 มีความยากลดลง คือ $b_2 = b_1 - 2$ ข้อสอบเทียมข้อที่ 3 ปฏิบัติคล้ายกัน ค่า a และ c ไม่เปลี่ยนแปลงระหว่างการทดสอบข้อสอบเทียม เมื่อดำเนินการจนครบ 3 ข้อ นำไปหาค่าประมาณความสามารถด้วยวิธีการประมาณค่าความเป็นไปได้สูงสุด (Maximum Likelihood Estimation : MLE) จะได้ค่าความสามารถเบื้องต้นใช้เป็นตัวกำหนดจุดเริ่มต้นการทดสอบต่อไป ถ้าผลการตอบข้อสอบเทียมทั้ง 3 ข้อถูกต้องหมดหรือผิดหมด จะต้องทำการทดสอบซ้ำเพื่อกำหนดจุดเริ่มต้นการทดสอบใหม่ด้วยข้อสอบเทียม 3 ข้อตามขั้นตอนที่กล่าวมา

ความน่าจะเป็นในการตอบข้อสอบถูกต้องใช้รูปแบบโลจิสติก 3 พารามิเตอร์ คำนวณได้จาก

$$P_i(\theta) = c_i + (1 - c_i) \frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}} \quad i = 1, 2, \dots, n$$

เมื่อ P_i แทน ความน่าจะเป็นที่ผู้สอบระดับความสามารถที่ θ ตอบข้อสอบข้อที่ i ได้ถูกต้อง

θ แทน ระดับความสามารถจริงของผู้สอบ มีค่าอยู่ระหว่าง -3.0 ถึง +3.0

a_i แทน ค่าอำนาจจำแนกของข้อสอบข้อที่ i

b_i แทน ค่าความยากของข้อสอบข้อที่ i มีค่าอยู่ระหว่าง -3.0 ถึง +3.0

c_i แทน ค่าการเดาของข้อสอบข้อที่ i

n แทน จำนวนข้อสอบในแบบทดสอบ

e แทน ค่าคงที่ของลอการิทึมฐานธรรมชาติ (Natural Log) มีค่าเท่ากับ 2.718

ค่า D เป็นค่าคงที่เพื่อใช้ปรับสเกลให้ฟังก์ชันโลจิสติกและฟังก์ชันนอร์มัลโอไบต์ใกล้เคียงกันมากที่สุดมีค่าเท่ากับ 1.7 จะไม่ใช้ในการวิจัยครั้งนี้ เพราะในทางปฏิบัติถือว่าค่าคงที่ D รวมอยู่ในค่าอำนาจจำแนก (a) อยู่แล้ว (Embretson & Reise, 2000 : 159)

ผลการตอบข้อสอบเมื่อกำหนดให้มี 2 ค่า คือตอบถูกต้องเท่ากับ 1 ตอบผิดเท่ากับ 0 (Dichotomous Scoring) หาได้จาก

ให้ความน่าจะเป็นของการตอบข้อสอบถูกต้อง $P_i(\theta)$ มีการแจกแจงเป็นความถี่สะสมจาก 0 ถึง 1 สร้างตัวเลขสุ่มจากการแจกแจงแบบเดียวกัน (0 ถึง 1) และใช้โปรแกรมคอมพิวเตอร์ภาษาคิวเบสิกทดสอบสุ่มตัวเลขสุ่มขึ้นมา

ถ้าผล $P_i(\theta) \geq$ ตัวเลขสุ่ม ให้ผลการตอบข้อสอบเป็น 1

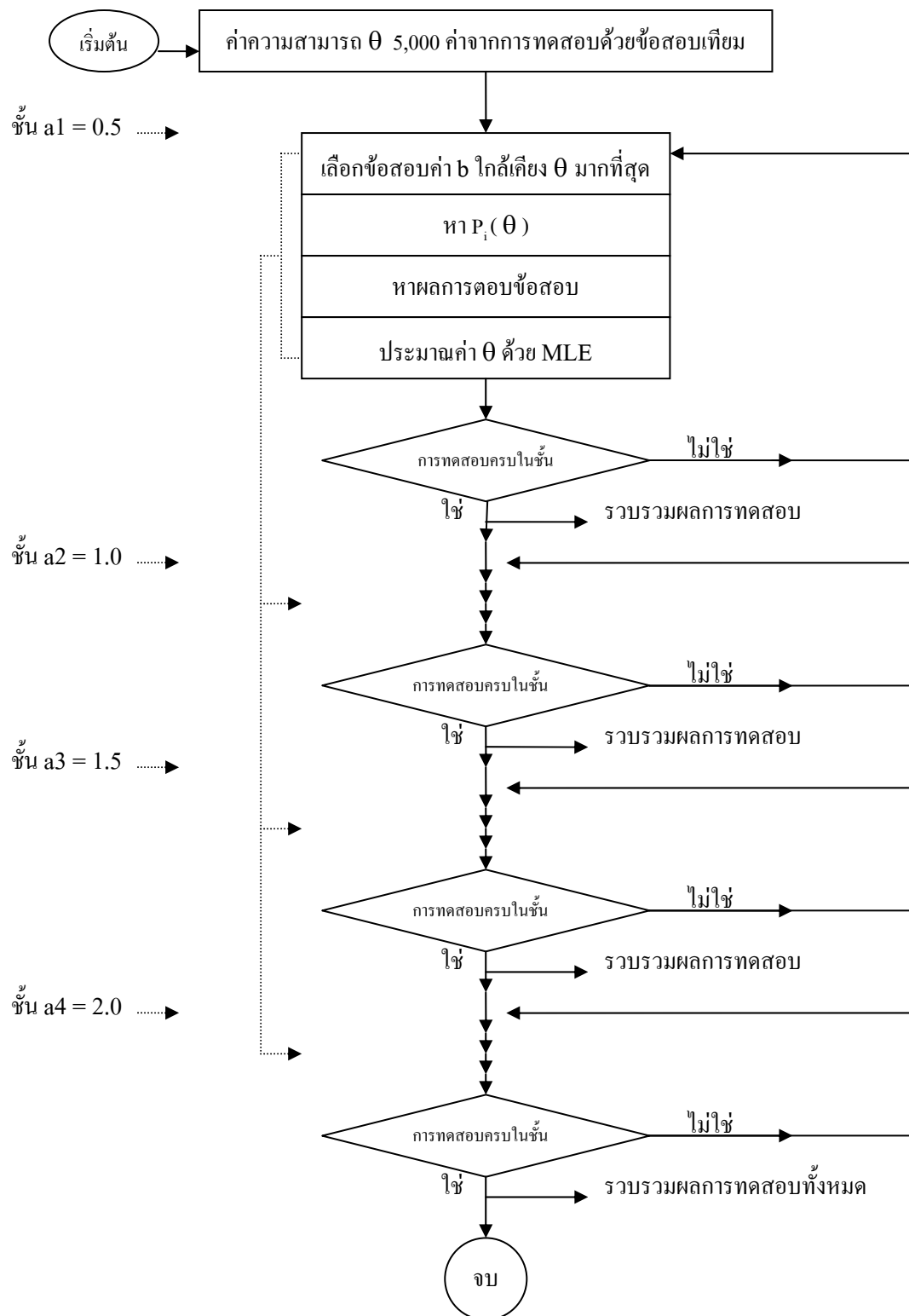
ถ้าผล $P_i(\theta) <$ ตัวเลขสุ่ม ให้ผลการตอบข้อสอบเป็น 0

4. กำหนดวิธีการประมาณค่าความสามารถ

การประมาณค่าความสามารถใช้วิธีการประมาณค่าความเป็นไปได้สูงสุด (Maximum Likelihood Estimation : MLE) เนื่องจากการใช้ข้อสอบเทียม 3 ข้อในการกำหนดจุดเริ่มต้นการทดสอบให้ผลการตอบข้อสอบไม่เป็น 0 ทั้งหมดหรือ 1 ทั้งหมด ดังนั้นจึงสามารถใช้วิธีการประมาณค่าความเป็นไปได้สูงสุดตลอดระยะเวลาการทดสอบ

5. กำหนดวิธีการเลือกข้อสอบ

การเลือกข้อสอบใช้วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก (a-Stratified Method : a-STR) ประยุกต์มาจากผลงานวิจัยของ ชาง และยิง (Chang; & Ying. 1999) มีขั้นตอนดำเนินการ ดังนี้



ภาพประกอบ 5 ขั้นตอนวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์
ตามระดับขั้นของค่าอำนาจจำแนก

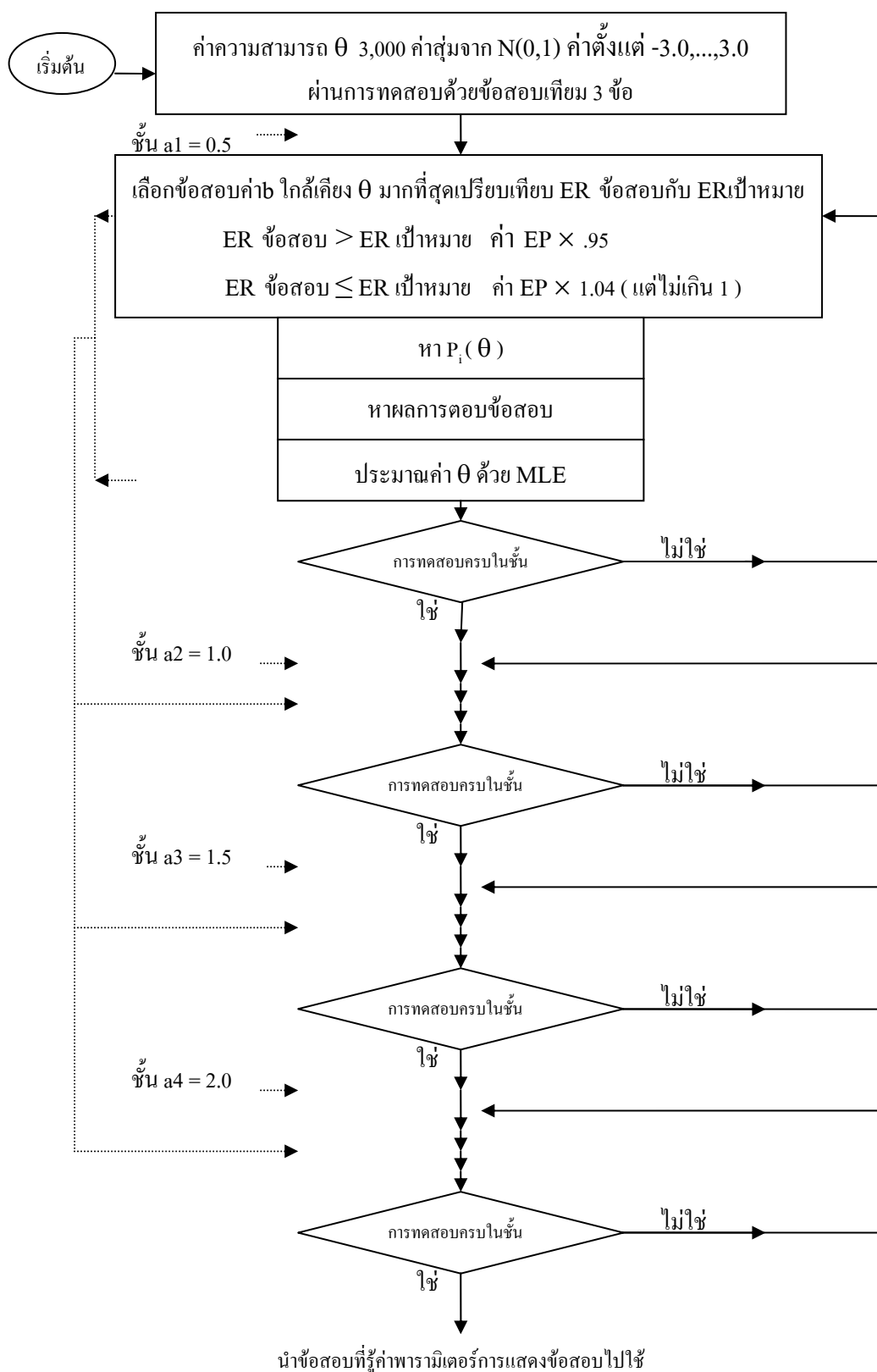
ขั้นตอนในภาพประกอบ 5 สรุปดังนี้

ขั้นตอนการเลือกข้อสอบของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก (a-Stratified Method : a-STR) เริ่มจากแบ่งคลังข้อสอบเป็น 4 ชั้นตามค่าอำนาจจำแนกข้อสอบ (พารามิเตอร์ a) ชั้นแรกบรรจุข้อสอบอำนาจจำแนกต่ำสุดเท่ากับ 0.5 ชั้นที่ 2 บรรจุข้อสอบอำนาจจำแนกเท่ากับ 1.0 ชั้นที่ 3 บรรจุข้อสอบอำนาจจำแนกเท่ากับ 1.5 ชั้นสุดท้ายบรรจุข้อสอบอำนาจจำแนกต่ำสุดเท่ากับ 2.0 แบ่งชั้นการทดสอบหรือความยาวแบบทดสอบต่างๆ กันตามการแบ่งชั้นของคลังข้อสอบ

แต่ละค่าความสามารถของผู้สอบ 5,000 ค่าหลังจากผ่านการทดสอบด้วยข้อสอบเทียม 3 ข้อ นำมากำหนดจุดเริ่มต้นการทดสอบ เลือกข้อสอบความยากใกล้เคียงกับค่าประมาณความสามารถของผู้สอบมากที่สุดและจัดให้ผู้สอบ หาผลการตอบข้อสอบและหาความน่าจะเป็นในการตอบข้อสอบ ถูกต้องนำไปคำนวณหาค่าประมาณความสามารถของผู้สอบเพื่อนำไปใช้เลือกข้อสอบข้อต่อไป เมื่อทำการทดสอบจนครบตามความยาวแบบทดสอบย่อยในขั้นนั้น จึงเลื่อนการทดสอบขึ้นไปยังขั้นและชั้นคลังข้อสอบต่อไป ทำการทดสอบจนครบทุกชั้นการทดสอบและทุกชั้นคลังข้อสอบ การทดสอบจึงยุติ

6. กำหนดวิธีการควบคุมการแสดงผลข้อสอบ

วิธีการควบคุมการแสดงผลข้อสอบในงานวิจัยครั้งนี้ ใช้วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกร่วมกับวิธีควบคุมการแสดงผลข้อสอบวิธีซิมสัน-สเตเทอร์หรือ วิธีขยายช่วงระดับชั้น (Enhanced Stratified Method) ประยุกต์มาจากผลงานวิจัยของ ลุง ชาง และฮัว (Leung; Chang; & Hua, 2002) เริ่มจากการเตรียมพารามิเตอร์การแสดงผลข้อสอบของข้อสอบแต่ละข้อในสถานการณ์จำลองก่อนนำไปใช้จริงกับวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกร่วมกับวิธีควบคุมการแสดงผลข้อสอบ ขั้นตอนการเตรียมพารามิเตอร์การแสดงผลข้อสอบดำเนินการ ดังนี้

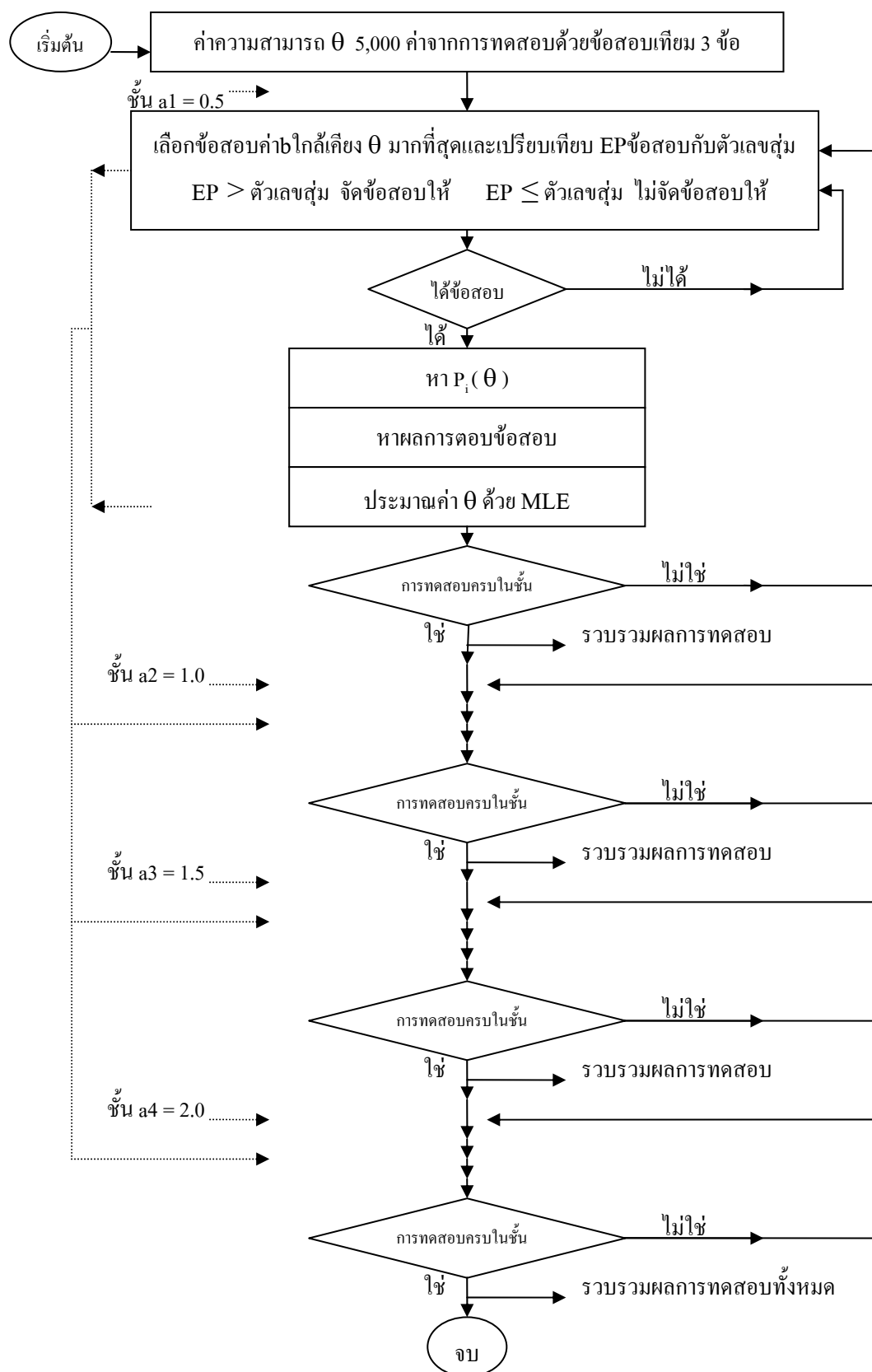


ภาพประกอบ 6 ขั้นตอนการเตรียมพารามิเตอร์การแสดงผลข้อสอบของวิธีการทดสอบแบบปรับ
เหมาะสมด้วยคอมพิวเตอร์ตามระดับขั้นของค่าอำนาจจำแนกร่วมกับการควบคุมการแสดงผลข้อสอบ

ขั้นตอนในภาพประกอบ 6 สรุปดังนี้

การเตรียมพารามิเตอร์การแสดงผล (Exposure Parameter : EP) ของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับขั้นของค่าอำนาจจำแนกร่วมกับการควบคุมการแสดงผล การเตรียมพารามิเตอร์การแสดงผลของข้อสอบแต่ละข้อทำในสถานการณ์จำลอง เริ่มจากกำหนดพารามิเตอร์การแสดงผลของข้อสอบทุกข้อเท่ากับ 1 กำหนดค่าอัตราการแสดงผล (Exposure Rate : ER) สูงสุดเป้าหมายเท่ากับ 0.2 เพื่อใช้เป็นจุดตัด แบ่งคลังข้อสอบเป็น 4 ขั้นตามค่าอำนาจจำแนกข้อสอบ (พารามิเตอร์ a) ขั้นแรกบรรจุข้อสอบอำนาจจำแนกต่ำสุดเท่ากับ 0.5 ขั้นที่ 2 บรรจุข้อสอบอำนาจจำแนกเท่ากับ 1.0 ขั้นที่ 3 บรรจุข้อสอบอำนาจจำแนกเท่ากับ 1.5 ขั้นสุดท้ายบรรจุข้อสอบอำนาจจำแนกต่ำสุดเท่ากับ 2.0 แบ่งขั้นการทดสอบเข้าใน 4 ขั้นของคลังข้อสอบ แต่ละค่าความสามารถของผู้สอบ 3,000 ค่าหลังจากผ่านการทดสอบด้วยข้อสอบเทียม 3 ข้อนำมากำหนดจุดเริ่มต้นการทดสอบ เลือกข้อสอบความยากใกล้เคียงกับค่าประมาณความสามารถของผู้สอบมากที่สุด และจัดให้ผู้สอบ หาผลการตอบข้อสอบและหาความน่าจะเป็นในการตอบข้อสอบถูกต้องนำไปคำนวณค่าประมาณความสามารถของผู้สอบก่อนนำไปใช้เลือกข้อสอบข้อต่อไป เมื่อทำการทดสอบจนครบตามความยาวแบบทดสอบย่อยในขั้นนั้น จึงเลื่อนการทดสอบขึ้นไปยังขั้นและชั้นคลังข้อสอบต่อไป ทำการทดสอบจนครบทุกขั้นการทดสอบและทุกชั้นคลังข้อสอบ การทดสอบจึงยุติ

การเลือกข้อสอบความยากใกล้เคียงกับค่าประมาณความสามารถของผู้สอบมากที่สุดและจัดให้ผู้สอบนั้น การใช้ข้อสอบสำหรับการทดสอบแต่ละครั้ง อัตราการแสดงผลข้อสอบจะนำไปเปรียบเทียบกับอัตราการแสดงผลข้อสอบสูงสุดเป้าหมาย (0.2) ถ้าข้อสอบมีอัตราการแสดงผลข้อสอบมากกว่าอัตราการแสดงผลข้อสอบสูงสุดเป้าหมาย พารามิเตอร์การแสดงผลข้อสอบจะต้องปรับให้ลดลงโดยคูณด้วยแฟกเตอร์น้อยกว่า 1 (เช่น 0.95) ถ้าข้อสอบมีอัตราการแสดงผลข้อสอบน้อยกว่าอัตราการแสดงผลข้อสอบสูงสุดเป้าหมาย พารามิเตอร์การแสดงผลข้อสอบจะต้องปรับให้สูงขึ้นแต่ค่าสูงสุดไม่เกิน 1 โดยคูณด้วยแฟกเตอร์มากกว่า 1 (เช่น 1.04) การใช้ค่าแฟกเตอร์น้อยกว่า 1 โดยใช้ค่า 0.95 หรือการใช้แฟกเตอร์มากกว่า 1 โดยใช้ค่า 1.04 ในการวิจัยนี้ เพื่อต้องการให้พารามิเตอร์การแสดงผลข้อสอบมีค่าใกล้เคียงกับการศึกษาของ ลุง ชาง และฮัว (Leung; Chang; & Hua. 2002) ใช้เป็นแนวทางหนึ่งในการเทียบเคียงผลของการควบคุมการแสดงผลข้อสอบ การปรับพารามิเตอร์การแสดงผลข้อสอบนี้จะดำเนินอย่างต่อเนื่องทุกครั้งที่ข้อสอบข้อนั้นได้รับเลือกและจัดให้ผู้สอบ เมื่อการทดสอบยุติ จะได้พารามิเตอร์การแสดงผลข้อสอบของข้อสอบทุกข้อในคลังข้อสอบ



ภาพประกอบ 7 ขั้นตอนวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของ
ค่าอำนาจจำแนกร่วมกับการควบคุมการแสดงข้อสอบ

ขั้นตอนในภาพประกอบ 7 สรุปดังนี้

ขั้นตอนการเลือกข้อสอบของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก (a-Stratified Method : a-STR) เริ่มจากแบ่งคลังข้อสอบเป็น 4 ชั้นตามค่าอำนาจจำแนกข้อสอบ (พารามิเตอร์ a) ชั้นแรกบรรจุข้อสอบอำนาจจำแนกต่ำสุดเท่ากับ 0.5 ชั้นที่ 2 บรรจุข้อสอบอำนาจจำแนกเท่ากับ 1.0 ชั้นที่ 3 บรรจุข้อสอบอำนาจจำแนกเท่ากับ 1.5 ชั้นสุดท้ายบรรจุข้อสอบอำนาจจำแนกต่ำสุดเท่ากับ 2.0 แบ่งชั้นการทดสอบหรือความยาวแบบทดสอบเท่าๆ กันตามการแบ่งชั้นของคลังข้อสอบ

แต่ละค่าความสามารถของผู้สอบ 5,000 ค่าหลังจากผ่านการทดสอบด้วยข้อสอบเทียม 3 ข้อนำไปกำหนดจุดเริ่มต้นการทดสอบ เลือกข้อสอบความยากใกล้เคียงกับค่าประมาณความสามารถของผู้สอบมากที่สุดและจัดให้ผู้สอบ หากผลการตอบข้อสอบและหาความน่าจะเป็นในการตอบข้อสอบถูกต้องนำไปคำนวณหาค่าประมาณความสามารถของผู้สอบก่อนนำไปใช้เลือกข้อสอบข้อต่อไป เมื่อทำการทดสอบจนครบตามความยาวแบบทดสอบย่อยในชั้นนั้น จึงเลื่อนการทดสอบขึ้นไปยังชั้นและชั้นคลังข้อสอบต่อไป ทำการทดสอบจนครบทุกชั้นการทดสอบและทุกชั้นคลังข้อสอบ การทดสอบจึงยุติ

การเลือกข้อสอบความยากใกล้เคียงกับค่าประมาณความสามารถของผู้สอบมากที่สุดและจัดให้ผู้สอบในแต่ละครั้ง พารามิเตอร์ควบคุมการแสดงข้อสอบของข้อสอบข้อนั้นนำไปเปรียบเทียบกับตัวเลขสุ่มที่สร้างจากการแจกแจงแบบเดียวกัน $U(0,1)$ ถ้าพารามิเตอร์ควบคุมการแสดงข้อสอบมีค่าสูงกว่าตัวเลขสุ่มจะจัดข้อสอบนั้นให้ แต่ถ้าพารามิเตอร์ควบคุมการแสดงข้อสอบมีค่าต่ำกว่าตัวเลขสุ่มจะไม่จัดข้อสอบนั้นให้ พารามิเตอร์ควบคุมการแสดงข้อสอบของข้อสอบที่ดีที่สุดข้อถัดไปนำมาเปรียบเทียบกับตัวเลขสุ่มค่าใหม่ก่อนพิจารณาจัดข้อสอบให้ ขั้นตอนนี้ต่อเนื่องจนกระทั่งหาข้อสอบจัดให้ผู้สอบได้ กรณีที่ไม่สามารถหาข้อสอบจัดให้แก่ผู้สอบได้และไม่มีข้อสอบอื่น การทดสอบจะทำซ้ำๆ ต่อไปเรื่อยๆ จนกระทั่งหาข้อสอบข้อใดข้อหนึ่งจัดให้ เมื่อความยาวแบบทดสอบย่อยครบตามกำหนดที่ชั้นการทดสอบนั้น การทดสอบจะต่อเนื่องไปยังชั้นต่อไป จนกระทั่งครบทุกชั้นการทดสอบและครบทุกชั้นคลังข้อสอบ การทดสอบจึงยุติ

7 กำหนดการยุติการทดสอบ

การยุติการทดสอบจะใช้ตามกำหนดความยาวแบบทดสอบ การวิจัยครั้งนี้กำหนดความยาวแบบทดสอบไว้ 3 ระดับ คือ 20 32 และ 44 ข้อ เมื่อการทดสอบดำเนินมาจนครบตามความยาวแบบทดสอบกำหนด การทดสอบจึงยุติ

บทที่ 4

ผลการวิเคราะห์ข้อมูล

งานวิจัยนี้มีจุดมุ่งหมายหลักเพื่อศึกษาประสิทธิภาพการประมาณค่าความสามารถของผู้สอบ และประสิทธิภาพสมมูลการใช้ข้อสอบในคลังข้อสอบของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก เมื่อใช้ จำนวนข้อสอบในแต่ละชั้นคลังข้อสอบ การควบคุมการแสดงผลข้อสอบ ขนาดคลังข้อสอบ และความยาวแบบทดสอบ ต่างกัน โดยใช้วิธีจำลองข้อมูล

ผลการวิจัยตามจุดมุ่งหมายหลักเบื้องต้นให้ผลการวิเคราะห์ข้อมูล จำแนกเป็น 3 ตอนหลัก ดังนี้

ตอนที่ 1 : ประสิทธิภาพการประมาณค่าความสามารถของผู้สอบ ซึ่งผู้วิจัยได้รายงานแยกประเด็นนี้ออกเป็น 5 ประเด็นย่อย ดังนี้

1.1 ประสิทธิภาพการประมาณค่าความสามารถของผู้สอบของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกวิธีที่ 1 (a-STR1)

1.2 ประสิทธิภาพการประมาณค่าความสามารถของผู้สอบของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกวิธีที่ 2 (a-STR2)

1.3 ประสิทธิภาพการประมาณค่าความสามารถของผู้สอบของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกวิธีที่ 3 (a-STR3)

1.4 ประสิทธิภาพการประมาณค่าความสามารถของผู้สอบของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกวิธีที่ 4 (a-STR4)

1.5 เปรียบเทียบประสิทธิภาพการประมาณค่าความสามารถของผู้สอบระหว่างวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก 4 วิธี

ตอนที่ 2 : ประสิทธิภาพสมมูลการใช้ข้อสอบในคลังข้อสอบ ซึ่งผู้วิจัยได้รายงานแยกประเด็นนี้ออกเป็น 5 ประเด็นย่อย ดังนี้

2.1 ประสิทธิภาพสมมูลการใช้ข้อสอบในคลังข้อสอบของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกวิธีที่ 1 (a-STR1)

2.2 ประสิทธิภาพสมมูลการใช้ข้อสอบในคลังข้อสอบของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกวิธีที่ 2 (a-STR2)

2.3 ประสิทธิภาพสมมูลการใช้ข้อสอบในคลังข้อสอบของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกวิธีที่ 3 (a-STR3)

2.4 ประสิทธิภาพสมมูลการใช้ข้อสอบในคลังข้อสอบของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกวิธีที่ 4 (a-STR4)

2.5 เปรียบเทียบประสิทธิภาพสมมูลการใช้ข้อสอบในคลังข้อสอบระหว่างวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก 4 วิธี

ตอนที่ 3 : สรุปวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ที่ไม่สามารถให้ผลการวิเคราะห์ได้

1. ประสิทธิภาพการประมาณค่าความสามารถของผู้สอบ

การวิเคราะห์ตอนนี้ ผู้วิจัยทำการทดสอบประสิทธิภาพการประมาณค่าความสามารถของผู้สอบที่ใช้การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก 4 วิธี ด้วยการวิเคราะห์หาค่าความคลาดเคลื่อนกำลังสองเฉลี่ยและความลำเอียงเฉลี่ย จำแนกตามขนาดคลังข้อสอบ 200 400 600 และ 800 ข้อ ความยาวแบบทดสอบ 20 32 และ 44 ข้อ และการควบคุมและไม่ควบคุมการแสดงผลข้อสอบ รายงานผลดังนี้

1.1 ประสิทธิภาพการประมาณค่าความสามารถของผู้สอบของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกวิธีที่ 1 (a-STR1)

วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกวิธีที่ 1 (a-STR1) ตามเงื่อนไข ขนาดคลังข้อสอบ ความยาวแบบทดสอบ และการควบคุมการแสดงผลข้อสอบ ให้ผลการวิเคราะห์แสดงเป็นความคลาดเคลื่อนกำลังสองเฉลี่ยและความลำเอียงเฉลี่ย ดังนี้

ตาราง 3 ความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) และความลำเอียงเฉลี่ย (Bias) ของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกวิธีที่ 1 (a-STR1)

ผลการวิเคราะห์	ความยาวแบบทดสอบ	ไม่ควบคุมการแสดงผลข้อสอบ			ควบคุมการแสดงผลข้อสอบ		
		20 ข้อ	32 ข้อ	44 ข้อ	20 ข้อ	32 ข้อ	44 ข้อ
ขนาดคลังข้อสอบ							
MSE	200 ข้อ	0.212	0.234	0.200	0.212	ทดสอบไม่ได้	ทดสอบไม่ได้
	400 ข้อ	0.192	0.176	0.168	0.201	0.170	0.158
	600 ข้อ	0.193	0.164	0.152	0.181	0.156	0.155
	800 ข้อ	0.183	0.161	0.144	0.187	0.153	0.134
Bias	200 ข้อ	-0.173	-0.226	-0.184	-0.176	ทดสอบไม่ได้	ทดสอบไม่ได้
	400 ข้อ	-0.146	-0.168	-0.180	-0.161	-0.168	-0.161
	600 ข้อ	-0.171	-0.169	-0.172	-0.165	-0.157	-0.168
	800 ข้อ	-0.165	-0.173	-0.166	-0.177	-0.146	-0.130

ผลการวิเคราะห์ตาราง 3

วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกวิธีที่ 1 (a-STR1) พบว่า เมื่อเพิ่มขนาดคลังข้อสอบจาก 200 400 600 และ 800 ข้อ พบว่า ความคลาดเคลื่อนกำลังสองเฉลี่ยเท่ากับ .212 .192 .193 และ .183 ตามลำดับการเพิ่มขึ้นของคลังข้อสอบที่ความยาวแบบทดสอบ 20 ข้อ ผลการวิเคราะห์ลักษณะคล้ายกันเมื่อใช้ความยาวแบบทดสอบ 32 ข้อและ 44 ข้อ

ขนาดคลังข้อสอบไม่ส่งผลต่อความลำเอียงเฉลี่ย ยกเว้นที่ความยาวแบบทดสอบ 44 ข้อ พบว่าความลำเอียงเฉลี่ยเท่ากับ -.184 -.180 -.172 และ -.161 ตามลำดับการเพิ่มขึ้นของคลังข้อสอบ

ขนาดคลังข้อสอบต่อความคลาดเคลื่อนกำลังสองเฉลี่ยและความลำเอียงเฉลี่ย แสดงผลการวิเคราะห์ลักษณะคล้ายกันทั้งขณะควบคุมและไม่ควบคุมการแสดงข้อสอบ

การใช้ความยาวแบบทดสอบระดับต่างกัน พบว่า เมื่อความยาวแบบทดสอบเพิ่มขึ้น 20 32 และ 44 ข้อ ความคลาดเคลื่อนกำลังสองเฉลี่ยเท่ากับ .192 .176 และ .168 ตามลำดับที่ขนาดคลังข้อสอบ 400 ข้อ และให้ผลลักษณะคล้ายกันเมื่อใช้ขนาดคลังข้อสอบ 600 และ 800 ข้อ ผลการวิเคราะห์

ให้ข้อสังเกตว่า การใช้ขนาดคลังข้อสอบ 200 ข้อ ถึงแม้เพิ่มความยาวแบบทดสอบ แต่ไม่สามารถลดความคลาดเคลื่อนกำลังสองเฉลี่ยตามลำดับการเพิ่มขึ้นของความยาวแบบทดสอบได้

ความยาวแบบทดสอบไม่มีผลต่อความลำเอียงเฉลี่ยทั้งขณะควบคุมและไม่ควบคุมการแสดงผลข้อสอบ ยกเว้นเมื่อวิธีที่ 1 ใช้ร่วมกับการควบคุมการแสดงผลข้อสอบที่ขนาดคลังข้อสอบ 800 ข้อ ให้ความลำเอียงเฉลี่ยเท่ากับ $-.177$ $-.146$ และ $-.130$ ตามลำดับการเพิ่มขึ้นของความยาวแบบทดสอบ

การควบคุมการแสดงผลข้อสอบ พบว่า ให้ความคลาดเคลื่อนกำลังสองเฉลี่ยและความลำเอียงเฉลี่ยน้อยกว่าขณะไม่ได้ควบคุมการแสดงผลข้อสอบ เมื่อใช้กับขนาดคลังข้อสอบเท่ากับหรือมากกว่า 400 ข้อขึ้นไป และความยาวแบบทดสอบเท่ากับหรือมากกว่า 32 ข้อขึ้นไป

การควบคุมการแสดงผลข้อสอบไม่สามารถให้ผลการวิเคราะห์ได้เมื่อใช้กับขนาดคลังข้อสอบ 200 ข้อ และที่ความยาวแบบทดสอบ 32 และ 44 ข้อ

1.2 ประสิทธิภาพการประมาณค่าความสามารถของผู้สอบของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับขั้นของค่าอำนาจจำแนกวิธีที่ 2 (a-STR2)

วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับขั้นของค่าอำนาจจำแนกวิธีที่ 2 (a-STR2) ร่วมกับเงื่อนไข ขนาดคลังข้อสอบ ความยาวแบบทดสอบ และการควบคุมการแสดงผลข้อสอบ ให้ผลการวิเคราะห์แสดงเป็นความคลาดเคลื่อนกำลังสองเฉลี่ยและความลำเอียงเฉลี่ย ดังนี้

ตาราง 4 ความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) และความลำเอียงเฉลี่ย (Bias) ของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกวิธีที่ 2 (a-STR2)

ผลการวิเคราะห์	ความยาวแบบทดสอบ	ไม่ควบคุมการแสดงผลข้อสอบ			ควบคุมการแสดงผลข้อสอบ		
		20 ข้อ	32 ข้อ	44 ข้อ	20 ข้อ	32 ข้อ	44 ข้อ
ขนาดคลังข้อสอบ							
MSE	200 ข้อ	0.211	0.195	0.195	ทดสอบไม่ได้	ทดสอบไม่ได้	ทดสอบไม่ได้
	400 ข้อ	0.201	0.172	0.162	0.193	ทดสอบไม่ได้	ทดสอบไม่ได้
	600 ข้อ	0.186	0.168	0.143	0.188	0.167	ทดสอบไม่ได้
	800 ข้อ	0.189	0.163	0.146	0.192	0.154	0.145
Bias	200 ข้อ	-0.184	-0.191	-0.155	ทดสอบไม่ได้	ทดสอบไม่ได้	ทดสอบไม่ได้
	400 ข้อ	-0.185	-0.175	-0.156	-0.152	ทดสอบไม่ได้	ทดสอบไม่ได้
	600 ข้อ	-0.157	-0.176	-0.163	-0.155	-0.171	ทดสอบไม่ได้
	800 ข้อ	-0.171	-0.164	-0.170	-0.168	-0.150	-0.157

ผลการวิเคราะห์ตามตาราง 4

วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกวิธีที่ 2 (a-STR2) พบว่า เมื่อเพิ่มขึ้นของขนาดคลังข้อสอบ 200 400 600 และ 800 ข้อ ความคลาดเคลื่อนกำลังสองเฉลี่ยเท่ากับ .211 .201 .186 และ .189 ตามลำดับที่ความยาวแบบทดสอบ 20 ข้อ ผลการวิเคราะห์ลักษณะคล้ายกันเมื่อใช้ความยาวแบบทดสอบ 32 ข้อและ 44 ข้อ

ขนาดคลังข้อสอบไม่มีผลต่อความลำเอียงเฉลี่ย สืบเนื่องจากความลำเอียงเฉลี่ยเท่ากับ -.184 -.185 -.157 และ -.171 ตามลำดับที่ขนาดความยาวแบบทดสอบ 20 ข้อ และความลำเอียงเฉลี่ยเท่ากับ -.155 -.156 -.163 และ -.170 ตามลำดับที่ขนาดความยาวแบบทดสอบ 44 ข้อ

วิธีที่ 2 ใช้ร่วมกับการควบคุมการแสดงผลข้อสอบ พบว่า ขนาดคลังข้อสอบไม่มีผลต่อคลาดเคลื่อนกำลังสองเฉลี่ยและความลำเอียงเฉลี่ยที่ความยาวแบบทดสอบ 20 ข้อ แต่เมื่อใช้ขนาดคลังข้อสอบ 600 และ 800 ข้อ พบว่าความลำเอียงเฉลี่ยเท่ากับ -.171 และ -.150 ตามลำดับที่ความยาวแบบทดสอบ 32 ข้อ

การใช้ความยาวแบบทดสอบ 20 32 และ 44 ข้อ พบว่า ความคลาดเคลื่อนกำลังสองเฉลี่ยเท่ากับ .201 .172 และ .162 ตามลำดับที่ขนาดคลังข้อสอบ 400 ข้อ ผลการวิเคราะห์มีลักษณะคล้ายกันเมื่อใช้ขนาดคลังข้อสอบ 600 และ 800 ข้อ จากการสังเกตพบว่า การใช้ความยาวแบบทดสอบเพิ่มขึ้นที่ขนาดคลังข้อสอบ 200 ข้อ ลดความคลาดเคลื่อนกำลังสองเฉลี่ยได้ไม่ชัดเจน

ความยาวแบบทดสอบไม่มีผลต่อความลำเอียงเฉลี่ยเมื่อพิจารณาตามขนาดคลังข้อสอบ ถึงแม้ว่าที่ขนาดคลังข้อสอบ 400 ข้อ การเพิ่มขึ้นของความยาวแบบทดสอบจะให้ความลำเอียงเฉลี่ยเท่ากับ -.185 -.175 และ -.156 ตามลำดับ

การเพิ่มความยาวแบบทดสอบเมื่อวิธีที่ 2 ใช้ร่วมกับการควบคุมการแสดงผลข้อสอบ พบว่า ให้ความคลาดเคลื่อนกำลังสองเฉลี่ยเท่ากับ .192 .154 และ .145 ตามลำดับ และความลำเอียงเฉลี่ยเท่ากับ -.168 -.150 และ -.157 ตามลำดับที่ขนาดคลังข้อสอบ 800 ข้อ

การควบคุมการแสดงผลข้อสอบ พบว่า ให้ความคลาดเคลื่อนกำลังสองเฉลี่ยและความลำเอียงเฉลี่ยใกล้เคียงกับขณะไม่ได้ควบคุมการแสดงผลข้อสอบ และคลาดเคลื่อนกำลังสองเฉลี่ยและความลำเอียงเฉลี่ยมีแนวโน้มลดลง เมื่อใช้ขนาดคลังข้อสอบเท่ากับหรือมากกว่า 600 ข้อขึ้นไป และความยาวแบบทดสอบเท่ากับหรือมากกว่า 32 ข้อขึ้นไป

การควบคุมการแสดงผลข้อสอบไม่สามารถให้ผลการวิเคราะห์เมื่อใช้กับขนาดคลังข้อสอบ 200 ข้อ ที่ความยาวแบบทดสอบ 20 32 และ 44 ข้อ ขนาดคลังข้อสอบ 400 ข้อ ที่ความยาวแบบทดสอบ 32 และ 44 ข้อ และขนาดคลังข้อสอบ 600 ข้อ ที่ความยาวแบบทดสอบ 44 ข้อ

1.3 ประสิทธิภาพการประมาณค่าความสามารถของผู้สอบของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับขั้นของค่าอำนาจจำแนกวิธีที่ 3 (a-STR3)

วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับขั้นของค่าอำนาจจำแนกวิธีที่ 3 (a-STR3) ร่วมกับเงื่อนไข ขนาดคลังข้อสอบ ความยาวแบบทดสอบ และการควบคุมการแสดงผลข้อสอบ ให้ผลการวิเคราะห์เป็นความคลาดเคลื่อนกำลังสองเฉลี่ยและความลำเอียงเฉลี่ย ดังนี้

ตาราง 5 ความคลาดเคลื่อนกำลังสองเฉลี่ย(MSE) และความลำเอียงเฉลี่ย(Bias) ของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกวิธีที่ 3 (a-STR3)

ผลการวิเคราะห์	ความยาวแบบทดสอบ	ไม่ควบคุมการแสดงผลข้อสอบ			ควบคุมการแสดงผลข้อสอบ		
		20 ข้อ	32 ข้อ	44 ข้อ	20 ข้อ	32 ข้อ	44 ข้อ
ขนาดคลังข้อสอบ							
MSE	200 ข้อ	0.206	0.173	0.173	ทดสอบไม่ได้	ทดสอบไม่ได้	ทดสอบไม่ได้
	400 ข้อ	0.190	0.166	0.140	0.186	0.164	ทดสอบไม่ได้
	600 ข้อ	0.185	0.154	0.142	0.195	0.158	ทดสอบไม่ได้
	800 ข้อ	0.182	0.155	0.134	0.185	0.147	0.133
Bias	200 ข้อ	-0.197	-0.156	-0.190	ทดสอบไม่ได้	ทดสอบไม่ได้	ทดสอบไม่ได้
	400 ข้อ	-0.170	-0.177	-0.163	-0.158	-0.161	ทดสอบไม่ได้
	600 ข้อ	-0.166	-0.179	-0.161	-0.178	-0.169	ทดสอบไม่ได้
	800 ข้อ	-0.168	-0.168	-0.169	-0.160	-0.159	-0.147

ผลการวิเคราะห์ตามทดสอบตาราง 5

วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกวิธีที่ 3 (a-STR3) เมื่อขนาดคลังข้อสอบเพิ่มขึ้น 200 400 600 และ 800 ข้อ พบว่า ความคลาดเคลื่อนกำลังสองเฉลี่ยเท่ากับ .206 .190 .185 และ .182 ตามลำดับที่ความยาวแบบทดสอบ 20 ข้อ ผลการวิเคราะห์ลักษณะคล้ายกันเมื่อใช้ความยาวแบบทดสอบ 32 ข้อและ 44 ข้อ

การเพิ่มขนาดคลังข้อสอบไม่มีผลต่อความลำเอียงเฉลี่ย สืบเนื่องจากขนาดคลังข้อสอบเพิ่มขึ้น 200 400 600 และ 800 ข้อ ให้ความลำเอียงเฉลี่ยเท่ากับ -.197 -.170 -.166 และ -.168 ตามลำดับที่ความยาวแบบทดสอบ 20 ข้อ ความลำเอียงเฉลี่ยเท่ากับ -.156 -.177 -.179 และ -.168 ตามลำดับที่ขนาดความยาวแบบทดสอบ 32 ข้อ และความลำเอียงเฉลี่ยเท่ากับ -.190 -.163 -.161 และ -.169 ตามลำดับที่ความยาวแบบทดสอบ 44 ข้อ

การเพิ่มขนาดคลังข้อสอบเมื่อวิธีนี้ใช้ร่วมกับการควบคุมการแสดงผลข้อสอบ พบว่า ขนาดคลังข้อสอบ 400 600 และ 800 ข้อ ที่ความยาวแบบทดสอบ 32 ข้อ ให้ความคลาดเคลื่อนกำลังสองเฉลี่ยเท่ากับ

.164 .158 และ .147 ตามลำดับ แต่ไม่สามารถลดความลำเอียงเฉลี่ยได้ ขณะที่ความยาวแบบทดสอบ 20 ข้อ ไม่ลดความคลาดเคลื่อนกำลังสองเฉลี่ยและความลำเอียงเฉลี่ยตามขนาดคลังข้อสอบที่เพิ่มขึ้น

การเพิ่มความยาวแบบทดสอบ 20 32 และ 44 ข้อ พบว่า ความคลาดเคลื่อนกำลังสองเฉลี่ยเท่ากับ .190 .166 และ .140 ตามลำดับที่ขนาดคลังข้อสอบ 400 ข้อ ผลการวิเคราะห์มีลักษณะคล้ายกันเมื่อใช้ขนาดคลังข้อสอบ 600 และ 800 ข้อ จากการสังเกตพบว่า การใช้ความยาวแบบทดสอบ เพิ่มขึ้นที่ขนาดคลังข้อสอบ 200 ข้อ ลดความคลาดเคลื่อนกำลังสองเฉลี่ยได้ไม่ชัดเจน

ความยาวแบบทดสอบไม่มีผลต่อความลำเอียงเฉลี่ย เมื่อใช้ความยาวแบบทดสอบ 20 32 และ 44 ข้อ ที่ขนาดคลังข้อสอบระดับต่างๆ

การเพิ่มความยาวแบบทดสอบเมื่อวิธีนี้ใช้ร่วมกับการควบคุมการแสดงผลข้อสอบ พบว่า ความคลาดเคลื่อนกำลังสองเฉลี่ยเท่ากับ .185 .147 และ .133 ตามลำดับและความลำเอียงเฉลี่ยเท่ากับ -.160 -.159 และ -.147 ตามลำดับที่ขนาดคลังข้อสอบ 800 ข้อ ผลการวิเคราะห์มีลักษณะคล้ายกันเมื่อใช้ขนาดคลังข้อสอบ 600 ข้อ แต่ที่ขนาดคลังข้อสอบ 400 ข้อ พบว่า ความยาวแบบทดสอบ 20 และ 32 ข้อ ให้ความคลาดเคลื่อนกำลังสองเฉลี่ยเท่ากับ .186 และ .164 ตามลำดับ แต่ไม่ลดความลำเอียงเฉลี่ย

การควบคุมการแสดงผลข้อสอบ พบว่า ให้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยและความลำเอียงเฉลี่ยใกล้เคียงกับขณะไม่ได้ควบคุมการแสดงผลข้อสอบ

การควบคุมการแสดงผลข้อสอบให้คลาดเคลื่อนกำลังสองเฉลี่ยและความลำเอียงเฉลี่ยมีแนวโน้มลดลง เมื่อใช้ขนาดคลังข้อสอบเท่ากับหรือมากกว่า 400 ข้อขึ้นไป และความยาวแบบทดสอบเท่ากับหรือมากกว่า 32 ข้อขึ้นไป

การควบคุมการแสดงผลข้อสอบไม่สามารถให้ผลการวิเคราะห์ได้เมื่อใช้กับขนาดคลังข้อสอบ 200 ข้อ ที่ระดับความยาวแบบทดสอบ 20 32 และ 44 ข้อ ขนาดคลังข้อสอบ 400 ข้อ ที่ความยาวแบบทดสอบ 44 ข้อ และขนาดคลังข้อสอบ 600 ข้อ ที่ความยาวแบบทดสอบ 44 ข้อ

1.4 ประสิทธิภาพการประมาณค่าความสามารถของผู้สอบของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับขั้นของค่าอำนาจจำแนกวิธีที่ 4 (a-STR4)

วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับขั้นของค่าอำนาจจำแนกวิธีที่ 4 (a-STR4) ร่วมกับเงื่อนไข ขนาดคลังข้อสอบ ความยาวแบบทดสอบ และการควบคุมการแสดงผลข้อสอบ ให้ผลการวิเคราะห์เป็นความคลาดเคลื่อนกำลังสองเฉลี่ยและความลำเอียงเฉลี่ย ดังนี้

ตาราง 6 ความคลาดเคลื่อนกำลังสองเฉลี่ย(MSE) และความลำเอียงเฉลี่ย (Bias) ของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกวิธีที่ 4 (a-STR4)

ผลการวิเคราะห์	ความยาวแบบทดสอบ	ไม่ควบคุมการแสดงผลข้อสอบ			ควบคุมการแสดงผลข้อสอบ		
		20 ข้อ	32 ข้อ	44 ข้อ	20 ข้อ	32 ข้อ	44 ข้อ
ขนาดคลังข้อสอบ							
MSE	200 ข้อ	0.210	0.197	0.172	ทดสอบไม่ได้	ทดสอบไม่ได้	ทดสอบไม่ได้
	400 ข้อ	0.195	0.167	0.143	0.197	ทดสอบไม่ได้	ทดสอบไม่ได้
	600 ข้อ	0.187	0.159	0.139	0.192	0.159	ทดสอบไม่ได้
	800 ข้อ	0.186	0.157	0.132	0.190	0.156	0.137
Bias	200 ข้อ	-0.185	-0.204	-0.193	ทดสอบไม่ได้	ทดสอบไม่ได้	ทดสอบไม่ได้
	400 ข้อ	-0.175	-0.170	-0.169	-0.165	ทดสอบไม่ได้	ทดสอบไม่ได้
	600 ข้อ	-0.151	-0.161	-0.164	-0.173	-0.162	ทดสอบไม่ได้
	800 ข้อ	-0.171	-0.167	-0.168	-0.163	-0.167	-0.159

ผลการวิเคราะห์ตามตาราง 6

วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกวิธีที่ 4 (a-STR4) เมื่อเพิ่มขนาดคลังข้อสอบ 200 400 600 และ 800 ข้อ พบว่า ความคลาดเคลื่อนกำลังสองเฉลี่ยเท่ากับ .210 .195 .187 และ .186 ตามลำดับที่ความยาวแบบทดสอบ 20 ข้อ ผลการวิเคราะห์ลักษณะคล้ายกันเมื่อใช้ความยาวแบบทดสอบ 32 ข้อและ 44 ข้อ

การเพิ่มขนาดคลังข้อสอบให้ความลำเอียงเฉลี่ยเท่ากับ -.204 -.170 -.161 และ -.167 ตามลำดับที่ความยาวแบบทดสอบ 32 ข้อ และเท่ากับ -.193 -.169 -.164 และ -.168 ตามลำดับที่ความยาวแบบทดสอบ 44 ข้อ

การเพิ่มขนาดคลังข้อสอบเมื่อวิธีนี้ใช้ร่วมกับการควบคุมการแสดงผลข้อสอบ พบว่าคลาดเคลื่อนกำลังสองเฉลี่ยเท่ากับ .197 .192 และ .190 ตามลำดับที่ความยาวแบบทดสอบ 20 ข้อ แต่ไม่มีผลต่อความลำเอียงเฉลี่ย ผลการวิเคราะห์มีลักษณะคล้ายกันเมื่อใช้ความยาวแบบทดสอบ 32 ข้อ

ความยาวแบบทดสอบเพิ่มขึ้น พบว่า ความคลาดเคลื่อนกำลังสองเฉลี่ยเท่ากับ .210 .197 และ .172 ตามลำดับที่ขนาดคลังข้อสอบ 200 ข้อ ผลการวิเคราะห์มีลักษณะคล้ายกันเมื่อใช้ขนาดคลังข้อสอบ 400 600 และ 800 ข้อ

ความยาวแบบทดสอบไม่มีผลต่อความลำเอียงเฉลี่ย ผลการวิเคราะห์พบว่า ความยาวแบบทดสอบ 20 32 และ 44 ข้อ ไม่สามารถลดความลำเอียงเฉลี่ยได้

การเพิ่มความยาวแบบทดสอบเมื่อวิธีที่ 4 ใช้ร่วมกับการควบคุมการแสดงผลข้อสอบ พบว่า ความคลาดเคลื่อนกำลังสองเฉลี่ยเท่ากับ .190 .156 และ .137 ตามลำดับ และความลำเอียงเฉลี่ยเท่ากับ -.163 -.167 และ -.159 ตามลำดับที่ขนาดคลังข้อสอบ 800 ข้อ ผลการวิเคราะห์มีลักษณะคล้ายกันเมื่อใช้ขนาดคลังข้อสอบ 600 ข้อ

การควบคุมการแสดงผลข้อสอบ พบว่า ให้ความคลาดเคลื่อนกำลังสองเฉลี่ยและความลำเอียงเฉลี่ยใกล้เคียงกับขณะไม่ได้ควบคุมการแสดงผลข้อสอบ เมื่อขนาดคลังข้อสอบเท่ากับหรือมากกว่า 600 ข้อขึ้นไป และความยาวแบบทดสอบเท่ากับหรือมากกว่า 32 ข้อขึ้นไป

การควบคุมการแสดงผลข้อสอบไม่สามารถให้ผลการวิเคราะห์ได้เมื่อใช้กับขนาดคลังข้อสอบ 200 ข้อที่ระดับความยาวแบบทดสอบ 20 32 และ 44 ข้อ ขนาดคลังข้อสอบ 400 ข้อที่ความยาวแบบทดสอบ 32 และ 44 ข้อ และขนาดคลังข้อสอบ 600 ข้อที่ความยาวแบบทดสอบ 44 ข้อ

1.5 เปรียบเทียบประสิทธิภาพการประมาณค่าความสามารถของผู้สอบระหว่าง วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับขั้นของค่าอำนาจจำแนก 4 วิธี

เปรียบเทียบประสิทธิภาพการประมาณค่าความสามารถของผู้สอบระหว่างวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับขั้นของค่าอำนาจจำแนก 4 วิธี โดยตรวจสอบวิธีที่ให้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยน้อยที่สุด และความลำเอียงเฉลี่ยน้อยที่สุด เมื่อการทดสอบครบตามความยาวแบบทดสอบกำหนด แสดงผลการเปรียบเทียบ ดังนี้

ตาราง 7 เปรียบเทียบความคลาดเคลื่อนกำลังสองเฉลี่ยและความลำเอียงเฉลี่ยของวิธีการทดสอบแบบ
ปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก 4 วิธี

คลังข้อสอบ (ข้อ)	ความยาว แบบทดสอบ (ข้อ)	ความคลาดเคลื่อนกำลังสองเฉลี่ย				ความลำเอียงเฉลี่ย			
		a-STR1	a-STR2	a-STR3	a-STR4	a-STR1	a-STR2	a-STR3	a-STR4
200	20			น้อยที่สุด		น้อยที่สุด			
400				น้อยที่สุด		น้อยที่สุด			
600				น้อยที่สุด					น้อยที่สุด
800				น้อยที่สุด		น้อยที่สุด			
200	32			น้อยที่สุด				น้อยที่สุด	
400				น้อยที่สุด		น้อยที่สุด			
600				น้อยที่สุด	น้อยที่สุด				น้อยที่สุด
800				น้อยที่สุด			น้อยที่สุด		
200	44				น้อยที่สุด	น้อยที่สุด			
400				น้อยที่สุด		น้อยที่สุด			
600					น้อยที่สุด			น้อยที่สุด	
800					น้อยที่สุด	น้อยที่สุด			

ผลการเปรียบเทียบตามตาราง 7

ผลการเปรียบเทียบประสิทธิภาพการประมาณค่าความสามารถของผู้สอบระหว่างวิธีการ
ทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก 4 วิธี วิธีที่ให้ความ
คลาดเคลื่อนกำลังสองเฉลี่ยน้อยที่สุด ในภาพรวมพบว่า

**a-STR3 หรือ วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจ
จำแนกวิธีที่ 1 มีแนวโน้มให้ความคลาดเคลื่อนกำลังสองเฉลี่ยน้อยที่สุด** ที่ขนาดคลังข้อสอบ 200 400
600 และ 800 ข้อ เมื่อใช้กับความยาวแบบทดสอบเท่ากับ 20 และ 32 ข้อ

a-STR4 หรือ วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกวิธีที่ 4 มีแนวโน้มให้ความคลาดเคลื่อนกำลังสองเฉลี่ยน้อยที่สุดรองลงมา ที่ขนาดคลังข้อสอบ 200 600 และ 800 ข้อ เมื่อใช้กับความยาวแบบทดสอบ 44 ข้อ

ผลการเปรียบเทียบประสิทธิภาพการประมาณค่าความสามารถของผู้สอบระหว่างวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก 4 วิธี วิธีที่ให้ความลำเอียงเฉลี่ยน้อยที่สุด ในภาพรวมพบว่า

วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกทั้ง 4 วิธี สามารถให้ความลำเอียงเฉลี่ยน้อยที่สุดเท่าเทียมกัน เนื่องจากทั้ง 4 วิธีสามารถให้ความลำเอียงเฉลี่ยน้อยที่สุดกระจายไปตามวิธีกำหนดต่างๆ

ผลการเปรียบเทียบเบื้องต้นนำไปสู่การค้นหาเหตุผลว่า วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกวิธีที่ 3 (a-STR3) และวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกวิธีที่ 4 (a-STR4) ให้ประสิทธิภาพการประมาณค่าความสามารถของผู้สอบดีที่สุดได้อย่างไร ผู้วิจัยได้ทำการศึกษาผลการวิเคราะห์ข้อมูลที่แต่ละระดับชั้นค่าอำนาจจำแนก เพื่อติดตามการเปลี่ยนแปลงความคลาดเคลื่อนกำลังสองเฉลี่ยและความลำเอียงเฉลี่ยของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก 4 วิธี ผลการวิเคราะห์อธิบายได้ดังนี้

วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกวิธีที่ 3 (a-STR3) มีอัตราส่วนจำนวนข้อสอบแต่ละชั้นคลังข้อสอบเท่ากับ 1: 0.5 : 0.5 : 2 ซึ่งชั้นสุดท้ายของคลังข้อสอบบรรจุข้อสอบอำนาจจำแนกสูงจำนวนมากกว่าชั้นอื่น ระยะนี้ค่าประมาณความสามารถเข้าใกล้ค่าความสามารถจริงของผู้สอบมากแล้ว การทดสอบด้วยข้อสอบอำนาจจำแนกสูงและเลือกข้อสอบความยากใกล้เคียงค่าประมาณความสามารถของผู้สอบมากที่สุดจากข้อสอบอำนาจจำแนกสูงที่มีให้เลือกจำนวนมาก สามารถลดความแตกต่างระหว่างค่าประมาณความสามารถและค่าความสามารถจริงของผู้สอบได้ดี ให้ความคลาดเคลื่อนกำลังสองเฉลี่ยและความลำเอียงเฉลี่ยลดลงอย่างรวดเร็ว

ดังนั้นมีความเป็นไปได้ว่า วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกวิธีที่ 3 (a-STR3) สามารถให้ความคลาดเคลื่อนกำลังสองเฉลี่ยและความลำเอียงเฉลี่ยลดต่ำกว่าความคลาดเคลื่อนกำลังสองเฉลี่ยและความลำเอียงเฉลี่ยของวิธีอื่นที่ชั้นอำนาจจำแนก 2.0

ตาราง 8 เปรียบเทียบความคลาดเคลื่อนกำลังสองเฉลี่ยและความลำเอียงเฉลี่ยของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก 4 วิธี ขนาดคลังข้อสอบ 200 ข้อ ความยาวแบบทดสอบ 32 ข้อ

ชั้นคลังข้อสอบตาม ค่าอำนาจจำแนก	ความคลาดเคลื่อนกำลังสองเฉลี่ย				ความลำเอียงเฉลี่ย			
	a-STR1	a-STR2	a-STR3	a-STR4	a-STR1	a-STR2	a-STR3	a-STR4
ชั้น 0.5	1.428	1.328	1.374	1.520	-.742	-.707	-.708	-.780
ชั้น 1.0	.562	.591	.812	.708	-.400	-.383	-.580	-.470
ชั้น 1.5	.295	.397	.352	.361	-.254	-.324	-.278	-.291
ชั้น 2.0	.234	.195	.173	.197	-.226	-.191	-.156	-.204

ผลการวิเคราะห์ตามตาราง 8

วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก 4 วิธีที่ชั้นอำนาจจำแนก 2.0 พบว่าวิธี a-STR3 ให้ความคลาดเคลื่อนกำลังสองเฉลี่ยเท่ากับ .173 ขณะที่วิธี a-STR1 a-STR2 และ a-STR4 ให้ความคลาดเคลื่อนกำลังสองเฉลี่ยเท่ากับ .234 .195 และ .197 ตามลำดับ

วิธี a-STR3 ให้ความลำเอียงเฉลี่ยเท่ากับ -.156 ขณะที่วิธี a-STR1 a-STR2 และ a-STR4 ให้ความลำเอียงเฉลี่ยเท่ากับ -.226 -.191 -.156 และ -.204 ตามลำดับ

การจัดจำนวนข้อสอบแต่ละชั้นคลังข้อสอบดังกล่าว นำไปอธิบายได้กับวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกวิธีที่ 4 (a-STR4) ซึ่งมีอัตราส่วนจำนวนข้อสอบแต่ละชั้นคลังข้อสอบเท่ากับ 1.5 : 0.5 : 0.5 : 1.5 ชั้นสุดท้ายของคลังข้อสอบบรรจุข้อสอบอำนาจจำแนกสูงจำนวนมากรองลงมาจากวิธี a-STR3

ดังนั้นมีความเป็นไปได้ว่าวิธี a-STR4 สามารถให้ความคลาดเคลื่อนกำลังสองเฉลี่ยและความลำเอียงเฉลี่ยลดลงรองลงมาจากวิธี a-STR3 เมื่อใช้กับขนาดคลังข้อสอบและความยาวแบบทดสอบมากเพียงพอ

ตาราง 9 เปรียบเทียบความคลาดเคลื่อนกำลังสองเฉลี่ยและความลำเอียงเฉลี่ยของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก 4 วิธี ที่ขนาดคลังข้อสอบ 600 ข้อ ความยาวแบบทดสอบ 32 ข้อ

ระดับคลังข้อสอบตาม ค่าอำนาจจำแนก	ความคลาดเคลื่อนกำลังสองเฉลี่ย				ความลำเอียงเฉลี่ย			
	a-STR1	a-STR2	a-STR3	a-STR4	a-STR1	a-STR2	a-STR3	a-STR4
ชั้น 0.5	1.382	1.351	1.474	1.471	-.716	-.710	-.752	-.752
ชั้น 1.0	.511	.572	.540	.529	-.378	-.434	-.401	-.393
ชั้น 1.5	.264	.282	.273	.281	-.248	-.250	-.241	-.256
ชั้น 2.0	.164	.168	.154	.159	-.169	-.176	-.179	-.161

ผลการวิเคราะห์ตามตาราง 9

วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก 4 วิธีที่ชั้นอำนาจจำแนก 2.0 เมื่อใช้ขนาดคลังข้อสอบ 600 ข้อ ความยาวแบบทดสอบ 32 ข้อ พบว่าวิธี a-STR4 ให้ความคลาดเคลื่อนกำลังสองเฉลี่ยเท่ากับ .159 รองลงมาจากวิธี a-STR3 ซึ่งให้ความคลาดเคลื่อนกำลังสองเฉลี่ยเท่ากับ .154 ขณะที่วิธี a-STR1 และ a-STR2 ให้ความคลาดเคลื่อนกำลังสองเฉลี่ยเท่ากับ .164 และ .168 ตามลำดับ

วิธี a-STR4 ให้ความลำเอียงเฉลี่ยเท่ากับ -.161 ขณะที่วิธี a-STR1 a-STR2 และ a-STR3 ให้ความลำเอียงเฉลี่ยเท่ากับ -.169 -.176 และ -.179 ตามลำดับ

การจัดข้อสอบอำนาจจำแนกต่ำจำนวนมากที่ชั้นค่าอำนาจจำแนก 0.5 ของวิธี a-STR4 ไม่สามารถลดความคลาดเคลื่อนกำลังสองเฉลี่ยและความลำเอียงเฉลี่ยเมื่อเปรียบเทียบกับวิธี a-STR1 และ a-STR3 ที่มีข้อสอบจำนวนน้อยกว่าที่ชั้นค่าอำนาจจำแนก 0.5 ผลการวิเคราะห์พบว่าวิธี a-STR4 ให้ความคลาดเคลื่อนกำลังสองเฉลี่ยเท่ากับ 1.471 ขณะที่วิธี a-STR1 และ a-STR3 ให้ความคลาดเคลื่อนกำลังสองเฉลี่ยเท่ากับ 1.382 และ 1.474 ตามลำดับ

วิธี a-STR4 ให้ความลำเอียงเฉลี่ยเท่ากับ -.752 ขณะที่วิธี a-STR1 และ a-STR3 ให้ความลำเอียงเฉลี่ยเท่ากับ -.716 และ -.752 ตามลำดับ

การจัดวางข้อสอบจำนวนมากที่ขึ้นค่าอำนาจจำแนก 0.5 ของวิธี a-STR2 ถึงแม้ว่ามีแนวโน้มให้ความคลาดเคลื่อนกำลังสองเฉลี่ยและความลำเอียงเฉลี่ยน้อยกว่าวิธีอื่นระยะเริ่มต้นการทดสอบ แต่ไม่สามารถลดความคลาดเคลื่อนกำลังสองเฉลี่ยและความลำเอียงเฉลี่ยได้ดีเท่าวิธี a-STR3 เมื่อสิ้นสุดการทดสอบ

วิธี a-STR2 a-STR3 และ a-STR4 บรรจุข้อสอบจำนวนน้อยในชั้นค่าอำนาจจำแนก 1.0 และ 1.5 การเลือกข้อสอบค่าความยากใกล้เคียงค่าความสามารถของผู้สอบเป็นไปได้ยาก ให้ความคลาดเคลื่อนกำลังสองเฉลี่ยและความลำเอียงเฉลี่ยลดลงช้า ผลการวิเคราะห์ตามตาราง 8 และตาราง 9 พบว่าวิธี a-STR2 a-STR3 และ a-STR4 ให้ความคลาดเคลื่อนกำลังสองเฉลี่ยและความลำเอียงเฉลี่ยลดลงที่ชั้นค่าอำนาจจำแนก 1.0 และ 1.5 ช้ากว่าวิธี a-STR1 ซึ่งมีจำนวนข้อสอบที่ชั้นค่าอำนาจจำแนกเท่ากันทุกชั้น

เปรียบเทียบประสิทธิภาพการประมาณค่าความสามารถของผู้สอบระหว่างวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก 4 วิธีเมื่อใช้ร่วมกับการควบคุมการแสดงข้อสอบ แสดงผลการเปรียบเทียบดังนี้

ตาราง 10 เปรียบเทียบความคลาดเคลื่อนกำลังสองเฉลี่ยและความลำเอียงเฉลี่ยของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกร่วมกับการควบคุมการแสดงผลข้อสอบ 4 วิธี

คลังข้อสอบ(ข้อ)	ความยาวแบบทดสอบ(ข้อ)	ความคลาดเคลื่อนกำลังสองเฉลี่ย				ความลำเอียงเฉลี่ย			
		a-STR1	a-STR2	a-STR3	a-STR4	a-STR1	a-STR2	a-STR3	a-STR4
200	20	น้อยที่สุด	ทดสอบไม่ได้	ทดสอบไม่ได้	ทดสอบไม่ได้	น้อยที่สุด	ทดสอบไม่ได้	ทดสอบไม่ได้	ทดสอบไม่ได้
400				น้อยที่สุด		น้อยที่สุด			
600		น้อยที่สุด				น้อยที่สุด			
800				น้อยที่สุด				น้อยที่สุด	
200	32	ทดสอบไม่ได้	ทดสอบไม่ได้	ทดสอบไม่ได้	ทดสอบไม่ได้	ทดสอบไม่ได้	ทดสอบไม่ได้	ทดสอบไม่ได้	ทดสอบไม่ได้
400			ทดสอบไม่ได้	น้อยที่สุด	ทดสอบไม่ได้		ทดสอบไม่ได้	น้อยที่สุด	ทดสอบไม่ได้
600		น้อยที่สุด				น้อยที่สุด			
800				น้อยที่สุด		น้อยที่สุด			
200	44	ทดสอบไม่ได้	ทดสอบไม่ได้	ทดสอบไม่ได้	ทดสอบไม่ได้	ทดสอบไม่ได้	ทดสอบไม่ได้	ทดสอบไม่ได้	ทดสอบไม่ได้
400		น้อยที่สุด	ทดสอบไม่ได้	ทดสอบไม่ได้	ทดสอบไม่ได้	น้อยที่สุด	ทดสอบไม่ได้	ทดสอบไม่ได้	ทดสอบไม่ได้
600		น้อยที่สุด	ทดสอบไม่ได้	ทดสอบไม่ได้	ทดสอบไม่ได้	น้อยที่สุด	ทดสอบไม่ได้	ทดสอบไม่ได้	ทดสอบไม่ได้
800				น้อยที่สุด		น้อยที่สุด			

ผลเปรียบเทียบตามตาราง 10

ผลการเปรียบเทียบประสิทธิภาพการประมาณค่าความสามารถของผู้สอบระหว่างวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก 4 วิธีรวมกับการควบคุมการแสดงผลข้อสอบ วิธีที่ให้ความคลาดเคลื่อนกำลังสองเฉลี่ยน้อยที่สุด ในภาพรวมพบว่า

a-STR1 หรือ วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกวิธีที่ 1 และ a-STR3 หรือ วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นค่าอำนาจจำแนกวิธีที่ 3 เมื่อใช้รวมกับการควบคุมการแสดงผลข้อสอบ มีแนวโน้มให้ความคลาดเคลื่อนกำลังสองเฉลี่ยน้อยที่สุด

การเปรียบเทียบแสดงข้อสังเกตว่า a-STR1 หรือ วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกวิธีที่ 1 รวมกับการควบคุมการแสดงผลข้อสอบสามารถให้ผลการวิเคราะห์ที่ขนาดคลังข้อสอบ 200 ข้อขึ้นไป แต่ a-STR3 หรือ วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกวิธีที่ 3 รวมกับการควบคุมการแสดงผลข้อสอบสามารถให้ผลการวิเคราะห์ที่ขนาดคลังข้อสอบตั้งแต่ 400 ข้อขึ้นไป

การเปรียบเทียบระหว่างวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก 4 วิธีรวมกับการควบคุมการแสดงผลข้อสอบ วิธีที่ให้ความลำเอียงเฉลี่ยน้อยที่สุดในภาพรวมพบว่า

a-STR1 หรือ วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นค่าอำนาจจำแนกวิธีที่ 1 รวมกับการควบคุมการแสดงผลข้อสอบมีแนวโน้มให้ความลำเอียงเฉลี่ยน้อยที่สุด

อย่างไรก็ตาม ผลการวิเคราะห์แสดงความลำเอียงเฉลี่ยน้อยที่สุดกระจายไปตามวิธีต่างๆ และวิธีกำหนดหลายวิธีไม่สามารถให้ผลการวิเคราะห์ได้ ดังนั้น การนำผลความลำเอียงเฉลี่ยของวิธี a-STR1 ไปใช้ ควรพิจารณาเงื่อนไขอื่นๆ ของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ประกอบการตัดสินใจ

ผลการเปรียบเทียบเบื้องต้นนำไปสู่การค้นหาเหตุผลว่า วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกวิธีที่ 1 (a-STR1) รวมกับการควบคุมการแสดงผลข้อสอบ และวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกวิธีที่ 3 (a-STR3) รวมกับการควบคุมการแสดงผลข้อสอบ ให้ประสิทธิภาพการประมาณค่าความสามารถของผู้สอบดีที่สุดได้อย่างไร ผู้วิจัยได้ทำการศึกษาผลการวิเคราะห์ข้อมูลที่แต่ละระดับชั้นค่าอำนาจจำแนก เพื่อติดตามการเปลี่ยนแปลงความคลาดเคลื่อนกำลังสองเฉลี่ยและความลำเอียงเฉลี่ยของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก 4 วิธีรวมกับการควบคุมการแสดงผลข้อสอบ ผลการวิเคราะห์อธิบายได้ดังนี้

ตาราง 11 เปรียบเทียบความคลาดเคลื่อนกำลังสองเฉลี่ยของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก 4 วิธี ร่วมกับการควบคุมและไม่ควบคุมการแสดงผลข้อสอบ ขนาดคลังข้อสอบ 800 ข้อ ความยาวแบบทดสอบ 32 ข้อ

ชั้นคลังข้อสอบตาม ค่าอำนาจจำแนก	ไม่ควบคุมการแสดงผลข้อสอบ				ควบคุมการแสดงผลข้อสอบ			
	a-STR1	a-STR2	a-STR3	a-STR4	a-STR1	a-STR2	a-STR3	a-STR4
ชั้น 0.5	1.342	1.320	1.514	1.373	1.253	1.254	1.225	1.260
ชั้น 1.0	.505	.531	.580	.508	.466	.489	.497	.453
ชั้น 1.5	.251	.282	.276	.280	.250	.267	.278	.283
ชั้น 2.0	.161	.163	.155	.157	.153	.154	.147	.156

ตาราง 12 เปรียบเทียบความลำเอียงเฉลี่ยของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก 4 วิธี ร่วมกับการควบคุมและไม่ควบคุมการแสดงผลข้อสอบ ขนาดคลังข้อสอบ 800 ข้อ ความยาวแบบทดสอบ 32 ข้อ

ชั้นคลังข้อสอบตาม ค่าอำนาจจำแนก	ไม่ควบคุมการแสดงผลข้อสอบ				ควบคุมการแสดงผลข้อสอบ			
	a-STR1	a-STR2	a-STR3	a-STR4	a-STR1	a-STR2	a-STR3	a-STR4
ชั้น 0.5	-.710	-.703	-.764	-.730	-.684	-.670	-.661	-.661
ชั้น 1.0	-.395	-.376	-.422	-.367	-.344	-.320	-.370	-.293
ชั้น 1.5	-.242	-.259	-.263	-.261	-.236	-.232	-.252	-.246
ชั้น 2.0	-.173	-.164	-.168	-.167	-.146	-.150	-.159	-.167

ผลการเปรียบเทียบตามตาราง 11 และตาราง 12

ผลการเปรียบเทียบวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก 4 วิธีร่วมกับควบคุมการแสดงผลข้อสอบ พบว่า วิธี a-STR2 a-STR3 และ a-STR4 ที่ชั้น 2 และ 3 ซึ่งเป็นชั้นที่บรรจุข้อสอบอัตราส่วนจำนวนเท่ากับ 0.5 ให้ความคลาดเคลื่อนกำลังสองเฉลี่ยและความลำเอียงเฉลี่ยลดลงต่ำกว่าวิธี a-STR1 ซึ่งบรรจุข้อสอบอัตราส่วนจำนวนเท่ากับ 1.0 เท่ากันทุกชั้นกำลังข้อสอบ

a-STR3 หรือ วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกวิธีที่ 3 ร่วมกับควบคุมการแสดงผลข้อสอบ พบว่า ให้ความคลาดเคลื่อนกำลังสองเฉลี่ยน้อยที่สุดเมื่อการทดสอบครบตามความยาวแบบทดสอบกำหนด เป็นผลมาจากการจัดวางข้อสอบอำนาจจำแนกสูงจำนวนมากที่ชั้นค่าอำนาจจำแนกเท่ากับ 2.0 ซึ่งเป็นชั้นสุดท้ายของการทดสอบ ระยะนี้ค่าประมาณความสามารถเข้าใกล้ค่าความสามารถจริงของผู้สอบมากแล้ว การเลือกข้อสอบความยากใกล้เคียงความสามารถของผู้สอบและเป็นข้อสอบอำนาจจำแนกสูงมิให้เลือกจำนวนมาก จะให้ความคลาดเคลื่อนกำลังสองเฉลี่ยลดลงมากกว่าวิธีอื่น

2. ประสิทธิภาพสมมูลการใช้ข้อสอบในคลังข้อสอบ

การวิเคราะห์ตอนนี้ ผู้วิจัยทำการทดสอบประสิทธิภาพสมมูลการใช้ข้อสอบในคลังข้อสอบที่ใช้วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นค่าอำนาจจำแนก 4 วิธี ด้วยการวิเคราะห์หาอัตราการทับซ้อนข้อสอบ ข้อสอบที่มีการแสดงมากเกินไป ข้อสอบที่มีการใช้ต่ำ และการแจกแจงอัตราการแสดงข้อสอบ จำแนกตามขนาดคลังข้อสอบ 200 400 600 และ 800 ข้อ ความยาวแบบทดสอบ 20 32 และ 44 ข้อ การควบคุมและไม่ควบคุมการแสดงข้อสอบ

ผลการวิเคราะห์อัตราการทับซ้อนข้อสอบ พบว่า ให้ค่าเป็นตัวเลขจำนวนน้อยหลังจุดทศนิยม ผู้วิจัยได้ทำการปรับค่าตัวเลขเพื่อให้การแสดงผลชัดเจนและเข้าใจได้ง่ายขึ้น ค่าอัตราการทับซ้อนข้อสอบในตารางแสดงเป็นตัวเลขจำนวนเต็มและหลังจุดทศนิยม 3 ตำแหน่ง ต้องคูณด้วย 10^{-6} และบวกด้วย 1.99×10^{-4} จึงจะได้ค่าอัตราการทับซ้อนข้อสอบตามผลการวิเคราะห์ที่เป็นจริง

2.1 ประสิทธิภาพสมมูลการใช้ข้อสอบในคลังข้อสอบของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกวิธีที่ 1 (a-STR1)

วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นค่าอำนาจจำแนกวิธีที่ 1 (a-STR1) ตามเงื่อนไข ขนาดคลังข้อสอบ ความยาวแบบทดสอบ และการควบคุมการแสดงข้อสอบ ให้ผลการวิเคราะห์แสดงเป็นอัตราการทับซ้อนข้อสอบ ข้อสอบที่มีการแสดงมากเกินไป และข้อสอบที่มีการใช้ต่ำ ดังนี้

ตาราง 13 อัตราการทับซ้อนข้อสอบ ข้อสอบที่มีการแสดงมากเกินไป และข้อสอบที่มีการใช้ต่ำของ
วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นค่าอำนาจจำแนกวิธีที่ 1 (a-STR1)

ผลการ วิเคราะห์	ความยาว แบบทดสอบ	ไม่ควบคุมการแสดงผลข้อสอบ			ควบคุมการแสดงผลข้อสอบ		
		20 ข้อ	32 ข้อ	44 ข้อ	20 ข้อ	32 ข้อ	44 ข้อ
ขนาดคลังข้อสอบ							
Overlap ¹	200 ข้อ	3.32	3.72	3.87	3.29	ทดสอบไม่ได้	ทดสอบไม่ได้
	400 ข้อ	2.55	3.07	3.39	2.54	3.13	3.43
	600 ข้อ	1.74	2.57	3.05	1.88	2.57	3.05
	800 ข้อ	0.89	2.17	2.74	1.02	2.19	2.76
O.Item ²	200 ข้อ	12	44	95	14	ทดสอบไม่ได้	ทดสอบไม่ได้
	400 ข้อ	4	18	38	7	14	29
	600 ข้อ	4	5	14	6	11	14
	800 ข้อ	3	3	9	5	9	11
U.Item ³	200 ข้อ	188	156	105	186	ทดสอบไม่ได้	ทดสอบไม่ได้
	400 ข้อ	396	382	362	393	386	371
	600 ข้อ	596	595	586	594	589	586
	800 ข้อ	797	797	791	795	791	789

¹ อัตราการทับซ้อนข้อสอบ (Item Overlap Rate : overlap)

² ข้อสอบที่มีการแสดงมากเกินไป (Overexposed Item : O.Item)

³ ข้อสอบที่มีการใช้ต่ำ (Underutilized Item : U.Item)

ผลการวิเคราะห์ตามตาราง 13

วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นค่าอำนาจจำแนกวิธีที่ 1 (a-STR1) เมื่อเพิ่มขนาดคลังข้อสอบ 200 400 600 และ 800 ข้อ พบว่า อัตราการทับซ้อนข้อสอบ เท่ากับ 3.32 2.55 1.74 และ 0.89 ตามลำดับ ที่ความยาวแบบทดสอบ 20 ข้อ ผลการวิเคราะห์มีลักษณะ คล้ายกันเมื่อใช้ความยาวแบบทดสอบ 32 ข้อและ 44 ข้อ

การเพิ่มขนาดคลังข้อสอบ 200 400 600 และ 800 ข้อ ให้ข้อสอบที่มีการแสดงมากเกินไป เท่ากับ 12 4 4 และ 3 ข้อตามลำดับ และข้อสอบที่มีการใช้ต่ำเท่ากับ 188 396 596 และ 797 ข้อตามลำดับที่ความยาวแบบทดสอบ 20 ข้อ ผลการวิเคราะห์มีลักษณะคล้ายกันเมื่อใช้ความยาวแบบทดสอบ 32 ข้อและ 44 ข้อ

ขนาดคลังข้อสอบส่งผลต่ออัตราการทับซ้อนข้อสอบ ข้อสอบที่มีการแสดงมากเกินไป และข้อสอบที่มีการใช้ต่ำ ผลการวิเคราะห์มีลักษณะคล้ายกันทั้งขณะควบคุมและไม่ควบคุมการแสดงผลข้อสอบ

การเพิ่มความยาวแบบทดสอบ 20 32 และ 44 ข้อ พบว่า อัตราการทับซ้อนข้อสอบเท่ากับ 3.32 3.72 และ 3.87 ตามลำดับที่ขนาดคลังข้อสอบ 200 ข้อ ผลการวิเคราะห์มีลักษณะคล้ายกันเมื่อใช้ขนาดคลังข้อสอบ 400 600 และ 800 ข้อ

การเพิ่มความยาวแบบทดสอบ พบว่า ให้ข้อสอบที่มีการแสดงมากเกินไปเท่ากับ 12 44 และ 95 ข้อตามลำดับ และจำนวนข้อสอบที่มีการใช้ต่ำเท่ากับ 188 156 และ 105 ข้อตามลำดับที่ขนาดคลังข้อสอบ 200 ข้อ ผลการวิเคราะห์มีลักษณะคล้ายกันเมื่อใช้ขนาดคลังข้อสอบ 400 600 และ 800 ข้อ

ความยาวแบบทดสอบส่งผลต่ออัตราการทับซ้อนข้อสอบ ข้อสอบที่มีการแสดงมากเกินไป และข้อสอบที่มีการใช้ต่ำ ผลการวิเคราะห์นี้เกิดขึ้นทั้งขณะควบคุมและไม่ควบคุมการแสดงผลข้อสอบ

การควบคุมการแสดงผลข้อสอบ พบว่า อัตราการทับซ้อนข้อสอบเพิ่มขึ้นเล็กน้อย ข้อสอบที่มีการแสดงมากเกินไปจำนวนเพิ่มขึ้น และลดจำนวนข้อสอบที่มีการใช้ต่ำ เมื่อเปรียบเทียบกับขณะไม่ได้ควบคุมการแสดงผลข้อสอบ

การควบคุมการแสดงผลข้อสอบไม่สามารถให้ผลการวิเคราะห์ได้เมื่อใช้กับขนาดคลังข้อสอบ 200 ข้อ และที่ความยาวแบบทดสอบ 32 และ 44 ข้อ

2.2 ประสิทธิภาพสมดุลการใช้ข้อสอบในคลังข้อสอบของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของคำอำนาจจำแนกวิธีที่ 2 (a-STR2)

วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นคำอำนาจจำแนกวิธีที่ 2 (a-STR2) ตามเงื่อนไข ขนาดคลังข้อสอบ ความยาวแบบทดสอบ และการควบคุมการแสดงผลข้อสอบ ให้ผลการวิเคราะห์แสดงเป็นอัตราการทับซ้อนข้อสอบ ข้อสอบที่มีการแสดงมากเกินไป และข้อสอบที่มีการใช้ต่ำ ดังนี้

ตาราง 14 อัตราการทับซ้อนข้อสอบ ข้อสอบที่มีการแสดงมากเกินไป ข้อสอบที่มีการใช้ต่ำ ของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นค่าอำนาจจำแนกวิธีที่ 2 (a-STR2)

ผลการ วิเคราะห์	ความยาว แบบทดสอบ	ไม่ควบคุมการแสดงผลข้อสอบ			ควบคุมการแสดงผลข้อสอบ		
		20 ข้อ	32 ข้อ	44 ข้อ	20 ข้อ	32 ข้อ	44 ข้อ
ขนาดคลังข้อสอบ							
Overlap ¹	200 ข้อ	3.35	3.64	3.78	ทดสอบไม่ได้	ทดสอบไม่ได้	ทดสอบไม่ได้
	400 ข้อ	2.54	3.25	3.37	2.54	ทดสอบไม่ได้	ทดสอบไม่ได้
	600 ข้อ	1.79	2.74	3.05	1.72	2.54	ทดสอบไม่ได้
	800 ข้อ	1.11	2.09	2.70	0.9	2.15	2.75
O.Item ²	200 ข้อ	24	61	86	ทดสอบไม่ได้	ทดสอบไม่ได้	ทดสอบไม่ได้
	400 ข้อ	11	30	70	11	ทดสอบไม่ได้	ทดสอบไม่ได้
	600 ข้อ	4	13	31	7	20	ทดสอบไม่ได้
	800 ข้อ	2	9	30	5	13	24
U.Item ³	200 ข้อ	176	139	114	ทดสอบไม่ได้	ทดสอบไม่ได้	ทดสอบไม่ได้
	400 ข้อ	389	370	330	389	ทดสอบไม่ได้	ทดสอบไม่ได้
	600 ข้อ	596	587	569	593	580	ทดสอบไม่ได้
	800 ข้อ	798	791	770	795	787	776

¹ อัตราการทับซ้อนข้อสอบ (Item Overlap Rate : overlap)

² ข้อสอบที่มีการแสดงมากเกินไป (Overexposed Item : O.Item)

³ ข้อสอบที่มีการใช้ต่ำ (Underutilized Item : U.Item)

ผลการวิเคราะห์ตามตาราง 14

วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นค่าอำนาจจำแนกวิธีที่ 2 (a-STR2) เมื่อเพิ่มขนาดคลังข้อสอบขนาด 200 400 600 และ 800 ข้อ พบว่า อัตราการทับซ้อนข้อสอบเท่ากับ 3.35 2.54 1.79 และ 1.11 ตามลำดับที่ความยาวแบบทดสอบ 20 ข้อ ผลการวิเคราะห์มีลักษณะคล้ายกันเมื่อใช้ความยาวแบบทดสอบ 32 ข้อและ 44 ข้อ

การเพิ่มขนาดคลังข้อสอบ พบว่า ให้ข้อสอบที่มีการแสดงมากเกินไปเท่ากับ 24 11 4 และ 2 ข้อตามลำดับ และข้อสอบที่มีการใช้ต่ำเท่ากับ 176 389 596 และ 798 ข้อตามลำดับที่ความยาวแบบทดสอบ 20 ข้อ ผลการวิเคราะห์มีลักษณะคล้ายกันเมื่อใช้ความยาวแบบทดสอบ 32 ข้อและ 44 ข้อ

ขนาดคลังข้อสอบเพิ่มขึ้นส่งผลต่ออัตราการทับซ้อนข้อสอบ ข้อสอบที่มีการแสดงมากเกินไป และข้อสอบที่มีการใช้ต่ำ แสดงผลการวิเคราะห์ลักษณะคล้ายกันทั้งขณะควบคุมและไม่ควบคุมการแสดงผลข้อสอบ

การเพิ่มความยาวแบบทดสอบ พบว่า อัตราการทับซ้อนข้อสอบเท่ากับ 3.35 3.64 และ 3.78 ตามลำดับที่ขนาดคลังข้อสอบ 200 ข้อ ผลการวิเคราะห์มีลักษณะคล้ายกันเมื่อใช้ขนาดคลังข้อสอบ 400 600 และ 800 ข้อ

การเพิ่มความยาวแบบทดสอบ พบว่า ให้ข้อสอบที่มีการแสดงมากเกินไปเท่ากับ 24 61 และ 86 ข้อตามลำดับ และจำนวนข้อสอบที่มีการใช้ต่ำเท่ากับ 176 139 และ 114 ข้อตามลำดับที่ขนาดคลังข้อสอบ 200 ข้อ ผลการวิเคราะห์มีลักษณะคล้ายกันเมื่อใช้ขนาดคลังข้อสอบ 400 600 และ 800 ข้อ

ความยาวแบบทดสอบเพิ่มขึ้นส่งผลต่ออัตราการทับซ้อนข้อสอบ ข้อสอบที่มีการแสดงมากเกินไป และข้อสอบที่มีการใช้ต่ำ ผลการวิเคราะห์นี้เกิดขึ้นทั้งขณะควบคุมและไม่ควบคุมการแสดงผลข้อสอบ

การควบคุมการแสดงผลข้อสอบ พบว่า ให้อัตราการทับซ้อนข้อสอบใกล้เคียงกัน ข้อสอบที่มีการแสดงมากเกินไปจำนวนเพิ่มขึ้น และลดจำนวนข้อสอบที่มีการใช้ต่ำ เมื่อเปรียบเทียบกับขณะไม่ได้ควบคุมการแสดงผลข้อสอบ

การควบคุมการแสดงผลข้อสอบไม่สามารถให้ผลการวิเคราะห์ได้เมื่อใช้กับขนาดคลังข้อสอบ 200 ข้อ ที่ความยาวแบบทดสอบ 20 32 และ 44 ข้อ ขนาดคลังข้อสอบ 400 ข้อ ที่ความยาวแบบทดสอบ 32 และ 44 ข้อ และขนาดคลังข้อสอบ 600 ข้อ ที่ความยาวแบบทดสอบ 44 ข้อ

2.3 ประสิทธิภาพสมมูลการใช้ข้อสอบในคลังข้อสอบของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับขั้นของค่าอำนาจจำแนกวิธีที่ 3 (a-STR3)

วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับขั้นค่าอำนาจจำแนกวิธีที่ 3 (a-STR3) ตามเงื่อนไข ขนาดคลังข้อสอบ ความยาวแบบทดสอบ และการควบคุมการแสดงผลข้อสอบ ให้ผลการวิเคราะห์แสดงเป็นอัตราการทับซ้อนข้อสอบ ข้อสอบที่มีการแสดงมากเกินไป และข้อสอบที่มีการใช้ต่ำ ดังนี้

ตาราง 15 อัตราการทับซ้อนข้อสอบ ข้อสอบที่มีการแสดงมากเกินไป ข้อสอบที่มีการใช้ต่ำของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นค่าอำนาจจำแนกวิธีที่ 3 (a-STR3)

ผลการ วิเคราะห์	ความยาว แบบทดสอบ	ไม่ควบคุมการแสดงผลข้อสอบ			ควบคุมการแสดงผลข้อสอบ		
		20 ข้อ	32 ข้อ	44 ข้อ	20 ข้อ	32 ข้อ	44 ข้อ
ขนาดคลังข้อสอบ							
Overlap ¹	200 ข้อ	3.37	3.71	3.77	ทดสอบไม่ได้	ทดสอบไม่ได้	ทดสอบไม่ได้
	400 ข้อ	2.54	3.16	3.38	2.67	3.14	ทดสอบไม่ได้
	600 ข้อ	1.87	2.75	2.98	1.72	2.79	ทดสอบไม่ได้
	800 ข้อ	0.87	2.26	2.77	0.90	2.14	2.76
O.Item ²	200 ข้อ	20	45	64	ทดสอบไม่ได้	ทดสอบไม่ได้	ทดสอบไม่ได้
	400 ข้อ	15	35	62	20	40	ทดสอบไม่ได้
	600 ข้อ	5	11	45	8	20	ทดสอบไม่ได้
	800 ข้อ	3	12	28	9	8	36
U.Item ³	200 ข้อ	180	155	136	ทดสอบไม่ได้	ทดสอบไม่ได้	ทดสอบไม่ได้
	400 ข้อ	385	365	338	380	360	ทดสอบไม่ได้
	600 ข้อ	595	589	555	592	580	ทดสอบไม่ได้
	800 ข้อ	797	788	772	791	792	764

¹ อัตราการทับซ้อนข้อสอบ (Item Overlap Rate : overlap)

² ข้อสอบที่มีการแสดงมากเกินไป (Overexposed Item : O.Item)

³ ข้อสอบที่มีการใช้ต่ำ (Underutilized Item : U.Item)

การทดสอบตามตาราง 15 สรุปดังนี้

วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นค่าอำนาจจำแนกวิธีที่ 3 (a-STR3) เมื่อเพิ่มขนาดคลังข้อสอบ 200 400 600 และ 800 ข้อ พบว่า อัตราการทับซ้อนข้อสอบ เท่ากับ 3.37 2.54 1.87 และ 0.87 ตามลำดับที่ความยาวแบบทดสอบ 20 ข้อ ผลการวิเคราะห์มีลักษณะ คล้ายกันเมื่อใช้ความยาวแบบทดสอบ 32 ข้อและ 44 ข้อ

การเพิ่มขนาดคลังข้อสอบ พบว่า ให้ข้อสอบที่มีการแสดงมากเกินไปเท่ากับ 20 15 5 และ 3 ข้อตามลำดับ และข้อสอบที่มีการใช้ต่ำเท่ากับ 180 385 595 และ 797 ข้อตามลำดับที่ความยาวแบบทดสอบ 20 ข้อ ผลการวิเคราะห์มีลักษณะคล้ายกันเมื่อใช้ความยาวแบบทดสอบ 32 ข้อและ 44 ข้อ

ขนาดคลังข้อสอบเพิ่มขึ้นส่งผลต่ออัตราการทับซ้อนข้อสอบ ข้อสอบที่มีการแสดงมากเกินไป และข้อสอบที่มีการใช้ต่ำ ผลการวิเคราะห์มีลักษณะคล้ายกันทั้งขณะควบคุมและไม่ควบคุมการแสดงผลข้อสอบ

การเพิ่มความยาวแบบทดสอบ 20 32 และ 44 ข้อ พบว่า ให้อัตราการทับซ้อนข้อสอบเท่ากับ 3.37 3.71 และ 3.77 ตามลำดับที่ขนาดคลังข้อสอบ 200 ข้อ ผลการวิเคราะห์มีลักษณะคล้ายกันเมื่อใช้ขนาดคลังข้อสอบ 400 600 และ 800 ข้อ

การเพิ่มความยาวแบบทดสอบ พบว่า ข้อสอบที่มีการแสดงมากเกินไปเท่ากับ 20 45 และ 64 ข้อตามลำดับ และข้อสอบที่มีการใช้ต่ำเท่ากับ 180 155 และ 136 ข้อตามลำดับที่ขนาดคลังข้อสอบ 200 ข้อ ผลการวิเคราะห์มีลักษณะคล้ายกันเมื่อใช้ขนาดคลังข้อสอบ 400 600 และ 800 ข้อ

ความยาวแบบทดสอบเพิ่มขึ้นให้อัตราการทับซ้อนข้อสอบเพิ่มขึ้น ข้อสอบที่มีการแสดงมากเกินไปจำนวนเพิ่มขึ้น และลดจำนวนข้อสอบที่มีการใช้ต่ำ ให้ผลการวิเคราะห์ทั้งขณะควบคุมและไม่ควบคุมการแสดงผลข้อสอบ

การควบคุมการแสดงผลข้อสอบ พบว่า ให้อัตราการทับซ้อนข้อสอบใกล้เคียงกัน ข้อสอบที่มีการแสดงมากเกินไปจำนวนเพิ่มขึ้น และลดจำนวนข้อสอบที่มีการใช้ต่ำ เมื่อเปรียบเทียบกับขณะไม่ได้ควบคุมการแสดงผลข้อสอบ

การควบคุมการแสดงผลข้อสอบไม่สามารถให้ผลการวิเคราะห์ได้เมื่อใช้กับขนาดคลังข้อสอบ 200 ข้อ ที่ความยาวแบบทดสอบ 20 32 และ 44 ข้อ ขนาดคลังข้อสอบ 400 ข้อ ที่ความยาวแบบทดสอบ 44 ข้อ และขนาดคลังข้อสอบ 600 ข้อ ที่ความยาวแบบทดสอบ 44 ข้อ

2.4 ประสิทธิภาพสมมูลการใช้ข้อสอบในคลังข้อสอบของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับขั้นของค่าอำนาจจำแนกวิธีที่ 4 (a-STR4)

วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับขั้นค่าอำนาจจำแนกวิธีที่ 4 (a-STR4) ตามเงื่อนไข ขนาดคลังข้อสอบ ความยาวแบบทดสอบ และการควบคุมการแสดงผลข้อสอบ ให้ผลการวิเคราะห์แสดงเป็นอัตราการทับซ้อนข้อสอบ ข้อสอบที่มีการแสดงมากเกินไป และข้อสอบที่มีการใช้ต่ำ ดังนี้

ตาราง 16 อัตราการทับซ้อนข้อสอบ ข้อสอบที่มีการแสดงมากเกินไป ข้อสอบที่มีการใช้ต่ำของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นค่าอำนาจจำแนกวิธีที่ 4 (a-STR4)

ผลการ วิเคราะห์	ความยาว แบบทดสอบ	ไม่ควบคุมการแสดงผลข้อสอบ			ควบคุมการแสดงผลข้อสอบ		
		20 ข้อ	32 ข้อ	44 ข้อ	20 ข้อ	32 ข้อ	44 ข้อ
ขนาดคลังข้อสอบ							
Overlap ¹	200 ข้อ	3.39	3.66	3.75	ทดสอบไม่ได้	ทดสอบไม่ได้	ทดสอบไม่ได้
	400 ข้อ	2.47	3.14	3.37	2.52	ทดสอบไม่ได้	ทดสอบไม่ได้
	600 ข้อ	1.70	2.51	3.03	1.71	2.59	ทดสอบไม่ได้
	800 ข้อ	0.90	1.98	2.62	1.08	2.27	2.66
O.Item ²	200 ข้อ	21	48	75	ทดสอบไม่ได้	ทดสอบไม่ได้	ทดสอบไม่ได้
	400 ข้อ	8	32	43	16	ทดสอบไม่ได้	ทดสอบไม่ได้
	600 ข้อ	4	13	38	7	19	ทดสอบไม่ได้
	800 ข้อ	5	11	14	7	14	23
U.Item ³	200 ข้อ	179	152	125	ทดสอบไม่ได้	ทดสอบไม่ได้	ทดสอบไม่ได้
	400 ข้อ	392	368	357	384	ทดสอบไม่ได้	ทดสอบไม่ได้
	600 ข้อ	596	587	562	593	581	ทดสอบไม่ได้
	800 ข้อ	795	789	786	793	786	777

¹ อัตราการทับซ้อนข้อสอบ (Item Overlap Rate : overlap)

² ข้อสอบที่มีการแสดงมากเกินไป (Overexposed Item : O.Item)

³ ข้อสอบที่มีการใช้ต่ำ (Underutilized Item : U.Item)

การทดสอบตามตาราง 16 สรุปดังนี้

วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นค่าอำนาจจำแนกวิธีที่ 4 (a-STR4) เมื่อเพิ่มขนาดคลังข้อสอบ 200 400 600 และ 800 ข้อ พบว่า อัตราการทับซ้อนข้อสอบ เท่ากับ 3.39 2.47 1.70 และ 0.90 ตามลำดับที่ความยาวแบบทดสอบ 20 ข้อ ผลการวิเคราะห์มีลักษณะคล้ายกันเมื่อใช้ความยาวแบบทดสอบ 32 ข้อและ 44 ข้อ

การเพิ่มขนาดคลังข้อสอบ พบว่า ข้อสอบที่มีการแสดงมากเกินไปจำนวนเท่ากับ 21 8 4 และ 5 ข้อตามลำดับ และข้อสอบที่มีการใช้ต่ำเท่ากับ 179 392 596 และ 795 ข้อตามลำดับที่ความยาวแบบทดสอบ 20 ข้อ ผลการวิเคราะห์มีลักษณะคล้ายกันเมื่อใช้ความยาวแบบทดสอบ 32 ข้อและ 44 ข้อ

ขนาดคลังข้อสอบเพิ่มขึ้น พบว่า ให้อัตราการทับซ้อนข้อสอบลดลง ข้อสอบที่มีการแสดงมากเกินไปลดลง และเพิ่มจำนวนข้อสอบที่มีการใช้ต่ำ ผลการวิเคราะห์มีลักษณะคล้ายกันทั้งขณะควบคุมและไม่ควบคุมการแสดงผลข้อสอบ

การเพิ่มความยาวแบบทดสอบ 20 32 และ 44 ข้อ พบว่า อัตราการทับซ้อนข้อสอบเท่ากับ 3.39 3.66 และ 3.75 ตามลำดับที่ขนาดคลังข้อสอบ 200 ข้อ ผลการวิเคราะห์มีลักษณะคล้ายกันเมื่อใช้ขนาดคลังข้อสอบ 400 600 และ 800 ข้อ

การเพิ่มความยาวแบบทดสอบ พบว่า ให้ข้อสอบที่มีการแสดงมากเกินไปเท่ากับ 21 48 และ 75 ข้อตามลำดับ และข้อสอบที่มีการใช้ต่ำเท่ากับ 179 152 และ 125 ข้อตามลำดับที่ขนาดคลังข้อสอบ 200 ข้อ ผลการวิเคราะห์มีลักษณะคล้ายกันเมื่อใช้ขนาดคลังข้อสอบ 400 600 และ 800 ข้อ

ความยาวแบบทดสอบเพิ่มขึ้นให้อัตราการทับซ้อนข้อสอบเพิ่มขึ้น ข้อสอบที่มีการแสดงมากเกินไปจำนวนเพิ่มขึ้น และลดจำนวนข้อสอบที่มีการใช้ต่ำ ผลการวิเคราะห์นี้เกิดขึ้นทั้งขณะควบคุมและไม่ควบคุมการแสดงผลข้อสอบ

การควบคุมการแสดงผลข้อสอบ พบว่า ให้อัตราการทับซ้อนข้อสอบเพิ่มขึ้นเล็กน้อย ข้อสอบที่มีการแสดงมากเกินไปจำนวนเพิ่มขึ้น และลดจำนวนข้อสอบที่มีการใช้ต่ำ เมื่อเปรียบเทียบกับขณะไม่ได้ควบคุมการแสดงผลข้อสอบ

การควบคุมการแสดงผลข้อสอบไม่สามารถให้ผลการวิเคราะห์ได้เมื่อใช้กับขนาดคลังข้อสอบ 200 ข้อ ที่ความยาวแบบทดสอบ 20 32 และ 44 ข้อ ขนาดคลังข้อสอบ 400 ข้อ ที่ความยาวแบบทดสอบ 32 และ 44 ข้อ และขนาดคลังข้อสอบ 600 ข้อ ที่ความยาวแบบทดสอบ 44 ข้อ

2.5 เปรียบเทียบประสิทธิภาพสมดุลการใช้ข้อสอบในคลังข้อสอบระหว่างวิธีการทดสอบแบบปรับ เหมาะสมด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก 4 วิธี

การเปรียบเทียบประสิทธิภาพสมดุลการใช้ข้อสอบในคลังข้อสอบระหว่างวิธีการทดสอบแบบปรับเหมาะสมด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก 4 วิธี โดยตรวจสอบวิธีที่ให้ค่าอัตราการทับซ้อนข้อสอบค่าน้อยที่สุด ข้อสอบที่มีการแสดงมากเกินไปจำนวนน้อยที่สุด จำนวนข้อสอบที่มีการใช้ต่ำมากที่สุด และการแจกแจงอัตราการแสดงผลข้อสอบสมดุลที่สุด เมื่อการทดสอบครบตามความยาวแบบทดสอบกำหนด แสดงผลเปรียบเทียบดังนี้

ตาราง 17 เปรียบเทียบอัตราการทับซ้อนข้อสอบและข้อสอบที่มีการแสดงมากเกินไปของวิธีการทดสอบ
แบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก 4 วิธี

คลังข้อสอบ	ความยาวแบบทดสอบ	อัตราการทับซ้อนข้อสอบ				ข้อสอบที่มีการแสดงมากเกินไป			
		a-STR1	a-STR2	a-STR3	a-STR4	a-STR1	a-STR2	a-STR3	a-STR4
200	20	น้อยที่สุด				น้อยที่สุด			
400						น้อยที่สุด			
600						น้อยที่สุด	น้อยที่สุด		น้อยที่สุด
800		น้อยที่สุด				น้อยที่สุด			
200	32	น้อยที่สุด				น้อยที่สุด			
400		น้อยที่สุด				น้อยที่สุด			
600						น้อยที่สุด			
800		น้อยที่สุด				น้อยที่สุด			
200	44					น้อยที่สุด		น้อยที่สุด	
400		น้อยที่สุด				น้อยที่สุด			
600		น้อยที่สุด				น้อยที่สุด			
800		น้อยที่สุด				น้อยที่สุด			

ตาราง 18 เปรียบเทียบข้อสอบที่มีการใช้ต่ำและการแจกแจงอัตราการแสดงข้อสอบของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก 4 วิธี

คลังข้อสอบ	ความยาวแบบทดสอบ	ข้อสอบที่มีการใช้ต่ำ				การแจกแจงอัตราการแสดงข้อสอบ			
		a-STR1	a-STR2	a-STR3	a-STR4	a-STR1	a-STR2	a-STR3	a-STR4
200	20	มากที่สุด				สมดุลที่สุด			
400		มากที่สุด				สมดุลที่สุด			
600		มากที่สุด	มากที่สุด		มากที่สุด	สมดุลที่สุด			
800			มากที่สุด			สมดุลที่สุด			
200	32	มากที่สุด				สมดุลที่สุด			
400		มากที่สุด				สมดุลที่สุด			
600		มากที่สุด				สมดุลที่สุด			
800		มากที่สุด				สมดุลที่สุด			
200	44	มากที่สุด				สมดุลที่สุด			
400		มากที่สุด				สมดุลที่สุด			
600		มากที่สุด				สมดุลที่สุด			
800		มากที่สุด				สมดุลที่สุด			

ผลการเปรียบเทียบตามตาราง 17 และ 18

ผลการเปรียบเทียบของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นค่าอำนาจจำแนก 4 วิธี พิจารณาวีธีที่ให้อัตราการทับซ้อนข้อสอบน้อยที่สุดในภาพรวม พบว่า

วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นค่าอำนาจจำแนก 4 วิธี ให้อัตราการทับซ้อนข้อสอบน้อยที่สุดได้เท่าเทียมกัน พิจารณาจากอัตราการทับซ้อนข้อสอบน้อยที่สุดมีการกระจายไปยังวิธีกำหนดอื่นๆ ที่ขนาดคั่งข้อสอบและความยาวข้อสอบต่างๆ

วิธี a-STR1 หรือ วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นค่าอำนาจจำแนกวิธีที่ 1 ให้ข้อสอบที่มีการแสดงมากที่สุดไปน้อยที่สุด และข้อสอบที่มีการใช้ต่ำมากที่สุด และให้การแจกแจงอัตราการแสดงข้อสอบสมดุลที่สุดที่ทุกขนาดคั่งข้อสอบและความยาวแบบทดสอบ

ผลการเปรียบเทียบประสิทธิภาพสมมูลการใช้ข้อสอบในคลังข้อสอบระหว่างวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก 4 วิธีร่วมกับการควบคุมการแสดงผลข้อสอบ พิจารณาวิธีที่ให้อัตราการทับซ้อนข้อสอบค่าน้อยที่สุด ข้อสอบที่มีการแสดงมากเกินไปจำนวนน้อยที่สุด จำนวนข้อสอบที่มีการใช้ต่ำมากที่สุด และการแจกแจงอัตราการแสดงผลข้อสอบสมมูลที่สุด เมื่อการทดสอบครบตามความยาวแบบทดสอบกำหนด แสดงผลดังนี้

ตาราง 19 เปรียบเทียบอัตราการทับซ้อนข้อสอบและข้อสอบที่มีการแสดงมากเกินไปของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกทั้ง 4 วิธีร่วมกับการควบคุมการแสดงผลข้อสอบ

คลังข้อสอบ	ความยาวแบบทดสอบ	อัตราการทับซ้อนข้อสอบ				ข้อสอบที่มีการแสดงมากเกินไป			
		a-STR1	a-STR2	a-STR3	a-STR4	a-STR1	a-STR2	a-STR3	a-STR4
200	20	น้อยที่สุด	ทดสอบไม่ได้	ทดสอบไม่ได้	ทดสอบไม่ได้	น้อยที่สุด	ทดสอบไม่ได้	ทดสอบไม่ได้	ทดสอบไม่ได้
400					น้อยที่สุด	น้อยที่สุด			
600					น้อยที่สุด	น้อยที่สุด			
800			น้อยที่สุด	น้อยที่สุด		น้อยที่สุด	น้อยที่สุด		
200	32	ทดสอบไม่ได้	ทดสอบไม่ได้	ทดสอบไม่ได้	ทดสอบไม่ได้	ทดสอบไม่ได้	ทดสอบไม่ได้	ทดสอบไม่ได้	ทดสอบไม่ได้
400		น้อยที่สุด	ทดสอบไม่ได้		ทดสอบไม่ได้	น้อยที่สุด	ทดสอบไม่ได้		ทดสอบไม่ได้
600			น้อยที่สุด				น้อยที่สุด		
800				น้อยที่สุด				น้อยที่สุด	
200	44	ทดสอบไม่ได้	ทดสอบไม่ได้	ทดสอบไม่ได้	ทดสอบไม่ได้	ทดสอบไม่ได้	ทดสอบไม่ได้	ทดสอบไม่ได้	ทดสอบไม่ได้
400		น้อยที่สุด	ทดสอบไม่ได้	ทดสอบไม่ได้	ทดสอบไม่ได้	น้อยที่สุด	ทดสอบไม่ได้	ทดสอบไม่ได้	ทดสอบไม่ได้
600		น้อยที่สุด	ทดสอบไม่ได้	ทดสอบไม่ได้	ทดสอบไม่ได้	น้อยที่สุด	ทดสอบไม่ได้	ทดสอบไม่ได้	ทดสอบไม่ได้
800				น้อยที่สุด	น้อยที่สุด				

ตาราง 20 เปรียบเทียบข้อสอบที่มีการใช้ค่าและการแจกแจงอัตราการแสดงข้อสอบของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกทั้ง 4 วิธีร่วมกับการควบคุมการแสดงผลข้อสอบ

คำสั่งข้อสอบ	ความยาวแบบทดสอบ	ข้อสอบที่มีการใช้ค่า				การแจกแจงอัตราการแสดงข้อสอบ			
		a-STR1	a-STR2	a-STR3	a-STR4	a-STR1	a-STR2	a-STR3	a-STR4
200	20	มากที่สุด	ทดสอบ ไม่ได้	ทดสอบ ไม่ได้	ทดสอบ ไม่ได้	สมมูลที่สุด	ทดสอบ ไม่ได้	ทดสอบ ไม่ได้	ทดสอบ ไม่ได้
400		มากที่สุด				สมมูลที่สุด			
600		มากที่สุด				สมมูลที่สุด			
800		มากที่สุด	มากที่สุด			สมมูลที่สุด			
200	32	ทดสอบ ไม่ได้	ทดสอบ ไม่ได้	ทดสอบ ไม่ได้	ทดสอบ ไม่ได้	ทดสอบ ไม่ได้	ทดสอบ ไม่ได้	ทดสอบ ไม่ได้	ทดสอบ ไม่ได้
400		มากที่สุด	ทดสอบ ไม่ได้		ทดสอบ ไม่ได้	สมมูลที่สุด	ทดสอบ ไม่ได้		ทดสอบ ไม่ได้
600			มากที่สุด			สมมูลที่สุด			
800				มากที่สุด		สมมูลที่สุด			
200	44	ทดสอบ ไม่ได้	ทดสอบ ไม่ได้	ทดสอบ ไม่ได้	ทดสอบ ไม่ได้	ทดสอบ ไม่ได้	ทดสอบ ไม่ได้	ทดสอบ ไม่ได้	ทดสอบ ไม่ได้
400		มากที่สุด	ทดสอบ ไม่ได้	ทดสอบ ไม่ได้	ทดสอบ ไม่ได้	สมมูลที่สุด	ทดสอบ ไม่ได้	ทดสอบ ไม่ได้	ทดสอบ ไม่ได้
600		มากที่สุด	ทดสอบ ไม่ได้	ทดสอบ ไม่ได้	ทดสอบ ไม่ได้	สมมูลที่สุด	ทดสอบ ไม่ได้	ทดสอบ ไม่ได้	ทดสอบ ไม่ได้
800		มากที่สุด					สมมูลที่สุด		

ผลการเปรียบเทียบตามตาราง 19 และ 20

วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นค่าอำนาจจำแนก กำหนดอัตราส่วนจำนวนข้อสอบต่างกัน 4 วิธีร่วมกับการควบคุมการแสดงผลข้อสอบ พิจารณาวิธีที่ให้อัตราการทำข้อสอบน้อยที่สุดในภาพรวม พบว่า

วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก 4 วิธี สามารถให้อัตราการทับซ้อนข้อสอบน้อยที่สุดเท่าเทียมกัน เนื่องจากอัตราการทับซ้อนข้อสอบน้อยที่สุดมีการกระจายไปยังวิธีกำหนดต่างๆ ที่ขนาดคลังข้อสอบและความยาวแบบทดสอบต่างๆ กัน

วิธี a-STR1 หรือ วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นค่าอำนาจจำแนกวิธีที่ 1 ร่วมกับการควบคุมการแสดงผลข้อสอบ ให้ข้อสอบที่มีการแสดงมากเกินไปน้อยที่สุด ข้อสอบที่มีการใช้ต่ำมากที่สุด และการแจกแจงอัตราการแสดงผลข้อสอบสมดุลที่สุด ที่ขนาดคลังข้อสอบและความยาวแบบทดสอบต่างๆ

ผลการเปรียบเทียบเบื้องต้นนำไปสู่การค้นหาคำตอบว่า วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกวิธีที่ 1 (a-STR1) ให้ประสิทธิภาพสมดุลการใช้ข้อสอบในคลังข้อสอบดีที่สุดได้อย่างไร ผู้วิจัยได้ทำการศึกษาผลการวิเคราะห์ข้อมูลในแต่ละระดับชั้นค่าอำนาจจำแนก เพื่อติดตามการเปลี่ยนแปลง อัตราการทับซ้อนข้อสอบ ข้อสอบที่มีการใช้มากเกินไป และข้อสอบที่มีการใช้ต่ำ ของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก 4 วิธี ผลการวิเคราะห์อธิบายได้ดังนี้

ตาราง 21 เปรียบเทียบอัตราการทับซ้อนข้อสอบของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก 4 วิธี ร่วมกับการควบคุมและไม่ควบคุมการแสดงผลข้อสอบ ขนาดคลังข้อสอบ 800 ข้อ ความยาวแบบทดสอบ 32 ข้อ

ชั้นค่าอำนาจจำแนก	ไม่ควบคุมการแสดงผลข้อสอบ				ควบคุมการแสดงผลข้อสอบ			
	a-STR1	a-STR2	a-STR3	a-STR4	a-STR1	a-STR2	a-STR3	a-STR4
ชั้น 0.5	.509	.059	.569	.219	.579	.009	.569	.339
ชั้น 1.0	.519	.749	.759	.790	.519	.840	.800	.840
ชั้น 1.5	.589	.770	.780	.749	.559	.820	.770	.810
ชั้น 2.0	.559	.519	.159	.229	.539	.489	.019	.289

ตาราง 22 เปรียบเทียบข้อสอบที่มีการแสดงมากเกินไปของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก 4 วิธี ร่วมกับการควบคุมและไม่ควบคุมการแสดงผลข้อสอบ ขนาดคลังข้อสอบ 800 ข้อ ความยาวแบบทดสอบ 32 ข้อ

ระดับชั้นข้อสอบตาม ค่าอำนาจจำแนก	ไม่ควบคุมการแสดงผลข้อสอบ				ควบคุมการแสดงผลข้อสอบ			
	a-STR1	a-STR2	a-STR3	a-STR4	a-STR1	a-STR2	a-STR3	a-STR4
ชั้น 0.5	4	1	4	3	4	7	5	5
ชั้น 1.0	1	7	5	6	7	6	7	7
ชั้น 1.5	0	5	2	4	0	7	8	7
ชั้น 2.0	0	0	0	0	0	0	0	0

ตาราง 23 เปรียบเทียบข้อสอบที่มีการใช้ต่ำของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก 4 วิธี ร่วมกับการควบคุมและไม่ควบคุมการแสดงผลข้อสอบ ขนาดคลังข้อสอบ 800 ข้อ ความยาวแบบทดสอบ 32 ข้อ

ระดับชั้นข้อสอบตาม ค่าอำนาจจำแนก	ไม่ควบคุมการแสดงผลข้อสอบ				ควบคุมการแสดงผลข้อสอบ			
	a-STR1	a-STR2	a-STR3	a-STR4	a-STR1	a-STR2	a-STR3	a-STR4
ชั้น 0.5	196	399	196	297	196	393	195	295
ชั้น 1.0	199	93	95	94	193	94	93	97
ชั้น 1.5	200	95	98	96	200	93	92	97
ชั้น 2.0	200	200	400	300	200	200	400	300
รวม	795	787	789	787	789	780	780	781

ผลการวิเคราะห์ตามตาราง 21 22 และ 23

วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก 4 วิธี พบว่า ชั้นที่มีจำนวนข้อสอบจัดวางในชั้นจำนวนมากจะให้อัตราการทับซ้อนข้อสอบต่ำ ข้อสอบที่มีการแสดงมากเกินไปจำนวนน้อย และให้ข้อสอบที่มีการใช้ต่ำจำนวนมาก ได้แก่ ชั้นที่ 1 ของวิธี a-STR2 ชั้นที่ 4 ของวิธี a-STR3 ชั้นที่ 1 และชั้นที่ 4 ของวิธี a-STR4

ในขณะเดียวกัน ชั้นที่มีอัตราส่วนจำนวนข้อสอบน้อยจะให้อัตราการทับซ้อนข้อสอบสูง ข้อสอบที่มีการแสดงมากเกินไปจำนวนมาก และข้อสอบที่มีการใช้ต่ำจำนวนลดลง ได้แก่ ชั้นที่ 2 และ 3 ของวิธี a-STR2 a-STR3 และ a-STR4

เปรียบเทียบที่ระดับชั้นค่าอำนาจจำแนกของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกวิธีที่ 1 (a-STR1) พบว่า ให้อัตราการทับซ้อนข้อสอบเท่าเทียมกันทุกระดับชั้นค่าอำนาจจำแนก แสดงถึงการกระจายการใช้ข้อสอบไปยังข้อสอบอื่นในชั้นอย่างทั่วถึง มีผลต่อการลดข้อสอบที่มีการแสดงมากเกินไป และให้ข้อสอบที่มีการใช้ต่ำจำนวนมาก ดังนั้น ผลการเปรียบเทียบจึงพบว่า วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกวิธีที่ 1 (a-STR1) สามารถให้ประสิทธิภาพสมดุลการใช้ข้อสอบในคลังข้อสอบได้ดีเมื่อเปรียบเทียบกับวิธีอื่นๆ

3. สรุปวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นค่าอำนาจจำแนกที่ไม่สามารถให้ผลการวิเคราะห์ได้

วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นค่าอำนาจจำแนก 4 วิธีร่วมกับการควบคุมการแสดงข้อสอบ เพื่อศึกษาประสิทธิภาพการประมาณค่าความสามารถของผู้สอบและประสิทธิภาพสมดุลการใช้ข้อสอบในคลังข้อสอบ พบว่า บางครั้งไม่สามารถให้ผลการวิเคราะห์ได้ รวบรวมและสรุปดังนี้

ตาราง 24 สรุปวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นค่าอำนาจจำแนก 4 วิธีไม่สามารถให้ผลการวิเคราะห์ได้

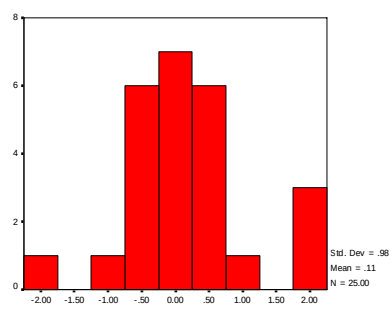
ขนาด คลัง ข้อสอบ	ความยาว แบบ ทดสอบ	วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นค่าอำนาจจำแนก							
		a-STR1		a-STR2		a-STR3		a-STR4	
		-SH	+SH	-SH	+SH	-SH	+SH	-SH	+SH
200 ข้อ	20 ข้อ				ทดสอบ ไม่ได้		ทดสอบ ไม่ได้		ทดสอบ ไม่ได้
	32 ข้อ		ทดสอบ ไม่ได้		ทดสอบ ไม่ได้		ทดสอบ ไม่ได้		ทดสอบ ไม่ได้
	44 ข้อ		ทดสอบ ไม่ได้		ทดสอบ ไม่ได้		ทดสอบ ไม่ได้		ทดสอบ ไม่ได้
400 ข้อ	20 ข้อ								
	32 ข้อ				ทดสอบ ไม่ได้				ทดสอบ ไม่ได้
	44 ข้อ				ทดสอบ ไม่ได้		ทดสอบ ไม่ได้		ทดสอบ ไม่ได้
600 ข้อ	20 ข้อ								
	32 ข้อ								
	44 ข้อ				ทดสอบ ไม่ได้		ทดสอบ ไม่ได้		ทดสอบ ไม่ได้
800 ข้อ	20 ข้อ								
	32 ข้อ								
	44 ข้อ								

-SH แทน ไม่ควบคุมการแสดงผลข้อสอบ +SH แทน ควบคุมการแสดงผลข้อสอบ

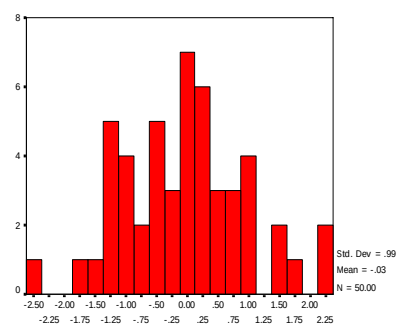
ผลการสรุปตามตาราง 24

วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นค่าอำนาจจำแนก 4 วิธีร่วมกับการควบคุมการแสดงผลข้อสอบ พบว่า เมื่อใช้ขนาดคลังข้อสอบ 200 ข้อ หรือเมื่อใช้ความยาวแบบทดสอบ 32 และ 44 ข้อ มีโอกาสที่จะไม่สามารถให้ผลการวิเคราะห์ได้

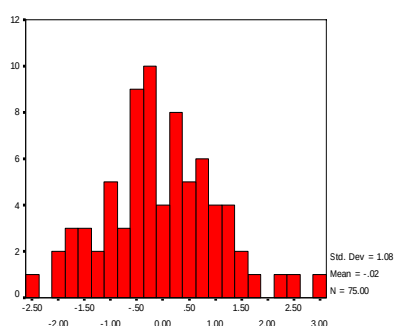
ผลการสรุปดังกล่าวนำไปสู่การค้นหาค่าเหตุของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก 4 วิธีที่ไม่สามารถให้ผลการวิเคราะห์ได้ ผู้วิจัยได้ทำการศึกษาเพิ่มเติมเกี่ยวกับโครงสร้างการทดสอบและขั้นตอนการทดสอบ ผลการศึกษาริขัยได้ดังนี้



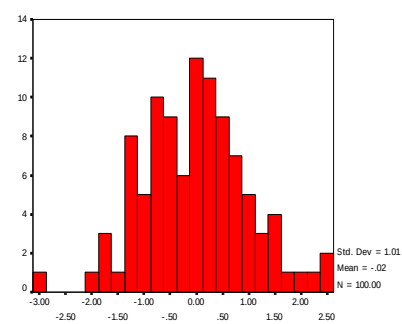
ค่าความยากข้อสอบ 25 ค่า



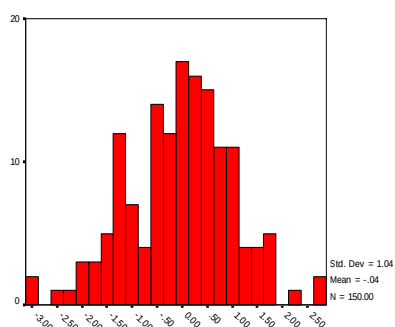
ค่าความยากข้อสอบ 50 ค่า



ค่าความยากข้อสอบ 75 ค่า



ค่าความยากข้อสอบ 100 ค่า



ค่าความยากข้อสอบ 150 ค่า

ภาพประกอบ 8 การแจกแจงโค้งปกติมาตรฐานของค่าความยากของข้อสอบ

ผลการศึกษาตามภาพประกอบ 8

ภาพประกอบ 8 แสดงตัวอย่างค่าความยากของข้อสอบซึ่งมีการแจกแจงเป็น โค้งปกติ มาตรฐาน 25 50 75 100 และ 150 ค่า ค่าความยากของข้อสอบเหล่านี้จะนำไปจัดวางที่แต่ละชั้นค่า อำนาจจำแนกตามวิธีกำหนด พบว่า ค่าความยากของข้อสอบมีลักษณะการกระจายไม่สม่ำเสมอบนเส้น ต่อเนื่องค่าความสามารถของผู้สอบตั้งแต่ -3.0 ถึง 3.0 โดยเฉพาะบริเวณปลายโค้งปกติมาตรฐาน ดังนั้น การสร้างค่าความยากของข้อสอบจากการสุ่มการแจกแจง โค้งปกติมาตรฐาน จะไม่รับรองว่าจะมี ค่าความยากของข้อสอบการกระจายสม่ำเสมอบนเส้นต่อเนื่องค่าความสามารถของผู้สอบ

วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับขั้นของค่าอำนาจจำแนกขณะ ไม่ได้ใช้การควบคุมการแสดงผลข้อสอบพบว่า การทดสอบมีการเลือกใช้ข้อสอบบริเวณปลายโค้งปกติ มาตรฐานสูงมากเกินไปอัตราการแสดงผลข้อสอบกำหนด เนื่องจากมีข้อสอบจำนวนน้อยและกระจายไม่ สมบูรณ์บริเวณปลายโค้งปกติมาตรฐานเมื่อเทียบกับจำนวนผู้สอบหรือค่าความสามารถจริง 5,000 ค่า การเลือกข้อสอบบริเวณปลายโค้งปกติมาตรฐานจึงมีโอกาสน้อยที่ผู้สอบจะได้รับข้อสอบความยาก ใกล้เคียงค่าประมาณความสามารถ การได้รับข้อสอบที่มีความยากห่างไกลจากค่าความสามารถของ ผู้สอบมีผลทำให้ความคลาดเคลื่อนกำลังสองเฉลี่ยและค่าความลำเอียงเฉลี่ยมีขนาดใหญ่ นำไปสู่การลด ประสิทธิภาพการประมาณค่าความสามารถของผู้สอบได้

วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับขั้นของค่าอำนาจจำแนกเมื่อใช้ ร่วมกับการควบคุมการแสดงผลข้อสอบ พบว่า มีการเลือกใช้ข้อสอบบริเวณปลายโค้งปกติมาตรฐานสูง มากเกินอัตราการแสดงผลข้อสอบกำหนด ทำให้พารามิเตอร์การแสดงผลข้อสอบของข้อสอบบริเวณปลาย โค้งปกติมาตรฐานมีค่าอยู่ในระดับต่ำมาก จนกระทั่งแบบทดสอบไม่สามารถเลือกข้อสอบบริเวณนี้จัด ให้แก่ผู้สอบซึ่งมีค่าความสามารถใกล้เคียงค่าความยากของข้อสอบบริเวณปลายโค้งปกติมาตรฐานได้ ข้อสอบตำแหน่งห่างออกไปที่มีค่าความยากใกล้เคียงค่าความสามารถของผู้สอบมากที่สุดจะได้รับเลือก และจัดให้แทน ส่งผลให้ความคลาดเคลื่อนกำลังสองเฉลี่ยและค่าความลำเอียงเฉลี่ยมากกว่าขณะไม่ ควบคุมการแสดงผลข้อสอบ นำไปสู่การลดประสิทธิภาพการประมาณค่าความสามารถของผู้สอบ และ อาจเป็นสาเหตุที่ไม่สามารถให้ผลการวิเคราะห์ได้ เมื่อแบบทดสอบไม่สามารถหาข้อสอบจัดให้แก่ ผู้สอบได้

การกระจายค่าความยากของข้อสอบไม่สม่ำเสมอบนเส้นต่อเนื่องค่าความสามารถของผู้สอบ ต่อการลดประสิทธิภาพการประมาณค่าความสามารถของผู้สอบ สามารถนำไปใช้อธิบายการลด ประสิทธิภาพสมมูลการใช้ข้อสอบในคลังข้อสอบเช่นกัน เมื่อข้อสอบมีจำนวนน้อยบริเวณปลายโค้ง ปกติมาตรฐานเมื่อเทียบกับจำนวนผู้สอบหรือค่าความสามารถจริง 5000 ค่า ข้อสอบปลายโค้งปกติ มาตรฐานจึงมีการใช้สูงมาก ให้อัตราการทับซ้อนข้อสอบสูง ข้อสอบที่มีการแสดงมากเกินไปจำนวน มาก และข้อสอบที่มีการใช้ต่ำจำนวนลดลง

วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกเมื่อใช้ร่วมกับการควบคุมการแสดงผลข้อสอบ พบว่า ข้อสอบปลายโค้งปกติมาตรฐานที่มีการใช้สูงมาก ทำให้พารามิเตอร์การแสดงผลข้อสอบของข้อสอบบริเวณปลายโค้งปกติมาตรฐานมีค่าอยู่ในระดับต่ำมาก จนกระทั่งแบบทดสอบไม่สามารถเลือกข้อสอบบริเวณนี้จัดให้แก่ผู้สอบซึ่งมีค่าความสามารถใกล้เคียงค่าความยากของข้อสอบบริเวณปลายโค้งปกติมาตรฐานได้ ข้อสอบตำแหน่งห่างออกไปที่มีค่าความยากใกล้เคียงค่าความสามารถของผู้สอบมากที่สุดจะได้รับเลือกและจัดให้แทน ส่งผลให้มีการกระจายการใช้ข้อสอบบริเวณปลายโค้งปกติมาตรฐานเข้าสู่ด้านใน แต่เนื่องจากข้อสอบปลายโค้งปกติมาตรฐานมีการใช้สูงมาก การใช้ข้อสอบถัดเข้ามาด้านในของโค้งปกติมาตรฐานจึงสูงตามไปด้วย ข้อสอบมีอัตราการแสดงผลข้อสอบสูงเพิ่มจำนวนมากขึ้น ผลนี้จะรุนแรงมากขึ้นเมื่อจัดวางข้อสอบจำนวนน้อยเกินไปในชั้นค่าอำนาจจำแนก และอาจเป็นสาเหตุที่ไม่สามารถให้ผลการวิเคราะห์ได้ ดังนั้น การควบคุมการแสดงผลข้อสอบจึงไม่สามารถลดอัตราการทับซ้อนข้อสอบ ลดข้อสอบที่มีการแสดงผลมากเกินไป และเพิ่มข้อสอบที่มีการใช้ต่ำได้ แต่สามารถให้การแจกแจงอัตราการแสดงผลข้อสอบได้ดีกว่าวิธีไม่ควบคุมการแสดงผลข้อสอบ ดังเช่นผลการวิจัยนี้เมื่อใช้วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกร่วมกับการควบคุมการแสดงผลข้อสอบ

บทที่ 5

สรุป อภิปรายผล และข้อเสนอแนะ

การวิจัยครั้งนี้ ผู้วิจัยได้เรียงลำดับการนำเสนอจาก ความมุ่งหมายการวิจัย สมมุติฐานการวิจัย วิธีดำเนินการวิจัย ตัวแปรศึกษา เครื่องมือใช้ในการวิจัย และ การวิเคราะห์และรวบรวมข้อมูล คิดตามด้วย การสรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ ตามลำดับ ดังนี้

ความมุ่งหมายของการวิจัย

การวิจัยมีความมุ่งหมายเพื่อศึกษาประสิทธิภาพการประมาณค่าความสามารถของผู้สอบและประสิทธิภาพสมมูลการใช้ข้อสอบในคลังข้อสอบของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก เมื่อใช้ จำนวนข้อสอบในแต่ละชั้นคลังข้อสอบ การควบคุมการแสดงผลข้อสอบ ขนาดคลังข้อสอบ ความยาวแบบทดสอบ ต่างกัน โดยใช้วิธีจำลองข้อมูล

สมมุติฐานการวิจัย

1. เงื่อนไขการทดสอบ ได้แก่ จำนวนข้อสอบในแต่ละชั้นคลังข้อสอบ การควบคุมการแสดงผลข้อสอบ ขนาดคลังข้อสอบ ความยาวแบบทดสอบ ที่เหมาะสม สามารถให้ประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
2. เงื่อนไขการทดสอบ ได้แก่ จำนวนข้อสอบในแต่ละชั้นคลังข้อสอบ การควบคุมการแสดงผลข้อสอบ ขนาดคลังข้อสอบ ความยาวแบบทดสอบ ที่เหมาะสม สามารถให้ประสิทธิภาพสมมูลการใช้ข้อสอบในคลังข้อสอบของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก

วิธีดำเนินการวิจัย

วิธีดำเนินการวิจัยเริ่มจากการจำลองข้อมูลสำหรับการวิจัย ได้แก่ ค่าความสามารถจริง (θ) ของกลุ่มตัวอย่างผู้สอบ และพารามิเตอร์ของข้อสอบ นำไปทดสอบตามขั้นตอนของวิธีทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกทั้งควบคุมและไม่ควบคุมการแสดงผลข้อสอบ โดยใช้โปรแกรมภาษาคิวเบสิกที่ผู้วิจัยสร้างขึ้น รายละเอียดดังนี้

การจำลองข้อมูลสำหรับการวิจัย โดยกำหนดค่าความสามารถจริง (θ) ของกลุ่มตัวอย่างผู้สอบ สร้างจากสุ่มการแจกแจงโค้งปกติมาตรฐาน ค่าเฉลี่ยเท่ากับ 0 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1 [$\theta \sim N(0,1)$] มีค่าตั้งแต่ -3.0 ถึง 3.0 จำนวน 5,000 ค่า เมื่อเปลี่ยนวิธีกำหนดจะทำการสุ่มค่าความสามารถจริงของกลุ่มตัวอย่างผู้สอบใหม่ทุกครั้ง

กำหนดพารามิเตอร์ข้อสอบ ได้แก่ ค่าการเดา (c) คงที่เท่ากับ 0 ค่าอำนาจจำแนก (a) มี 4 ค่า คือ 0.5 1.0 1.5 และ 2.0 ตามลำดับชั้นคลังข้อสอบ 4 ชั้น ค่าความยาก (b) สร้างโดยสุ่มการแจกแจงโค้งปกติมาตรฐาน ค่าเฉลี่ยเท่ากับ 0 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1 มีค่าตั้งแต่ -3.0 ถึง 3.0 บนมาตรวัดเดียวกันกับเส้นต่อเนื่องค่าความสามารถจริงของกลุ่มตัวอย่างผู้สอบ

การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก มีขั้นตอนดังนี้ เริ่มจากแบ่งชั้นคลังข้อสอบออกเป็น 4 ชั้นสอดคล้องกับค่าอำนาจจำแนกข้อสอบคือ 0.5 1.0 1.5 และ 2.0 ชั้นแรกบรรจุด้วยข้อสอบอำนาจจำแนกต่ำสุดเท่ากับ 0.5 ชั้นต่อไปบรรจุด้วยข้อสอบอำนาจจำแนกสูงขึ้นเรื่อยๆ จนกระทั่งถึงชั้นสุดท้ายบรรจุด้วยข้อสอบอำนาจจำแนกสูงสุด จำนวนข้อสอบในชั้นคลังข้อสอบต่างกันตามวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก 4 วิธี แบ่งชั้นการทดสอบหรือความยาวแบบทดสอบตามการแบ่งชั้นของคลังข้อสอบ เลือกข้อสอบซึ่งมีความยากใกล้เคียงกับค่าความสามารถของผู้สอบมากที่สุดและจัดให้ผู้สอบ ความน่าจะเป็นในการตอบข้อสอบถูกต้องและผลการตอบข้อสอบนำไปคำนวณหาค่าประมาณความสามารถของผู้สอบ ก่อนทำการเลือกข้อสอบข้อต่อไป เมื่อการทดสอบครบตามความยาวแบบทดสอบในชั้นการทดสอบและชั้นคลังข้อสอบ การทดสอบจึงยุติ

เมื่อใช้การควบคุมการแสดงผลข้อสอบ การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก มีขั้นตอนดังนี้ เริ่มจากกำหนดค่าอัตราการแสดงผลข้อสอบสูงสุดเป้าหมายเท่ากับ 0.2 เพื่อใช้เป็นจุดตัด แบ่งชั้นการทดสอบเข้าในชั้นของคลังข้อสอบ ขั้นตอนการทดสอบทำตามอธิบายไว้ในวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก ข้อสอบแต่ละข้อนำไปหาพารามิเตอร์การแสดงผลข้อสอบในสถานการณ์จำลองล่วงหน้า ต่อจากนั้นทำการทดสอบโดยเลือกข้อสอบความยากใกล้เคียงค่าประมาณความสามารถของผู้สอบครั้งหลังสุดและจัดให้ผู้สอบ พารามิเตอร์ควบคุมการแสดงผลข้อสอบของข้อสอบข้อนั้นนำไปเปรียบเทียบกับตัวเลขสุ่มที่สร้างให้มีการแจกแจงแบบเดียวกัน $U(0,1)$ ถ้าพารามิเตอร์ควบคุมการแสดงผลข้อสอบสูงกว่าตัวเลขสุ่มจะจัดข้อสอบให้ แต่ถ้าพารามิเตอร์ควบคุมการแสดงผลข้อสอบต่ำกว่าตัวเลขสุ่มจะไม่จัดข้อสอบนั้นให้ ต้องทำการเลือกข้อสอบใหม่ที่มีค่าความยากใกล้เคียงค่าประมาณความสามารถของผู้สอบอันดับรองลงมา และพิจารณาพารามิเตอร์ควบคุมการแสดงผลข้อสอบเปรียบเทียบกับตัวเลขสุ่มค่าใหม่ก่อนจัดข้อสอบให้ ขั้นตอนนี้ต่อเนื่องจนกระทั่งสามารถหาข้อสอบจัดให้แก่ผู้สอบได้ การทดสอบดำเนินไปเรื่อยๆ จนกระทั่งครบตามความยาวแบบทดสอบทุกชั้นและทุกชั้นคลังข้อสอบ การทดสอบจึงยุติ

ตัวแปรศึกษา

ตัวแปรต้นในการวิจัยครั้งนี้มี 4 กลุ่ม ได้แก่

1. วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
4. วิธี ได้แก่

วิธีที่ 1 หรือวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกเมื่อกำหนดอัตราส่วนจำนวนข้อสอบเท่ากันทุกชั้น (1 : 1 : 1 : 1)

วิธีที่ 2 หรือวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกเมื่อกำหนดอัตราส่วนจำนวนข้อสอบชั้นแรกมากกว่าชั้นอื่น (2 : 0.5 : 0.5 : 1)

วิธีที่ 3 หรือวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกเมื่อกำหนดอัตราส่วนจำนวนข้อสอบชั้นสุดท้ายมากกว่าชั้นอื่น (1 : 0.5 : 0.5 : 2)

วิธีที่ 4 หรือวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกเมื่อกำหนดอัตราส่วนจำนวนข้อสอบชั้นแรกและชั้นสุดท้ายมากกว่าชั้นอื่น (1.5 : 0.5 : 0.5 : 1.5)

2. การควบคุมการแสดงผลข้อสอบ 2 ระดับ ได้แก่ ไม่ควบคุมการแสดงผลข้อสอบและควบคุมการแสดงผลข้อสอบด้วยวิธีซิมตัน-เฮดเทอร์ หรือ วิธีขยายช่วงระดับชั้น (Enhanced a-Stratified Method)

3. ขนาดคลังข้อสอบ 4 ระดับ ได้แก่ ขนาดคลังข้อสอบ 200 400 600 และ 800 ข้อ

4. ความยาวแบบทดสอบ 3 ระดับ ได้แก่ ความยาวแบบทดสอบ 20 32 และ 44 ข้อ

ตัวแปรตามสำหรับติดตามผลของตัวแปรต้น ได้แก่

ประสิทธิภาพการประมาณค่าความสามารถของผู้สอบ พิจารณาจาก ความคลาดเคลื่อนกำลังสองเฉลี่ย และความลำเอียงเฉลี่ย

ประสิทธิภาพสมมูลการใช้ข้อสอบในคลังข้อสอบ พิจารณาจาก ข้อสอบที่มีการแสดงมากเกินไป ข้อสอบที่มีการใช้ต่ำ อัตราการทับซ้อนข้อสอบ และการแจกแจงอัตราการใช้ข้อสอบ

เครื่องมือใช้ในการวิจัย

การทดสอบทำโดยการจำลองข้อมูล ใช้โปรแกรมคอมพิวเตอร์เขียนด้วยภาษาคิวเบสิก (Q-Basic) จัดกระทำข้อมูลในการทดสอบ โดยผ่านขั้นตอนดังนี้คือ ศึกษาความมุ่งหมายการวิจัยและรวบรวมข้อมูลที่เกี่ยวข้อง เขียนผังขั้นตอนการทดสอบ ออกแบบโปรแกรมภาษาคิวเบสิก (Q-Basic) ตามลำดับผังขั้นตอนการทดสอบ ทดสอบปรับปรุงโปรแกรมกับข้อมูลที่รู้ค่าผลลัพธ์และทดสอบซ้ำจนแน่ใจว่าการทำงานของโปรแกรมถูกต้องและผลการทดสอบแม่นยำเพียงพอจึงนำไปใช้ทดสอบ

การวิเคราะห์และรวบรวมข้อมูล

โปรแกรมภาษาคิวเบสิกที่พัฒนาขึ้นจะทำการวิเคราะห์และรวบรวมข้อมูลตัวแปรตามเมื่อผ่านการทดสอบในแต่ละชั้นของคลังข้อสอบ และภายหลังการทดสอบเสร็จสิ้นของแต่ละวิธีกำหนดตรวจสอบความสมบูรณ์ครบถ้วนก่อนนำข้อมูลไปดำเนินการต่อไป

สรุปผลการวิจัย

ผู้วิจัยได้สรุปผลการวิจัยแบ่งเป็น 2 ตอน ได้แก่

ตอนที่ 1 : ประสิทธิภาพการประมาณค่าความสามารถของผู้สอบ

ตอนที่ 2 : ประสิทธิภาพสมมูลการใช้ข้อสอบในคลังข้อสอบ

1. ประสิทธิภาพการประมาณค่าความสามารถของผู้สอบ

วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นค่าอำนาจจำแนก 4 วิธี ตามเงื่อนไข ขนาดคลังข้อสอบ ความยาวแบบทดสอบ การควบคุมการแสดงข้อสอบ ให้ความคลาดเคลื่อนกำลังสองเฉลี่ยและความลำเอียงเฉลี่ย ผลสรุปดังนี้

วิธีที่ 1 (a-STR1) พบว่า ให้ความคลาดเคลื่อนกำลังสองเฉลี่ยลดลงเมื่อเพิ่มขนาดคลังข้อสอบ และเพิ่มความยาวแบบทดสอบ ใช้ได้ดีกับขนาดคลังข้อสอบ 400 ข้อขึ้นไป ความลำเอียงเฉลี่ยลดลงเมื่อใช้ความยาวแบบทดสอบ 44 ข้อตามขนาดคลังข้อสอบที่เพิ่มขึ้น การควบคุมการแสดงข้อสอบให้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยและความลำเอียงเฉลี่ยน้อยกว่าขณะไม่ได้ควบคุมการแสดงข้อสอบเมื่อใช้ขนาดคลังข้อสอบเท่ากับหรือมากกว่า 400 ข้อขึ้นไป และความยาวแบบทดสอบเท่ากับหรือมากกว่า 32 ข้อขึ้นไป

วิธีที่ 2 (a-STR2) พบว่า ให้ความคลาดเคลื่อนกำลังสองเฉลี่ยลดลง เมื่อเพิ่มขนาดคลังข้อสอบ และเพิ่มความยาวแบบทดสอบ ใช้ได้ดีกับขนาดคลังข้อสอบ 400 ข้อขึ้นไป แต่การเพิ่มขึ้นของขนาดคลังข้อสอบและความยาวแบบทดสอบไม่มีผลต่อความลำเอียงเฉลี่ย การควบคุมการแสดงข้อสอบให้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยและความลำเอียงเฉลี่ยน้อยกว่าขณะไม่ได้ควบคุมการแสดงข้อสอบ เมื่อใช้ขนาดคลังข้อสอบเท่ากับหรือมากกว่า 600 ข้อขึ้นไป และความยาวแบบทดสอบเท่ากับหรือมากกว่า 32 ข้อขึ้นไป

วิธีที่ 3 (a-STR3) พบว่า ให้ความคลาดเคลื่อนกำลังสองเฉลี่ยลดลง เมื่อเพิ่มขนาดคลังข้อสอบ และเพิ่มความยาวแบบทดสอบ ใช้ได้ดีกับขนาดคลังข้อสอบ 400 ข้อขึ้นไป แต่การเพิ่มขึ้นของขนาดคลังข้อสอบและความยาวแบบทดสอบไม่มีผลต่อความลำเอียงเฉลี่ย การควบคุมการแสดงข้อสอบ

ให้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยและความลำเอียงเฉลี่ยน้อยกว่าขณะไม่ได้ควบคุมการแสดงผลข้อสอบ เมื่อใช้ขนาดคลังข้อสอบเท่ากับหรือมากกว่า 400 ข้อขึ้นไป และความยาวแบบทดสอบเท่ากับหรือมากกว่า 32 ข้อขึ้นไป

วิธีที่ 4 (a-STR4) พบว่า ให้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยลดลง เมื่อเพิ่มขนาดคลังข้อสอบและความยาวแบบทดสอบ ใช้ได้ดีกับขนาดคลังข้อสอบ 400 ข้อขึ้นไป แต่การเพิ่มขึ้นของขนาดคลังข้อสอบและความยาวแบบทดสอบไม่มีผลต่อความลำเอียงเฉลี่ย การควบคุมการแสดงผลข้อสอบให้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยและความลำเอียงเฉลี่ยใกล้เคียงกับขณะไม่ได้ควบคุมการแสดงผลข้อสอบ เมื่อใช้ขนาดคลังข้อสอบเท่ากับหรือมากกว่า 600 ข้อขึ้นไป และความยาวแบบทดสอบเท่ากับหรือมากกว่า 32 ข้อขึ้นไป

เปรียบเทียบประสิทธิภาพการประมาณค่าความสามารถของผู้สอบระหว่างวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับขั้นของค่าอำนาจจำแนก 4 วิธีขณะไม่ควบคุมการแสดงผลข้อสอบ พบว่า วิธีที่ 3 (a-STR3) ให้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยน้อยที่สุด รองลงมาได้แก่ วิธีที่ 4 (a-STR4) และทุกวิธีสามารถให้ความความลำเอียงเฉลี่ยน้อยที่สุดได้เท่าเทียมกัน

เปรียบเทียบประสิทธิภาพการประมาณค่าความสามารถของผู้สอบระหว่างวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับขั้นของค่าอำนาจจำแนก 4 วิธีร่วมกับการควบคุมการแสดงผลข้อสอบ พบว่า วิธีที่ 1 (a-STR1) และวิธีที่ 3 (a-STR3) ให้ความคลาดเคลื่อนกำลังสองเฉลี่ยน้อยที่สุด ขณะที่วิธีที่ 1(a-STR1) ให้ความลำเอียงเฉลี่ยน้อยที่สุด

2. ประสิทธิภาพสมมูลการใช้ข้อสอบในคลังข้อสอบ

วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับขั้นค่าอำนาจจำแนก 4 วิธี ตามเงื่อนไข ขนาดคลังข้อสอบ ความยาวแบบทดสอบ การควบคุมการแสดงผลข้อสอบ ส่งผลต่ออัตราการทับซ้อนข้อสอบ ข้อสอบที่มีการแสดงมากเกินไป และข้อสอบที่มีการใช้ต่ำ สรุปได้ดังนี้

วิธีที่ 1 (a-STR1) พบว่า ให้อัตราการทับซ้อนข้อสอบลดลง ข้อสอบที่มีการแสดงมากเกินไปลดลง และเพิ่มจำนวนข้อสอบที่มีการใช้ต่ำเมื่อเพิ่มขนาดคลังข้อสอบ แต่การเพิ่มความยาวแบบทดสอบกลับให้อัตราการทับซ้อนข้อสอบเพิ่มขึ้น ข้อสอบที่มีการแสดงมากเกินไปเพิ่มขึ้น และข้อสอบที่มีการใช้ต่ำจำนวนลดลง การควบคุมการแสดงผลข้อสอบให้อัตราการทับซ้อนข้อสอบเพิ่มขึ้นเล็กน้อย ข้อสอบที่มีการแสดงมากเกินไปจำนวนเพิ่มขึ้น และจำนวนข้อสอบที่มีการใช้ต่ำลดลง เมื่อเปรียบเทียบกับขณะไม่ควบคุมการแสดงผลข้อสอบ

วิธีที่ 2 (a-STR2) พบว่า ให้อัตราการทับซ้อนข้อสอบลดลง ข้อสอบที่มีการแสดงมากเกินไปลดลง และเพิ่มจำนวนข้อสอบที่มีการใช้ต่ำเมื่อเพิ่มขนาดคลังข้อสอบ แต่การเพิ่มความยาวแบบทดสอบกลับให้อัตราการทับซ้อนข้อสอบเพิ่มขึ้น ข้อสอบที่มีการแสดงมากเกินไปเพิ่มขึ้น และข้อสอบที่มีการ

ใช้ต่ำจำนวนลดลง การควบคุมการแสดงข้อสอบให้อัตราการทับซ้อนข้อสอบใกล้เคียงกัน ข้อสอบที่มีการแสดงมากเกินไปจำนวนเพิ่มขึ้น และจำนวนข้อสอบที่มีการใช้ต่ำลดลง เมื่อเปรียบเทียบกับขณะไม่ควบคุมการแสดงข้อสอบ

วิธีที่ 3 (a-STR3) พบว่า ให้อัตราการทับซ้อนข้อสอบลดลง ข้อสอบที่มีการแสดงมากเกินไปลดลง และเพิ่มจำนวนข้อสอบที่มีการใช้ต่ำเมื่อเพิ่มขนาดคลังข้อสอบ แต่การเพิ่มความยาวแบบทดสอบกลับให้อัตราการทับซ้อนข้อสอบเพิ่มขึ้น ข้อสอบที่มีการแสดงมากเกินไปเพิ่มขึ้น และข้อสอบที่มีการใช้ต่ำจำนวนลดลง การควบคุมการแสดงข้อสอบให้อัตราการทับซ้อนข้อสอบใกล้เคียงกัน ข้อสอบที่มีการแสดงมากเกินไปจำนวนเพิ่มขึ้น และจำนวนข้อสอบที่มีการใช้ต่ำลดลง เมื่อเปรียบเทียบกับขณะไม่ควบคุมการแสดงข้อสอบ

วิธีที่ 4 (a-STR4) พบว่า ให้อัตราการทับซ้อนข้อสอบลดลง ข้อสอบที่มีการแสดงมากเกินไปลดลง และเพิ่มจำนวนข้อสอบที่มีการใช้ต่ำเมื่อเพิ่มขนาดคลังข้อสอบ แต่การเพิ่มความยาวแบบทดสอบกลับให้อัตราการทับซ้อนข้อสอบเพิ่มขึ้น ข้อสอบที่มีการแสดงมากเกินไปเพิ่มขึ้น และข้อสอบที่มีการใช้ต่ำจำนวนลดลง การควบคุมการแสดงข้อสอบให้อัตราการทับซ้อนข้อสอบเพิ่มขึ้นเล็กน้อย ข้อสอบที่มีการแสดงมากเกินไปจำนวนเพิ่มขึ้น และจำนวนข้อสอบที่มีการใช้ต่ำลดลง เมื่อเปรียบเทียบกับขณะไม่ควบคุม การแสดงข้อสอบ

เปรียบเทียบประสิทธิภาพสมมูลการใช้ข้อสอบในคลังข้อสอบระหว่างวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก 4 วิธีขณะไม่ควบคุมการแสดงข้อสอบพบว่า ทุกวิธีมีแนวโน้มให้อัตราการทับซ้อนข้อสอบน้อยที่สุดเท่าเทียมกัน และวิธีที่ 1 (a-STR1) ให้ข้อสอบที่มีการแสดงมากเกินไปน้อยที่สุด ข้อสอบที่มีการใช้ต่ำมากที่สุด และให้อัตราการแจกแจงการแสดงผลข้อสอบสมมูลที่สุด ผลการเปรียบเทียบนี้เหมือนกันทั้งขณะควบคุมและไม่ควบคุมการแสดงข้อสอบ

อภิปรายผล

1 ประสิทธิภาพการประมาณค่าความสามารถของผู้สอบ

การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นค่าอำนาจจำแนก เป็นวิธีการเลือกข้อสอบวิธีหนึ่งซึ่งให้ประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบและให้สมมูลการใช้ข้อสอบในคลังข้อสอบไปพร้อมกัน เนื่องจากข้อสอบอำนาจจำแนกระดับต่างๆ มีโอกาสได้รับการเลือกใช้เท่าเทียมกัน

ผู้วิจัยให้ความสนใจกับการจัดข้อสอบในชั้นคลังข้อสอบจำนวนต่างกันว่า ให้ผลต่อประสิทธิภาพการประมาณค่าความสามารถของผู้สอบและประสิทธิภาพสมมูลการใช้ข้อสอบในคลังข้อสอบอย่างไร นอกจากนี้ยังได้ศึกษาเงื่อนไข ขนาดคลังข้อสอบ ความยาวแบบทดสอบ และการควบคุมการแสดงผลข้อสอบ ระดับต่างๆ กัน จากผลสรุปทั้งหมดนำมาอภิปรายโดยแบ่งเป็นหัวข้อตามสิ่งที่น่าสนใจศึกษา ดังนี้

1.1 การจัดวางข้อสอบในชั้นคลังข้อสอบ

การเปรียบเทียบวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก 4 วิธี พบว่า วิธีที่ 3 (a-STR3) ให้ความคลาดเคลื่อนกำลังสองเฉลี่ยน้อยที่สุดและวิธีที่ 4 (a-STR4) ให้ความคลาดเคลื่อนกำลังสองเฉลี่ยน้อยที่สุดอันดับรองลงมา และทุกวิธีสามารถให้ความความลำเอียงเฉลี่ยน้อยที่สุดได้เท่าเทียมกัน

วิธีที่ 3 (a-STR3) และ วิธีที่ 4 (a-STR4) ดังกล่าวเป็นวิธีที่มีการจัดวางข้อสอบอำนาจจำแนกสูงจำนวนมากที่ระดับชั้นค่าอำนาจจำแนก 2.0 พบว่า การทดสอบด้วยข้อสอบอำนาจจำแนกสูงและเลือกข้อสอบความยากใกล้เคียงค่าประมาณความสามารถของผู้สอบ สามารถลดความแตกต่างระหว่างค่าประมาณความสามารถและค่าความสามารถจริงของผู้สอบได้มากที่สุด ผลการวิจัยนี้เป็นไปตามการศึกษาของ ลุง ชาง และฮัว (Leung; Chang; & Hau. 2002) ซึ่งใช้ข้อสอบอำนาจจำแนกสูงที่ระยะสุดท้ายของการทดสอบ เมื่อค่าประมาณความสามารถเข้าใกล้ค่าความสามารถจริงของผู้สอบ การใช้ข้อสอบอำนาจจำแนกสูงจะให้สารสนเทศข้อสอบสูงสุด แต่การจัดข้อสอบอำนาจจำแนกสูงจำนวนมากที่ชั้นสุดท้ายของคลังข้อสอบกลับให้ผลดียิ่งขึ้น ช่วยให้แบบทดสอบมีโอกาสเลือกข้อสอบอำนาจจำแนกสูงที่ดีที่สุดและจัดให้ผู้สอบ สามารถลดความคลาดเคลื่อนกำลังสองเฉลี่ยและให้ประสิทธิภาพการประมาณค่าความสามารถของผู้สอบได้มากกว่าวิธีที่ 1(a-STR1) ซึ่งเป็นรูปแบบเดิมที่เคยมีผู้ศึกษามา

ในทางกลับข้าม วิธีที่มีข้อสอบจำนวนน้อยในชั้นค่าอำนาจจำแนก ได้แก่ ชั้นค่าอำนาจจำแนก 1.0 และ 1.5 ของวิธีที่ 2 (a-STR2) วิธีที่ 3 (a-STR3) และวิธีที่ 4 (a-STR4) ข้อสอบจำนวนน้อยในชั้น

ค่าอำนาจจำแนกมีผลทำให้แบบทดสอบมีโอกาสเลือกข้อสอบค่าความยากใกล้เคียงค่าความสามารถของผู้สอบได้น้อย ประสิทธิภาพในการลดความคลาดเคลื่อนกำลังสองเฉลี่ยและความลำเอียงเฉลี่ยย่อมลดลงตามไปด้วย สามารถสังเกตได้โดยเปรียบเทียบทั้ง 3 วิธีดังกล่าวกับวิธีที่ 1 (a-STR1) ซึ่งมีอัตราส่วนจำนวนข้อสอบที่ขึ้นค่าอำนาจจำแนก 1.0 และ 1.5 มากกว่าวิธีอื่น

พิจารณาวิธีที่ 2 (a-STR2) ซึ่งมีการจัดวางข้อสอบอำนาจจำแนกค่าจำนวนมากที่ขึ้นค่าอำนาจจำแนก 0.5 สามารถให้ความคลาดเคลื่อนกำลังสองเฉลี่ยและความลำเอียงเฉลี่ยน้อยกว่าวิธีอื่น แต่เมื่อสิ้นสุดการทดสอบกลับไม่สามารถลดความคลาดเคลื่อนกำลังสองเฉลี่ยและความลำเอียงเฉลี่ยได้เท่ากับวิธีที่ 1 (a-STR1)

จากที่กล่าวมาสรุปได้ว่า การจัดข้อสอบจำนวนมากในชั้นคลังข้อสอบสามารถลดความคลาดเคลื่อนกำลังสองเฉลี่ยและความลำเอียงเฉลี่ยได้มากกว่าการจัดข้อสอบจำนวนน้อย และการจัดวางข้อสอบอำนาจจำแนกสูงจำนวนมากที่ชั้นสุดท้ายของคลังข้อสอบ มีผลต่อการลดความคลาดเคลื่อนกำลังสองเฉลี่ยและความลำเอียงเฉลี่ยสูงสุดเมื่อสิ้นสุดการทดสอบ การวิจัยครั้งนี้ ผู้วิจัยมีความเห็นว่าการใช้ข้อสอบจำนวนมากกว่า 100 ข้อขึ้นไปในแต่ละชั้นคลังข้อสอบ เป็นจำนวนเพียงพอที่จะให้ประสิทธิภาพการประมาณค่าความสามารถของผู้สอบสำหรับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับขั้นของค่าอำนาจจำแนก

1.2 ขนาดคลังข้อสอบ

ผลการวิจัยแสดงถึงการเพิ่มขนาดคลังข้อสอบสามารถลดความคลาดเคลื่อนกำลังสองเฉลี่ยและความลำเอียงเฉลี่ย ให้ประสิทธิภาพการประมาณค่าความสามารถของผู้สอบได้ดียิ่งขึ้น เหตุผลมาจากคลังข้อสอบขนาดใหญ่เพิ่มโอกาสให้แบบทดสอบสามารถเลือกข้อสอบที่ดีที่สุดได้ดีกว่าคลังข้อสอบขนาดเล็ก ผลดังกล่าวสอดคล้องกับการศึกษาของ ซิงและแฮมเบิลตัน (Xing; & Hambleton, 2004) ซึ่งรายงานถึงการทดสอบด้วยคลังข้อสอบขนาดใหญ่ทำให้สารสนเทศแบบทดสอบเพิ่มขึ้นตามไปด้วย

การวิจัยนี้ ผู้วิจัยมีความเห็นว่า ขนาดคลังข้อสอบที่ควรใช้กับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับขั้นของค่าอำนาจจำแนกต้องมีขนาดมากกว่า 400 ข้อขึ้นไป แต่มีข้อแม้ว่าจำเป็นต้องพิจารณาเงื่อนไขการทดสอบอื่นๆ เพื่อใช้เป็นข้อมูลประกอบในการเลือกขนาดคลังข้อสอบด้วย เนื่องจากเงื่อนไขการทดสอบมีปฏิสัมพันธ์ซึ่งกันและกัน การเปลี่ยนแปลงเงื่อนไขใดๆ จำเป็นต้องทำการศึกษาผลที่เกิดขึ้นใหม่ทุกครั้งของการเปลี่ยนแปลง

1.3 ความยาวแบบทดสอบ

การวิจัยแสดงให้เห็นว่า การเพิ่มความยาวแบบทดสอบสามารถลดความคลาดเคลื่อนกำลังสองเฉลี่ยและความลำเอียงเฉลี่ย ให้ประสิทธิภาพการประมาณค่าความสามารถของผู้สอบได้ดียิ่งขึ้น เนื่องจากการเพิ่มความยาวแบบทดสอบจะให้สารสนเทศแบบทดสอบเพิ่มตามไปด้วยภายหลังจากการตอบข้อสอบแต่ละข้อ ความคลาดเคลื่อนของการประมาณค่าความสามารถของผู้สอบจะลดลง ให้ค่าประมาณความสามารถเข้าใกล้ค่าความสามารถจริงของผู้สอบมากที่สุด (Hambleton; & Swaminathan 1985 : 101-124) ผลการวิจัยสอดคล้องกับการศึกษาของ ลุง ชาง และฮัว (Leung; Chang; & Hau. 2003 : Online) ซึ่งรายงานว่า ความยาวแบบทดสอบเพิ่มขึ้น มีผลทำให้ความสัมพันธ์ของค่าประมาณความสามารถและค่าความสามารถจริงของผู้สอบสูงขึ้น ผลตามมามีคือ ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยและค่าความลำเอียงเฉลี่ยจะลดลง

การใช้ความยาวแบบทดสอบ 20 32 และ 44 ข้อในการวิจัยนี้ สามารถลดความคลาดเคลื่อนกำลังสองเฉลี่ยและความลำเอียงเฉลี่ยตามความยาวแบบทดสอบที่เพิ่มขึ้น นับว่าความยาวแบบทดสอบ 44 ข้อไม่มากเกินไปที่จะทำให้สารสนเทศแบบทดสอบมีลักษณะโค้งแบนตามรายงานของทิสเซน (Thissen. 1990 : 169) ดังนั้น การใช้ความยาวแบบทดสอบ 44 ข้อจะไม่ลดประสิทธิภาพการประมาณค่าความสามารถของผู้สอบและยังสามารถใช้ได้กับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของคำอ่านจำแนก

1.4 การควบคุมการแสดงข้อสอบ

การควบคุมการแสดงข้อสอบซิมตัน-เฮเตอร์ หรือวิธีขยายช่วงระดับชั้นนี้ มีคุณลักษณะให้การแจกแจงอัตราการแสดงข้อสอบมีความสมดุลมากขึ้นและอยู่ในอัตราการแสดงข้อสอบกำหนดขณะเดียวกันยังสามารถลดความคลาดเคลื่อนกำลังสองเฉลี่ยและความลำเอียงเฉลี่ยได้ (Leung; Chang; & Hau. 2002) การวิจัยครั้งนี้ให้ผลสอดคล้องกับรายงานตามที่เคยมีผู้ศึกษามา กล่าวคือทั้งวิธีที่ 1 (a-STR1) วิธีที่ 2 (a-STR2) และวิธีที่ 3 (a-STR3) เมื่อใช้ร่วมกับการควบคุมการแสดงข้อสอบจะให้ความคลาดเคลื่อนกำลังสองเฉลี่ยและความลำเอียงเฉลี่ยลดลงมากกว่าขณะไม่ควบคุมการแสดงข้อสอบถึงแม้ว่าวิธีที่ 4 (a-STR4) จะให้ผลใกล้เคียงเท่านั้น

การเปรียบเทียบวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของคำอ่านจำแนก 4 วิธีร่วมกับการควบคุมการแสดงข้อสอบพบว่า วิธีที่ 3 (a-STR3) และวิธีที่ 1 (a-STR1) มีแนวโน้มให้ความคลาดเคลื่อนกำลังสองเฉลี่ยและความลำเอียงเฉลี่ยน้อยที่สุด ผลที่ได้นี้อาจเป็นผลจากลักษณะการจัดวางข้อสอบในชั้นคลังข้อสอบร่วมกับการควบคุมการแสดงข้อสอบของวิธีดังกล่าว

อย่างไรก็ตาม การลดลงของความคลาดเคลื่อนกำลังสองเฉลี่ยและความลำเอียงเฉลี่ยบางครั้งไม่เป็นไปตามขนาดคลังข้อสอบและความยาวแบบทดสอบที่เพิ่มขึ้น ผลนี้เกิดขึ้นทั้งขณะควบคุมและไม่ควบคุมการแสดงผลข้อสอบ โดยเฉพาะอย่างยิ่ง ค่าความลำเอียงเฉลี่ยซึ่งมักพบค่ามากน้อยไม่แน่นอน ผู้วิจัยจึงได้ทำการศึกษาเพิ่มเติมในด้านโครงสร้างการทดสอบและขั้นตอนการทดสอบ อธิบายได้ดังนี้

ปัจจัยสำคัญที่ส่งผลต่อประสิทธิภาพการประมาณค่าความสามารถของผู้สอบ ได้แก่ การจัดวางข้อสอบในชั้นคลังข้อสอบให้ครอบคลุมค่าความยาก -3.0 ถึง +3.0 การวิจัยครั้งนี้กำหนดพารามิเตอร์ค่าความยากของข้อสอบสร้างโดยสุ่มการแจกแจงโค้งปกติมาตรฐาน ค่าเฉลี่ยเท่ากับ 0 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1 ความถี่ของค่าความยากข้อสอบจึงหนาแน่นอยู่บริเวณตรงกลางของโค้งปกติมาตรฐาน ค่าความยากข้อสอบซึ่งมีค่ามากหรือน้อยมีจำนวนน้อยและอยู่บริเวณปลายโค้งปกติมาตรฐาน เมื่อทำการสุ่มค่าความยากข้อสอบเข้าในแต่ละชั้นคลังข้อสอบด้วยโปรแกรม SPSS version 11 พบว่าค่าความยากของข้อสอบบริเวณปลายโค้งปกติมาตรฐานไม่สามารถกระจายครอบคลุมตั้งแต่ -3.0 ถึง +3.0 ได้ เช่น มีค่าตั้งแต่ -2.7 ถึง 2.6 เป็นต้น นอกจากนี้ การกระจายตัวของค่าความยากข้อสอบจะหายเป็นช่วงๆ ไม่สม่ำเสมอบนเส้นต่อเนื่องค่าความสามารถของผู้สอบ อ้างได้ว่าไม่มีข้อสอบความยากบริเวณดังกล่าวสำหรับจัดให้แก่ผู้สอบที่มีค่าความสามารถบนตำแหน่งเดียวกัน ดังนั้น การแจกแจงโค้งปกติมาตรฐานนี้ไม่รับรองว่าจะมีค่าความยากข้อสอบกระจายสม่ำเสมออยู่บนเส้นต่อเนื่องค่าความสามารถตั้งแต่ -3.0 ถึง +3.0

การทดสอบด้วยกลุ่มตัวอย่างขนาดใหญ่ ได้แก่ ค่าความสามารถจริงของกลุ่มตัวอย่างผู้สอบ 5,000 ค่า ค่าความสามารถจริงของผู้สอบตำแหน่งเดียวกันกับค่าความยากข้อสอบบริเวณปลายโค้งปกติมีจำนวนมาก โอกาสที่ค่าความสามารถจริงของผู้สอบเหล่านี้จะเลือกใช้ข้อสอบความยากตำแหน่งปลายโค้งปกติจึงมีสูง ถ้าค่าความยากของข้อสอบกระจายไม่สม่ำเสมอบนเส้นต่อเนื่องค่าความสามารถ ผู้สอบจะได้รับข้อสอบที่มีความยากห่างไกลจากค่าความสามารถของตนเองจัดให้แทน ทำให้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยและค่าความลำเอียงเฉลี่ยมีขนาดใหญ่ และเมื่อการทดสอบกำหนดให้ผู้สอบไม่ได้รับข้อสอบซ้ำข้อเดิม แบบทดสอบจะทำการเลือกข้อสอบความยากลำดับรองลงมาซึ่งไกลจากค่าความสามารถของผู้สอบออกไปอีกจัดให้แทน จากเหตุผลนี้ ทำให้เพิ่มค่าความคลาดเคลื่อนกำลังสองเฉลี่ยและค่าความลำเอียงเฉลี่ยที่มีขนาดใหญ่อยู่แล้วให้มีขนาดใหญ่มากขึ้น และสามารถทำให้ความลำเอียงเฉลี่ยมีค่ามากน้อยไม่แน่นอนได้

ปรากฏการณ์ดังกล่าว อ้างได้กับผลการวิเคราะห์ที่ให้ความคลาดเคลื่อนกำลังสองเฉลี่ยและความลำเอียงเฉลี่ยสูงในชั้นค่าอำนาจจำแนก 0.5 หรือระยะเริ่มต้นการทดสอบ และชั้นค่าอำนาจจำแนกซึ่งบรรจุด้วยข้อสอบจำนวนน้อย ดังนั้น การจัดข้อสอบในชั้นค่าอำนาจจำแนกให้ครอบคลุมทุกระดับบนเส้นต่อเนื่องค่าความสามารถของผู้สอบ นับเป็นหลักสำคัญสำหรับการทดสอบแบบปรับเหมาะ ช่วยให้ความคลาดเคลื่อนกำลังสองเฉลี่ยและความลำเอียงเฉลี่ยลดลงได้

วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกเมื่อใช้ร่วมกับการควบคุมการแสดงผลข้อสอบ บางครั้งผลการวิเคราะห์ทำให้ความคลาดเคลื่อนกำลังสองเฉลี่ยและความลำเอียงเฉลี่ยสูงกว่าขณะที่ไม่ควบคุมการแสดงผลข้อสอบ อธิบายได้ดังนี้

การหาพารามิเตอร์การแสดงผลข้อสอบเพื่อใช้ในการควบคุมการแสดงผลข้อสอบ พบว่า มีการเลือกใช้ข้อสอบบริเวณปลายโค้งปกติมาตรฐานในแต่ละชั้นค่าอำนาจจำแนกบ่อยครั้งเกินไป ทำให้ข้อสอบมีค่าพารามิเตอร์การแสดงผลข้อสอบต่ำกว่า 1 จำนวนมาก คิดเป็นจำนวนข้อสอบประมาณ 8-10 เปอร์เซนต์ของจำนวนข้อสอบทั้งหมดในคลังข้อสอบ นอกจากนี้ยังพบว่า ข้อสอบได้รับการเลือกใช้บ่อยเกินไปจนกระทั่งพารามิเตอร์การแสดงผลข้อสอบมีค่าหลังจุดทศนิยมหลายตำแหน่ง ข้อสอบเหล่านี้เมื่อเข้าสู่ขั้นตอนการควบคุมการแสดงผลข้อสอบโอกาสได้รับการเลือกใช้น้อยมาก แบบทดสอบจะทำการเลือกข้อสอบความยากใกล้เคียงค่าความสามารถของผู้สอบรองลงมาและจัดให้แทน กระบวนการใช้ข้อสอบไปยังข้อสอบข้างเคียงมากขึ้น ค่าความยากข้อสอบซึ่งอยู่ห่างจากค่าความสามารถของผู้สอบเป็นผลทำให้ความคลาดเคลื่อนกำลังสองเฉลี่ยและความลำเอียงเฉลี่ยมีขนาดใหญ่ ผลนี้จะรุนแรงมากขึ้นถ้ามีข้อสอบจำนวนน้อยสำหรับการทดสอบ ดังนั้นผลของการควบคุมการแสดงผลข้อสอบจะสังเกตไม่ชัดเจนที่ขนาดคลังข้อสอบเล็กหรือความยาวแบบทดสอบมากเกินไป อาจเป็นสาเหตุทำให้ไม่สามารถให้ผลการวิเคราะห์ได้ เมื่อเพิ่มขนาดคลังข้อสอบเป็น 600 ข้อขึ้นไป ความยาวแบบทดสอบไม่มากนักสามารถสังเกตได้ว่า วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกทั้ง 4 วิธีร่วมกับการควบคุมการแสดงผลข้อสอบ ให้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยและความลำเอียงเฉลี่ยมีแนวโน้มลดลงกว่าขณะที่ไม่ควบคุมการแสดงผลข้อสอบ

เมื่อจัดข้อสอบให้มีมากเพียงพอในแต่ละชั้นคลังข้อสอบ หรือการเพิ่มขนาดคลังข้อสอบจำนวนข้อสอบที่เพิ่มขึ้นในแต่ละชั้นค่าอำนาจจำแนกจะทำให้แบบทดสอบมีโอกาสในการเลือกข้อสอบความยากใกล้เคียงค่าความสามารถของผู้สอบมากขึ้น อีกประการหนึ่งคือ การเพิ่มความยาวแบบทดสอบจะทำให้ความคลาดเคลื่อนกำลังสองเฉลี่ยและค่าความลำเอียงเฉลี่ยค่อยๆ ลดลง ค่าประมาณความสามารถจะเข้าใกล้ค่าความสามารถจริงของผู้สอบมากที่สุด แต่ต้องระวังเมื่อเพิ่มความยาวแบบทดสอบมากเกินไป ทำให้ต้องใช้ข้อสอบจำนวนมากในการทดสอบ จนขาดแคลนข้อสอบความยากที่เหมาะสมกับค่าความสามารถของผู้สอบ แบบทดสอบอาจทำการเลือกข้อสอบความยากใกล้เคียงค่าความสามารถของผู้สอบข้อถัดไปและจัดให้แทน มีผลต่อการเพิ่มความคลาดเคลื่อนกำลังสองเฉลี่ยและความลำเอียงเฉลี่ยตามมา

การวิจัยนี้ขอเสนอการใช้ขนาดคลังข้อสอบมากกว่า 600 ข้อขึ้นไป และความยาวแบบทดสอบ 32 ข้อขึ้นไป เมื่อต้องใช้การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกร่วมกับการควบคุมการแสดงผลข้อสอบ

2. ประสิทธิภาพสมมูลการใช้ข้อสอบในคลังข้อสอบ

2.1 การจัดวางข้อสอบในชั้นคลังข้อสอบ

เปรียบเทียบวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก 4 วิธี พบว่า ทั้ง 4 วิธีสามารถให้อัตราการทับซ้อนข้อสอบน้อยที่สุดได้เท่าเทียมกัน และวิธีที่ 1 (a-STR1) ให้ข้อสอบที่มีการใช้มากที่สุดไปน้อยที่สุด ให้ข้อสอบที่มีการใช้ต่ำมากที่สุด และการแจกแจงอัตราการแสดงข้อสอบสมมูลที่สุด การเปรียบเทียบนี้ให้ผลเหมือนกันทั้งขณะควบคุมและไม่ควบคุมการแสดงผลข้อสอบ สามารถอธิบายได้ตามลักษณะการจัดวางข้อสอบจำนวนต่างๆ กันตามระดับชั้นค่าอำนาจจำแนก ดังนี้

ชั้นที่มีอัตราส่วนข้อสอบจำนวนน้อย ได้แก่ ชั้นค่าอำนาจจำแนก 1.0 และ 1.5 ของวิธีที่ 2 (a-STR2) วิธีที่ 3 (a-STR3) และวิธีที่ 4 (a-STR4) จะให้อัตราการทับซ้อนข้อสอบสูง ข้อสอบที่มีการแสดงมากเกินไปจำนวนมากขึ้น และข้อสอบที่มีการใช้ต่ำจำนวนลดลง ผลการวิเคราะห์เป็นเช่นนี้เนื่องจาก ข้อสอบจำนวนน้อยในชั้นค่าอำนาจจำแนก 1.0 และ 1.5 ทำให้แบบทดสอบมีโอกาสเลือกข้อสอบความยากใกล้เคียงค่าความสามารถของผู้สอบได้ลดลง ข้อสอบแต่ละข้อที่มีอยู่เข้าไปใช้บ่อยเกินไป เมื่อพิจารณาเปรียบเทียบกับวิธีที่ 1 (a-STR1) พบว่า วิธีนี้สามารถให้อัตราการทับซ้อนข้อสอบเท่าเทียมกันทุกระดับชั้นค่าอำนาจจำแนก กระจายการใช้ข้อสอบไปยังข้อสอบอื่นในชั้นอย่างทั่วถึง มีผลต่อการลดข้อสอบที่มีการแสดงมากเกินไป และให้ข้อสอบที่มีการใช้ต่ำจำนวนมาก การแจกแจงอัตราการแสดงข้อสอบในคลังข้อสอบมีความสมมูล ดังนั้นวิธีที่ 1 (a-STR1) จึงสามารถให้ประสิทธิภาพสมมูลการใช้ข้อสอบในคลังข้อสอบได้ดีกว่าเมื่อเปรียบเทียบกับวิธีอื่น

ชั้นที่มีอัตราส่วนข้อสอบจำนวนมาก ได้แก่ ชั้นที่ 1 ของวิธีที่ 2 (a-STR2) และวิธีที่ 4 (a-STR4) ชั้นที่ 4 ของวิธีที่ 3 (a-STR3) และวิธีที่ 4 (a-STR4) จะให้อัตราการทับซ้อนข้อสอบลดลง ข้อสอบที่มีการแสดงมากเกินไปจำนวนน้อย และข้อสอบที่มีการใช้ต่ำจำนวนมาก เป็นเช่นนี้เนื่องจากแบบทดสอบมีข้อสอบให้เลือกใช้จำนวนมาก โอกาสที่จะเลือกข้อสอบความยากใกล้เคียงค่าความสามารถของผู้สอบมีมากขึ้น และกระจายการแสดงผลข้อสอบไปยังข้อสอบอื่นได้ดี ให้สมมูลการใช้ข้อสอบอย่างทั่วถึง แต่วิธีที่ 2 (a-STR2) วิธีที่ 3 (a-STR3) และวิธีที่ 4 (a-STR4) มีข้อสอบจำนวนน้อยที่ชั้นค่าอำนาจจำแนก 1.0 และ 1.5 เพราะฉะนั้น การให้อัตราการทับซ้อนข้อสอบ ข้อสอบที่มีการแสดงมากเกินไป และข้อสอบที่มีการใช้ต่ำ และสมมูลการแจกแจงอัตราการแสดงข้อสอบ เมื่อสิ้นสุดการทดสอบจะไม่ดีเท่ากับวิธีที่ 1 (a-STR1)

จากที่กล่าวมาสรุปได้ว่า การจัดข้อสอบจำนวนมากในแต่ละชั้นคลังข้อสอบสามารถให้อัตราการทับซ้อนข้อสอบลดลง ข้อสอบที่มีการแสดงมากเกินไปจำนวนน้อย และข้อสอบที่มีการใช้ต่ำจำนวนมาก และให้สมมูลการแจกแจงอัตราการแสดงข้อสอบ ได้มากกว่าการจัดข้อสอบจำนวนน้อย

การวิจัยครั้งนี้ เสนอให้ใช้ข้อสอบจำนวนมากว่า 100 ข้อขึ้นไปในแต่ละชั้นคลังข้อสอบ เพื่อให้ได้ประสิทธิภาพสมมูลการใช้ข้อสอบในคลังข้อสอบสำหรับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกต่อไป

2.2 ขนาดคลังข้อสอบ

ผลการวิจัยแสดงการเพิ่มขนาดคลังข้อสอบให้อัตราการทับซ้อนข้อสอบลดลง ข้อสอบที่มีการแสดงมากเกินไปจำนวนลดลง เพิ่มจำนวนข้อสอบที่มีการใช้ต่ำ และให้สมมูลการแจกแจงอัตราการแสดงข้อสอบ เป็นเช่นนี้เนื่องจากขนาดคลังข้อสอบเพิ่มขึ้น แบบทดสอบมีข้อสอบให้เลือกใช้จำนวนมาก ทำให้การกระจายการใช้ข้อสอบไปยังข้อสอบอื่นๆ ในคลังข้อสอบอย่างทั่วถึง สามารถลดอัตราการแสดงข้อสอบ ในทางตรงข้าม ถ้าใช้คลังข้อสอบขนาดเล็กเกินไป จะให้อัตราการทับซ้อนข้อสอบเพิ่มขึ้น ข้อสอบที่มีการแสดงมากเกินไปจำนวนมากขึ้น ข้อสอบที่มีการใช้ต่ำจำนวนลดลง และลดสมมูลการแจกแจงอัตราการแสดงข้อสอบ

ในทางปฏิบัติการสร้างคลังข้อสอบขนาดใหญ่ทำได้ยาก โดยเฉพาะข้อสอบอำนาจจำแนกสูง ดังนั้น การพิจารณาเลือกใช้ขนาดคลังข้อสอบต่างๆ สามารถศึกษาเงื่อนไขการทดสอบอื่นประกอบการตัดสินใจ เลือกขนาดคลังข้อสอบไม่ใหญ่เกินไปแต่ยังคงรักษาประสิทธิภาพสมมูลการใช้ข้อสอบในคลังข้อสอบไว้ได้

2.3 ความยาวแบบทดสอบ

การวิจัยแสดงการเพิ่มความยาวแบบทดสอบจะให้อัตราการทับซ้อนข้อสอบเพิ่มขึ้น ข้อสอบที่มีการแสดงมากเกินไปจำนวนมาก และจำนวนข้อสอบที่มีการใช้ต่ำจำนวนลดลง และลดสมมูลการแจกแจงอัตราการแสดงข้อสอบ ผลนี้อ้างได้ถึงการใช้ข้อสอบในคลังข้อสอบเพิ่มขึ้น และข้อสอบอาจมีไม่เพียงพอสำหรับการทดสอบ ทำให้การแสดงข้อสอบสูงเกินอัตราการแสดงข้อสอบกำหนดในขณะเดียวกัน เมื่อใช้ความยาวแบบทดสอบน้อยเกินไปจะให้อัตราการทดสอบแบบทดสอบลดลงลดประสิทธิภาพการประมาณค่าความสามารถของผู้สอบ

ดังนั้นความยาวแบบทดสอบจึงเป็นปัจจัยสำคัญมีผลต่อประสิทธิภาพการประมาณค่าความสามารถของผู้สอบและประสิทธิภาพสมมูลการใช้ข้อสอบในคลังข้อสอบ การเลือกใช้ความยาวแบบทดสอบควรพิจารณาเงื่อนไขต่างๆ ในการทดสอบประกอบการตัดสินใจ และทำการทดสอบเบื้องต้นก่อนนำไปใช้ในการปฏิบัติต่อไป

2.4 การควบคุมการแสดงผลข้อสอบ

ผลการวิจัยแสดงว่า การควบคุมการแสดงผลข้อสอบสามารถให้อัตราการทั้บซ้อนข้อสอบสูงขึ้น ข้อสอบที่มีการแสดงผลมากเกินไปจำนวนมากขึ้น ข้อสอบที่มีการใช้ต่ำจำนวนลดลง แต่สามารถให้การแจ้แจงอัตรการแสดงผลข้อสอบได้ดี นอกจากนี้ บางครั้งไม่สามารถให้ผลการวิเคราะห์ได้ ผลการวิเคราะห์จึงไม่เป็นไปตามรายงานการศึกษาของ ลุง ชาง และฮัว (Leung; Chang; & Hau. 2002) ที่อ้างถึงวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก เมื่อใช้ร่วมกับการควบคุมการแสดงผลข้อสอบ สามารถให้สมมูลการใช้ข้อสอบในคลังข้อสอบได้ดีกว่าขณะไม่ใช้การควบคุมการแสดงผลข้อสอบ

จากผลการวิเคราะห์นี้ ผู้วิจัยได้ทำการศึกษาข้อมูลการกระจายค่าพารามิเตอร์การแสดงผลข้อสอบ และการควบคุมการแสดงผลข้อสอบในแต่ละชั้นคลังข้อสอบ ด้วยการควบคุมโปรแกรมการทดสอบให้แสดงผลทุกขั้นตอนการทดสอบ สามารถอธิบายเหตุผลดังนี้

การทดสอบของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก ขณะไม่ได้ใช้การควบคุมการแสดงผลข้อสอบพบว่า การทดสอบมีการเลือกใช้ข้อสอบบริเวณปลายโค้งปกติมาตรฐานสูงมากเกินไปอัตราการแสดงผลข้อสอบกำหนด สามารถพบได้บ่อยในการทดสอบเมื่อใช้คลังข้อสอบขนาดเล็ก และเมื่อแต่ละชั้นคลังข้อสอบบรรจุค่าความยากของข้อสอบลักษณะการกระจายไม่สม่ำเสมอบนเส้นต่อเนื่องค่าความสามารถของผู้สอบ โดยเฉพาะบริเวณปลายโค้งปกติมาตรฐาน เมื่อนำการควบคุมการแสดงผลข้อสอบเข้ามาใช้ ขบวนการควบคุมการแสดงผลข้อสอบจะทำการเลือกข้อสอบจาก 1. ข้อสอบความยากใกล้เคียงค่าประมาณความสามารถของผู้สอบ 2. ข้อสอบข้อใหม่ที่ผู้สอบคนนั้นไม่เคยได้ทำมาก่อน และ 3. พารามิเตอร์การแสดงผลข้อสอบสูงกว่าตัวเลขสุ่ม ผลการทดสอบพบว่า พารามิเตอร์การแสดงผลข้อสอบของข้อสอบบริเวณปลายโค้งปกติมาตรฐานมีค่าอยู่ในระดับต่ำมาก จนกระทั่งแบบทดสอบไม่สามารถเลือกข้อสอบบริเวณนี้จัดให้แก่ผู้สอบที่มีค่าความสามารถใกล้เคียงค่าความยากบริเวณปลายโค้งปกติมาตรฐานได้ ข้อสอบห้างออกไปที่มีค่าความยากใกล้เคียงค่าความสามารถของผู้สอบมากที่สุดจะได้รับเลือกและจัดให้แทน ส่งผลให้มีการกระจายการใช้ข้อสอบบริเวณปลายโค้งปกติมาตรฐานเข้าสู่ด้านใน แต่เนื่องจากข้อสอบปลายโค้งปกติมาตรฐานมีการใช้สูงมาก การใช้ข้อสอบถัดเข้ามาด้านในของโค้งปกติมาตรฐานจึงสูงตามไปด้วย ข้อสอบมีอัตราการแสดงผลข้อสอบสูงเพิ่มจำนวนมากขึ้น ผลนี้จะรุนแรงขึ้นเมื่อจัดวางข้อสอบจำนวนน้อยเกินไปในชั้นค่าอำนาจจำแนก และอาจเป็นสาเหตุที่ไม่สามารถให้ผลการทดสอบได้

ดังนั้น การควบคุมการแสดงผลข้อสอบจึงไม่สามารถลดอัตรการทั้บซ้อนข้อสอบ ลดข้อสอบที่มีการแสดงผลมากเกินไป และเพิ่มข้อสอบที่มีการใช้ต่ำได้ แต่สามารถให้การแจ้แจงอัตรการแสดงผลข้อสอบได้ดีกว่าขณะไม่ควบคุมการแสดงผลข้อสอบ ดังเช่นผลการวิจัยนี้เมื่อใช้วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกร่วมกับการควบคุมการแสดงผลข้อสอบ

การกระจายค่าพารามิเตอร์ข้อสอบในแต่ละชั้นคลังข้อสอบสามารถจัดกระทำได้ การวิจัยนี้ได้เพิ่มการทดสอบด้วยขนาดคลังข้อสอบ 800 ข้อ พบว่า สามารถทำการทดสอบด้วยการควบคุมการแสดงผลข้อสอบได้ทุกระดับความยาวแบบทดสอบ อย่างไรก็ตาม การแก้ไขด้วยวิธีนี้ยังไม่เป็นการแก้ปัญหาโดยตรง สิ่งสำคัญคือ การจัดวางข้อสอบในชั้นคลังข้อสอบให้ครอบคลุมค่าความยาก -3.0 ถึง 3.0 และจัดข้อสอบจำนวนเพียงพอทุกระดับค่าความสามารถของผู้สอบบนเส้นต่อเนื่อง เป็นปัจจัยสำคัญในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกร่วมกับการควบคุมการแสดงผลข้อสอบ เพื่อให้ได้ประสิทธิภาพสมมูลการใช้ข้อสอบในคลังข้อสอบสูงสุด

การวิจัยสรุปได้ว่า การศึกษาประสิทธิภาพการประมาณค่าความสามารถของผู้สอบและประสิทธิภาพสมมูลการใช้ข้อสอบในคลังข้อสอบของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก เมื่อใช้ จำนวนข้อสอบในแต่ละชั้นคลังข้อสอบ การควบคุมการแสดงผลข้อสอบ ขนาดคลังข้อสอบ ความยาวแบบทดสอบ ต่างกัน พบว่า วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกวิธีที่ 3 (a-STR3) เมื่อกำหนดอัตราส่วนจำนวนข้อสอบชั้นสุดท้ายมากกว่าชั้นอื่นให้ประสิทธิภาพการประมาณค่าความสามารถของผู้สอบ โดยให้ความคลาดเคลื่อนกำลังสองเฉลี่ยน้อยที่สุด แต่ไม่สามารถให้ประสิทธิภาพสมมูลการใช้ข้อสอบในคลังข้อสอบได้ดีพอ ขณะที่วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกวิธีที่ 1(a-STR1) เมื่อกำหนดอัตราส่วนจำนวนข้อสอบกันทุกชั้น สามารถให้ประสิทธิภาพการประมาณค่าความสามารถของผู้สอบในระดับยอมรับได้ และให้ประสิทธิภาพสมมูลการใช้ข้อสอบในคลังข้อสอบดีที่สุด โดยให้ข้อสอบที่มีการแสดงมากเกินไปน้อยที่สุด ข้อสอบที่มีการใช้ต่ำมากที่สุด และการแจกแจงการแสดงผลข้อสอบสมมูลที่สุด ทั้งขณะควบคุมและไม่ควบคุมการแสดงผลข้อสอบ และให้ประสิทธิภาพเมื่อใช้กับขนาดคลังข้อสอบ 400 ข้อขึ้นไป และความยาวข้อสอบ 32 ข้อขึ้นไป

ดังนั้น วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกวิธีที่ 1(a-STR1) เป็นวิธีที่ควรนำมาพิจารณาศึกษาเพิ่มเติมและเลือกใช้ในทางปฏิบัติ ในขณะเดียวกัน ควรต้องพิจารณาเงื่อนไขอื่นๆ เพื่อจัดวางโครงสร้างการทดสอบให้สามารถส่งเสริมประสิทธิภาพการประมาณค่าความสามารถของผู้สอบและประสิทธิภาพสมมูลการใช้ข้อสอบในคลังข้อสอบต่อไป

ข้อเสนอแนะ

1. ข้อเสนอแนะในการนำผลการวิจัยไปใช้

การศึกษาประสิทธิภาพการประมาณค่าความสามารถของผู้สอบ และประสิทธิภาพสมมูลการใช้ข้อสอบในคลังข้อสอบ ของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นค่าอำนาจจำแนก เงื่อนไขต่างๆ ในการทดสอบเป็นสิ่งสำคัญในการเพิ่มหรือลดประสิทธิภาพดังกล่าว การวิจัยนี้

ผู้วิจัยให้ความสนใจการจัดจำนวนข้อสอบในแต่ละชั้นของการทดสอบ ขนาดคลังข้อสอบ ความยาวแบบทดสอบ และการควบคุมการแสดงข้อสอบ พบสิ่งที่น่าสนใจสำหรับการนำผลการวิจัยไปใช้ ดังนี้

1.1 การจัดจำนวนข้อสอบในชั้นค่าอำนาจจำแนก

การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนกเมื่อจัดวางจำนวนข้อสอบในชั้นคลังข้อสอบแตกต่างกันมีผลให้ประสิทธิภาพการทดสอบแตกต่างกัน ได้แก่ การจัดจำนวนข้อสอบชั้นสุดท้ายมากกว่าชั้นอื่น สามารถให้ประสิทธิภาพการประมาณค่าความสามารถของผู้สอบ ในขณะเดียวกันก็สามารถลดประสิทธิภาพสมมูลการใช้ข้อสอบในคลังข้อสอบเนื่องจากชั้นที่ 2 และชั้นที่ 3 มีอัตราส่วนจำนวนข้อสอบน้อย วิธีนี้จึงมีข้อจำกัดสำหรับการพิจารณาเลือกใช้

วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นค่าอำนาจจำแนกเมื่อกำหนดอัตราส่วนจำนวนข้อสอบเท่ากันทุกชั้น ให้ประสิทธิภาพการประมาณค่าความสามารถของผู้สอบ ขณะเดียวกันยังสามารถให้ประสิทธิภาพสมมูลการใช้ข้อสอบในคลังข้อสอบไปพร้อมกัน เมื่อใช้กับขนาดคลังข้อสอบ 400 ข้อ ความยาวแบบทดสอบ 32 ข้อขึ้นไป ทั้งขณะควบคุมและไม่ควบคุมการแสดงผลข้อสอบ ดังนั้น วิธีนี้จึงเป็นทางเลือกที่ดีในการนำไปใช้ในทางปฏิบัติต่อไป

หลักสำคัญของการจัดวางจำนวนข้อสอบในชั้นค่าอำนาจจำแนก ได้แก่ การจัดวางข้อสอบให้มีจำนวนมากเพียงพอในแต่ละชั้นคลังข้อสอบ การวิจัยนี้เสนอการใช้จำนวนข้อสอบในแต่ละชั้นคลังข้อสอบเท่ากับหรือมากกว่า 100 ข้อขึ้นไป แต่ควรพิจารณาเงื่อนไขอื่นๆ ในการทดสอบด้วย เช่น ขนาดคลังข้อสอบ ความยาวแบบทดสอบ การควบคุมการแสดงผลข้อสอบ จำนวนกลุ่มตัวอย่างผู้สอบ เป็นต้น ดังนั้น จึงควรศึกษาความสัมพันธ์ของเงื่อนไขต่างๆ ก่อนจัดโครงสร้างการทดสอบ

1.2 ขนาดคลังข้อสอบ

การเพิ่มประสิทธิภาพการประมาณค่าความสามารถของผู้สอบและประสิทธิภาพสมมูลการใช้ข้อสอบในคลังข้อสอบ สามารถกระทำโดยการใช้ขนาดคลังข้อสอบเพิ่มขึ้น ถ้าการทดสอบใช้คลังข้อสอบขนาดเล็กและความยาวแบบทดสอบมากเกินไป อาจส่งผลต่อการลดประสิทธิภาพการประมาณค่าความสามารถของผู้สอบและประสิทธิภาพสมมูลการใช้ข้อสอบในคลังข้อสอบ แก้ไขได้ด้วยการปรับขนาดคลังข้อสอบให้มีขนาดใหญ่ขึ้น ผลจากการวิจัยครั้งนี้ แนะนำการใช้ขนาดคลังข้อสอบสำหรับวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นค่าอำนาจจำแนกควรใช้ขนาดคลังข้อสอบเท่ากับหรือมากกว่า 400 ข้อขึ้นไป และข้อสอบในแต่ละชั้นค่าอำนาจจำแนกควรเท่ากับหรือมากกว่า 100 ข้อขึ้นไป อย่างไรก็ตาม การนำไปใช้ควรพิจารณาเงื่อนไขการทดสอบอื่นๆ ประกอบการตัดสินใจด้วย

1.3 ความยาวแบบทดสอบ

การเพิ่มประสิทธิภาพการประมาณค่าความสามารถของผู้สอบและประสิทธิภาพสมมูลการใช้ข้อสอบในคลังข้อสอบ สามารถกระทำโดยความยาวแบบทดสอบเพิ่มขึ้น แต่ต้องติดตามการเพิ่มขึ้นของความยาวแบบทดสอบมากเกินไป อาจมีผลทำให้ประสิทธิภาพสมมูลการใช้ข้อสอบในคลังข้อสอบลดลง จากการเพิ่มขึ้นของอัตราการทับซ้อนข้อสอบ ข้อสอบที่มีการแสดงมากเกินไปจำนวนมากขึ้น ข้อสอบที่มีการใช้ต่ำจำนวนน้อยลง และลดสมมูลการแจกแจงอัตราการแสดงข้อสอบ ในการวิจัยครั้งนี้ ผู้วิจัยมีความเห็นว่า ความยาวแบบทดสอบเท่ากับ 32 และ 44 ข้อ สามารถให้ประสิทธิภาพการประมาณค่าความสามารถของผู้สอบและประสิทธิภาพสมมูลการใช้ข้อสอบในคลังข้อสอบ อย่างไรก็ตาม การเลือกใช้ความยาวแบบทดสอบควรคำนึงถึงขนาดคลังข้อสอบเป็นหลัก รวมทั้งเงื่อนไขการทดสอบอื่นๆ ด้วย

1.4 การควบคุมการแสดงข้อสอบ

การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นค่าอำนาจจำแนกร่วมกับการควบคุมการแสดงข้อสอบ สามารถเพิ่มประสิทธิภาพการประมาณค่าความสามารถของผู้สอบและประสิทธิภาพสมมูลการใช้ข้อสอบในคลังข้อสอบ เมื่อขนาดคลังข้อสอบและความยาวข้อสอบมากเพียงพอ ขนาดคลังข้อสอบ 200 ข้อ และความยาวแบบทดสอบ 44 ข้อ อาจพบปัญหาให้ผลการทดสอบไม่ได้เมื่อใช้ร่วมกับการควบคุมการแสดงข้อสอบ ดังนั้น ควรจัดข้อสอบเข้าในแต่ละชั้นค่าอำนาจจำแนกไม่ต่ำกว่า 100 ข้อและใช้ขนาดคลังข้อสอบ 400 ข้อขึ้นไป จะสามารถทำการทดสอบร่วมกับการควบคุมการแสดงข้อสอบและยังสามารถใช้ความยาวแบบทดสอบ 44 ข้อได้

2. ข้อเสนอแนะในการวิจัยครั้งต่อไป

จากการวิจัยครั้งนี้ วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นค่าอำนาจจำแนกยังพบปัจจัยสำคัญอื่นที่คาดว่าจะมีผลต่อประสิทธิภาพการประมาณค่าความสามารถของผู้สอบและประสิทธิภาพสมมูลการใช้ข้อสอบในคลังข้อสอบ ซึ่งสามารถศึกษาได้โดยวิธีจำลองข้อมูล ได้แก่ การจัดวางข้อสอบในแต่ละชั้นคลังข้อสอบให้ครอบคลุมค่าความยากและมีจำนวนข้อสอบเพียงพอทุกตำแหน่งค่าความสามารถของผู้สอบบนเส้นต่อเนื่อง เพื่อให้แบบทดสอบมีข้อสอบให้เลือกใช้ได้ทุกระดับค่าความสามารถของผู้สอบ ดังนั้น รูปแบบการจัดวางข้อสอบในแต่ละชั้นค่าอำนาจจำแนกเป็นสิ่งที่สำคัญที่ควรศึกษาเพิ่มเติมเพื่อให้การทดสอบมีประสิทธิภาพสูงสุด ก่อนที่จะทำการออกแบบโครงสร้างการทดสอบของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นค่าอำนาจจำแนกอื่นๆ ต่อไป

บรรณานุกรม

บรรณานุกรม

- Ackerman, Terry A. (1991, March). The Use of Unidimensional Parameter Estimates of Multidimensional Items in Adaptive Testing. *Applied Psychological Measurement*. 15(1) : 13-24.
- American Psychological Association. (1985). *Standards for Educational and Psychological Testing*. Washington, DC.
- Baker, Frank B. (1992). Item Characteristic Curve : Dichotomous Response. in *Item Response Theory : Parameter Estimation Technique*. pp. 1-23. New York: Marcel Dekker.
- Ban, Jae Chun; & et al. (2001, Fall). A Comparative Study of On-Line Pretest Item-Calibration /Scaling Methods in Computerized Adaptive Testing. *Journal of Educational Measurement*. 38(3) : 191-212.
- Ban, Jae-Chun; & et al. (2002, Fall). Data Sparseness and On-line Pretest Item Calibration-Scaling Methods in CAT. *Journal of Educational Measurement*. 39(3) : 207-218.
- Bergstrom, Betty A.; & Lunz, Mary E. (1999). CAT for Certification and Licensure. in *Innovations in Computerized Assessment*. by Fritz Drasgow and Julie B. Olson-Buchnan. pp. 67-91. New Jersey : Lawrence Erlbaum.
- Blais, Jean-Guy; & Raiche, Gilles. (2002, April). *Some Features of The Sampling Distribution of The Ability Estimate in Computerized Adaptive Testing according to Two Stopping Rules*. (Online). Available: <http://www.Umoncton.ca/raicheg/recherche/IOMW/Text1.htm>. Retrieved January 10, 2004.
- Boyd, Aimee Michelle. (2003, May). *Strategies for Controlling Testlet Exposure Rates in Computerized Adaptive Testing Systems*. (Online). Available: <http://www.ic2.org/publications/testlet%20Exposure%20Rates.pdf>. Retrieved January 10, 2004.
- Chalhoub-Deville, Micheline; Alcaya, Cheryl; & Lozier Vashti McCollum. (1996, July). *An Operational Framework for Constructing a Computer-Adaptive Test of L2 Reading Ability : Theoretical and Practical Issues*. (Online). Available: <http://www.carla.umn.edu/resources/working-papers/samples/CAT.pdf>. Retrieved January 10, 2004.

- Chang, Hua-Hua; Qian, Jiahe; & Ying, Zhiliang. (2001, December). a-Stratified Multistage Computerized Adaptive Testing with b Blocking. *Applied Psychological Measurement*. 25(4) : 333-341.
- Chang, Hua-Hua; & van der Linden, Wim J. (2003, July). Optimal Stratification of Item Pools in a-Stratified Computerized Adaptive Testing. *Applied Psychological Measurement*. 27(4) : 262-274.
- Chang, Hua-Hua; & Ying, Zhiliang. (1996, September). A Global Information Approach to Computerized Adaptive Testing. *Applied Psychological Measurement*. 20(3) : 213-229.
- Chang, Hua-Hua; & Ying, Zhiliang. (1999, September). a-Stratified Multistage Computerized Adaptive Testing. *Applied Psychological Measurement*. 23(3) : 211-222.
- Chang, Yuan-chin Ivan; & Ying Zhiliang. (2001, July 13). *Sequential Estimation in Variable Length Computerized Adaptive Testing*. (Online). Available: http://www.stat.sinica.edu.tw/library/c_tee_rep/c_2001_10.pdf. Retrieved January 10, 2004.
- Chen Shu-Ying; Ankenmann Robert D.; & Chang Hua-Hua. (2000, September). A Comparison of Item Selection Rules of the Early Stages of Computerized Adaptive Testing. *Applied Psychological Measurement*. 24(3) : 241-255.
- Cheng, Philip E.; & Liou, Michelle. (2000, September). Estimation of Trait Level in Computerized Adaptive Testing. *Applied Psychological Measurement*. 24(3) : 257-263.
- Cheng, Philip E.; & Liou, Michelle. (2003, May). Computerized Adaptive Testing Using the Nearest-Neighbors Criterion. *Applied Psychological Measurement*. 27(3) : 204-216.
- Davis, Laurie Laughlin; & Dodd, Barbara G. (2003, September). Item Exposure Constraints for Testlets in the Verbal Reasoning Section of the MCAT. *Applied Psychological Measurement*. 27(5) : 335-356.
- De Ayala, R.J.; Dodd, Barbara G.; & Koch, William R. (1990, Fall). A Simulation and Comparison of Flexilevel and Bayesian Computerized Adaptive Testing. *Journal of Educational Measurement*. 27(3) : 227-239.
- Deng, Hui. (2001). *a-Stratified Computerized Adaptive Testing with Unequal Item Exposure Across Strata*. (Online). Available: <http://www.uiowa.edu/~fyi/issues2001-v39/07052002/defenses.html>. Retrieved January 10, 2004.

- Deng, Hui; & Ansley, Tim. (2003, April). *To Stratify or Not : An Investigation of CAT Item Selection Procedures Under Practical Constraints*. (Online). Available:
<http://www.ets.org/research/dload/ncme03-deng2.pdf>. Retrieved January 10, 2004.
- Dodd Barbara G.; De Ayala R.J.; & Koch William R. (1995, March). Computerized Adaptive Testing with Polytomous Items. *Applied Psychological Measurement*. 19(1) : 5-22.
- Eggen, T.J.H.M. (1999, September). Item Selection in Adaptive with the Sequential Probability Ratio Test. *Applied Psychological Measurement*. 23(3) : 249-261.
- Eggen, T.J.H.M. (2001, January). *Overexposure and Underexposure of Items in Computerized Adaptive Testing*. (Online). Available:
<http://www.download.citogroep.nl/pub/pok/reports/Report01-01.pdf>. Retrieved January 10, 2004.
- Eggen, T.J.H.M.; & Straetmans, G.J.J.M. (2000, October). Computerized Adaptive Testing for Classifying Examinees into Three Categories. *Educational and Psychological Measurement*. 60(5) : 713-734.
- Embretson, Susan E.; & Reise, Steven P. (2000). *Introduction*. in *Item Response Theory for Psychologists*. pp.3-9. New Jersey : Lawrence Erlbaum.
- Embretson, Susan E.; & Reise, Steven P. (2000). *Calibrating Items: Estimation*. in *Item Response Theory for Psychologists*. pp. 187-225. New Jersey : Lawrence Erlbaum.
- Fan, Meichu; & Zhu Daming. (2001). *A Further Study on Adjusting CAT Item Selection Starting Point for Individual Examinees*. (Online). Available:
<http://www.edtech.connect.msu.edu/Searchaera2002/viewproposaltext.htm>. Retrieved January 10, 2004.
- Fischer, Gerhard H.; & Molenaar, Ivo W. (1995). Some Background for Item Response Theory and the Rasch Model. in *Rasch Model Foundation, Recent Developments, and Applications*. pp. 3-14. New York : Springer - Verlag.
- Flaughner, Ronald. (1990). Item Pools. in *Computerized Adaptive Testing : A Primer*. by Howard Wainer & et al. pp. 41-63. New Jersey : Lawrence Erlbaum.
- Gialluca, Kathleen A.; & Weiss, David J. (1979). Efficiency of a Adaptive Inter-Subtest Branching Strategy in *The Measurement of Classroom Achievement*. Research Report 79-6. Minneapolis : University of Minnesota, Department of Psychology Psychometric Methods Program.

- Green, Bert F.; & et al. (1984, Winter). Technical Guidelines for Assessing Computerized Adaptive Testings. *Journal of Educational Measurement*. 21(4) : 347-360.
- Hambleton, Ronald K.; Swaminathan, Hariharan. (1985). *Item Response Theory : Principles and Applications*. USA: Kluwer Nijhoff.
- Hambleton, Ronald K.; Swaminathan, Hariharan; & Rogers, H. Jane. (1991). *Fundamentals of Item Response Theory*. USA: SAGE.
- Harwell, Michael R. (1997, April). Analyzing The Results of Monte Carlo Studies in Item Response Theory. *Educational and Psychological Measurement*. 57(2) : 266-279.
- Harwell, Michael; & et al. (1996, June). Monte Carlo Studies in Item Response Theory. *Applied Psychological Measurement*. 20(2) : 101-125.
- Harwell, Michael R.; & Janosky, J.E. (1991). An Empirical Study of the Effect of Small Datasets and Varying Prior Variances on Item Parameter Estimation in BILOG. *Applied Psychological Measurement*. 15 : 279-291.
- Hau, Kit-Tai; & Chang, Hua-Hua. (2001, Fall). Item Selection in Computerized Adaptive Testing : Should More Discriminating Items be Used First ?. *Journal of Educational Measurement*. 38(3) : 249-266.
- Hulin, Charles L.; Drasgow, Fritz; & Parsons, Charles K. (1983). *Item Response Theory : Application to Psychological Measurement*. Illinois: Dow Jones-Irwin.
- Jacobs, Lucy Cheser; & Chase, Clinton I. (1992). Computer-Assisted Testing. in *Developing and Using Tests Effectively : A Guide for Faculty*. pp. 168-177. San Francisco: Jossey – Bass.
- Leung, Chi-Keung; Chang, Hua-Hua; & Hau, Kit-Tai. (2001, March). *Making a-Stratified Computerized Adaptive Testing Design More Practical : Imposing Non-Statistical Constraints*. (On Line). Available:
<http://www.ied.edu.hk/scast/staff/math/mathslleungck.html>. Retrieved January 10, 2004.
- Leung, Chi-Keung; Chang, Hua-Hua; & Hau, Kit-Tai. (2002, December). Item Selection in Computerized Adaptive Testing : Improving the a-Stratified Design with the Symson-Hetter Algorithm. *Applied Psychological Measurement*. 26(4) : 376-392.
- Leung, Chi-Keung; Chang Hau-Hau; & Hau, Kit-Tai. (2003). *Computerized Adaptive Testing : A Comparison of Three Content Balancing Methods*. (Online). Available:
<http://www.bc.edu/research/intase/jtla/journal/pdf/v2n2-jtla.pdf>. Retrieved January 10, 2004.

- Leung, Chi-Keung; Chang, Hua-Hua; & Hau, Kit-Tai. (2003, April). Incorporation of Content Balancing Requirements in Stratification Designs for Computerized Adaptive Testing. *Educational and Psychological Measurement*. 63(2) : 257-270.
- Lord, Frederic M. (1980). *Applications of Item Response Theory to Practical Testing Problems*. New Jersey: Lawrence Erlbaum.
- Lord, Frederic M.; & Novick Melvin R. (1968). Latent Traits and Item Characteristic Functions. in *Statistical Theory of Mental Test Scores*. pp. 358-394. New York: Addison-Wesley.
- Lunz, Mary E.; Bergstrom, Betty A.; Wright Benjamin D. (1992, March). The Effect of Review on Student Ability and Test Efficiency for Computerized Adaptive Tests. *Applied Psychological Measurement*. 16(1) : 33-40.
- Lunz, Mary E., & Bergstrom, Betty A. (1994, Fall). An Empirical Study of Computerized Adaptive Test Administration Conditions. *Journal of Educational Measurement*. 31(3) : 251-263.
- McLeod, Lori Davis; & Lewis, Charles. (1999, June). Detecting Item Memorization in the CAT Environment. *Applied Psychological Measurement*. 23(2) : 147-160.
- McLeod, Lori; Lewis, Charles; & Thissen, David. (2003, March). A Bayesian Method for the Detection of Item Preknowledge in Computerized Adaptive Testing. *Applied Psychological Measurement*. 27(2) : 121-137.
- Meijer, Rob R.; & Nering, Michael L. (1999, September). Computerized Adaptive Testing : Overview and Introduction. *Applied Psychological Measurement*. 23(3) : 187-194.
- Nicewander, W. Alan; Thomasson, Gary L. (1999, September). Some Reliability Estimates for Computerized Adaptive Testing. *Applied Psychological Measurement*. 23(3) : 239-247.
- Parshall, Cynthia; Harmes, J. Christine; & Kromrey, Jeffrey D. (2000, Fall). Item Exposure Control in Computer-Adaptive Testing : The Use of Freezing to Augment Stratification. *Florida Journal of Educational Research*. 40(1) : 28-52.
- Pastor, Dena A.; Dodd, Barbara G., & Chang, Hua-Hua. (2002, June). A Comparison of Item Selection Techniques and Exposure Control Mechanisms in CATs Using the Generalized Partial Credit Model. *Applied Psychological Measurement*. 26(2) : 147-163.
- Reise, S.P.; & Yu, J. (1990). Parameter Recovery in the Graded Response Model Using MULTILOG. *Journal of Educational Measurement*. 27 : 133-144.

- Revuelta, Javier; & Ponsoda, Vicente. (1998, Winter). A Comparison of Item Exposure Control Methods in Computerized Adaptive Testing. *Journal of Educational Measurement*. 35(4) : 311-327.
- Schnipke, Deborah L.; & Green, Bert F. (1995, Fall). A Comparison of Item Selection Routines in Linear and Adaptive Tests. *Journal of Educational Measurement*. 32(3) : 227-242.
- Segall, Daniel O. (2003?) *Computerized Adaptive Testing*. (Online). Available: <http://www.segalld.home.neteom.com/CAT%20ESM.pdf>. Retrieved January 10, 2004.
- Steinberg, Lynne; Thissen, David; & Wainer, Howard. (1990). Validity. in *Computerized Adaptive Testing : A Primer*. by Howard Wainer, & et al. pp. 187-231. New Jersey: Lawrence Erlbaum.
- Stocking, Martha L. (1997, June). Revising Item Responses in Computerized Adaptive Tests : A Comparison of Three Models. *Applied Psychological Measurement*. 21(2) : 129-142.
- Stocking, Martha L.; & Lewis, Charles. (1998, Spring). Controlling Item Exposure Conditional on Ability in Computerized Adaptive Testing. *Journal of Educational and Behavioral Statistics*. 23(1) : 57-75.
- Stocking, Martha L.; & Lewis, Charles. (2000). Methods of Controlling The Exposure of Items in CAT. in *Computerized Adaptive Testing : Theory and practice*. by Wim J. van der Linden and Cees A.W. Glass. pp. 163-182. Netherlands: Kluwer Academic.
- Stocking, Martha L.; & Swanson, Len. (1993, September). A method for severely constrained item selection in adaptive testing. *Applied Psychological Measurement*. 17(3) : 277- 292.
- Stocking, Martha L.; & Swanson, Len. (1998, September). Optimal Design of Item Banks for Computerized Adaptive Tests. *Applied Psychological Measurement*. 22(3) : 271-279.
- Thissen, David. (1990). Reliability and Measurement Precision. in *Computerized Adaptive Testing : A Primer*. by Howard Wainer, & et al. pp. 161-186. New Jersey: Lawrence Erlbaum.
- Thissen, David; & Mislevy Robert J. (1990). Testing Algorithms. in *Computerized Adaptive Testing : A Primer*. by Howard Wainer, & et al. pp. 103-136. New Jersey: Lawrence Erlbaum.
- van der Linden, Wim J. (1998, June). Bayesian Item Selection Criteria for Adaptive Testing. *Psychometrika*. 63(2) : 201-216.

- van der Linden, Wim J. (1998, September). Optimal Assembly of Psychological and Educational Tests. *Applied Psychological Measurement*. 22(3) : 195-211.
- van der Linden, Wim J. (1999, March). Empirical Initialization of the Trait Estimator in Adaptive Testing. *Applied Psychological Measurement*. 23(1) : 21-29.
- van der Linden, Wim J.; & Chang, Hua-Hua. (2003, March). Implementing Content Constraints in Alpha-Stratified Adaptive Testing Using a Shadow Test Approach. *Applied Psychological Measurement*. 27(2) : 107-120.
- van Krimpen-Stoop Edith M.L.A.; & Meijer Rob R. (2001, Summer). CUSUM-Based Person-Fit Statistic for Adaptive Testing. *Journal of Educational and Behavioral Statistics*. 26(2) : 199-218.
- van Rijn, P.W.; & et al. (2002, December). Evaluation of Selection Procedures for Computerized Adaptive Testing with Polytomous Items. *Applied Psychological Measurement*. 26(4) : 393-411.
- Veerkamp, Wim J.J.; & Berger, Martijn P.F. (1997, Summer). Some New Item Selection Criteria for Adaptive Testing. *Journal of Educational and Behavioral Statistics*. 22(2) : 203-226.
- Veerkamp, Wim J.J.; & Glas, Cees A. (2000, Winter). Detection of Known Items in Adaptive Testing with a Statistical Quality Control Method. *Journal of Educational and Behavioral Statistics*. 25(4) : 373-389.
- Vispoel, Waiter P. (1999). Creating Computerized Adaptive Tests of Music Aptitude: Problems, Solutions, and Future Directions. in *Innovations in Computerized Assessment*. by Drasgow Fritz and Olson-Buchanan Julia B. pp. 151-176. New Jersey: Lawrence Erlbaum.
- Vispoel, Waiter P.; Hendrickson, Amy B.; & Bleiler, Timothy. (2000, Spring). Limiting Answer Review and Change on Computerized Adaptive Vocabulary Tests : Psychometric and Attitudinal Results. *Journal of Educational Measurement*. 37(1) : 21-38.
- Wainer, Howard. (1990). Introduction and History. in *Computerized Adaptive Testing : A Primer*. by Howard Wainer, & et al. pp. 1-21. New Jersey: Lawrence Erlbaum.
- Wainer, Howard; & Mislevy, Robert J. (1990). Item Response Theory, Item Calibration and Proficiency Estimation. in *Computerized Adaptive Testing : A Primer*. by Howard Wainer, & et al. pp. 65-102. New Jersey: Lawrence Erlbaum.
- Wang, Tianyou; Kolen, Michael L. (2001, Spring). Evaluation Comparability in Computerized Adaptive Testing: Issues, Criteria and an Example. *Journal of Educational Measurement*.

38(1) : 19-49.

- Weiss, David J. (1973). *The Stratified Adaptive Computerized Ability Test*. Research Report 73-3. Minneapolis: University of Minnesota, Department of Psychology Psychometric Methods Program.
- Weiss, David J. (1974). *Strategies of Adaptive Ability Measurement*. Research Report 74-5. Minneapolis: University of Minnesota, Department of Psychology Psychometric Methods Program.
- Weiss, David J. (1985). Adaptive Testing by Computer. *Journal of Consulting and Clinical Psychology*. 53(6) : 774-789.
- Weiss, David J.; & Kingsbury, G. Gage. (1984, Winter). Application of Computerized Adaptive Testing to Educational Problems. *Journal of Educational Measurement*. 21(4) : 361-375.
- Wen, Jian-Bing; Chang Hua-Hua; & Hua Kit-Tai. (2001). *Optimum Number of Strata in the a-Stratified Computerized Adaptive Testing Design*. (Online). Available: http://www.edtech.connect.msu.edu/Searchaera2002/viewproposal_text.asp?propID=1853. Retrieved January 10, 2004.
- Xing, Dehui; & Hambleton Ronald K. (2004, February). 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.
- Yi Qing; Wang Tianyon; & Ban Jae-Chun. (2001, Fall). Effect of Scale Transformation and Test-Termination Rule on the Precision of Ability Estimation in Computerized Adaptive Testing. *Journal of Educational Measurement*. 38(3) : 267-292.
- Zwick, R.; & Thayer, D.T. (2002). Application of an Empirical Bayes Enhancement of Mantel-Haenszel Differential Item Functioning Analysis to a Computerized Adaptive Testing. *Applied Psychological Measurement*. 26(1) : 57-76.

ภาคผนวก

ตาราง 25 การทดสอบของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจ
จำแนก ขนาดคลังข้อสอบ 200 ข้อ ความยาวแบบทดสอบ 20 ข้อ

เงื่อนไขการทดสอบ		ผลการทดสอบ					χ^2
		MSE	A.Bias	Overlap*	U.item	O.item	
ขนาดคลังข้อสอบ 200 ข้อ ความยาวข้อสอบ 20 ข้อ	a=0.5	1.646	-.801	.848	46	4	15.503
	a-STR1 a=1.0	.579	-.362	.848	47	3	
	1:1:1:1 a=1.5	.333	-.241	.816	47	3	
	a=2.0	.212	-.173	.816	48	2	
	a=0.5	1.522	-.755	.623	98	2	28.872
	a-STR2 a=1.0	.644	-.398	.912	14	11	
	2:0.5:0.5:1 a=1.5	.387	-.269	.944	14	11	
	a=2.0	.211	-.184	.880	50	0	
	a=0.5	1.758	-.830	.848	48	2	48.502
	a-STR3 a=1.0	.618	-.383	.928	18	7	
	1:0.5:0.5:2 a=1.5	.397	-.297	.944	14	11	
	a=2.0	.206	-.197	.655	100	0	
	a=0.5	1.750	-.823	.752	74	1	31.918
	a-STR4 a=1.0	.646	-.407	.944	16	9	
	1.5:0.5:0.5:1.5 a=1.5	.395	-.284	.944	14	11	
	a=2.0	.210	-.185	.752	75	0	

MSE = ความคลาดเคลื่อนกำลังสองเฉลี่ย

A.Bias = ความลำเอียงเฉลี่ย

Overlap = อัตราการทับซ้อนข้อสอบ * หน่วย $\times 10^{-6} + 1.99 \times 10^{-4}$

U.item = ข้อสอบที่มีการแสดงมากเกินไป

O.item = ข้อสอบที่มีการใช้ต่ำ

χ^2 = ค่าไคสแควร์แสดงการแจกแจงอัตราการแสดงข้อสอบ

a = ค่าอำนาจจำแนก

a-STR1 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
กำหนดอัตราส่วนจำนวนข้อสอบเท่ากันทุกชั้น

a-STR2 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
กำหนดอัตราส่วนจำนวนข้อสอบชั้นแรกมากกว่าชั้นอื่น

a-STR3 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
กำหนดอัตราส่วนจำนวนข้อสอบชั้นสุดท้ายมากกว่าชั้นอื่น

a-STR4 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
กำหนดอัตราส่วนจำนวนข้อสอบชั้นแรกและชั้นสุดท้ายมากกว่าชั้นอื่น

ตาราง 26 การทดสอบของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจ
จำแนกร่วมกับการควบคุมการแสดงผลข้อสอบ ขนาดคลังข้อสอบ 200 ข้อ ความยาวแบบทดสอบ 20 ข้อ

เงื่อนไขการทดสอบ			ผลการทดสอบ					χ^2
			MSE	A.Bias	Overlap*	U.item	O.item	
ขนาดคลังข้อสอบ 200 ข้อ ความยาวข้อสอบ 20 ข้อ	a-STR1	a=0.5	1.334	-.652	.848	45	5	17.424
		a=1.0	.531	-.259	.832	45	5	
		1:1:1:1 a=1.5	.340	-.250	.816	46	4	
		a=2.0	.212	-.176	.800	50	0	

MSE = ความคลาดเคลื่อนกำลังสองเฉลี่ย

A.Bias = ความลำเอียงเฉลี่ย

Overlap = อัตราการทับซ้อนข้อสอบ * หน่วย $\times 10^{-6} + 1.99 \times 10^{-4}$

U.item = ข้อสอบที่มีการแสดงมากเกินไป

O.item = ข้อสอบที่มีการใช้ค่า

χ^2 = ค่าไคสแควร์แสดงการแจกแจงอัตราการแสดงข้อสอบ

a = ค่าอำนาจจำแนก

a-STR1 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก

กำหนดอัตราส่วนจำนวนข้อสอบเท่ากันทุกชั้น

ตาราง 27 การทดสอบของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจ
จำแนก ขนาดคลังข้อสอบ 200 ข้อ ความยาวแบบทดสอบ 32 ข้อ

เงื่อนไขการทดสอบ		ผลการทดสอบ					χ^2
		MSE	A.Bias	Overlap*	U.item	O.item	
ขนาดคลังข้อสอบ 200 ข้อ ความยาวข้อสอบ 32 ข้อ	a=0.5	1.428	-.742	.940	46	4	23.074
	a-STR1 a=1.0	.562	-.400	.930	35	15	
	1:1:1:1 a=1.5	.295	-.254	.930	34	16	
	a=2.0	.234	-.226	.920	41	9	
	a=0.5	1.328	-.707	.770	93	7	37.967
	a-STR2 a=1.0	.591	-.383	.990	8	17	
	2:0.5:0.5:1 a=1.5	.397	-.324	.950	2	23	
	a=2.0	.195	-.191	.930	3	14	
	a=0.5	1.374	-.708	.940	43	7	68.165
	a-STR3 a=1.0	.812	-.580	.980	9	16	
	1:0.5:0.5:2 a=1.5	.352	-.278	1.00	4	21	
	a=2.0	.173	-.156	.790	99	1	
	a=0.5	1.520	-.780	.870	69	6	45.033
	a-STR4 a=1.0	.708	-.470	1.00	5	20	
	1.5:0.5:0.5:1.5 a=1.5	.361	-.291	.980	7	18	
	a=2.0	.197	-.204	.810	71	4	

MSE = ความคลาดเคลื่อนกำลังสองเฉลี่ย

A.Bias = ความลำเอียงเฉลี่ย

Overlap = อัตราการทับซ้อนข้อสอบ * หน่วย $\times 10^{-6} + 1.99 \times 10^{-4}$

U.item = ข้อสอบที่มีการแสดงมากเกินไป

O.item = ข้อสอบที่มีการใช้ต่ำ

χ^2 = ค่าไคสแควร์แสดงการแจกแจงอัตราการแสดงข้อสอบ

a = ค่าอำนาจจำแนก

a-STR1 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
กำหนดอัตราส่วนจำนวนข้อสอบเท่ากันทุกชั้น

a-STR2 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
กำหนดอัตราส่วนจำนวนข้อสอบชั้นแรกมากกว่าชั้นอื่น

a-STR3 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
กำหนดอัตราส่วนจำนวนข้อสอบชั้นสุดท้ายมากกว่าชั้นอื่น

a-STR4 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
กำหนดอัตราส่วนจำนวนข้อสอบชั้นแรกและชั้นสุดท้ายมากกว่าชั้นอื่น

ตาราง 28 การทดสอบของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจ
จำแนก ขนาดคลังข้อสอบ 200 ข้อ ความยาวแบบทดสอบ 44 ข้อ

ผลการทดสอบ		MSE	A.Bias	Overlap*	U.item	O.item	χ^2
เงื่อนไขการทดสอบ							
ขนาดคลังข้อสอบ 200 ข้อ ความยาวข้อสอบ 44 ข้อ	a=0.5	1.214	-.650	.952	37	13	24.055
	a-STR1 a=1.0	.507	-.360	.967	28	22	
	1:1:1:1 a=1.5	.297	-.270	.981	20	30	
	a=2.0	.200	-.184	.974	20	30	
	a=0.5	1.362	-.730	.843	89	11	43.789
	a-STR2 a=1.0	.578	-.385	.996	3	22	
	2:0.5:0.5:1 a=1.5	.436	-.331	.989	1	24	
	a=2.0	.195	-.155	.960	21	29	
	a=0.5	1.429	-.759	.945	38	12	84.241
	a-STR3 a=1.0	.659	-.385	.981	4	21	
	1:0.5:0.5:2 a=1.5	.427	-.302	.989	1	24	
	a=2.0	.173	-.190	.858	93	7	
	a=0.5	1.266	-.686	.894	60	15	49.991
	a-STR4 a=1.0	.624	-.430	.974	1	24	
	1.5:0.5:0.5:1.5 a=1.5	.403	-.279	.989	3	22	
	a=2.0	.172	-.193	.894	61	14	

MSE = ความคลาดเคลื่อนกำลังสองเฉลี่ย

A.Bias = ความลำเอียงเฉลี่ย

Overlap = อัตราการทับซ้อนข้อสอบ * หน่วย $\times 10^{-6} + 1.99 \times 10^{-4}$

U.item = ข้อสอบที่มีการแสดงมากเกินไป

O.item = ข้อสอบที่มีการใช้ต่ำ

χ^2 = ค่าไคสแควร์แสดงการแจกแจงอัตราการแสดงข้อสอบ

a = ค่าอำนาจจำแนก

a-STR1 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
กำหนดอัตราส่วนจำนวนข้อสอบเท่ากันทุกชั้น

a-STR2 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
กำหนดอัตราส่วนจำนวนข้อสอบชั้นแรกมากกว่าชั้นอื่น

a-STR3 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
กำหนดอัตราส่วนจำนวนข้อสอบชั้นสุดท้ายมากกว่าชั้นอื่น

a-STR4 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
กำหนดอัตราส่วนจำนวนข้อสอบชั้นแรกและชั้นสุดท้ายมากกว่าชั้นอื่น

ตาราง 29 การทดสอบของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจ
จำแนก ขนาดคลังข้อสอบ 400 ข้อ ความยาวแบบทดสอบ 20 ข้อ

เงื่อนไขการทดสอบ		ผลการทดสอบ					χ^2
		MSE	A.Bias	Overlap*	U.item	O.item	
ขนาดคลังข้อสอบ 400 ข้อ ความยาวข้อสอบ 20 ข้อ	a=0.5	1.538	-.762	.639	98	2	36.763
	a-STR1 a=1.0	.592	-.385	.591	99	1	
	1:1:1:1 a=1.5	.300	-.205	.687	99	1	
	a=2.0	.192	-.146	.639	100	0	
	a=0.5	1.567	-.777	.223	199	1	39.914
	a-STR2 a=1.0	.607	-.369	.864	43	7	
	2:0.5:0.5:1 a=1.5	.339	-.262	.832	47	3	
	a=2.0	.201	-.185	.623	100	0	
	a=0.5	1.415	-.697	.623	99	1	63.072
	a-STR3 a=1.0	.625	-.416	.800	45	5	
	1:0.5:0.5:2 a=1.5	.341	-.264	.832	41	9	
	a=2.0	.190	-.170	.287	200	0	
	a=0.5	1.526	-.751	.415	149	1	46.519
	a-STR4 a=1.0	.573	-.359	.800	48	2	
	1.5:0.5:0.5:1.5 a=1.5	.319	-.233	.784	45	5	
	a=2.0	.195	-.175	.479	150	0	

MSE = ความคลาดเคลื่อนกำลังสองเฉลี่ย

A.Bias = ความลำเอียงเฉลี่ย

Overlap = อัตราการทับซ้อนข้อสอบ * หน่วย $\times 10^{-6} + 1.99 \times 10^{-4}$

U.item = ข้อสอบที่มีการแสดงมากเกินไป

O.item = ข้อสอบที่มีการใช้ต่ำ

χ^2 = ค่าไคสแควร์แสดงการแจกแจงอัตราการแสดงข้อสอบ

a = ค่าอำนาจจำแนก

a-STR1 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
กำหนดอัตราส่วนจำนวนข้อสอบเท่ากันทุกชั้น

a-STR2 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
กำหนดอัตราส่วนจำนวนข้อสอบชั้นแรกมากกว่าชั้นอื่น

a-STR3 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
กำหนดอัตราส่วนจำนวนข้อสอบชั้นสุดท้ายมากกว่าชั้นอื่น

a-STR4 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
กำหนดอัตราส่วนจำนวนข้อสอบชั้นแรกและชั้นสุดท้ายมากกว่าชั้นอื่น

ตาราง 30 การทดสอบของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจ
จำแนกร่วมกับการควบคุมการแสดงผลข้อสอบขนาดคลังข้อสอบ 400 ข้อ ความยาวแบบทดสอบ 20 ข้อ

ผลการทดสอบ		MSE	A.Bias	Overlap*	U.item	O.item	χ^2
เงื่อนไขการทดสอบ	เงื่อนไขการทดสอบ						
ขนาดคลังข้อสอบ 400 ข้อ ความยาวข้อสอบ 20 ข้อ	a=0.5	1.185	-.589	.639	95	5	18.959
	a-STR1 a=1.0	.510	-.305	.591	98	2	
	1:1:1:1 a=1.5	.304	-.194	.623	100	0	
	a=2.0	.201	-.161	.687	100	0	
	a=0.5	1.455	-.739	.303	198	2	39.469
	a-STR2 a=1.0	.544	-.294	.832	45	5	
	2:0.5:0.5:1 a=1.5	.334	-.243	.816	46	4	
	a=2.0	.193	-.152	.591	100	0	
	a=0.5	1.248	-.630	.655	95	5	57.194
	a-STR3 a=1.0	.558	-.324	.848	44	6	
	1:0.5:0.5:2 a=1.5	.390	-.265	.896	41	9	
	a=2.0	.186	-.158	.271	200	0	
	a=0.5	1.341	-.664	.431	146	4	42.112
	a-STR4 a=1.0	.551	-.323	.800	44	6	
	1.5:0.5:0.5:1.5 a=1.5	.350	-.235	.896	44	6	
	a=2.0	.197	-.165	.399	150	0	

MSE = ความคลาดเคลื่อนกำลังสองเฉลี่ย

A.Bias = ความลำเอียงเฉลี่ย

Overlap = อัตราการทับซ้อนข้อสอบ * หน่วย $\times 10^{-6} + 1.99 \times 10^{-4}$

U.item = ข้อสอบที่มีการแสดงมากเกินไป

O.item = ข้อสอบที่มีการใช้ต่ำ

χ^2 = ค่าไคสแควร์แสดงการแจกแจงอัตราการแสดงข้อสอบ

a = ค่าอำนาจจำแนก

a-STR1 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
กำหนดอัตราส่วนจำนวนข้อสอบเท่ากันทุกชั้น

a-STR2 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
กำหนดอัตราส่วนจำนวนข้อสอบชั้นแรกมากกว่าชั้นอื่น

a-STR3 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
กำหนดอัตราส่วนจำนวนข้อสอบชั้นสุดท้ายมากกว่าชั้นอื่น

a-STR4 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
กำหนดอัตราส่วนจำนวนข้อสอบชั้นแรกและชั้นสุดท้ายมากกว่าชั้นอื่น

ตาราง 31 การทดสอบของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจ
จำแนก ขนาดคลังข้อสอบ 400 ข้อ ความยาวแบบทดสอบ 32 ข้อ

ผลการทดสอบ		MSE	A.Bias	Overlap*	U.item	O.item	χ^2
เงื่อนไขการทดสอบ	เงื่อนไขการทดสอบ						
ขนาดคลังข้อสอบ 400 ข้อ ความยาวข้อสอบ 32 ข้อ	a=0.5	1.471	-.755	.780	94	6	32.882
	a-STR1 a=1.0	.508	-.356	.759	90	10	
	1:1:1:1 a=1.5	.278	-.254	.790	98	2	
	a=2.0	.176	-.168	.749	100	0	
	a=0.5	1.513	-.793	.579	197	3	46.815
	a-STR2 a=1.0	.542	-.402	.930	37	13	
	2:0.5:0.5:1 a=1.5	.288	-.255	.930	38	12	
	a=2.0	.172	-.175	.820	98	2	
	a=0.5	1.280	-.670	.780	94	6	81.913
	a-STR3 a=1.0	.574	-.404	.900	36	14	
	1:0.5:0.5:2 a=1.5	.294	-.247	.920	35	15	
	a=2.0	.166	-.177	.569	200	0	
	a=0.5	1.341	-.700	.639	147	3	54.529
	a-STR4 a=1.0	.557	-.398	.910	37	13	
	1.5:0.5:0.5:1.5 a=1.5	.314	-.267	.930	34	16	
	a=2.0	.167	-.170	.669	150	0	

MSE = ความคลาดเคลื่อนกำลังสองเฉลี่ย

A.Bias = ความลำเอียงเฉลี่ย

Overlap = อัตราการทับซ้อนข้อสอบ * หน่วย $\times 10^{-6} + 1.99 \times 10^{-4}$

U.item = ข้อสอบที่มีการแสดงมากเกินไป

O.item = ข้อสอบที่มีการใช้ต่ำ

χ^2 = ค่าไคสแควร์แสดงการแจกแจงอัตราการแสดงข้อสอบ

a = ค่าอำนาจจำแนก

a-STR1 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
กำหนดอัตราส่วนจำนวนข้อสอบเท่ากันทุกชั้น

a-STR2 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
กำหนดอัตราส่วนจำนวนข้อสอบชั้นแรกมากกว่าชั้นอื่น

a-STR3 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
กำหนดอัตราส่วนจำนวนข้อสอบชั้นสุดท้ายมากกว่าชั้นอื่น

a-STR4 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
กำหนดอัตราส่วนจำนวนข้อสอบชั้นแรกและชั้นสุดท้ายมากกว่าชั้นอื่น

ตาราง 32 การทดสอบของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจ
จำแนกร่วมกับการควบคุมการแสดงผลข้อสอบขนาดคลังข้อสอบ 400 ข้อ ความยาวแบบทดสอบ 32 ข้อ

เงื่อนไขการทดสอบ		ผลการทดสอบ					
		MSE	A.Bias	Overlap*	U.item	O.item	χ^2
ขนาดคลังข้อสอบ 400 ข้อ ความยาวข้อสอบ 32 ข้อ	a=0.5	1.112	-.539	.820	94	6	22.137
	a-STR1 a=1.0	.456	-.298	.790	94	6	
	1:1:1:1 a=1.5	.278	-.238	.749	98	2	
	a=2.0	.170	-.168	.780	100	0	
	a=0.5	1.163	-.604	.770	93	7	73.642
	a-STR3 a=1.0	.518	-.323	.930	34	16	
	1:0.5:0.5:2 a=1.5	.327	-.218	.930	33	17	
	a=2.0	.164	-.161	.519	200	0	

MSE = ความคลาดเคลื่อนกำลังสองเฉลี่ย

A.Bias = ความลำเอียงเฉลี่ย

Overlap = อัตราการทับซ้อนข้อสอบ * หน่วย $\times 10^{-6} + 1.99 \times 10^{-4}$

U.item = ข้อสอบที่มีการแสดงมากเกินไป

O.item = ข้อสอบที่มีการใช้ต่ำ

χ^2 = ค่าไคสแควร์แสดงการแจกแจงอัตราการแสดงข้อสอบ

a = ค่าอำนาจจำแนก

a-STR1 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก

กำหนดอัตราส่วนจำนวนข้อสอบเท่ากันทุกชั้น

a-STR3 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก

กำหนดอัตราส่วนจำนวนข้อสอบชั้นสุดท้ายมากกว่าชั้นอื่น

ตาราง 33 การทดสอบของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจ
จำแนก ขนาดคลังข้อสอบ 400 ข้อ ความยาวแบบทดสอบ 44 ข้อ

ผลการทดสอบ		MSE	A.Bias	Overlap*	U.item	O.item	χ^2
เงื่อนไขการทดสอบ ขนาดคลังข้อสอบ 400 ข้อ ความยาวข้อสอบ 44 ข้อ	เงื่อนไขการทดสอบ						
	a=0.5	1.324	-.703	.843	92	8	39.618
	a-STR1 a=1.0	.449	-.353	.858	88	12	
	1:1:1:1 a=1.5	.253	-.237	.865	88	12	
	a=2.0	.168	-.180	.829	94	6	
	a=0.5	1.239	-.674	.639	192	8	48.263
	a-STR2 a=1.0	.514	-.377	.938	23	27	
	2:0.5:0.5:1 a=1.5	.289	-.258	.952	20	30	
	a=2.0	.162	-.156	.850	95	5	
	a=0.5	1.261	-.700	.858	93	7	89.602
	a-STR3 a=1.0	.479	-.352	.952	30	20	
	1:0.5:0.5:2 a=1.5	.315	-.285	.960	15	35	
	a=2.0	.140	-.163	.618	200	0	
	a=0.5	1.288	-.698	.756	144	6	66.661
	a-STR4 a=1.0	.508	-.387	.938	21	29	
	1.5:0.5:0.5:1.5 a=1.5	.288	-.270	.930	22	28	
	a=2.0	.143	-.169	.749	150	0	

MSE = ความคลาดเคลื่อนกำลังสองเฉลี่ย

A.Bias = ความลำเอียงเฉลี่ย

Overlap = อัตราการทับซ้อนข้อสอบ * หน่วย $\times 10^{-6} + 1.99 \times 10^{-4}$

U.item = ข้อสอบที่มีการแสดงมากเกินไป

O.item = ข้อสอบที่มีการใช้ต่ำ

χ^2 = ค่าไคสแควร์แสดงการแจกแจงอัตราการแสดงข้อสอบ

a = ค่าอำนาจจำแนก

a-STR1 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
กำหนดอัตราส่วนจำนวนข้อสอบเท่ากันทุกชั้น

a-STR2 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
กำหนดอัตราส่วนจำนวนข้อสอบชั้นแรกมากกว่าชั้นอื่น

a-STR3 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
กำหนดอัตราส่วนจำนวนข้อสอบชั้นสุดท้ายมากกว่าชั้นอื่น

a-STR4 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
กำหนดอัตราส่วนจำนวนข้อสอบชั้นแรกและชั้นสุดท้ายมากกว่าชั้นอื่น

ตาราง 34 การทดสอบของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจ
จำแนกร่วมกับการควบคุมการแสดงผลข้อสอบ ขนาดคลังข้อสอบ 400 ข้อ ความยาวแบบทดสอบ
44 ข้อ

		ผลการทดสอบ	MSE	A.Bias	Overlap*	U.item	O.item	χ^2
ขนาดคลังข้อสอบ 400 ข้อ	ความยาวข้อสอบ 44 ข้อ	a=0.5	1.034	-.509	.865	90	10	21.879
		a-STR1	a=1.0	.475	-.353	.843	90	
		1:1:1:1	a=1.5	.272	-.271	.872	92	
			a=2.0	.158	-.161	.850	99	

MSE = ความคลาดเคลื่อนกำลังสองเฉลี่ย

A.Bias = ความลำเอียงเฉลี่ย

Overlap = อัตราการทับซ้อนข้อสอบ * หน่วย $\times 10^{-6} + 1.99 \times 10^{-4}$

U.item = ข้อสอบที่มีการแสดงมากเกินไป

O.item = ข้อสอบที่มีการใช้ต่ำ

χ^2 = ค่าไคสแควร์แสดงการแจกแจงอัตราการแสดงข้อสอบ

a = ค่าอำนาจจำแนก

a-STR1 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก

กำหนดอัตราส่วนจำนวนข้อสอบเท่ากันทุกชั้น

ตาราง 35 การทดสอบของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจ
จำแนก ขนาดคลังข้อสอบ 600 ข้อ ความยาวแบบทดสอบ 20 ข้อ

ผลการทดสอบ		MSE	A.Bias	Overlap*	U.item	O.item	χ^2
เงื่อนไขการทดสอบ							
ขนาดคลังข้อสอบ 600 ข้อ ความยาวข้อสอบ 20 ข้อ	a=0.5	1.681	-.817	.399	148	2	39.313
	a-STR1 a=1.0	.590	-.387	.431	148	2	
	1:1:1:1 a=1.5	.307	-.249	.463	150	0	
	a=2.0	.193	-.171	.447	150	0	
	a=0.5	1.565	-.779	-.121	299	1	47.078
	a-STR2 a=1.0	.602	-.393	.768	72	3	
	2:0.5:0.5:1 a=1.5	.312	-.246	.736	75	0	
	a=2.0	.186	-.157	.415	150	0	
	a=0.5	1.571	-.764	.479	149	1	65.942
	a-STR3 a=1.0	.619	-.406	.752	71	4	
	1:0.5:0.5:2 a=1.5	.317	-.241	.784	75	0	
	a=2.0	.185	-.166	-.144	300	0	
	a=0.5	1.552	-.765	.143	224	1	56.972
	a-STR4 a=1.0	.546	-.345	.719	73	2	
	1.5:0.5:0.5:1.5 a=1.5	.317	-.242	.736	74	1	
	a=2.0	.187	-.151	.111	225	0	

MSE = ความคลาดเคลื่อนกำลังสองเฉลี่ย

A.Bias = ความลำเอียงเฉลี่ย

Overlap = อัตราการทับซ้อนข้อสอบ * หน่วย $\times 10^{-6} + 1.99 \times 10^{-4}$

U.item = ข้อสอบที่มีการแสดงมากเกินไป

O.item = ข้อสอบที่มีการใช้ต่ำ

χ^2 = ค่าไคสแควร์แสดงการแจกแจงอัตราการแสดงข้อสอบ

a = ค่าอำนาจจำแนก

a-STR1 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
กำหนดอัตราส่วนจำนวนข้อสอบเท่ากันทุกชั้น

a-STR2 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
กำหนดอัตราส่วนจำนวนข้อสอบชั้นแรกมากกว่าชั้นอื่น

a-STR3 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
กำหนดอัตราส่วนจำนวนข้อสอบชั้นสุดท้ายมากกว่าชั้นอื่น

a-STR4 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
กำหนดอัตราส่วนจำนวนข้อสอบชั้นแรกและชั้นสุดท้ายมากกว่าชั้นอื่น

ตาราง 36 การทดสอบของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจ
จำแนกร่วมกับการควบคุมการแสดงผลข้อสอบ ขนาดคลังข้อสอบ 600 ข้อ ความยาวแบบทดสอบ
20 ข้อ

เงื่อนไขการทดสอบ		ผลการทดสอบ					χ^2		
		MSE	A.Bias	Overlap*	U.item	O.item			
ขนาดคลังข้อสอบ 600 ข้อ ความยาวข้อสอบ 20 ข้อ	a-STR1	a=0.5	1.538	-.760	.479	146	4	20.541	
		a=1.0	.552	-.353	.479	148	2		
		1:1:1:1	a=1.5	.291	-.218	.431	150		0
		a=2.0	.181	-.165	.495	150	0		
	a-STR2	a=0.5	1.502	-.758	-.225	298	2	36.633	
		a=1.0	.578	-.361	.719	70	5		
		2:0.5:0.5:1	a=1.5	.306	-.226	.768	75		0
		a=2.0	.188	-.155	.463	150	0		
	a-STR3	a=0.5	1.383	-.697	.399	147	3	51.643	
		a=1.0	.569	-.357	.719	70	5		
		1:0.5:0.5:2	a=1.5	.323	-.248	.768	75		0
		a=2.0	.195	-.178	-.160	300	0		
	a-STR4	a=0.5	1.420	-.716	.127	222	3	41.163	
		a=1.0	.575	-.355	.752	71	4		
		1.5:0.5:0.5:1.5	a=1.5	.327	-.258	.768	75		0
		a=2.0	.192	-.173	.063	225	0		

MSE = ความคลาดเคลื่อนกำลังสองเฉลี่ย

A.Bias = ความลำเอียงเฉลี่ย

Overlap = อัตราการทับซ้อนข้อสอบ * หน่วย $\times 10^{-6} + 1.99 \times 10^{-4}$

U.item = ข้อสอบที่มีการแสดงมากเกินไป

O.item = ข้อสอบที่มีการใช้ต่ำ

χ^2 = ค่าไคสแควร์แสดงการแจกแจงอัตราการศึกษาข้อสอบ

a = ค่าอำนาจจำแนก

a-STR1 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
กำหนดอัตราส่วนจำนวนข้อสอบเท่ากันทุกชั้น

a-STR2 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
กำหนดอัตราส่วนจำนวนข้อสอบชั้นแรกมากกว่าชั้นอื่น

a-STR3 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
กำหนดอัตราส่วนจำนวนข้อสอบชั้นสุดท้ายมากกว่าชั้นอื่น

a-STR4 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
กำหนดอัตราส่วนจำนวนข้อสอบชั้นแรกและชั้นสุดท้ายมากกว่าชั้นอื่น

ตาราง 37 การทดสอบของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจ
จำแนก ขนาดคลังข้อสอบ 600 ข้อ ความยาวแบบทดสอบ 32 ข้อ

ผลการทดสอบ		MSE	A.Bias	Overlap*	U.item	O.item	χ^2
เงื่อนไขการทดสอบ							
ขนาดคลังข้อสอบ 600 ข้อ ความยาวข้อสอบ 32 ข้อ	a=0.5	1.382	-.716	.619	146	4	42.528
	a-STR1 a=1.0	.511	-.378	.649	149	1	
	1:1:1:1 a=1.5	.264	-.248	.659	150	0	
	a=2.0	.164	-.169	.649	150	0	
	a=0.5	1.351	-.710	.259	299	1	57.283
	a-STR2 a=1.0	.572	-.434	.850	68	7	
	2:0.5:0.5:1 a=1.5	.282	-.259	.890	70	5	
	a=2.0	.168	-.176	.749	150	0	
	a=0.5	1.474	-.752	.679	146	4	85.551
	a-STR3 a=1.0	.540	-.401	.900	70	5	
	1:0.5:0.5:2 a=1.5	.273	-.241	.870	73	2	
	a=2.0	.154	-.179	.309	300	0	
	a=0.5	1.471	-.752	.449	222	3	65.760
	a-STR4 a=1.0	.529	-.393	.840	69	6	
	1.5:0.5:0.5:1.5 a=1.5	.281	-.256	.860	71	4	
	a=2.0	.159	-.161	.369	225	0	

MSE = ความคลาดเคลื่อนกำลังสองเฉลี่ย

A.Bias = ความลำเอียงเฉลี่ย

Overlap = อัตราการทับซ้อนข้อสอบ * หน่วย $\times 10^{-6} + 1.99 \times 10^{-4}$

U.item = ข้อสอบที่มีการแสดงมากเกินไป

O.item = ข้อสอบที่มีการใช้ค่า

χ^2 = ค่าไคสแควร์แสดงการแจกแจงอัตราการแสดงข้อสอบ

a = ค่าอำนาจจำแนก

a-STR1 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
กำหนดอัตราส่วนจำนวนข้อสอบเท่ากันทุกชั้น

a-STR2 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
กำหนดอัตราส่วนจำนวนข้อสอบชั้นแรกมากกว่าชั้นอื่น

a-STR3 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
กำหนดอัตราส่วนจำนวนข้อสอบชั้นสุดท้ายมากกว่าชั้นอื่น

a-STR4 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
กำหนดอัตราส่วนจำนวนข้อสอบชั้นแรกและชั้นสุดท้ายมากกว่าชั้นอื่น

ตาราง 38 การทดสอบของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจ
 จำแนกร่วมกับการควบคุมการแสดงผลข้อสอบ ขนาดคลังข้อสอบ 600 ข้อ ความยาวแบบทดสอบ
 32 ข้อ

เงื่อนไขการทดสอบ		ผลการทดสอบ					χ^2		
		MSE	A.Bias	Overlap*	U.item	O.item			
ขนาดคลังข้อสอบ 600 ข้อ ความยาวข้อสอบ 32 ข้อ	a-STR1	a=0.5	1.244	-.651	.629	146	4	25.010	
		a=1.0	.453	-.321	.659	143	7		
		1:1:1:1	a=1.5	.254	-.236	.669	150		0
		a=2.0	.156	-.157	.619	150	0		
	a-STR2	a=0.5	1.361	-.713	.259	293	7	40.912	
		a=1.0	.508	-.372	.860	69	6		
		2:0.5:0.5:1	a=1.5	.291	-.261	.850	68		7
		a=2.0	.167	-.171	.579	150	0		
	a-STR3	a=0.5	1.225	-.634	.709	145	5	83.378	
		a=1.0	.477	-.285	.870	68	7		
		1:0.5:0.5:2	a=1.5	.281	-.214	.850	67		8
		a=2.0	.158	-.169	.369	300	0		
	a-STR4	a=0.5	1.172	-.622	.439	220	5	62.346	
		a=1.0	.484	-.301	.850	68	7		
		1.5:0.5:0.5:1.5	a=1.5	.272	-.223	.880	68		7
		a=2.0	.159	-.162	.429	225	0		

MSE = ความคลาดเคลื่อนกำลังสองเฉลี่ย

A.Bias = ความลำเอียงเฉลี่ย

Overlap = อัตราการทับซ้อนข้อสอบ * หน่วย $\times 10^{-6} + 1.99 \times 10^{-4}$

U.item = ข้อสอบที่มีการแสดงมากเกินไป

O.item = ข้อสอบที่มีการใช้ต่ำ

χ^2 = ค่าไคสแควร์แสดงการแจกแจงอัตราการศึกษาข้อสอบ

a = ค่าอำนาจจำแนก

a-STR1 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
 กำหนดอัตราส่วนจำนวนข้อสอบเท่ากันทุกชั้น

a-STR2 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
 กำหนดอัตราส่วนจำนวนข้อสอบชั้นแรกมากกว่าชั้นอื่น

a-STR3 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
 กำหนดอัตราส่วนจำนวนข้อสอบชั้นสุดท้ายมากกว่าชั้นอื่น

a-STR4 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
 กำหนดอัตราส่วนจำนวนข้อสอบชั้นแรกและชั้นสุดท้ายมากกว่าชั้นอื่น

ตาราง 39 การทดสอบของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจ
จำแนก ขนาดคลังข้อสอบ 600 ข้อ ความยาวแบบทดสอบ 44 ข้อ

เงื่อนไขการทดสอบ		ผลการทดสอบ					χ^2
		MSE	A.Bias	Overlap*	U.item	O.item	
ขนาดคลังข้อสอบ 600 ข้อ ความยาวข้อสอบ 44 ข้อ	a=0.5	1.250	-.658	.749	144	6	51.045
	a-STR1 a=1.0	.437	-.340	.763	143	7	
	1:1:1:1 a=1.5	.240	-.248	.814	149	1	
	a=2.0	.152	-.172	.727	150	0	
	a=0.5	1.294	-.698	.472	294	6	59.241
	a-STR2 a=1.0	.488	-.365	.909	63	12	
	2:0.5:0.5:1 a=1.5	.269	-.252	.894	62	13	
	a=2.0	.143	-.163	.778	150	0	
	a=0.5	1.199	-.637	.727	144	6	116.575
	a-STR3 a=1.0	.445	-.314	.909	57	18	
	1:0.5:0.5:2 a=1.5	.270	-.246	.880	54	21	
	a=2.0	.142	-.161	.465	300	0	
	a=0.5	1.263	-.698	.625	219	6	74.412
	a-STR4 a=1.0	.465	-.366	.901	58	17	
	1.5:0.5:0.5:1.5 a=1.5	.265	-.248	.916	60	15	
	a=2.0	.139	-.164	.596	225	0	

MSE = ความคลาดเคลื่อนกำลังสองเฉลี่ย

A.Bias = ความลำเอียงเฉลี่ย

Overlap = อัตราการทับซ้อนข้อสอบ * หน่วย $\times 10^{-6} + 1.99 \times 10^{-4}$

U.item = ข้อสอบที่มีการแสดงมากเกินไป

O.item = ข้อสอบที่มีการใช้ต่ำ

χ^2 = ค่าไคสแควร์แสดงการแจกแจงอัตราการแสดงข้อสอบ

a = ค่าอำนาจจำแนก

a-STR1 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
กำหนดอัตราส่วนจำนวนข้อสอบเท่ากันทุกชั้น

a-STR2 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
กำหนดอัตราส่วนจำนวนข้อสอบชั้นแรกมากกว่าชั้นอื่น

a-STR3 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
กำหนดอัตราส่วนจำนวนข้อสอบชั้นสุดท้ายมากกว่าชั้นอื่น

a-STR4 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
กำหนดอัตราส่วนจำนวนข้อสอบชั้นแรกและชั้นสุดท้ายมากกว่าชั้นอื่น

ตาราง 40 การทดสอบของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจ
 จำแนกร่วมกับการควบคุมการแสดงผลข้อสอบ ขนาดคลังข้อสอบ 600 ข้อ ความยาวแบบทดสอบ
 44 ข้อ

		ผลการทดสอบ	MSE	A.Bias	Overlap*	U.item	O.item	χ^2
ขนาดคลังข้อสอบ 600 ข้อ	ความยาวข้อสอบ 44 ข้อ	a=0.5	1.064	-.553	.756	142	8	27.637
		a-STR1	a=1.0	.470	-.370	.756	144	6
		1:1:1:1	a=1.5	.241	-.226	.763	150	0
			a=2.0	.155	-.168	.778	150	0

MSE = ความคลาดเคลื่อนกำลังสองเฉลี่ย

A.Bias = ความลำเอียงเฉลี่ย

Overlap = อัตราการทับซ้อนข้อสอบ * หน่วย $\times 10^{-6} + 1.99 \times 10^{-4}$

U.item = ข้อสอบที่มีการแสดงมากเกินไป

O.item = ข้อสอบที่มีการใช้ต่ำ

χ^2 = ค่าไคสแควร์แสดงการแจกแจงอัตราการแสดงข้อสอบ

a = ค่าอำนาจจำแนก

a-STR1 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก

กำหนดอัตราส่วนจำนวนข้อสอบเท่ากันทุกชั้น

ตาราง 41 การทดสอบของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจ
จำแนก ขนาดคลังข้อสอบ 800 ข้อ ความยาวแบบทดสอบ 20 ข้อ

ผลการทดสอบ		MSE	A.Bias	Overlap*	U.item	O.item	χ^2
เงื่อนไขการทดสอบ							
ขนาดคลังข้อสอบ 800 ข้อ ความยาวข้อสอบ 20 ข้อ	a=0.5	1.837	-.874	.223	199	1	53.277
	a-STR1 a=1.0	.607	-.399	.159	198	2	
	1:1:1:1 a=1.5	.289	-.225	.207	200	0	
	a=2.0	.183	-.165	.303	200	0	
	a=0.5	1.712	-.824	-.481	399	1	56.997
	a-STR2 a=1.0	.587	-.394	.671	99	1	
	2:0.5:0.5:1 a=1.5	.319	-.241	.639	100	0	
	a=2.0	.189	-.171	.287	200	0	
	a=0.5	1.625	-.797	.207	199	1	80.543
	a-STR3 a=1.0	.582	-.380	.655	99	1	
	1:0.5:0.5:2 a=1.5	.307	-.238	.687	99	1	
	a=2.0	.182	-.168	-.673	400	0	
	a=0.5	1.782	-.857	-.161	298	2	46.974
	a-STR4 a=1.0	.605	-.400	.655	97	3	
	1.5:0.5:0.5:1.5 a=1.5	.311	-.243	.591	100	0	
	a=2.0	.186	-.171	-.177	300	0	

MSE = ความคลาดเคลื่อนกำลังสองเฉลี่ย

A.Bias = ความลำเอียงเฉลี่ย

Overlap = อัตราการทับซ้อนข้อสอบ * หน่วย $\times 10^{-6} + 1.99 \times 10^{-4}$

U.item = ข้อสอบที่มีการแสดงมากเกินไป

O.item = ข้อสอบที่มีการใช้ต่ำ

χ^2 = ค่าไคสแควร์แสดงการแจกแจงอัตราการแสดงข้อสอบ

a = ค่าอำนาจจำแนก

a-STR1 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
กำหนดอัตราส่วนจำนวนข้อสอบเท่ากันทุกชั้น

a-STR2 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
กำหนดอัตราส่วนจำนวนข้อสอบชั้นแรกมากกว่าชั้นอื่น

a-STR3 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
กำหนดอัตราส่วนจำนวนข้อสอบชั้นสุดท้ายมากกว่าชั้นอื่น

a-STR4 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
กำหนดอัตราส่วนจำนวนข้อสอบชั้นแรกและชั้นสุดท้ายมากกว่าชั้นอื่น

ตาราง 42 การทดสอบของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจ
 จำแนกร่วมกับการควบคุมการแสดงผลข้อสอบ ขนาดคลังข้อสอบ 800 ข้อ ความยาวแบบทดสอบ
 20 ข้อ

เงื่อนไขการทดสอบ		ผลการทดสอบ					χ^2		
		MSE	A.Bias	Overlap*	U.item	O.item			
ขนาดคลังข้อสอบ 800 ข้อ ความยาวข้อสอบ 20 ข้อ	a-STR1	a=0.5	1.481	-.755	.223	195	5	27.178	
		a=1.0	.579	-.380	.271	200	0		
		1:1:1:1	a=1.5	.304	-.262	.287	200		0
		a=2.0	.187	-.177	.239	200	0		
	a-STR2	a=0.5	1.351	-.690	-.577	397	3	42.186	
		a=1.0	.585	-.383	.655	98	2		
		2:0.5:0.5:1	a=1.5	.311	-.249	.687	100		0
		a=2.0	.192	-.168	.143	200	0		
	a-STR3	a=0.5	1.671	-.828	.223	195	5	57.428	
		a=1.0	.580	-.388	.703	96	4		
		1:0.5:0.5:2	a=1.5	.309	-.238	.623	100		0
		a=2.0	.185	-.160	-.641	400	0		
	a-STR4	a=0.5	1.556	-.767	-.145	296	4	49.838	
		a=1.0	.549	-.337	.655	97	3		
		1.5:0.5:0.5:1.5	a=1.5	.316	-.244	.655	100		0
		a=2.0	.190	-.163	-.081	300	0		

MSE = ความคลาดเคลื่อนกำลังสองเฉลี่ย

A.Bias = ความลำเอียงเฉลี่ย

Overlap = อัตราการทับซ้อนข้อสอบ * หน่วย $\times 10^{-6} + 1.99 \times 10^{-4}$

U.item = ข้อสอบที่มีการแสดงมากเกินไป

O.item = ข้อสอบที่มีการใช้ต่ำ

χ^2 = ค่าไคสแควร์แสดงการแจกแจงอัตราการศึกษาข้อสอบ

a = ค่าอำนาจจำแนก

a-STR1 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
 กำหนดอัตราส่วนจำนวนข้อสอบเท่ากันทุกชั้น

a-STR2 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
 กำหนดอัตราส่วนจำนวนข้อสอบชั้นแรกมากกว่าชั้นอื่น

a-STR3 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
 กำหนดอัตราส่วนจำนวนข้อสอบชั้นสุดท้ายมากกว่าชั้นอื่น

a-STR4 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
 กำหนดอัตราส่วนจำนวนข้อสอบชั้นแรกและชั้นสุดท้ายมากกว่าชั้นอื่น

ตาราง 43 การทดสอบของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจ
จำแนก ขนาดคลังข้อสอบ 800 ข้อ ความยาวแบบทดสอบ 32 ข้อ

ผลการทดสอบ		MSE	A.Bias	Overlap*	U.item	O.item	χ^2
เงื่อนไขการทดสอบ							
ขนาดคลังข้อสอบ 800 ข้อ ความยาวข้อสอบ 32 ข้อ	a=0.5	1.342	-.710	.509	198	2	59.832
	a-STR1 a=1.0	.505	-.395	.519	199	1	
	1:1:1:1 a=1.5	.251	-.242	.589	200	0	
	a=2.0	.161	-.173	.559	200	0	
	a=0.5	1.320	-.703	.059	399	1	71.797
	a-STR2 a=1.0	.531	-.376	.749	94	6	
	2:0.5:0.5:1 a=1.5	.282	-.259	.770	98	2	
	a=2.0	.163	-.164	.519	200	0	
	a=0.5	1.514	-.764	.569	197	3	94.042
	a-STR3 a=1.0	.580	-.422	.759	92	8	
	1:0.5:0.5:2 a=1.5	.276	-.263	.780	99	1	
	a=2.0	.155	-.168	.159	400	0	
	a=0.5	1.373	-.730	.219	298	2	84.740
	a-STR4 a=1.0	.508	-.367	.790	91	9	
	1.5:0.5:0.5:1.5 a=1.5	.280	-.261	.749	100	0	
	a=2.0	.157	-.167	.229	300	0	

MSE = ความคลาดเคลื่อนกำลังสองเฉลี่ย

A.Bias = ความลำเอียงเฉลี่ย

Overlap = อัตราการทับซ้อนข้อสอบ * หน่วย $\times 10^{-6} + 1.99 \times 10^{-4}$

U.item = ข้อสอบที่มีการแสดงมากเกินไป

O.item = ข้อสอบที่มีการใช้ค่า

χ^2 = ค่าไคสแควร์แสดงการแจกแจงอัตราการแสดงข้อสอบ

a = ค่าอำนาจจำแนก

a-STR1 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
กำหนดอัตราส่วนจำนวนข้อสอบเท่ากันทุกชั้น

a-STR2 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
กำหนดอัตราส่วนจำนวนข้อสอบชั้นแรกมากกว่าชั้นอื่น

a-STR3 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
กำหนดอัตราส่วนจำนวนข้อสอบชั้นสุดท้ายมากกว่าชั้นอื่น

a-STR4 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
กำหนดอัตราส่วนจำนวนข้อสอบชั้นแรกและชั้นสุดท้ายมากกว่าชั้นอื่น

ตาราง 44 การทดสอบของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจ
 จำแนกร่วมกับการควบคุมการแสดงผลข้อสอบ ขนาดคลังข้อสอบ 800 ข้อ ความยาวแบบทดสอบ
 32 ข้อ

เงื่อนไขการทดสอบ		ผลการทดสอบ					χ^2		
		MSE	A.Bias	Overlap*	U.item	O.item			
ขนาดคลังข้อสอบ 800 ข้อ ความยาวข้อสอบ 32 ข้อ	a-STR1	a=0.5	1.253	-.684	.579	193	7	34.371	
		a=1.0	.466	-.344	.519	198	2		
		1:1:1:1	a=1.5	.250	-.236	.559	200		0
		a=2.0	.153	-.146	.539	200	0		
	a-STR2	a=0.5	1.254	-.670	.009	396	4	56.412	
		a=1.0	.489	-.320	.840	91	9		
		2:0.5:0.5:1	a=1.5	.267	-.232	.820	100		0
		a=2.0	.154	-.150	.489	200	0		
	a-STR3	a=0.5	1.225	-.661	.569	197	3	72.142	
		a=1.0	.497	-.370	.800	97	3		
		1:0.5:0.5:2	a=1.5	.278	-.252	.770	98		2
		a=2.0	.147	-.159	.019	400	0		
	a-STR4	a=0.5	1.260	-.661	.339	296	4	61.972	
		a=1.0	.453	-.293	.840	93	7		
		1.5:0.5:0.5:1.5	a=1.5	.283	-.246	.810	97		3
		a=2.0	.156	-.167	.289	300	0		

MSE = ความคลาดเคลื่อนกำลังสองเฉลี่ย

A.Bias = ความลำเอียงเฉลี่ย

Overlap = อัตราการทับซ้อนข้อสอบ * หน่วย $\times 10^{-6} + 1.99 \times 10^{-4}$

U.item = ข้อสอบที่มีการแสดงมากเกินไป

O.item = ข้อสอบที่มีการใช้ต่ำ

χ^2 = ค่าไคสแควร์แสดงการแจกแจงอัตราการแสดงข้อสอบ

a = ค่าอำนาจจำแนก

a-STR1 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
 กำหนดอัตราส่วนจำนวนข้อสอบเท่ากันทุกชั้น

a-STR2 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
 กำหนดอัตราส่วนจำนวนข้อสอบชั้นแรกมากกว่าชั้นอื่น

a-STR3 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
 กำหนดอัตราส่วนจำนวนข้อสอบชั้นสุดท้ายมากกว่าชั้นอื่น

a-STR4 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
 กำหนดอัตราส่วนจำนวนข้อสอบชั้นแรกและชั้นสุดท้ายมากกว่าชั้นอื่น

ตาราง 45 การทดสอบของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจ
จำแนก ขนาดคลังข้อสอบ 800 ข้อ ความยาวแบบทดสอบ 44 ข้อ

ผลการทดสอบ		MSE	A.Bias	Overlap*	U.item	O.item	χ^2
เงื่อนไขการทดสอบ							
ขนาดคลังข้อสอบ 800 ข้อ ความยาวข้อสอบ 44 ข้อ	a=0.5	1.189	-.666	.661	196	4	61.418
	a-STR1 a=1.0	.427	-.331	.719	195	5	
	1:1:1:1 a=1.5	.239	-.245	.654	200	0	
	a=2.0	.144	-.166	.712	200	0	
	a=0.5	1.293	-.715	.290	396	4	75.060
	a-STR2 a=1.0	.458	-.364	.850	83	17	
	2:0.5:0.5:1 a=1.5	.259	-.249	.850	91	9	
	a=2.0	.146	-.170	.712	200	0	
	a=0.5	1.270	-.690	.719	195	5	123.287
	a-STR3 a=1.0	.457	-.358	.858	85	15	
	1:0.5:0.5:2 a=1.5	.251	-.240	.836	92	8	
	a=2.0	.134	-.169	.363	400	0	
	a=0.5	1.293	-.730	.450	296	4	92.311
	a-STR4 a=1.0	.461	-.374	.829	91	9	
	1.5:0.5:0.5:1.5 a=1.5	.250	-.256	.872	99	1	
	a=2.0	.132	-.168	.472	300	0	

MSE = ความคลาดเคลื่อนกำลังสองเฉลี่ย

A.Bias = ความลำเอียงเฉลี่ย

Overlap = อัตราการทับซ้อนข้อสอบ * หน่วย $\times 10^{-6} + 1.99 \times 10^{-4}$

U.item = ข้อสอบที่มีการแสดงมากเกินไป

O.item = ข้อสอบที่มีการใช้ค่า

χ^2 = ค่าไคสแควร์แสดงการแจกแจงอัตราการแสดงข้อสอบ

a = ค่าอำนาจจำแนก

a-STR1 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
กำหนดอัตราส่วนจำนวนข้อสอบเท่ากันทุกชั้น

a-STR2 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
กำหนดอัตราส่วนจำนวนข้อสอบชั้นแรกมากกว่าชั้นอื่น

a-STR3 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
กำหนดอัตราส่วนจำนวนข้อสอบชั้นสุดท้ายมากกว่าชั้นอื่น

a-STR4 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
กำหนดอัตราส่วนจำนวนข้อสอบชั้นแรกและชั้นสุดท้ายมากกว่าชั้นอื่น

ตาราง 46 การทดสอบของวิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจ
 จำแนกร่วมกับการควบคุมการแสดงผลข้อสอบ ขนาดคลังข้อสอบ 800 ข้อ ความยาวแบบทดสอบ
 44 ข้อ

เงื่อนไขการทดสอบ		ผลการทดสอบ					χ^2		
		MSE	A.Bias	Overlap*	U.item	O.item			
ขนาดคลังข้อสอบ 800 ข้อ ความยาวข้อสอบ 44 ข้อ	a-STR1	a=0.5	1.133	-.599	.683	193	7	35.506	
		a=1.0	.472	-.391	.698	196	4		
		1:1:1:1	a=1.5	.226	-.228	.676	200		0
		a=2.0	.134	-.130	.712	200	0		
	a-STR2	a=0.5	1.146	-.655	.290	395	5	58.001	
		a=1.0	.423	-.298	.887	93	7		
		2:0.5:0.5:1	a=1.5	.256	-.210	.887	88		12
		a=2.0	.145	-.157	.690	200	0		
	a-STR3	a=0.5	1.265	-.687	.719	191	9	105.944	
		a=1.0	.428	-.301	.880	82	18		
		1:0.5:0.5:2	a=1.5	.265	-.235	.850	91		9
		a=2.0	.133	-.147	.319	400	0		
	a-STR4	a=0.5	1.115	-.619	.450	295	5	78.961	
		a=1.0	.436	-.271	.858	87	13		
		1.5:0.5:0.5:1.5	a=1.5	.262	-.245	.858	95		5
		a=2.0	.137	-.159	.501	300	0		

MSE = ความคลาดเคลื่อนกำลังสองเฉลี่ย

A.Bias = ความลำเอียงเฉลี่ย

Overlap = อัตราการทับซ้อนข้อสอบ * หน่วย $\times 10^{-6} + 1.99 \times 10^{-4}$

U.item = ข้อสอบที่มีการแสดงมากเกินไป

O.item = ข้อสอบที่มีการใช้ต่ำ

χ^2 = ค่าไคสแควร์แสดงการแจกแจงอัตราการศึกษาข้อสอบ

a = ค่าอำนาจจำแนก

a-STR1 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
 กำหนดอัตราส่วนจำนวนข้อสอบเท่ากันทุกชั้น

a-STR2 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
 กำหนดอัตราส่วนจำนวนข้อสอบชั้นแรกมากกว่าชั้นอื่น

a-STR3 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
 กำหนดอัตราส่วนจำนวนข้อสอบชั้นสุดท้ายมากกว่าชั้นอื่น

a-STR4 = วิธีการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
 กำหนดอัตราส่วนจำนวนข้อสอบชั้นแรกและชั้นสุดท้ายมากกว่าชั้นอื่น

ตาราง 47 ความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Square Error : MSE) เมื่อสิ้นสุดการทดสอบของวิธีกำหนด

วิธี / การควบคุม การแสดงผลข้อสอบ		a-STR1		a-STR2		a-STR3		a-STR4	
		-SH	+SH	-SH	+SH	-SH	+SH	-SH	+SH
ขนาดคลังข้อสอบ / ความยาวแบบทดสอบ									
200	20	.212	.212	.211	-	.206	-	.210	-
	32	.234	-	.195	-	.173	-	.197	-
	44	.200	-	.195	-	.173	-	.172	-
400	20	.192	.201	.201	.193	.190	.186	.195	.197
	32	.176	.170	.172	-	.166	.164	.167	-
	44	.168	.158	.162	-	.140	-	.143	-
600	20	.193	.181	.186	.188	.185	.195	.187	.192
	32	.164	.156	.168	.167	.154	.158	.154	.159
	44	.152	.155	.143	-	.142	-	.139	-
800	20	.183	.187	.189	.192	.182	.185	.186	.190
	32	.161	.153	.163	.154	.155	.147	.157	.156
	44	.144	.134	.146	.145	.134	.133	.132	.137

-SH = ไม่ควบคุมการแสดงผลข้อสอบ

+SH = ควบคุมการแสดงผลข้อสอบ

ตาราง 48 ความลำเอียงเฉลี่ย (Average Bias) เมื่อสิ้นสุดการทดสอบของวิธีกำหนด

วิธี / การควบคุม การแสดงข้อสอบ		a-STR1		a-STR2		a-STR3		a-STR4	
		-SH	+SH	-SH	+SH	-SH	+SH	-SH	+SH
ขนาดคลังข้อสอบ / ความยาวแบบทดสอบ									
200	20	-.173	-.176	-.184	-	-.197	-	-.185	-
	32	-.226	-	-.191	-	-.156	-	-.204	-
	44	-.184	-	-.155	-	-.190	-	-.193	-
400	20	-.146	-.161	-.185	-.152	-.170	-.158	-.175	-.165
	32	-.168	-.168	-.175	-	-.177	-.161	-.170	-
	44	-.180	-.161	-.156	-	-.163	-	-.169	-
600	20	-.171	-.165	-.157	-.155	-.166	-.178	-.151	-.173
	32	-.169	-.157	-.176	-.171	-.179	-.169	-.161	-.162
	44	-.172	-.168	-.163	-	-.161	-	-.164	-
800	20	-.165	-.177	-.171	-.168	-.168	-.160	-.171	-.163
	32	-.173	-.146	-.164	-.150	-.168	-.159	-.167	-.167
	44	-.166	-.130	-.170	.157	-.169	-.147	-.168	-.159

-SH = ไม่ควบคุมการแสดงข้อสอบ

+SH = ควบคุมการแสดงข้อสอบ

ตาราง 49 จำนวนข้อสอบที่มีการแสดงมากเกินไป (Overexpose Item) เมื่อสิ้นสุดการทดสอบของวิธีกำหนด

วิธี / การควบคุม การแสดงผลข้อสอบ		a-STR1		a-STR2		a-STR3		a-STR4	
		-SH	+SH	-SH	+SH	-SH	+SH	-SH	+SH
ขนาดคลังข้อสอบ / ความยาวแบบทดสอบ									
200	20	12	14	24	-	20	-	21	-
	32	44	-	61	-	45	-	48	-
	44	95	-	86	-	64	-	75	-
400	20	4	7	11	11	15	20	8	16
	32	18	14	30	-	35	40	32	-
	44	38	29	70	-	62	-	43	-
600	20	4	6	4	7	5	8	4	7
	32	5	11	13	20	11	20	13	19
	44	14	14	31	-	45	-	38	-
800	20	3	5	2	5	3	9	5	7
	32	3	9	9	13	12	8	11	14
	44	9	11	30	24	28	36	14	23

-SH = ไม่ควบคุมการแสดงผลข้อสอบ +SH = ควบคุมการแสดงผลข้อสอบ

ตาราง 50 จำนวนข้อสอบที่มีการใช้ต่ำ (Underutilized Item) เมื่อสิ้นสุดการทดสอบของวิธีกำหนด

วิธี / การควบคุม การแสดงผลข้อสอบ		a-STR1		a-STR2		a-STR3		a-STR4	
		-SH	+SH	-SH	+SH	-SH	+SH	-SH	+SH
ขนาดคลังข้อสอบ / ความยาวแบบทดสอบ									
200	20	188	186	176	-	180	-	179	-
	32	156	-	139	-	155	-	152	-
	44	105	-	114	-	136	-	125	-
400	20	396	393	389	389	385	380	392	384
	32	382	386	370	-	365	360	368	-
	44	362	371	330	-	338	-	357	-
600	20	596	594	596	593	595	592	596	593
	32	595	589	587	580	589	580	587	581
	44	586	586	569	-	555	-	562	-
800	20	797	795	798	795	797	791	795	793
	32	797	791	791	787	788	792	789	786
	44	791	789	770	776	772	764	786	777

-SH = ไม่ควบคุมการแสดงผลข้อสอบ +SH = ควบคุมการแสดงผลข้อสอบ

ตาราง 51 อัตราการทับซ้อนข้อสอบ (Item Overlap Rate)* เมื่อสิ้นสุดการทดสอบของวิธีกำหนด

วิธี / การควบคุม การแสดงผลข้อสอบ		a-STR1		a-STR2		a-STR3		a-STR4	
		-SH	+SH	-SH	+SH	-SH	+SH	-SH	+SH
ขนาดคลังข้อสอบ / ความยาวแบบทดสอบ									
200	20	3.32	3.29	3.35	-	3.37	-	3.39	-
	32	3.72	-	3.64	-	3.71	-	3.66	-
	44	3.87	-	3.78	-	3.77	-	3.75	-
400	20	2.55	2.54	2.54	2.54	2.54	2.67	2.47	2.52
	32	3.07	3.13	3.25	-	3.16	3.14	3.14	-
	44	3.39	3.43	3.37	-	3.38	-	3.37	-
600	20	1.74	1.88	1.79	1.72	1.87	1.72	1.70	1.71
	32	2.57	2.57	2.74	2.54	2.75	2.79	2.51	2.59
	44	3.05	3.05	3.05	-	2.98	-	3.03	-
800	20	0.89	1.02	1.11	0.90	0.87	0.90	0.90	1.08
	32	2.17	2.19	2.09	2.15	2.26	2.14	1.98	2.27
	44	2.74	2.76	2.70	2.75	2.77	2.76	2.62	2.66

* ค่าในตารางคูณด้วย 10^{-6} บวกด้วย 1.99×10^{-4}

-SH = ไม่ควบคุมการแสดงผลข้อสอบ

+SH = ควบคุมการแสดงผลข้อสอบ

ตาราง 52 การแจกแจงอัตราการแสดงผลข้อสอบ (Item Exposure Rate Distribution) ด้วยไคสแควร์ เมื่อสิ้นสุดการทดสอบของวิธีกำหนด

วิธี / การควบคุม การแสดงผลข้อสอบ		a-STR1		a-STR2		a-STR3		a-STR4	
		-SH	+SH	-SH	+SH	-SH	+SH	-SH	+SH
ขนาดคลังข้อสอบ / ความยาวแบบทดสอบ									
200	20	15.50	17.00	28.87	-	48.50	-	31.91	-
	32	23.07	-	37.96	-	68.16	-	45.03	-
	44	24.05	-	43.78	-	84.24	-	49.99	-
400	20	36.76	18.95	39.91	39.46	63.07	57.19	46.51	42.11
	32	32.88	22.13	46.81	-	81.91	73.64	54.52	-
	44	39.61	21.87	48.26	-	89.60	-	66.66	-
600	20	39.31	20.54	47.07	36.63	65.94	51.64	56.97	41.16
	32	42.52	25.01	57.28	40.91	85.55	83.37	65.76	62.34
	44	51.04	27.63	59.24	-	116.57	-	74.41	-
800	20	53.27	27.17	56.99	42.18	80.54	57.42	46.97	49.83
	32	59.83	34.37	71.79	56.41	94.04	72.14	84.74	61.97
	44	61.41	35.50	75.06	58.00	123.28	105.94	92.31	78.96

-SH = ไม่ควบคุมการแสดงผลข้อสอบ +SH = ควบคุมการแสดงผลข้อสอบ

โปรแกรมภาษากิวเบสิก (Q-Basic)

การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก

สำหรับขนาดคลังข้อสอบ 400 ข้อและ ความยาวแบบทดสอบ 20 ข้อ

```
1000 CLS
1020 REM "ss1-72. dat" PREPARATION
1040 DIM ftheta(5000)
1060 OPEN "ss1.dat" FOR INPUT AS #1
1080 d = 1
1100 IF EOF(1) THEN 1180
1120 INPUT #1, ftheta(d)
1140 d = d + 1
1160 GOTO 1100
1180 CLOSE #1
1200 REM "sc1-72. dat" PREPARATION
1220 DIM theta(5000)
1240 OPEN "sc1.dat" FOR INPUT AS #1
1260 d = 1
1280 IF EOF(1) THEN 1360
1300 INPUT #1, theta(d)
1320 d = d + 1
1340 GOTO 1280
1360 CLOSE #1
1380 REM "bs.dat" PREPARATION
1400 DIM b1(3000)
1420 OPEN "bs.dat" FOR INPUT AS #1
1440 e = 1
1460 IF EOF(1) THEN 1540
1480 INPUT #1, b1(e)
```

```
1500 e = e + 1
1520 GOTO 1460
1540 CLOSE #1
1560 REM ITEM PREPARATION
1580 DIM bb(4, 100)
1600 FOR i = 1 TO 4
1620 IF i = 1 THEN OPEN "bb1.dat" FOR INPUT AS #1
1640 IF i = 2 THEN OPEN "bb2.dat" FOR INPUT AS #1
1660 IF i = 3 THEN OPEN "bb3.dat" FOR INPUT AS #1
1680 IF i = 4 THEN OPEN "bb4.dat" FOR INPUT AS #1
1700 g = 1
1720 IF EOF(1) THEN 1800
1740 INPUT #1, bb(i, g)
1760 g = g + 1
1780 GOTO 1720
1800 CLOSE #1
1820 NEXT i
1840 REM ADJUSTED FREQUENCY AS ZERO
1860 DIM ffb(4, 100)
1880 FOR i = 1 TO 4: REM LEVEL
1900 IF i = 1 THEN gg = 100
1920 IF i = 2 THEN gg = 100
1940 IF i = 3 THEN gg = 100
1960 IF i = 4 THEN gg = 100
1980 FOR g = 1 TO gg
2000 ffb(i, g) = 0
2020 NEXT g
2040 NEXT i
2060 FOR i = 1 TO 4
2080 sdiff(i) = 0
2100 mdiff(i) = 0
```

```

2120 share(i) = 0
2140 un(i) = 0
2160 ov(i) = 0
2180 NEXT i
2200 REM ENTRY POINT DETERMINATION
2220 FOR d = 1 TO 5000
2240 a = 1
2260 c = .2
2280 big = .5
2300 snu = 0
2320 sud = 0
2340 the = theta(d)
2345 REM FIND THE NEAREST VALUE FOR THETA-BB
2350 k = 1
2355 min = ABS(the - b1(k))
2360 FOR e = 1 TO 3000
2365 diff0 = ABS(the - b1(e))
2370 IF min < diff0 THEN GOTO 2385
2375 IF min >= diff0 THEN k = e
2380 IF min >= diff0 THEN min = diff0
2382 ee = e
2385 NEXT e
2400 dev1 = a * (the - b1(ee))
2420 phat1 = 1! / (1! + EXP(-dev1))
2440 pt1 = c + (1! - c) * phat1
2460 REM PROTECT AGAINST DIVIDE BY ZERO
2480 IF (pt1 < .00001) THEN pt1 = .00001
2500 IF (pt1 > .99999) THEN pt1 = .99999
2520 RANDOMIZE TIMER: rnum1 = RND(1#)
2540 IF rnum1 >= pt1 THEN uij1 = 1
2560 IF rnum1 < pt1 THEN uij1 = 0

```

```

2580 IF uij1 = 1 THEN b2 = b1(ee) + 2
2600 IF uij1 = 0 THEN b2 = b1(ee) - 2
2620 IF b2 > 3 THEN b2 = 3
2640 IF b2 < -3 THEN b2 = -3
2660 dev2 = a * (the - b2)
2680 phat2 = 1! / (1! + EXP(-dev2))
2700 pt2 = c + (1! - c) * phat2
2720 REM PROTECT AGAINST DIVIDE BY ZERO
2740 IF (pt2 < .00001) THEN pt2 = .00001
2760 IF (pt2 > .99999) THEN pt2 = .99999
2780 RANDOMIZE TIMER: rnum2 = RND(1#)
2800 IF rnum2 >= pt2 THEN uij2 = 1
2820 IF rnum2 < pt2 THEN uij2 = 0
2840 IF uij2 = 1 THEN b3 = b2 + 2
2860 IF uij2 = 0 THEN b3 = b2 - 2
2880 IF b3 > 3 THEN b3 = 3
2900 IF b3 < -3 THEN b3 = -3
2920 dev3 = a * (the - b3)
2940 phat3 = 1! / (1! + EXP(-dev3))
2960 pt3 = c + (1! - c) * phat3
2980 REM PROTECT AGAINST DIVIDE BY ZERO
3000 IF (pt3 < .00001) THEN pt3 = .00001
3020 IF (pt3 > .99999) THEN pt3 = .99999
3040 RANDOMIZE TIMER: rnum3 = RND(1#)
3060 IF rnum3 >= pt3 THEN uij3 = 1
3080 IF rnum3 < pt3 THEN uij3 = 0
3100 IF (uij1 = 0) AND (uij2 = 0) AND (uij3 = 0) THEN GOTO 2300
3120 IF (uij1 = 1) AND (uij2 = 1) AND (uij3 = 1) THEN GOTO 2300
3140 zz = 0
3160 dev = a * (the - b1(ee))
3180 phat = 1! / (1! + EXP(-dev))

```

```

3200 pt = c + (1! - c) * phat
3220 REM PROTECT AGAINST DEVIDE BY ZERO
3240 IF (pt < .00001) THEN pt = .00001
3260 IF (pt > .99999) THEN pt = .99999
3280 wij = pt * (1 - pt)
3300 vij = uij1 - pt
3320 psp = phat / pt
3340 snu = snu + a * vij * psp
3360 sud = sud + a ^ 2 * wij * psp * psp
3380 dev = a * (the - b2)
3400 phat = 1! / (1! + EXP(-dev))
3420 pt = c + (1! - c) * phat
3460 REM PROTECT AGAINST DEVIDE BY ZERO
3480 IF (pt < .00001) THEN pt = .00001
3500 IF (pt > .99999) THEN pt = .99999
3520 wij = pt * (1 - pt)
3540 vij = uij2 - pt
3560 psp = phat / pt
3580 snu = snu + a * vij * psp
3600 sud = sud + a ^ 2 * wij * psp * psp
3620 dev = a * (the - b3)
3640 phat = 1! / (1! + EXP(-dev))
3660 pt = c + (1! - c) * phat
3680 REM PROTECT AGAINST DEVIDE BY ZERO
3700 IF (pt < .00001) THEN pt = .00001
3720 IF (pt > .99999) THEN pt = .99999
3740 wij = pt * (1 - pt)
3760 vij = uij3 - pt
3780 psp = phat / pt
3800 snu = snu + a * vij * psp
3820 sud = sud + a ^ 2 * wij * psp * psp

```

```

3840 REM PROTECT AGAINST DEVIDE BY ZERO
3860 IF (sud < .00001) THEN sud = .00001
3880 delta = snu / sud
3900 IF ABS(delta < big) GOTO 3960
3920 REM PROTECT AGAINST BIG CHANGE IN THETA
3940 IF delta > 0! THEN
    delta = big
ELSE
    delta = -big
END IF
3960 the = ftheta(d) + delta
3980 IF ABS(delta < .05) THEN GOTO 4100
4000 zz = zz + 1
4020 IF zz = 5 GOTO 4100
4040 snu = 0
4060 sud = 0
4080 GOTO 3160
4100 IF the > 3! THEN the = 3!
4120 IF the < -3! THEN the = -3!
4140 REM DATA PREPARATION
4160 DIM ever(4, 100)
4180 FOR i = 1 TO 4
4200 FOR g = 1 TO 100
4220 ever(i, g) = 0
4240 NEXT g
4260 NEXT i
4280 c = 0!
4300 pl = 400: REM ITEM POOL
4320 f = .5
4340 FOR i = 1 TO 4: REM LEVEL
4360 IF i = 1 THEN gg = 100

```

```

4380 IF i = 2 THEN gg = 100
4400 IF i = 3 THEN gg = 100
4420 IF i = 4 THEN gg = 100
4440 a = f
4460 tl = 5
4480 FOR l = 1 TO tl: REM TEST LENGTH
4500 REM FIND THE NEAREST VALUE FOR THETA-BB
4520 k = 1
4540 min = ABS(the - bb(i, k))
4560 FOR g = 1 TO gg
4580 IF ever(i, g) = 1 THEN GOTO 4700
4600 diff0 = ABS(the - bb(i, g))
4620 IF min < diff0 THEN GOTO 4700
4640 IF min >= diff0 THEN k = g
4660 IF min >= diff0 THEN min = diff0
4680 gx = g
4700 NEXT g
4720 ever(i, gx) = 1
4740 REM a-STR EXAMINATION
4760 dev = a * (the - bb(i, gx))
4780 phat = 1! / (1! + EXP(-dev))
4800 pt = c + (1! - c) * phat
4820 REM PROTECT AGAINST DIVIDE BY ZERO
4840 IF (pt < .00001) THEN pt = .00001
4860 IF (pt > .99999) THEN pt = .99999
4880 RANDOMIZE TIMER: rnum = RND(1#)
4900 IF rnum >= pt THEN uij = 1
4920 IF rnum < pt THEN uij = 0
4940 zx = 0
4960 dev = a * (the - bb(i, gx))
4980 phat = 1! / (1! + EXP(-dev))

```

```

5000  $pt = c + (1! - c) * phat$ 
5020 REM PROTECT AGAINST DIVIDE BY ZERO
5040 IF ( $pt < .00001$ ) THEN  $pt = .00001$ 
5060 IF ( $pt > .99999$ ) THEN  $pt = .99999$ 
5080  $wij = pt * (1 - pt)$ 
5100  $vij = uij - pt$ 
5120  $psp = phat / pt$ 
5140  $snu = snu + a * vij * psp$ 
5160  $sud = sud + a^2 * wij * psp * psp$ 
5180 REM PROTECT AGAINST DIVIDE BY ZERO
5200 IF ( $sud < .00001$ ) THEN  $sud = .00001$ 
5220  $delta = snu / sud$ 
5240 IF  $ABS(delta < big)$  GOTO 5300
5260 REM PROTECT AGAINST BIG CHANGE IN THETA
5280 IF  $delta > 0!$  THEN
    delta = big
ELSE
    delta = -big
END IF
5300  $the = ftheta(d) + delta$ 
5320 IF  $ABS(delta < .05)$  THEN GOTO 5440
5340  $zx = zx + 1$ 
5360 IF  $zx = 5$  GOTO 5440
5380  $snu = 0$ 
5400  $sud = 0$ 
5420 GOTO 4960
5440 IF  $the > 3!$  THEN  $the = 3!$ 
5460 IF  $the < -3!$  THEN  $the = -3!$ 
5480 REM CALCULATE FREQUENCY
5500 IF  $ever(i, gx) = 1$  THEN  $ffb(i, gx) = ffb(i, gx) + 1$ 
5520  $g = gx$ 

```



```

5540 NEXT I
5560 REM DEPENDENT VARIABLE VALUES
5580 REM BIAS CALCULATION
5600 sdiff(i) = sdiff(i) + (the - ftheta(d))
5620 IF d = 5000 THEN bias(i) = sdiff(i) / 5000
5640 IF d = 5000 THEN PRINT TAB(1); "level = "; i; TAB(15); "bias = "; bias(i); TAB(40);
      "sdiff = "; sdiff(i)
5680 IF d = 5000 THEN LPRINT TAB(1); "level = "; i; TAB(15); "bias = "; bias(i); TAB(40);
      "sdiff = "; sdiff(i)
5700 REM MSE CALCULATION
5720 mdiff(i) = mdiff(i) + ((the - ftheta(d)) ^ 2)
5740 IF d = 5000 THEN mse(i) = mdiff(i) / 5000
5760 IF d = 5000 THEN PRINT TAB(1); "level = "; i; TAB(15); "mse = "; mse(i); TAB(40);
      "mdiff = "; mdiff(i)
5780 IF d = 5000 THEN LPRINT TAB(1); "level = "; i; TAB(15); "mse = "; mse(i); TAB(40);
      "mdiff = "; mdiff(i)
5800 IF d <> 5000 THEN GOTO 6160
5820 REM ITEM EXPOSURE CALCULATION
5840 FOR g = 1 TO gg
5860 tg = (ffb(i, g) / 5000)
5880 IF tg <= .2 THEN un(i) = un(i) + 1
5900 IF tg > .2 THEN ov(i) = ov(i) + 1
5920 NEXT g
5940 PRINT TAB(1); "level = "; i; TAB(15); "underrate = "; un(i); TAB(40); "overrate = "; ov(i)
5960 LPRINT TAB(1); "level = "; i; TAB(15); "underrate = "; un(i); TAB(40); "overrate = "; ov(i)
5980 REM OVERLAP CALCULATION
6000 FOR g = 1 TO gg
6020 sv = FIX(ffb(i, g) / 2)
6040 share(i) = share(i) + sv
6060 NEXT g
6080 np(i) = (tl * d * (d - 1) / 2)

```

```
6100 overlap(i) = share(i) / np(i)
6120 PRINT TAB(1); "level = "; i; TAB(15); "share = "; share(i); TAB(40); "np = "; np(i);
      TAB(60); "overlap = "; overlap(i)
6140 LPRINT TAB(1); "level = "; i; TAB(15); "share = "; share(i); TAB(40); "np = "; np(i);
      TAB(60); "overlap = "; overlap(i)
6160 f = f + .5
6180 NEXT i
6200 NEXT d
6220 REM CHI-SQUARE CALCULATION
6240 schi = 0
6260 tr = ((tl * 4) / pl)
6280 FOR i = 1 TO 4
6300 FOR g = 1 TO gg
6320 tg = (ffb(i, g) / 5000)
6340 wr = ((tg - tr) ^ 2) / tr
6360 schi = schi + wr
6380 NEXT g
6400 NEXT i
6420 PRINT TAB(1); "chi-square = "; schi
6440 LPRINT TAB(1); "chi-square = "; schi
6460 END
```

โปรแกรมภาษากิวเบสิก (Q-Basic)

การเตรียมพารามิเตอร์การแสดงผลข้อสอบ

สำหรับขนาดคลังข้อสอบ 400 ข้อและความยาวแบบทดสอบ 20 ข้อ

```
1000 CLS
1020 REM "ee1-36. dat" PREPARATION
1040 DIM ftheta(3000)
1060 OPEN "ee1.dat" FOR INPUT AS #1
1080 d = 1
1100 IF EOF(1) THEN 1180
1120 INPUT #1, ftheta(d)
1140 d = d + 1
1160 GOTO 1100
1180 CLOSE #1
1200 REM "cc1-36. dat" PREPARATION
1220 DIM theta(3000)
1240 OPEN "cc1.dat" FOR INPUT AS #1
1260 d = 1
1280 IF EOF(1) THEN 1360
1300 INPUT #1, theta(d)
1320 d = d + 1
1340 GOTO 1280
1360 CLOSE #1
1380 REM "bs.dat" PREPARATION
1400 DIM b1(3000)
1420 OPEN "bs.dat" FOR INPUT AS #1
1440 OPEN "ep.dat" FOR OUTPUT AS #2
1460 e = 1
1480 IF EOF(1) THEN 1560
```

```
1500 INPUT #1, b1(e)
1520 e = e + 1
1540 GOTO 1480
1560 CLOSE #1
1580 REM ITEM PREPARATION
1600 DIM bb(4, 100)
1620 FOR i = 1 TO 4
1640 IF i = 1 THEN OPEN "bb1.dat" FOR INPUT AS #1
1660 IF i = 2 THEN OPEN "bb2.dat" FOR INPUT AS #1
1680 IF i = 3 THEN OPEN "bb3.dat" FOR INPUT AS #1
1700 IF i = 4 THEN OPEN "bb4.dat" FOR INPUT AS #1
1720 g = 1
1740 IF EOF(1) THEN 1820
1760 INPUT #1, bb(i, g)
1780 g = g + 1
1800 GOTO 1740
1820 CLOSE #1
1840 NEXT i
1860 REM ADJUSTED FBB EP AS ZERO
1880 DIM fbb(4, 100)
1900 DIM ep(4, 100)
1920 FOR i = 1 TO 4: REM LEVEL
1940 IF i = 1 THEN gg = 100
1960 IF i = 2 THEN gg = 100
1980 IF i = 3 THEN gg = 100
2000 IF i = 4 THEN gg = 100
2020 FOR g = 1 TO gg
2040 fbb(i, g) = 0
2060 ep(i, g) = 1
2080 NEXT g
2100 NEXT i
```

```

2120 REM ENTRY POINT DETERMINATION
2140 FOR d = 1 TO 3000
2160 a = 1
2180 c = .2
2200 big = .5
2220 snu = 0
2240 sud = 0
2260 the = theta(d)
2265 REM FIND THE NEAREST VALUE FOR THETA-BB
2270 k = 1
2275 min = ABS(the - b1(k))
2280 FOR e = 1 TO 3000
2285 diff0 = ABS(the - b1(e))
2290 IF min < diff0 THEN GOTO 2310
2295 IF min >= diff0 THEN k = e
2300 IF min >= diff0 THEN min = diff0
2305 ee = e
2310 NEXT e
2320 dev1 = a * (the - b1(ee))
2340 phat1 = 1! / (1! + EXP(-dev1))
2360 pt1 = c + (1 - c) * phat1
2380 REM PROTECT AGAINST DIVIDE BY ZERO
2400 IF (pt1 < .00001) THEN pt1 = .00001
2420 IF (pt1 > .99999) THEN pt1 = .99999
2440 RANDOMIZE TIMER: rnum1 = RND(1#)
2460 IF rnum1 >= pt1 THEN uij1 = 1
2480 IF rnum1 < pt1 THEN uij1 = 0
2500 IF uij1 = 1 THEN b2 = b1(ee) + 2
2520 IF uij1 = 0 THEN b2 = b1(ee) - 2
2540 IF b2 > 3 THEN b2 = 3
2560 IF b2 < -3 THEN b2 = -3

```

```

2580 dev2 = a * (the - b2)
2600 phat2 = 1! / (1! + EXP(-dev2))
2620 pt2 = c + (1! - c) * phat2
2640 REM PROTECT AGAINST DIVIDE BY ZERO
2660 IF (pt2 < .00001) THEN pt2 = .00001
2680 IF (pt2 > .99999) THEN pt2 = .99999
2700 RANDOMIZE TIMER: rnum2 = RND(1#)
2720 IF rnum2 >= pt2 THEN uij2 = 1
2740 IF rnum2 < pt2 THEN uij2 = 0
2760 IF uij2 = 1 THEN b3 = b2 + 2
2780 IF uij2 = 0 THEN b3 = b2 - 2
2800 IF b3 > 3 THEN b3 = 3
2820 IF b3 < -3 THEN b3 = -3
2840 dev3 = a * (the - b3)
2860 phat3 = 1! / (1! + EXP(-dev3))
2880 pt3 = c + (1! - c) * phat3
2900 REM PROTECT AGAINST DIVIDE BY ZERO
2920 IF (pt3 < .00001) THEN pt3 = .00001
2940 IF (pt3 > .99999) THEN pt3 = .99999
2960 RANDOMIZE TIMER: rnum3 = RND(1#)
2980 IF rnum3 >= pt3 THEN uij3 = 1
3000 IF rnum3 < pt3 THEN uij3 = 0
3020 IF (uij1 = 0) AND (uij2 = 0) AND (uij3 = 0) THEN GOTO 2220
3040 IF (uij1 = 1) AND (uij2 = 1) AND (uij3 = 1) THEN GOTO 2220
3060 zz = 0
3080 dev = a * (the - b1(ee))
3100 phat = 1! / (1! + EXP(-dev))
3120 pt = c + (1! - c) * phat
3140 REM PROTECT AGAINST DEVIDE BY ZERO
3160 IF (pt < .00001) THEN pt = .00001
3180 IF (pt > .99999) THEN pt = .99999

```

```

3200 wij = pt * (1 - pt)
3220 vij = uij1 - pt
3240 psp = phat / pt
3260 snu = snu + a * vij * psp
3280 sud = sud + a ^ 2 * wij * psp * psp
3300 dev = a * (the - b2)
3320 phat = 1! / (1! + EXP(-dev))
3340 pt = c + (1! - c) * phat
3360 REM PROTECT AGAINST DEVIDE BY ZERO
3380 IF (pt < .00001) THEN pt = .00001
3400 IF (pt > .99999) THEN pt = .99999
3420 wij = pt * (1 - pt)
3440 vij = uij2 - pt
3460 psp = phat / pt
3480 snu = snu + a * vij * psp
3500 sud = sud + a ^ 2 * wij * psp * psp
3520 dev = a * (the - b3)
3540 phat = 1! / (1! + EXP(-dev))
3560 pt = c + (1! - c) * phat
3580 REM PROTECT AGAINST DEVIDE BY ZERO
3600 IF (pt < .00001) THEN pt = .00001
3620 IF (pt > .99999) THEN pt = .99999
3640 wij = pt * (1 - pt)
3660 vij = uij3 - pt
3680 psp = phat / pt
3700 snu = snu + a * vij * psp
3720 sud = sud + a ^ 2 * wij * psp * psp
3740 REM PROTECT AGAINST DEVIDE BY ZERO
3760 IF (sud < .00001) THEN sud = .00001
3780 delta = snu / sud
3800 IF ABS(delta < big) GOTO 3860

```

3820 REM PROTECT AGAINST BIG CHANGE IN THETA

3840 IF delta > 0! THEN

 delta = big

 ELSE

 delta = -big

 END IF

3860 the = ftheta(d) + delta

3880 IF ABS(delta < .05) THEN GOTO 4000

3900 zz = zz + 1

3920 IF zz = 5 GOTO 4000

3940 snu = 0

3960 sud = 0

3980 GOTO 3080

4000 IF the > 3! THEN the = 3!

4020 IF the < -3! THEN the = -3!

4040 REM DATA PREPARATION

4060 DIM ever(4, 100)

4080 FOR i = 1 TO 4

4100 FOR g = 1 TO 100

4120 ever(i, g) = 0

4140 NEXT g

4160 NEXT i

4180 c = 0!

4200 f = .5

4220 FOR i = 1 TO 4: REM LEVEL

4240 IF i = 1 THEN gg = 100

4260 IF i = 2 THEN gg = 100

4280 IF i = 3 THEN gg = 100

4300 IF i = 4 THEN gg = 100

4320 a = f

4340 tl = 5


```

4360 FOR l = 1 TO tl: REM TEST LENGHT
4380 REM FIND THE NEAREST VALUE FOR THETAS-BB
4400 k = 1
4420 min = ABS(the - bb(i, k))
4440 FOR g = 1 TO gg
4460 IF ever(i, g) = 1 THEN GOTO 4580
4480 diff0 = ABS(the - bb(i, g))
4500 IF min < diff0 THEN GOTO 4580
4520 IF min >= diff0 THEN k = g
4540 IF min >= diff0 THEN min = diff0
4560 gx = g
4580 NEXT g
4600 ever(i, gx) = 1
4620 REM a-STR EXAMINATION
4640 dev = a * (the - bb(i, gx))
4660 phat = 1! / (1! + EXP(-dev))
4680 pt = c + (1! - c) * phat
4700 REM PROTECT AGAINST DIVIDE BY ZERO
4720 IF (pt < .00001) THEN pt = .00001
4740 IF (pt > .99999) THEN pt = .99999
4760 RANDOMIZE TIMER: rnum = RND(1#)
4780 IF rnum >= pt THEN uij = 1
4800 IF rnum < pt THEN uij = 0
4820 zx = 0
4840 dev = a * (the - bb(i, gx))
4860 phat = 1! / (1! + EXP(-dev))
4880 pt = c + (1! - c) * phat
4900 REM PROTECT AGAINST DIVIDE BY ZERO
4920 IF (pt < .00001) THEN pt = .00001
4940 IF (pt > .99999) THEN pt = .99999
4960 wij = pt * (1 - pt)

```

```

4980 vij = uij - pt
5000 psp = phat / pt
5020 snu = snu + a * vij * psp
5040 sud = sud + a ^ 2 * wij * psp * psp
5060 REM PROTECT AGAINST DIVIDE BY ZERO
5080 IF (sud < .00001) THEN sud = .00001
5100 delta = snu / sud
5120 IF ABS(delta < big) GOTO 5180
5140 REM PROTECT AGAINST BIG CHANGE IN THETA
5160 IF delta > 0! THEN
    delta = big
ELSE
    delta = -big
END IF
5180 the = ftheta(d) + delta
5200 IF ABS(delta < .05) THEN GOTO 5320
5220 zx = zx + 1
5240 IF zx = 5 GOTO 5320
5260 snu = 0
5280 sud = 0
5300 GOTO 4840
5320 IF the > 3! THEN the = 3!
5340 IF the < -3! THEN the = -3!
5360 REM CALCULATE FREQUENCY
5380 IF ever(i, gx) = 1 THEN fbb(i, gx) = fbb(i, gx) + 1
5400 eb = fbb(i, gx) / 3000
5420 IF eb <= .2 THEN ep(i, gx) = ep(i, gx) * 1.04
5440 IF ep(i, gx) > 1 THEN ep(i, gx) = 1
5460 IF eb > .2 THEN ep(i, gx) = ep(i, gx) * .95
5480 g = gx
5500 NEXT i

```

```
5520 f = f + .5
5540 NEXT i
5560 NEXT d
5580 FOR i = 1 TO 4
5600 IF i = 1 THEN gg = 100
5620 IF i = 2 THEN gg = 100
5640 IF i = 3 THEN gg = 100
5660 IF i = 4 THEN gg = 100
5680 FOR g = 1 TO gg
5700 PRINT TAB(1); "i="; i; TAB(20); "g="; g; TAB(40); "ep="; ep(i, g)
5720 PRINT #2, ep(i, g)
5740 NEXT g
5760 NEXT i
5780 CLOSE #2
5800 END
```

โปรแกรมภาษากิวเบสิก (Q-Basic)
การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ตามระดับชั้นของค่าอำนาจจำแนก
ร่วมกับการควบคุมการแสดงผลข้อสอบ
สำหรับขนาดคลังข้อสอบ 400 ข้อและความยาวแบบทดสอบ 20 ข้อ

```

1000 CLS
1010 REM "ss1-72. dat" PREPARATION
1020 DIM ftheta(5000)
1030 OPEN "ss1.dat" FOR INPUT AS #1
1040 d = 1
1050 IF EOF(1) THEN 1090
1060 INPUT #1, ftheta(d)
1070 d = d + 1
1080 GOTO 1050
1090 CLOSE #1
1100 REM "sc1-72. dat" PREPARATION
1110 DIM theta(5000)
1120 OPEN "sc1.dat" FOR INPUT AS #1
1130 d = 1
1140 IF EOF(1) THEN 1180
1150 INPUT #1, theta(d)
1160 d = d + 1
1170 GOTO 1140
1180 CLOSE #1
1190 REM "bs.dat" PREPARATION
1200 DIM b1(3000)
1210 OPEN "bs.dat" FOR INPUT AS #1
1220 e = 1

```

```
1230 IF EOF(1) THEN 1270
1240 INPUT #1, b1(e)
1250 e = e + 1
1260 GOTO 1230
1270 CLOSE #1
1280 REM ITEM PREPARATION
1290 DIM bb(4, 100)
1300 FOR i = 1 TO 4
1310 IF i = 1 THEN OPEN "bb1.dat" FOR INPUT AS #1
1320 IF i = 2 THEN OPEN "bb2.dat" FOR INPUT AS #1
1330 IF i = 3 THEN OPEN "bb3.dat" FOR INPUT AS #1
1340 IF i = 4 THEN OPEN "bb4.dat" FOR INPUT AS #1
1350 g = 1
1360 IF EOF(1) THEN 1400
1370 INPUT #1, bb(i, g)
1380 g = g + 1
1390 GOTO 1360
1400 CLOSE #1
1410 NEXT i
1420 REM "ep.dat" PREPARATION
1430 DIM ep(4, 100)
1440 OPEN "ep.dat" FOR INPUT AS #1
1450 FOR i = 1 TO 4
1460 IF i = 1 THEN gg = 100
1470 IF i = 2 THEN gg = 100
1480 IF i = 3 THEN gg = 100
1490 IF i = 4 THEN gg = 100
1500 g = 1
1510 IF EOF(1) THEN 1580
1520 INPUT #1, ep(i, g)
1530 IF g = gg THEN GOTO 1570
```

```
1540 g = g + 1
1560 GOTO 1510
1570 NEXT i
1580 CLOSE #1
1590 REM ADJUSTED FREQUENCY AS ZERO
1600 DIM ffb(4, 100)
1610 FOR i = 1 TO 4: REM LEVEL
1620 IF i = 1 THEN gg = 100
1630 IF i = 2 THEN gg = 100
1640 IF i = 3 THEN gg = 100
1650 IF i = 4 THEN gg = 100
1660 FOR g = 1 TO gg
1670 ffb(i, g) = 0
1680 NEXT g
1690 NEXT i
1700 REM ADJUSTED DV AS ZERO
1710 FOR i = 1 TO 4
1720 sdiff(i) = 0
1730 mdiff(i) = 0
1740 share(i) = 0
1750 un(i) = 0
1760 ov(i) = 0
1770 NEXT i
1780 REM ENTRY POINT DETERMINATION
1790 FOR d = 1 TO 5000
1800 a = 1
1810 c = .2
1820 big = .5
1830 snu = 0
1840 sud = 0
1850 the = theta(d)
```

```

1852 REM FIND THE NEAREST VALUE FOR THETA-BB
1854 k = 1
1856 min = ABS(the - b1(k))
1858 FOR e = 1 TO 3000
1860 diff0 = ABS(the - b1(e))
1862 IF min < diff0 THEN GOTO 1875
1864 IF min >= diff0 THEN k = e
1866 IF min >= diff0 THEN min = diff0
1870 ee = e
1875 NEXT e
1880 dev1 = a * (the - b1(ee))
1890 phat1 = 1! / (1! + EXP(-dev1))
1900 pt1 = c + (1! - c) * phat1
1910 REM PROTECT AGAINST DIVIDE BY ZERO
1920 IF (pt1 < .00001) THEN pt1 = .00001
1930 IF (pt1 > .99999) THEN pt1 = .99999
1940 RANDOMIZE TIMER: rnum1 = RND(1#)
1950 IF rnum1 >= pt1 THEN uij1 = 1
1960 IF rnum1 < pt1 THEN uij1 = 0
1970 IF uij1 = 1 THEN b2 = b1(ee) + 2
1980 IF uij1 = 0 THEN b2 = b1(ee) - 2
1990 IF b2 > 3 THEN b2 = 3
2000 IF b2 < -3 THEN b2 = -3
2010 dev2 = a * (the - b2)
2020 phat2 = 1! / (1! + EXP(-dev2))
2030 pt2 = c + (1! - c) * phat2
2040 REM PROTECT AGAINST DIVIDE BY ZERO
2050 IF (pt2 < .00001) THEN pt2 = .00001
2060 IF (pt2 > .99999) THEN pt2 = .99999
2070 RANDOMIZE TIMER: rnum2 = RND(1#)

```

```

2080 IF rnum2 >= pt2 THEN uij2 = 1
2090 IF rnum2 < pt2 THEN uij2 = 0
2100 IF uij2 = 1 THEN b3 = b2 + 2
2110 IF uij2 = 0 THEN b3 = b2 - 2
2120 IF b3 > 3 THEN b3 = 3
2130 IF b3 < -3 THEN b3 = -3
2140 dev3 = a * (the - b3)
2150 phat3 = 1! / (1! + EXP(-dev3))
2160 pt3 = c + (1! - c) * phat3
2170 REM PROTECT AGAINST DIVIDE BY ZERO
2180 IF (pt3 < .00001) THEN pt3 = .00001
2190 IF (pt3 > .99999) THEN pt3 = .99999
2200 RANDOMIZE TIMER: rnum3 = RND(1#)
2210 IF rnum3 >= pt3 THEN uij3 = 1
2220 IF rnum3 < pt3 THEN uij3 = 0
2230 IF (uij1 = 0) AND (uij2 = 0) AND (uij3 = 0) THEN GOTO 1830
2240 IF (uij1 = 1) AND (uij2 = 1) AND (uij3 = 1) THEN GOTO 1830
2250 zz = 0
2260 dev = a * (the - b1(ee))
2270 phat = 1! / (1! + EXP(-dev))
2280 pt = c + (1! - c) * phat
2290 REM PROTECT AGAINST DEVIDE BY ZERO
2300 IF (pt < .00001) THEN pt = .00001
2310 IF (pt > .99999) THEN pt = .99999
2320 wij = pt * (1 - pt)
2330 vij = uij1 - pt
2340 psp = phat / pt
2350 snu = snu + a * vij * psp
2360 sud = sud + a ^ 2 * wij * psp * psp
2370 dev = a * (the - b2)
2380 phat = 1! / (1! + EXP(-dev))

```



```

2390 pt = c + (1! - c) * phat
2400 REM PROTECT AGAINST DEVIDE BY ZERO
2410 IF (pt < .00001) THEN pt = .00001
2420 IF (pt > .99999) THEN pt = .99999
2430 wij = pt * (1 - pt)
2440 vij = uij2 - pt
2450 psp = phat / pt
2460 snu = snu + a * vij * psp
2470 sud = sud + a ^ 2 * wij * psp * psp
2480 dev = a * (the - b3)
2490 phat = 1! / (1! + EXP(-dev))
2500 pt = c + (1! - c) * phat
2510 REM PROTECT AGAINST DEVIDE BY ZERO
2520 IF (pt < .00001) THEN pt = .00001
2530 IF (pt > .99999) THEN pt = .99999
2540 wij = pt * (1 - pt)
2550 vij = uij3 - pt
2560 psp = phat / pt
2570 snu = snu + a * vij * psp
2580 sud = sud + a ^ 2 * wij * psp * psp
2590 REM PROTECT AGAINST DEVIDE BY ZERO
2600 IF (sud < .00001) THEN sud = .00001
2610 delta = snu / sud
2620 IF ABS(delta < big) GOTO 2650
2630 REM PROTECT AGAINST BIG CHANGE IN THETA
2640 IF delta > 0! THEN
    delta = big
ELSE
    delta = -big
END IF
2650 the = ftheta(d) + delta

```

```
2670 IF ABS(delta < .05) THEN GOTO 2730
2680 zz = zz + 1
2690 IF zz = 5 GOTO 2730
2700 snu = 0
2710 sud = 0
2720 GOTO 2260
2730 IF the > 3! THEN the = 3!
2740 IF the < -3! THEN the = -3!
2750 REM DATA PREPARATION
2760 DIM ve(4, 100)
2770 FOR i = 1 TO 4
2780 FOR g = 1 TO 100
2790 ve(i, g) = 0
2800 NEXT g
2810 NEXT i
2820 c = 0!
2830 pl = 400: REM ITEM POOL
2840 f = .5
2850 FOR i = 1 TO 4: REM LEVEL
2860 IF i = 1 THEN gg = 100
2870 IF i = 2 THEN gg = 100
2880 IF i = 3 THEN gg = 100
2890 IF i = 4 THEN gg = 100
2900 a = f
2910 tl = 5
2920 FOR l = 1 TO tl: REM TEST LENGHT
2930 REM FIND THE NEAREST VALUE FOR THETA-BB
2940 k = 1
2950 ms = ABS(the - bb(i, k))
2960 FOR g = 1 TO gg
2970 IF ve(i, g) = 1 THEN GOTO 3030
```

```

2980 diff0 = ABS(the - bb(i, g))
2990 IF ms < diff0 THEN GOTO 3030
3000 IF ms >= diff0 THEN k = g
3010 IF ms >= diff0 THEN ms = diff0
3020 gx = g
3030 NEXT g
3040 RANDOMIZE TIMER: ran = RND(1#)
3050 IF ran < ep(i, gx) THEN GOTO 3560
3060 IF ran > ep(i, gx) THEN GOTO 3070
3070 si = 1
3080 IF gx = 1 THEN GOTO 3120
3090 IF gx = gg THEN GOTO 3170
3100 IF ABS(bb(i, gx) - bb(i, gx - si)) < ABS(bb(i, gx + si) - bb(i, gx)) THEN GOTO 3220
3110 IF ABS(bb(i, gx) - bb(i, gx - si)) > ABS(bb(i, gx + si) - bb(i, gx)) THEN GOTO 3400
3120 gx = gx + si
3130 IF ve(i, gx) = 1 THEN GOTO 3120
3140 RANDOMIZE TIMER: ran = RND(1#)
3150 IF ran < ep(i, gx) THEN GOTO 3560
3160 IF ran > ep(i, gx) THEN GOTO 3120
3170 gx = gx - si
3180 IF ve(i, gx) = 1 THEN GOTO 3170
3190 RANDOMIZE TIMER: ran = RND(1#)
3200 IF ran < ep(i, gx) THEN GOTO 3560
3210 IF ran > ep(i, gx) THEN GOTO 3170
3220 gx = gx - si
3240 IF gx = 1 THEN GOTO 3120
3250 IF ve(i, gx) = 1 THEN si = si + 1
3260 IF ve(i, gx) = 1 THEN GOTO 3310
3270 RANDOMIZE TIMER: ran = RND(1#)
3280 IF ran < ep(i, gx) THEN GOTO 3560
3290 IF ran > ep(i, gx) THEN si = si + 1

```

```

3300 IF ran > ep(i, gx) THEN GOTO 3310
3310 gx = gx + si
3320 IF gx = gg THEN GOTO 3170
3340 IF ve(i, gx) = 1 THEN si = si + 1
3350 IF ve(i, gx) = 1 THEN GOTO 3220
3360 RANDOMIZE TIMER: ran = RND(1#)
3370 IF ran < ep(i, gx) THEN GOTO 3560
3380 IF ran > ep(i, gx) THEN si = si + 1
3390 IF ran > ep(i, gx) THEN GOTO 3220
3400 gx = gx + si
3410 IF gx = gg THEN GOTO 3170
3420 IF ve(i, gx) = 1 THEN si = si + 1
3430 IF ve(i, gx) = 1 THEN GOTO 3480
3440 RANDOMIZE TIMER: ran = RND(1#)
3450 IF ran < ep(i, gx) THEN GOTO 3560
3460 IF ran > ep(i, gx) THEN si = si + 1
3470 IF ran > ep(i, gx) THEN GOTO 3480
3480 gx = gx - si
3490 IF gx = 1 THEN GOTO 3120
3500 IF ve(i, gx) = 1 THEN si = si + 1
3510 IF ve(i, gx) = 1 THEN GOTO 3400
3520 RANDOMIZE TIMER: ran = RND(1#)
3530 IF ran < ep(i, gx) THEN GOTO 3560
3540 IF ran > ep(i, gx) THEN si = si + 1
3550 IF ran > ep(i, gx) THEN GOTO 3400
3560 ve(i, gx) = 1
3570 REM a-STR EXAMINATION
3580 dev = a * (the - bb(i, gx))
3590 phat = 1! / (1! + EXP(-dev))
3600 pt = c + (1! - c) * phat
3610 REM PROTECT AGAINST DIVIDE BY ZERO

```

```

3620 IF (pt < .00001) THEN pt = .00001
3630 IF (pt > .99999) THEN pt = .99999
3640 RANDOMIZE TIMER: rnum = RND(1#)
3650 IF rnum >= pt THEN uij = 1
3660 IF rnum < pt THEN uij = 0
3670 zx = 0
3680 dev = a * (the - bb(i, gx))
3690 phat = 1! / (1! + EXP(-dev))
3700 pt = c + (1! - c) * phat
3710 REM PROTECT AGAINST DEVIDE BY ZERO
3720 IF (pt < .00001) THEN pt = .00001
3730 IF (pt > .99999) THEN pt = .99999
3740 wij = pt * (1 - pt)
3750 vij = uij - pt
3760 psp = phat / pt
3770 snu = snu + a * vij * psp
3780 sud = sud + a ^ 2 * wij * psp * psp
3790 REM PROTECT AGAINST DEVIDE BY ZERO
3800 IF (sud < .00001) THEN sud = .00001
3810 delta = snu / sud
3820 IF ABS(delta < big) GOTO 3850
3830 REM PROTECT AGAINST BIG CHANGE IN THETA
3840 IF delta > 0! THEN
    delta = big
ELSE
    delta = -big
END IF
3850 the = ftheta(d) + delta
3860 IF ABS(delta < .05) THEN GOTO 3920
3870 zx = zx + 1
3880 IF zx = 5 GOTO 3920

```

```

3890 snu = 0
3900 sud = 0
3910 GOTO 3680
3920 IF the > 3! THEN the = 3!
3930 IF the < -3! THEN the = -3!
3940 REM CALCULATE FREQUENCY
3950 IF ve(i, gx) = 1 THEN ffb(i, gx) = ffb(i, gx) + 1
3960 g = gx
3970 NEXT i
3980 REM DEPENDENT VARIABLE VALUES
3990 REM BIAS CALCULATION
4000 sdiff(i) = sdiff(i) + (the - ftheta(d))
4010 IF d = 5000 THEN bias(i) = sdiff(i) / 5000
4020 IF d = 5000 THEN PRINT TAB(1); "level = "; i; TAB(15); "bias = "; bias(i); TAB(40);
    "sdiff = "; sdiff(i)
4030 IF d = 5000 THEN LPRINT TAB(1); "level = "; i; TAB(15); "bias = "; bias(i); TAB(40);
    "sdiff = "; sdiff(i)
4040 REM MSE CALCULATION
4050 mdiff(i) = mdiff(i) + ((the - ftheta(d)) ^ 2)
4060 IF d = 5000 THEN mse(i) = mdiff(i) / 5000
4070 IF d = 5000 THEN PRINT TAB(1); "level = "; i; TAB(15); "mse = "; mse(i); TAB(40);
    "mdiff = "; mdiff(i)
4080 IF d = 5000 THEN LPRINT TAB(1); "level = "; i; TAB(15); "mse = "; mse(i); TAB(40);
    "mdiff = "; mdiff(i)
4090 IF d <> 5000 THEN GOTO 4270
4100 REM ITEM EXPOSE CALCULATION
4110 FOR g = 1 TO gg
4120 tg = (ffb(i, g) / 5000)
4130 IF tg <= .2 THEN un(i) = un(i) + 1
4140 IF tg > .2 THEN ov(i) = ov(i) + 1
4150 NEXT g

```

```

4160 PRINT TAB(1); "level = "; i; TAB(15); "underrate = "; un(i); TAB(40); "overrate = "; ov(i)
4170 LPRINT TAB(1); "level = "; i; TAB(15); "underrate = "; un(i); TAB(40); "overrate = "; ov(i)
4180 REM OVERLAP CALCULATION
4190 FOR g = 1 TO gg
4200 sv = FIX(ffb(i, g) / 2)
4210 share(i) = share(i) + sv
4220 NEXT g
4230 np(i) = (tl * d * (d - 1) / 2)
4240 overlap(i) = share(i) / np(i)
4250 PRINT TAB(1); "level = "; i; TAB(15); "share = "; share(i); TAB(40); "np = "; np(i);
      TAB(60); "overlap = "; overlap(i)
4260 LPRINT TAB(1); "level = "; i; TAB(15); "share = "; share(i); TAB(40); "np = "; np(i);
      TAB(60); "overlap = "; overlap(i)
4270 f = f + .5
4280 NEXT i
4290 NEXT d
4300 REM CHI-SQUARE CALCULATION
4310 schi = 0
4320 tr = ((tl * 4) / pl)
4330 FOR i = 1 TO 4
4340 FOR g = 1 TO gg
4350 tg = (ffb(i, g) / 5000)
4360 wr = ((tg - tr) ^ 2) / tr
4370 schi = schi + wr
4380 NEXT g
4390 NEXT i
4400 PRINT TAB(1); "chi-square = "; schi
4410 LPRINT TAB(1); "chi-square = "; schi
4420 END

```

ประวัติย่อผู้วิจัย

ประวัติย่อผู้วิจัย

ชื่อ-ชื่อสกุล	นางสาวสุภารัตน์ หวลมุกดา
วันเดือนปีเกิด	31 ธันวาคม 2503
สถานที่เกิด	กรุงเทพมหานคร
สถานที่อยู่ปัจจุบัน	681 ถนนรามอินทรา เขตคันนายาว กรุงเทพฯ 10230
ตำแหน่งหน้าที่การงานปัจจุบัน	วิทยากรฯ 7
สถานที่ทำงานปัจจุบัน	วิทยาลัยพยาบาลบรมราชชนนี นพรัตน์วชิระ สถาบันพระบรมราชชนก สำนักงานปลัดกระทรวง กระทรวงสาธารณสุข
ประวัติการศึกษา	
พ.ศ. 2524	ปศ. (พยาบาลอนามัยและผดุงครรภ์) จากวิทยาลัยพยาบาลเกื้อการุณย์
พ.ศ. 2527	วทบ. (โภชนวิทยา) จากมหาวิทยาลัยมหิดล
พ.ศ. 2534	วทม. (โภชนศาสตร์) จากมหาวิทยาลัยมหิดล
พ.ศ. 2550	กศด. (การทดสอบและวัดผลการศึกษา) จากมหาวิทยาลัยศรีนครินทรวิโรฒ