การเปรียบเทียบคุณภาพของการกำหนดมาตรฐาน ระหว่างวิธีแองกอฟที่ได้รับการปรับปรุงกับวิธีบุ๊คมาร์ค

นางสุรีพร อนุศาสนนันท์

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาครุศาสตรดุษฎีบัณฑิต
สาขาวิชาการวัดและประเมินผลการศึกษา ภาควิชาวิจัยและจิตวิทยาการศึกษา
คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2550

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

A COMPARISON OF THE QUALITY OF STANDARD SETTING BETWEEN THE MODIFIED ANGOFF METHOD AND THE BOOKMARK METHOD

Mrs. Sureeporn Anusasananun

A Dissertation Submitted in Partial Fulfillment of the Requirements

for the Degree of Doctor of Philosophy Program in Educational Measurement and Evaluation

Department of Educational Research and Psychology

Faculty of Education

Chulalongkorn University

Academic year 2007

Copyright of Chulalongkorn University

หัวข้อวิทยานิพนธ์	การเปรียบเทียบคุณภาพของการกำหนดมาตรฐานระหว่างวิธีแองกอฟที่ได้รับการปรับปรุงกับวิธีบุ๊คมาร์ค
โดย	นางสุรีพร อนุศาสนนันท์
สาขาวิชา	การวัดและประเมินผลการศึกษา
อาจารย์ที่ปรึกษา	รองศาสตราจารย์ ดร.ศิริเดช สุชีวะ
อาจารย์ที่ปรึกษาร่วม	ดร. ชูศักดิ์ ขัมภลิขิต

คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาดุษฎีบัณฑิต

..... คณบดีคณะครุศาสตร์
(รองศาสตราจารย์ ดร.พฤทธิ์ ศิริบรรณพิทักษ์)

คณะกรรมการสอบวิทยานิพนธ์

.....ประธานกรรมการ
(ศาสตราจารย์ ดร.ศิริชัย กาญจนวาสี)

.....อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
(รองศาสตราจารย์ ดร. ศิริเดช สุชีวะ)

.....อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม
(ดร. ชูศักดิ์ ขัมภลิขิต)

.....กรรมการภายนอกมหาวิทยาลัย
(รองศาสตราจารย์ ดร.พรทิพย์ ไชยโส)

.....กรรมการ
(รองศาสตราจารย์ ดร.เอมอร จังศิริพรปกรณ์)

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สุรีพร อนุศาสนนันท์ : การเปรียบเทียบคุณภาพของการกำหนดมาตรฐานระหว่างวิธีแองกอฟที่ได้รับการปรับปรุงกับวิธีบุ๊คมาร์ค. (A COMPARISON OF THE QUALITY OF STANDARD SETTING BETWEEN THE MODIFIED ANGOFF METHOD AND THE BOOKMARK METHOD). อ. ที่ปรึกษา : รศ. ดร. ศิริเดช สุชีวะ, อ. ที่ปรึกษาร่วม : ดร. ชูศักดิ์ ขัมภลิขิต, 401 หน้า.

การวิจัยครั้งนี้มีวัตถุประสงค์ 1) เพื่อเปรียบเทียบคุณภาพของการกำหนดมาตรฐานระหว่างวิธีการแองกอฟที่ได้รับการปรับปรุงกับวิธีการบุ๊คมาร์ค โดยพิจารณาจาก ความตรง และความเที่ยง ความเหมาะสมและความเป็นไปได้ 2) เพื่อศึกษาผลของจำนวนผู้ตัดสิน และจำนวนครั้งที่มีต่อการกำหนดมาตรฐานด้วยวิธีการแองกอฟที่ได้รับการปรับปรุง และวิธีการบุ๊คมาร์ค กลุ่มตัวอย่างประกอบด้วย ครูผู้สอนวิชาคณิตศาสตร์ ซึ่งเป็นผู้ตัดสินจำนวน 12 คน และนักเรียนชั้นมัธยมศึกษาปีที่ 3 ปีการศึกษา 2549 จำนวน 987 คน ในจังหวัดชลบุรี เครื่องมือที่ใช้ในการวิจัยคือ แบบสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ภาคเรียนที่ 1 จำนวน 4 ฉบับ รวม 100 ข้อ ผลการสอบนำมาวิเคราะห์ค่าความยาก อำนาจจำแนก โดยใช้ทฤษฎีตอบสนองข้อสอบ ข้อมูลการกำหนดมาตรฐานทั้งสองวิธีจากผู้ตัดสินนำมาวิเคราะห์หาค่าความเที่ยงโดยใช้ทฤษฎีสรุปอ้างอิง วิเคราะห์ความตรงคือ หาสหสัมพันธ์ ส่วนความเหมาะสมและความเป็นไปได้ในการนำไปใช้วิเคราะห์โดย มัธยฐาน และพิสัยระหว่างควอไทล์

ผลการวิจัยพบว่า

1. แบบสอบวัดผลสัมฤทธิ์ทางการเรียน มีคะแนนจุดตัดที่กำหนดด้วยวิธีแองกอฟที่ได้รับการปรับปรุง 7 ระดับคือ ระดับดีเยี่ยม (A) ระดับดีมาก (B+) ระดับดี (B) ระดับดีพอใช้ (C+) ระดับพอใช้ (C) ระดับอ่อน (D+) และระดับอ่อนมาก (D) ครั้งที่ 1 มีค่า 72.86, 64.31, 56.18, 43.37, 33.59, 21.34 และ 12.98 ตามลำดับ ครั้งที่ 2 มีค่า 73.28, 63.74, 55.28, 45.11, 35.67, 22.94, 12.51 ตามลำดับ ครั้งที่ 3 มีค่า 73.49, 63.25, 52.82, 52.82, 41.46, 31, 19.07, 11.72 ตามลำดับ และมีคะแนนจุดตัดด้วยวิธีบุ๊คมาร์ค ครั้งที่ 1 มีค่า 80.42, 65.75 47.5, 33.17, 24, 14.75, 7.83 ตามลำดับ ครั้งที่ 2 เท่ากับ 80.58, 66.67, 47.5, 33.75, 24.58, 14.75, 7.83 ตามลำดับ และครั้งที่ 3 มีค่า 80.92, 64.08, 47.33, 34, 24.08, 14.83, 7.92 ตามลำดับ

2. การกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์คมีค่าความเที่ยงสูงกว่าวิธีแองกอฟที่ได้รับการปรับปรุงอย่างมีนัยสำคัญทางสถิติที่ระดับนัยสำคัญ .01 ทั้ง 7 ระดับ

3. ค่าความตรงของการกำหนดมาตรฐาน โดยพิจารณาค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างระดับผลการเรียนที่นักเรียนได้รับจากโรงเรียน กับระดับมาตรฐานที่กำหนดด้วยวิธีแองกอฟที่ได้รับการปรับปรุง เท่ากับ .661 - .678 อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างระดับผลการเรียนที่นักเรียนได้รับจากโรงเรียน กับระดับมาตรฐานที่กำหนดด้วยวิธีบุ๊คมาร์ค เท่ากับ .533 อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และเมื่อทดสอบความแตกต่างที่ระดับนัยสำคัญทางสถิติ .05 พบว่า ไม่แตกต่างกัน

4. ค่าเฉลี่ยความคิดเห็นของผู้ตัดสินเกี่ยวกับความเหมาะสมและความเป็นไปได้ในการนำไปใช้ของวิธีบุ๊คมาร์คมากกว่าวิธีแองกอฟที่ได้รับการปรับปรุงอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

5. ค่าความเที่ยงของการกำหนดมาตรฐานมีค่าเพิ่มขึ้น เมื่อจำนวนผู้ตัดสิน และจำนวนครั้งเพิ่มขึ้น

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาควิชา.....วิจัยและจิตวิทยาการศึกษา..... ลายมือชื่อนิสิต.....

สาขาวิชา.....การวัดและประเมินผลการศึกษา... ลายมือชื่ออาจารย์ที่ปรึกษา.....

ปีการศึกษา.....2550..... ลายมือชื่ออาจารย์ที่ปรึกษาร่วม.....

##4684659127 : MAJOR EDUCATIONAL MEASUREMENT AND EVALUATION
KEY WORDS: STANDARD SETTING / BOOKMARK METHOD / MODIFIED ANGOFF METHOD / CUT SCORES

SUREEPORN ANUSASANANUN: A COMPARISON OF THE QUALITY OF STANDARD SETTING BETWEEN MODIFIED ANGOFF AND BOOKMARK STANDARD SETTING METHODS. THESIS ADVISOR: ASSOC. PROF. SIRIDEJ SUJIVA, Ph.D., THESIS COADVISOR: CHOOSAK KHAMPALIKIT, Ph.D., 401pp.

The objectives of this study are 1) to compare the quality of setting standards assessed by the modified Angoff method with that of the Bookmark standard-setting method, in terms of validity, reliability, propriety, and feasibility, and 2) to investigate the effect of the number of judges and the number of occations on the standard setting of the two methods. The research sample is comprised of 12 mathematics teachers (also serving as judges) and 987 Matthayom Suksa 3 students, Academic Year B.E. 2549, in Chonburi. The research instruments include 4 multiplechoice mathematical achievement tests for the students' first semester, totaling altogether 100 items. Their scores were taken into thorough analysis—namely a) degree of difficulty and b) power of discrimination—on the basis of the IRT model. The data on the two standard-setting approaches from these teachers were analyzed in relation to reliability under the generalizability theory and analyzed in validity under correlation. As for their propriety and feasibility, the analysis of their median as well as inter-quartile range was performed.

The following are the research findings:

1. The achievement tests attain 7 levels of cut scores, i.e. excellent (A), very good (B+), good (B), rather good (C+), fair (C), poor (D+), and very good (D). Via the modified Angoff method, the scores range from 72.86, 64.31, 56.18, 43.37, 33.59, 21.34 to 12.98 respectively for the first occation, from 73.28, 63.74, 55.28, 45.11, 35.67, 22.94 to 12.51 respectively for the second occation, and from 73.49, 63.25, 52.82, 52.82, 41.46, 31, 19.07 to 11.72 respectively for the third occation. By the Bookmark method, the scores range from 80.42, 65.75, 47.5, 33.17, 24, 14.75 to 7.83 respectively for the first occation, from 80.58, 66.67, 47.5, 33.75, 24.58, 14.75 to 7.83 respectively for the second occation, and from 80.92, 64.08, 47.33, 34, 24.08, 14.83 to 7.92 respectively for the third occation.

2. The reliability derived from the Bookmark scheme is significantly higher at the .01 level for all the 7 cut score ranks than that from the modified Angoff technique.

3. The correlation coefficients between the students usual achievement scores and their standardized scores obtained from the Angoff approach are .661-.678 and are statistically significant at the .01 level. The correlation coefficients between their usual scores provided by the school and those determined by the Bookmark method are .533 and carry a .01 significance but no significance at all at the .05 level.

4. The mean of the judges' opinion-ranking scores as to the propriety and feasibility of the Bookmark approach is significantly higher that that of the modified Angoff scheme.

5. The greater the number of judges and number of occations, the higher the reliability coefficients.

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Department Educational Research and Psychology Student's signature
Field of study Educational Measurement and Evaluation Advisor's signature
Academic year 2007 Co-advisor's signature

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ด้วยความกรุณาจาก รองศาสตราจารย์ ดร.ศิริเดช สุชีวะ อาจารย์ที่ปรึกษาวิทยานิพนธ์ และ ดร.ชูศักดิ์ ขัมภลิขิต อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม ที่ให้คำปรึกษาและข้อเสนอแนะที่เป็นประโยชน์อย่างยิ่งในการทำวิจัยด้วยดีตลอดมา ผู้วิจัยมีความซาบซึ้งในความกรุณาของท่านอาจารย์ และขอกราบขอบพระคุณอย่างสูงมา ณ โอกาสนี้

ขอกราบขอบพระคุณ ศาสตราจารย์ ดร.ศิริชัย กาญจนวาสี ประธานกรรมการสอบวิทยานิพนธ์ รองศาสตราจารย์ ดร.เอมอร จังศิริพรปกรณ์ ผู้ทรงคุณวุฒิ และรองศาสตราจารย์ ดร.พรทิพย์ ไชยโส ผู้ทรงคุณวุฒิภายนอก ที่ช่วยให้ข้อเสนอแนะที่เป็นประโยชน์ต่อการปรับปรุงแก้ไขวิทยานิพนธ์ให้มีความสมบูรณ์มากยิ่งขึ้น รวมทั้งคณาจารย์ภาควิชาวิจัยและจิตวิทยาการศึกษาที่ได้ประสิทธิ์ประสาทวิชาความรู้ต่างๆ ให้แก่ผู้วิจัย และขอขอบคุณบัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัยที่ให้ทุนอุดหนุนการเดินทางไปเผยแพร่ผลงานวิจัยในที่ประชุม Asia Pacific Educational Research Association International Conference 2006 ณ ประเทศฮ่องกง

ขอขอบคุณ รองศาสตราจารย์ ดร.ไพรัตน์ วงษ์นาม ที่ให้คำแนะนำและให้คำปรึกษาเกี่ยวกับวิธีการใช้โปรแกรม GENOVA ขอบคุณ Prof.Dr.Larry Nelson ที่ให้ความอนุเคราะห์ โปรแกรม XCALIBRE for Windows95/NT (Version 1.10) เพื่อใช้ในการศึกษาครั้งนี้

และขอขอบคุณ คุณวิเรขา ปัญจมานนท์ ที่ให้ความช่วยเหลือผู้วิจัยในการเก็บข้อมูล ตลอดจน คุณพิกุล เอกวรางกุร คุณวารุณี ลัภนโชคดี คุณวัชราภรณ์ จิตรมาศ และคุณวันเพ็ญ ผ่องกาย และน้องๆ รุ่นพี่ นิสิตภาควิชาวิจัย และจิตวิทยาการศึกษา ที่ให้ความช่วยเหลือ และเป็นกำลังใจด้วยดีมาตลอด

ท้ายนี้ ขอกราบขอบพระคุณคุณพ่อและคุณแม่ นายอำพล และนางสมานจิต ศิริมาลย์ ผู้เป็นที่รักยิ่ง ผู้ที่เอาใจใส่ ส่งเสริมสนับสนุนและให้กำลังใจด้านการศึกษาแก่ผู้วิจัยมาโดยตลอดตั้งแต่เล็กจนสำเร็จการศึกษาระดับปริญญาเอก และขอบคุณสามี นาวาอากาศเอก ปริญญา อนุศาสนนันท์ ที่คอยช่วยเหลือ ให้กำลังใจ ในขณะที่ผู้วิจัยทำวิทยานิพนธ์

สารบัญ

หน้า

บทคัดย่อภาษาไทย..... ง

บทคัดย่อภาษาอังกฤษ..... จ

กิตติกรรมประกาศ..... ฉ

สารบัญ..... ช

สารบัญตาราง..... ฌ

สารบัญภาพ..... ถ

บทที่

1 บทนำ..... 1

ความเป็นมาและความสำคัญของปัญหา..... 1

คำถามวิจัย..... 13

วัตถุประสงค์การวิจัย..... 14

สมมุติฐานการวิจัย..... 14

ขอบเขตการวิจัย..... 14

คำจำกัดความที่ใช้ในการวิจัย..... 16

ประโยชน์ที่คาดว่าจะได้รับ..... 19

2 เอกสารและงานวิจัยที่เกี่ยวข้อง..... 20

ตอนที่ 1 มโนทัศน์ของมาตรฐาน และคะแนนจุดตัด..... 21

ตอนที่ 2 วิธีการกำหนดมาตรฐาน..... 22

ตอนที่ 3 ทฤษฎีการตอบสนองข้อสอบกับการกำหนดคะแนนจุดตัด..... 47

ตอนที่ 4 การตรวจสอบความเที่ยง และความตรงของการกำหนดมาตรฐาน..... 53

ตอนที่ 5 แนวคิดเกณฑ์ประเมินกระบวนการกำหนดมาตรฐานเพื่อใช้ประเมินความเหมาะสม และความเป็นไปได้ในการนำไปใช้ 63

ตอนที่ 6 งานวิจัยที่เกี่ยวข้อง..... 64

3 วิธีดำเนินการวิจัย..... 73

ขั้นตอนการดำเนินงานวิจัย..... 73

ประชากรและกลุ่มตัวอย่าง..... 76

เครื่องมือที่ใช้ในการวิจัย..... 80

การเก็บรวบรวมข้อมูล..... 106

บทที่		หน้า
	การวิเคราะห์ข้อมูล.....	116
4	ผลการวิเคราะห์ข้อมูล	124
	ตอนที่ 1 ผลการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุงและวิธีบุ๊คมาร์ค.....	125
	ตอนที่ 2 ผลการหาคุณภาพของการกำหนดมาตรฐานด้านความเที่ยง ด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และวิธีบุ๊คมาร์ค จำแนกตามจำนวนผู้ตัดสิน.....	140
	ตอนที่ 3 ผลการหาคุณภาพของการกำหนดมาตรฐานด้านความตรงด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และวิธีบุ๊คมาร์ค.....	207
	ตอนที่ 4 ผลการหาความเหมาะสมและความเป็นไปได้ในการนำไปใช้ด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และวิธีบุ๊คมาร์ค.....	209
	ตอนที่ 5 ผลการเปรียบเทียบความเที่ยงของการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และวิธีบุ๊คมาร์ค จำแนกตามจำนวนผู้ตัดสิน..	227
	ตอนที่ 6 ผลของจำนวนผู้ตัดสิน และจำนวนครั้งที่มีต่อการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และวิธีการบุ๊คมาร์ค.....	245
5	สรุปผล อภิปรายและข้อเสนอแนะ.....	281
	สรุปผลการวิจัย.....	283
	อภิปรายผลการวิจัย.....	288
	ข้อเสนอแนะในการนำผลการวิจัยไปใช้.....	297
	ข้อเสนอแนะในการวิจัยครั้งต่อไป.....	298
รายการอ้างอิง.....		299
ภาคผนวก.....		308
	ภาคผนวก ก.....	309
	ภาคผนวก ข.....	313
	ภาคผนวก ค.....	316
	ภาคผนวก ง.....	342
	ภาคผนวก จ.....	367
	ภาคผนวก ฉ.....	376
ประวัติผู้เขียนวิทยานิพนธ์.....		401

สารบัญตาราง

ตาราง		หน้า
1	งานวิจัยเกี่ยวกับการกำหนดมาตรฐานวิธีการแองกอฟที่ได้รับการปรับปรุงและวิธีบุ๊คมาร์ค	8
2	ตัวอย่างการคำนวณหาคะแนนจุดตัดด้วยวิธีแองกอฟจากเปอร์เซ็นต์การตอบข้อสอบแต่ละข้อถูก ของผู้สอบที่มีความสามารถขั้นต่ำ	25
3	ตัวอย่างการคำนวณหาคะแนนจุดตัดด้วยวิธีแองกอฟที่ได้รับการปรับปรุงจาก. เปอร์เซนต์การตอบข้อสอบแต่ละข้อถูกของผู้สอบที่มีความสามารถขั้นต่ำ	26
4	รายชื่อโรงเรียนและจำนวนกลุ่มตัวอย่างนักเรียนชั้นมัธยมศึกษาปีที่ 3 ที่ใช้เก็บข้อมูลเพื่อทดลองแบบสอบครั้งที่ 1 ครั้งที่ 2 และครั้งที่ 3	78
5	ตารางโครงสร้างกำหนดน้ำหนักความสำคัญของเนื้อหาและวัตถุประสงค์การเรียนรู้ของวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ภาคเรียนที่ 1	83
6	ผลการตรวจสอบความสอดคล้องระหว่างข้อสอบกับจุดประสงค์การเรียนรู้	88
7	ค่าสถิติพื้นฐานของแบบสอบผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ภาคเรียนที่ 1 จำนวน 4 ฉบับ รวม 100 ข้อ จำนวนผู้สอบ 264 คน	90
8	ค่าความยาก และอำนาจจำแนกของแบบสอบผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 จำนวน 4 ฉบับ รวม 100 ข้อ	90
9	ข้อสอบที่ไม่มีคุณภาพตามเกณฑ์ที่กำหนด ต้องได้รับการปรับปรุงข้อสอบ	91
10	ผลการวิเคราะห์ตัวประกอบของแบบสอบเลือกตอบฉบับที่ 1 เรื่องพื้นที่ผิวและปริมาตร	92
11	ผลการวิเคราะห์ตัวประกอบของแบบสอบเลือกตอบฉบับที่ 2 เรื่อง กราฟ	92
12	ผลการวิเคราะห์ตัวประกอบของแบบสอบเลือกตอบฉบับที่ 3 เรื่อง ระบบสมการเชิงเส้น	93
13	ผลการวิเคราะห์ตัวประกอบของแบบสอบเลือกตอบฉบับที่ 4 เรื่อง ความคล้าย	93
14	ผลการวิเคราะห์ตัวประกอบของแบบสอบเลือกตอบทั้งฉบับ	94
15	ค่าสถิติพื้นฐานของแบบสอบผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 จำนวน 4 ฉบับ รวม 100 ข้อ จำนวนผู้สอบทั้งสิ้น 987 คน	95

ตาราง **หน้า**

16 ค่าความยาก (b) ค่าอำนาจจำแนก (a) และค่าการเดา (c) ของแบบสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 จำนวน 4ฉบับ..... 96

17 โครงสร้างแบบสอบถามความเหมาะสม และความเป็นไปได้ในกระบวนการการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และ วิธีการบุ๊คมาร์ค..... 104

18 โครงสร้างแบบสอบถามกระบวนการการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และ วิธีการบุ๊คมาร์ค..... 105

19 คะแนนจุดตัดของผู้ตัดสินระดับดีเยี่ยม (A) ด้วยวิธีแองกอฟที่ได้รับการปรับปรุงและวิธีบุ๊คมาร์ค..... 126

20 คะแนนจุดตัดของผู้ตัดสินระดับดีมาก (B+) ด้วยวิธีแองกอฟที่ได้รับการปรับปรุงและวิธีบุ๊คมาร์ค..... 128

21 คะแนนจุดตัดของผู้ตัดสินระดับดี (B) ด้วยวิธีแองกอฟที่ได้รับการปรับปรุงและวิธีบุ๊คมาร์ค..... 129

22 คะแนนจุดตัดของผู้ตัดสินระดับดีพอใช้ (C+) ด้วยวิธีแองกอฟที่ได้รับการปรับปรุงและวิธีบุ๊คมาร์ค..... 131

23 คะแนนจุดตัดของผู้ตัดสินระดับพอใช้ (C) ด้วยวิธีแองกอฟที่ได้รับการปรับปรุงและวิธีบุ๊คมาร์ค..... 132

24 คะแนนจุดตัดของผู้ตัดสินระดับอ่อน (D+) ด้วยวิธีแองกอฟที่ได้รับการปรับปรุงและวิธีบุ๊คมาร์ค..... 134

25 คะแนนจุดตัดของผู้ตัดสินระดับอ่อนมาก (D) ด้วยวิธีแองกอฟที่ได้รับการปรับปรุงและวิธีบุ๊คมาร์ค..... 135

26 จำนวนนักเรียนผ่านแต่ละระดับมาตรฐานครั้งที่ 1 ครั้งที่ 2 และครั้งที่ 3 ด้วยการกำหนดมาตรฐานวิธีการแองกอฟที่ได้รับการปรับปรุง..... 136

27 จำนวนนักเรียนผ่านแต่ละระดับมาตรฐานครั้งที่ 1 ครั้งที่ 2 และครั้งที่ 3 ด้วยการกำหนดมาตรฐานวิธีการบุ๊คมาร์ค..... 137

ตาราง **หน้า**

28 การวิเคราะห์ความแปรปรวนและการประมาณค่าความแปรปรวนของการกำหนดมาตรฐาน ด้วยวิธีแองกอฟที่ได้รับการปรับปรุง ที่คะแนนจุดตัดระดับดีเยี่ยม (A) ในขั้นสรุปอ้างอิง (G-study) 141

29 การประมาณค่าความแปรปรวนในขั้นการสรุปอ้างอิง (G-study) ขั้นการตัดสินใจ (D-study) และสัมประสิทธิ์การอ้างอิงสรุปของการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุงที่คะแนนจุดตัดระดับดีเยี่ยม (A) จำแนกตามจำนวนผู้ตัดสิน..... 143

30 การวิเคราะห์ความแปรปรวนและการประมาณค่าความแปรปรวนของการกำหนดมาตรฐาน ด้วยวิธีแองกอฟที่ได้รับการปรับปรุง ที่คะแนนจุดตัดระดับดีมาก (B+)ในขั้นสรุปอ้างอิง (G-study) 145

31 การประมาณค่าความแปรปรวนในขั้นการสรุปอ้างอิง (G-study) ขั้นการตัดสินใจ (D-study) และสัมประสิทธิ์การอ้างอิงสรุปของการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุงที่คะแนนจุดตัดระดับดีมาก (B+) จำแนกตามจำนวนผู้ตัดสิน..... 147

32 การวิเคราะห์ความแปรปรวนและการประมาณค่าความแปรปรวนของการกำหนดมาตรฐาน ด้วยวิธีแองกอฟที่ได้รับการปรับปรุง ที่คะแนนจุดตัดระดับดี (B)ในขั้นสรุปอ้างอิง (G-study) 149

33 การประมาณค่าความแปรปรวนในขั้นการสรุปอ้างอิง (G-study) ขั้นการตัดสินใจ (D-study) และสัมประสิทธิ์การอ้างอิงสรุปของการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุงที่คะแนนจุดตัดระดับดี (B) จำแนกตามจำนวนผู้ตัดสิน..... 151

34 การวิเคราะห์ความแปรปรวนและการประมาณค่าความแปรปรวนของการกำหนดมาตรฐาน ด้วยวิธีแองกอฟที่ได้รับการปรับปรุง ที่คะแนนจุดตัดระดับดีพอใช้ (C+) ในขั้นสรุปอ้างอิง (G-study) 153

35 การประมาณค่าความแปรปรวนในขั้นการสรุปอ้างอิง (G-study) ขั้นการตัดสินใจ (D-study) และสัมประสิทธิ์การอ้างอิงสรุปของการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุงที่คะแนนจุดตัดระดับดีพอใช้ (C+) จำแนกตามจำนวนผู้ตัดสิน..... 155

ตาราง **หน้า**

36 การวิเคราะห์ความแปรปรวนและการประมาณค่าความแปรปรวนของการกำหนดมาตรฐาน ด้วยวิธีแองกอฟที่ได้รับการปรับปรุง ที่คะแนนจุดตัดระดับพอใช้ (C) ในขั้นสรุปอ้างอิง (G-study) 157

37 การประมาณค่าความแปรปรวนในขั้นการสรุปอ้างอิง (G-study) ขั้นการตัดสินใจ (D-study) และสัมประสิทธิ์การอ้างอิงสรุปของการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุงที่คะแนนจุดตัดระดับพอใช้ (C) จำแนกตามจำนวนผู้ตัดสิน..... 159

38 การวิเคราะห์ความแปรปรวนและการประมาณค่าความแปรปรวนของการกำหนดมาตรฐาน ด้วยวิธีแองกอฟที่ได้รับการปรับปรุง ที่คะแนนจุดตัดระดับอ่อน (D+) ในขั้นสรุปอ้างอิง (G-study) 161

39 การประมาณค่าความแปรปรวนในขั้นการสรุปอ้างอิง (G-study) ขั้นการตัดสินใจ (D-study) และสัมประสิทธิ์การอ้างอิงสรุปของการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุงที่คะแนนจุดตัดระดับอ่อน (D+) จำแนกตามจำนวนผู้ตัดสิน..... 163

40 การวิเคราะห์ความแปรปรวนและการประมาณค่าความแปรปรวนของการกำหนดมาตรฐาน ด้วยวิธีแองกอฟที่ได้รับการปรับปรุง ที่คะแนนจุดตัดระดับอ่อนมาก (D) ในขั้นสรุปอ้างอิง (G-study) 165

41 การประมาณค่าความแปรปรวนในขั้นการสรุปอ้างอิง (G-study) ขั้นการตัดสินใจ (D-study) และสัมประสิทธิ์การอ้างอิงสรุปของการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุงที่คะแนนจุดตัดระดับอ่อนมาก (D) จำแนกตามจำนวนผู้ตัดสิน..... 167

42 การวิเคราะห์ความแปรปรวนและการประมาณค่าความแปรปรวนของการกำหนดมาตรฐาน ด้วยวิธีบุ๊คมาร์ค ที่คะแนนจุดตัด ระดับดีเยี่ยม (A) ในขั้นสรุปอ้างอิง (G-study) 169

43 การประมาณค่าความแปรปรวนในขั้นการสรุปอ้างอิง (G-study) ขั้นการตัดสินใจ (D-study) และสัมประสิทธิ์การอ้างอิงสรุปของการกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์คที่คะแนนจุดตัดระดับดีเยี่ยม (A) จำแนกตามจำนวนผู้ตัดสิน..... 171

ตาราง **หน้า**

44 การวิเคราะห์ความแปรปรวนและการประมาณค่าความแปรปรวนของการกำหนดมาตรฐาน ด้วยวิธีบุ๊คมาร์ค ที่คะแนนจุดตัด ระดับดีมาก (B+) ในขั้นสรุปอ้างอิง (G-study) 173
45 การประมาณค่าความแปรปรวนในขั้นการสรุปอ้างอิง (G-study) ขั้นการตัดสินใจ (D-study) และสัมประสิทธิ์การอ้างอิงสรุปของการกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์คที่คะแนนจุดตัดระดับดีมาก (B+) จำแนกตามจำนวนผู้ตัดสิน..... 175
46 การวิเคราะห์ความแปรปรวนและการประมาณค่าความแปรปรวนของการกำหนดมาตรฐาน ด้วยวิธีบุ๊คมาร์คที่คะแนนจุดตัด ระดับดี (B) ในขั้นสรุปอ้างอิง (G-study) 177
47 การประมาณค่าความแปรปรวนในขั้นการสรุปอ้างอิง (G-study) ขั้นการตัดสินใจ (D-study) และสัมประสิทธิ์การอ้างอิงสรุปของการกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์คที่คะแนนจุดตัดระดับดี (B) จำแนกตามจำนวนผู้ตัดสิน..... 179
48 การวิเคราะห์ความแปรปรวนและการประมาณค่าความแปรปรวนของการกำหนดมาตรฐาน ด้วยวิธีบุ๊คมาร์ค ที่คะแนนจุดตัด ระดับดีพอใช้ (C+) ในขั้นสรุปอ้างอิง (G-study) 181
49 การประมาณค่าความแปรปรวนในขั้นการสรุปอ้างอิง (G-study) ขั้นการตัดสินใจ (D-study) และสัมประสิทธิ์การอ้างอิงสรุปของการกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์คที่คะแนนจุดตัดระดับดีพอใช้ (C+) จำแนกตามจำนวนผู้ตัดสิน..... 183
50 การวิเคราะห์ความแปรปรวนและการประมาณค่าความแปรปรวนของการกำหนดมาตรฐาน ด้วยวิธีบุ๊คมาร์ค ที่คะแนนจุดตัด ระดับพอใช้ (C) ในขั้นสรุปอ้างอิง (G-study) 185
51 การประมาณค่าความแปรปรวนในขั้นการสรุปอ้างอิง (G-study) ขั้นการตัดสินใจ (D-study) และสัมประสิทธิ์การอ้างอิงสรุปของการกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์คที่คะแนนจุดตัดระดับพอใช้ (C) จำแนกตามจำนวนผู้ตัดสิน..... 187
52 การวิเคราะห์ความแปรปรวนและการประมาณค่าความแปรปรวนของการกำหนดมาตรฐาน ด้วยวิธีบุ๊คมาร์ค ที่คะแนนจุดตัด ระดับอ่อน (D+) ในขั้นสรุปอ้างอิง (G-study) 189
53 การประมาณค่าความแปรปรวนในขั้นการสรุปอ้างอิง (G-study) ขั้นการตัดสินใจ (D-study) และสัมประสิทธิ์การอ้างอิงสรุปของการกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์คที่คะแนนจุดตัดระดับอ่อน (D+) จำแนกตามจำนวนผู้ตัดสิน..... 191

ตาราง **หน้า**

54 การวิเคราะห์ความแปรปรวนและการประมาณค่าความแปรปรวนของการกำหนดมาตรฐาน ด้วยวิธีบุ๊คมาร์ค ที่คะแนนจุดตัด ระดับอ่อนมาก (D) ในขั้นสรุปอ้างอิง (G-study) 193

55 การประมาณค่าความแปรปรวนในขั้นการสรุปอ้างอิง (G-study) ขั้นการตัดสินใจ (D-study) และสัมประสิทธิ์การอ้างอิงสรุปของการกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์คที่คะแนนจุดตัดระดับอ่อนมาก (D) จำแนกตามจำนวนผู้ตัดสิน..... 195

56 ค่าความเที่ยงหรือสัมประสิทธิ์การสรุปอ้างอิงที่กำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และวิธีบุ๊คมาร์ค จำแนกตามจำนวนผู้ตัดสิน..... 197

57 ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างระดับผลการเรียนที่นักเรียนได้รับจากโรงเรียน กับระดับมาตรฐานที่กำหนดด้วยวิธีการแองกอฟที่ได้รับการปรับปรุง และวิธีบุ๊คมาร์ค จำนวน 67 คน 208

58 ผลประเมินของผู้ตัดสินเกี่ยวกับกระบวนการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง..... 211

59 ผลประเมินของผู้ตัดสินเกี่ยวกับกระบวนการกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์ค..... 213

60 ค่ามัธยฐาน (Median : Mdn) และพิสัยระหว่างควอไทล์ (Inter quartile range : IR) จากการประเมินความเหมาะสมและความเป็นไปได้ในการใช้วิธีการ กำหนดคะแนนจุดตัดด้วยวิธีแองกอฟที่ได้รับการปรับปรุง..... 216

61 ค่ามัธยฐาน (Median : Mdn) และพิสัยระหว่างควอไทล์ (Inter quartile range: IR) จากการประเมินความเหมาะสมและความเป็นไปได้ในการใช้วิธีการกำหนดคะแนนจุดตัดด้วยวิธีบุ๊คมาร์ค..... 219

62 การเปรียบเทียบความเหมาะสมระหว่างวิธีแองกอฟที่ได้รับการปรับปรุงและวิธีบุ๊คมาร์ค..... 222

63 การเปรียบเทียบความเป็นไปได้ในการนำไปใช้ระหว่างวิธีแองกอฟที่ได้รับการปรับปรุง และวิธีบุ๊คมาร์ค..... 222

64 การเปรียบเทียบค่าความเที่ยงตามทฤษฎีสรุปอ้างอิง ระหว่างวิธีแองกอฟที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์ค ที่ระดับดีเยี่ยม (A)..... 227

65 การเปรียบเทียบค่าความเที่ยงตามทฤษฎีสรุปอ้างอิง ระหว่างวิธีแองกอฟที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์ค ที่ระดับดีมาก (B+)..... 230

66 การเปรียบเทียบค่าความเที่ยงตามทฤษฎีสรุปอ้างอิง ระหว่างวิธีแองกอฟที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์ค ที่ระดับดี (B)..... 232

ตาราง **หน้า**

67 การเปรียบเทียบค่าความเที่ยงตามทฤษฎีสรุปอ้างอิง ระหว่างวิธีแองกอฟ ที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์ค ที่ระดับดีพอใช้ (C+)..... 235
68 การเปรียบเทียบค่าความเที่ยงตามทฤษฎีสรุปอ้างอิง ระหว่างวิธีแองกอฟ ที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์ค ที่ระดับพอใช้ (C)..... 237
69 การเปรียบเทียบค่าความเที่ยงตามทฤษฎีสรุปอ้างอิง ระหว่างวิธีแองกอฟ ที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์ค ที่ระดับอ่อน (D+)..... 240
70 การเปรียบเทียบค่าความเที่ยงตามทฤษฎีสรุปอ้างอิง ระหว่างวิธีแองกอฟ ที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์ค ที่ระดับอ่อนมาก(D..... 242
71 การวิเคราะห์ความแปรปรวนและการประมาณค่าความแปรปรวน ในขั้น ศึกษา G-study โดยทำการกำหนดมาตรฐานวิธีแองกอฟที่ได้รับการปรับปรุง ที่คะแนนจุดตัดระดับดีเยี่ยม (A)..... 245
72 การเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิง (Genaralizability coefficient) ในขั้นศึกษา D-study โดยทำการกำหนดมาตรฐานวิธีแองกอฟที่ได้รับการ ปรับปรุง มีรูปแบบ i x r x o design ที่คะแนนจุดตัดระดับดีเยี่ยม (A)..... 246
73 การวิเคราะห์ความแปรปรวนและการประมาณค่าความแปรปรวน ในขั้น ศึกษา G-study โดยทำการกำหนดมาตรฐานวิธีแองกอฟที่ได้รับการปรับปรุง ที่คะแนนจุดตัดระดับดีมาก (B+)..... 247
74 การเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิง (Genaralizability coefficient) ในขั้นศึกษา D-study โดยทำการกำหนดมาตรฐานวิธีแองกอฟที่ได้รับ การปรับปรุง มีรูปแบบ i x r x o design ที่คะแนนจุดตัดระดับดีมาก (B+)..... 248
75 การวิเคราะห์ความแปรปรวนและการประมาณค่าความแปรปรวน ในขั้นศึกษา G-study โดยทำการกำหนดมาตรฐานวิธีแองกอฟที่ได้รับการปรับปรุง ที่คะแนนจุดตัดระดับดี (B)..... 249
76 การเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิง (Genaralizability coefficient) ในขั้นศึกษา D-study โดยทำการกำหนดมาตรฐานวิธีแองกอฟที่ได้รับ การปรับปรุง มีรูปแบบ i x r x o design ที่คะแนนจุดตัดระดับดี (B)..... 250
77 การวิเคราะห์ความแปรปรวนและการประมาณค่าความแปรปรวน ในขั้น ศึกษา G-study โดยทำการกำหนดมาตรฐานวิธีแองกอฟที่ได้รับการปรับปรุง ที่คะแนนจุดตัดระดับดีพอใช้ (C+)..... 251

ตาราง **หน้า**

78 การเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิง (Genaralizability coefficient) ในขั้นศึกษา D-study โดยทำการกำหนดมาตรฐานวิธีแองกอฟที่ได้รับการปรับปรุง มีรูปแบบ i x r x o design ที่คะแนนจุดตัดระดับดีพอใช้ (C+)..... 252
79 การวิเคราะห์ความแปรปรวนและการประมาณค่าความแปรปรวน ในขั้นศึกษา G-study โดยทำการกำหนดมาตรฐานวิธีแองกอฟที่ได้รับการปรับปรุงที่คะแนนจุดตัดระดับพอใช้ (C)..... 253
80 การเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิง (Genaralizability coefficient) ในขั้นศึกษา D-study โดยทำการกำหนดมาตรฐานวิธีแองกอฟที่ได้รับการปรับปรุง มีรูปแบบ i x r x o design ที่คะแนนจุดตัดระดับพอใช้ (C)..... 254
81 การวิเคราะห์ความแปรปรวนและการประมาณค่าความแปรปรวน ในขั้นศึกษา G-study โดยทำการกำหนดมาตรฐานวิธีแองกอฟที่ได้รับการปรับปรุงที่คะแนนจุดตัดระดับอ่อน (D+)..... 255
82 การเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิง (Genaralizability coefficient) ในขั้นศึกษา D-study โดยทำการกำหนดมาตรฐานวิธีแองกอฟที่ได้รับการปรับปรุง มีรูปแบบ i x r x o design ที่คะแนนจุดตัดระดับอ่อน (D+)..... 256
83 การวิเคราะห์ความแปรปรวนและการประมาณค่าความแปรปรวน ในขั้นศึกษา G-study โดยทำการกำหนดมาตรฐานวิธีแองกอฟที่ได้รับการปรับปรุงที่คะแนนจุดตัดระดับอ่อนมาก (D)..... 257
84 การเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิง (Genaralizability coefficient) ในขั้นศึกษา D-study โดยทำการกำหนดมาตรฐานวิธีแองกอฟที่ได้รับการปรับปรุง มีรูปแบบ i x r x o design ที่คะแนนจุดตัดระดับอ่อนมาก (D)..... 258
85 การวิเคราะห์ความแปรปรวนและการประมาณค่าความแปรปรวน ในขั้นศึกษา G-study โดยทำการกำหนดมาตรฐานวิธีบุ๊คมาร์ค ที่คะแนนจุดตัดระดับดีเยี่ยม (A)..... 260
86 การเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิง (Genaralizability coefficient) . ในขั้นศึกษา D-study โดยทำการกำหนดมาตรฐานวิธีบุ๊คมาร์ค มีรูปแบบ i x r x o design ที่คะแนนจุดตัดระดับ ดีเยี่ยม (A)..... 261
87 การวิเคราะห์ความแปรปรวนและการประมาณค่าความแปรปรวน ในขั้นศึกษา G-study โดยทำการกำหนดมาตรฐานวิธีบุ๊คมาร์ค ที่คะแนนจุดตัดระดับดีมาก(B+)..... 262

ตาราง **หน้า**

88 การเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิง (Genaralizability coefficient) .
ในขั้นศึกษา D-study โดยทำการกำหนดมาตรฐานวิธีบุ๊คมาร์ค
มีรูปแบบ i x r x o design ที่คะแนนจุดตัดระดับดีมาก (B+)..... 263
89 การวิเคราะห์ความแปรปรวนและการประมาณค่าความแปรปรวน
ในขั้นศึกษา G-study โดยทำการกำหนดมาตรฐานวิธีบุ๊คมาร์ค
ที่คะแนนจุดตัดระดับดี(B)..... 264
90 การเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิง (Genaralizability coefficient) .
ในขั้นศึกษา D-study โดยทำการกำหนดมาตรฐานวิธีบุ๊คมาร์ค
มีรูปแบบ i x r x o design ที่คะแนนจุดตัดระดับดี (B)..... 265
91 การวิเคราะห์ความแปรปรวนและการประมาณค่าความแปรปรวน
ในขั้นศึกษา G-study โดยทำการกำหนดมาตรฐานวิธีบุ๊คมาร์ค
ที่คะแนนจุดตัดระดับดีพอใช้(C+)..... 266
92 การเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิง (Genaralizability coefficient)
ในขั้นศึกษา D-study โดยทำการกำหนดมาตรฐานวิธีบุ๊คมาร์ค
มีรูปแบบ i x r x o design ที่คะแนนจุดตัดระดับดีพอใช้ (C+)..... 267
93 การวิเคราะห์ความแปรปรวนและการประมาณค่าความแปรปรวน
ในขั้นศึกษา G-study โดยทำการกำหนดมาตรฐานวิธีบุ๊คมาร์ค
ที่คะแนนจุดตัดระดับพอใช้(C)..... 268
94 การเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิง (Genaralizability coefficient)
ในขั้นศึกษา D-study โดยทำการกำหนดมาตรฐานวิธีบุ๊คมาร์ค
มีรูปแบบ i x r x o design ที่คะแนนจุดตัดระดับพอใช้ (C)..... 269
95 การวิเคราะห์ความแปรปรวนและการประมาณค่าความแปรปรวน
ในขั้นศึกษา G-study โดยทำการกำหนดมาตรฐานวิธีบุ๊คมาร์ค
ที่คะแนนจุดตัดระดับอ่อน(D+)..... 270
96 การเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิง (Genaralizability coefficient)
ในขั้นศึกษา D-study โดยทำการกำหนดมาตรฐานวิธีบุ๊คมาร์ค
มีรูปแบบ i x r x o design ที่คะแนนจุดตัดระดับอ่อน (D+)..... 271
97 การวิเคราะห์ความแปรปรวนและการประมาณค่าความแปรปรวน
ในขั้นศึกษา G-study โดยทำการกำหนดมาตรฐานวิธีบุ๊คมาร์ค
ที่คะแนนจุดตัดระดับอ่อนมาก (D)..... 272

ตาราง **หน้า**

98 การเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิง (Genaralizability coefficient) . ในขั้นศึกษา D-study โดยทำการกำหนดมาตรฐานวิธีบุ๊คมาร์ค มีรูปแบบ i x r x o design ที่คะแนนจุดตัดระดับอ่อนมาก (D)..... 273

99 ค่าความเที่ยงหรือสัมประสิทธิ์การสรุปอ้างอิงที่กำหนดมาตรฐานด้วย วิธีแองกอฟที่ได้รับการปรับปรุง และวิธีบุ๊คมาร์ค มีรูปแบบ i x r x o design..... 274

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญภาพ

ภาพประกอบ		หน้า
1	กรอบแนวคิดในการวิจัย	16
2	คะแนนการแจกแจงผู้สอบในกลุ่มคาบเส้น	32
3	คะแนนการแจกแจงของผู้สอบกลุ่มที่ต่ำและสูงกว่ามาตรฐาน	33
4	คะแนนการแจกแจงกลุ่มผู้สอบคาบเส้นสำหรับระดับการปฏิบัติหลายระดับ	42
5	ขั้นตอนการดำเนินงานวิจัย	75
6	โค้งสารสนเทศของแบบสอบฉบับที่ 1 เรื่อง พื้นที่ผิวและปริมาตร	98
7	โค้งลักษณะแบบสอบฉบับที่ 1 เรื่อง พื้นที่ผิวและปริมาตร	98
8	โค้งสารสนเทศของแบบสอบฉบับที่ 2 เรื่อง กราฟ	99
9	โค้งลักษณะแบบสอบฉบับที่ 2 เรื่อง กราฟ	99
10	โค้งสารสนเทศของแบบสอบ ฉบับที่3 ระบบสมการเชิงเส้น	100
11	โค้งลักษณะแบบสอบ ฉบับที่3 ระบบสมการเชิงเส้น	100
12	โค้งสารสนเทศของแบบสอบ ฉบับที่ 4 ความคล้าย	101
13	โค้งลักษณะแบบสอบ ฉบับที่ 4 ความคล้าย	101
14	โค้งสารสนเทศของแบบสอบรวม 4 ฉบับ 100 ข้อ	102
15	โค้งลักษณะแบบสอบรวม 4 ฉบับ 100 ข้อ	102
16	ลำดับขั้นการพิจารณาความสามารถของนักเรียน 8 ระดับ	110
17	ตัวอย่างรูปแบบหน้าในคู่มือจัดเรียงข้อสอบ	112
18	ตัวอย่างการเรียงข้อสอบจากข้อง่ายสุดไปยังข้อยากสุดในคู่มือจัดเรียงข้อสอบ	113
19	ตัวอย่างการกำหนดบุ๊คมาร์ค (การคั่นหนังสือ) 7 ระดับ	115
20	การเปรียบเทียบค่าความเที่ยงของการกำหนดมาตรฐานด้วยวิธีแองกอฟ ที่ได้รับการปรับปรุงกับวิธีบุ๊คมาร์ค จำแนกตามจำนวนผู้ตัดสินที่คะแนนจุดตัดระดับดีเยี่ยม (A)	200
21	การเปรียบเทียบค่าความเที่ยงของการกำหนดมาตรฐานด้วยวิธีแองกอฟ ที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์คจำแนกตามจำนวนผู้ตัดสินที่คะแนนจุดตัดระดับดีมาก (B+)	201
22	การเปรียบเทียบค่าความเที่ยงของการกำหนดมาตรฐานด้วยวิธีแองกอฟ ที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์ค จำแนกตามจำนวนผู้ตัดสินที่คะแนนจุดตัดระดับดี (B)	202

ภาพประกอบ **หน้า**

23 การเปรียบเทียบค่าความเที่ยงของการกำหนดมาตรฐานด้วยวิธีแองกอฟ ที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์ค จำแนกตามจำนวนผู้ตัดสินที่คะแนนจุดตัดระดับดีพอใช้ (C+)..... 203
24 การเปรียบเทียบค่าความเที่ยงของการกำหนดมาตรฐานด้วยวิธีแองกอฟ ที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์ค จำแนกตามจำนวนผู้ตัดสินที่คะแนนจุดตัดระดับพอใช้ (C)..... 204
25 การเปรียบเทียบค่าความเที่ยงของการกำหนดมาตรฐานด้วยวิธีแองกอฟ ที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์ค จำแนกตามจำนวนผู้ตัดสินที่คะแนนจุดตัดระดับอ่อน (D+)..... 205
26 การเปรียบเทียบค่าความเที่ยงของการกำหนดมาตรฐานด้วยวิธีแองกอฟ ที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์ค จำแนกตามจำนวนผู้ตัดสินที่คะแนนจุดตัดระดับอ่อนมาก (D)..... 206
27 การเปรียบเทียบค่าสัมประสิทธิ์สรุปอ้างอิงของการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์ค ที่คะแนนจุดตัดระดับดีเยี่ยม (A)..... 277
28 การเปรียบเทียบค่าสัมประสิทธิ์สรุปอ้างอิงของการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์ค ที่คะแนนจุดตัดระดับดีมาก (B+)..... 277
29 การเปรียบเทียบค่าสัมประสิทธิ์สรุปอ้างอิงของการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์ค ที่คะแนนจุดตัดระดับดี (B)..... 278
30 การเปรียบเทียบค่าสัมประสิทธิ์สรุปอ้างอิงของการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์ค ที่คะแนนจุดตัดระดับดีพอใช้ (C+)..... 278
31 การเปรียบเทียบค่าสัมประสิทธิ์สรุปอ้างอิงของการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์ค ที่คะแนนจุดตัดระดับพอใช้ (C)..... 279
32 การเปรียบเทียบค่าสัมประสิทธิ์สรุปอ้างอิงของการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์ค ที่คะแนนจุดตัดระดับอ่อน (D+)..... 279

ภาพประกอบ **หน้า**

33 การเปรียบเทียบค่าสัมประสิทธิ์สรุปอ้างอิงของการกำหนดมาตรฐาน
ด้วยวิธีแองกอฟที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์ค ที่คะแนนจุดตัด
ระดับอ่อนมาก (D)..... 280

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 1

บทนำ

ความเป็นมาและความสำคัญของปัญหา

การวัดผลการเรียนการสอนมี 2 ลักษณะคือ การวัดผลแบบอิงกลุ่ม และการวัดผลแบบอิงเกณฑ์ การวัดผลแบบอิงกลุ่ม เป็นการวัดความสามารถของบุคคลโดยการเปรียบเทียบกับคนอื่นๆ ที่วัดด้วยแบบสอบฉบับเดียวกันว่า มีมาตรฐานอยู่ในระดับใดของกลุ่ม แต่การวัดผลแบบอิงกลุ่มมีข้อจำกัดที่สำคัญ คือ ไม่สามารถแปลความหมายเพื่อบ่งบอกระดับผลสัมฤทธิ์ของนักเรียนแต่ละคนตามจุดประสงค์ของเนื้อหาหลักสูตรว่าได้บรรลุตามที่ต้องการหรือไม่ จากข้อจำกัดดังกล่าว จึงมีการพัฒนาการวัดผลแบบอิงเกณฑ์ขึ้นมาเพื่อการพัฒนาความสามารถของนักเรียน หลักการของการวัดผลแบบอิงเกณฑ์เป็นการวัดผลเพื่อสำรวจความก้าวหน้าและวินิจฉัยความสามารถทางการเรียนของแต่ละคนตามวัตถุประสงค์ที่กำหนดไว้ เป็นการวัดเพื่อจะแยกผู้ที่ได้รับการเรียนรู้แล้วเป็น ผู้รอบรู้ (Mastery) ออกจากผู้ที่ไม่รอบรู้ (Nonmastery) ดังนั้นในการวัดผลแบบอิงเกณฑ์ สิ่งที่ต้องนำมากำหนดเพื่อใช้ในการพิจารณาตัดสินแยกความสามารถของผู้เรียนว่ามีความรอบรู้หรือไม่คือ “การกำหนดมาตรฐาน” (Standard Setting)

สำหรับความเป็นมาของการกำหนดมาตรฐาน เริ่มมาจากมีการทดสอบแบบอิงเกณฑ์ที่มีมาตั้งแต่ ค.ศ.1951 เมื่อฟลานาแกน (Flanagan) ได้อธิบายคำว่า “เกณฑ์ปกติ” และ “มาตรฐาน” ฟลานาแกนอธิบายถึงการทดสอบแบบอิงเกณฑ์ในลักษณะที่ผู้เรียนจะแปลหนังสือได้โดยผิดไม่เกิน 3 แห่ง อีเบล(Ebel,1962) ก็มีแนวคิดเช่นเดียวกันว่า การทดสอบใดๆ ก็ตามจะมีความหมายอย่างสมบูรณ์ก็ต่อเมื่อผลการสอบนั้นต้องบอกได้ว่า ผู้สอบมีความรู้อยู่ในระดับใดของเนื้อหาทั้งหมดที่ทำการสอบ ในปี ค.ศ. 1962 เกลเซอร์และเคลาซ์ (Glaser and Klaus) เป็นผู้ใช้คำว่า การวัดผลแบบอิงเกณฑ์ (Criterion–reference Measurement) เป็นครั้งแรก (Berk,1980 อ้างถึงใน สุรินทร์ แพ่งจันทึก, 2528)

การศึกษาค้นคว้าในช่วงแรกๆ การใช้คะแนนจุดตัด มักใช้การกำหนดขึ้นจากการสร้างระดับเกณฑ์ที่สูงๆ ไว้ และใช้เทคนิคแบบธรรมดาในการกำหนดระดับเกณฑ์ขึ้นมา เป็นต้นว่า การกำหนดมาตรฐานความรอบรู้โดยใช้เปอร์เซนต์การตอบถูก 80% 85% 90% หรือ 100% การกำหนดระดับเกณฑ์ดังกล่าวมักจะใช้การคาดคะเนจากความรู้สึก โดยไม่คำนึงถึงสถานการณ์เงื่อนไขอื่นๆ ที่อาจแทรกซ้อนเข้ามาตลอดจนคุณสมบัติของข้อสอบ เช่น ระดับความยากที่มีผลต่อการกำหนดระดับเกณฑ์เป็นอย่างมาก

การกำหนดระดับเกณฑ์ในยุคเริ่มต้น เกิดจากการเคลื่อนไหวของการเขียนจุดประสงค์การสอนในรูปจุดประสงค์เชิงพฤติกรรม เมเกอร์ (Mager, 1962 อ้างถึงใน บุญเชิด ภิญโญอนันตพงษ์ , 2526) เป็นผู้นำในการเขียนจุดประสงค์เชิงพฤติกรรม ได้เสนอการกำหนดมาตรฐานโดยเขียนเป็นองค์ประกอบหนึ่งของจุดประสงค์เชิงพฤติกรรม ตัวอย่างเช่น

"นักเรียนต้องสามารถแก้สมการเชิงเส้นได้ถูกต้องภายในเวลา 30 นาที"

จากตัวอย่างจะเห็นว่า การกำหนดระดับเกณฑ์ว่า "ถูกต้องภายใน 30 นาที" เป็นการกำหนดมาตรฐานการปฏิบัติที่ใช้ภาษาแต่ไม่ใช่ค่าปริมาณที่แท้จริง

นักวัดผลอีกท่านหนึ่งคือ บลูม (Bloom, 1968 อ้างถึงใน บุญเชิด ภิญโญอนันตพงษ์, 2526) มีส่วนเข้ามาเกี่ยวข้องกับการพัฒนาการกำหนดมาตรฐาน โดยเสนอแนวคิดเกี่ยวกับการเรียนการสอนเพื่อรอบรู้ว่า "ถ้าให้เวลาอย่างเพียงพอ และให้ความช่วยเหลือที่เหมาะสมแล้ว นักเรียน 95% ของทั้งหมดจะสามารถเรียนรู้เนื้อหาวิชาได้ถึงระดับความรอบรู้ที่กำหนดไว้ได้" อย่างไรก็ตามแนวความคิดของบลูมก็ไม่ได้ให้วิธีการกำหนดระดับความรอบรู้ที่ชัดเจนไว้

ผู้นำทางการทดสอบแบบอิงเกณฑ์อีกท่านคือ โพแฮม(Popham, 1969) ได้ยืนยันซ้ำถึงความสำคัญของมาตรฐานการปฏิบัติว่า "การสร้างมาตรฐานการปฏิบัติโดยกำหนดเป็นระดับผลสัมฤทธิ์ของนักเรียนขั้นต่ำก่อนที่จะทำการสอน จะเป็นประโยชน์อย่างมากต่อการวางแผนประเมินผลการสอน" และได้ยกตัวอย่างคล้ายเมเกอร์ดังนี้

"นักเรียนต้องสามารถแก้โจทย์ปัญหาเกี่ยวกับเส้นรอบรูปได้ถูกต้อง10 ใน 15 ข้อ"

นอกจากนี้ยังมีนักประเมินผลอีกท่านหนึ่งคือ ไทเลอร์ (Tyler, 1973 อ้างถึงใน บุญเชิด ภิญโญอนันตพงษ์, 2526) ได้เน้นถึงการกำหนดมาตรฐานในการชี้บอกระดับความรอบรู้ตามโปรแกรม และได้ให้ตัวอย่างไว้ดังนี้

"นักเรียนที่แสดงว่ามีความรอบรู้ในเนื้อหาวิชาเฉพาะที่เรียนรู้ไปแล้ว จะต้องมีความสามารถในการตอบปัญหาที่กำหนดให้ได้ถูกต้อง 85% ของทั้งหมด ภายในเวลาที่กำหนด"

การนำเอาระดับเกณฑ์ ระบุไว้ในจุดประสงค์การสอนดังกล่าว เป็นการกำหนดเกณฑ์โดยใช้ภาษาธรรมดาซึ่งยังไม่เป็นค่าเชิงปริมาณที่แท้จริง และลักษณะการกำหนดระดับเกณฑ์มักจะใช้ประสบการณ์ในการคาดคะเน และเป็นการกำหนดตามประเพณีนิยมมากกว่าจะเป็นกระบวนการทางวิทยาศาสตร์ จึงได้มีวิธีการกำหนดค่าเกณฑ์ในรูปคะแนนจุดตัดที่เป็นระบบขึ้นมา

วิธีการในการกำหนดคะแนนจุดตัดมีด้วยกันหลายวิธี เช่น วิธีการนับลดจาก 100% (counting backwards from 100%) ซึ่งกลาส (Glass, 1978) เสนอไว้โดยให้ผู้เชี่ยวชาญหรือครูผู้สอนเป็นผู้พิจารณาข้อสอบและจุดประสงค์เชิงพฤติกรรมว่า จุดประสงค์เชิงพฤติกรรมแต่ละ

ข้อมีความสำคัญมากน้อยเพียงใด แล้วพิจารณากำหนดคะแนนจุดตัดของแบบทดสอบจากเกณฑ์ที่ตั้งไว้ ซึ่งแล้วแต่ว่าผู้ตั้งเกณฑ์จะลดลงมากี่เปอร์เซนต์ วิธีการนี้เป็นการกำหนดเกณฑ์ที่ค่อนข้างเป็นอัตนัย เนื่องจากใช้ความคิดเห็นส่วนบุคคลในการพิจารณา ต่อมาจึงมีผู้พัฒนาวิธีการกำหนดคะแนนจุดตัดโดยการพิจารณาเนื้อหาวิชาของแบบสอบเป็นรายข้อ แล้วระบุว่าผู้สอบที่มีความสามารถหรือสมรรถภาพขั้นต่ำที่จะยอมรับว่าผ่านได้นั้นควรทำแบบสอบได้เป็นจำนวนเท่าไร ก็จะทำให้ได้คะแนนจุดตัดที่เหมาะสมขึ้น ซึ่งวิธีการดังกล่าวนี้คือ การพิจารณาตัดสินจากความสามารถต่ำสุดของผู้สอบ (judging minimal competence) ตามที่กลาส (Glass, 1978) เสนอไว้ หรือเป็นวิธีการพิจารณาจากข้อคำถามในแบบสอบตามที่ลิฟวิงตัน และซีกี (Livingston and Zieky, 1982) เสนอไว้ วิธีการนี้มีหลายเทคนิคด้วยกันคือ เทคนิคของแองกอฟ (Angoff) เทคนิคของอีเบล (Ebel) และเทคนิคของนีเดลสกี (Nedelsky) ทั้ง 3 เทคนิคต่างก็มีวิธีการที่เหมือนกันคือ ผู้เชี่ยวชาญจะต้องพิจารณาถึงความรู้และทักษะของผู้สอบที่อยู่ในระดับต่ำสุดหรือที่เรียกว่า ระดับคาบเส้น (borderline) ว่าสมควรให้ผ่านหรือไม่ ซึ่งเป็นเทคนิคที่สะดวกแก่ผู้ใช้มากเนื่องจากสามารถกำหนดคะแนนจุดตัดได้ทั้งก่อนการสอบ และหลังการสอบเสร็จสิ้นแล้ว โดยที่ผู้เชี่ยวชาญจะต้องพิจารณาตัดสินข้อความในแบบทดสอบร่วมกัน

จะเห็นว่าเทคนิคทั้ง 3 ในทางทฤษฎีเป็นสิ่งที่ดี แต่ในทางปฏิบัติยังมีปัญหาคือ 1) ผู้เชี่ยวชาญแต่ละคนมีความคิดที่แตกต่างกันเกี่ยวกับความสามารถต่ำสุดของผู้สอบ เพราะฉะนั้นคะแนนจุดตัดที่ผู้เชี่ยวชาญกำหนดจะสะท้อนถึงความแตกต่างของผู้สอบด้วย 2) การประมาณความสามารถต่ำสุดของผู้สอบจากการพิจารณาข้อสอบจำนวนมากเช่น จำนวน 50 ข้อ นับว่าเป็นงานที่ยาก เพราะว่าผู้สอบบางคนจะได้คะแนนสูงในบางข้อ และบางคนได้คะแนนต่ำในบางข้อ เพราะฉะนั้นการคาดคะเนจากความสามารถต่ำสุดจึงเป็นสิ่งที่เป็นไปไม่ได้ แต่ถ้าผู้สอบที่อยู่ในระดับความสามารถต่ำสุดอยู่ในระดับเดียวกัน การประมาณความสามารถต่ำสุดของผู้สอบจากการทำแบบสอบทั้งฉบับจะทำได้ง่ายกว่าการพิจารณาแต่ละข้อคำถาม (Rowley, 1982) ดังนั้น ลิเวลตัน และซิกี (Livingston & Zieky, 1982) จึงเสนอวิธีการกำหนดจุดตัดวิธีใหม่คือ การใช้วิธีกลุ่มตรงกันข้าม (contrasting group method) ซึ่งเป็นวิธีการตัดสินที่ขึ้นอยู่กับผู้สอบมากกว่าการตัดสินเกี่ยวกับข้อคำถามในแบบทดสอบ วิธีการนี้มีข้อดีคือ ผู้ตัดสินจะคุ้นเคยกับการพิจารณาทักษะของผู้เรียน และสามารถพิจารณาได้อย่างมีความหมาย นอกจากนี้ข้อสนเทศที่ได้จากวิธีการนี้จะสอดคล้องกับสภาพความเป็นจริงมากกว่าการพิจารณาจากข้อคำถามในแบบทดสอบ การใช้วิธีกลุ่มตรงกันข้ามในการกำหนดคะแนนจุดตัดมีหลายวิธี วิธีที่ซับซ้อนน้อยที่สุดคือ การเขียนกราฟแสดงการแจกแจงของคะแนนของกลุ่มรอบรู้ และกลุ่มไม่รอบรู้ แล้วพิจารณาเส้นกราฟของผู้สอบทั้งสองกลุ่มตัดกัน ณ จุดใด จุดนั้นก็คือ คะแนนจุดตัด (Mills, 1983)

ในการใช้วิธีกลุ่มตรงข้ามเพื่อกำหนดคะแนนจุดตัดนั้นปัญหาที่พบก็คือ ไม่สามารถแบ่งได้อย่างแน่ชัดระหว่างผู้ที่รอบรู้กับผู้ที่ไม่รอบรู้ ดังนั้น ลิเวลตัน และซิกี (Livingston & Zieky, 1982) ได้เสนอวิธีการกำหนดจุดตัดอีกวิธีหนึ่งคือ วิธีกลุ่มคาบเส้น (boraderline group method) ซึ่งวิธีนี้มาจากแนวคิดที่ว่า มาตรฐานจะเป็นคะแนนที่ได้จากผู้สอบซึ่งมีความสามารถอยู่ในระดับคาบเส้น วิธีการนี้แทนที่จะให้ผู้เชี่ยวชาญเดาถึงความสามารถของผู้สอบที่อยู่ในระดับคาบเส้นก็ใช้วิธีการให้ผู้เชี่ยวชาญระบุความรู้ความสามารถของผู้สอบที่อยู่ในระดับคาบเส้น และพิจารณาคะแนนจุดตัดจากผู้สอบดังกล่าว มาตรฐานที่กำหนดสำหรับการผ่านจะใช้มัธยฐานของคะแนนผู้สอบที่อยู่ในกลุ่มคาบเส้น ประโยชน์สำคัญของวิธีการนี้คือง่ายในการนำไปใช้และง่ายในการอธิบาย (กัญจนา ลินทรัตนศิริกุล, 2534)

เมื่อเข้าสู่ ค.ศ.1990 ในอเมริกาได้เกิดแนวโน้มจากการประเมินระดับประเทศ (large scale) 2 วิธีคือ 1) มีการใช้ข้อสอบที่ให้คะแนนมากกว่า 2 ค่า เพิ่มมากขึ้น 2) มีความต้องการคะแนนจุดตัดที่หลากหลายระดับ และแล้วในเดือนกรกฎาคม ปี 1995 แนวโน้มดังกล่าวได้รับการยืนยันโดยข้อบังคับแห่งชาติสำหรับการดำเนินโปรแกรมที่อยู่ภายใต้ Title I of Elementary and Secondary Education Act of 1965 และได้รับการเปลี่ยนแปลงแก้ไขโดย The Improving America's School Act of 1994 (U.S. Department of Education, 1995, cited in Berk, 1996) ข้อบังคับดังกล่าวต้องการให้มลรัฐต่างๆ ได้ใช้วิธีการที่หลากหลายในการประเมินซึ่งประกอบด้วยการประเมินทักษะการคิดที่ซับซ้อนและความเข้าใจในเนื้อหา และอธิบายระดับการปฏิบัติหรือจุดตัด 3 ระดับคือ ระดับขั้นสูง (advanced) ระดับชำนาญการ (proficient) และระดับชำนาญการบางส่วน (partially proficient) ในการประเมินนักเรียนทุกคน Berk (1996) และ Zieky (2001) ได้กล่าวว่า ได้เกิดสิ่งที่น่าท้าทายในการกำหนดมาตรฐานที่จะนำมาตรฐานไปใช้กับแบบสอบเขียนตอบ แบบสอบภาคปฏิบัติ และการทดสอบปรับเหมาะในคอมพิวเตอร์ นับจากนั้นจึงมีงานวิจัยเกี่ยวกับวิธีการกำหนดมาตรฐานแนวใหม่ออกมาเป็นจำนวนมาก เช่น **วิธีแองกอฟแบบปรับขยาย** (Extented Angoff) (Hambleton and Plake, 1995; Impara and Plake ,1997; Goodwin, 1999; Brandon, 2004) **วิธีการบุ๊คมาร์ค (Bookmark)** (Kiplinger, 1997; Huynh, 1998; Zwick, Senturk, Wang & Barbara, Loomis, 2001; Van der Schoot, 2002; Buckendahl, Smith, Impara and Plake, 2002; Wang, 2003; Green, Trimble and Lewis, 2003; Beretvas, 2004) **วิธีการเส้นภาพเด่น (Dominant Profile)** (Putnam, Pence and Jaeger, 1995; Plake, Hambleton and Jaeger, 1997) **วิธีการใช้การตัดสินเชิงนโยบาย (Judgmental Policy Capturing)** (Jaeger, 1995) เป็นต้น

วิธีแองกอฟแบบปรับขยาย วิธีนี้ถูกพัฒนาสำหรับการประเมินการปฏิบัติที่ให้คะแนนมากกว่า 2 ค่า โดยกำหนดให้คณะผู้ตัดสินทำการประมาณคะแนนคาดหวังของผู้สอบที่คาบเส้นหรือผู้สอบที่เกือบไม่ได้รับการรับรอง (borderline candidates) ในแบบฝึกหัดแต่ละฉบับที่ให้คะแนนมากกว่า 2 ค่า งานของผู้ตัดสินก็คือ กำหนดความน่าจะเป็น (เช่น คะแนนที่คาดหวัง) ที่ผู้สอบคาบเส้นสามารถตอบข้อสอบถูก คะแนนคาดหวังจากแบบฝึกหัดของผู้ตัดสินแต่ละคนสามารถนำไปใช้ในมาตรฐานการปฏิบัติในชุดของแบบฝึดหัด จากนั้นมาตรฐานของผู้ตัดสินจะถูกนำมาเฉลี่ยเพื่อให้ได้มาตรฐานตามที่ต้องการสำหรับใช้ในการรับรอง วิธีการแองกอฟแบบปรับขยายเป็นวิธีที่ง่ายมีความชัดเจน กระบวนการกำหนดน้ำหนักทำให้ผู้ตัดสินกำหนดค่าของคำถามแต่ละข้อแตกต่างกัน ซึ่งการกำหนดน้ำหนักมากในบางคำถามแสดงว่า ผู้ตัดสินให้ความสำคัญกับข้อนั้นมาก อย่างไรก็ตาม วิธีการแองกอฟแบบปรับขยายอาจใช้ไม่ดีเมื่อนำไปใช้กับการปฏิบัติที่เป็นธรรมชาติซึ่งมีความยืดหยุ่นสูง

วิธีการบุ๊คมาร์ค วิธีการบุ๊คมาร์คเป็นการรวมเนื้อหาข้อสอบกับผลการสอบของผู้สอบเข้าด้วยกัน โดยมีคู่มือที่มีการเรียงข้อสอบจากข้องง่ายไปข้อยาก (Mitzel, Lewis, Patz, & Green, 2001, cite in Beretvas, 2004) ซึ่งข้อสอบที่ถูกจัดเรียงตามค่าความยากจากข้อสอบง่ายสุดไปยังข้อสอบที่ยากสุด จะใช้กระบวนการ IRT ในการวิเคราะห์เพื่อให้ทราบถึงค่าความยากของข้อสอบและความสามารถของนักเรียนที่อยู่บนมาตรเดียวกัน (same continuum) ข้อสอบแบบเขียนตอบก็จะถูกนำมาไว้ในมาตรเดียวกับข้อสอบแบบเลือกตอบโดยคะแนนแต่ละคะแนนในข้อสอบแบบเขียนตอบจะถูกนำมากำหนดบนมาตร แล้วจัดเรียงลำดับลงในคู่มือ คู่มือที่ได้จะนำมาให้ผู้ตัดสินพิจารณาประกอบการหาคะแนนจุดตัด โดยผู้ตัดสินจะพิจารณาข้อสอบทีละข้อ จากข้องง่ายไปยังข้อยาก โดยให้ผู้ตัดสินตอบคำถามว่า “ผู้สอบที่คาบเส้นมีโอกาสตอบข้อสอบข้อนั้นถูก 67% หรือไม่” พิจารณาทีละข้อไปเรื่อยๆ จนถึงข้อที่ผู้สอบกลุ่มคาบเส้นไม่มีโอกาสตอบถูก ก็จะใช้ที่คั่นหนังสือคั่นตรงหน้านั้นไว้ จากนั้นนำผลการตัดสินของแต่ละคนมาอภิปรายร่วมกัน เปิดโอกาสให้แต่ละคนหาคะแนนจุดตัดอีกครั้งแล้วจึงนำมาหาค่าเฉลี่ย จะเห็นได้ว่าวิธีนี้ ผู้ตัดสินจะพิจารณาแบบสอบในภาพรวมโดยอยู่บนพื้นฐานการพิจารณาจากผลคะแนนของนักเรียน นอกจากนี้วิธีบุ๊คมาร์คยังถูกออกแบบมาให้ใช้กับข้อสอบแบบเลือกตอบที่ให้คะแนนแบบ 2 ค่า และข้อสอบแบบเขียนตอบที่ให้คะแนนมากกว่า 2 ค่า อย่างเท่าเทียมกันเพราะข้อสอบทุกข้ออยู่บนสเกล IRT เหมือนกัน อย่างไรก็ตามวิธีนี้ควรมีข้อสอบจำนวนมากพอ สำหรับพิจารณานักเรียนที่อยู่ในมาตรฐานระดับก้าวหน้า หรือระดับตก ซึ่งทั้งสองระดับเป็นระดับสุดโต่ง(เก่งมาก กับอ่อนมาก) ซึ่งอาจทำให้เกิด floor and ceiling effect ได้

วิธีการเส้นภาพเด่น (Dominant Profile Method) วิธีนี้จะใช้ยุทธวิธีสร้างฉันทมติ (consensus-building) ผู้ตัดสินจะถูกถามเพื่อให้เกิดกฎที่เป็นนโยบาย (ตัดสินใจ) เพื่อใช้

กำหนดคะแนนงาน หรือ แบบฝึกหัดที่ใช้ประกอบการประเมินเป็นค่าต่ำสุดที่ต้องการให้ผ่าน องค์ประกอบการทดแทน (compensatory) และองค์ประกอบที่เชื่อมโยง (conjunctive) มีความยืดหยุ่นสำหรับกฎการตัดสินใจ ดังนั้นผู้ตัดสินสามารถสร้างกฎการตัดสินใจที่มีลักษณะซับซ้อนเพื่อให้มีความหมายถึงผู้สอบที่มีลักษณะคุณภาพผ่านการประเมิน วิธีนี้มีความซับซ้อนในการใช้มากกว่ามาตรฐานการปฏิบัติที่มีเพียงมาตรฐานเดียว

วิธีการใช้การตัดสินเชิงนโยบาย (Judgmental Policy Capturing Method) วิธีการคือ คณะผู้ตัดสินต้องได้รับการฝึกอย่างเข้มข้นเพื่อให้เข้าใจความหมายของแต่ละคะแนนในแต่ละแบบฝึกหัด เมื่อแสดงเส้นภาพของคะแนนนักเรียนจากทุกแบบฝึกหัดแล้ว ผู้ตัดสินก็จะสร้างรูปภาพทางความคิดของคุณภาพผู้สอบด้านการปฏิบัติ คะแนนเส้นภาพเหล่านี้จะถูกนำมาจำแนกเพื่อกำหนดเป็นคะแนนเส้นภาพลงในแต่ละประเภทที่ระบุข้อความมาตรฐานการปฏิบัตินั้นๆ (เช่น ต่ำมาก ค่อนข้างต่ำ ต่ำเล็กน้อย สูงเล็กน้อย ค่อนข้างสูง สูงมาก) ผลการตัดสินจะถูกนำมาวิเคราะห์ โดยแต่ละประเภทจะถูกนำมากำหนดเป็นค่าตัวเลข กระบวนการทางสถิติ (เช่น การวิเคราะห์การถดถอยพหุ) จะถูกนำมาใช้เพื่อกำหนดความสัมพันธ์ทางคณิตศาสตร์ระหว่างเส้นภาพคะแนนผู้สอบ กับประเภทการปฏิบัติ (performance categories) ความสัมพันธ์ดังกล่าวจะกำหนดน้ำหนักเพื่อนำไปใช้กับคะแนนของผู้สอบในแต่ละแบบฝึกหัดและสำหรับคำนวณคะแนนการปฏิบัติในผู้สอบแต่ละคน คะแนนการปฏิบัติทั้งหมดจะถูกคำนวณสำหรับผู้สอบแต่ละคน เส้นภาพที่เป็นคะแนนแบบฝึกหัดจะนำมาแสดงให้ผู้ตัดสินพิจารณา เส้นภาพเหล่านี้ก็จะถูกจัดเรียงลำดับคะแนนการปฏิบัติสูงสุดไปยังคะแนนการปฏิบัติต่ำสุด ผู้ตัดสินจะถูกถามเพื่อกำหนดคะแนนการปฏิบัติต่ำสุดที่สามารถประเมินผ่านได้ ข้อเสนอแนะจากผู้ตัดสินทั้งหมดจะถูกนำมาจัดเป็นตาราง การแจกแจงค่ามัธยฐานจะถูกนำมากำหนดเป็นมาตรฐานการปฏิบัติ ขั้นตอนทั้งสองจะถูกนำมาทำซ้ำโดยให้โอกาสผู้ตัดสินอธิบายเหตุผลจากข้อเสนอแนะเดิมของพวกเขา มาตรฐานการปฏิบัติขั้นสุดท้ายจะยึดเสียงข้างมากของผู้ตัดสิน วิธีนี้ใช้เวลาในการฝึกผู้ตัดสินเป็นระยะเวลานานและเมื่อใช้เวลาฝึกนาน ค่าใช้จ่ายก็จะสูงตามไปด้วย

จากที่กล่าวมาจะเห็นว่า วิธีการกำหนดมาตรฐานส่วนใหญ่ยังคงเป็นการตัดสินที่ใช้ดุลยพินิจของผู้ตัดสินซึ่งเป็นข้อจำกัดที่สำคัญ จึงได้มีความพยายามพัฒนาปรับปรุงวิธีการกำหนดมาตรฐานขึ้นมาเพื่อลดข้อจำกัดเกี่ยวกับตัวผู้ตัดสินเช่น การให้ข้อมูลสารสนเทศประกอบการตัดสินที่เพียงพอ การจัดอภิปรายกลุ่ม การดำเนินการตัดสินหลายรอบ เป็นต้น

ปัจจุบันวิธีการกำหนดมาตรฐานที่อาศัยดุลยพินิจของผู้ตัดสินที่ได้รับความนิยมใช้กันอย่างกว้างขวางมี 2 วิธีคือ วิธีการกำหนดคะแนนจุดตัดแนวใหม่ด้วยวิธีบุ๊คมาร์ค (Bookmark)

เนื่องจากในสหรัฐอเมริกาวิธีนี้ได้รับความนิยมใช้กันอย่างกว้างขวางใน 28 มลรัฐ (Egan, 2001, cited in Beretvas, 2004) และในประเทศไทยยังไม่มีการศึกษาวิธีการดังกล่าว วิธีนี้มีจุดเด่นคือช่วยจัดระบบการคิดแก่ผู้ตัดสินทำให้ผู้ตัดสินตัดสินได้ง่ายขึ้น (McKinley, Newman, & Wiser, 1996) เนื่องจากมีการจัดเรียงข้อสอบที่เป็นระบบในรูปของคู่มือ ส่วนอีกวิธีหนึ่งที่นิยมใช้คือวิธีการแองกอฟที่ได้รับการปรับปรุง (Modified Angoff) เป็นวิธีที่ได้รับความนิยมใช้กันอย่างกว้างขวางในการกำหนดมาตรฐาน (Shepard, Glaser, Linn, & Bohrnstedt, 1993) เนื่องจากเป็นวิธีที่ใช้ง่ายไม่ต้องจัดเตรียมสารสนเทศมากเหมือนกับวิธีการบุ๊คมาร์ค จากที่กล่าวมาจะเห็นว่าทั้งสองวิธีต่างมีจุดเด่น จุดอ่อนแตกต่างกัน ด้วยเหตุนี้ผู้วิจัยจึงมีความสนใจเปรียบเทียบวิธีการกำหนดมาตรฐานบุ๊คมาร์ค และวิธีการแองกอฟที่ได้รับการปรับปรุงว่าวิธีการใดมีคุณภาพดีกว่ากัน

สำหรับงานวิจัยในต่างประเทศที่เกี่ยวข้องกับการเปรียบเทียบวิธีการกำหนดมาตรฐานวิธีแองกอฟที่การปรับปรุง และวิธีบุ๊คมาร์ค ผู้วิจัยรวบรวมและนำเสนอในตาราง 1 ดังต่อไปนี้

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 1 งานวิจัยเกี่ยวกับการกำหนดมาตรฐานวิธีการแองกอฟที่ได้รับการปรับปรุง และวิธีบุ๊คมาร์ค

ชื่อผู้วิจัย	Buckendahl (2002)	Wang (2003)	Beretvas (2004)
ชื่อเรื่อง	การเปรียบเทียบการกำหนดมาตรฐานด้วยวิธีการแองกอฟและวิธีบุ๊คมาร์ค	การใช้โมเดล Rasch IRT ในการกำหนดมาตรฐาน : วิธีการผังข้อสอบ(item mapping)	การเปรียบเทียบการกำหนดค่าความยากในวิธีการบุ๊คมาร์คภายใต้โมเดล IRT ที่แตกต่างกัน
วัตถุประสงค์	เพื่อเปรียบเทียบจุดตัดที่ใช้วิธีแองกอฟและวิธีบุ๊คมาร์ค	เพื่อกำหนดจุดตัดสำหรับการสอบใบรับรองและประกาศนียบัตรโดยวิธีผังข้อสอบ	เพื่อเปรียบเทียบค่าความยากด้วยวิธีบุ๊คมาร์คที่ใช้ IRT modle ที่ต่างกัน และค่า RP ที่ต่างกันจากการใช้กระบวนการผังข้อสอบ
กลุ่มตัวอย่าง	นักเรียนเกรด 7 จำนวน 448 คน	ผู้สอบขอใบรับรองและประกาศนียบัตร	ฐานข้อมูลนักเรียนเกรด 4 จำนวน 63,533 คน ปี 1998
เครื่องมือ	แบบทดสอบคณิตศาสตร์เกรด 7 แบบเลือกคำตอบที่มีเนื้อหาทางทักษะคณิตศาสตร์ 6 อย่าง	แบบสอบแบบเลือกคำตอบสำหรับ professional licensure or certification examinations โดยมีการสอบ 4 ครั้ง	แบบสอบย่อยคณิตศาสตร์นักเรียนเกรด 4 ของ WASL โดยข้อสอบเป็นรูปแบบผสม 40 ข้อ แบ่งเป็นแบบเลือกตอบ 24 ข้อ และแบบเขียนตอบ 16 ข้อ ซึ่งแบ่งเป็นแบบตอบสั้น 13 ข้อ ให้คะแนนแบบ 0 , 1 และ 2 และข้อสอบแสดงวิธีทำให้คะแนนเป็น 0, 1 , 2 ,3 และ 4 คะแนนรวมเท่ากับ 62 คะแนน

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 1 (ต่อ)

ชื่อผู้วิจัย	Buckendahl (2002)	Wang (2003)	Beretvas (2004)
วิธีดำเนินการวิจัย	1) เชิญคณะผู้ตัดสินมาสัมมนาเชิงปฏิบัติการ 1 วัน เพื่อสร้างความเข้าใจในวิธีการทั้งสองวิธี 2) แบ่งกลุ่มผู้ตัดสินเป็น 2 กลุ่มคือ สำหรับวิธีแองกอฟ และวิธีบุ๊คมาร์ค 3) แต่ละกลุ่มพิจารณาหาจุดตัด 2 รอบ โดยรอบแรกพิจารณาอย่างอิสระ รอบสองอภิปรายร่วมกันและเปิดโอกาสให้ผู้ตัดสินแต่ละคนเปลี่ยนจุดตัดของตน 4) หาค่าเฉลี่ยจุดตัด	1) ผู้ตัดสินร่วมกันอธิบายลักษณะผู้สอบที่มีความสามารถต่ำ (MCC) 2) วิธีแองกอฟ ผู้ตัดสินจะถูกถามว่าเปอร์เซนต์ของ MCC ที่ตอบข้อสอบข้อนี้ถูกคือเท่าไร จากนั้นให้ค่าสถิติของข้อสอบพิจารณาโดยเปิดโอกาสให้ผู้ตัดสินอภิปรายและปรับเปลี่ยนจุดตัดของตน หาค่าเฉลี่ยคะแนนจุดตัด 3) นำผังข้อสอบให้ผู้ตัดสินพิจารณาว่า “MCC มีโอกาสตอบข้อสอบถูกอย่างน้อย .50 หรือไม่” จัดหา p-value และเปิดโอกาสให้อภิปรายและปรับเปลี่ยนคะแนนจุดตัดของตนได้ กระบวนการนี้ถูกทำซ้ำจนกระทั่งคอลัมน์ข้อสอบถูกกำหนดให้เป็นจุดที่ข้อสอบส่วนใหญ่ในคอลัมน์ได้รับฉันทมติว่ามีความน่าจะเป็นที่ MCC จะตอบถูก .50 และระดับตรงกลางของค่าความยากข้อสอบภายใต้คอลัมน์จะถูกกำหนดคะแนนเป็นจุดตัด	นำผลการสอบจากฐานข้อมูลมาวิเคราะห์ข้อมูลตามโมเดล IRT และค่า RP ดังนี้ 1PL-PC RP 1/2 1PL-PC RP 2/3 1PL-PC RP 4/5 3PL-GPC RP 1/2 3PL-GPC RP 2/3 3PL-GPC RP 4/5 3PL-GR RP 1/2 3PL-GR RP 2/3 3PL-GR RP 4/5

ตารางที่ 1 (ต่อ)

ชื่อผู้วิจัย	Buckendahl (2002)	Wang (2003)	Beretvas (2004)
การวิเคราะห์ข้อมูล	1) หาค่าความยากโดย CTT 2) เปรียบเทียบคะแนนจุดตัดและส่วนเบี่ยงเบนมาตรฐาน ของรอบที่ 1 และรอบที่ 2 ในแต่ละวิธี	1) วิเคราะห์ Rasch IRT ด้วยโปรแกรม BIGSTEPS 2) ใช้ G-theory วิเคราะห์หา Inter-judge consistency (i x r) ใช้โปรแกรม GENOVA วิเคราะห์ G และ D studies	นำค่าพารามิเตอร์ซึ่งวิเคราะห์ด้วยโปรแกรม Parscale หรือ Mulitilog มาแทนค่าในสมการเพื่อคำนวณหาค่า BDL (bookmark difficulty locations) ตามโมเดล IRT และ RP ที่แตกต่างกัน โดยใช้โปรแกรม Mathematica ช่วยในการคำนวณ จากนั้นข้อสอบถูกจัดเรียงตามค่าความยากจากน้อยไปมาก เพื่อนำมาเปรียบเทียบกัน โดย spearman rank order correlation จะถูกคำนวณสำหรับแต่ละโมเดลเป็นคู่ๆ กับแต่ละค่า RP สำหรับในกระบวนการบุ๊คมาร์คจะนำ BDL ที่ใช้โมเดล 1PL-PC และ RP2/3 มาใช้ในคู่มือการจัดเรียงข้อสอบ เพื่อให้คณะกรรมการใช้พิจารณาหาคะแนนจุดตัดซึ่งแบ่งได้เป็น 4 ระดับ

จากตารางที่ 1 จะเห็นว่างานวิจัยของเบเรทวาส (Beretvas, 2004) เป็นการนำฐานข้อมูลที่เก็บข้อมูลจากแบบสอบแบบเลือกตอบ และแบบเขียนตอบ มาใช้วิเคราะห์หาค่าความยาก (BDL) โดยใช้โมเดล IRT ต่างกัน และ RP ต่างกัน จากนั้นนำค่า BDL มาพิจารณาจัดเรียงข้อสอบ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

เพื่อเปรียบเทียบความแตกต่างของการจัดเรียงข้อสอบที่มีโมเดล IRT ต่างกัน และ RP ต่างกัน ส่วนงานวิจัยของบาเคนดัลส (Buckendahl, 2002) ใช้แบบสอบแบบเลือกหลายคำตอบ และวิเคราะห์ค่าความยากโดยใช้ทฤษฎี CTT ซึ่งทำให้ค่าความยากที่ได้ใช้เฉพาะกลุ่มที่ทำการวิเคราะห์ ถ้าไปใช้กับกลุ่มอื่นอาจเกิดการคลาดเคลื่อนได้ ส่วนแวง (Wang, 2003) ใช้คะแนนจากแบบสอบแบบหลายตัวเลือกเพื่อหาคะแนนจุดตัดจากวิธีผังข้อสอบ (Item mapping) เพียงอย่างเดียวไม่ได้นำกระบวนการบุ๊คมาร์คเข้ามาใช้ (Lewis, Mitzel, & Green, 1996; Mitzel, Lewis, Patz, & Green, 2000, cite in Zwick, Senturk, Wang, Barbara & Loomis, 2001) แต่จุดเด่นของงานวิจัยนี้คือมีการวิเคราะห์ด้วย Rasch model ทำให้ผลการตัดสินมีความน่าเชื่อมากกว่าการวิเคราะห์ด้วยทฤษฎีการทดสอบแบบดั้งเดิม (CTT)

จากงานวิจัยทั้ง 3 เรื่องที่กล่าวข้างต้น งานวิจัยของเบเรทวาส (Beretvas, 2004) ศึกษาวิธีการบุ๊คมาร์คโดยเปรียบเทียบความสอดคล้องของการจัดเรียงข้อสอบเมื่อใช้โมเดล IRT ต่างกัน และ RP ต่างกัน ส่วนแวง (Wang, 2003) ได้เปรียบเทียบการกำหนดมาตรฐานด้วยวิธีผังข้อสอบกับวิธีการแองกอฟ จะเห็นว่างานวิจัยทั้งสองเรื่องยังไม่มีการเปรียบเทียบวิธีการบุ๊คมาร์คกับวิธีการแองกอฟ เมื่อพิจารณางานวิจัยของบาเคนดัลส (Buckendahl, 2002) ที่ทำการเปรียบเทียบวิธีการบุ๊คมาร์คและ แองกอฟ แต่เป็นการวิเคราะห์ด้วยทฤษฎีการทดสอบแบบดั้งเดิม (CTT) ซึ่งCTT มีข้อจำกัดที่สำคัญคือ การที่ค่าพารามิเตอร์ของข้อสอบผันแปรตามกลุ่มข้อสอบ และคะแนนที่สังเกตได้หรือค่าประมาณความสามารถของผู้สอบไม่เป็นอิสระ หรือขึ้นอยู่กับข้อสอบและแบบสอบที่นำมาใช้ (ศิริชัย กาญจนวาสี, 2545) จากข้อจำกัดดังกล่าวเมื่อนำแบบสอบไปทดสอบเพื่อหาค่าความยากจะทำให้มีความคลาดเคลื่อนสูง และค่าความยากที่ได้ก็จะใช้ได้เฉพาะกลุ่มที่ทดสอบจึงใช้ได้ในวงแคบ ดังนั้นผู้วิจัยจึงมีความสนใจที่จะเปรียบเทียบวิธีการของบุ๊คมาร์คและวิธีการของแองกอฟว่าวิธีการใดมีคุณภาพดีกว่ากัน โดยการนำทฤษฎีการตอบสนองข้อสอบมาใช้ในการวิเคราะห์ เนื่องจากทฤษฎีนี้มีจุดเด่นคือ การประมาณค่าพารามิเตอร์ของข้อสอบจะต้องกระทำพร้อมๆ ไปกับการประมาณค่าความสามารถจริงของผู้สอบจึงจะทำให้ค่าพารามิเตอร์มีนัยทั่วไป มีความน่าเชื่อถือและไม่แปรผันไปตามความสามารถของผู้สอบ (ศิริชัย กาญจนวาสี, 2545) นอกจากนี้งานวิจัยของบาเคนดัลส ยังมีจุดอ่อนอีกประการหนึ่งคือ งานวิจัยนี้เปรียบเทียบวิธีการบุ๊คมาร์ค และแองกอฟโดยการพิจารณาเพียงความแตกต่างของคะแนนจุดตัดและส่วนเบี่ยงเบนมาตรฐาน ว่าแต่ละรอบของการกำหนดคะแนนจุดตัดมีความแตกต่างของคะแนนจุดตัดจากทั้งสองวิธีมากน้อยเพียงใด ซึ่งการเปรียบเทียบดังกล่าวยังขาดความเป็นปรนัย ดังนั้นเพื่อให้เกิดความชัดเจนในการเปรียบเทียบคุณภาพของวิธีการกำหนดคะแนนจุดตัด ผู้วิจัยจึงสนใจที่เปรียบเทียบคุณภาพของการกำหนดมาตรฐานระหว่างวิธีแองกอฟที่ได้รับการปรับปรุงกับวิธี

บุ๊คมาร์ค โดยการหาค่าความสอดคล้องหรือสัมประสิทธิ์ความเที่ยงในการตัดสินคะแนนจุดตัดโดยทฤษฎีการสรุปอ้างอิง (Generalizability theory) Kane (1987) ได้สนับสนุนให้ใช้วิธีที่วิเคราะห์ตามทฤษฎีการสรุปอ้างอิงมากกว่าวิธีการวิเคราะห์แบบดั้งเดิม เพราะการวัดตามทฤษฎีการทดสอบแบบดั้งเดิมนั้นการประมาณค่าความเที่ยงของแบบทดสอบจะใช้เฉพาะกลุ่มบุคคลภายใต้เงื่อนไขของการทดสอบที่เฉพาะ ส่วนทฤษฎีการสรุปอ้างอิงนั้นได้ทำการวิเคราะห์ความเที่ยงในสถานการณ์หรือเงื่อนไขต่างๆ ของการวัด ทำให้ทราบและสามารถควบคุมแหล่งความคลาดเคลื่อนได้ตรงประเด็น เพื่อผลการวัดมีความน่าเชื่อถือหรือมีความเที่ยงสูงถึงระดับที่ต้องการ (ศิริชัย กาญจนวาสี, 2545) และขณะเดียวกันผู้วิจัยจะกำหนดคะแนนจุดตัดทั้งสองวิธีโดยการประยุกต์ทฤษฎีการตอบสนองข้อสอบเพื่อหาคะแนนความสามารถ

นอกจากนี้ผู้วิจัยยังได้นำกระบวนการพิจารณากำหนดคะแนนจุดตัดมากกว่า 1 รอบมาใช้ในการวิจัยครั้งนี้ด้วย โดยแต่ละรอบให้ผู้ตัดสินทั้งหมดร่วมกันอภิปรายเป็นการเพิ่มความเที่ยงในการตัดสินของผู้ตัดสิน (Hurtz and Auerbach, 2003) ซึ่งสอดคล้องกับงานวิจัยของเฟรดแมนและโฮ (Friedman and Ho, 1990) ที่สนับสนุนกระบวนการดังกล่าว โดยเฟรดแมนและโฮ ให้ผู้ตัดสินจำนวน 11 คน จาก 11 รัฐของอเมริกาตัดสินคะแนนจุดตัดโดยใช้ข้อสอบจำนวน 65 ข้อ หลังจากนั้นนำผลการตัดสินที่ได้มาให้ผู้ตัดสินทั้งหมดร่วมกันอภิปราย จากนั้นผู้ตัดสินทำการตัดสินคะแนนจุดตัดของข้อสอบชุดเดิมอีกครั้งพบว่า สามารถเพิ่มความสอดคล้องในการตัดสินระหว่างกลุ่มผู้ตัดสิน (Interjudges consistency) และความสอดคล้องในการตัดสินภายในของผู้ตัดสิน (Intrajudges consistency)

จากงานวิจัยของ แพทริเซีย (Patricia, 1991) พบว่า การกำหนดมาตรฐานโดยใช้ดุลยพินิจของผู้ตัดสินขึ้นอยู่กับคุณภาพของผู้ตัดสิน และข้อมูลสารสนเทศประกอบการตัดสินที่เพียงพอ ดังนั้นในงานวิจัยนี้ผู้วิจัยจึงได้เพิ่มสารสนเทศในวิธีการทั้งสองเพื่อให้ผู้ตัดสินใช้ในการกำหนดมาตรฐาน ประกอบด้วย สัดส่วนผู้ที่ตอบถูกในแต่ละข้อ เข้ามาเป็นนวัตกรรมใหม่ในคู่มือจัดเรียงข้อสอบ และกราฟแจกแจงความถี่ของคะแนนสอบ เข้ามาให้ผู้ตัดสินพิจารณาในการกำหนดมาตรฐานแองกอฟที่ได้รับการปรับปรุง

และเนื่องจากจำนวนผู้ตัดสิน และจำนวนครั้งในการกำหนดมาตรฐานน่าจะเป็นปัจจัยหนึ่งที่ส่งผลต่อคุณภาพการกำหนดมาตรฐาน ดังเช่น เจเกอร์ (Jaeger, 1991, cited in Berk, 1996) กล่าวว่า งานวิจัยในช่วง 15 ปีที่ผ่านมา การกำหนดมาตรฐานใช้ผู้ตัดสินระหว่าง 5-20 คน ดังนั้นผู้วิจัยจึงได้สนใจที่ศึกษาว่าจำนวนผู้ตัดสิน และจำนวนครั้งส่งผลต่อการกำหนดมาตรฐานอย่างไร โดยการใช้ทฤษฎีการสรุปอ้างอิง ดังนั้นผู้วิจัยจึงมีความสนใจที่จะศึกษาว่าจำนวน

ผู้ตัดสินที่ต่างกัน มีผลต่อการกำหนดมาตรฐานทั้งสองวิธีอย่างไร เพื่อจะได้เป็นแนวทางในการกำหนดจำนวนผู้ตัดสิน และจำนวนครั้งที่เหมาะสมต่อไป

นอกจากนี้กระทรวงศึกษาธิการได้ให้อิสระแก่สถานศึกษาในการกำหนดหลักสูตร และเกณฑ์การประเมินในโรงเรียนเพื่อให้สอดคล้องกับการจัดการเรียนการสอนตามสภาพจริง การประเมินตามสภาพจริง และบริบทของชุมชน แต่กลับพบว่า สถานศึกษาหลายแห่งยังคงใช้เกณฑ์เดิมของกระทรวงศึกษาธิการ อาจเนื่องมาจากขาดแนวทางชี้แนะในการกำหนดเกณฑ์ประเมิน ดังนั้นในการวิจัยครั้งนี้ ผู้วิจัยจึงได้นำวิธีการกำหนดมาตรฐานทั้งสองวิธีมาใช้กำหนดมาตรฐานในโรงเรียน และเพื่อให้สอดคล้องกับบริบทการวัดและประเมินผลของประเทศไทยที่มีการตัดเกรดเป็น 8 ระดับ และจากการศึกษาเอกสารงานวิจัยของ เบเรทวาส (Beretvas, 2004) แวง (Wang, 2003) และบาเคนดัลส (Buckendahl, 2002) เป็นงานวิจัยที่กำหนดคะแนนจุดตัดเพียงจุดเดียว การวิจัยนี้จึงให้ผู้ตัดสินพิจารณาคะแนนจุดตัดของนักเรียนเป็น 7 ระดับ คือ ระดับดีเยี่ยม (A) ระดับดีมาก (B+) ระดับดี (B) ระดับดีพอใช้ (C+) ระดับพอใช้ (C) ระดับอ่อน (D+) และ ระดับอ่อนมาก (D) งานวิจัยนี้น่าจะเป็นประโยชน์ต่อวงการศึกษาของไทยในการพัฒนาการกำหนดมาตรฐานที่มีคุณภาพ ในการสอบลักษณะต่างๆ โดยเฉพาะการสอบที่ส่งผลต่ออนาคตของผู้สอบ เช่น การสอบในโรงเรียนที่ต้องนำผลการสอบมาพิจารณาเพื่อรับทุน หรือเข้าศึกษาต่อในระดับสูง การสอบเข้ามหาวิทยาลัย การสอบใบประกอบวิชาชีพ

คำถามวิจัย

1. คุณภาพของการกำหนดมาตรฐานด้านความตรง ความเที่ยง และ ความเหมาะสม และความเป็นไปได้ระหว่างวิธีการแองกอฟที่ได้รับการปรับปรุง กับวิธีการบุ๊คมาร์ค มีความแตกต่างกันหรือไม่

2. จำนวนผู้ตัดสิน และจำนวนครั้งที่ต่างกัน มีผลต่อการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และวิธีบุ๊คมาร์คอย่างไร

วัตถุประสงค์การวิจัย

1. เพื่อเปรียบเทียบคุณภาพของการกำหนดมาตรฐานระหว่างวิธีการแองกอฟที่ได้รับการปรับปรุงกับวิธีการบุ๊คมาร์ค โดยพิจารณาจาก

   1.1 ความตรง และความเที่ยง

   1.2 ความเหมาะสมและความเป็นไปได้

2. เพื่อศึกษาผลของจำนวนผู้ตัดสิน และจำนวนครั้ง ที่มีต่อการกำหนดมาตรฐานด้วยวิธีการแองกอฟที่ได้รับการปรับปรุง และวิธีการบุ๊คมาร์ค

สมมุติฐานการวิจัย

เนื่องจากวิธีการบุ๊คมาร์คถูกออกแบบมาให้ง่ายต่อการตัดสินโดยมีคู่มือที่มีการเรียงข้อสอบจากข้อง่ายไปข้อยาก (Mitzel, Lewis, Patz, & Green, 2001, cite in Beretvas, 2004) ทำให้ผู้ตัดสินได้บูรณาการมโนทัศน์ (Integrated conceptualization) ขณะเดียวกันเป็นเสมือนเครื่องมือนำไปสู่การตัดสินคะแนนจุดตัด (Lewis et al, 1998)

นอกจากนี้จากการศึกษาเอกสารงานวิจัยที่เกี่ยวข้อง บาเคนดัลส (Buckendahl, 2002) ได้เปรียบเทียบการกำหนดมาตรฐานด้วยวิธีการแองกอฟและวิธีการบุ๊คมาร์คโดยใช้แบบสอบแบบเลือกหลายคำตอบและวิเคราะห์ค่าความยากโดยใช้ทฤษฎีการทดสอบแบบดั้งเดิม (CTT) พบว่าวิธีบุ๊คมาร์ค ในรอบที่สองมีส่วนเบี่ยงเบนมาตรฐานลดลง เมื่อเทียบกับวิธีแองกอฟ ซึ่งการลดลงของส่วนเบี่ยงเบนมาตรฐานในรอบสองของวิธีบุ๊คมาร์คนี้ แสดงให้เห็นว่าพิสัยของคะแนนจุดตัดลดลง และชี้ให้เห็นว่า การตัดสินในระหว่างผู้ตัดสิน (inter-judge) มีระดับสูงขึ้น ส่วน แวง (Wang, 2003) ใช้คะแนนจากแบบสอบแบบหลายตัวเลือกเพื่อหาคะแนนจุดตัดด้วยวิธีผังข้อสอบ (Item mapping) และวิธีแองกอฟมาเปรียบเทียบกันโดยใช้ Rasch Model พบว่า วิธีผังข้อสอบมีความสอดคล้องในการตัดสินระหว่างผู้ตัดสิน (inter-judge consistency) มากกว่าวิธีการแองกอฟ ผู้วิจัยจึงกำหนดสมมุติฐานการวิจัยดังนี้

วิธีการกำหนดมาตรฐานด้วยวิธีการบุ๊คมาร์คน่าจะมีคุณภาพของการกำหนดมาตรฐานสูงกว่า วิธีการแองกอฟที่ได้รับการปรับปรุง กล่าวคือ การกำหนดมาตรฐานด้วยวิธีการบุ๊คมาร์คมีความตรง และความเที่ยงสูงกว่า วิธีการแองกอฟที่ได้รับการปรับปรุง ขณะเดียวกันการกำหนดมาตรฐานด้วยวิธีการบุ๊คมาร์คมีความเหมาะสม และความเป็นไปได้สูงกว่าวิธีการแองกอฟที่ได้รับการปรับปรุง

ขอบเขตการวิจัย

1. ประชากรที่ศึกษา

1.1 ประชากรผู้ตัดสิน คือ ครูผู้สอนวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ในโรงเรียนสังกัดสำนักงานการศึกษาขั้นพื้นฐาน สำนักงานเขตพื้นที่การศึกษาชลบุรี เขต 1

1.2 ประชากรผู้สอบ คือ นักเรียนชั้นมัธยมศึกษาปีที่ 3 ปีการศึกษา 2549 ภาคการศึกษาที่ 1 ในโรงเรียนสังกัดสำนักงานการศึกษาขั้นพื้นฐาน สำนักงานเขตพื้นที่การศึกษาชลบุรี เขต 1

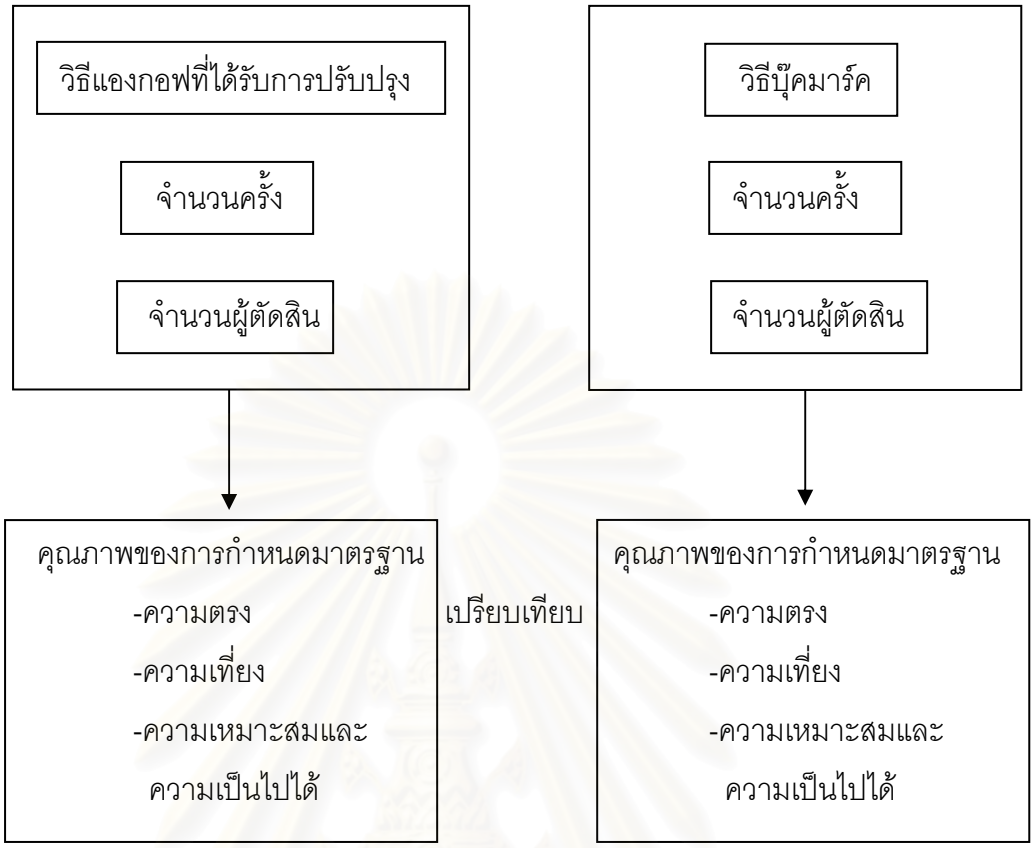
2. ตัวแปรที่ศึกษา

2.1 ตัวแปรต้น คือ วิธีการกำหนดมาตรฐาน ได้แก่ วิธีการแองกอฟที่ได้รับการปรับปรุงและวิธีการบุ๊คมาร์ค จำนวนผู้ตัดสิน และจำนวนครั้ง

2.2 ตัวแปรตาม คือ คุณภาพของการกำหนดมาตรฐาน ได้แก่ ความตรง ความเที่ยง ความเหมาะสมและความเป็นไปได้

จากตัวแปรที่ศึกษาดังกล่าว ผู้วิจัยขอนำเสนอ แผนภาพที่ 1 กรอบแนวคิดการวิจัย ดังนี้

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย



แผนภาพที่ 1 กรอบแนวคิดในการวิจัย

คำจำกัดความที่ใช้ในการวิจัย

1. **มาตรฐาน** (Standard) หมายถึง ระดับที่บ่งบอกคุณภาพของนักเรียนในวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ภาคเรียนที่ 1 ในการศึกษาครั้งนี้แบ่งนักเรียนออกเป็น 8 ระดับ คือ ระดับดีเยี่ยม (A) ระดับดีมาก (B+) ระดับดี (B) ระดับดีพอใช้ (C+) ระดับพอใช้ (C) ระดับอ่อน (D+) ระดับอ่อนมาก (D) และ ระดับตก (F) และในแต่ละระดับจะมีคำอธิบายระดับการปฏิบัติ (performance level descriptor) ซึ่งเป็นข้อความเพื่อแสดงว่า นักเรียนควรมีความรู้ ทักษะหรือคุณลักษณะในแต่ละระดับอย่างไร

2. **คะแนนจุดตัด** (Cut score) หมายถึง จุดคะแนนที่ใช้เป็นจุดแบ่งนักเรียนในวิชาคณิตศาสตร์ชั้นมัธยมศึกษาปีที่ 3 ภาคเรียนที่ 1 ตามความสามารถของนักเรียน งานวิจัยนี้แบ่งคะแนนจุดตัดออกเป็น 7 จุด คือ 1) จุดคะแนนที่แบ่งความสามารถนักเรียนที่มีความสามารถอ่อนมาก ออกจากนักเรียนระดับตก 2) จุดคะแนนที่แบ่งนักเรียนที่มีความสามารถระดับอ่อนมาก ออกจากนักเรียนระดับอ่อน 3) จุดคะแนนที่แบ่งนักเรียนที่มีความสามารถระดับอ่อน ออกจาก

นักเรียนระดับพอใช้ 4) จุดคะแนนที่แบ่งนักเรียนที่มีความสามารถระดับพอใช้ ออกจากนักเรียนระดับดีพอใช้ 5) จุดคะแนนที่แบ่งนักเรียนที่มีความสามารถระดับดีพอใช้ ออกจากนักเรียนระดับดี 6) จุดคะแนนที่แบ่งนักเรียนที่มีความสามารถระดับดี ออกจากนักเรียนระดับดีมาก 7) จุดคะแนนที่แบ่งนักเรียนที่มีความสามารถระดับดีมาก ออกจากนักเรียนระดับดีเยี่ยม

3. **วิธีการกำหนดมาตรฐาน** (Standard setting) หมายถึง กระบวนการหามาตรฐานเพื่อแบ่งความสามารถของนักเรียนออกเป็น 8 ระดับ คือ ระดับดีเยี่ยม(A) ระดับดีมาก(B+) ระดับดี(B) ระดับดีพอใช้ (C+) ระดับพอใช้ (C) ระดับอ่อน (D+) ระดับอ่อนมาก (D) และระดับตก (F) ในการวิจัยครั้งนี้ใช้วิธีการกำหนดมาตรฐาน 2 วิธีคือ

3.1 **วิธีการกำหนดมาตรฐานบุ๊คมาร์ค** (Bookmark method) หมายถึง วิธีการหามาตรฐาน โดยจัดให้มีคู่มือการเรียงข้อสอบจากข้อง่ายสุดไปข้อยากสุด ข้อละหนึ่งหน้า มีการวิเคราะห์ข้อสอบเพื่อหาค่าความยาก และความสามารถของผู้สอบในแต่ละข้อโดยนำทฤษฎีการตอบสนองข้อสอบ (IRT) มาใช้ในการวิเคราะห์ คู่มือที่ได้จะนำมาให้ผู้ตัดสินพิจารณาประกอบการหาคะแนนจุดตัด ซึ่งผู้ตัดสินจะพิจารณาทีละข้อ จากข้อง่ายไปยังข้อยาก โดยให้ผู้ตัดสินตอบคำถามว่า *ผู้ที่คาบเส้นระดับพอใช้จะมีโอกาสตอบข้อสอบข้อนั้นถูก 67%”* พิจารณาทีละข้อไปเรื่อยๆ จนถึงข้อที่ผู้สอบกลุ่มคาบเส้นระดับพอใช้ไม่มีโอกาสตอบถูก ก็จะใช้ที่คั่นหนังสือคั่นตรงหน้านั้นไว้” ซึ่งเป็นคะแนนจุดตัดระดับพอใช้ (จุดแรก) จากนั้นผู้ตัดสินดำเนินการหาคะแนนจุดตัดระดับดี ระดับดีมาก ระดับดีเยี่ยม ระดับอ่อน ระดับอ่อนมาก และระดับตกตามลำดับ โดยใช้กระบวนการคล้ายกับที่กล่าวมา จากนั้นนำผลการตัดสินของแต่ละคนมาอภิปรายร่วมกัน แล้วจึงเปิดโอกาสให้ผู้ตัดสินแต่ละคนหาคะแนนจุดตัดอีกครั้ง นำคะแนนจุดตัดเหล่านั้นมาหาค่าเฉลี่ย

3.2 **วิธีการแองกอฟที่ได้รับการปรับปรุง** (Modified Angoff method) หมายถึง วิธีการหามาตรฐาน โดยให้ผู้ตัดสินประมาณระดับความสามารถขั้นต่ำของผู้สอบในข้อสอบแต่ละข้อจนครบทุกข้อ หลังจากนั้นผู้ตัดสินแต่ละคนร่วมกันอภิปรายผลการตัดสินของแต่ละคน ซึ่งระหว่างการอภิปรายผู้อำนวยความสะดวกแสดงรูปการแจกแจงผลการตัดสินของผู้ตัดสินแต่ละคน เมื่อผู้ตัดสินอภิปรายเสร็จก็เปิดโอกาสให้ตัดสินคะแนนอีกครั้งหนึ่ง ผลการตัดสินที่ได้จากผู้ตัดสินแต่ละคนในรอบที่สองนำมาหาค่าเฉลี่ยซึ่งจะเป็นคะแนนจุดตัด

4. **ผู้ตัดสิน** หมายถึง ครูผู้สอนวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ในโรงเรียนสังกัดสำนักงานการศึกษาขั้นพื้นฐาน สังกัดสำนักงานเขตพื้นที่การศึกษา ชลบุรีเขต 1 ที่มีคุณสมบัติดังนี้ 1) มีความเชี่ยวชาญในเนื้อหาคณิตศาสตร์มัธยมศึกษาตอนต้น คือ จบการศึกษาระดับปริญญาตรีขึ้นไปในสาขาทางคณิตศาสตร์ และ 2) มีความเชี่ยวชาญในการสอน และการ

วัดผล คือ มีประสบการณ์การสอนมาไม่น้อยกว่า 5 ปี และเป็นครูชำนาญการ หรือชำนาญการพิเศษ

5. **ความเที่ยงของการกำหนดมาตรฐาน** หมายถึง ความคงเส้นคงวาในการตัดสินจำแนกนักเรียนออกเป็น 8 ระดับคือ ระดับดีเยี่ยม ระดับดีมาก ระดับดี ระดับดีพอใช้ ระดับพอใช้ ระดับอ่อน ระดับอ่อนมาก และ ระดับตก วัดได้โดยใช้ทฤษฎีการสรุปอ้างอิง (Generalizability theory) ด้วยการวิเคราะห์ความแปรปรวนเพื่อนำมาใช้หาสัมประสิทธิ์การสรุปอ้างอิง หรือสัมประสิทธ์ความเที่ยง นอกจากนี้ยังมีคำจำกัดความที่ใช้ทฤษฎีการสรุปอ้างอิงในการหาความเที่ยงของการกำหนดมาตรฐานดังนี้

5.1 เอกภพ หมายถึง เงื่อนไขทั้งหมดในการวัด สำหรับการวิจัยนี้ได้แก่ ผู้ตัดสินทุกคน และข้อสอบทุกข้อ

5.2 ฟาเซท หมายถึง องค์ประกอบหรือลักษณะร่วมกันของกลุ่มเงื่อนไขการวัดชนิดเดียวกัน ในการวิจัยนี้มี 2 ฟาเซท คือ ผู้ตัดสิน และข้อสอบ

5.3 เงื่อนไขการวัด หมายถึง สถานการณ์ของการตัดสิน สำหรับการวิจัยครั้งนี้คือ ผู้ตัดสินแต่ละคน และข้อสอบแต่ละข้อ ที่สุ่มมาใช้

5.4 เอกภพของค่าที่ได้จากการสังเกตทั้งหมด (Universe of Admissible Observation) หมายถึง เงื่อนไขการวัดที่สนใจทั้งหมดคือ ผู้ตัดสินทุกคน และข้อสอบทุกข้อ ซึ่งอยู่ในเอกภพของแต่ละฟาเซทที่ผู้วิจัยสามารถสังเกตและวัดได้

5.5 เอกภพของการสรุปอ้างอิง (Universe of Generalization) หมายถึง เงื่อนไขการวัดทั้งหมดที่เป็นเป้าหมายของการสรุปอ้างอิง ในการวิจัยนี้ได้แก่ ผู้ตัดสิน และข้อสอบ

6. **ความตรงของการกำหนดมาตรฐาน** หมายถึง คุณสมบัติของมาตรฐานที่สามารถแบ่งนักเรียนออกเป็น 8 ระดับคือ ระดับดีเยี่ยม ระดับดีมาก ระดับดี ระดับดีพอใช้ ระดับพอใช้ ระดับอ่อน ระดับอ่อนมาก และระดับตก วัดได้โดยใช้สหสัมพันธ์แบบ Spearman' s Rho

7. **ความตรงเชิงเนื้อหาของแบบสอบ** หมายถึง ความสามารถของแบบสอบที่ใช้วัดในเนื้อหาที่ต้องการวัด ความตรงเชิงเนื้อหาของแบบสอบได้จากการประเมินความสอดคล้องกันระหว่างเนื้อหาของแบบสอบกับขอบเขตเนื้อหาที่ต้องการวัด ซึ่งได้มาจากการสรุปผลการพิจารณาตัดสินความสอดคล้องโดยคณะผู้เชี่ยวชาญ

8. **ความเหมาะสมของการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุงและวิธีบุ๊คมาร์ค** หมายถึง คุณสมบัติที่แสดงถึงความถูกต้องของการกำหนดมาตรฐานสองวิธีคือ วิธีแองกอฟที่ได้รับการปรับปรุง และวิธีการบุ๊คมาร์ค วัดโดยใช้เกณฑ์ประเมินกระบวนการกำหนดมาตรฐานที่ประกอบด้วย ความชัดเจน การนำไปปฏิบัติได้ การดำเนินการ ผลย้อนกลับ และเอกสารที่ใช้ในกระบวนการ

9. **ความเป็นไปได้ของการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุงและวิธีบุ๊คมาร์ค** หมายถึง คุณสมบัติที่แสดงถึงความสอดคล้องกับสภาพความเป็นจริง สามารถนำไปปฏิบัติได้ของการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และวิธีบุ๊คมาร์ค วัดโดยใช้เกณฑ์ประเมินกระบวนการกำหนดมาตรฐานที่ประกอบด้วย ความชัดเจน การนำไปปฏิบัติได้ การดำเนินการ ผลย้อนกลับ และเอกสารที่ใช้ในกระบวนการ

ประโยชน์ที่คาดว่าจะได้รับ

1. **ประโยชน์ในทางวิชาการ**

1.1 ได้ข้อความรู้เกี่ยวกับวิธีการกำหนดมาตรฐานแนวใหม่ คือวิธีการบุ๊คมาร์ค วิธีนี้มีจุดเด่น คือ ช่วยจัดระบบการคิดแก่ผู้ตัดสินทำให้ผู้ตัดสินกำหนดคะแนนจุดตัดได้ง่ายขึ้น (McKinley, Newman, & Wiser, 1996) เนื่องจากมีการจัดเรียงข้อสอบเป็นระบบในรูปของคู่มือ นอกจากนี้วิธีการนี้เหมาะสำหรับการกำหนดมาตรฐานแบบสอบที่มีลักษณะที่ให้คะแนนสองค่าและที่ให้คะแนนแบบหลายค่า

1.2 การกำหนดมาตรฐานทั้งวิธีบุ๊คมาร์ค และวิธีแองกอฟที่ได้รับการปรับปรุงในการวิจัยครั้งนี้ได้นำกระบวนการกลุ่มมาใช้ในการกำหนดมาตรฐาน โดยการจัดให้มีการอภิปรายภายในกลุ่ม และเปิดโอกาสให้มีการปรับเปลี่ยนคะแนนจุดตัดหลังมีการอภิปราย กระบวนการดังกล่าวนี้มีข้อดีคือ เป็นการเพิ่มความตรงในการกำหนดมาตรฐานนั้นเอง

2. **ประโยชน์ในทางปฏิบัติ**

การวิจัยครั้งนี้ได้คู่มือสำหรับผู้ตัดสินในการกำหนดมาตรฐานด้วยวิธีการแองกอฟที่ได้รับการปรับปรุง และวิธีการบุ๊คมาร์ค ผู้ตัดสินสามารถนำคู่มือดังกล่าวมาใช้ในการกำหนดมาตรฐานในสถานการณ์ต่างๆ เช่น การสอบที่มีลักษณะส่งผลกระทบต่ออนาคตผู้สอบสูง (high stake) กล่าวคือ การสอบเข้ามหาวิทยาลัยหรือสถาบันการศึกษาต่างๆ การสอบในแต่ละช่วงชั้น การสอบเพื่อให้ผลการเรียนในแต่ละระดับชั้น

บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อ 1) เพื่อเปรียบเทียบคุณภาพของการกำหนดมาตรฐานด้านความเที่ยง ความตรง ความเหมาะสมและความเป็นไปได้ระหว่างวิธีการกำหนดมาตรฐานแองกอฟที่ได้รับการปรับปรุงกับวิธีการบุ๊คมาร์ค 2) เพื่อศึกษาผลของจำนวนผู้ตัดสิน และจำนวนครั้งที่มีต่อการกำหนดมาตรฐานด้วยวิธีการแองกอฟที่ได้รับการปรับปรุง และวิธีการบุ๊คมาร์ค ผู้วิจัยได้ศึกษาเอกสารและงานวิจัยที่เกี่ยวข้อง โดยแบ่งการนำเสนอออกเป็น 6 ตอน ดังนี้

ตอนที่ 1 มโนทัศน์ของมาตรฐาน และคะแนนจุดตัด

ตอนที่ 2 วิธีการกำหนดมาตรฐาน

1. วิธีการกำหนดมาตรฐานแบบดั้งเดิม
2. วิธีการกำหนดมาตรฐานแนวใหม่

ตอนที่ 3 ทฤษฎีการตอบสนองข้อสอบกับการกำหนดคะแนนจุดตัด

1 การกำหนดคะแนนจุดตัดด้วยวิธีแองกอฟที่ได้รับการปรับปรุงโดยการประยุกต์ทฤษฎีการตอบสนองข้อสอบ

2 การกำหนดคะแนนจุดตัดด้วยวิธีบุ๊คมาร์คโดยการประยุกต์ทฤษฎีการตอบสนองข้อสอบ

ตอนที่ 4 การตรวจสอบความเที่ยง และความตรงของการกำหนดมาตรฐาน

1 มโนทัศน์ทฤษฎีการสรุปอ้างอิง

2 ความตรงของการกำหนดมาตรฐาน

ตอนที่ 5 แนวคิดเกณฑ์ประเมินกระบวนการกำหนดมาตรฐานเพื่อใช้ประเมินความเหมาะสมและความเป็นไปได้ในการนำไปใช้

ตอนที่ 6 งานวิจัยที่เกี่ยวข้อง

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตอนที่ 1 มโนทัศน์ของมาตรฐาน (Standards) และคะแนนจุดตัด (Cut score)

คำว่า “มาตรฐาน” (standards) มีนักวิชาการให้ความหมายไว้ดังนี้

พจนานุกรมลองแมน (Longman Dictionary of Contemporary English, 1999) ได้ให้คำนิยาม มาตรฐาน (Standards) ว่า ระดับคุณภาพที่เป็นทักษะ ความสามารถ หรือความสำเร็จของคน หรือสิ่งของที่ได้รับการตัดสินให้เป็นที่ยอมรับในสถานการณ์หนึ่งๆ

เคน (Kane, 1994, cited in Cizek; Bunch and Koons, 2004) มาตรฐาน หมายถึง ระดับความสามารถในการปฏิบัติขั้นต่ำสำหรับบางจุดประสงค์

โคเฮน เคน และครุ๊ค (Cohen, Kane and Crooks, 1999) มาตรฐานการปฏิบัติ (Performance standard) เป็นการระบุชนิดของการปฏิบัติ หรือความสำเร็จของผู้สอบในกลุ่มที่แตกต่างกัน มาตรฐานการปฏิบัติเกี่ยวข้องกับขอบเขตในการจัดจำพวกที่ระบุว่า คนที่อยู่ในกลุ่มสูงสามารถทำได้ และคนที่อยู่ในกลุ่มต่ำไม่สามารถทำได้ มาตรฐานการปฏิบัติอยู่ในรูปภาษาที่อธิบายคุณภาพของสิ่งหนึ่งที่ควรจะเป็น

กูลิคซัน (Gullickson, 2003) ให้ความหมาย มาตรฐาน (standards) ว่า หลักการที่ตกลงโดยคนที่เกี่ยวข้องในการปฏิบัติทางวิชาชีพของการประเมิน ถ้าได้มาตรฐานก็จะเสริมให้การประเมินมีคุณภาพ และเป็นไปอย่างยุติธรรมยิ่งขึ้น

ศิริชัย กาญจนวาสี (2545) กล่าวว่า มาตรฐาน หมายถึง คุณลักษณะ หรือระดับที่ถือเป็นคุณภาพ, ความสำเร็จ หรือความเหมาะสมอันเป็นที่ยอมรับกันทางวิชาชีพโดยทั่วไป

สรุปได้ว่า มาตรฐาน หมายถึง ระดับที่บ่งบอกถึงคุณภาพของความสามารถของผู้สอบขั้นต่ำโดยระดับดังกล่าวจะแบ่งผู้สอบออกเป็นระดับ เช่น แบ่งเป็น 2 ระดับ คือ ระดับรอบรู้ และระดับไม่รอบรู้ หรือ แบ่งเป็น 4 ระดับ คือ ระดับต่ำกว่าพื้นฐาน ระดับพื้นฐาน ระดับชำนาญการ ระดับขั้นสูง เป็นต้น มาตรฐานจะมีคำอธิบายการปฏิบัติงานซึ่งจะอยู่ในรูปภาษาที่อธิบายคุณภาพของสิ่งหนึ่งที่ควรจะเป็น

ส่วนคำว่า “คะแนนจุดตัด” (cut score) มีนักวิชาการให้ความหมายไว้ดังนี้

เยาวดี วิบูลย์ศรี (2539) ให้ความหมายคำว่า คะแนนจุดตัด หมายถึง จุด หรือระดับค่าของคะแนนระดับใดระดับหนึ่งซึ่งอยู่ระหว่าง คะแนนเต็ม และคะแนนศูนย์ ที่ผู้วัดผลใช้เป็นจุดตัดเพื่อแบ่งความสามารถของผู้สอบออกเป็น “ผู้รอบรู้” และ “ผู้ไม่รอบรู้” “ผู้รอบรู้” ในที่นี้หมายถึง

ผู้ที่สอบได้คะแนนสูงถึงระดับที่กำหนดไว้ตามเกณฑ์มาตรฐานนั้นๆ ส่วน “ผู้ไม่รอบรู้” หมายถึง ผู้ที่สอบได้คะแนนต่ำกว่าระดับที่กำหนดไว้ตามเกณฑ์มาตรฐานนั้นๆ เช่นเดียวกัน

เบอร์ก (Berk, 1980) คะแนนจุดตัด คือ จุดบนสเกลคะแนนของแบบทดสอบที่แบ่งผู้สอบออกเป็น 2 กลุ่ม คือ กลุ่มรอบรู้ และกลุ่มไม่รอบรู้

โคเฮน เคน และครุ๊ค (Cohen, Kane and Crooks, 1999) กล่าวว่า คะแนนจุดตัดเป็นขอบเขตระหว่างรอยต่อของแต่ละกลุ่มซึ่งก็คือ จุดบนสเกลคะแนนนั้นเอง ตัวอย่างเช่น มีคะแนนจุดตัด 1 จุดที่แบ่งออกเป็น 2 กลุ่มคือ กลุ่มผ่านและกลุ่มตก ซึ่งคะแนนจุดตัดนั้นเป็นการกำหนดขึ้นในทางปฏิบัติสำหรับกำหนดกลุ่มที่แตกต่างกัน

เคน (Kane, 1994, cited in Cizek; Bunch and Koons, 2004) คะแนนจุดตัด คือ จุดบนสเกลคะแนน

สรุปได้ว่า คะแนนจุดตัด หมายถึง จุดบนสเกลคะแนน เพื่อใช้แบ่งนักเรียนออกเป็นกลุ่มๆ ตามความสามารถของนักเรียน

การที่จะทราบว่ามาตรฐาน ในการประเมินหนึ่งๆ เป็นอย่างไร จำเป็นต้องกำหนดมาตรฐาน ซึ่งวิธีการกำหนดมาตรฐานมีด้วยกันหลายวิธี สำหรับตอนที่ 2 นำเสนอวิธีการกำหนดมาตรฐานแบบดั้งเดิม และวิธีการกำหนดมาตรฐานแนวใหม่

ตอนที่ 2 วิธีการกำหนดมาตรฐาน(Standard setting methods)

การกำหนดมาตรฐาน (Standard setting) ได้มีนักวิชาการให้ความหมายไว้ดังนี้

เจเกอร์ (Jaeger, 1989, cited in Kane, 1998) ได้กล่าวว่า การกำหนดมาตรฐานเป็นการออกแบบมาเพื่อให้มีการตัดสินที่ชัดเจนเกี่ยวกับการจัดให้ผู้สอบแต่ละคนเข้ากลุ่มหนึ่ง หรือหลายกลุ่มโดยอยู่บนพื้นฐานของคะแนนผู้สอบแต่ละคน

ซิเซค (Cizek, 1993, cited in Cizek; Bunch and Koons, 2004) กล่าวว่า การกำหนดมาตรฐาน คือ ข้อกำหนดที่มีความเหมาะสมโดยใช้กระบวนการที่เป็นเหตุเป็นผลเพื่อให้ได้ระดับการปฏิบัติ 2 ระดับ หรือมากกว่านี้

สำหรับวิธีการกำหนดมาตรฐาน ผู้วิจัยได้ศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องสามารถแบ่งออกเป็น 2 แนวใหญ่ๆ คือ

1) วิธีการกำหนดมาตรฐานแบบดั้งเดิม วิธีแบบสอบเป็นศูนย์กลาง (test-center methods) กับวิธีผู้สอบเป็นศูนย์กลาง (Examinee-centered methods)

2) วิธีการกำหนดมาตรฐานแนวใหม่ เป็นแนวการกำหนดมาตรฐานที่ประยุกต์วิธีการกำหนดมาตรฐานแบบดั้งเดิม และสร้างวิธีการแบบใหม่ขึ้นมา เพื่อนำไปใช้กำหนดมาตรฐานการประเมินผลการเรียนรู้ที่มีหลากหลายรูปแบบ

1. วิธีกำหนดมาตรฐานแบบดั้งเดิม

วิธีการกำหนดมาตรฐานแบบดั้งเดิม แบ่งออกเป็น 2 วิธี คือ วิธีแบบสอบเป็นศูนย์กลาง (test-center methods) กับวิธีผู้สอบเป็นศูนย์กลาง (Examinee-centered methods) (Jaeger, 1989, cited in Kane,1998) เจเกอร์ได้แบ่งวิธีการกำหนดมาตรฐานเป็น 2 วิธีดังกล่าวข้างต้น การแบ่งแบบนี้ได้มีการนำมาใช้อ้างอิงในงานวิจัยเป็นจำนวนมาก (Berk, 1996; Kiplinger, 1997; Sireci and Robin, 1999; Ben-David, 2000; Green, 2000)

1.1 วิธีแบบสอบเป็นศูนย์กลาง (test-center methods) ผู้ตัดสินจะพิจารณางาน และข้อสอบที่อยู่ในแบบสอบ แล้วตัดสินระดับการปฏิบัติของงานนั้นซึ่งจะชี้ถึงความสำเร็จในมาตรฐานการปฏิบัติ วิธีแบบสอบเป็นศูนย์กลางมี 4 วิธีคือ

1.1.1 วิธีแองกอฟ (Angoff method) ผู้คิดค้นวิธีนี้คือ William H. Angoff ในปีค.ศ. 1971 ผู้ตัดสินจะถูกถามเพื่อให้นึกถึงความสามารถของผู้สอบที่มีความสามารถขั้นต่ำ (minimally acceptable person) หรือนักวัดผลบางคนเรียกว่า ผู้สอบคาบเส้น ซึ่งผู้สอบที่มีความสามารถขั้นต่ำคือ คนที่มีระดับความรู้ขั้นต่ำในเรื่องที่จำเป็นต้องปฏิบัตินั้นเอง หลังจากนั้นผู้ตัดสินก็ทำการตัดสินความเป็นไปได้ ที่เรียกว่า “ระดับการผ่านขั้นต่ำ” (minimum pass level) นั้นคือ ผู้ตัดสินจะกำหนดความสามารถขั้นต่ำของผู้สอบที่มีความเป็นไปหรือระดับการผ่านขั้นต่ำโดยกำหนดในรูปของเปอร์เซนต์การตอบข้อสอบแต่ละข้อถูก เมื่อผู้ตัดสินทำการตัดสินความเป็นไปได้ในการตอบถูกของผู้ที่มีความสามารถครบทุกข้อแล้ว ระดับการผ่านขั้นต่ำจะถูกเฉลี่ยจากผู้ตัดสินเป็นระดับการผ่านขั้นต่ำรายข้อ ระดับการผ่านขั้นต่ำรายข้อจะถูกรวมจากทุกข้อในแบบสอบเพื่อกำหนดเป็นจุดตัด วิธีการแองกอฟเป็นวิธีที่ใช้กันอย่างกว้างขวางสำหรับการกำหนดมาตรฐานในแบบสอบวัดผลสัมฤทธิ์ทางการเรียนที่ส่งผลกระทบต่ออนาคตของนักเรียนอย่างสูง (high stakes) (Angoff, 1971, cited in Kane, 1998)

ดังนั้นเพื่อให้ง่ายต่อความเข้าใจถึงวิธีการแองกอฟในการหาจุดตัดคะแนน ผู้วิจัยขอสรุปขั้นตอนการดำเนินการพร้อมทั้งยกตัวอย่างประกอบเพื่อให้เห็นภาพที่ชัดเจนยิ่งขึ้น

1) ผู้ตัดสินนึกถึงลักษณะความสามารถของผู้สอบที่มีความสามารถขั้นต่ำ

2) ผู้ตัดสินพิจารณาข้อสอบทีละข้อโดยแต่ละข้อตอบคำถามที่ว่า “ผู้สอบที่มีความสามารถขั้นต่ำมีโอกาสการตอบข้อสอบข้อนั้นถูกเท่าไร”

3) ผู้ตัดสินทุกคนตอบคำถามตามข้อ 2 ทุกข้อ

4) นำเปอร์เซนต์การตอบถูกของผู้สอบที่มีความสามารถขั้นต่ำของข้อสอบข้อนั้นจากผู้ตัดสินทุกคนมาบวกกันแล้วหารด้วยจำนวนผู้ตัดสิน แล้วนำค่าเฉลี่ยเปอร์เซนต์การตอบในแต่ละข้อมาหาค่าเฉลี่ยก็จะได้คะแนนจุดตัด จากดังตารางที่ 2 จะเห็นว่าจุดตัดคะแนนคือ 75

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 2 ตัวอย่างการคำนวณหาคะแนนจุดตัดด้วยวิธีแองกอฟจากเปอร์เซนต์การตอบข้อสอบแต่ละข้อถูกของผู้สอบที่มีความสามารถขั้นต่ำ

	ผู้ตัดสิน(คนที่)					
ข้อที่	1	2	3	4	5	ค่าเฉลี่ย
1	90	85	95	90	90	90
2	85	80	70	75	80	78
3	80	80	85	85	85	83
4	60	65	70	70	60	65
5	85	80	75	75	80	79
6	80	75	75	80	80	78
7	65	65	70	65	65	66
8	75	70	70	70	70	71
9	80	70	75	75	75	75
10	65	65	65	60	65	64
ค่าเฉลี่ย	76.5	73.5	75	74.5	75	75

จุดตัดคะแนน

ต่อมา Livingston และ Zieky (1982) ได้ปรับปรุงวิธีแองกอฟใหม่ มีชื่อว่า **Modified Angoff Method** สำหรับสาเหตุที่ต้องปรับปรุงวิธีการแองกอฟก็เนื่องจากว่าต้องการเพิ่มความเที่ยงของผู้ตัดสินในการประมาณโอกาสในการตอบข้อสอบถูกโดยการเพิ่มความสอดคล้องในการตัดสินของผู้ตัดสิน (intra-judge consistency) และความสอดคล้องในการตัดสินระหว่างผู้ตัดสิน (inter-judge consistency) ซึ่งการเพิ่มความเที่ยงของผู้ตัดสินก็จะทำให้ค่าความผิดพลาดของคะแนนจุดตัดที่ได้ลดลง (Hurtz and Auerbach, 2003)

สำหรับวิธีการกำหนดจุดตัดทำโดยการกำหนดให้ผู้ตัดสินพิจารณาโอกาสการตอบถูกของผู้สอบที่มีความสามารถขั้นต่ำจำนวน 2 รอบ โดยรอบแรก ผู้ตัดสินกำหนดโอกาสการตอบถูกของผู้สอบที่มีความสามารถขั้นต่ำในข้อสอบข้อแรก เมื่อกำหนดเสร็จก็นำผลการตอบของผู้ตัดสินแต่ละ

คนมาเขียนบนกระดาษ หรือกระดาษแผ่นใหญ่เพื่อให้ผู้ตัดสินทุกคนได้เห็น ถ้าผลการตอบของผู้ตัดสินคล้ายกันคือมีค่าห่างกันไม่เกิน 10 % ก็ให้ผู้ตัดสินพิจารณาข้ออื่นถัดไป สำหรับในรอบที่สอง จะมีการอภิปรายร่วมกัน กรณีที่ผลการตัดสินของแต่ละคนไม่คล้ายกัน โดยจะให้ผู้ตัดสินที่ให้เปอร์เซ็นต์สูงสุด และต่ำสุดอธิบายเหตุผลที่ให้เปอร์เซนต์ต่ำสุด และสูงสุด จากนั้นเปิดโอกาสให้ผู้ตัดสินสามารถปรับเปลี่ยนผลการตัดสินของแต่ละคนได้ จากนั้นนำผลการตัดสินของแต่ละคนมาหาค่าเฉลี่ยก็จะได้คะแนนจุดตัด ดังตารางที่ 3 ตัวอย่างการคำนวณหาคะแนนจุดตัด

ตารางที่ 3 ตัวอย่างการคำนวณหาคะแนนจุดตัดด้วยวิธีแองกอฟที่ได้รับการปรับปรุงจากเปอร์เซนต์การตอบข้อสอบแต่ละข้อถูกของผู้สอบที่มีความสามารถขั้นต่ำ

ข้อสอบ	ผู้ตัดสิน(คนที่)											
	1	2	3	4	5	6	7	8	9	10	11	12
1	90	90	100	100	100	90	90	90	90	60	90	100
	80	90	90	100	90	90	90	90	60	70	90	90
2	80	90	90	40	100	80	100	70	80	90	100	70
	60	70	90	60	100	80	90	80	70	80	80	80
3	90	70	80	80	100	60	80	80	80	60	50	90
	90	80	90	70	80	60	70	80	80	60	60	90
4	70	60	70	80	90	80	80	70	70	60	50	90
	70	70	60	70	80	80	70	70	70	70	70	80
5	90	60	90	40	80	50	80	70	60	60	90	70
	80	70	90	60	80	60	70	70	70	70	80	70
6	60	60	80	60	70	70	80	80	60	50	70	80
	70	60	70	70	70	70	70	80	60	50	70	80

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 3 (ต่อ)

ข้อสอบ	ผู้ตัดสิน(คนที่)											
	1	2	3	4	5	6	7	8	9	10	11	12
7	90	50	80	60	60	70	70	70	70	60	80	60
	80	60	80	70	60	70	60	80	60	50	80	70
8	80	50	70	80	40	90	70	70	60	60	70	70
	70	50	80	70	50	90	70	80	70	70	70	80
9	80	70	60	70	60	80	50	60	60	30	50	60
	90	70	70	70	60	80	50	60	60	30	50	60
10	60	80	50	60	70	90	70	60	30	40	40	50
	70	80	60	70	80	90	80	70	40	50	60	60
ค่าเฉลี่ย												
	79	68	77	67	77	77	77	72	66	57	69	76
	78	70	78	71	75	77	74	77	69	63	72	72

จุดตัดคะแนน

วิธีแองกอฟได้รับการปรับปรุงอีกหลายครั้ง ซึ่ง Reckase (2000a, cite in Brandon, 2004) กล่าวว่า ยังไม่มีข้อยุติเกี่ยวกับนิยามของกระบวนการแองกอฟ แต่สามารถสรุปขั้นตอนกว้างๆ ได้ ดังนี้ เลือกผู้ตัดสิน ฝึกผู้ตัดสิน กำหนดนิยามและอธิบายระดับการปฏิบัติที่ผู้สอบควรพึงมี และผู้ตัดสินจะประมาณระดับความสามารถขั้นต่ำของผู้สอบในข้อสอบแต่ละข้อจนครบทุกข้อ หลังจากนั้นผู้ตัดสินจะได้รับข้อมูลเช่น ค่าความยาก(p-value) ซึ่งจะนำมาใช้ประกอบการอภิปรายร่วมกันในแต่ละรอบ โดยปกติประมาณ 2–3 รอบ จากนั้นจึงคำนวณหาคะแนนจุดตัด

จากที่กล่าวมาผู้วิจัยขอสรุปวิธีการแองกอฟที่ได้รับการปรับปรุง (Modified Angoff) ดังนี้

1) ผู้ตัดสินช่วยกันพิจารณาคุณลักษณะของกลุ่มผู้สอบที่มีความสามารถขั้นต่ำหรือกลุ่มผู้สอบคาบเส้นว่ามีลักษณะการตอบวิชานั้นๆอย่างไร

2) ผู้ตัดสินพิจารณาข้อสอบทีละข้อ แล้วตอบคำถามว่า กลุ่มผู้สอบที่มีความสามารถขั้นต่ำมีโอกาสการตอบถูกในแต่ละข้อ (Livingston and Ziekey, 1982) เป็นเท่าไร แต่ก็มีบางงานวิจัยตั้งคำถามว่า ข้อสอบแต่ละข้อกลุ่มผู้สอบคาบเส้นมีโอกาสตอบถูกหรือไม่ (ผู้ตัดสินจะตอบว่า ใช่ หรือ ไม่ใช่) แล้วจึงรวมคะแนนที่ผู้ตัดสินแต่ละคนที่ตัดสินในแต่ละข้อ (Ben-David, 2000) ได้มีงานวิจัยของ Impara and Plake (1997) ได้ทำการศึกษาเปรียบเทียบวิธีการแองกอฟที่ตั้งคำถามผู้ตัดสินเกี่ยวกับเปอร์เซนต์ในการตอบถูกของผู้สอบกลุ่มคาบเส้น กับวิธีการแองกอฟที่ตั้งคำถามผู้ตัดสินว่า ผู้สอบคาบเส้นมีโอกาสตอบถูกในข้อสอบข้อนั้นหรือไม่ (ผู้ตัดสินจะตอบว่า ใช่ หรือ ไม่ใช่) ผลการวิจัยพบว่าวิธีการแองกอฟที่ตั้งคำถามทั้งสองแบบให้คะแนนจุดตัดที่เท่ากัน และผู้ตัดสินยังกล่าวอีกว่า วิธีการตอบแบบใช่/ไม่ใช่ มีความสะดวกสบายกว่าวิธีการที่ให้ผู้ตัดสินตอบแบบกำหนดโอกาสในการตอบถูกในข้อสอบแต่ละข้อ

3) ผู้ตัดสินแต่ละคนร่วมกันอภิปรายผลการตัดสินของแต่ละคน โดยเฉพาะผู้ตัดสินที่มีผลตัดสินต่ำสุด และสูงสุดให้อธิบายแสดงความคิดเห็นเหตุผลที่ให้ผลการตัดสินต่ำสุดและสูงสุด และระหว่างการอภิปรายผู้อำนวยความสะดวกจะแสดงรูปการแจกแจงผลการตัดสินของผู้ตัดสินแต่ละคน

4) เมื่อผู้ตัดสินอภิปรายเสร็จ ก็เปิดโอกาสให้ตัดสินคะแนนอีกครั้งหนึ่ง ซึ่งการให้คะแนนอีกครั้งเป็นการเปิดโอกาสให้ผู้ตัดสินปรับเปลี่ยนคะแนน หรือยังคงคะแนนเดิมก็ได้ ผลการตัดสินที่ได้จากผู้ตัดสินแต่ละคนในรอบที่สองนำมาหาค่าเฉลี่ยก็คือ คะแนนจุดตัดนั้นเอง

จะเห็นได้ว่าวิธีการแองกอฟแบบเดิม และวิธีแองกอฟที่ปรับปรุงมีความเหมือนและความแตกต่างกันหลายประการ ผู้วิจัยจึงขอสรุปดังนี้

ความเหมือน

1) ทั้งสองวิธีต่างก็เป็นวิธีที่พิจารณาคะแนนจุดตัดจากข้อคำถามในแบบสอบ (test-center methods)

2) ทั้งสองวิธีเป็นวิธีการหาคะแนนจุดตัดจากกลุ่มผู้สอบที่มีความสามารถต่ำสุด หรือผู้สอบกลุ่มคาบเส้นเพื่อหาเปอร์เซ็นการตอบข้อสอบข้อนั้นถูก

3) ทั้งสองวิธีใช้กับแบบสอบแบบเลือกตอบ

ความแตกต่าง

1) วิธีแองกอฟแบบเดิมพิจารณาหาคะแนนจุดตัดเพียงรอบเดียว ในขณะที่วิธีแองกอฟที่ได้รับการปรับปรุงพิจารณาหาคะแนนจุดตัด 2-3 รอบ โดยจัดให้มีการอภิปรายร่วมกันระหว่างผู้ตัดสินและเปิดโอกาสให้ผู้ตัดสินเปลี่ยนผลการตัดสินในรอบแรกได้ ซึ่งจะทำให้คะแนนจุดตัดที่ได้มีความเที่ยงตรงมากขึ้น

2) วิธีแองกอฟที่ปรับปรุงเสนอข้อมูลสารสนเทศสำหรับพิจารณาประกอบการตัดสินใจแก่ผู้ตัดสิน เช่น ค่าความยาก หรือ ผลการตัดสินของผู้ตัดสินแต่ละคน

1.1.2 **วิธีของอีเบล** (Ebel's procedure) อีเบลได้กล่าวถึงจุดอ่อนของการนิยามคะแนนจุดตัด หรือคะแนนผ่านโดยการพิจารณาจากร้อยละของคะแนนรวมทำให้ขาดรายละเอียดอื่นๆ ที่จะนำมาใช้ในการพิจารณาด้วย เนื่องจากในบางครั้งข้อสอบอาจจะมีความยากมาก หรือง่ายเกินไป หรือมีอำนาจจำแนกต่ำมากเกินกว่าที่ผู้ออกข้อสอบคาดหวังได้ ดังนั้นการที่ผู้สอบจะสอบผ่านหรือไม่ จึงขึ้นอยู่กับคำถามในแบบสอบมากกว่าการพิจารณาจากระดับความสามารถของผู้สอบ อีเบลจึงเสนอวิธีการแก้ไขจุดอ่อนดังกล่าวโดยพิจารณาความเกี่ยวข้อง (Relevance) หรือความสำคัญ (Important) ของแต่ละข้อคำถามในแบบสอบ สำหรับในเรื่องของความเกี่ยวข้องอีเบลได้แบ่งออกเป็น 4 ประเภทคือ มีความจำเป็น (Essential) มีความสำคัญ (Important) เป็นที่ยอมรับได้ (Acceptable) และไม่แน่ใจ (Questionable) ส่วนในเรื่องระดับความยาก อีเบลได้แบ่งออกเป็น 3 ระดับ คือ ง่าย ปานกลาง และยาก และกำหนดร้อยละที่คาดหวังที่จะให้ผ่านข้อคำถามในแต่ละกลุ่มซึ่งร้อยละที่คาดหวังดังกล่าวถือว่าเป็นปริมาณความคาดหวังของผู้ที่มีความสามารถขั้นต่ำที่จะสอบผ่าน (Ebel, 1972, cited in Kane, 1998)

วิธีของอีเบล สามารถสรุปได้ดังนี้ (กัญจนา ลินทรัตนศิริกุล, 2534)

1) ให้ผู้ตัดสินแต่ละคนพิจารณาข้อคำถาม แล้วจัดแยกเป็น 12 ประเภท (คือลักษณะที่เกี่ยวข้องมี 4 ลักษณะ และลักษณะความยากมี 3 ระดับ)

2) อภิปรายแต่ละข้อคำถามดังนี้

- ถามผู้ตัดสินว่ามีกี่คนที่จัดคำถามอยู่ในระดับง่าย ปานกลาง และยาก ถ้าผู้ตัดสินจัดไม่เหมือนกัน ก็ควรจะให้อธิบายเหตุผล
- ถามผู้ตัดสินว่ามีกี่คนที่แยกประเภทของข้อคำถามออกเป็น "มีความจำเป็น มีความสำคัญ เป็นที่ยอมรับ และไม่แน่ใจ" ถ้าผู้เชี่ยวชาญแยกประเภทของข้อคำถามไม่เหมือนกัน ก็ควรให้อธิบายเหตุผลเช่นเดียวกัน

- ให้โอกาสแก่ผู้ตัดสินจัดประเภทของข้อคำถามใหม่ถ้าต้องการเปลี่ยนแปลง

3) ให้ผู้ตัดสินกำหนดร้อยละของผู้สอบที่คาดหวังว่าจะตอบในแต่ละข้อคำถามได้ถูกต้องทั้ง 12 ประเภท

4) รวบรวมผลการตัดสิน

วิธีการหาคะแนนจุดตัดของอีเบลคือ นำจำนวนข้อสอบในแต่ลักษณะคูณกับค่าร้อยละที่คาดหวังแล้วนำผลคูณที่ได้มาหาค่าเฉลี่ย ค่าเฉลี่ยที่ได้จะเป็นคะแนนจุดตัดหรือคะแนนที่ผู้มีความสามารถขั้นต่ำจะสอบผ่าน (Jaeger, 1989, cited in Kane, 1998)

1.1.3 **วิธีของนีเดลสกี (Nedelsky procedure)** นีเดลกีได้เสนอเทคนิคนี้ในปี ค.ศ. 1954 ซึ่งเป็นเทคนิคที่เหมาะกับข้อสอบแบบเลือกตอบโดยข้อสอบแต่ละข้อผู้ตัดสินจะตัดสินว่ามีจำนวนตัวเลือกตอบจำนวนเท่าไรที่ความสามารถของผู้สอบขั้นต่ำควรจะตอบผิด ระดับการผ่านขั้นต่ำสำหรับข้อสอบถูกคำนวณจากจำนวนตัวเลือกที่ยังคงเหลืออยู่ หลังจากตัวเลือกผิดถูกนำออกไปไม่ใช้ในการพิจารณา ถ้าผู้ตัดสินตัดสินว่า ความสามารถขั้นต่ำของผู้สอบควรกำจัดออกทั้งหมดยกเว้นตัวเลือก 3 ตัวเลือกในแบบสอบเลือกตอบในข้อหนึ่ง ระดับการผ่านขั้นต่ำในนีเดสสกี ควรเป็น 1/3 หรือ 0.33 ระดับการผ่านขั้นต่ำในทุกข้อจะถูกนำมารวมกันเพื่อกำหนดคะแนนจุดตัด (Nedelsky, 1954 ,cited in Kane,1998)

ดังนั้นหากกล่าวโดยสรุปวิธีของนีเดลสกี มีดังนี้ (กัญจนา ลินทรัตนศิริกุล, 2534)

1) ให้ผู้ตัดสินแต่ละคนพิจารณาคำถามทั้งหมด และใช้ดินสอขีดลงไปที่ตัวเลือกที่ผิด ที่เห็นว่าผู้สอบจะไม่เลือก

2) อภิปรายคำถามแต่ละข้อ ตามแนวทางดังนี้

- ให้ผู้ตัดสินพิจารณาคำตอบที่ผิดก่อน และดูว่ามีผู้ตัดสินกี่ท่านที่คิดว่า ผู้สอบจะไม่เลือก และมีกี่ท่านไม่เห็นด้วย
- ถ้าผู้ตัดสินมีความเห็นไม่ตรงกัน จะต้องให้แต่ละคนอธิบายเหตุผลแล้วถามผู้ตัดสินที่ไม่เห็นด้วย ซึ่งตอนนี้ยังไม่ต้องมีการสรุปจนกว่าผู้ตัดสินจะแน่ใจว่าได้เข้าใจความหมายของคำว่า ผู้สอบที่มีความสามารถขั้นต่ำได้ตรงกัน
- พิจารณาตัวเลือกที่เป็นตัวผิดตัวต่อไป

3) หลังจากที่มีการอภิปรายคำถามทุกข้อในลักษณะนี้แล้ว ขั้นตอนต่อมาคือ จะต้องให้ผู้ตัดสินทบทวนถึงการตัดสินโดยที่ผู้ตัดสินจะต้องทำเครื่องหมายกากบาทลงบนตัวเลือกที่เขาแน่ใจว่า ผู้สอบที่มีความสามารถต่ำจะไม่เลือก

4) รวบรวมผลการตัดสิน แล้วดำเนินการเช่นเดียวกันในคำถามต่อไป วิธีของนีเดลสกีนี้อาจให้ผู้ตัดสินแต่ละคนตัดสินเองที่ได้โดยปราศจากการอภิปรายร่วมกัน แต่ปัญหาที่จะตามมาก็คือ ผู้ตัดสินจะยังไม่คุ้นเคยกับวิธีการตัดสินเพราะฉะนั้นถ้าเป็นไปได้ก็ควรจะให้โอกาสแก่ผู้ตัดสินในการฝึกตัดสินใจ จากตัวอย่างของคำถามและอภิปรายซึ่งกันและกันก่อนที่จะมีการพิจารณาคำถามที่เหลือด้วยตนเองต่อไป

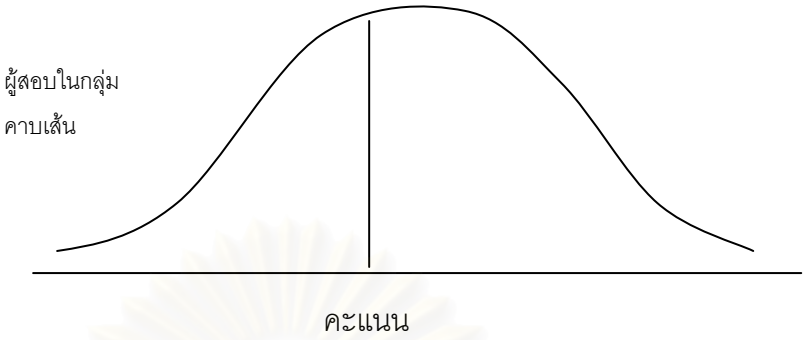
1.2 **วิธีผู้สอบเป็นศูนย์กลาง** (Examinee-centered methods) การกำหนดมาตรฐานด้วยวิธีผู้สอบเป็นศูนย์กลาง ผู้ตัดสินจะจัดจำพวกผู้สอบ จุดตัดจะถูกกำหนดโดยการกำหนดจุดที่สเกลคะแนน (score scale) ที่มีความคงที่มากที่สุดในกลุ่มนั้น วิธีผู้สอบเป็นศูนย์กลาง มี 2 วิธี ดังนี้

1.2.1 **วิธีกลุ่มคาบเส้น** (Borderline-group method) (Livingston & Zieky, 1982, cited in Kane, 1998) ผู้ตัดสินระบุผู้สอบแต่ละคนที่มีความสามารถระดับคาบเส้น ซึ่งผู้สอบคาบเส้นจะเป็นที่การปฏิบัติงานของพวกเขาผ่านมาตรฐานขึ้นมาเล็กน้อย (just meet the performance standard) ผู้ตัดสินอาจเป็นครู หรือผู้ควบคุมดูแล (supervisors) ฯลฯ ซึ่งพวกเขาเหล่านั้นจะมีประสบการณ์ต่อตัวผู้สอบแต่ละคน โดยผู้ตัดสินจะระบุผู้สอบที่มีความสามารถระดับคาบเส้น จากนั้นก็จะคำนวณมัธยฐานของคะแนนสอบของคนกลุ่มคาบเส้นก็จะได้คะแนนจุดตัด ข้อดีของวิธีนี้คือ ง่ายในการนำไปใช้และง่ายในการอธิบายแต่ข้อจำกัดของวิธีนี้ก็คือ ค่อนข้างยากที่จะตัดสินว่าผู้สอบอยู่ในกลุ่มที่มีลักษณะเป็นกลุ่มคาบเส้นจริงๆ

อย่างไรก็ตามหากนำวิธีนี้ไปใช้ มีขั้นตอนโดยสรุปดังนี้ (กัญจนา ลินทรัตนศิริกุล, 2534)

1) เลือกผู้ตัดสิน
2) นิยามความหมายของคำว่าระดับความรู้และทักษะของผู้สอบที่อยู่ในระดับคาบเส้น
3) ระบุกลุ่มผู้สอบที่เป็นกลุ่มคาบเส้น
4) เก็บรวบรวมคะแนนที่ได้จากการสอบของผู้สอบในกลุ่มคาบเส้น
5) กำหนดคะแนนมัธยฐานของผู้สอบในกลุ่มคาบเส้นให้เป็นคะแนนจุดตัด

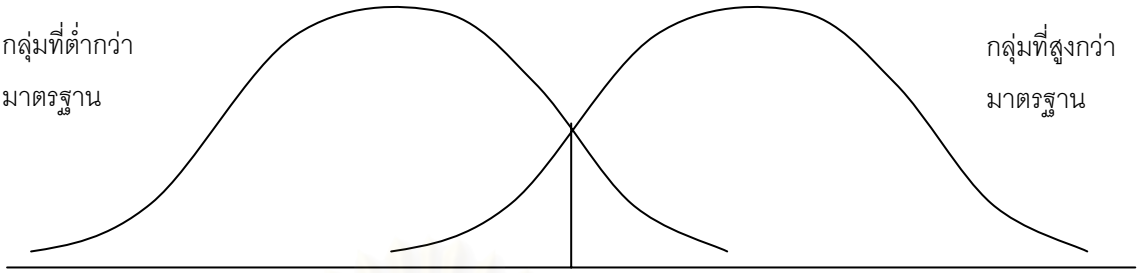
สามารถเขียนคะแนนการแจกแจงผู้สอบในกลุ่มคาบเส้น ดังแผนภาพที่ 2



แผนภาพที่ 2 คะแนนการแจกแจงผู้สอบในกลุ่มคาบเส้น

1.2.2 **วิธีกลุ่มตรงข้าม** (Contrasting-group method) (Livingston & Zieky, 1982, cited in Kane, 1998) ผู้ตัดสินจัดประเภทผู้สอบออกเป็น 2 กลุ่มคือ กลุ่มที่สูงกว่ามาตรฐาน และกลุ่มที่ต่ำกว่ามาตรฐาน วิธีกลุ่มตรงข้าม ผู้ตัดสินที่จะจัดเรียงอันดับ (rating) คือ ครู ผู้ควบคุมดูแล หรือบางคนที่มีประสบการณ์เกี่ยวกับการจัดระดับการปฏิบัติงานของผู้สอบ หรือคนที่มีประสบการณ์การวัดเกณฑ์ภายนอก หรือประเมินแบบสอบการปฏิบัติ หลังจากแจกแจงคะแนนสำหรับสองกลุ่มเพื่อใช้ตัดสินแล้วคะแนนจุดตัดจะถูกเลือกเพื่อจำแนกความเป็นไปได้ระหว่างกลุ่มสูง และกลุ่มต่ำโดยจุดที่คะแนนการแจกแจงของผู้สอบทั้งสองกลุ่มตัดกันจะเป็นคะแนนจุดตัด ดังแผนภาพที่ 3

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย



แผนภาพที่ 3 คะแนนการแจกแจงของผู้สอบกลุ่มที่ต่ำและสูงกว่ามาตรฐาน

จากที่กล่าวมาการกำหนดมาตรฐานแบบดั้งเดิมมี 2 วิธีคือ วิธีแบบสอบเป็นศูนย์กลาง (test-center methods) กับ วิธีผู้สอบเป็นศูนย์กลาง (Examinee-centered methods) วิธีแบบสอบเป็นศูนย์กลางเป็นวิธีที่ใช้ง่าย เหมาะกับแบบสอบปรนัย หรือแบบตอบสั้น ซึ่งแบบสอบลักษณะดังกล่าวทำเป็นข้อสอบจำนวนมากๆได้ แต่วิธีแบบสอบเป็นศูนย์กลางไม่เหมาะกับแบบสอบการปฏิบัติ ส่วนวิธีผู้สอบเป็นศูนย์กลาง เหมาะสำหรับการปฏิบัติในสภาพจริง (authentic performances) โดยวิธีนี้ผู้ตัดสินจะประเมินการปฏิบัติงานของนักเรียนแต่ละคน ซึ่งทำให้ผู้ตัดสินเสียเวลามาก (Kane, 1998) สำหรับวิธีย่อยๆทั้ง 3 วิธีในแบบสอบเป็นศูนย์กลาง วิธีที่ได้รับความนิยมใช้อย่างแพร่หลายคือ วิธีของแองกอฟ (Berk, 1996) แต่เนื่องจากในช่วงประมาณปี 1980 นักการศึกษาทั้งของอังกฤษและอเมริกา เริ่มตระหนักถึงความจำเป็นที่จะต้องปฏิรูปการศึกษา ความคิดที่จะเปลี่ยนแปลงในเรื่องเป้าหมายของการเรียนรู้ สาระการเรียนรู้ การจัดประสบการณ์การเรียนรู้ตลอดจนการวัดและประเมินผล ทำให้เกิดการศึกษาวิจัยเพื่อหาแนวทางในการปฏิรูปการศึกษา สำหรับการวัดและประเมินผลได้เริ่มมีการศึกษาถึงทางเลือกใหม่ของการประเมิน (alternative assessment) อาทิเช่น การประเมินการปฏิบัติงาน (performance assessment) การประเมินตามสภาพจริง (authentic assessment) เป็นต้น (Hart, 1994 อ้างถึงใน พวงแก้ว ปุณยกนก, 2546) ซึ่งทำให้วิธีการกำหนดมาตรฐานโดยวิธีแบบสอบเป็นศูนย์กลางเป็นวิธีการกำหนดมาตรฐานที่ไม่เหมาะสมกับการประเมินการปฏิบัติงาน และการประเมินตามสภาพจริง ดังนั้นในช่วงทศวรรษ 1990 ได้มีนักการศึกษาหลายท่านได้ปรับปรุงวิธีการกำหนดมาตรฐานเดิม หรือเสนอวิธีการกำหนดมาตรฐานใหม่เกิดขึ้นมากมาย

2. วิธีการกำหนดมาตรฐานแบบใหม่

ปัจจุบันในด้านการศึกษาและการรับรองได้ใช้การประเมินที่ประกอบด้วยข้อสอบที่มีทั้งแบบเขียนตอบ (constructed–response) และแบบเลือกคำตอบ (selected-response) (Parshall, Davey & Pashley, 2000, cited in Hambleton, Jaeger, Plake & Mills, 2000) เพื่อกำหนดมาตรฐานการเรียนรู้ของผู้สอบออกเป็นระดับ เช่น 5 ระดับคือ ตก, ต่ำกว่าพื้นฐาน, พื้นฐาน , เชี่ยวชาญ (proficient) และก้าวหน้า (advance) หรือถ้าแบ่งตามการรับรองสามารถแบ่งออกเป็น 2 ระดับคือ ได้รับการรับรอง และไม่ได้รับการรับรอง ส่วนการประเมินการปฏิบัติที่มีการให้คะแนนหลายระดับในงานๆหนึ่ง หรือเป็นงานที่ต้องใช้หลายๆ ทักษะเพื่อบ่งบอกถึงความสำเร็จ สิ่งเหล่านี้นับเป็นปัญหาในวิธีการกำหนดมาตรฐาน (standard setting methods) (Berk, 1986; Jaeger, 1989; Livingston & Zieky, 1982; Plake, 1998, cited in Hambleton, Jaeger, Plake & Mills, 2000) เนื่องจากวิธีการกำหนดมาตรฐานหลายวิธีในแบบดั้งเดิมไม่สามารถนำมาใช้กับการประเมินการเรียนรู้ที่ให้คะแนนมากกว่า 2 ค่าได้ (polytomously score) ทำให้มีการปรับ หรือพัฒนาวิธีการใหม่ขึ้นมาเพื่อใช้กับการประเมินการเรียนรู้

วิธีการกำหนดมาตรฐานแบบใหม่นี้ใช้สำหรับการประเมินที่ให้คะแนนแบบหลายค่า ซึ่งใช้ในการประเมินการเรียนรู้ (Cizek, 2001; Reckase, 2000, cited in Hambleton, Jaeger, Plake & Mills, 2000) มีด้วยกันหลายวิธีคือ

2.1 วิธีแองกอฟแบบปรับขยาย (Extended Angoff Method)

2.2 วิธีบุ๊คมาร์ค (Bookmark Method)

2.3 วิธีกลุ่มตรงข้าม (Contrasting Groups Method)

2.4 วิธีเลือกงานผู้สอบ (Examinee Paper Selections Method)

2.5 วิธีองค์รวม (Holistic Method)

2.6 วิธีเส้นภาพเด่น (Dominant Profile Method)

2.7 วิธีการนำนโยบายมาตัดสิน (Judgmental Policy Capturing Method)

แต่ละวิธีมีรายละเอียดดังนี้

2.1 วิธีการแองกอฟแบบปรับขยาย (Extended Angoff Method)

กระบวนการกำหนดมาตรฐานด้วยวิธีการแองกอฟมีมาตั้งแต่ปี 1971 โดยวิลเลียม แองกอฟ (William H. Angoff) ได้เสนอวิธีนี้ขึ้นมา และต่อมาแฮมเบลตัน และเพลค (Hambleton and Plake, 1995) ได้นำวิธีการแองกอฟมาปรับปรุงเพื่อใช้ในการประเมินการปฏิบัติ นั้นคือ วิธีแองกอฟแบบปรับขยายถูกพัฒนาสำหรับการประเมินการปฏิบัติที่ให้คะแนนมากกว่า 2 ค่านั้นเอง วิธีนี้กำหนดให้คณะผู้ตัดสินทำการประมาณคะแนนคาดหวังของผู้สอบคาบเส้น (ผู้สอบที่เกือบไม่ได้รับการรับรอง) (borderline candidates) ในแบบฝึกหัดแต่ละฉบับที่ให้คะแนนมากกว่า 2 ค่า งานของผู้ตัดสินก็คือ กำหนดความน่าจะเป็น (เช่น คะแนนที่คาดหวัง) ที่ผู้สอบที่คาบเส้นสามารถตอบข้อสอบถูก คะแนนที่คาดหวังจากแบบฝึกหัดของผู้ตัดสินแต่ละคนสามารถนำไปใช้ในมาตรฐานการปฏิบัติในชุดของแบบฝึดหัด จากนั้นมาตรฐานของผู้ตัดสินจะถูกนำมาเฉลี่ยเพื่อให้ได้มาตรฐานตามที่ต้องการสำหรับใช้ในการรับรองเพื่อให้เกิดความเข้าใจมากขึ้นจะขอยกตัวอย่างงานวิจัยของแฮมเบลตัน และเพลค (Hambleton and Plake, 1995)

แฮมเบลตัน และเพลค (Hambleton and Plake, 1995) กล่าวในงานวิจัยว่า วิธีนี้คณะผู้ตัดสิน (Panelists) จะประมาณคะแนนที่คาดหวังของผู้สอบที่คาบเส้น (borderline candidates) โดยใช้มาตรประมาณค่า 4 ระดับ สำหรับแบบฝึกหัดแต่ละฉบับซึ่งมีแบบฝึกหัดทั้งหมด 7 ฉบับ ที่รวมเป็นการประเมินการปฏิบัติหนึ่งงาน คณะผู้ตัดสินจะถูกถามเพื่อกำหนดน้ำหนักในแบบฝึกหัดเพื่อสะท้อนความเข้าใจถึงความสำคัญที่สัมพันธ์ของแบบฝึกหัดในการทำการตัดสิน ผ่าน-ตกในรอบสุดท้าย การตัดสินคะแนนที่คาดหวัง และน้ำหนักที่สัมพันธ์กันจะถูกทำหลายๆครั้งโดยการอภิปรายร่วมกันระหว่างผู้ตัดสินในแต่ละขั้นของกระบวนการกำหนดมาตรฐาน การอภิปรายจะช่วยให้คณะผู้ตัดสินมีฉันทมติ (consensus) ในมาตรฐานที่ออกมา

กล่าวโดยสรุป ขั้นตอนการกำหนดมาตรฐานด้วยวิธีการแองกอฟแบบปรับขยายตามแนวคิดของ Hambleton and Plake (1995) ดังต่อไปนี้

1. คณะผู้ตัดสินร่วมกันอภิปรายถึงลักษณะการปฏิบัติงานในแต่ละงานของผู้สอบคาบเส้นว่าควรเป็นแบบใด จนกระทั่งชัดเจน

2. ฝึกผู้ตัดสินให้เข้าใจถึงวิธีการให้คะแนน ลักษณะแบบฝึกหัดของนักเรียนแต่ละแบบฝึกหัดซึ่งแต่ละแบบฝึกหัดมีมิติย่อยๆ ที่ผู้ตัดสินต้องพิจารณาด้วย หลังจากนั้นผู้ตัดสินแต่ละคนกำหนดมาตรฐาน (คะแนนคาดหวังของผู้สอบคาบเส้น)ในมิติย่อยๆที่อยู่ในแบบฝึกหัดแต่ละแบบฝึกหัด จากนั้นผู้ตัดสินแต่ละคนก็กำหนดระดับความเชื่อมั่นของพวกเขาที่ได้กำหนด

มาตรฐานของเขาขึ้นมา เมื่อดำเนินการเรียบร้อยแล้ว นำมาตรฐานและระดับความเชื่อมั่นของผู้ตัดสินแต่ละคนมาหาค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐาน

3. นำผู้ตัดสินทั้งหมดมาอภิปรายร่วมกันถึงการกำหนดมาตรฐานของพวกเขา ซึ่งผู้อำนวยความสะดวกจะแสดงมาตรฐานที่ได้เฉลี่ยออกมาในแต่ละมิติของแบบฝึกหัดให้ตัดสินพิจารณาประกอบการอภิปราย และควรให้ผู้ตัดสินที่ให้ค่ามาตรฐานต่ำสุด และสูงสุดอภิปรายถึงการกำหนดมาตรฐานของพวกเขา แล้วให้ผู้ตัดสินคนอื่นร่วมกันอภิปรายถึงสิ่งที่ผู้ตัดสินให้ค่ามาตรฐานต่ำสุด และสูงสุดกล่าวถึง การดำเนินการอภิปรายจะดำเนินการจนกระทั่งผู้ตัดสินรู้สึกว่าการอภิปรายได้แสดงความคิดเห็นโดยสมบูรณ์แล้ว ก็จะดำเนินการในขั้นตอนต่อไป

4. ผู้ตัดสินแต่ละคนดำเนินการกำหนดมาตรฐาน (คะแนนคาดหวังของผู้สอบคาบเส้น) เป็นรอบที่ 2 และกำหนดระดับความเชื่อมั่นในมาตรฐานที่พวกเขากำหนดขึ้นมา

5. ผู้ตัดสินจะถูกถามเพื่อกำหนดน้ำหนักสัมพัทธ์ (relative weightings) ของแต่ละมิติโดยนำมาตรฐานที่กำหนดขึ้นประกอบการพิจารณาให้น้ำหนัก ขณะเดียวกันผู้ตัดสินก็กำหนดระดับความเชื่อมั่นที่พวกเขากำหนดน้ำหนักสัมพัทธ์ จากนั้นนำน้ำหนักสัมพัทธ์ของผู้ตัดสินแต่ละคนมาหาค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐาน

6. เมื่อผู้ตัดสินแต่ละคนกำหนดน้ำหนักสัมพัทธ์ในแต่ละมิติของแต่ละแบบฝึกหัดเรียบร้อยแล้วก็มาอภิปรายร่วมกัน โดยนำค่าเฉลี่ยของน้ำหนักสัมพัทธ์ของแต่ละมิติมาให้ผู้ตัดสินพิจารณาประกอบการอภิปราย และให้ผู้ตัดสินที่กำหนดน้ำหนักสัมพัทธ์สูงสุดและต่ำสุดอภิปรายถึงการกำหนดน้ำหนักสัมพัทธ์ของพวกเขา การอภิปรายดำเนินไปจนกระทั่งผู้ตัดสินทั้งหมดเห็นว่าการอภิปรายสมบูรณ์แล้วจึงดำเนินการขั้นต่อไป

7. ผู้ตัดสินจะถูกถามเพื่อกำหนดน้ำหนักสัมพัทธ์เป็นรอบที่ 2 ในแต่ละมิติของแต่ละแบบฝึกหัด และให้ผู้ตัดสินกำหนดระดับความเชื่อมั่นของการกำหนดน้ำหนักสัมพัทธ์

8. ผู้อำนวยความสะดวกนำค่ามาตรฐานในแต่ละมิติของแบบฝึกหัดมาคูณกับค่าน้ำหนักสัมพัทธ์ แล้วนำค่าของแต่ละมิติมาบวกกันก็จะได้ค่ามาตรฐานรวมในแบบฝึกหัดชิ้นนั้น ตัวอย่างเช่น แบบฝึกหัดชิ้นหนึ่งมี 3 มิติ จะมีมาตรฐานแต่ละมิติคือ 3, 4 และ 2 และมีค่าน้ำหนักสัมพัทธ์ในแต่ละมิติคือ .25 , .35 และ .40 ดังนั้นมาตรฐานของแบบฝึกหัดชิ้นนี้คือ (3 x .25)+(4 x .35)+ (2 x .40) = 2.95 หาค่ามาตรฐานรวมของทุกแบบฝึกหัดตามวิธีการข้างต้น

9. ผู้ตัดสินจะถูกถามเพื่อกำหนดน้ำหนักสัมพัทธ์ของแต่ละแบบฝึกหัดเช่น ถ้ามีแบบฝึกหัด 5 ฉบับ ก็จะมีน้ำหนักสัมพัทธ์ 5 ค่า (น้ำหนักสัมพัทธ์เฉลี่ยของผู้สอบแต่ละคน) และพิจารณาระดับความเชื่อมั่นถึงการกำหนดน้ำหนักสัมพัทธ์เหล่านั้น

10. ผู้อำนวยความสะดวกนำค่ามาตรฐานของแบบฝึกหัดแต่ละชิ้น (คำนวณได้ตามข้อ 8) มาคูณกับค่าน้ำหนักสัมพัทธ์ของแบบฝึกหัดแต่ละชิ้น (จากข้อ9) แล้วจึงนำผลคูณที่ได้ในแต่ละแบบฝึกหัดมาบวกกัน ก็จะได้ค่ามาตรฐานของแบบฝึกหัดทั้งหมดในภาพรวม

จุดแข็งและข้อบกพร่อง

จุดแข็ง วิธีการแองกอฟแบบปรับขยายเป็นวิธีที่ง่ายชัดเจน กระบวนการให้น้ำหนักทำให้ผู้ตัดสินกำหนดค่าของคำถามแต่ละข้อแตกต่างกัน การที่ให้น้ำหนักมากในบางคำถามแสดงว่า ผู้ตัดสินรู้สึกว่ามีความสำคัญมาก (Hambleton, Jaeger, Plake& Mills, 2000) นอกจากนี้ ผู้ตัดสินยังเกิดความรู้สึกมั่นใจเกี่ยวกับความสามารถของพวกเขาว่า มีความถูกต้องแม่นยำในขั้นตอนของกระบวนการ (Plake, Hambleton& Jaeger, 1997)

ข้อบกพร่อง วิธีการแองกอฟแบบปรับขยายอาจใช้ได้ไม่ดี เมื่อนำไปใช้กับการปฏิบัติที่เป็นธรรมชาติที่มีความยืดหยุ่นสูง (Hambleton, Jaeger, Plake& Mills., 2000) นอกจากนี้ในงานวิจัยของแฮมเบลตัน และเพลค (Hambleton and Plake, 1995) คณะผู้ตัดสินได้กล่าวถึงปัญหาด้านความตรงของมาตรฐานรอบสุดท้าย โดยคณะผู้ตัดสินต้องการให้มีการใช้นโยบายในการกำหนดมาตรฐานที่ซับซ้อนมากกว่านี้เพื่อให้เกิดการเชื่อมโยง (conjunctive) และ/หรือ กฎการทดแทนการตัดสินใจ (compensatory decision rules) โดยวิธีการแองกอฟแบบปรับขยายเป็นการทดแทน (compensatory) เพราะคะแนนการผ่าน (passing score) รอบสุดท้ายเป็นการรวมน้ำหนักของคะแนนแบบฝึกหัดแต่ละแบบฝึกหัด ดังนั้นผู้สอบที่ทำแบบฝึกหัดหลายแบบฝึกหัด ซึ่งพบว่ามีบางแบบฝึกหัดได้คะแนนการปฏิบัติต่ำ แต่ผู้สอบคนนี้ก็ยังคงผ่านการประเมิน เนื่องจากว่าการปฏิบัติของนักเรียนคนนั้นในแบบฝึกหัดอื่นที่เหลือได้คะแนนสูง สิ่งนี้เป็นหลักการของวิธีการทดแทน (compensatory approach)

2.2 วิธีบุ๊คมาร์ค (Bookmark Method)

วิธีการบุ๊คมาร์คเป็นหนึ่งในกระบวนการผังข้อสอบ (item-mapping procedure) ที่มีหลายกระบวนการ (Cizek; Bunch & Koons, 2004) กระบวนการกำหนดมาตรฐานบุ๊คมาร์ค (bookmark standard setting) ได้รับการเสนอขึ้นเป็นครั้งแรกโดยลิวิส มิเซล และกรีน (Lewis, Mizel and Green) ในปี ค.ศ. 1996 วิธีการบุ๊คมาร์คได้รับการพัฒนาเพื่อให้ผู้ตัดสินมีกระบวนการคิด(cognitive) เพื่อตัดสินได้ง่ายขึ้น วิธีนี้ได้รับความนิยมอย่างรวดเร็ว และได้

มีการใช้บุ๊คมาร์คในการกำหนดคะแนนจุดตัดที่เป็นการประเมินระดับประเทศ (large-scale assessment) วิธีบุ๊คมาร์คถูกนำมาใช้กับแบบสอบที่ประกอบด้วยแบบเลือกตอบ และแบบเขียนตอบ

กระบวนการบุ๊คมาร์คเป็นการรวมเนื้อหาข้อสอบกับผลการสอบของผู้สอบเข้าด้วยกัน กระบวนการนี้เปิดโอกาสให้ผู้ตัดสินอภิปรายและให้โอกาสเปลี่ยนคะแนนจุดตัดที่เคยกำหนดในรอบแรกเป็นคะแนนจุดตัดอันใหม่ได้ ประโยชน์อีกประการหนึ่งของบุ๊คมาร์คคือสามารถใช้กับแบบสอบที่มีรูปแบบผสมที่ประกอบด้วย ข้อสอบแบบเลือกตอบ (ให้คะแนนแบบ 2 ค่า) และข้อสอบแบบเขียนตอบ (ให้คะแนนมากกว่า 2 ค่า) นอกจากนี้บุ๊คมาร์คยังออกแบบมาให้ง่ายต่อการตัดสินโดยมีคู่มือจัดเรียงข้อสอบจากข้อง่ายไปข้อยาก (Mitzel, Lewis, Patz, & Green, 2001, cite in Beretvas, 2004)

คู่มือจัดเรียงข้อสอบ (ordered item booklet) เป็นสิ่งสำคัญที่ใช้ในกระบวนการบุ๊คมาร์ค ข้อสอบจะถูกจัดเรียงตามค่าความยากจากข้อสอบง่ายสุดไปยังข้อสอบยากสุด โดยใช้กระบวนการ IRT เพื่อให้ทราบถึงค่าความยากของข้อสอบ และความสามารถของนักเรียนที่อยู่บนมาตรเดียวกัน (same continuum) ข้อสอบแบบเขียนตอบก็จะถูกนำมาไว้ในมาตรเดียวกับข้อสอบแบบเลือกตอบโดยคะแนนแต่ละคะแนนในข้อสอบแบบเขียนตอบจะถูกนำมากำหนดบนมาตร แล้วจัดเรียงลำดับลงในคู่มือ (Egan, 2002; Kiplinger, 1997) ตัวอย่างเช่น ข้อสอบข้อที่ 37 เป็นข้อสอบเขียนตอบที่ให้คะแนนแบบ 0, 1, 2 ซึ่งคะแนนเหล่านี้จะถูกนำมาเป็นตัวแทนลงบนสเกลความสามารถ (ability scale) และในคู่มือ โดยจะมีข้อสอบข้อที่ 37.1 แทนข้อที่ให้คะแนน 1 คะแนน ซึ่งจะถูกกำหนดให้ใกล้กับจุดเริ่มต้นของคู่มือมากกว่า 2 คะแนน ขณะที่ข้อสอบข้อที่ 37.2 แทนข้อสอบที่ให้คะแนน 2 คะแนน จะถูกกำหนดให้ใกล้ท้ายเล่มของคู่มือเพราะว่า 2 คะแนนทำได้ยากกว่า 1 ส่วนคะแนน 0 คะแนนจะไม่ถูกนำมาใช้เพราะว่า มันแสดงถึงการไม่มีคะแนนคำตอบ เป็นต้น

เมื่อเตรียมคู่มือจัดเรียงข้อสอบเรียบร้อยแล้ว ขั้นตอนต่อไปเป็นกระบวนการหาคะแนนจุดตัดโดยผู้ตัดสินแต่ละคนได้รับคู่มือจัดเรียงข้อสอบซึ่งลักษณะของคู่มือคือ มีข้อสอบเหมือนกับแบบสอบที่ใช้สอบ เพียงแต่ว่าคู่มือมีการจัดเรียงข้อสอบใหม่ตามความยากของข้อสอบ (ข้อสอบ 1 ข้อต่อ 1 หน้า) เมื่อผู้ตัดสินได้รับคู่มือแล้ว ผู้ตัดสินก็เริ่มพิจารณาข้อสอบทีละข้อไปเรื่อยๆ โดยพิจารณาว่าผู้สอบกลุ่มคาบเส้นมีโอกาสตอบข้อสอบข้อนั้นถูก 2 ใน 3 หรือ 67% เกณฑ์ตรงนี้เรียกว่าค่า RP (response probability) ค่า RP 2/3 เหมาะนำมาใช้กับวิธีการกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์ค (Buckendahl, Smith, Impara, & Plake, 2002; Mizel et al, 2001; Reckase, 2000; Skaggs& Tessema, 2001) โดย Huynh (1998) แสดงให้เห็นว่า โมเดล IRT แบบ 3 พารามิเตอร์ ฟังก์ชันสารสนเทศของข้อสอบ (Item information function) จะมีค่าสูงสุดที่

θ สำหรับ $P(\theta)$ = (c+2)/3 เมื่อไม่ได้พิจารณาการเดา (c = 0) ค่า RP จะมีค่าเท่ากับ 2/3 นั้นคือ Huynh พบว่า RP=.67 ให้ค่าฟังก์ชันสารสนเทศข้อสอบสูงสุดนั้นเอง

สำหรับการกำหนดบุ๊คมาร์ค (กำหนดคะแนนจุดตัด) ดำเนินการซ้ำกัน 3 รอบ แต่ละรอบทำขึ้นเพื่อช่วยเพิ่มฉันทามติ และลดความแตกต่างของผู้ตัดสิน

รอบที่ 1

เป้าหมายหลักของรอบ 1 เพื่อให้ผู้ตัดสินคุ้นเคยกับคู่มือการจัดเรียงข้อสอบซึ่งเป็นการดำเนินการขั้นต้นของการกำหนดบุ๊คมาร์ค ในรอบนี้ผู้ตัดสินทำงานเป็นกลุ่มย่อยโดยการอภิปรายข้อสอบเป็นรายข้อว่า อะไรที่ทำให้ข้อสอบข้อนี้ยากกว่าข้อสอบข้อก่อน การบรรยายการปฏิบัติที่มีระดับต่างกัน (เช่น ระดับพื้นฐาน ระดับชำนาญการ และระดับสูง) จะถูกนำมาแสดงและอภิปรายร่วมกัน จากนั้นผู้ตัดสินถูกถามเพื่ออภิปราย และตัดสินเนื้อหาว่า นักเรียนควรมีความรอบรู้ที่จะอยู่ในระดับการปฏิบัติใด การตัดสินของผู้ตัดสินจะทำการตัดสินคะแนนจุดตัดอิสระจากกันโดยการนำที่คั่นหนังสือมาคั่นระหว่างข้อสอบที่ถูกตัดสินว่าเป็นตัวแทนของคะแนนจุดตัด ที่คั่นหนังสือ 1 อันจะถูกคั่นสำหรับแต่ละคะแนนจุดตัดที่ต้องการ ข้อสอบก่อนหน้าที่คั่นหนังสือนี้จะได้รับการพิจารณาว่า นักเรียนทั้งหมดถูกคาดหวังว่ารู้ และสามารถปฏิบัติงานได้สำเร็จที่ความน่าจะเป็นอย่างน้อย 0.67 ในทางตรงกันข้ามนักเรียนเหล่านี้ถูกคาดหวังว่าสามารถปฏิบัติงานได้สำเร็จในข้อสอบที่อยู่หลังจากข้อสอบที่คั่นหนังสือที่ความน่าจะเป็นน้อยกว่า 0.67

รอบที่ 2

กิจกรรมอย่างแรกในรอบที่ 2 เกี่ยวข้องกับผู้ตัดสินแต่ละคนทำการคั่นหนังสือลงในคู่มือจัดเรียงข้อสอบของแต่ละคน ตัวอย่างเช่น ผู้ตัดสินกลุ่มหนึ่งมี 6 คน คู่มือจัดเรียงข้อสอบของผู้ตัดสินทั้ง 6 คน มีที่คั่นหนังสือ 6 อันสำหรับแต่ละคะแนนจุดตัด ในการอภิปรายจะให้เน้นข้อสอบที่อยู่ระหว่างที่คั่นหนังสืออันแรก และอันสุดท้ายสำหรับแต่ละระดับการปฏิบัติ เมื่อการอภิปรายสิ้นสุดลงผู้ตัดสินทำการคั่นหนังสือลงในคู่มือจัดเรียงข้อสอบอีกครั้ง ค่าเฉลี่ยในการกำหนดบุ๊คมาร์ค (คั่นหนังสือ) ในรอบที่ 2 สำหรับแต่ละคะแนนจุดตัด จะถูกนำมาเป็นข้อเสนอแนะของกลุ่มเพื่อใช้กำหนดคะแนนจุดตัดต่อไป

รอบที่ 3

รอบที่ 3 ผู้ตัดสินรวมกันเป็นกลุ่มใหญ่ จากนั้นเริ่มจากการแสดงข้อมูลผลกระทบ (impact data) ให้ผู้ตัดสินพิจารณา เปอร์เซนต์การตกของนักเรียนแต่ละระดับการปฏิบัติถูกนำมาแสดง ให้ค่าเฉลี่ยคะแนนจุดตัดของแต่ละกลุ่มที่ได้จากรอบที่ 2 กับข้อมูลของนักเรียนว่ามีการปฏิบัติอย่างไร ผู้ตัดสินอภิปรายจุดที่คั่นหนังสือในกลุ่มใหญ่ จากนั้นจึงให้ผู้ตัดสินดำเนินการ

ตัดสินอีกครั้งอย่างอิสระจากกันโดยการให้ผู้ตัดสินคั่นหนังสือ ค่าเฉลี่ยคะแนนจุดตัดของกลุ่มใหญ่ถูกพิจารณาเพื่อเป็นคะแนนจุดตัดขั้นสุดท้ายสำหรับทำเป็นระดับการปฏิบัติ

เมื่อได้คะแนนจุดตัดในรอบสุดท้ายแล้ว ขั้นตอนต่อไปคือ การบรรยายระดับการปฏิบัติ (performance level descriptors) โดยผู้ตัดสินช่วยกันเขียนการบรรยายการปฏิบัติในประเด็น ความรู้เฉพาะ ทักษะ และความสามารถของนักเรียน ณ ระดับการปฏิบัติที่กำหนดไว้ ข้อสอบก่อนที่ทำการคั่นหนังสือถูกนำมาพิจารณาเนื้อหาว่า นักเรียน ณ ระดับการปฏิบัตินี้ถูกคาดหวังว่า มีความสามารถในการตอบคำถามถูกด้วยความน่าจะเป็นอย่างน้อย 0.67 ความรู้และทักษะที่นำมาตอบคำถามได้สำเร็จในข้อสอบเหล่านี้ถูกสังเคราะห์เข้าอย่างเป็นระบบเป็นการบรรยายสำหรับการปฏิบัตินี้ ด้วยเหตุนี้การบรรยายระดับการปฏิบัติเกิดการขยายออกมาของกระบวนการกำหนดคะแนนจุดตัด

จุดดีและข้อบกพร่อง

จุดดี

1) วิธีนี้ผู้ตัดสินพิจารณาแบบสอบในภาพรวมโดยอยู่บนพื้นฐานการพิจารณาจากผลคะแนนของนักเรียน

2) วิธีบุ๊คมาร์คถูกออกแบบมาให้ใช้กับข้อสอบแบบเลือกตอบที่ให้คะแนนแบบ 2 ค่า และ ข้อสอบแบบเขียนตอบที่ให้คะแนนมากกว่า 2 ค่า อย่างเท่าเทียมกันเพราะข้อสอบทุกข้ออยู่บนสเกล IRT เหมือนกัน

3) กระบวนการบุ๊คมาร์คอยู่บนพื้นฐานที่มีข้อมูลเชิงประจักษ์ และทำให้การตัดสินของผู้ตัดสินเป็นไปด้วยความง่าย เพราะ วิธีการนี้ผู้ตัดสินจะไม่ได้ถูกถามเพื่อประมาณอย่างคาดคะเนเองถึงระดับความยากของข้อสอบแต่ละข้อในกลุ่มนักเรียนที่สมมุติขึ้นเอง แต่ผู้ตัดสินจะถูกถามเพื่อให้คั่นหน้าในคู่มือที่เรียงลำดับข้อสอบไว้ว่าหน้าไหนที่พวกเขารู้สึกว่า ข้อสอบหน้านั้นมีความยากของข้อสอบเกินจากระดับที่คาดว่า นักเรียนที่มีความรู้จะสามารถทำได้ นั้นคือ ผู้ตัดสินกำหนดคะแนนจุดตัดตามที่การปฏิบัติงานของนักเรียนควรจะเป็นและตามเงื่อนไขการปฏิบัติที่นักเรียนแสดงออกมา

ข้อบกพร่อง

1) บางครั้งไม่มีข้อสอบจำนวนมากพอสำหรับพิจารณานักเรียนที่อยู่ในมาตรฐานระดับ ก้าวหน้า หรือระดับตก ซึ่งทั้งสองระดับเป็นระดับสุดโต่ง(เก่งมาก กับอ่อนมาก) ซึ่งอาจทำให้เกิด floor and ceiling effect ได้

2) วิธีการนี้ไม่ได้แสดงข้อมูลการปฏิบัติงานของนักเรียนในสเกลรายงานผลซึ่งการรายงานนี้นำมาใช้ตัดสินมาตรฐานการปฏิบัติรอบสุดท้าย ตัวอย่างเช่น ข้อสอบอาจจะกำหนดสเกลรายงานผลที่ระดับความสามารถที่นักเรียนมีโอกาสทำได้สำเร็จ 50% หรือกำหนดสเกลรายงานผลที่ระดับความสามารถของนักเรียนที่มีโอกาสทำได้สำเร็จ 65% ลักษณะแบบนี้เป็นปัญหาการกำหนดที่มีความหลากหลายซึ่งอาจส่งผลกระทบต่อมาตรฐานการปฏิบัติได้ ดังนั้นจึงต้องมีการวิจัยในประเด็นดังกล่าว

ดังนั้นเพื่อให้เกิดความชัดเจน ผู้วิจัยได้สรุปความเหมือน และความแตกต่างระหว่างวิธีการกำหนดมาตรฐานด้วยวิธีแองกอฟ และวิธีบุ๊คมาร์คดังนี้

ความเหมือน

1) ทั้งสองวิธีต่างก็เป็นวิธีที่พิจารณาหาคะแนนจุดตัดจากข้อคำถามในแบบสอบ(test-center methods)

2) ทั้งสองวิธีเป็นวิธีการหาคะแนนจุดตัดจากกลุ่มผู้สอบที่มีความสามารถต่ำสุด หรือผู้สอบกลุ่มคาบเส้นเพื่อหาโอกาสการตอบข้อสอบข้อนั้นถูก

ความแตกต่าง

1) วิธีบุ๊คมาร์คเป็นวิธีที่มีการจัดเตรียมสารสนเทศให้กับผู้ตัดสินเพื่อช่วยพิจารณากำหนดคะแนนจุดตัด โดยจัดเตรียมในรูปคู่มือจัดเรียงข้อสอบตามความยาก ในขณะที่วิธีแองกอฟผู้ตัดสินพิจารณาจากตัวแบบทดสอบโดยไม่มีการจัดเรียงข้อสอบตามความยาก

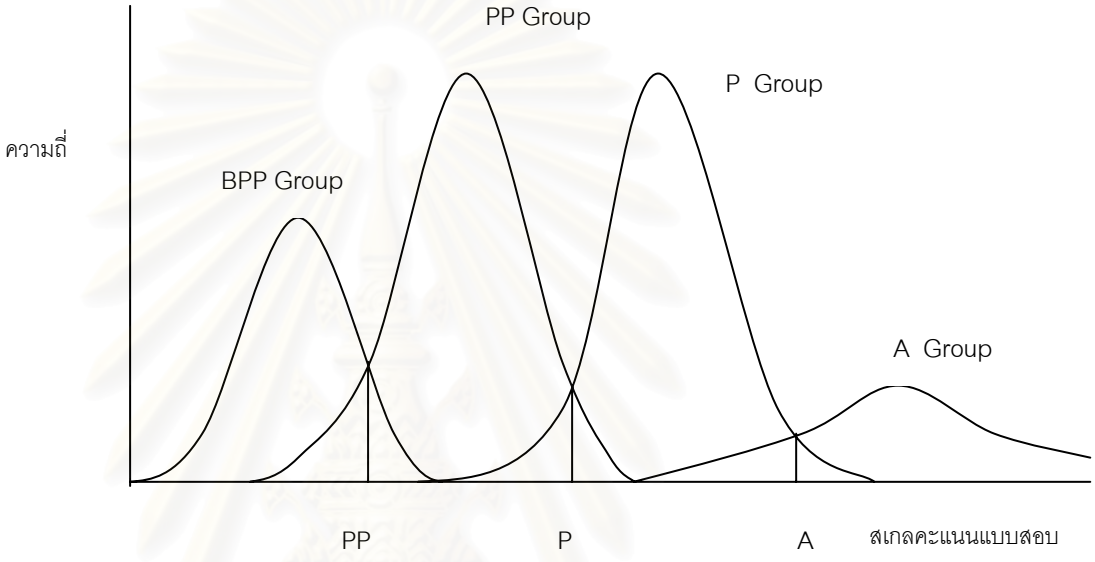
2) วิธีบุ๊คมาร์คสามารถนำไปใช้กับแบบทดสอบที่มีรูปแบบผสมคือ มีทั้งข้อสอบแบบเลือกตอบและข้อสอบแบบเขียนตอบอยู่ในฉบับเดียวกัน ในขณะที่วิธีการแองกอฟเหมาะกับแบบสอบแบบเลือกตอบ

3) วิธีการบุ๊คมาร์คสามารถนำไปใช้หาคะแนนจุดตัดที่มีหลายระดับได้ ส่วนวิธีแองกอฟไม่เหมาะที่จะนำไปใช้หาคะแนนจุดตัดที่มีหลายระดับ

2.3 วิธีกลุ่มตรงข้าม (Contrasting Groups Method)

การกำหนดมาตรฐานการปฏิบัติวิธีนี้ทำโดยการเปรียบเทียบการแจกแจงคะแนนสอบของผู้สอบโดยผู้สอบจะถูกแบ่งออกเป็นกลุ่มๆ ผู้ตัดสินเป็นผู้ที่ทราบว่านักเรียนแต่ละคนมีความสามารถเป็นอย่างไร (เช่น ครูทราบความสามารถทางคณิตศาสตร์ของนักเรียน) ผู้ตัดสินถูกถามเพื่อให้จำแนกนักเรียนออกเป็นกลุ่มตามมาตรฐานที่กำหนด (เช่น ทำงานต่ำกว่า

ระดับเกณฑ์ ทำงานได้ตามระดับเกณฑ์ ทำงานได้สูงกว่าระดับเกณฑ์) ผู้สอบถูกจัดจำพวกโดยไม่ใช้ความรู้จากคะแนนสอบของพวกเขา แต่คะแนนสอบของผู้สอบแต่ละคนจะถูกนำมาจัดจำพวกในรูปการแจกแจงความถี่ ผลจากการแจกแจงความถี่นำมาเปรียบเทียบเพื่อตัดสินระดับของการซ้อนกัน (overlap) คะแนนจุดตัดจะถูกจำแนกเพื่อใช้แบ่งระหว่างกลุ่มให้เป็นมาตรฐานการปฏิบัติ



BPP = ระดับต่ำกว่าความชำนาญการขั้นต่ำ (Below Partially Proficient)
PP = ระดับความชำนาญการขั้นต่ำ (Partially Proficient)
P = ระดับชำนาญการ (Proficient)
A = ระดับขั้นสูง (Advanced)

แผนภาพที่ 4 คะแนนการแจกแจงกลุ่มผู้สอบคาบเส้นสำหรับระดับการปฏิบัติหลายระดับ

จุดดีและข้อบกพร่อง

ข้อดีของวิธีนี้คือ เป็นวิธีที่ต้องการให้ผู้ตัดสินทำการตัดสินนักเรียนเป็นรายบุคคล โดยจัดให้กลุ่มที่คล้ายๆ กันอยู่ระดับเดียวกัน แต่การจัดจำพวกผู้สอบดังกล่าวมีความเป็นไปได้ที่จะผิดพลาดเพราะการแจกแจงความถี่มีการทับซ้อนกันทำให้การกำหนดมาตรฐานการปฏิบัติทำได้ยาก นอกจากนี้ครูอาจจะลังเลที่จะให้นักเรียนของพวกเขาเข้ากลุ่มต่ำ เนื่องจากครูเหล่านี้ให้เกรดนักเรียนเหล่านี้ผ่านมาแล้ว ปัญหาใหญ่อีกประการหนึ่ง คือ การทำให้การกำหนดมาตรฐานสามารถสรุปอ้างอิง (generalizability) ไปยังประชากรผู้สอบเป็นไปได้ยาก ถ้ากลุ่มตัวอย่างผู้สอบ

ที่ใช้ในการวิเคราะห์ไม่เป็นตัวแทนที่ดีแล้ว มาตรฐานการปฏิบัติก็จะไม่เหมาะสม (Hambleton, Jaeger, Plake & Mills, 2000)

2.4 วิธีเลือกงานผู้สอบ (Examinee Paper Selections Method)

วิธีการนี้ใช้กับการประเมินการปฏิบัติงานที่ให้คะแนนมากกว่า 2 ค่า วิธีนี้คล้ายคลึงกับวิธีกลุ่มคาบเส้น (borderline group method) ซึ่งเป็นวิธีที่ใช้กับการให้คะแนนแบบ 2 ค่า เพียงแต่ว่าคณะผู้ตัดสินพิจารณาการตอบของผู้สอบมากกว่าการจัดเรียงอันดับผู้สอบอย่างเป็นอิสระ คณะผู้ตัดสินจะเลือกชิ้นงาน (papers) ที่พวกเขารู้สึกว่าเป็นตัวแทนงานของกลุ่มผู้สอบที่คาบเส้นสำหรับในแต่ละข้อ คะแนนเฉลี่ยจากชิ้นงานที่ถูกเลือกในแต่ละระดับมาตรฐานการปฏิบัติงานจะถูกใช้เป็น “ค่าการผ่านขั้นต่ำ” (minimum passing value) สำหรับแต่ละข้อคำถาม ค่าขั้นต่ำจะถูกรวมเพื่อใช้ตัดสินมาตรฐานการปฏิบัติแต่ละมาตรฐานในแบบทดสอบ

จุดดีและข้อบกพร่อง

คณะผู้ตัดสิน (โดยเฉพาะครู) ชอบวิธีนี้เพราะพวกเขามีโอกาสในการตัดสินใจโดยมีงานของผู้สอบประกอบการพิจารณา อย่างไรก็ตามวิธีนี้นำมาใช้กับผู้สอบกลุ่มเล็กๆ ได้ยากเนื่องจากการแสดงชิ้นงานจากการแจกแจงงานของผู้สอบเป็นไปได้ยาก ดังนั้นจึงต้องใช้ความพยายามอย่างมากเพื่อเลือกชิ้นงานของผู้สอบ ระบุระบบรหัส จัดระบบกล่องบรรจุชิ้นงานของผู้สอบแก่ผู้ตัดสิน และใช้ชิ้นงานจำนวนมากเพื่อให้คณะผู้ตัดสินประเมินเนื่องจากว่าไม่มีการกำหนดน้ำหนักองค์ประกอบในแต่ละส่วนให้แตกต่างกัน ซึ่งข้อความแต่ละข้อจะถูกนำไปกำหนดเป็นแต่ละมาตรฐานการปฏิบัติ แม้ว่ามีบางองค์ประกอบที่อาจมีค่าน้อยกว่าองค์ประกอบอื่น นอกจากนี้ยังไม่มีเหตุผลว่า ทำไมการให้น้ำหนักจึงไม่ได้นำมาใช้กับกระบวนการเลือกชิ้นงานซึ่งทำให้การนำใช้ปฏิบัติได้ไม่ดี (Hambleton, Jaeger, Plake& Mills, 2000)

2.5 วิธีองค์รวม (Holistic Method)

วิธีนี้คล้ายกับวิธีเลือกงานผู้สอบ (examinee paper selection method) ถูกเสนอโดย the National Academy of Education (1993) เพื่อกำหนดมาตรฐานของ NAGB และ ACT วิธีองค์รวม (holistic) คณะผู้ตัดสินถูกถามเพื่อให้พิจารณางานที่สัมบูรณ์ (ประกอบด้วยแบบฝึกหัดทั้งหมด หรืองานในการประเมินผลของผู้สอบ) และตัดสินว่าคู่มือเล่มไหนที่เป็นตัวแทนกลุ่มผู้สอบที่คาบเส้น (หรือผู้สอบรอบรู้และไม่รอบรู้ หรือ พื้นฐาน ชำนาญการ และก้าวหน้า) ในบางกรณีมาตรประมาณค่าจะนำมาใช้สำหรับคณะผู้ตัดสินเพื่อแบ่งชิ้นงานของผู้สอบออกเป็นกลุ่มๆ เช่น ต่ำกว่าพื้นฐาน แยกเป็นสูง กลาง ต่ำ และสูงกว่าพื้นฐานแยกเป็นสูง

กลาง ต่ำ วิธีนี้ถูกนำไปใช้เป็นทางเลือกเช่นเดียวกับวิธีแองกอฟในแบบสอบแบบเลือกตอบโดยตรวจสอบข้อวิจารณ์ว่า วิธีแองกอฟ (เน้นระดับข้อสอบ) มีจุดอ่อนตรงการปฏิบัติงานของผู้สอบ NAGB และ ACT ได้นำวิธีนี้มาทดสอบกับข้อมูลของ NAEP ในหลายวิชา ระดับชั้นเกรด 4 ,8 และ 12

จุดดีและข้อบกพร่อง

ข้อดี คือ คณะผู้ตัดสินใช้งานของผู้สอบจริงๆ นอกจากนี้คณะผู้ตัดสินยังได้พิจารณางานของผู้สอบอย่างยืดหยุ่นโดยให้ผู้ตัดสินมีความรู้สึกเมตตา (forgiving) มากกว่าวิธีที่จัดเรียงลำดับงานต่องาน อย่างไรก็ตามข้อคำถามที่เหลือคณะผู้ตัดสินสามารถนำมาใช้เพื่อหาข้อมูลจากผู้สอบได้ แต่ถ้าคู่มือมีความหนาหรือยาวมาก คณะผู้ตัดสินจะทำให้มีความตรง และความเที่ยงตรงในการตัดสินเกี่ยวกับคุณภาพของงานได้หรือไม่ (Hambleton, Jaeger, Plake& Mills, 2000)

2.6 วิธีเส้นภาพเด่น (Dominant Profile Method)

วิธีนี้ถูกใช้สำหรับการประเมินงานหรือแบบฝึกหัดจำนวนจำกัด แต่ละงานมีการให้คะแนนแบบหลายค่า วิธีนี้ใช้ยุทธวิธีสร้างฉันทมติ (consensus-building) ผู้ตัดสินจะถูกถามเพื่อให้เกิดกฎที่เป็นนโยบาย (ตัดสินใจ) เพื่อใช้กำหนดคะแนนงานหรือแบบฝึกหัดที่ใช้ประกอบการประเมินเป็นค่าต่ำสุดที่ต้องการให้ผ่าน องค์ประกอบการทดแทน (compensatory) และองค์ประกอบที่เชื่อมโยง (conjunctive) มีความยืดหยุ่นสำหรับกฎการตัดสินใจ ดังนั้นผู้ตัดสินสามารถสร้างกฎการตัดสินใจที่ซับซ้อนเพื่อให้มีความหมายถึงผู้สอบที่มีคุณภาพที่ผ่านการประเมิน วิธีนี้มีความซับซ้อนในการใช้มากกว่ามาตรฐานการปฏิบัติที่มีเพียงมาตรฐานเดียว

หลังจากผู้ตัดสินได้รับการฝึกเกี่ยวกับงานหรือแบบฝึกหัดที่ใช้ประกอบการประเมินเช่น พิจารณาความหมายของคะแนนที่มีความเป็นไปได้หลายๆ ค่า ในแต่ละงานหรือแบบฝึกหัด จากนั้นผู้ตัดสินจะถูกถามอย่างอิสระเพื่อสร้างกฎในการตัดสินใจในการผ่านขึ้นมา กฎการตัดสินใจนี้จะถูกวิเคราะห์อย่างมีคุณภาพเพื่อใช้กำหนดงานว่ามีความเป็นตัวแทนคล้ายกับองค์ประกอบ จากนั้นข้อสรุปจากกฎการตัดสินใจจะถูกนำมาแสดงให้คณะผู้ตัดสินพิจารณาอภิปราย และปรับปรุง การใช้ในบางครั้งได้มีการเพิ่มรอบของกฎการตัดสินเป็นรายบุคคลขึ้นมาเพื่อมีเป้าหมายในการสร้างฉันทมติร่วมกันในผู้ตัดสินเกี่ยวกับกฎการตัดสินเพียงกฎเดียว บ่อยครั้งข้อมูลผู้สอบถูกนำมาให้ผู้ตัดสินพิจารณาประกอบผลกระทบของกฎการตัดสินหลายๆ แบบ กฎการตัดสินใจถูกลงคะแนนเสียงเพื่อรับรองจากผู้ตัดสิน และเกิดการปฏิบัติของผู้สอบขั้น

ต่ำที่มีคุณภาพที่จะผ่านนั้นคือ เป็นเส้นภาพ (profiles) ทั้งหมดที่มีค่าคะแนน (score value) แสดงถึงจุดที่เท่ากับหรือเหนือกว่าเส้นภาพขั้นต่ำซึ่งเป็นเส้นที่ถูกลงความเห็นว่าผ่าน และเส้นภาพที่มีค่าคะแนนไม่ถึง กฎการตัดสินใจก็ตกไป

จุดดีและข้อบกพร่อง

วิธีนี้ยอมให้ผู้ตัดสินมีความยืดหยุ่นสูงในการสร้างกฎการตัดสินใจ สำหรับข้อสรุปของกระบวนการ ถ้ากลุ่มมีฉันทามติร่วมกันผู้ตัดสินก็มีแนวโน้มที่จะเกิดความรู้สึกมั่นใจว่ากฎการตัดสินใจมีความเหมาะสมสำหรับใช้ในการตัดสินใจว่าผ่าน/ตก ในทางตรงข้ามวิธีนี้ไม่มีความจำเป็นว่ากลุ่มจะต้องมีฉันทมติทั้งหมด เสียงข้างน้อยก็ได้รับการพิจารณาด้วย ผู้ตัดสินบางคนอาจจะเข้มงวดต่อกฎการตัดสินใจขั้นสุดท้ายและบางครั้งไม่เต็มใจที่จะยอมรับความคิดเห็นของกลุ่ม สถานการณ์เกิดเมื่อค่ามีจุดที่ต่ำสุด (minimum point values) เช่น ไม่ยอมให้คะแนนที่ต่ำมากๆ จากแบบฝึกหัดเข้าสู่กฎการตัดสินใจ บางผู้ตัดสินอาจรู้สึกมั่นใจว่า การปฏิบัติที่ได้คะแนนต่ำจะไม่คงเส้นคงวากับการรับรอง ในขณะที่ผู้ตัดสินบางคนอาจรู้สึกว่ามีบางแบบฝึกหัดยังไม่ดีพอ แต่ก็ยังพอยอมรับได้ ถ้าในภาพรวมของแบบฝึกหัดทั้งหมดมีระดับสูงเพียงพอ จะเห็นว่า ข้อบกพร่องคือ กระบวนการกำหนดมาตรฐานบางครั้งจะทำให้ได้ผลฉันทมติของกลุ่มที่ไม่ดีพอ นอกจากนี้วิธีนี้มีศักยภาพสำหรับองค์ประกอบที่เชื่อมโยง (conjunctive component) ในส่วนของกฎการตัดสินใจเพราะว่า คุณภาพการวัดมีจำกัด (โดยเฉพาะความเที่ยงของคะแนน) สำหรับรูปแบบของแบบฝึกหัด

2.7 วิธีการตัดสินเชิงนโยบาย (Judgmental Policy Capturing Method)

วิธีนี้เกี่ยวข้องกับการตัดสินเส้นภาพ (profile) ของคะแนนนักเรียนที่ได้จากการทำแบบฝึกหัด วิธีการใช้การตัดสินเชิงนโยบายมีประโยชน์มากเมื่อแต่ละแบบฝึกหัดถูกให้คะแนนที่แสดงคะแนนอย่างน้อย 3 ค่า ซึ่งดีกว่าการให้คะแนนแบบผ่าน/ตก

วิธีดำเนินการคณะผู้ตัดสินต้องได้รับการฝึกอย่างเข้มข้นเพื่อให้พวกเขาเข้าใจความหมายของแต่ละคะแนนที่นักเรียนทำในแต่ละแบบฝึกหัด เมื่อแสดงเส้นภาพคะแนนนักเรียนจากทุกแบบฝึกหัดแล้ว ผู้ตัดสินก็จะสร้างรูปภาพทางความคิดของคุณภาพผู้สอบด้านการปฏิบัติ คะแนนเส้นภาพจะถูกจำแนกอย่างอิสระเพื่อกำหนดแต่ละคะแนนเส้นภาพลงในแต่ละประเภทที่ระบุข้อความมาตรฐานการปฏิบัตินั้นๆ (เช่น ต่ำมาก ค่อนข้างต่ำ ต่ำเล็กน้อย สูงเล็กน้อย ค่อนข้างสูง สูงมาก)

ผลการตัดสินจะถูกนำมาวิเคราะห์ แต่ละประเภทจะถูกนำมากำหนดเป็นค่าตัวเลข กระบวนการทางสถิติ (เช่นการวิเคราะห์การถดถอยพหุ) จะถูกนำมาใช้เพื่อกำหนดความสัมพันธ์ทางคณิตศาสตร์ระหว่างเส้นภาพคะแนนผู้สอบ กับประเภทการปฏิบัติ (performance categories) ความสัมพันธ์ดังกล่าวจะกำหนดน้ำหนักเพื่อนำไปใช้กับคะแนนของผู้สอบในแต่ละแบบฝึกหัด และสำหรับคำนวณคะแนนการปฏิบัติในผู้สอบแต่ละคน

คะแนนการปฏิบัติทั้งหมดถูกคำนวณสำหรับผู้สอบแต่ละคน เส้นภาพที่เป็นคะแนนแบบฝึกหัดถูกแสดงกับผู้ตัดสิน เส้นภาพเหล่านี้ถูกเรียงลำดับคะแนนปฏิบัติสูงสุดไปยังคะแนนปฏิบัติต่ำสุด ผู้ตัดสินจะถูกถามเพื่อกำหนดคะแนนปฏิบัติต่ำสุดที่สามารถประเมินผ่านได้ ข้อเสนอแนะจากผู้ตัดสินทั้งหมดก็ถูกนำมาจัดเป็นตาราง และการแจกแจงค่ามัธยฐานจะถูกนำมากำหนดเป็นมาตรฐานการปฏิบัติ

ขั้นตอนทั้งสองถูกนำมาทำซ้ำโดยให้โอกาสผู้ตัดสินอธิบายเหตุผลจากข้อเสนอแนะเดิมของพวกเขา มาตรฐานการปฏิบัติขั้นสุดท้ายจะยึดเสียงข้างมากของผู้ตัดสินและบางครั้งมาตรฐานการปฏิบัติอาจมีการเปลี่ยนแปลงเพียงเล็กน้อย

จุดดีและข้อบกพร่อง

จุดดีคือ ผลลัพธ์สำคัญที่ได้จากการใช้การตัดสินเชิงนโยบาย (judgmental policy capturing) คือ ชุดของน้ำหนักที่กำหนดความสำคัญของแบบฝึกหัด แม้ว่ามันจะมีความเป็นไปได้ที่ใช้วิธีนี้กับโมเดลที่ประมาณการกำหนดมาตรฐานการปฏิบัติสำหรับแต่ละแบบฝึกหัดในการประเมิน ซึ่งโมเดลการวิเคราะห์การถดถอยพหุจะถูกนำมาใช้บ่อยกับวิธีที่เป็นการทดแทน (compensatory) วิธีนี้ผู้สอบสามารถทดแทนงานการปฏิบัติที่ได้คะแนนต่ำในบางแบบฝึกหัดโดยแสดงคะแนนปฏิบัติที่ได้คะแนนสูงในแบบฝึกหัดอื่นแทน มาตรฐานการปฏิบัติทดแทน (compensatory performance standards) ได้แสดงถึงความเที่ยงของการตัดสินใจที่เพิ่มขึ้น

สำหรับข้อบกพร่องสำคัญคือ ใช้เวลาในการฝึกผู้ตัดสินเป็นระยะเวลานานและเมื่อใช้เวลาฝึกนาน ค่าใช้จ่ายก็จะสูงตามไปด้วย

โมเดลการใช้การตัดสินเชิงนโยบาย ผู้ตัดสินจะมีนโยบายการกำหนดมาตรฐาน ส่วนโมเดลกระบวนการผู้ตัดสินจะทำการแยกประเภทการปฏิบัติของผู้สอบมากกว่าที่จะถามผู้ตัดสิน เกี่ยวกับนโยบายของพวกเขา ซึ่งบางผู้ตัดสินจะรู้สึกหงุดหงิดกับวิธีการแบบนี้เพราะการกระทำของพวกเขาไม่ได้สะท้อนสิ่งที่พวกเขาอ้างอิงนโยบายออกมาแต่ก็มีผู้สนับสนุนวิธีนี้เห็นว่าการกระทำของผู้ตัดสินเป็นตัวบ่งชี้การอ้างอิงนโยบายของพวกเขา

จากที่กล่าวมาสรุปได้ว่า วิธีกำหนดมาตรฐานแบบดั้งเดิม มี 2 วิธีคือ 1) วิธีแบบสอบเป็นศูนย์กลาง (test-center methods) ซึ่งแบ่งเป็น 4 วิธีย่อยคือ วิธีแองกอฟ วิธีของอีเบล วิธีของนีเดลสกี 2) วิธีผู้สอบเป็นศูนย์กลาง (Examinee-centered methods) แบ่งเป็น 2 วิธีย่อยคือ วิธีกลุ่มคาบเส้น และ วิธีกลุ่มตรงข้าม วิธีการกำหนดมาตรฐานแนวใหม่ ซึ่งเป็นวิธีการที่เหมาะสำหรับการประเมินผลการปฏิบัติที่มีความหลากหลายซึ่งมีด้วยกัน 7 วิธีคือ วิธีแองกอฟแบบปรับขยาย (Extended Angoff Method) วิธีบุ๊คมาร์ค (Bookmark Method) วิธีกลุ่มตรงข้าม (Contrasting Groups Method) วิธีเลือกงานผู้สอบ (Examinee Paper Selections Method) วิธีองค์รวม (Holistic Method) วิธีเส้นภาพเด่น (Dominant Profile Method) และวิธีการนำนโยบายมาช่วยตัดสิน (Judgmental Policy Capturing Method)

ตอนที่ 3 ทฤษฎีการตอบสนองข้อสอบกับการกำหนดคะแนนจุดตัด

1. การกำหนดคะแนนจุดตัดด้วยวิธีแองกอฟโดยการประยุกต์ทฤษฎีการตอบสนองข้อสอบ

เคน (Kane 1987 อ้างถึงใน วิมลมาศ ยิ้มละมัย, 2541) ได้รวบรวมวิธีการแปลงคะแนนจุดตัดที่ได้จากวิธีการกำหนดคะแนนจุดตัดโดยใช้ดุลพินิจของผู้เชี่ยวชาญตัดสิน ให้เป็นระดับความสามารถขั้นต่ำสุดที่ยอมให้ผ่าน (θ^*) โดยมีข้อตกลงว่า จะต้องทราบค่าโค้งลักษณะข้อสอบ $(P_i(\theta))$ ซึ่งเป็นหลักของทฤษฎีการตอบสนองข้อสอบในการประมาณความน่าจะเป็นของผู้สอบซึ่งมีระดับความสามารถ θ ต่างๆ จะสามารถตอบข้อสอบได้ถูก ซึ่งมีอยู่ 3 วิธี คือ

<u>วิธีที่ 1</u> ประมาณค่า θ^* ของข้อสอบแต่ละข้อโดยอาศัยโค้งลักษณะข้อสอบรายข้อ แล้วจึงคำนวณหา θ^* เฉลี่ยของข้อสอบทุกข้อในแบบสอบ มี 3 ขั้นตอน ดังนี้

1. คำนวณหาค่าเฉลี่ยของความน่าจะเป็นในการตอบข้อสอบ i ได้ถูกที่ใช้ดุลพินิจของผู้เชี่ยวชาญตัดสิน (M_{iR})

2. ให้ M_{iR} เป็นค่าประมาณของความน่าจะเป็นที่ผู้สอบซึ่งมีระดับความสามารถขั้นต่ำสุดที่ยอมให้ผ่าน สามารถตอบข้อสอบได้ถูกในรูปแบบโลจิสติก ซึ่งแทนได้ด้วยโค้งลักษณะข้อสอบที่ระดับความสามารถ θ^*_{iR} ดังสมการ

$$M_{iR} = P_i(\theta^*_{iR})$$

3. ประมาณค่าความแปรปรวนของ θ^*_{iR} ในข้อ i

$$\text{กำหนดให้} \quad P_i(\theta^*_{iR}) = \frac{\Delta M_{iR}}{\Delta \theta_{iR}} \tag{6}$$

เป็นความชันของโค้งลักษณะข้อสอบ ข้อ i ที่ความสามารถ (θ^*_{iR}) แทนด้วยความเบี่ยงเบนมาตรฐาน

$$\frac{P_i'(\theta^*_{iR})}{\sigma_i(M_{iR})} \cong \frac{1}{\sigma_i(\theta^*_{iR})} \tag{7}$$

$$\frac{[P_i'(\theta^*_{iR})]^2}{\sigma_i^2(M_{iR})} \cong \frac{1}{\sigma_i^2(\theta^*_{iR})} \tag{8}$$

4. ประมาณค่า θ^* จากค่าเฉลี่ยของ θ^*_{iR}

$$\theta^* = \frac{1}{n}\sum_i \theta^*_{iR} \tag{9}$$

ซึ่งมีความแปรปรวน $\sigma^2(\theta^*) = \frac{1}{n^2}\sum_i \sigma_i^2(\theta^*_{iR})$ (10)

โดยมีข้อตกลงเบื้องต้นว่าการตัดสินความน่าจะเป็นในการตอบถูกของข้อสอบแต่ละข้อเป็นอิสระกัน แต่การที่ผู้ตัดสินคนหนึ่งตัดสินข้อสอบทุกข้อในแบบสอบอาจทำให้ผลการตัดสินข้อสอบรายข้อไม่เป็นอิสระกัน จึงแทนสมการที่ (9) ด้วยการประมาณแบบถ่วงน้ำหนัก ดังนี้

$$\theta^2_w = \sum_i w_i \theta^*_{iR} \tag{11}$$

เมื่อ $\sum_i w_i = 1$ (12)

และมีความแปรปรวนเป็น

$$\sigma^2(\theta^*_w) = \sum_i w_i^2 \sigma_i^2(\theta^*_{iR}) \tag{13}$$

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ค่าของ $\mathbf{w_i}$ ที่จะทำให้ความแปรปรวนในสมการที่ (13) มีค่าน้อยที่สุดและมีคุณสมบัติดังในสมการที่ (12) ทำได้โดยการใช้ Lagrange multipliers

$$\mathbf{w_i} = \frac{1}{\sigma_i^2(\theta_{iR}^*)} / \sum_i \frac{1}{\sigma_i^2(\theta_{iR}^*)} \quad (14)$$

การประมาณแบบถ่วงน้ำหนักนี้จะช่วยลดความคลาดเคลื่อนมาตรฐานของ θ^* ข้อสอบที่ผู้ตัดสิน ตัดสินคล้ายๆ กัน จะถูกถ่วงน้ำหนักมาก ส่วนข้อที่ผู้ตัดสิน ตัดสินต่างกันออกไปก็จะถูกถ่วงน้ำหนักน้อย

<u>วิธีที่ 2</u> เป็นวิธีที่ใช้กันมากที่สุด เป็นการประมาณ θ^* โดยอาศัยโค้งลักษณะแบบสอบมี 3 ขั้นตอน คือ

1. คำนวณหาค่าเฉลี่ยของความน่าจะเป็นในการตอบข้อ i ถูก ที่ตัดสินโดยผู้ตัดสินทั้งหมด (M_{iR})

2. คำนวณหาคะแนนจุดตัดของแบบสอบซึ่งอยู่บนสเกลของคะแนนจริงจากผลรวมของ M_{iR} ในแต่ละข้อของแบบสอบ

3. ใช้โค้งลักษณะแบบสอบ (test characticsistic curve) หาค่าระดับความสามารถ θ ซึ่งสัมพันธ์กับคะแนนจริงในขั้นที่ 2 จากสมการ

$$\sum M_{iR} = \sum p_i(\theta^*) \quad (15)$$

$$\sum_i [M_{iR} - P_i(\theta^*)] = 0 \quad (16)$$

จากสมการที่ (5) $$\sum_i [P_i(\theta_{iR}^*) - P_i(\theta^*)] = 0 \quad (17)$$

แต่ $\theta_{iR}^* \rightarrow \theta^*$ ความสัมพันธ์สามารถประมาณได้ด้วย linear function (Lord 1980 : 69)

$$\sum_i P_i(\theta_{iR}^*)(\theta_{iR}^* - \theta^*)] \cong 0 \quad (18)$$

$$\theta^* \cong \frac{1}{\sum p'_i(\theta_{iR}^*)} \sum p'_i(\theta_{iR}^*)\theta_{iR}^* \quad (19)$$

จากสมการที่ (7) และ (19)

$$\theta^* = \frac{1}{\sum \sigma_i(M_{iR}) / \sigma_i(\theta^*_{iR})} \sum \frac{\sigma(M_{iR})}{\sigma_i(\theta^*_{iR})} \theta^*_{iR} \quad (20)$$

การประมาณ θ^* จากสมการที่ 20 ยังไม่ใช่ค่าที่ดีที่สุด สามารถปรับปรุงได้โดยการถ่วงน้ำหนักด้วย $1 / \sigma_i(M_{iR})$ ในสมการที่ 17 จะได้

$$\theta^* = \frac{1}{\sum_i [P_i'(\theta^*_{iR}) / \sigma_i(M_{iR})]} \sum_i \frac{P_i'(\theta^*_{iR})}{\sigma_i(M_{iR})} \theta^*_{iR} \quad (21)$$

จากสมการที่ 7 จะได้ว่า

$$\theta^*_w = \frac{1}{\sum_i [1 / \sigma_i(\theta^*_{iR})]} \sum_i \frac{1}{\sigma_i(\theta^*_{iR})} \theta^*_{iR} \quad (22)$$

<u>วิธีที่ 3</u> เป็นการประมาณ θ^* ที่เหมาะสมที่สุดระหว่างการใช้ดุลพินิจในการตัดสินความน่าจะเป็นในการตอบข้อสอบได้ถูก((M_{iR})) กับโค้งลักษณะข้อสอบมี 3 ขั้นตอน คือ

1. คำนวณหาค่าเฉลี่ยของความน่าจะเป็นในการตอบข้อ i ถูก ที่ตัดสินโดยผู้ตัดสินทั้งหมด (M_{iR}) 2. คำนวณหาความแปรปรวนของค่าเฉลี่ยความน่าจะเป็นในการตอบข้อสอบ i ถูก ที่ตัดสินโดย ผู้ตัดสินทั้งหมด $\sigma_i^2(M_{iR})$

3. คำนวณหา θ^* ซึ่งมีผลบวกของความเบี่ยงเบนกำลังสองระหว่าง (M_{iR}) และ $P_i(\theta^*)$ มีค่าน้อยที่สุดจากสมการ

$$\sum_i \frac{(M_{iR} - p_i(\theta^*)^2}{\sigma^2(M_{iR})} = \sum_i \frac{[P_i(\theta^*_{iR}) - P_i(\theta^*)]^2}{\sigma_i^2(M_{iR})} \quad (23)$$

แต่ θ^*_{IR} มีค่าใกล้เคียงกับ θ^* และสมการที่ 18 จะได้

$$\sum_i \frac{(M_{iR} - P_i(\theta^*))^2}{\sigma^2(M_{iR})} \cong \sum_i \frac{1}{\sigma_i^2(M_{iR})} [P_i'(\theta^*_{iR})(\theta^*_{iR} - \theta^*)]^2$$

$$\cong \quad \sum_i \frac{P_i^{'2}(\theta_{iR}^*)}{\sigma_i^2(M_{iR})}(\theta_{iR}^* - \theta^*)^2 \qquad (24)$$

กำหนดให้สมการที่ 24 เท่ากับ 0 จะได้

$$\sum \frac{P_i^{'2}(\theta_{iR}^*)}{\sigma_i^2(M_{iR})}(\theta_{iR}^* - \theta^*) \cong \quad 0 \qquad (25)$$

เมื่อแก้สมการแล้วหาค่า θ^* จะได้

$$\theta^* \quad \cong \quad \frac{1}{\sum_i [P_i^{'2}(\theta_{iR}^*) / \sigma_i^2(M_{iR})]} \sum_i \frac{P_i^{'2}(\theta_{iR}^*)}{\sigma_i^2(M_{iR})}\theta_{iR}^* \qquad (26)$$

จากสมการที่ 8 จะได้ว่า

$$\theta^* \quad \cong \quad \frac{1}{\sum_i [1 / \sigma_i^2(\theta_{iR}^*)]} \sum_i \frac{1}{\sigma_i^2(\theta_{iR}^*)}\theta_{iR}^* \qquad (27)$$

สรุปวิธีการประมาณ θ^* ของทั้ง 3 วิธี จะเห็นว่า วิธีที่ 1 และ 3 จะอาศัยผลการสอบของข้อสอบ แต่ละข้อประกอบกับการใช้ดุลพินิจของผู้เชี่ยวชาญ ส่วนวิธีที่ 2 จะหาคะแนนจุดตัดจากดุลพินิจของผู้เชี่ยวชาญก่อนแล้วจึงอาศัยโค้งลักษณะแบบสอบเพื่อแปลงคะแนนจุดตัดในสเกลของคะแนนโดเมนให้อยู่บนสเกลของความสามารถซึ่งเป็นวิธีที่นิยมกัน ดังนั้นในการวิจัยครั้งนี้ ผู้วิจัยจึงกำหนดจุดตัดด้วยวิธีที่ 2 เพราะต้องการคะแนนจุดตัดจากการตัดสินของผู้เชี่ยวชาญแล้วจึงแปลงให้เป็นระดับความสามารถ θ^*

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

2. การกำหนดคะแนนจุดตัดด้วยวิธีบุ๊คมาร์คโดยการประยุกต์ทฤษฎีการตอบสนองข้อสอบ

ในการจัดเตรียมคู่มือการจัดเรียงข้อสอบนั้น ต้องหาค่าความสามารถ (θ) โดยการนำค่าพารามิเตอร์ความยาก อำนาจจำแนก และการเดาจากการวิเคราะห์ด้วยทฤษฎี IRT มาแทนค่าในสูตรเพื่อวิเคราะห์หาค่า θ

สำหรับสูตรหาการค่า θ ในงานวิจัยนี้มาจาก Mitzel et al.(2001 , cited in Cizek, Bunch & Koons, 2004) ซึ่งได้เสนอสูตรดั้งเดิมแบบโมเดล 3 พารามิเตอร์ สำหรับข้อสอบแบบเลือกตอบ ดังนี้

$$P_j(\theta) = \frac{c_j + (1 - c_j)}{1 + \exp[-1.7a_j(\theta - b_j)]} \quad (1)$$

เมื่อ

P_j = ความน่าจะเป็นของการตอบถูก

θ = ฟังก์ชันความสามารถของผู้สอบ

b_j = ความยากของข้อสอบ

a_j = อำนาจจำแนกข้อสอบ

c_j = threshold หรือ ตัวแปรการเดา (chance variable)

exp = natural logarithm e (2.71828…)

อย่างไรก็ตาม Mitzel et al. (2001, cited in Cizek, Bunch & Koons, 2004) ได้กำหนดให้ $c_j = 0$ จากสมการ (1) จะได้

$$P_j(\theta) = \frac{1}{1 + \exp[-1.7a_j(\theta - b_j)]} \quad (2)$$

สำหรับงานวิจัยนี้ แบบสอบที่ใช้เป็นการให้คะแนนแบบ 2 ค่า (เช่น แบบเลือกตอบ) และเนื่องจากว่างานวิจัยนี้ได้ตั้งคำถามว่า นักเรียนที่คาบเส้นมีโอกาสตอบข้อสอบถูกที่ความน่าจะเป็น .67 ซึ่งความน่าจะเป็น .67 นี้นำมาแก้สมการในสมการ (2) สำหรับค่า θ ที่ต้องการให้ตอบข้อสอบถูก ดังนี้

$$0.67 = \frac{1}{1+e^{-1.7a_j(\theta-b_j)}}$$

$$1+e^{-1.7a_j(\theta-b_j)} = \frac{1}{0.67}$$

$$e^{-1.7a_j(\theta-b_j)} = 1.4925 - 1$$

take ln ทั้งสองข้าง $-1.7a_j(\theta-b_j) = \ln 0.4925$

$$\theta - b_j = \frac{\ln 0.4925}{-1.7a_j}$$

$$\theta = b_j + \frac{\ln 0.4925}{-1.7a_j}$$

$$\theta = b_j + \frac{.708}{1.7a_j} \quad (3)$$

ตอนที่ 4 การตรวจสอบความเที่ยง และความตรงของการกำหนดมาตรฐาน

การตรวจสอบความเที่ยง และความตรงของการกำหนดมาตรฐานในงานวิจัยนี้ ทำโดยใช้ทฤษฎีการสรุปอ้างอิง (Generalizability Theory) ดังนั้นจะขอเสนอมโนทัศน์ทฤษฎีการสรุปอ้างอิงดังนี้

1. มโนทัศน์ทฤษฎีการสรุปอ้างอิง

ครอนบาค และคณะ (Gronbach, Gleser, Nanda and Rajaratnam, 1972) ได้เสนอหลักการเกี่ยวกับทฤษฎีการสรุปอ้างอิง (Generalizability Theory) หรือ G–Theory ขึ้นเป็นครั้งแรกและพัฒนาเพื่อประมาณค่าความเที่ยงของแบบสอบในสถานการณ์ของการสอบต่างๆ และสามารถอธิบายแหล่งความคลาดเคลื่อนต่างๆ ได้มากขึ้น ตามแนวคิดของ G–theory คะแนนที่ได้จากการวัดภายใต้สถานการณ์ หรือเงื่อนไขเดียวด้วยการสอบชุดเดียว และทำการสอบครั้งเดียว คะแนนที่ได้ไม่สามารถเชื่อถือได้อย่างเต็มที่ เพราะคะแนนที่ได้จะไม่เป็นตัวแทนที่ดีของคะแนนจริง ซึ่งเป็นคะแนนเฉลี่ยที่ได้จากการทดสอบในหลายๆ สถานการณ์ด้วยแบบทดสอบ

หลายๆ ชุด และทำการสอบหลายๆ ครั้ง แต่ G–Theory ได้พิจารณาถึงแหล่งความคลาดเคลื่อนพหุ (multiple sources of error) สามารถวิเคราะห์ความคลาดเคลื่อนจากหลายแหล่งพร้อมกันไปได้ ผู้บริหารการสอบจึงสามารถตัดสินใจได้ว่าควรใช้ในสถานการณ์หรือเงื่อนไขการวัดแบบใดจึงจะได้คะแนนที่เชื่อถือถึงระดับที่ต้องการ ผลการวิเคราะห์ด้วย G–Theory จะให้ค่าที่เรียกว่าสัมประสิทธิ์การสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (Generalizability Coefficient) ซึ่งคล้ายกับสัมประสิทธิ์ความเที่ยง (Reliability Coefficient) ในทฤษฎีการทดสอบแบบดั้งเดิม (ศิริชัย กาญจนวาสี, 2545) ทฤษฎีการวัดแบบดั้งเดิมนั้นอธิบายความคลาดเคลื่อนเป็นค่าคงที่เพียงค่าเดียวไม่สามารถแบ่งแยกได้ ส่วนทฤษฎีการสรุปอ้างอิงนั้นไม่ต้องมีข้อตกลงเบื้องต้นเกี่ยวกับแบบสอบคู่ขนาน (Nunnally, 1978) และสามารถประมาณค่าความคลาดเคลื่อนได้ครอบคลุมเงื่อนไขการวัดโดยใช้การวิเคราะห์ความแปรปรวน (Analysis of Variance)

คำศัพท์ที่เกี่ยวข้องกับทฤษฎีการสรุปอ้างอิง

1. ประชากร (Population) หมายถึง สิ่งที่มุ่งวัดทั้งหมดในสถานการณ์ของการสอบ ได้แก่ บุคคลหรือผู้สอบ

2. องค์ประกอบ หรือ ฟาเซทของการวัด (Facet) หมายถึง ชุดหรือกลุ่มของเงื่อนไขของการวัดที่มีลักษณะเดียวกันเป็นองค์ประกอบที่คาดว่ามีผลต่อความคลาดเคลื่อนของการวัด เช่น ความยาวของข้อสอบ วิธีการวัดจำนวนครั้งของการสอบ จำนวนผู้ตรวจให้คะแนน เป็นต้น

3. เงื่อนไขของการวัด (Condition of measurement) เป็นระดับของฟาเซทของการวัด เช่น ฟาเซทของจำนวนผู้ตรวจอาจกำหนดจำนวนระดับเป็น 1 , 2 และ 3 คน ฟาเซทของความยาวแบบสอบอาจกำหนดความยาวของแบบสอบเป็น 10 , 20 และ 30 ข้อ เป็นต้น อาจสรุปได้ว่า เป็นขนาดของกลุ่มตัวอย่างขององค์ประกอบนั้น

4. เอกภพ (Universe) หมายถึง เงื่อนไขของการวัดที่สนใจทั้งหมดของแต่ละฟาเซท (Facet) เช่น จำนวนข้อสอบทั้งหมด จำนวนผู้ตรวจทั้งหมด นักเรียนทุกคน

เอกภพของค่าที่ได้จากการสังเกตทั้งหมด (admissible observation) เป็นกลุ่มเงื่อนไขของการวัดที่เป็นไปได้ ซึ่งสามารถวัดหรือสังเกตได้จากการเก็บข้อมูลการทดสอบของประชากรภายใต้เงื่อนไขของการวัดทั้งหมดในแต่ละฟาเซท เช่น ฟาเซทของข้อสอบ ฟาเซทของผู้ตรวจ

เอกภพของการสรุปอ้างอิง (Universe of generalization) เป็นเงื่อนไขการวัดทั้งหมดที่เป็นเป้าหมายของการสรุปอ้างอิง และเป็นการวัดที่ครอบคลุมเงื่อนไขที่สนใจทั้งหมด

5. คะแนนเอกภพ (universe scores) หมายถึง คะแนนของเงื่อนไขการวัดทั้งหมดของแต่ละฟาเซทซึ่งคล้ายกับคะแนนจริงของสิ่งที่วัดในทฤษฎีการวัดแบบดั้งเดิม

6. รูปแบบความสัมพันธ์ระหว่างองค์ประกอบ หมายถึง องค์ประกอบที่เป็นสิ่งที่ชี้ถึงแหล่งความแปรปรวนของรูปแบบการวัด ในแบบจำลองการวิเคราะห์ความแปรปรวนซึ่งรูปแบบความสัมพันธ์แบ่งออกเป็น 3 ลักษณะ คือ

6.1 ความสัมพันธ์แบบไขว้ (crossed) หมายถึง ความสัมพันธ์แต่ละลักษณะที่แต่ละระดับของสิ่งที่ถูกวัด ถูกวัดภายใต้เงื่อนไขเดียวกันทั้งหมด สัญลักษณ์คือ " X" อ่านว่า "crossed with" เช่น กำหนดรูปแบบความสัมพันธ์ระหว่างองค์ประกอบเป็น p x i x r หมายถึง นักเรียน (p) ทำข้อสอบ (i) ทุกข้อ และผู้ตรวจ (r) ตรวจข้อสอบของนักเรียนทุกคน ทุกๆ ข้อ

6.2 ความสัมพันธ์แบบแฝง (nested) หมายถึง ความสัมพันธ์ในลักษณะที่แต่ละระดับของสิ่งที่ถูกวัด ถูกวัดภายใต้เงื่อนไขที่แตกต่างกัน สัญลักษณ์คือ " : " อ่านว่า "nested within" เช่น รูปแบบความสัมพันธ์ระหว่างองค์ประกอบเป็น i : r หมายถึง ผู้ตรวจข้อสอบ (r) ตรวจข้อสอบ (i) ต่างข้อกัน

6.3 ความสัมพันธ์แบบผสม (confounded) หมายถึง ความสัมพันธ์ที่มีทั้งความสัมพันธ์แบบไขว้และความสัมพันธ์แบบแฝงปนกันอยู่ เช่น p x (i : t) หมายถึง นักเรียน (p) ทำข้อสอบ (i) บางข้อที่อยู่ในแบบสอบ (t)

ความแปรปรวนของคะแนนตามทฤษฎีการสรุปอ้างอิง

ตามทฤษฎีการวัดแบบดั้งเดิม (Classical Test Theory) คะแนนจริงของผู้สอบ (true score : T_{pi}) คือ ค่าเฉลี่ยของคะแนนจากการสอบซ้ำๆ กันด้วยแบบสอบคู่ขนาน จึงเป็นความแปรปรวนของค่าเฉลี่ยของการสอบซ้ำ และความแปรปรวนของคะแนนที่สังเกตได้ เป็นผลรวมของความแปรปรวนของคะแนนจริงกับความแปรปรวนของความคลาดเคลื่อน

$$X = T_{pi} + E_{pi}$$

$$\sigma^2_X = \sigma^2_T + \sigma^2_E$$

สำหรับทฤษฎีการสรุปอ้างอิง คะแนนเอกภพ (Universe score ; Mp) คือค่าเฉลี่ยของคะแนนการวัดซ้ำหลายๆ ครั้งตามเงื่อนไขการวัดในเอกภพของการสรุปอ้างอิง ค่าความคลาดเคลื่อนของการวัด ถูกจำแนกออกเป็นความคลาดเคลื่อนจากฟาเซทหรือกลุ่มเงื่อนไขของการวัดและความคลาดเคลื่อนจากแหล่งที่เหลืออื่นๆ และการวัดแต่ละครั้งไม่จำเป็นต้องใช้แบบสอบคู่ขนาน เหมือนทฤษฎีการทดสอบแบบดั้งเดิม ส่วนความแปรปรวนของคะแนนที่สังเกตได้เป็น

ผลรวมของการแปรปรวนของคะแนนเอกภพกับความแปรปรวนของคะแนนความคลาดเคลื่อนจากฟาเซทหรือองค์ประกอบต่างๆของการวัด และความแปรปรวนของความคลาดเคลื่อนแหล่งอื่นๆดังนี้

$$X_{pi} = T_{pi} + E_{pi} + e_{pi}$$

$$\sigma^2 x_p = \sigma^2 \mu_p + \sigma^2 E_i + \sigma^2 e_p$$

ความแปรปรวนของคะแนนความคลาดเคลื่อนจากองค์ประกอบต่างๆของการวัดแบ่งออกเป็น 2 ประเภท ได้แก่ (ศิริชัย กาญจนวาสี, 2545)

1. ความแปรปรวนของคะแนนความคลาดเคลื่อนสัมบูรณ์ (Absolute Error Variance ; σ^2_{ABS}) คือ ความแปรปรวนของ $\mu_p - X_p$ ซึ่งคำนวณได้จากผลรวมของความแปรปรวนของคะแนนจากแหล่งต่างๆ ยกเว้น $\sigma^2_{\mu_p}$ หรือ σ^2_p

2. ความแปรปรวนของคะแนนความคลาดเคลื่อนสัมพันธ์ (Relative Error Variance ; σ^2_{REL}) คือ ความแปรปรวนของ $\mu_p - X_p$ ที่คำนวณได้จากผลรวมของความแปรปรวนของคะแนนจากแหล่งต่างๆ ที่มีปฏิสัมพันธ์ร่วมกับผู้สอบ (p)

สัมประสิทธิ์การสรุปอ้างอิง

สัมประสิทธิ์การสรุปอ้างอิง (Generalizability coefficient : ρ^2_i) เป็นอัตราส่วนระหว่างความ แปรปรวนของคะแนนเอกภพ และความแปรปรวนของคะแนนที่สังเกตได้ ดังสมการดังต่อไปนี้

$$\text{G - coefficient} = \frac{\sigma_P^2}{\sigma_P^2 + \text{Error Variance}}$$

สัมประสิทธิ์การสรุปอ้างอิง จึงเป็นดัชนีที่ใช้อธิบายความแม่นยำของการวัดเช่นเดียวกับค่าสัมประสิทธิ์ความเที่ยงแบบดั้งเดิม สามารถใช้คำนวณช่วงของความเชื่อมั่นของคะแนนเอกภพหรือใช้ในสมการถดถอยในการประมาณค่าคะแนนเอกภพและใช้ในการปรับแก้ค่าสหสัมพันธ์ที่ลดลงอันเนื่องมาจากความคลาดเคลื่อน เนื่องจากความแปรปรวนของคะแนนความคลาดเคลื่อนจากองค์ประกอบต่างๆ ของการวัดแบ่งเป็น 2 ประเภท จึงทำให้ค่าสัมประสิทธิ์การสรุปอ้างอิงประกอบด้วย 2 ประเภท ดังนี้

1. สัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์ (Generalizability coefficient for absolute decision) สัมประสิทธิ์นี้แสดงถึงความเที่ยงของแบบสอบในสถานการณ์ของการตัดสินใจที่ขึ้นกับคะแนนผู้สอบตามลำพัง ไม่มีการเปรียบเทียบภายในกลุ่มหรือระหว่างกลุ่ม เช่น แบบสอบอิงเกณฑ์ซึ่งใช้ตรวจความสามารถของผู้สอบกับเกณฑ์ที่กำหนดไว้ล่วงหน้า

2. สัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมพันธ์ (Generalizability coefficient for relative decision) สัมประสิทธิ์นี้แสดงถึงความเที่ยงของแบบสอบในสถานการณ์การตัดสินใจที่มีการเปรียบเทียบคะแนนระหว่างผู้สอบ เช่นแบบสอบอิงกลุ่มซึ่งใช้เปรียบเทียบคะแนนระหว่างผู้สอบด้วยการแจกแจงคะแนนเป็นโค้งปกติและตัดเกรด หรือให้ระดับผลการเรียนหรือการสอบแข่งขันเพื่อหาผู้ที่ได้คะแนนสูงเข้าศึกษา

ขั้นตอนของทฤษฎีการสรุปอ้างอิง

เบรนเนน (Brennan, 1983 อ้างถึงใน วิภารัตน์ ศรีบุตรตา, 2541) ได้เสนอว่าทฤษฎีการสรุปอ้างอิง ประกอบด้วยขั้นตอนที่สำคัญ 2 ขั้นตอน คือ การศึกษาเพื่อการสรุปอ้างอิง หรือ การศึกษา G (Generalizability study or G–Study) และการศึกษาเพื่อการตัดสินใจ หรือการศึกษา D (Decision study or D–Study)

การศึกษาเพื่อการสรุปอ้างอิง (G–Study) เป็นการสรุปอ้างอิงผลที่ได้จากการศึกษาตัวอย่างการวัดตามเงื่อนไขที่น่าสนใจ บรรยายความแปรปรวนของความคลาดเคลื่อนจากแหล่งต่างๆ เพื่อสรุปอ้างอิงไปยังเอกภพของการวัด จุดประสงค์ของการศึกษาเพื่อการสรุปอ้างอิงคือ ต้องการประมาณค่าความแปรปรวนของคะแนนจริง และความแปรปรวนของคะแนนความคลาดเคลื่อนจากแหล่งความคลาดเคลื่อนต่างๆที่สนใจ และใช้เป็นข้อมูลสำหรับวางแผนเพื่อความสนใจ ดังนั้นการออกแบบการศึกษาเพื่อการสรุปอ้างอิงจึงควรครอบคลุมเงื่อนไขของการวัดที่ต้องการตัดสินใจนำแบบสอบไปใช้ในการศึกษาเพื่อการตัดสินใจ

การศึกษาเพื่อการตัดสินใจ (D–Study) เป็นการใช้ข้อมูลจากการศึกษาเพื่อการสรุปอ้างอิงที่สอดคล้องกับจุดประสงค์เฉพาะของการวัด ตัดสินใจเลือกใช้แบบ หรือวิธีวัดในสถานการณ์ต่างๆ ของการวัด

1. การศึกษาเพื่อสรุปอ้างอิง (G–Study)

การศึกษาเพื่อการสรุปอ้างอิงเป็นกระบวนการที่มุ่งประมาณความแปรปรวนจากแหล่งต่างๆ ภายใต้เอกภพของคะแนนที่สังเกตได้ โดยมีขั้นตอนดังนี้

1. กำหนดประชากร หรือ สิ่งที่มุ่งวัดทั้งหมด เช่น นักเรียนหรือผู้สอบ
2. กำหนดเอกภพของค่าที่ได้จากการสังเกตทั้งหมด เช่น ฟาเซทความยาวข้อสอบ ฟาเซทจำนวนผู้ตัดสิน
3. กำหนดฟาเซทของการวัดหรือองค์ประกอบที่มีผลต่อความคลาดเคลื่อนของการวัด เช่น กำหนดให้มี 2 ฟาเซท คือ ความยาวข้อสอบ และจำนวนผู้ตัดสิน

4. ออกแบบความสัมพันธ์ระหว่างฟาเซท ว่าจะเป็นแบบใด เช่น แบบ Crossed (แทนด้วย x) หรือแบบ Nested (แทนด้วย :) ถ้าต้องการศึกษาแบบให้ผู้ตรวจทุกคนตรวจข้อสอบทุกข้อ จะมีรูปแบบ Crossed (i x r) แต่ถ้าให้ตรวจแต่ละคนตรวจชุดข้อสอบต่างกันก็จะเป็นแบบ Nested (i : r) ทั้งนี้แล้วแต่ความสนใจของผู้ศึกษา

5. การเก็บรวบรวมข้อมูล โดยการสุ่มตัวอย่างมาศึกษา เช่น สุ่มตัวอย่างนักเรียน n_p คน และตัวอย่างผู้ตัดสิน n_r คน

6. การวิเคราะห์ข้อมูล นำข้อมูลที่เก็บรวบรวมได้ไปวิเคราะห์ทางสถิติ โดยใช้วิธีการวิเคราะห์ความแปรปรวน (Analysis of Variance : ANOVA) เพื่อประมาณค่าองค์ประกอบความแปรปรวนจากแหล่งต่างๆ ตัวอย่างเช่น การศึกษา G ในรูปแบบ Cross design (p x i x r) จะมีแหล่งความแปรปรวนหลัก (main effect) 3 แหล่ง คือ ผู้สอบ (p) ข้อสอบ (r) และผู้ตัดสิน (r) นอกจากนี้ยังมีผลร่วมหรือปฏิสัมพันธ์สองระดับมี 4 ค่า คือ ปฏิสัมพันธ์ระหว่างผู้สอบกับข้อสอบ (pi) ปฏิสัมพันธ์ระหว่างผู้สอบและผู้ตัดสิน (pr) ปฏิสัมพันธ์ระหว่างข้อสอบและผู้ตัดสิน (ir) และปฏิสัมพันธ์ทั้งหมดของผู้สอบ ข้อสอบ และผู้ตัดสิน

ค่าความแปรปรวนของแหล่งต่างๆ เหล่านี้ คำนวณมาจากค่าเฉลี่ยกำลังสอง (Mean Square : MS) ปกติในการประมาณค่าเฉลี่ยกำลังสอง (MS) ก็เพื่อทดสอบความมีนัยสำคัญของสัดส่วนความแปรปรวน โดยการใช้สถิติทดสอบ F (F-Test) และใช้ F-ratio ในการตัดสินว่าจะคงไว้หรือจะปฏิเสธสมมติฐานนั้น แต่ในการประยุกต์การวิเคราะห์ความแปรปรวนกับทฤษฎีการสรุปอ้างอิงนั้น F-ratio เป็นสิ่งที่ไม่มีความจำเป็น เพราะต้องการเพียงค่าเฉลี่ยกำลังสองจากการวิเคราะห์ความแปรปรวนเท่านั้น

องค์ประกอบของความแปรปรวนสามารถประมาณค่าได้จากค่าเฉลี่ยกำลังสอง ดังนี้

$$\sigma^2_{pir} = MS_{pir}$$
$$\sigma^2_{pi} = (MS_{pi} - MS_{pir}) / n_r$$
$$\sigma^2_{pr} = (MS_{pr} - MS_{pir}) / n_i$$
$$\sigma^2_{ir} = (MS_{ir} - MS_{pir}) / n_p$$

$$\sigma_r^2 = (MS_r - MS_{pr} - MS_{ir} + MS_{pir}) / n_p n_i$$

$$\sigma_i^2 = (MS_i - MS_{ir} - MS_{pi} + MS_{pir}) / n_p n_r$$

$$\sigma_p^2 = (MS_p - MS_{pi} - MS_{pr} + MS_{pir}) / n_i n_r$$

เมื่อ

σ_{pir}^2 = ค่าประมาณองค์ประกอบความแปรปรวนของความคลาดเคลื่อน

σ_{pi}^2 = ค่าประมาณองค์ประกอบความแปรปรวนของปฏิสัมพันธ์ระหว่างผู้สอบกับข้อสอบ

σ_{pr}^2 = ค่าประมาณองค์ประกอบความแปรปรวนของปฏิสัมพันธ์ระหว่างผู้สอบกับผู้ตัดสิน

σ_{ir}^2 = ค่าประมาณองค์ประกอบความแปรปรวนของปฏิสัมพันธ์ระหว่างข้อสอบกับผู้ตัดสิน

σ_r^2 = ค่าประมาณองค์ประกอบความแปรปรวนของผู้ตัดสิน

σ_i^2 = ค่าประมาณองค์ประกอบความแปรปรวนของข้อสอบ

σ_p^2 = ค่าประมาณองค์ประกอบความแปรปรวนของผู้สอบ

2. การศึกษาเพื่อการตัดสินใจ (D-Study)

การศึกษาเพื่อการตัดสินใจ (D-Study) เป็นขั้นตอนการนำผลที่ได้จากการประมาณค่าองค์ประกอบความแปรปรวนในการศึกษาเพื่อสรุปอ้างอิง (G-Study) มาตัดสินใจเลือกวิธีการวัดที่เฉพาะเหมาะสมเพื่อให้ได้ค่าสัมประสิทธิ์การสรุปอ้างอิง (G – Coefficient) ที่ดี ในการศึกษาการตัดสินใจ (D-Study) มีขั้นตอนดังนี้

1. กำหนดเอกภพของการสรุปอ้างอิง (Universe of generalizability) ที่ผู้ศึกษาต้องการสรุปอ้างอิง ซึ่งเอกภพของการสรุปอ้างอิงอาจประกอบด้วยเงื่อนไขทั้งหมด หรือเป็นเพียงเงื่อนไขย่อยในเอกภพของการสังเกตที่ยอมรับได้ ในขั้นนี้ผู้วิจัยต้องเลือกว่าจะใช้โมเดลใด จะเป็นโมเดลเชิงสุ่ม (Random Model) โมเดลกำหนด (fixed Model) หรือโมเดลผสม (Mix Model) ซึ่งโดยหลักการแล้ว ควรใช้โมเดลแบบสุ่มดีที่สุดเพราะจะทำให้การสรุปอ้างอิงได้กว้างขวางกว่าโมเดลแบบผสมและโมเดลแบบกำหนดตามลำดับ

2. กำหนดขนาดตัวอย่าง (Sample Size) จำนวนเงื่อนไขของตัวอย่างฟาเซทในการศึกษาเพื่อสรุปอ้างอิง(G-Study) ไม่จำเป็นต้องเป็นตัวอย่างในการศึกษาเพื่อตัดสินใจ (D-Study) ทั้งหมด ทั้งนี้การกำหนดขนาดตัวอย่างขึ้นอยู่กับความสนใจของผู้ศึกษา

3. โครงสร้างของแบบที่ศึกษา (Design Structure) ใน D-Study อาจใช้โครงสร้างของแบบที่ศึกษาเหมือนหรือต่างจากการศึกษาเพื่ออ้างอิง (G-Study) ก็ได้ เช่น ถ้าในการศึกษาเพื่อตัดสินใจ (D-Study) ตัดสินใจให้นักเรียนทุกคนทำข้อสอบเหมือนกัน จะมีโครงสร้างลักษณะ Cross design

4. การประมาณค่าความแปรปรวน (Estimated D study Variance Conponent)มีขั้นดังนี้

4.1 ระบุรูปแบบและขนาดกลุ่มตัวอย่าง

4.2 คำนวณหา universe variance และ error variance แล้วคำนวณค่าสัมประสิทธิ์การสรุปอ้างอิง หรือค่าสัมประสิทธิ์ความเที่ยง

2 ความตรงของการกำหนดมาตรฐาน

เมื่อกล่าวถึงเรื่องการหาความตรง สิ่งที่มักจะคู่กันก็คือเกณฑ์ Jaeger, Kane, Norcini และMehrens (Kane ,1998) ได้แบ่งเกณฑ์ที่ใช้สำหรับหาความตรงในการกำหนดมาตรฐานคือ เกณฑ์กระบวนการ (procedural criteria) เกณฑ์ความคงที่ภายใน (internal-consistency criteria) และเกณฑ์ภายนอก (external criteria) รายละเอียดมีดังนี้

2.1 **เกณฑ์กระบวนการ** เป็นเรื่องเกี่ยวกับการเลือกและการฝึกผู้ตัดสิน การกำหนดนิยาม มาตรฐานการปฏิบัติ และการออกแบบการเก็บรวบรวมข้อมูลเพราะเราจะเน้นว่าวิธีที่ใช้ดีอย่างไร อย่างไรก็ตามเกณฑ์กระบวนการก็มีประโยชน์เพราะเป็นแนวทางการใช้วิธีการกำหนดมาตรฐาน

2.2 **เกณฑ์ความคงที่ภายใน** ใช้ข้อมูลภายในที่ได้จากการศึกษาการกำหนดมาตรฐาน โดยการตรวจสอบความคงที่ของกระบวนการ ซึ่งสามารถตรวจสอบจากความสอดคล้องกันระหว่างผลลัพธ์ที่แตกต่างกันที่ได้จากการศึกษา

ในการตรวจสอบความตรงตามเกณฑ์ภายในเกี่ยวข้องกับความแม่นยำของการประมาณคะแนนการผ่าน (Kane, 1994) โดยจะประมาณค่า standard error ของคะแนนการผ่าน (passing score) โดยใช้ทฤษฎีการสรุปอ้างอิง (generalizability theory) ซึ่งสามารถกำหนดเอกภพของการสรุปอ้างอิง (universe of generalization) กลุ่มผู้ตัดสินต่างกลุ่มกันจะถูกนำมาศึกษา และข้อมูลจะถูกเก็บจากจำนวนครั้งของการสอบ (occasions) ที่แตกต่างกัน ดังนั้นเราก็จะสรุปอ้างอิงไปยังฟาเซทผู้ตัดสิน และฟาเซทจำนวนครั้งของการสอบ สำหรับผู้ตัดสินควรเลือกผู้ที่มีภูมิหลังคล้ายกัน (เช่น ครู สมาชิกองค์กรวิชาชีพ) ส่วนฟาเซทข้อสอบนั้นจะมีความ

ยุ่งยากซับซ้อนในบางจุด นั้นคือแบบสอบผลสัมฤทธิ์ทางการเรียนส่วนใหญ่กำหนดว่าข้อสอบที่ถูกสุ่มจากโดเมนของข้อสอบ และบ่อยครั้งที่คะแนนการผ่านถูกสรุปอ้างอิงข้ามไปยังฟอร์มข้อสอบที่แตกต่างกันโดยใช้วิธีวิทยาการปรับเทียบ (equating methodology) ถึงแม้ว่าคะแนนการผ่านจะเป็นชุดข้อสอบบางชุด

ในการประมาณ standard error สามารถทำได้ 2 วิธี วิธีแรกสามารถประมาณ standard error ได้โดยตรง โดยนำกลุ่มผู้ตัดสินที่ต่างกลุ่มกัน จากการสอบ 2 ครั้งหรือมากกว่า 2 ครั้ง หรือเป็นกลุ่มผู้ตัดสิน 2 กลุ่ม หรือมากกว่า 2 กลุ่มจากการสอบครั้งเดียวแล้วนำผลมาเปรียบเทียบกัน ข้อบกพร่องวิธีนี้คือค่าใช้จ่ายสูง และความเป็นอิสระในการศึกษา ส่วนข้อดีคือสามารถเห็นความแตกต่างที่ชัดเจนจากการศึกษาเรื่องหนึ่งไปยังการศึกษาอีกเรื่องหนึ่งที่ใช้รูปแบบเหมือนกัน นอกจากนี้มันยังสามารถประยุกต์ใช้ในวิธีการกำหนดมาตรฐาน เช่น วิธีแองกอฟ วิธีนีเดลสกี วิธีเจลเกอร์ วิธี อีเบล วิธีกลุ่มตรงข้าม วิธีกลุ่มคาบเส้น และอื่นๆ ถ้าวิธีการกำหนดมาตรฐานใช้วิธีการแองกอฟ นีเดลสกี เจลเกอร์ ก็ยังสามารถวิเคราะห์ข้อมูลโดยใช้ทฤษฎีการสรุปอ้างอิง (generalizability theory) เพื่อประมาณองค์ประกอบของความแปรปรวน (variance components) สำหรับผู้ตัดสิน ข้อสอบ และวิธีการ องค์ประกอบของความแปรปรวนสำหรับการศึกษา (studies) จะถูกนำมาเป็นตัวบ่งชี้ความแปรปรวนเกี่ยวกับจำนวนครั้งที่สอบ และความแปรปรวนเกี่ยวกับความแตกต่างในการศึกษาที่ต่างกัน

อีกวิธีหนึ่งถ้ามีข้อมูลจากการศึกษาเพียงเรื่องเดียวเกี่ยวกับวิธีการแองกอฟ นีเดลสกี หรือ เจลเกอร์ เราสามารถใช้ทฤษฎีการสรุปอ้างอิง เพื่อประมาณองค์ประกอบของความแปรปรวนสำหรับผู้ตัดสิน (judges) และข้อสอบ (items) และถ้าข้อมูลที่เก็บมาเป็นการสอบที่มากกว่า 1 ครั้ง เราสามารถประมาณองค์ประกอบของความแปรปรวนสำหรับจำนวนครั้งในการสอบได้ด้วย การประมาณองค์ประกอบของความแปรปรวน สามารถนำมาใช้ในการประมาณ standard error ของคะแนนการผ่าน (Brennan & Lockwood, 1980; Kane & Wilson, 1984 ,cited in Kane, 1994) วิธีนี้มีข้อดีคือ ง่ายในการใช้มากกว่าวิธีการศึกษาหลายๆ วิธี ส่วนข้อบกพร่องมันไม่ได้รวมแหล่งความคลาดเคลื่อนที่เป็นไปได้ในวิธีการศึกษาหลายๆ วิธี โดยเฉพาะอย่างยิ่ง การประมาณ standard error ที่ไม่ได้รวมความแปรปรวนที่แตกต่างกันในการนำไปใช้ข้ามวิธีการที่แตกต่างกัน

2.3 **เกณฑ์ภายนอก** เกี่ยวข้องกับการเปรียบเทียบกับผลของแหล่งข้อมูลเกี่ยวกับผลสัมฤทธิ์ทางการเรียนของนักเรียนที่เป็นอิสระกัน การเปรียบเทียบนี้คล้ายกับความตรงลู่เข้า (convergent validity) สำหรับสเกลคะแนน และพบว่าไม่มีการเปรียบเทียบเพียงแบบเดียว

ที่เป็นตัวชี้ขาดแต่ความคงเส้นคงวาของรูปแบบผลลัพธ์ที่ออกมาจะเป็นตัวสนับสนุนความเหมาะสมของคะแนนการผ่าน การเปรียบเทียบมีด้วยกัน 5 แบบคือ (Kane, 1994)

(1) **วิธีทางตรงซึ่งเป็นวิธีเกณฑ์สัมพัทธ์** (The direct criterion-related approach) คะแนนการผ่านใช้ในการตัดสินใจในกิจกรรมบางอย่างเช่น โรงเรียน งาน การปฏิบัติทางวิชาชีพ วิธีทางตรงส่วนใหญ่ตรวจสอบความตรงของการตัดสินใจจากกลุ่มผู้สอบที่มีการประเมินที่สมบูรณ์ และกลุ่มนี้ได้รับการรับรองในกิจกรรมนั้น ถ้าผู้สอบได้คะแนนสูงก็มีแนวโน้มสูงที่จะทำได้ดีในกิจกรรมนั้นๆ และผู้สอบที่ได้คะแนนต่ำก็มีแนวโน้มทำกิจกรรมนั้นได้ไม่ดี ถ้าอัตราการผ่านมีความเหมาะสมเทียบเท่ากับการประเมินและเกณฑ์การปฏิบัติของกิจกรรมนั้น เราก็มีหลักฐานที่บอกว่าคะแนนการผ่านนั้นเหมาะสม วิธีนี้ไม่ค่อยนิยมนำมาใช้เพราะ การหาเกณฑ์เป็นไปได้ยากว่ามีความตรงหรือไม่ และการระบุนักเรียนที่ผ่านและตกในกิจกรรมนั้นทำไม่ได้หลังจากแบบสอบนั้นถูกนำไปใช้ มันจะไม่ได้รับการอนุญาตจากคนที่ไม่ได้เตรียมตัวในการสอบ

(2) **การเปรียบเทียบผลของวิธีการกำหนดมาตรฐานด้วยวิธีอื่น** (Comparisons to results of other standard-setting methods) ทางหนึ่งในการตรวจสอบความเหมาะสมของผลคะแนนการผ่านจากการศึกษาการกำหนดมาตรฐานโดยการเปรียบเทียบกับการศึกษาการกำหนดมาตรฐานอื่นๆ ที่มีการใช้แบบสอบเดียวกันแต่ใช้วิธีการแตกต่างกันเช่น ถ้าใช้วิธีการแองกอฟในการศึกษาเดิม การศึกษาใหม่อาจใช้วิธีการนีเดลสกี อีเบล หรือเจลเกอร์ เป็นต้น

(3) **การเปรียบเทียบการตัดสินใจผ่าน-ตก ที่ทำในแบบสอบต่างกัน** (Comparison to pass-fail decisions made with a different test) คะแนนสำหรับผู้สอบคนเดียวจากแบบสอบต่างกันสามารถใช้ในการตรวจสอบมาตรฐานโดยผลของแบบสอบฉบับที่ 2 จะถูกพิจารณาความตรงเพื่อใช้ตัดสิน วิธีนี้น่าสนใจ ถ้าผลของแบบสอบฉบับที่ 2 มีความคงทน(exist) และได้มาง่าย แต่สิ่งนี้นำมาซึ่งอันตรายที่นำแบบสอบฉบับที่ 2 มาใช้เพื่อความสะดวกมากกว่า เพื่อความตรงประเด็น

(4) **การเปรียบเทียบด้วยวิธีการประเมินอื่น** (Comparisons involving other assessment methods) ความตรงของมาตรฐานสำหรับแบบสอบฉบับหนึ่งจะถูกตรวจสอบโดยเปรียบเทียบการตัดสินใจผ่าน-ตก จากกลุ่มตัวอย่างผู้สอบเพื่อทำการตัดสินผ่าน-ตก ในผู้สอบเดียวกันโดยใช้การประเมินระดับผลสัมฤทธิ์ทางการเรียนของผู้สอบ ระดับผลสัมฤทธิ์ทางการเรียนของผู้สอบแต่ละคนอาจจะประเมินจากประสบการณ์ของครูเพื่อจะตรวจสอบคะแนนการผ่านที่เหมาะสมที่ใช้วิธีการแองกอฟ one-on–one assessments หรือการ

ตัดสินของครูเพื่อระบุนักเรียนที่คาบเส้น ถ้านักเรียนคาบเส้นได้คะแนนจากแบบสอบโดยได้คะแนนใกล้เคียงคะแนนการผ่าน สิ่งนี้จะเป็นตัวสนับสนุนว่าคะแนนการผ่านมีความเหมาะสม

(5) **การเปรียบเทียบการแจกแจงของกลุ่ม** (Comparisons of group distributions) การตรวจสอบความตรงภายนอกทั้งหมดนั้นเน้นการวิเคราะห์คะแนนรายบุคคล ความเหมาะสมของคะแนนการผ่านถูกประเมินโดยการวิเคราะห์ข้อมูลเกี่ยวกับการแจกแจงคะแนนโดยเฉพาะอัตราการผ่านที่ใช้เสนอคะแนนการผ่านจะถูกเปรียบเทียบกับอัตราการผ่านที่พบในสถานการณ์อื่น ตัวอย่าง ถ้าการแจกแจงผลสัมฤทธิ์ทางการเรียนในประชากรที่สนใจสามารถกำหนดอย่างหยาบๆ เหมือนกับการแจกแจงในประชากรอื่น และถ้าสัดส่วนของประชากรในประชากรอื่นที่ตัดสินว่ามีความสามารถเป็นกลุ่มที่รู้ ก็ควรคาดหวังว่าอัตราการผ่านคล้ายกันสำหรับประชากรที่สนใจ ถ้าอัตราการผ่านที่ใช้เสนอคะแนนการผ่านเป็นอัตราการผ่านที่รู้อยู่แล้ว ความเหมาะสมของคะแนนการผ่านจะได้รับการสนับสนุน อย่างไรก็ตามความเชื่อมั่นในความเหมาะสมของคะแนนการผ่านจะลดลง ดังตัวอย่าง ถ้าอัตราการผ่านในการสอบขอใบรับรองมี 90% สำหรับในปีนั้น และถ้าระหว่างช่วงนั้นการขอใบรับรองใหม่เป็นที่น่าพอใจในการปฏิบัติ และไม่มีสิ่งใดเกิดขึ้นที่เป็นสาเหตุให้ความสามารถของผู้สมัครลดลง หรือเพิ่มขึ้นในระดับของทักษะที่ต้องการในการปฏิบัติ แล้วผลของการศึกษาการกำหนดมาตรฐานตัวใหม่จะถูกใช้เป็นคะแนนการผ่านตัวใหม่สำหรับการสอบโดยคาดหวังว่าอัตราการผ่านน่าจะประมาณ 90% แต่ถ้าคะแนนการผ่านตัวใหม่มีอัตราการผ่านเป็น 60% แสดงว่าความเหมาะสมของคะแนนการผ่านตัวใหม่มีความน่าสงสัย

ตอนที่ 5 แนวคิดเกณฑ์ประเมินกระบวนการกำหนดมาตรฐานเพื่อใช้ประเมินความเหมาะสม และความเป็นไปได้ในการนำไปใช้

แนวคิดเกณฑ์ประเมินกระบวนการกำหนดมาตรฐานของ Pitoniak (Cizek; Bunch; Koons, 2004) และของ Jaeger, Kane, Norcini, Mehrens (Kane ,1998) ประกอบด้วย

1) **ความชัดเจน** (Explicitness) คือ ระดับความชัดเจนเกี่ยวกับจุดประสงค์ของการกำหนดมาตรฐาน และกระบวนการ

2) **การนำไปปฏิบัติได้** (Practicability) คือ ความง่ายในการนำกระบวนการไปใช้ และการวิเคราะห์ข้อมูลแสดงถึงระดับกระบวนการที่มีความน่าเชื่อถือ และแปลผลได้ตรงกัน

3) **การดำเนินการ** (Implementation) คือ ระดับแสดงถึงกระบวนการที่มีความสมเหตุสมผล เป็นระบบ และได้ผลแม่นยำ เช่น การเลือก และการฝึกผู้ตัดสิน การกำหนดมาตรฐานการปฏิบัติ และการรวบรวมข้อมูล

4) **ผลย้อนกลับ** (feedback) คือ ผู้ตัดสินมีความเชื่อมั่นในกระบวนการการตัดสินคะแนนจุดตัด

5) **เอกสารที่ใช้ในกระบวนการ** (Documentation) คือ จุดเด่นในการศึกษาที่เกิดจากการทบทวนเอกสารต่างๆ สำหรับการประเมิน และจุดประสงค์ของการสื่อสาร

นอกจากนี้ Jaeger, Kane, Norcini, Mehrens (Kane ,1998) ได้กล่าวถึง เกณฑ์กระบวนการในการกำหนดมาตรฐานไว้ว่า เป็นเรื่องเกี่ยวกับการเลือกและการฝึกผู้ตัดสิน การกำหนดนิยาม มาตรฐานการปฏิบัติ และการออกแบบการเก็บรวบรวมข้อมูล

จะเห็นได้ว่า แนวคิดเกณฑ์ประเมินกระบวนการกำหนดมาตรฐานของ Jaeger, Kane, Norcini และ Mehrens เป็นส่วนหนึ่งของแนวคิดเกณฑ์ประเมินกระบวนการกำหนดมาตรฐานของ Pitoniak

ตอนที่ 6 งานวิจัยที่เกี่ยวข้อง

1. งานวิจัยที่เกี่ยวข้องกับการกำหนดจุดตัดด้วยวิธีแองกอฟ

จากการศึกษางานวิจัยที่เกี่ยวข้องกับการกำหนดมาตรฐาน หรือคะแนนจุดตัดด้วยวิธีแองกอฟ ผู้วิจัยได้แบ่งเป็น 2 กลุ่มดังนี้

1.1 การเปรียบเทียบวิธีการกำหนดมาตรฐาน หรือคะแนนจุดตัด

เบรนเนนและล็อกวูด (Brennan and Lockwood, 1980) ศึกษาเปรียบเทียบการกำหนดจุดตัดด้วยวิธีของแองกอฟและนีเดลสกี โดยใช้ทฤษฎีสรุปอ้างอิง ดำเนินการวิจัยโดยใช้กลุ่มตัวอย่างเป็นผู้ตัดสิน 5 คน ตัดสินข้อสอบ 126 ข้อ ด้วยวิธีของแองกอฟ และวิธีของนีเดลสกี พบว่า 1) วิธีของแองกอฟ และวิธีของนีเดลสกี ให้คะแนนจุดตัดแตกต่างกันอย่างมีนัยสำคัญทางสถิติ และความเบี่ยงเบนมาตรฐานของคะแนนจุดตัดของวิธีนีเดลสกีจะเป็น 2 เท่าของวิธีแองกอฟ 2) เมื่อทำการวิเคราะห์ความแปรปรวน (ANOVA) แบบ 3 ทาง คือ วิธีกำหนดคะแนนจุดตัด 2 วิธี x ผู้ตัดสิน 5 คน x ข้อสอบ 126 ข้อ พบว่า ส่วนเบี่ยงเบนมาตรฐานของคะแนนจุดตัดของวิธีนีเดลสกีใกล้เคียงกับส่วนเบี่ยงเบนมาตรฐานของการตัดสินโดยวิธีแองกอฟ 3) การใช้ผู้ตัดสิน

กลุ่มเดียวกันทำการตัดสินข้อสอบทั้งวิธีแองกอฟ และวิธีนีเดลสกี พบว่า วิธีแองกอฟ กับวิธีนีเดลสกีไม่สัมพันธ์กัน (r=-.052) คะแนนจุดตัดของวิธีแองกอฟสูงกว่าของวิธีนีเดลสกี และส่วนเบี่ยงเบนมาตรฐานของคะแนนจุดตัดของวิธีแองกอฟต่ำกว่าวิธีนีเดลสกี

ครอสและคณะ (Cross, et al., 1984) ศึกษาความสอดคล้องในการกำหนดคะแนนจุดตัด (Indics of consistency) ระหว่างวิธีของแองกอฟ วิธีของนีเดลสกี และวิธีของเจลเกอร์ โดยให้ผู้เชี่ยวชาญกำหนดคะแนนจุดตัด 3 ครั้ง และวิเคราะห์ความสอดคล้องในการกำหนดคะแนนจุดตัดด้วยวิธีการประยุกต์ตามทฤษฎีการสรุปอ้างอิง (Generalizability Theory) พบว่า วิธีของแองกอฟมีความสอดคล้องในการกำหนดคะแนนจุดตัดสูงสุด ส่วนวิธีนีเดลสกีมีความสอดคล้องในการกำหนดคะแนนจุดตัดต่ำสุด

เคน (Kane, 1987) ใช้ทฤษฎีการตอบสนองข้อสอบในการวิเคราะห์ผลการตัดสินคะแนนจุดตัดด้วยวิธีของแองกอฟ และเปรียบเทียบวิธีวิเคราะห์ 3 วิธี ได้แก่ วิธีการวิเคราะห์แบบดั้งเดิม วิธีวิเคราะห์ตามทฤษฎีการสรุปอ้างอิง และวิธีวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบ พบว่า วิธีการวิเคราะห์แบบดั้งเดิมไม่ใช่วิธีที่ดีที่สุด และได้สนับสนุนให้ใช้วิธีที่วิเคราะห์ตามทฤษฎีการสรุปอ้างอิง

เพรคและอิมพารา (Plake and Impara, 1996) ศึกษาความสอดคล้องในการตัดสินคะแนนจุดตัดด้วยวิธีของแองกอฟโดยให้ผู้เชี่ยวชาญกำหนดสัดส่วนการตอบข้อสอบได้ถูกต้องของผู้สอบที่มีความสามารถต่ำสุดที่ยอมรับได้ โดยใช้ข้อสอบ 24 ข้อ และให้ผู้เชี่ยวชาญตัดสิน 2 ครั้ง ผลสรุปชี้ให้เห็นว่า มีความสอดคล้องในการตัดสินคะแนนจุดตัดของผู้เชี่ยวชาญสูง ในกรณีแสดงความแตกต่างของค่าเฉลี่ยสมบูรณ์สำหรับการประมาณค่าการปฏิบัติข้อสอบ และแสดงให้เห็นคุณสมบัติของแบบสอบมีผลต่อระดับความสอดคล้องในการตัดสินคะแนนจุดตัด

อิมพาราและเพรค (Impara and Plake, 1997) ได้เปรียบเทียบวิธีการกำหนดคะแนนจุดตัดระหว่างวิธีแองกอฟแบบดั้งเดิม กับวิธีของอิมพาราและเพรค หรือวิธีตัดสินว่าได้ หรือตก(yes-no method) พบว่าวิธีกำหนดคะแนนจุดตัดทั้งสองให้คะแนนที่ใกล้เคียงกัน และผู้ตัดสินมีความเห็นว่าวิธีของอิมพาราและเพรคมีความสะดวกในการปฏิบัติมากกว่าวิธีของแองกอฟแบบดั้งเดิม

กัญจนา ลินทรัตนศิริกุล (2534) ได้เปรียบเทียบคะแนนจุดตัด จำนวนผู้รอบรู้ ความเชื่อมั่น และความเที่ยงตรง ในการจำแนกผู้สอบเป็นผู้รอบรู้ โดยการกำหนดคะแนนจุดตัด 2 วิธี คือ วิธีการของแวน เดอร์ ลินเดน และวิธีการของเคน จากการกำหนดเกณฑ์มาตรฐาน 2 วิธี คือ เทคนิคของนีเดลสกีกับวิธีการใช้กลุ่มคาบเส้น โดยนำทฤษฎีการตอบสนองข้อสอบมาประยุกต์ใช้ในการกำหนดคะแนนจุดตัด ผลการวิจัยพบว่า 1) เมื่อกำหนดเกณฑ์

มาตรฐานตามเทคนิคของนีเดลสกีที่นักเรียนระดับพอใช้ แบบสอบส่วนใหญ่มีคะแนนจุดตัดตามวิธีการของแวน เดอร์ ลินเดน สูงกว่าวิธีการของเคน แต่นักเรียนระดับดี พบว่า แบบสอบส่วนใหญ่มีคะแนนจุดตัดตามวิธีการของแวน เดอร์ ลินเดน เท่ากับวิธีการของเคน 2) เมื่อกำหนดเกณฑ์มาตรฐานโดยวิธีการใช้กลุ่มคาบเส้นในกลุ่มนักเรียนที่มีความสามารถระดับพอใช้ และในระดับดีพบว่า คะแนนจุดตัดตามวิธีการของแวน เดอร์ ลินเดนสูงกว่าวิธีการของเคนในแบบสอบทุกฉบับ 3) คะแนนจุดตัดตามวิธีการของแวน เดอร์ ลินเดน จากการกำหนดมาตรฐานตามเทคนิคของนีเดลสกี พบว่า แบบสอบส่วนใหญ่มีคะแนนจุดตัดสูงกว่าวิธีการใช้กลุ่มคาบเส้นในนักเรียนระดับพอใช้ และคะแนนจุดตัดจากการกำหนดเกณฑ์มาตรฐานตามเทคนิคของนีเดลสกี พบว่า แบบสอบส่วนใหญ่มีคะแนนจุดตัดเท่ากับวิธีการใช้กลุ่มคาบเส้นในกลุ่มนักเรียนที่มีความสามารถระดับดี 4) คะแนนจดตัดตามวิธีการของเคนจากการกำหนดเกณฑ์มาตรฐานตามเทคนิคของนีเดลสกี สูงกว่าวิธีการใช้กลุ่มคาบเส้นทั้งในนักเรียนระดับพอใช้ และระดับดี 5) จำนวนผู้รอบรู้ขึ้นอยู่กับคะแนนจุดตัด 6) ความเชื่อมั่นในการจำแนกผู้สอบเป็นผู้รอบรู้ขึ้นอยู่กับคะแนนจุดตัด 7) ความเที่ยงตรงในการจำแนกผู้สอบเป็นผู้รอบรู้ขึ้นอยู่กับความเชื่อมั่นในการจำแนกผู้สอบเป็นผู้รอบรู้

สาลิกา เมธนาวิน (2537) ได้เปรียบเทียบการประมาณค่าความเชื่อมั่นของแบบสอบอิงเกณฑ์ ตามวิธีของฮวน วิธีของเบรนนอนและเคน และวิธีของราชู เมื่อกำหนดคะแนนจุดตัดต่างกัน 4 วิธีคือ วิธีของนีเดลสกี แองกอฟ อีเบลและวิธีนับลดจาก 100% พบว่าค่าความเชื่อมั่นของแบบทดสอบประมาณค่าตามวิธีของฮวน เมื่อกำหนดคะแนนจุดตัด 4 วิธี วิธีนับลดจาก 100% ให้ค่าความเชื่อมั่นสูงสุด ส่วนที่เหลือ 3 วิธีให้ค่าความเชื่อมั่นต่ำมาก ส่วนค่าความเชื่อมั่นของแบบทดสอบประมาณค่าตามวิธีของเบรนนอนและเคน พบว่าวิธีนับลดจาก 100% ให้ค่าความเชื่อมั่นสูงที่สุด รองลงมาได้แก่เทคนิคของอีเบล แองกอฟ และนีเดลสกีตามลำดับ และค่าความเชื่อมั่นของแบบสอบประมาณค่าตามวิธีของราชูเมื่อกำหนดคะแนนจุดตัด 4 วิธี วิธีนับลดจาก 100% ให้ค่าความเชื่อมั่น รองลงมาได้แก่เทคนิคของอีเบล แองกอฟ และนีเดลสกีตามลำดับ

เสมอ ตรีมงคล (2539) ได้ศึกษาสัมประสิทธิ์การสรุปอ้างอิงของวิธีการหาคะแนนจุดตัดตามเทคนิคของนีเดลสกีและแองกอฟ โดยมีจุดมุ่งหมายดังนี้ 1) เพื่อเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิงที่ได้จากวิธีกำหนดคะแนนจุดตัด ตามเทคนิคนีเดลสกีและแองกอฟ 2) เพื่อหาขนาดของกลุ่มตัวอย่างของข้อสอบ ผู้เชี่ยวชาญ และกลุ่มโรงเรียนมาใช้ในการกำหนดคะแนนจุดตัด ตามเทคนิคของนีเดลสกี และเทคนิคของแองกอฟที่ให้ค่าสัมประสิทธิ์การสรุปอ้างอิงอย่างน้อยที่สุดเป็น 0.80 สำหรับองค์ประกอบของการศึกษามี

ความสัมพันธ์แบบ i x (r:g) คือ ผู้เชี่ยวชาญทุกคน ประเมินข้อสอบทุกข้อ และสุ่มมาจากกลุ่มโรงเรียนแต่ละกลุ่มจำนวนเท่ากัน ผลการศึกษาพบว่า 1) ค่าสัมประสิทธิ์การสรุปอ้างอิงของวิธีการหาคะแนนจุดตัด ตามเทคนิคแองกอฟและนีเดลสกี รูปแบบการวัดที่สรุปผลการวัดไปยังเอกภพของผู้เชี่ยวชาญและกลุ่มโรงเรียนพร้อมกัน ให้ค่าสูงกว่ารูปแบบการวัดที่สรุปผลการวัดไปยังเอกภพของผู้เชี่ยวชาญอย่างเดียว และเปรียบเทียบค่าสัมประสิทธ์การสรุปอ้างอิงแต่ละค่าพบว่ามี 2 คู่ที่แตกต่างอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 2) ขนาดของกลุ่มตัวอย่างของแต่ละองค์ประกอบที่สามารถปฏิบัติได้ และทำให้ค่าสัมประสิทธิ์การสรุปอ้างอิงมีค่าอย่างน้อยเป็น 0.80 เมื่อมีข้อสอบ 40 ข้อ สำหรับวิธีการหาคะแนนจุดตัด ตามเทคนิคแองกอฟรูปแบบการวัดที่สรุปการวัดไปยังเอกภพของผู้เชี่ยวชาญและกลุ่มโรงเรียนพร้อมกัน ต้องใช้กลุ่มโรงเรียน 6 กลุ่ม และผู้เชี่ยวชาญกลุ่มโรงเรียนละ 6 คน และรูปแบบการวัดที่สรุปผลการวัดไปยังเอกภพของผู้เชี่ยวชาญอย่างเดียวต้องใช้กลุ่มโรงเรียน 6 กลุ่ม และผู้เชี่ยวชาญกลุ่มโรงเรียนละ 5 คน ส่วนเทคนิคนีเดลสกีรูปแบบการวัดที่สรุปผลการวัดไปยังเอกภพของผู้เชี่ยวชาญและกลุ่มโรงเรียนพร้อมกัน และรูปแบบการวัดที่สรุปผลการวัดไปยังเอกภพของผู้เชี่ยวชาญอย่างเดียวต้องใช้กลุ่มโรงเรียน 20 กลุ่ม และผู้เชี่ยวชาญกลุ่มโรงเรียนละ 15 คน

วิภารัตน์ ศรีบุตรตา (2541) ได้เปรียบเทียบความสอดคล้องในการตัดสินคะแนนจุดตัดระหว่างวิธีของแองกอฟ วิธีของนีเดลสกี และวิธีของอิมพาราและเพลค พบว่า ค่าความสอดคล้องในการตัดสินคะแนนจุดตัดของผู้เชี่ยวชาญเชิงสัมพัทธ์ด้วยวิธีของแองกอฟ วิธีของนีเดลสกี และวิธีของอิมพาราและเพลคครั้งที่ 1 เท่ากับ 0.7418, 0.3275, และ 0.337 ตามลำดับ สำหรับครั้งที่ 2 เท่ากับ 0.6617, 0.3281 และ 0.384 ตามลำดับ จากการทดสอบพบว่า ค่าความสอดคล้องในการตัดสินไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนค่าความสอดคล้องในการตัดสินคะแนนจุดตัดของผู้เชี่ยวชาญเชิงสัมบูรณ์ด้วยวิธีของแองกอฟ วิธีของนีเดลสกี และวิธีของอิมพาราและเพลคครั้งที่ 1 เท่ากับ 0.5688, 0.2759, และ 0.329 ตามลำดับ สำหรับครั้งที่ 2 เท่ากับ 0.4593, 0.2700 และ 0.384 ตามลำดับ จากการทดสอบพบว่า ค่าความสอดคล้องในการตัดสินไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

วิมลมาศ ยิ้มละมัย (2541) เปรียบเทียบคุณภาพของการกำหนดจุดตัดระหว่างวิธีวิลคอกซ์ กับวิธีแองกอฟสำหรับแบบสอบอิงโดเมนประเภทเลือกตอบและตอบสั้น พบว่า ความตรงของการกำหนดจุดตัดคำนวณด้วยวิธีของคาร์เวอร์พบว่า ความตรงของการกำหนดจุดตัดด้วยวิธีการทั้งสองไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนค่าความเที่ยงของการกำหนดจุดตัดด้วยวิธีของสวามินาทาน แฮมเบิลตัน และอัลจินา พบว่า

ความเที่ยงของการกำหนดจุดตัดด้วยวิธีการทั้งสองไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 นั้นคือคุณภาพของการกำหนดจุดตัดด้วยวิธีวิลคอกซ์ กับวิธีแองกอฟตามทฤษฎีการตอบสนองข้อสอบไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

จะเห็นได้ว่าในการเปรียบเทียบด้วยวิธีต่างๆ ซึ่งส่วนใหญ่เป็นวิธีที่ใช้แบบสอบเป็นศูนย์กลาง หรือเป็นการนำแบบสอบมาช่วยในการพิจารณาพบว่า วิธีแองกอฟเป็นวิธีที่ดี แต่ไม่สะดวกในการใช้เมื่อเทียบกับวิธีอิมพาราและเพรค นอกจากนี้วิธีที่นำมาใช้ในการเปรียบเทียบคะแนนจุดตัดส่วนใหญ่ใช้ทฤษฎีการสรุปอ้างอิงซึ่ง Kane (1987) ได้สนับสนุนให้ใช้วิธีที่วิเคราะห์ตามทฤษฎีการสรุปอ้างอิงมากกว่าวิธีการวิเคราะห์แบบดั้งเดิม

1.2 ปัจจัยที่ส่งผลต่อการกำหนดมาตรฐานด้วยวิธีแองกอฟ

แพทริเซีย (Patricia, 1991) ศึกษาความเหมาะสมของการกำหนดจุดตัดด้วยวิธีของแองกอฟโดยใช้ผู้เชี่ยวชาญเฉพาะวิชา และได้รับการฝึกเกี่ยวกับการกำหนดคะแนนจุดตัด พบว่า การกำหนดคะแนนจุดตัดโดยใช้ดุลยพินิจของผู้เชี่ยวชาญไม่ได้ตัดสินโดยปราศจากเหตุผล และคะแนนจุดตัดที่ได้ขึ้นอยู่กับคุณภาพของผู้เชี่ยวชาญ

นอร์ซินิ และคณะ (Norcini, et. al., 1987) ตรวจสอบความแปรผันในการกำหนดคะแนนจุดตัดด้วยวิธีของแองกอฟที่มีผลต่อข้อสรุปของความสอดคล้องโดยใช้กระบวนการกลุ่ม พบว่า การตัดสินด้วยกระบวนการกลุ่มเพิ่มประสิทธิภาพในการกำหนดคะแนนจุดตัดด้วยวิธีของแองกอฟ

เบอร์ก (Berk, 1980) ได้ศึกษาวิจัยถึงความคลาดเคลื่อนประเภท I (type I error) คือ ผู้รอบรู้แต่ถูกตัดสินให้เป็นผู้ไม่รอบรู้ และความคลาดเคลื่อนประเภทที่ 2 (type II error) คือผู้ไม่รอบรู้ แต่ถูกตัดสินให้เป็นผู้รอบรู้ โดยสรุปได้ว่า ถ้าใช้จำนวนข้อสอบ 15 ข้อ ในแต่ละจุดประสงค์ กับการสอบที่ส่งผลย้อนกลับทันทีที่สอบเสร็จ จะทำให้การตัดสินถูกต้องแม่นยำ แต่ถ้าใช้ 2-3 ข้อ ในแต่ละจุดประสงค์จะเกิดความคลาดเคลื่อนในการตัดสินมาก และเบอร์กได้เสนอแนะว่า การกำหนดจำนวนข้อสอบสำหรับการประเมินผลความก้าวหน้า หรือการประเมินผลเพื่อจัดตำแหน่งนั้น ควรใช้ข้อสอบ 5-10 ข้อ แต่ถ้าเป็นการประเมินผลสรุปรวม หรือเป็นการตัดสินแบบถาวร (permanent decision) จำเป็นต้องใช้ข้อสอบจำนวนมากขึ้นเป็น 2 เท่านั้นคือ จำนวน 10–20 ข้อ

นภา นิ่มนาค (2538) ศึกษาคะแนนจุดตัด และความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ เมื่อกำหนดคะแนนจุดตัดด้วยวิธีของแองกอฟ วิธีของนีเดลสกี และวิธีของ

อีเบล โดยผู้เชี่ยวชาญ 2 กลุ่มคือ กลุ่มที่มีประสบการณ์ 1-5 ปี และ มากกว่า 5 ปี ซึ่งพบว่า วิธีของแองกอฟ และวิธีของอีเบล ผู้เชี่ยวชาญที่มีประสบการณ์การสอน 1-5 ปีกำหนดคะแนนจุดตัดสูงกว่า ผู้เชี่ยวชาญที่มีประสบการณ์การสอน มากกว่า 5 ปี ส่วนวิธีของนีเดลสกี ผู้เชี่ยวชาญที่มีประสบการณ์การสอน 1-5 ปี กำหนดคะแนนจุดตัดต่ำกว่า ผู้เชี่ยวชาญที่มีประสบการณ์การสอนมากกว่า 5 ปี นอกจากนี้ค่าความเชื่อมั่นของแบบสอบอิงเกณฑ์ เมื่อกำหนดคะแนนจุดตัดด้วยวิธีเดียวกัน แต่ประสบการณ์ของผู้เชี่ยวชาญต่างกัน พบว่าค่าความเชื่อมั่นแตกต่างอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ยกเว้นวิธีของนีเดลสกีค่าความเชื่อมั่นแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ และเมื่อกำหนดจุดตัดต่างวิธีกัน พบว่า กรณีผู้เชี่ยวชาญทีมีประสบการณ์ 1 – 5 ปี ทั้ง 3 วิธีค่าความเชื่อมั่นแตกต่างกัน กรณีผู้เชี่ยวชาญมีประสบการณ์การสอนมากกว่า 5 ปี พบว่าวิธีแองกอฟและอีเบล วิธีนีเดลสกีและอีเบล ค่าความเชื่อมั่นแตกต่างกัน ส่วนแองกอฟและนีเดลสกี ค่าความเชื่อมั่นไม่แตกต่างกัน

อัจฉริยา ปราบอริพ่าย (2531) ได้เปรียบเทียบคะแนนจุดตัดของแบบสอบ และดัชนีความคงที่ของการตัดสินคะแนนจุดตัดซึ่งตัดสินโดยผู้เชี่ยวชาญที่ไม่มีสารสนเทศและผู้เชี่ยวชาญที่มีสารสนเทศ เกี่ยวกับค่าพารามิเตอร์ของข้อสอบ (ค่าอำนาจจำแนก ค่าความยาก และค่าการเดา) ความถี่ของคะแนนสอบ สัดส่วนของผู้สอบที่ตอบข้อสอบแต่ละข้อได้ถูกและความน่าจะเป็นที่จะตอบข้อสอบแต่ละข้อได้ถูกซึ่งประมาณโดยผู้เชี่ยวชาญที่ไม่มีสารสนเทศ ผลการวิจัยพบว่า คะแนนจุดตัดที่ใช้เป็นเกณฑ์ตัดสินความสามารถขั้นต่ำที่ยอมรับได้ของผู้สอบซึ่งอยู่บนสเกลของคะแนนโดเมนที่ได้จากการตัดสินของผู้เชี่ยวชาญที่ไม่มีสารสนเทศ และผู้เชี่ยวชาญที่มีสารสนเทศเท่ากับ .66 และ .53 ตามลำดับ ซึ่งแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 สำหรับค่าดัชนีความคงที่ของการตัดสินคะแนนจุดตัดซึ่งตัดสินโดยผู้เชี่ยวชาญที่มีสารสนเทศ เท่ากับ .78 และ .76 ตามลำดับ ซึ่งไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

จากงานวิจัยดังกล่าวจะเห็นว่า คุณภาพของผู้เชี่ยวชาญ เช่น ประสบการณ์ของผู้เชี่ยวชาญ จำนวนข้อสอบ การให้สารสนเทศกับผู้เชี่ยวชาญ และการใช้กระบวนการกลุ่มในการพิจารณาหาคะแนนจุดตัด สิ่งเหล่านี้ต่างส่งผลต่อคะแนนจุดตัด

2. งานวิจัยที่เกี่ยวข้องกับการกำหนดจุดตัดด้วยวิธีบุ๊คมาร์ค

บาเคนดัลส (Buckendahl, 2002) ได้ทำการเปรียบเทียบการกำหนดมาตรฐานด้วยวิธีการแองกอฟ และ วิธีบุ๊คมาร์ค เพื่อเปรียบเทียบจุดตัดที่ใช้วิธีแองกอฟและวิธีบุ๊คมาร์ค นักเรียนเกรด 7 จำนวน 448 คน เครื่องมือที่ใช้ในการวิจัยคือ แบบทดสอบคณิตศาสตร์เกรด 7 แบบเลือกคำตอบที่มีเนื้อหาทางทักษะคณิตศาสตร์ 6 อย่าง โดยมีวิธีดำเนินการดังนี้ 1) นำคณะผู้ตัดสินมาทำสัมมนาเชิงปฏิบัติการ 1 วัน เพื่อสร้างความเข้าใจในวิธีการทั้งสองวิธี 2) แบ่งกลุ่มผู้ตัดสินเป็น 2 กลุ่มคือ สำหรับวิธีแองกอฟ และวิธีบุ๊คมาร์ค 3) แต่ละกลุ่มพิจารณาหาจุดตัด 2 รอบ โดยรอบแรกพิจารณาอย่างอิสระ รอบสองอภิปรายร่วมกันและเปิดโอกาสให้ผู้ตัดสินแต่ละคนเปลี่ยนจุดตัดของตนได้ 4) หาค่าเฉลี่ยจุดตัด โดยมีวิธีการวิเคราะห์ข้อมูลดังนี้ 1) หาค่าความยากโดย CTT 2) เปรียบเทียบจุดตัดและส่วนเบี่ยงเบนมาตรฐานของรอบที่ 1 และรอบที่ 2 ในแต่ละวิธี ผลการวิจัยพบว่า วิธีแองกอฟ รอบแรกได้คะแนนจุดตัด 34.92 ส่วนเบี่ยงเบนมาตรฐาน 7.79 และเปอร์เซนต์การตกของนักเรียน 8.9 รอบสองได้คะแนนจุดตัด 33.42 ส่วนเบี่ยงเบนมาตรฐาน 10.96 และเปอร์เซนต์การตก 7.6 สำหรับวิธี บุ๊คมาร์ค รอบแรกได้คะแนนจุดตัด 33.64 ส่วนเบี่ยงเบนมาตรฐาน 11.03 เปอร์เซนต์การตก 7.6 รอบที่สองได้คะแนนจุดตัด 35.64 ส่วนเบี่ยงเบนมาตรฐาน 8.66 เปอร์เซนต์การตก 9.4 จะเห็นได้ว่า คะแนนจุดตัดในรอบสุดท้ายของทั้งสองวิธีมีความแตกต่างกันเพียงเล็กน้อย (คะแนนห่างกัน 2 คะแนน และเปอร์เซนต์การตกห่างกัน 1.8% ของนักเรียนที่ทำการสอบ) และพบว่าวิธีบุ๊คมาร์ค ในรอบที่สองมีส่วนเบี่ยงเบนมาตรฐานลดลงเมื่อเทียบกับวิธีแองกอฟ ซึ่งการลดลงของส่วนเบี่ยงเบนมาตรฐานในรอบสองของวิธีบุ๊คมาร์คนี้ แสดงให้เห็นว่าพิสัยของคะแนนจุดตัดลดลง และชี้ให้เห็นว่า การตัดสินในระหว่างผู้ตัดสิน (inter-judge) มีระดับสูงขึ้น

แวง (Wang, 2003) ได้นำโมเดล Rasch IRT มาใช้ในการกำหนดมาตรฐานวิธีการผังข้อสอบ (item mapping) เพื่อกำหนดหามาตรฐาน หรือคะแนนจุดตัดสำหรับการสอบใบรับรองและประกาศนียบัตรด้วยวิธีผังข้อสอบ กลุ่มตัวอย่างเป็นผู้สอบขอใบรับรองและประกาศนียบัตร โดยใช้เครื่องมือคือ แบบสอบแบบเลือกคำตอบสำหรับ professional licensure or certification examinations ซึ่งมีการสอบ 4 ครั้ง สอบครั้งที่ 1 ethics examination of a state insurance licensing program ในเดือนมิถุนายน 1999 สอบครั้งที่ 2 property examination of a state insurance licensing program ในเดือนสิงหาคม 1999 สอบครั้งที่ 3 casualty examination of a state insurance licensing program ในเดือนสิงหาคม 1999 สอบครั้งที่ 4 national nurse aide certification examination ในเดือนธันวาคม 2000 วิธีดำเนินการวิจัยมีดังนี้ 1) ผู้ตัดสิน

ร่วมกันอธิบายลักษณะของผู้สอบที่มีความสามารถต่ำ (MCC) สำหรับขอใบรับรอง 2) วิธีแองกอฟ ผู้ตัดสินจะถูกถามว่าเปอร์เซนต์ของ MCC ที่ตอบข้อสอบข้อนี้ถูกคือเท่าไร จากนั้นให้ค่าสถิติของข้อสอบพิจารณาโดยเปิดโอกาสให้ผู้ตัดสินอภิปรายและปรับเปลี่ยนจุดตัดของตนได้ จากนั้นหาค่าเฉลี่ยจุดตัด 3) วิธีผังข้อสอบจะทำผังข้อสอบให้ผู้ตัดสินพิจารณา และผู้ตัดสินจะถูกถามว่า "ผู้สอบกลุ่มคาบเส้น มีโอกาสตอบข้อสอบถูกอย่างน้อย .50 หรือไม่" จากนั้นจัดหา p-value และเปิดโอกาสให้อภิปรายและปรับเปลี่ยนจุดตัดของตนได้ ปกติข้อสอบหลายข้อในแต่ละคอลัมน์จะถูกเลือกเพื่อให้ผู้ตัดสินหา ฉันทมติของข้อสอบ กระบวนการนี้จะถูกทำซ้ำจนกระทั่งคอลัมน์ข้อสอบถูกกำหนดให้เป็นจุดที่ข้อสอบส่วนใหญ่ในคอลัมน์ได้รับฉันทมติจากผู้ตัดสินส่วนใหญ่ว่ามีความน่าจะเป็นที่ ผู้สอบกลุ่มคาบเส้นจะตอบถูก .50 และระดับตรงกลางของค่าความยากของข้อสอบภายใต้คอลัมน์จะถูกกำหนดเป็นจุดตัด วิธีการวิเคราะห์ข้อมูลมีดังนี้ 1) วิเคราะห์ Rasch IRT ด้วยโปรแกรม BIGSTEPS 2) ใช้ G-theory วิเคราะห์หา Inter-judge consistency (r x i) ใช้โปรแกรม GENOVA วิเคราะห์ G และ D studies ผลการวิจัยพบว่า วิธีผังข้อสอบมีความสอดคล้องในการตัดสินระหว่างผู้ตัดสิน (inter-judge consistency) มากกว่าวิธีการแองกอฟ

บิเรทวาส (Beretvas, 2004) ได้เปรียบเทียบการกำหนดค่าความยากในวิธีการบุ๊คมาร์ค โดยการวิเคราะห์ด้วยโมเดล IRT ที่แตกต่างกัน กลุ่มตัวอย่าง ผู้วิจัยได้ใช้ฐานข้อมูลนักเรียนเกรด 4 จำนวน 63,533 คน ปี 1998 ส่วนเครื่องมือการวิจัยคือ แบบสอบย่อยคณิตศาสตร์นักเรียนเกรด 4 ของ Washington Assessment of student Learning (WASL) โดยข้อสอบเป็นรูปแบบผสม 40 ข้อ แบ่งเป็นแบบเลือกตอบ 24 ข้อ และแบบเขียนตอบ 16 ข้อ ซึ่งแบ่งเป็นแบบตอบสั้น 13 ข้อ ให้คะแนนแบบ 0 , 1 และ 2 และข้อสอบแสดงวิธีทำให้คะแนนเป็น 0, 1 , 2 ,3 และ 4 โดยมีคะแนนรวมเท่ากับ 62 คะแนน ผู้วิจัยนำผลการสอบจากฐานข้อมูลมาวิเคราะห์ข้อมูลหาค่าพารามิเตอร์ข้อสอบตามโมเดล IRT คือ โมเดล IRT ที่ให้คะแนนแบบ 2 ค่า แบบ 1 พารามิเตอร์ และ 2 พารามิเตอร์ และโมเดล IRT ที่ให้คะแนนแบบหลายค่าคือ partial credit (PC), generalized partial credit (GPC) และ graded response (GR) ซึ่งวิเคราะห์ด้วยโปรแกรม Parscale หรือ Mulitilog ค่าพารามิเตอร์ที่ได้มาคำนวณหา BDL (bookmark difficulty locations) ตามโมเดล IRT และ RP ที่แตกต่างกัน โดยใช้โปรแกรม Mathematica ช่วยในการคำนวณ โดยใช้ โมเดล IRT และค่า RP ที่แตกต่างกัน ดังนี้ 1PL-PC RP ½,1PL-PC RP 2/3, 1PL-PC RP 4/5, 3PL-GPC RP ½, 3PL-GPC RP 2/3, 3PL-GPC RP 4/5, 3PL-GR RP ½, 3PL-GR RP 2/3, 3PL-GR RP 4/5 จากนั้นนำค่า BDL มาใช้พิจารณาจัดเรียงข้อสอบตามค่าความยากจากน้อยไปมาก เพื่อนำโมเดล IRT และค่า RP ที่แตกต่างกัน มาเปรียบเทียบกัน โดยใช้ spearman rank order correlation ซึ่งผลการวิเคราะห์จากโมเดล IRT และค่า RP ทั้ง 9 แบบ ค่าสหสัมพันธ์มีค่าระหว่าง .897 - .998 ที่ระดับนัยสำคัญ .001 โดยโมเดล 1PL-PC มีค่าสหสัมพันธ์ระหว่าง .900 - .974

โมเดล 3PL-GPC มีค่าสหสัมพันธ์ระหว่าง .933 - .991 และโมเดล 3PL-GR มีค่าสหสัมพันธ์ระหว่าง .976 - .995 และพบว่าทั้ง 3 โมเดลค่าสหสัมพันธ์ระหว่าง ค่า RP 2/3 และ 4/5 สูงที่สุด สำหรับในกระบวนการบุ๊คมาร์คจะนำ BDL ที่ใช้โมเดล 1PL-PC, 3PL-GPC, 3PL-GR ที่ RP 2/3 มาใช้ในคู่มือการจัดเรียงข้อสอบตามโมเดลการวิเคราะห์แต่ละโมเดล ผลการจัดเรียงข้อสอบพบว่าโมเดล 3PL-GPC, 3PL-GR มีการจัดเรียงข้อสอบใกล้เคียงกันมาก มีความแตกต่างกันเพียง 2 ข้อ ในขณะที่ 1PL-PC มีผลการจัดเรียงแตกต่างจากโมเดลทั้งสองแบบ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 3

วิธีดำเนินการวิจัย

การวิจัยครั้งนี้มีวัตถุประสงค์ 1) เพื่อเปรียบเทียบคุณภาพของการกำหนดมาตรฐานด้านความเที่ยง ความตรง และความเหมาะสมและความเป็นไปได้ระหว่างวิธีการกำหนดมาตรฐานแองกอฟที่ได้รับการปรับปรุงกับวิธีการบุ๊คมาร์ค 2) เพื่อศึกษาผลของจำนวนผู้ตัดสิน และจำนวนครั้งที่มีต่อการกำหนดมาตรฐานด้วยวิธีการแองกอฟที่ได้รับการปรับปรุง และวิธีการบุ๊คมาร์ค ซึ่งมีรายละเอียดของการดำเนินการวิจัยดังนี้

ขั้นตอนการดำเนินการวิจัย

1. ศึกษาเอกสาร งานวิจัยที่เกี่ยวข้องกับการกำหนดมาตรฐาน เช่น ความหมาย วิธีการกำหนดมาตรฐานทั้งแบบดั้งเดิม และแบบใหม่ การตรวจสอบหาความเที่ยง และความตรงของการกำหนดมาตรฐาน

2. กำหนดกรอบแนวคิดการวิจัย โดยมีตัวแปรที่ศึกษาดังนี้

2.1 ตัวแปรอิสระ คือ วิธีการกำหนดมาตรฐาน ได้แก่ วิธีการบุ๊คมาร์ค และวิธีการแองกอฟที่ได้รับการปรับปรุง จำนวนผู้ตัดสิน จำนวนครั้ง

2.2 ตัวแปรตาม คือ คุณภาพของการกำหนดมาตรฐาน ได้แก่ ความตรง ความเที่ยง ความเหมาะสมและความเป็นไปได้

3. กำหนดประชากร และกลุ่มตัวอย่าง แบ่งออกเป็น 2 กลุ่มคือ

3.1 ผู้ตัดสิน คือ อาจารย์ที่สอนในวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 สังกัดสำนักงานการศึกษาขั้นพื้นฐาน

3.2 ผู้สอบ คือ นักเรียนชั้นมัธยมศึกษาปีที่ 3 ปีการศึกษา 2549 ในโรงเรียนสังกัดสำนักงานการศึกษาขั้นพื้นฐาน สำนักงานเขตพื้นที่การศึกษา ชลบุรีเขต 1

4 สร้างและพัฒนาแบบสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ภาคเรียนที่ 1 แบบเลือกตอบ 4 ตัวเลือก จำนวน 4 ฉบับ คือ ฉบับที่ 1 พื้นที่และปริมาตร ฉบับที่ 2 กราฟ ฉบับที่ 3 ระบบสมการเชิงเส้น และฉบับที่ 4 ความคล้าย

5 ดำเนินการเก็บรวบรวมข้อมูล ระยะที่ 1 ดำเนินการสอบกับผู้สอบ คือนักเรียน ชั้นมัธยมศึกษาปีที่ 3

6 ดำเนินการวิเคราะห์ข้อมูล ระยะที่ 1 โดยการวิเคราะห์ข้อสอบเพื่อหาค่าความยาก อำนาจจำแนก และการเดาของข้อสอบโดยใช้โปรแกรม Xcalibre

7 ดำเนินการเก็บรวบรวมข้อมูล ระยะที่ 2 ดำเนินการกำหนดมาตรฐานโดยกลุ่มผู้ตัดสินที่เป็นอาจารย์สอนวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 โดยมีขั้นตอนดังนี้

7.1 กำหนดผู้ตัดสินคือ ครูที่สอนวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ที่มีความเชี่ยวชาญในเนื้อหาคณิตศาสตร์ชั้นมัธยมศึกษาตอนต้น คือ จบการศึกษาระดับปริญญาตรีขึ้นไป สาขาทางคณิตศาสตร์ และมีความเชี่ยวชาญในการสอนและการวัดผล คือ มีประสบการณ์การสอนวิชาคณิตศาสตร์ชั้นมัธยมศึกษาตอนต้น มาไม่น้อยกว่า 5 ปี และ เป็นครูชำนาญการ หรือชำนาญการพิเศษ จำนวนทั้งสิ้น 12 คน ขั้นตอนต่อมานำผู้ตัดสินทั้งหมดมาร่วมสัมมนาเชิงปฏิบัติการเพื่อรับฟังวิธีการกำหนดมาตรฐานทั้งวิธีการแองกอฟที่ได้รับการปรับปรุงและวิธีการบุ๊คมาร์ค พร้อมทั้งลองฝึกปฏิบัติ

7.2 แบ่งกลุ่มผู้ตัดสินเป็นกลุ่มย่อย 2 กลุ่มย่อย กลุ่มย่อยละ 6 คน

7.3 การดำเนินงานคือให้ผู้ตัดสินทั้งหมดกำหนดมาตรฐาน ด้วยวิธีการ 2 วิธี ดังนี้

- วิธีกำหนดมาตรฐานแองกอฟที่ได้รับการปรับปรุง(Modified Angoff)
- วิธีกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์ค (Bookmark)

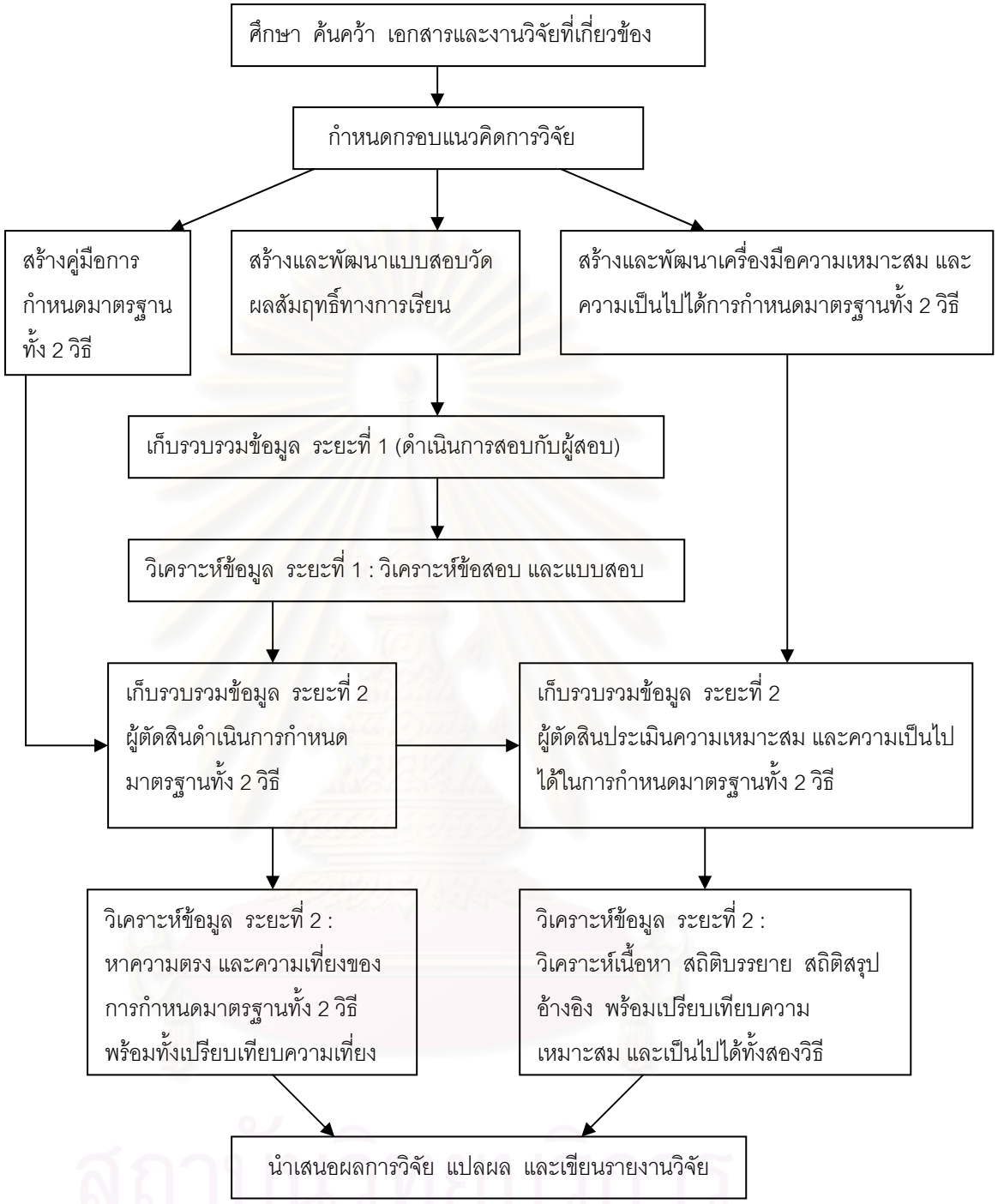
8 ดำเนินการวิเคราะห์ข้อมูล ระยะที่ 2 หาความตรงของการกำหนดมาตรฐาน และความเที่ยงของการกำหนดมาตรฐาน โดยใช้โปรแกรม GENOVA ในการวิเคราะห์

9 ศึกษาวิธีการกำหนดมาตรฐานวิธีแองกอฟที่ได้รับการปรับปรุง และวิธีบุ๊คมาร์คเกี่ยวกับความเหมาะสม และความเป็นไปได้ในการนำมาใช้ โดยผู้วิจัยแจกแบบสอบถามให้ผู้ตัดสินทุกคนทำพร้อมทั้งสัมภาษณ์ผู้ตัดสิน จากนั้นดำเนินการวิเคราะห์ข้อมูล โดยการวิเคราะห์เนื้อหา และคำนวณหาสถิติบรรยาย เช่น ความถี่ ร้อยละ ค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐาน และคำนวณหาสถิติสรุปอ้างอิง คือ สถิติทดสอบ t แบบ dependent

10 นำเสนอผลการวิจัย แปลผล และเขียนรายงานการวิจัย

ขั้นตอนการดำเนินงานวิจัยทั้งหมดข้างต้น เสนอเป็นแผนภาพที่ 5 ได้ดังนี้

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย



แผนภาพที่ 5 ขั้นตอนการดำเนินงานวิจัย

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ประชากรและกลุ่มตัวอย่าง

ประชากร

ในการวิจัยนี้ได้แบ่งประชากรออกเป็น 2 กลุ่ม คือ

1. **ผู้ตัดสิน** คือ ครูผู้สอนวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ในโรงเรียนสังกัดสำนักงานการศึกษาขั้นพื้นฐาน สำนักงานเขตพื้นที่การศึกษาชลบุรี เขต 1

2. **ผู้สอบ** คือ นักเรียนชั้นมัธยมศึกษาปีที่ 3 ปีการศึกษา 2549 ในโรงเรียนสังกัดการศึกษาขั้นพื้นฐานเขตพื้นที่การศึกษาชลบุรี เขต 1 จำนวน 35 โรง จำนวน 5,423 คน (http://doc.obec.go.th/web 2548)

กลุ่มตัวอย่าง

1. **ผู้ตัดสิน** คือ ครูสอนวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 สังกัดสำนักงานการศึกษาขั้นพื้นฐาน สำนักงานเขตพื้นที่การศึกษาชลบุรี เขต 1 ที่มีคุณสมบัติดังนี้

1.1 มีความเชี่ยวชาญในเนื้อหาคณิตศาสตร์มัธยมศึกษาตอนต้น คือ จบการศึกษาระดับปริญญาตรีขึ้นไปในสาขาทางคณิตศาสตร์

1.2 มีความเชี่ยวชาญในการสอน และการวัดผล ดังนี้

1.2.1 มีประสบการณ์การสอนมาไม่น้อยกว่า 5 ปี

1.2.2 เป็นครูชำนาญการ หรือชำนาญการพิเศษ

ผู้วิจัยใช้วิธีการสุ่มตัวอย่างอย่างง่ายเพื่อให้ได้กลุ่มตัวอย่างผู้ตัดสินจำนวน 12 คน จากนั้นผู้วิจัยทำการแบ่งผู้ตัดสินออกเป็น 2 กลุ่ม กลุ่มละ 6 คน แล้วให้ผู้ตัดสินทุกคนทำการกำหนดมาตรฐานทั้ง 2 วิธี คือ วิธีการกำหนดมาตรฐานแองกอฟที่ได้รับการปรับปรุง และวิธีการกำหนดมาตรฐานบุ๊คมาร์ค

2. **ผู้สอบ** คือ นักเรียนชั้นมัธยมศึกษาปีที่ 3 ปีการศึกษา 2549 ในโรงเรียนสังกัดสำนักงานการศึกษาขั้นพื้นฐานจังหวัดชลบุรี สำนักงานเขตพื้นที่การศึกษาชลบุรีเขต 1 โดยในการวิจัยนี้ทำการสุ่มตัวอย่างแบบหลายขั้นตอน (Multi-stage sampling) คือ

(1) **การทดลองแบบสอบครั้งที่ 1** เพื่อตรวจสอบความเป็นปรนัยของข้อสอบ ผู้วิจัยสุ่มตัวอย่างโรงเรียนโดยการสุ่มอย่างง่ายจำนวน 1 โรงเรียนคือ โรงเรียนชลกันยานุกูล แสนสุข จากนั้นสุ่มตัวอย่างห้องเรียนโดยใช้วิธีการสุ่มตัวอย่างอย่างง่ายได้ห้องเรียน 1 ห้อง รวมกลุ่มตัวอย่างนักเรียนทั้งสิ้น 10 คน

(2) **การทดลองแบบสอบครั้งที่** 2 เพื่อพัฒนาแบบสอบ ผู้วิจัยสุ่มตัวอย่างโรงเรียนโดยการสุ่มแบบแบ่งชั้นตามขนาดของโรงเรียนคือขนาดเล็ก กลาง และใหญ่ จำนวน 4 โรงเรียน จากนั้นสุ่มตัวอย่างห้องเรียนในแต่ละโรงเรียนที่สุ่มมาโดยใช้วิธีการสุ่มตัวอย่างอย่างง่าย ได้ห้องเรียนจำนวน 8 ห้องเรียน รวมกลุ่มตัวอย่างนักเรียนทั้งสิ้น 290 คน

(3) **การทดลองแบบสอบครั้งที่** 3 เพื่อพัฒนาแบบสอบและใช้เก็บข้อมูลจริง ผู้วิจัยสุ่มตัวอย่างโรงเรียนโดยการสุ่มแบบแบ่งชั้นตามขนาดของโรงเรียนคือขนาดเล็ก กลาง และใหญ่ จำนวน 8 โรงเรียน จากนั้นสุ่มตัวอย่างห้องเรียนในแต่ละโรงเรียนที่สุ่มมาโดยใช้วิธีการสุ่มตัวอย่างอย่างง่ายได้ห้องเรียนจำนวน 25 ห้องเรียน รวมกลุ่มตัวอย่างนักเรียนทั้งสิ้น 987 คน สำหรับการคำนวณขนาดกลุ่มตัวอย่างดังนี้

สูตรหาขนาดกลุ่มตัวอย่าง กรณีต้องการศึกษาค่าสัดส่วน ผู้วิจัยใช้สูตรของโคชแรน (Cochran, 1963 อ้างถึงใน ศิริชัย กาญจนวาสี, 2545)

$$n = \frac{NZ^2P(1-P)}{NE^2 + Z^2P(1-P)}$$

เมื่อ

n = ขนาดของกลุ่มตัวอย่าง

N = ขนาดของประชากร

Z = ค่าที่กำหนดจากค่าความเชื่อมั่นที่ผู้วิจัยต้องการจะใช้เพื่อการสรุปผล ซึ่งระบุเป็นร้อยละ โดยทั่วไปแล้ว เช่น ถ้าความเชื่อมั่น 99% จะเท่ากับ 2.58 หรือประมาณ 3

P = ค่าสัดส่วนของลักษณะที่ต้องการศึกษา

$P(1-P)$ = ค่าความแปรปรวนของลักษณะที่ต้องการศึกษา มีค่าสูงสุดเป็น 0.25

E = ค่าความคลาดเคลื่อนที่ผู้วิจัยยอมเสี่ยงในการสรุปผล

ในการวิจัยนี้ผู้วิจัยกำหนดระดับความเชื่อมั่น 99% ความคลาดเคลื่อน 5% และขนาดประชากร 5,423 คน

$$n = \frac{5423(3)^2(0.25)}{5423(0.05)^2 + (3)^2(0.25)}$$

$$= 771.90$$

$$\approx 772$$

นั้นคือ ในการวิจัยครั้งนี้ควรใช้ขนาดกลุ่มตัวอย่างอย่างน้อย 772 คน สำหรับการวิจัยครั้งนี้ใช้กลุ่มตัวอย่าง 987 คน โดยแบ่งตามขนาดโรงเรียน คือ โรงเรียนขนาดเล็ก (นักเรียน 1-500 คน) โรงเรียนขนาดกลาง (นักเรียน 501-1499 คน) และโรงเรียนขนาดใหญ่ (นักเรียน 1500 คนขึ้นไป) ดังตารางที่ 4

ตารางที่ 4 รายชื่อโรงเรียนและจำนวนกลุ่มตัวอย่างนักเรียนชั้นมัธยมศึกษาปีที่ 3 ที่ใช้เก็บข้อมูลเพื่อทดลองแบบสอบครั้งที่ 1 ครั้งที่ 2 และครั้งที่ 3

โรงเรียน	ประชากร		กลุ่มตัวอย่าง	
	จำนวนห้องเรียน	จำนวนนักเรียน	จำนวนห้องเรียน	จำนวนนักเรียน
<u>การทดลองแบบสอบครั้งที่ 1</u>				
1) โรงเรียนชลกันยานุกูล แสนสุข	7	322	1	10
รวม	7	322	1	10

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 4 (ต่อ)

โรงเรียน	ประชากร		กลุ่มตัวอย่าง	
	จำนวนห้องเรียน	จำนวนนักเรียน	จำนวนห้องเรียน	จำนวนนักเรียน
<u>การทดลองแบบสอบครั้งที่ 2</u>				
โรงเรียนขนาดเล็ก				
1) วอนนภาศัพท์	1	29	1	20
2) วัดเขาเชิงเทียนเทพาราม	1	41	1	35
โรงเรียนขนาดกลาง				
3) วัดราษฎร์ศรัทธา	3	104	2	63
โรงเรียนขนาดใหญ่				
4) ชลบุรี(สุขบท)	14	549	5	172
รวม	19	723	9	290
<u>การทดลองแบบสอบครั้งที่ 3</u>				
โรงเรียนขนาดเล็ก				
1) หนองใหญ่ศิริวรวาทวิทยา	2	108	2	54
2) หนองรีมงคลสุขสวัสดิ์	2	123	2	69
โรงเรียนขนาดกลาง				
3) บ้านสวนอุดม	2	93	1	51
4) ชลราษฎรอำรุง 2	8	303	5	194
5) ชลกันยานุกูล แสนสุข	7	322	3	84
โรงเรียนขนาดใหญ่				
6) ชลกันยานุกูล	16	746	5	178
7) ชลราษฎรอำรุง	13	584	4	211
8) บ้านสวน (จั่น อนุสรณ์)	12	479	4	146
รวม	54	2,758	26	987

เครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้ในการวิจัยนี้มีดังนี้

1. แบบสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ภาคเรียนที่ 1

2. เครื่องมือเกี่ยวกับความเหมาะสมและความเป็นไปได้ในการกำหนดมาตรฐานด้วยวิธีการบุ๊คมาร์คประกอบด้วย แบบประเมิน แบบสัมภาษณ์

3. คู่มือวิธีการกำหนดมาตรฐานวิธีแองกอฟที่ได้รับการปรับปรุง และวิธีการบุ๊คมาร์ค

1. แบบสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์

แบบสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ภาคเรียนที่ 1 เป็นแบบสอบที่ผู้วิจัยได้ดำเนินการสร้างและพัฒนาขึ้นมาเอง แบบสอบเป็นแบบเลือกตอบ 4 ตัวเลือก จำนวน 4 ฉบับ คือ ฉบับที่ 1 พื้นที่และปริมาตร ฉบับที่ 2 กราฟ ฉบับที่ 3 ระบบสมการเชิงเส้น และฉบับที่ 4 ความคล้าย สำหรับเหตุผลที่ผู้วิจัยเลือกสร้างแบบสอบในวิชาคณิตศาสตร์ เนื่องจากว่า วิชาคณิตศาสตร์เป็นวิชาที่มีความสำคัญสำหรับนักเรียนในการดำเนินชีวิตประจำวัน และเป็นวิชาที่มีความสัมพันธ์กับกระบวนการคิดขั้นพื้นฐานทางการศึกษาต่อวิชาอื่นๆ นอกจากนี้ยังเป็นวิชาที่ฝึกให้ผู้เรียนคิดอย่างเป็นระบบอีกด้วย

สำหรับขั้นตอนการสร้างและพัฒนาแบบสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 มีลำดับขั้นดังนี้

1) ศึกษาหลักสูตร คู่มือครู และเนื้อหาในหนังสือแบบเรียนวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ภาคเรียนที่ 1

2) วิเคราะห์วัตถุประสงค์การเรียนรู้เป็นรายข้อ และเนื้อหาวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 เพื่อสร้างตารางโครงสร้างเนื้อหา จากนั้นนำตารางโครงสร้างเนื้อหาไปให้ครูผู้สอนวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ที่มีประสบการณ์การสอนไม่ต่ำกว่า 3 ปี จำนวน 3 คน กำหนดน้ำหนักเนื้อหา และพิจารณาความครอบคลุมของเนื้อหา ตลอดจนความเหมาะสมของวัตถุประสงค์การเรียนรู้ ตัวอย่างแบบประเมินความครอบคลุมของวัตถุประสงค์การเรียนรู้ในขอบเขตเนื้อหา และแบบประเมินการกำหนดน้ำหนักความสำคัญของแต่ละจุดประสงค์การเรียนรู้ดังต่อไปนี้

ตัวอย่างแบบประเมินความครอบคลุมของวัตถุประสงค์การเรียนรู้ในขอบเขตเนื้อหา

<u>คำชี้แจง</u> ขอให้ท่านพิจารณาว่าวัตถุประสงค์เหล่านี้วัดได้ครอบคลุมในขอบเขตเนื้อหานั้นหรือไม่ โดยพิจารณาตามเกณฑ์ดังต่อไปนี้

1 หมายถึง แน่ใจว่าวัตถุประสงค์นี้วัดได้ครอบคลุมขอบเขตเนื้อหานั้นจริง

0 หมายถึง ไม่แน่ใจว่าวัตถุประสงค์นี้วัดได้ครอบคลุมขอบเขตเนื้อหานั้นจริง

-1 หมายถึง แน่ใจว่าวัตถุประสงค์นี้วัดได้ไม่ครอบคลุมขอบเขตเนื้อหานั้นจริง

วัตถุประสงค์	คะแนนการตัดสิน			ข้อเสนอแนะ
	- 1	0	1	
พื้นที่ผิวและปริมาตร **<u>รูปเรขาคณิตสามมิติ</u>** 1) นักเรียนสามารถบอกชื่อปริซึม และพีระมิด ชนิดต่างๆ ได้				

ตัวอย่างแบบประเมิน การกำหนดน้ำหนักความสำคัญของแต่ละจุดประสงค์การเรียนรู้

<u>คำชี้แจง</u>

ขอให้ท่านกำหนดน้ำหนักความสำคัญของแต่ละวัตถุประสงค์การเรียนรู้วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ภาคเรียนที่ 1 ลงบนช่องว่างในตารางที่กำหนดให้ โดยให้คะแนนเต็มของแต่ละวัตถุประสงค์เป็น 10 คะแนน และให้มีหลักเกณฑ์การประเมินดังต่อไปนี้

ถ้าเห็นว่า วัตถุประสงค์นี้มีความสำคัญมากให้คะแนน 7- 10 คะแนน

ถ้าเห็นว่า วัตถุประสงค์นี้มีความสำคัญปานกลางให้คะแนน 4- 6 คะแนน

ถ้าเห็นว่า วัตถุประสงค์นี้มีความสำคัญน้อยให้คะแนน 1- 3 คะแนน

พร้อมกันนี้ ผู้วิจัยได้นำข้อมูลจำนวนคาบเรียนในแต่ละเนื้อหาจากคู่มือครู มาให้ท่านประกอบการพิจารณาด้วย

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

เนื้อหา/ วัตถุประสงค์การเรียนรู้	จำนวนคาบ	น้ำหนักความสำคัญ
1. พื้นที่ผิวและปริมาตร	(23)	
1.1 รูปเรขาคณิตสามมิติ	2	
1) นักเรียนสามารถบอกชื่อปริซึม และพีระมิดชนิดต่างๆ ได้		_____

ผลการกำหนดความครอบคลุมของวัตถุประสงค์การเรียนรู้ในขอบเขตเนื้อหา ของผู้เชี่ยวชาญทั้ง 3 ท่านเมื่อนำมาเฉลี่ย ปรากฏว่าวัตถุประสงค์ทุกข้อวัดได้ครอบคลุมเนื้อหา (ค่าเฉลี่ยคะแนนตัดสินเท่ากับ 1 ทุกข้อ) นอกจากนี้ผู้วิจัยยังได้ปรับวัตถุประสงค์ตามคำแนะนำของผู้เชี่ยวชาญ เช่น

นักเรียนสามารถหาคำตอบของระบบสมการเชิงเส้นสองตัวแปร แก้ไขเป็น นักเรียนสามารถหาคำตอบของระบบสมการเชิงเส้นสองตัวแปร**โดยวิธีการแทนค่าและขจัดตัวแปร**

นักเรียนสามารถบอกนิยามและทฤษฎีบทของรูปสามเหลี่ยมสองรูปที่คล้ายกัน แก้ไขเป็น นักเรียนสามารถบอก**สมบัติของการคล้ายกันของรูปสามเหลี่ยมและบอกเงื่อนไขที่ทำให้รูปสามเหลี่ยมสองรูปคล้ายกันได้**

สำหรับตารางโครงสร้างเนื้อหา ผู้เชี่ยวชาญได้กำหนดน้ำหนักความสำคัญของเนื้อหาของแต่ละจุดประสงค์การเรียนรู้ เมื่อผู้วิจัยนำมาเฉลี่ย ปรากฏดังตารางที่ 5

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 5 ตารางโครงสร้างกำหนดน้ำหนักความสำคัญของเนื้อหาและวัตถุประสงค์การเรียนรู้ของวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ภาคเรียนที่ 1

เนื้อหา	น้ำหนักเฉลี่ย	จำนวนข้อ	เลขที่ข้อ
แบบสอบฉบับที่ 1 พื้นที่ผิวและปริมาตร			
1.1รูปเรขาคณิตสามมิติ	(13)	(4)	
1) นักเรียนสามารถบอกชื่อปริซึม และพีระมิดชนิดต่างๆ ได้	6	2	1-2
2) นักเรียนสามารถบอกส่วนต่างๆของปริซึม ทรงกระบอก พีระมิด กรวยกลม และทรงกลมได้	7	2	3-4
1.2 ปริมาตรของปริซึมและทรงกระบอก	(15.3)	(6)	
1) นักเรียนสามารถหาปริมาตรของปริซึมได้อย่างถูกต้อง	7.3	3	5-7
2) นักเรียนสามารถหาปริมาตรของทรงกระบอก	8	3	8-10
1.3 ปริมาตรของพีระมิดและกรวย	(17.4)	(6)	
1) นักเรียนสามารถหาปริมาตรของพีระมิด	8.7	3	11-13
2) นักเรียนสามารถหาปริมาตรทรงกรวยกลม	8.7	3	14-16
1.4 ปริมาตรของทรงกลม	(7.3)	(2)	
1) นักเรียนสามารถหาปริมาตรของทรงกลม	7.3	2	17-18
1.5 พื้นที่ผิวของปริซึมและทรงกระบอก	(15.7)	(6)	
1)) นักเรียนสามารถหาพื้นที่ผิวของปริซึมได้อย่างถูกต้อง	7.7	3	19-21
2) นักเรียนสามารถหาพื้นที่ผิวของทรงกระบอก	8	3	22-24
1.6 การเปรียบเทียบหน่วยปริมาตร	(15)	(6)	
1) เปรียบเทียบหน่วยความจุ หรือหน่วยปริมาตรในระบบเดียวกัน หรือต่างระบบได้	8	3	25-27
2) นักเรียนสามารถใช้หน่วยวัดเกี่ยวกับความจุ หรือปริมาตรและคาดคะเนเกี่ยวกับการวัดในสถานการณ์ต่างๆ ได้	7	3	28-30
รวม	**83.7**	**30**	

ตารางที่ 5 (ต่อ)

เนื้อหา	น้ำหนักเฉลี่ย	จำนวนข้อ	เลขที่ข้อ
แบบสอบฉบับที่ 2. กราฟ			
2.1 กราฟแสดงความสัมพันธ์ระหว่างปริมาณที่มีความสัมพันธ์เชิงเส้น	(23.3)	(9)	
1) นักเรียนสามารถเขียนกราฟแสดงความสัมพันธ์ระหว่างปริมาณสองปริมาณที่กำหนดให้ได้	7.3	3	31-33
2) นักเรียนสามารถบอกได้ว่าจุดที่แทนคู่อันดับที่กำหนดให้อยู่บนกราฟเส้นตรงที่กำหนดให้หรือไม่	7	2	34-35
3) นักเรียนสามารถแปลความหมายของกราฟเส้นตรงที่กำหนดให้	9	4	36-39
2.2 กราฟของสมการเชิงเส้นสองตัวแปร	(48.3)	(18)	
1) นักเรียนสามารถเขียนกราฟของสมการเชิงเส้นสองตัวแปร	8	3	40-42
2) เมื่อกำหนดกราฟเส้นตรงนักเรียนสามารถเขียนสมการเชิงเส้นได้	7.6	3	43-45
3) นักเรียนสามารถหาสมาชิกที่หนึ่ง หรือสมาชิกตัวที่สองของคู่อันดับจากกราฟเส้นตรงได้ เมื่อกำหนดสมาชิกตัวใดตัวหนึ่งได้	8.3	3	46-48
4) นักเรียนสามารถบอกได้ว่ากราฟของสมการเชิงเส้นสองตัวแปรคู่หนึ่งตัดกันหรือขนานกัน	8	3	49-51
5) นักเรียนสามารถหาจุดตัดของกราฟ จากกราฟของสมการเชิงเส้นสองตัวแปรสองสมการได้	8.7	3	52-54
6) นักเรียนสามารถบอกลักษณะมุมที่กราฟของสมการเชิงเส้นทำกับแกน X	7.7	3	55-57
2.3 กราฟกับการนำไปใช้	(8.7)	(3)	
1) นักเรียนสามารถอ่านค่าต่างๆ จากกราฟอื่นๆ ที่ไม่เป็นเส้นตรงที่กำหนดให้	8.7	3	58-60
รวม	**80.3**	**30**	

ตารางที่ 5 (ต่อ)

เนื้อหา	น้ำหนักเฉลี่ย	จำนวนข้อ	เลขที่ข้อ
แบบสอบฉบับที่ 3 ระบบสมการเชิงเส้น			
3.1 ระบบสมการเชิงเส้นสองตัวแปร	(16)	(6)	
1) นักเรียนสามารถบอกได้ว่าจากกราฟของระบบสมการเชิงเส้นสองตัวแปรมีคำตอบเดียว มีหลายคำตอบหรือไม่มีคำตอบ	8	3	61-63
2) นักเรียนสามารถเขียนกราฟสมการเชิงเส้นสองตัวแปร	8	3	64-66
3.2 การแก้ระบบสมการเชิงเส้นสองตัวแปรและกราฟ	(16)	(6)	
1) นักเรียนสามารถหาคำตอบของระบบสมการเชิงเส้นสองตัวแปร	8.7	3	67-69
2) นักเรียนสามารถหาคำตอบของระบบสมการเชิงเส้นจากกราฟ	7.3	3	70-72
3.3 โจทย์สมการเชิงเส้นสองตัวแปร	(9.3)	(4)	
1) นักเรียนสามารถใช้ระบบสมการเชิงเส้นสองตัวแปรแก้โจทย์ปัญหา	9.3	4	73-76
รวม	**41.3**	**16**	
แบบสอบฉบับที่ 4 ความคล้าย			
4.1 รูปเรขาคณิตที่คล้ายกัน	(30.3)	(9)	
1) นักเรียนสามารถบอกนิยามของรูปหลายเหลี่ยมสองรูปที่คล้ายกันได้	7.3	2	77-78
2) นักเรียนสามารถบอกได้ว่ารูปหลายเหลี่ยมสองรูปที่กำหนดให้คล้ายกันหรือไม่	7.3	3	79-81
3) นักเรียนสามารถนำบทนิยามรูปหลายเหลี่ยมสองรูปที่คล้ายกันไปใช้หาขนาดของมุมที่เหลือของรูปหลายเหลี่ยมได้	8.7	1	82
4) นักเรียนสามารถนำบทนิยามรูปหลายเหลี่ยมสองรูปที่คล้ายกันไปใช้หาความยาวของด้านที่เหลือของรูปหลายเหลี่ยมได้	7	3	83-85

ตารางที่ 5 (ต่อ)

เนื้อหา	น้ำหนักเฉลี่ย	จำนวนข้อ	เลขที่ข้อ
4.2 รูปสามเหลี่ยมที่คล้ายกัน	(33)	(12)	
1) นักเรียนสามารถบอกสมบัติของการคล้ายกันของรูปสามเหลี่ยมและบอกเงื่อนไขที่ทำให้รูปสามเหลี่ยมสองรูปคล้ายกันได้	8	2	86-87
2) นักเรียนสามารถบอกได้ว่ารูปสามเหลี่ยมสองรูปที่กำหนดให้คล้ายกันหรือไม่	7.7	3	88-90
3) นักเรียนสามารถนำบทนิยาม และทฤษฎีของรูปสามเหลี่ยมที่คล้ายกันไปใช้หาขนาดของมุมที่เหลือของรูปสามเหลี่ยมได้	9.3	4	91-94
4) นักเรียนสามารถนำบทนิยามรูปสามเหลี่ยมสองรูปที่คล้ายกันไปใช้หาความยาวของด้านที่เหลือของรูปสามเหลี่ยมได้	8	3	95-97
4.3 การนำไปใช้	(9)	(3)	
1) นักเรียนสามารถนำสมบัติของรูปสามเหลี่ยมที่คล้ายกันไปใช้แก้โจทย์ปัญหา	9	3	98-100
รวม	72.3	24	

3) สร้างข้อสอบแบบเลือกคำตอบ 4 ตัวเลือกให้สอดคล้องกับตารางโครงสร้างกำหนดน้ำหนักเนื้อหา จำนวน 4 ฉบับ คือ ฉบับที่ 1 พื้นที่และปริมาตร ฉบับที่ 2 กราฟ ฉบับที่ 3 ระบบสมการเชิงเส้น และฉบับที่ 4 ความคล้าย ตามระดับพุทธิพิสัยของแนวคิดของวิลสันที่มี 4 ระดับ คือ ความรู้ความจำเกี่ยวกับการคิดคำนวณ ความรู้ความจำเกี่ยวกับศัพท์และนิยาม ความรู้ความจำเกี่ยวกับการใช้กระบวนการคิดคำนวณ (พร้อมพรรณ อุดมสิน, 2544)

4) ตรวจสอบความตรงเชิงเนื้อหาของแบบสอบ โดยผู้วิจัยนำแบบสอบ ทั้ง 4 ฉบับมาให้ผู้เชี่ยวชาญจำนวน 5 คน ที่มีประสบการณ์ในการสอนมาไม่ต่ำกว่า 3 ปี พิจารณาตัดสินความสอดคล้องระหว่างข้อสอบและจุดประสงค์การเรียนรู้ โดยพิจารณาตามเกณฑ์ดังนี้

+1 หมายถึง แน่ใจว่าข้อสอบข้อนี้สอดคล้องกับวัตถุประสงค์การเรียนรู้

0 หมายถึง ไม่แน่ใจว่าข้อสอบข้อนี้สอดคล้องกับวัตถุประสงค์การเรียนรู้

-1 หมายถึง แน่ใจว่าข้อสอบข้อนี้ไม่สอดคล้องกับวัตถุประสงค์การเรียนรู้

เมื่อผู้เชี่ยวชาญพิจารณาตัดสินความสอดคล้องระหว่างข้อสอบและวัตถุประสงค์การเรียนรู้เรียบร้อยแล้ว นำผลประเมินมาหาค่าดัชนีความสอดคล้องตามวิธีของโรวิเนลลีและแฮมเบิลตัน (Rovinelli and Hambleton, 1977, cited in Hambleton, 1978) แบบประเมินความสอดคล้องระหว่างข้อสอบกับจุดประสงค์การเรียนรู้ มีตัวอย่างดังต่อไปนี้

แบบประเมินความสอดคล้องระหว่างข้อสอบกับจุดประสงค์การเรียนรู้

คำชี้แจง ขอให้ท่านพิจารณาว่าวัตถุประสงค์เหล่านี้วัดได้ครอบคลุมในขอบเขตเนื้อหานั้นหรือไม่ โดยพิจารณาตามเกณฑ์ดังต่อไปนี้

1 หมายถึง แน่ใจว่าข้อสอบเหล่านี้วัดได้สอดคล้องกับจุดประสงค์นั้นจริง

0 หมายถึง ไม่แน่ใจว่าข้อสอบเหล่านี้วัดได้สอดคล้องกับจุดประสงค์นั้นจริง

-1 หมายถึง แน่ใจว่าข้อสอบเหล่านี้วัดได้สอดคล้องกับจุดประสงค์นั้นจริง

วัตถุประสงค์	คะแนนการตัดสิน			ข้อเสนอแนะ
	- 1	0	1	
พื้นที่ผิวและปริมาตร **รูปเรขาคณิตสามมิติ** 1) นักเรียนสามารถบอกชื่อปริซึม และพีระมิดชนิดต่างๆ ได้				
2) นักเรียนสามารถบอกส่วนต่างๆ ของปริซึม ทรงกระบอก พีระมิด กรวย และทรงกลม				
ปริมาตรของปริซึมและทรงกระบอก 3) นักเรียนสามารถหาปริมาตรของปริซึมได้อย่างแม่นยำและรวดเร็ว				
4) นักเรียนสามารถหาปริมาตรของทรงกระบอก				

หลังจากนั้นผู้วิจัยนำคะแนนความคิดเห็นของผู้เชี่ยวชาญมาคำนวณหาค่าดัชนี IOC รายละเอียดดังตารางที่ 6

ตารางที่ 6 ผลการตรวจสอบความสอดคล้องระหว่างข้อสอบกับจุดประสงค์การเรียนรู้

ข้อที่	IOC	ข้อที่	IOC	ข้อที่	IOC	ข้อที่	IOC	ข้อที่	IOC
1	1.0	21	1.0	41	1.0	61	1.0	81	1.0
2	0.8	22	1.0	42	1.0	62	0.6	82	1.0
3	0.6	23	1.0	43	1.0	63	0.4*	83	1.0
4	1.0	24	-0.2*	44	0.8	64	0.6	84	1.0
5	1.0	25	1.0	45	0.6	65	0.8	85	1.0
6	1.0	26	1.0	46	1.0	66	1.0	86	1.0
7	1.0	27	1.0	47	0.2*	67	1.0	87	1.0
8	0.6	28	1.0	48	0.8	68	1.0	88	1.0
9	1.0	29	1.0	49	0.8	69	1.0	89	1.0
10	0.4	30	1.0	50	1.0	70	0.6	90	1.0
11	0.8	31	0.8	51	1.0	71	0.2*	91	1.0
12	1.0	32	0.6	52	1.0	72	1.0	92	1.0
13	1.0	33	1.0	53	1.0	73	0.8	93	1.0
14	1.0	34	1.0	54	1.0	74	1.0	94	1.0
15	0.2*	35	0.8	55	0.2*	75	1.0	95	1.0
16	0.6	36	1.0	56	1.0	76	1.0	96	0.8
17	0.6	37	1.0	57	1.0	77	1.0	97	1.0
18	0.8	38	1.0	58	1.0	78	1.0	98	1.0
19	1.0	39	1.0	59	1.0	79	1.0	99	1.0
20	1.0	40	1.0	60	1.0	80	1.0	100	1.0

* ข้อสอบที่มีค่า IOC ต่ำกว่า 0.5

หมายเหตุ แบบสอบฉบับที่ 1 คือ ข้อ 1 – 30 แบบสอบฉบับที่ 2 คือ ข้อ 31 – 60
แบบสอบฉบับที่ 3 คือ ข้อ 61 – 76 แบบสอบฉบับที่ 4 คือ ข้อ 77 – 100

จากตารางที่ 6 ค่าดัชนีความสอดคล้อง (IOC) อยู่ในช่วง 0.2 – 1.0 โดยมีข้อสอบ 7 ข้อ มีค่า IOC ต่ำกว่า 0.5 คือ ข้อที่ 10,15, 24, 47, 55, 63 และ 71 ผู้วิจัยจึงดำเนินการปรับปรุง

ข้อสอบตามคำแนะนำของผู้เชี่ยวชาญ ส่วนข้อสอบที่เหลืออีก 93 ข้อ มีค่าดัชนีความสอดคล้องสูงกว่า 0.5 แสดงว่าผู้เชี่ยวชาญมีความคิดเห็นและพิจารณาตัดสินแล้วว่าข้อสอบทั้งหมดนั้นวัดได้ตรงตามวัตถุประสงค์

5) การนำแบบสอบไปทดลองใช้เพื่อตรวจสอบคุณภาพของแบบสอบ ผู้วิจัยดำเนินการทดลองใช้แบบสอบ 3 ครั้ง ดังนี้

5.1 **การทดลองใช้แบบสอบครั้งที่** 1 เพื่อตรวจสอบด้านภาษาเกี่ยวกับคำชี้แจงในการสอบ และข้อคำถาม ว่านักเรียนมีความเข้าใจตรงกันหรือไม่ มีภาษาตรงจุดไหนมีความกำกวมเพื่อนำไปปรับปรุงแก้ไข ตลอดจนเพื่อนำข้อมูลที่ได้มากำหนดเวลาในการสอบที่มีความเหมาะสม สำหรับการดำเนินงานมีดังนี้

(1) ผู้วิจัยสุ่มตัวอย่างอย่างง่ายนักเรียนชั้นมัธยมศึกษาปีที่ 3 ปีการศึกษา 2549 ของโรงเรียนชลกันยานุกูล แสนสุข จำนวน 10 คน

(2) นำนักเรียนมาทำแบบสอบโดยจับเวลาในการทำข้อสอบของนักเรียนแต่ละคน เพื่อนำเวลาที่นักเรียนแต่ละคนทำมาหาค่าเฉลี่ยในการทำแบบสอบ

(3) เมื่อนักเรียนทำแบบสอบเสร็จ ผู้วิจัยสัมภาษณ์นักเรียนแต่ละคนเกี่ยวกับภาษาที่ใช้ในแบบทดสอบว่ามีคำสั่ง และข้อคำถามชัดเจน เข้าใจตรงกันหรือไม่

(4) ปรับปรุงภาษาที่ใช้ในแบบทดสอบ

ผลการทดลองใช้แบบสอบครั้งที่ 1 ปรากฏว่า นักเรียนมีความเข้าใจในคำชี้แจงและข้อคำถามแต่ละข้อดี โดยเฉลี่ยนักเรียนใช้เวลาทำแบบสอบ 4 ฉบับ รวม 100 ข้อ ดังนี้ ฉบับที่ 1 พื้นที่ผิวและปริมาตร 50 นาที ฉบับที่ 2 กราฟ 50 นาที ฉบับที่ 3 ระบบสมการเชิงเส้น 30 นาที และฉบับที่ 4 ความคล้าย 50 นาที

5.2 **การทดลองใช้แบบสอบครั้งที่** 2 เพื่อนำผลการสอบไปคำนวณเพื่อหาคุณภาพข้อสอบ และแบบสอบ นั้นคือ ผู้วิจัยคำนวณหาค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน ความเบ้ ความโด่ง ของคะแนนสอบ และหาคุณภาพข้อสอบคือ ค่าความยาก ค่าอำนาจจำแนก และหาคุณภาพแบบสอบคือ ความเที่ยงแบบสอดคล้องภายในวิธีคูเดอร์-ริชาร์ดสัน (KR-20) วิเคราะห์โดยใช้โปรแกรม CTIA

ผลการทดลองใช้ในครั้งที่ 2 ปรากฏค่าสถิติ ดังต่อไปนี้

(1) **ค่าสถิติพื้นฐาน** มีดังต่อไปนี้

ตารางที่ 7 ค่าสถิติพื้นฐานของแบบสอบผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ภาคเรียนที่ 1 จำนวน 4 ฉบับ รวม 100 ข้อ

ค่าสถิติพื้นฐาน	ฉบับที่ 1 พื้นที่ผิวและปริมาตร	ฉบับที่ 2 กราฟ	ฉบับที่3 ระบบสมการเชิงเส้น	ฉบับที่ 4 ความคล้าย
คะแนนเต็ม	30	30	16	24
คะแนนเฉลี่ย	9.61	9.67	5.53	7.27
ส่วนเบี่ยงเบนมาตรฐาน	3.13	3.86	2.15	3.19
ความเบ้	-.08	-.14	.50	.32
ความโด่ง	-.05	1.26	.25	.45
คะแนนสูงสุด	19	21	12	16
คะแนนต่ำสุด	4	3	1	2

จากตารางที่ 7 พบว่า ผู้สอบส่วนใหญ่ได้คะแนนสอบฉบับที่ 1 ฉบับที่ 2 ฉบับที่ 3 และฉบับที่ 4 ต่ำกว่าครึ่งหนึ่งของคะแนนเต็ม ค่าเฉลี่ยคือ 9.61, 9.67, 5.53 และ 7.27 ตามลำดับ และเมื่อพิจารณาค่าความเบ้ พบว่า ฉบับที่ 1 และ 2 ติดค่าลบ หรือโค้งเบ้ซ้าย ฉบับที่ 3 และ 4 มีค่ามากกว่า 0 หรือโค้งเบ้ขวา

(2) ค่าความยาก ค่าอำนาจจำแนก และค่าความเที่ยง

เมื่อพิจารณาค่าความยาก อำนาจจำแนก และค่าความเที่ยง ของแบบสอบทั้ง 4 ฉบับมีรายละเอียดดังนี้

ตารางที่ 8 ค่าความยาก และอำนาจจำแนกของแบบสอบผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 จำนวน 4 ฉบับ รวม 100 ข้อ

แบบสอบเลือกตอบ	จำนวนข้อ	ค่าอำนาจจำแนกเฉลี่ย ($\bar{r}$)	ค่าความยากเฉลี่ย ($\bar{p}$)	ความเที่ยง (KR-20)
ฉบับที่ 1	30	0.192	0.328	.613
ฉบับที่ 2	30	0.241	0.321	.649
ฉบับที่ 3	16	0.203	0.304	.287
ฉบับที่ 4	24	0.211	0.283	.557

จากตารางที่ 8 แสดงว่า แบบสอบทั้ง 4 ฉบับ มีค่าความยากง่ายเฉลี่ยค่อนข้างยาก แบบสอบฉบับที่ 2 ฉบับที่ 3 และฉบับที่ 4 มีค่าอำนาจจำแนกเฉลี่ยพอใช้ แต่แบบสอบฉบับที่ 1 มีค่าอำนาจจำแนกต่ำ แบบสอบฉบับที่ 1 ฉบับที่ 2 และฉบับที่ 4 มีค่าความเที่ยงระดับปานกลาง คือ.613 , .649 และ .557 ตามลำดับ แบบสอบฉบับที่ 3 มีค่าความเที่ยงต่ำ คือ .287

สำหรับเกณฑ์ที่ใช้คัดเลือกข้อสอบที่มีคุณภาพดีคือ ค่าความยากอยู่ระหว่าง 0.20 – 0.80 และค่าอำนาจจำแนกตั้งแต่ 0.20 ขึ้นไป (ศิริชัย กาญจนวาสี, 2544) ส่วนข้อสอบที่ไม่เป็นไปตามเกณฑ์ดังกล่าว ผู้วิจัยได้ดำเนินการปรับปรุงข้อสอบ เช่น ข้อสอบที่ยากมาก ผู้วิจัยจะปรับปรุงข้อสอบข้อนั้นให้ง่ายขึ้นโดยการลดความซับซ้อนของโจทย์ เป็นต้น สำหรับข้อสอบที่ไม่เป็นไปตามเกณฑ์ดังกล่าว ดังตารางที่ 9

ตารางที่ 9 ข้อสอบที่ไม่มีคุณภาพตามเกณฑ์ที่กำหนด ต้องได้รับการปรับปรุงข้อสอบ

แบบสอบ เลือกตอบ	เลขที่ข้อ แต่ละฉบับ	ข้อสอบที่ต้องปรับปรุง
ฉบับที่ 1	1-30	ข้อที่ 1,3,6,7,10,12,13,14,15,16,17,19,20,21,22,23,24,29
ฉบับที่ 2	31-60	ข้อที่ 31,32,35,36,37,39,42,43,4752,54,56,57,59
ฉบับที่ 3	61-76	-
ฉบับที่ 4	77-100	ข้อที่ 77,79,80,82,83,84,87,89,90,96,98,100

เนื่องจากในการทดลองครั้งที่ 3 วิเคราะห์ผลการสอบโดยใช้โมเดลการตอบสนองข้อสอบแบบ 3 พารามิเตอร์ โดยมีข้อตกลงเบื้องต้นว่า แบบสอบมุ่งวัดคุณลักษณะเดียว (unidimensionality) และมีความเป็นอิสระหว่างข้อสอบ (independence) ดังนั้นจึงต้องทำการตรวจสอบก่อนการทดลองใช้แบบสอบครั้งที่ 3 โดยการวิเคราะห์องค์ประกอบ (factor analysis) ด้วยวิธีการวิเคราะห์องค์ประกอบหลัก (principal components analysis) และหมุนแกนแบบ varimax เพื่อหาค่าไอเกน (eigen Value) วิเคราะห์โดยใช้โปรแกรม SPSS

การพิจารณาค่าไอเกนจะพิจารณาเฉพาะค่าไอเกนมากกว่า 1 ซึ่งจะถือเป็น 1 องค์ประกอบ การที่ต้องมีค่าไอเกนมากกว่า 1 หรือ องค์ประกอบหนึ่งต้องอธิบายได้มากกว่า 1 ตัวแปร การอธิบายได้มากกว่าการผันแปรของ 1 ตัวแปร หมายความว่า จะต้องอธิบายได้มากกว่า 1 Eigen Value เพราะ 1 Eigen Value คือ ค่าการผันแปรมาตรฐานของ 1 ตัวแปร (สุชาติ ประสิทธิ์รัฐสินธุ, 2545) ดังนั้นจะพิจารณาแบบสอบที่สร้างขึ้นมุ่งวัดคุณลักษณะเดียวโดยพิจารณาว่ามีองค์ประกอบหนึ่งมีค่าไอเกนสูงมากๆ และสูงกว่า ค่าไอเกนขององค์ประกอบอื่นๆ

อย่างมาก นั้นคือแบบสอบนั้นมุ่งวัดเพียงองค์ประกอบเดียวนั้นเอง โดยมีเกณฑ์ของ Rackase (Raju, 1993) เสนอแนะว่า การใช้ทฤษฎีการตอบสนองข้อสอบเพื่อวิเคราะห์ข้อสอบนั้น ค่าความแปรปรวนจากตัวประกอบหลักตัวแรกของแบบสอบควรมีค่าไม่ต่ำกว่าร้อยละ 20 ผลการวิเคราะห์ความเป็นมิติเดียว ดังตารางที่ 10 ถึง 13

ตารางที่ 10 ผลการวิเคราะห์ตัวประกอบของแบบสอบเลือกตอบฉบับที่ 1 เรื่อง พื้นที่ผิวและปริมาตร

ตัวประกอบ	ค่าไอเกน	ค่าร้อยละของความแปรปรวน
1	8.884	29.614
2	1.765	5.885
3	1.200	3.999
4	1.010	3.3661

จากตารางที่ 10 พบว่า ตัวประกอบที่มีค่าไอเกนมากกว่า 1.00 อยู่ 4 ตัวประกอบ โดยตัวประกอบที่ 1 มีค่าไอเกนสูงสุดคือ 8.884 ค่าความแปรปรวนร้อยละ 29.614 ตัวประกอบที่ 2 มีค่า 1.765 ค่าความแปรปรวนร้อยละ 5.885 จะเห็นว่า ค่าความแปรปรวนจากตัวประกอบหลักตัวแรกของแบบสอบมีค่าสูงกว่าร้อยละ 20 แสดงว่าแบบสอบมุ่งวัดคุณลักษณะเด่นเพียงลักษณะเดียว ซึ่งเป็นไปตามข้อตกลงเบื้องต้นของทฤษฎีตอบสนองข้อสอบว่า แบบสอบมุ่งวัดคุณลักษณะเดียว (unidimensionality) นอกจากนี้ยังทำให้แบบสอบฉบับที่ 1 มีคุณสมบัติความเป็นอิสระของข้อสอบด้วย

ตารางที่ 11 ผลการวิเคราะห์ตัวประกอบของแบบสอบเลือกตอบฉบับที่ 2 เรื่อง กราฟ

ตัวประกอบ	ค่าไอเกน	ค่าร้อยละของความแปรปรวน
1	9.716	32.387
2	1.666	5.552

จากตารางที่ 11 พบว่า ตัวประกอบที่มีค่าไอเกนมากกว่า 1.00 อยู่ 2 ตัวประกอบ โดยตัวประกอบที่ 1 มีค่าไอเกนสูงสุดคือ 9.716 ค่าความแปรปรวนร้อยละ 32.387 ตัวประกอบที่ 2 มีค่า 1.666 ค่าความแปรปรวนร้อยละ 5.552 จะเห็นว่า ค่าความแปรปรวนจากตัวประกอบหลักตัวแรกของแบบสอบมีค่าสูงกว่าร้อยละ 20 แสดงว่าแบบสอบมุ่งวัดคุณลักษณะเด่นเพียงลักษณะเดียว ซึ่งเป็นไปตามข้อตกลงเบื้องต้นของทฤษฎีตอบสนองข้อสอบว่า แบบสอบมุ่งวัดคุณลักษณะเดียว (unidimensionality) นอกจากนี้ยังทำให้แบบสอบฉบับที่ 2 มีคุณสมบัติความเป็นอิสระของข้อสอบด้วย

ตารางที่ 12 ผลการวิเคราะห์ตัวประกอบของแบบสอบเลือกตอบฉบับที่ 3 เรื่อง ระบบสมการเชิงเส้น

ตัวประกอบ	ค่าไอเกน	ค่าร้อยละของความแปรปรวน
1	12.265	76.653
2	.463	2.894

จากตารางที่ 12 พบว่า ตัวประกอบที่มีค่าไอเกนมากกว่า 1.00 อยู่ 1 ตัวประกอบ โดยตัวประกอบที่ 1 มีค่าไอเกน คือ 12.265 ร้อยละความแปรปรวนถึง 76.653 จะเห็นว่า ค่าความแปรปรวนจากตัวประกอบหลักตัวแรกของแบบสอบมีค่าสูงกว่าร้อยละ 20 แสดงว่าแบบสอบมุ่งวัดคุณลักษณะเด่นเพียงลักษณะเดียว ซึ่งเป็นไปตามข้อตกลงเบื้องต้นของทฤษฎีตอบสนองข้อสอบว่า แบบสอบมุ่งวัดคุณลักษณะเดียว (unidimensionality) นอกจากนี้ยังทำให้แบบสอบฉบับที่ 3 มีคุณสมบัติความเป็นอิสระของข้อสอบด้วย

ตารางที่ 13 ผลการวิเคราะห์ตัวประกอบของแบบสอบเลือกตอบฉบับที่ 4 เรื่อง ความคล้าย

ตัวประกอบ	ค่าไอเกน	ค่าร้อยละของความแปรปรวน
1	8.485	35.356
2	1.195	4.980
3	1.036	4.316

จากตารางที่ 13 พบว่า ตัวประกอบที่มีค่าไอเกนมากกว่า 1.00 อยู่ 3 ตัวประกอบ โดยตัวประกอบที่ 1 มีค่าไอเกนคือ 8.485 และค่าความแปรปรวนร้อยละ 35.356 ตัวประกอบที่ 2 มีค่า 1.195 ค่าความแปรปรวนร้อยละ 4.980 จะเห็นว่า ค่าความแปรปรวนจากตัวประกอบหลักตัวแรกของแบบสอบมีค่าสูงกว่าร้อยละ 20 แสดงว่าแบบสอบมุ่งวัดคุณลักษณะเด่นเพียงลักษณะเดียว ซึ่งเป็นไปตามข้อตกลงเบื้องต้นของทฤษฎีตอบสนองข้อสอบว่า แบบสอบมุ่งวัดคุณลักษณะเดียว (unidimensionality) นอกจากนี้ยังทำให้แบบสอบฉบับที่ 4 มีคุณสมบัติความเป็นอิสระของข้อสอบด้วย

ตารางที่ 14 ผลการวิเคราะห์ตัวประกอบของแบบสอบเลือกตอบทั้งฉบับ

ตัวประกอบ	ค่าไอเกน	ค่าร้อยละของความแปรปรวน
1	42.692	42.692
2	20.686	20.686
3	12.768	12.768
4	1.998	1.998
5	1.217	1.217

จากตารางที่ 14 พบว่า ตัวประกอบที่มีค่าไอเกนมากกว่า 1.00 อยู่ 5 ตัวประกอบ โดยตัวประกอบที่ 1 มีค่าไอเกนสูงสุดคือ 42.692 ความแปรปรวนร้อยละ 42.692 ตัวประกอบที่ 2 มีค่า 20.686 ความแปรปรวนร้อยละ 20.686 จะเห็นว่า ค่าความแปรปรวนจากตัวประกอบหลักตัวแรกของแบบสอบมีค่าสูงกว่าร้อยละ 20 แสดงว่าแบบสอบมุ่งวัดคุณลักษณะเด่นเพียงลักษณะเดียว ซึ่งเป็นไปตามข้อตกลงเบื้องต้นของทฤษฎีตอบสนองข้อสอบว่า แบบสอบมุ่งวัดคุณลักษณะเดียว (unidimensionality) นอกจากนี้ยังทำให้แบบสอบฉบับที่ 2 มีคุณสมบัติความเป็นอิสระของข้อสอบด้วย

5.3 การทดลองใช้แบบสอบครั้งที่ 3 เป็นการนำแบบสอบที่ได้รับการปรับปรุงในการทดลองใช้ครั้งที่ 2 นำมาใช้จริงนั้นคือ เพื่อคำนวณหาคุณภาพของข้อสอบ และแบบสอบ หาค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน ค่ามัธยฐาน ฐานนิยม พิสัย ความเบ้ ความโด่งของคะแนนสอบของกลุ่มตัวอย่างนักเรียน สำหรับการหาคุณภาพของข้อสอบคือ ค่าความยาก (b) ค่าอำนาจจำแนก (a) และค่าการเดา (c) และการหาคุณภาพแบบสอบคือ ฟังก์ชัน

สารสนเทศของแบบสอบ (test information) หรือ ความเที่ยงของแบบสอบนั้นเอง โดยใช้โมเดลการตอบสนองข้อสอบ (IRT) แบบ 3 พารามิเตอร์ สามารถวิเคราะห์โดยใช้โปรแกรม Xcalibre สำหรับเกณฑ์ที่ใช้คัดเลือกข้อสอบที่มีคุณภาพดีคือ ค่าความยากอยู่ระหว่าง –2.50 ถึง +2.50 ค่าอำนาจจำแนกอยู่ระหว่าง +0.50 ถึง +2.50 และค่าการเดาไม่เกิน 0.30 (ศิริชัย กาญจนวาสี, 2545) จากนั้นผู้วิจัยนำผลวิเคราะห์ข้อสอบใช้เป็นข้อมูลในการสร้างคู่มือจัดเรียงอันดับในการกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์คต่อไป นอกจากนี้ยังนำคะแนนผลการสอบของนักเรียนเป็นข้อมูลส่วนหนึ่งในการคำนวณหาค่าความตรงในการกำหนดมาตรฐานทั้งวิธีการบุ๊คมาร์ค และวิธีการแองกอฟที่ได้รับการปรับปรุง

ผลการทดลองใช้ในครั้งที่ 3 ปรากฏค่าสถิติ ดังต่อไปนี้

(1) **ค่าสถิติพื้นฐาน** มีดังต่อไปนี้

ตารางที่ 15 ค่าสถิติพื้นฐานของแบบสอบผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 จำนวน 4 ฉบับ รวม 100 ข้อ จำนวนผู้สอบทั้งสิ้น 987 คน

ค่าสถิติพื้นฐาน	ฉบับที่ 1 พื้นที่ผิวและปริมาตร	ฉบับที่ 2 กราฟ	ฉบับที่3 ระบบสมการเชิงเส้น	ฉบับที่ 4 ความคล้าย
คะแนนเต็ม	30.00	30.00	16.00	24.00
คะแนนเฉลี่ย	12.39	11.50	4.33	5.40
ส่วนเบี่ยงเบนมาตรฐาน	5.46	5.65	4.12	4.52
ความเบ้	0.67	0.88	0.71	0.19
ความโด่ง	0.09	0.49	-0.24	-1.04
คะแนนสูงสุด	28.00	29.00	16.00	16.00
คะแนนต่ำสุด	0.00	0.00	0.00	0.00

จากตารางที่ 15 พบว่า ผู้สอบส่วนใหญ่ได้คะแนนสอบทั้ง 4 ฉบับต่ำ (12.39, 11.50 , 4.33 และ 5.40 ตามลำดับ) และเมื่อพิจารณาค่าความเบ้ พบว่า แบบสอบทั้ง 4 ฉบับ มีค่ามากกว่า 0 หรือโค้งเบ้ขวา

(2) ค่าความยาก (b) ค่าอำนาจจำแนก (a) ค่าการเดา (c) และค่าความเที่ยง

ก. ค่าความยาก (b) ค่าอำนาจจำแนก (a) และค่าการเดา(c)

สำหรับผลการวิเคราะห์คุณภาพของข้อสอบคือ ค่าความยาก (b) ค่าอำนาจจำแนก (a) และค่าการเดา (c) ของแบบสอบทั้ง 4 ฉบับ คือ ฉบับที่ 1 พื้นที่ผิวและปริมาตร ฉบับที่ 2 กราฟ ฉบับที่3 ระบบสมการเชิงเส้น และฉบับที่ 4 ความคล้าย ดังตารางที่ 16

ตารางที่ 16 ค่าความยาก (b) ค่าอำนาจจำแนก (a) และค่าการเดา (c) ของแบบสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 จำนวน 4 ฉบับ

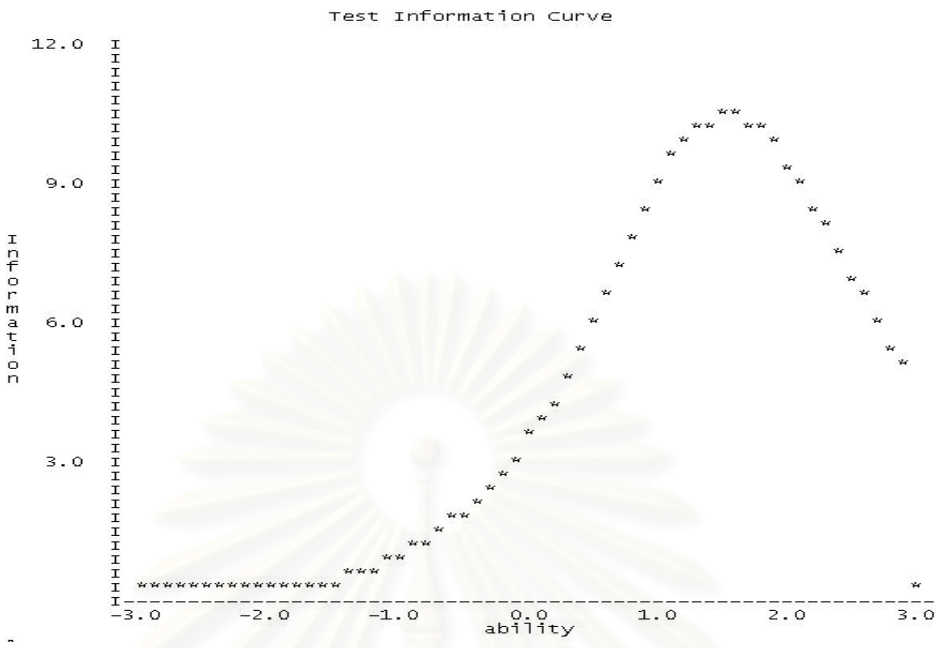
ค่าพารามิเตอร์	ฉบับที่ 1	ฉบับที่ 2	ฉบับที่ 3	ฉบับที่ 4
พิสัยค่าความยาก	-0.81 – 2.07	0.23 – 2.88	-0.13 – 1.90	-0.04 – 2.50
ค่าความยากเฉลี่ย	1.31	1.42	1.26	1.42
พิสัยค่าอำนาจจำแนก	0.93 – 1.21	0.87 – 1.18	0.82 – 1.13	0.90 – 1.11
ค่าอำนาจจำแนกเฉลี่ย	1.03	1.01	1.00	1.02
พิสัยค่าการเดา	0.10 – 0.45	0.12 -0.35	0.08 – 0.24	0.13 – 0.30
ค่าการเดาเฉลี่ย	0.25	0.22	0.15	0.21

จากตารางที่ 16 พบว่า **แบบสอบฉบับที่ 1** มีค่าความยากอยู่ในช่วง -0.81 – 2.07 คิดเป็นค่าความยากเฉลี่ยเท่ากับ 1.31 ค่าอำนาจจำแนกอยู่ในช่วง 0.93 – 1.21 คิดเป็นค่าอำนาจจำแนกเฉลี่ยเท่ากับ 1.03 และค่าการเดาอยู่ในช่วง 0.10 – 0.45 คิดเป็นค่าเฉลี่ยการเดาเท่ากับ 0.25 **แบบสอบฉบับที่ 2** มีค่าความยากอยู่ในช่วง 0.23 – 2.88 คิดเป็นค่าความยากเฉลี่ยเท่ากับ 1.42 ค่าอำนาจจำแนกอยู่ในช่วง 0.87 – 1.18 คิดเป็นค่าอำนาจจำแนกเฉลี่ยเท่ากับ 1.01 และค่าการเดาอยู่ในช่วง 0.12 -0.35 คิดเป็นค่าเฉลี่ยการเดาเท่ากับ 0.22 **แบบสอบฉบับที่ 3** มีค่าความยากอยู่ในช่วง -0.13 – 1.90 คิดเป็นค่าความยากเฉลี่ยเท่ากับ 1.26 ค่าอำนาจจำแนกอยู่ในช่วง 0.82 – 1.13 คิดเป็นค่าอำนาจจำแนกเฉลี่ยเท่ากับ 1.00 และค่าการเดาอยู่ในช่วง 0.08 – 0.24 คิดเป็นค่าเฉลี่ยการเดาเท่ากับ 0.15 และ**แบบสอบฉบับที่ 4** มีค่าความยากอยู่ในช่วง -0.04 – 2.50 คิดเป็นค่าความยากเฉลี่ยเท่ากับ 1.42 ค่าอำนาจจำแนกอยู่ในช่วง 0.90 – 1.11 คิดเป็นค่าอำนาจจำแนกเฉลี่ยเท่ากับ 1.02 และค่าการเดาอยู่ในช่วง 0.13 – 0.30 คิดเป็นค่าเฉลี่ยการเดาเท่ากับ 0.21 จะเห็นว่า แบบสอบทั้ง 4 ฉบับนั้น ข้อสอบส่วนใหญ่ค่อนข้างยาก และสามารถจำแนกผู้สอบที่มีความสามารถแตกต่างกันได้ดี

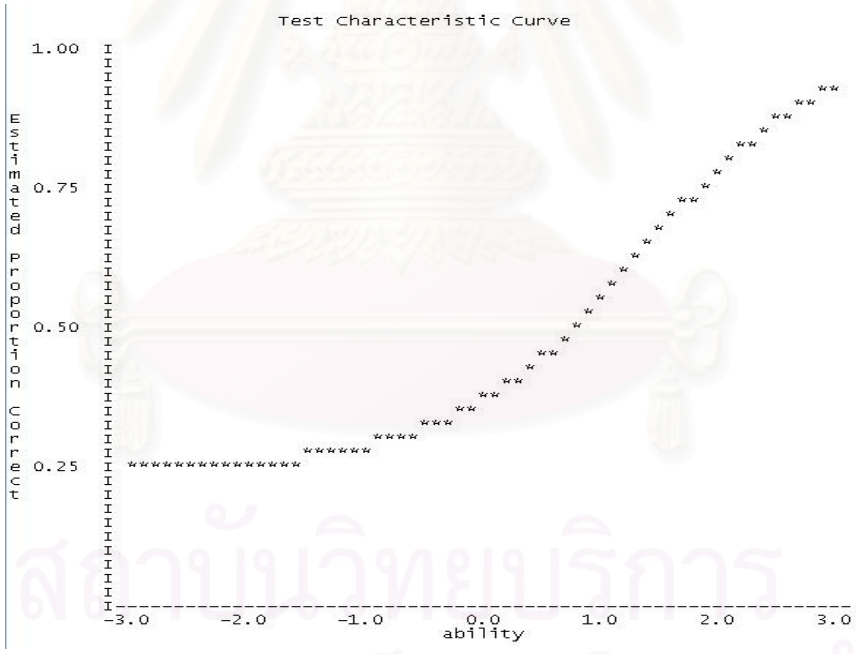
ข. ค่าความเที่ยงของแบบสอบ (Information Function)

จากการวิเคราะห์ข้อสอบตามทฤษฎีการตอบสนองข้อสอบ ผลการคำนวณค่าสารสนเทศของแบบสอบที่ระดับความสามารถต่างๆ พบว่า **แบบสอบฉบับที่ 1** ค่าสารสนเทศของแบบสอบอยู่ในช่วงประมาณ 0.5 ถึง 10 และมีค่าสูงสุดอยู่ที่ระดับความสามารถ (θ) ประมาณ 1.5 **แบบสอบฉบับที่ 2** ค่าสารสนเทศของแบบสอบอยู่ในช่วงประมาณ 0.5 ถึง 12 และมีค่าสูงสุดอยู่ที่ระดับความสามารถ (θ) ประมาณ 1.8 **แบบสอบฉบับที่ 3** ค่าสารสนเทศของแบบสอบอยู่ในช่วงประมาณ 0.5 ถึง 7.0 และมีค่าสูงสุดอยู่ที่ระดับความสามารถ (θ) ประมาณ 1.3 และ **แบบสอบฉบับที่ 4** ค่าสารสนเทศของแบบสอบอยู่ในช่วงประมาณ 0.5 ถึง 8.5 และมีค่าสูงสุดอยู่ที่ระดับความสามารถ (θ)ประมาณ 1.5 และเมื่อรวมข้อสอบทั้ง 4 ฉบับ 100 ข้อ ค่าสารสนเทศของแบบสอบอยู่ในช่วงประมาณ 2.0 ถึง 39.0 และมีค่าสูงสุดอยู่ที่ระดับความสามารถ (θ) ประมาณ 1.5 จะเห็นว่าแบบสอบทั้ง 4 ฉบับ และเมื่อรวมทั้ง 4 ฉบับเป็น 100 ข้อ พบว่า เหมาะสำหรับนักเรียนเก่ง ดังแผนภาพที่ 6 ถึง 15

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

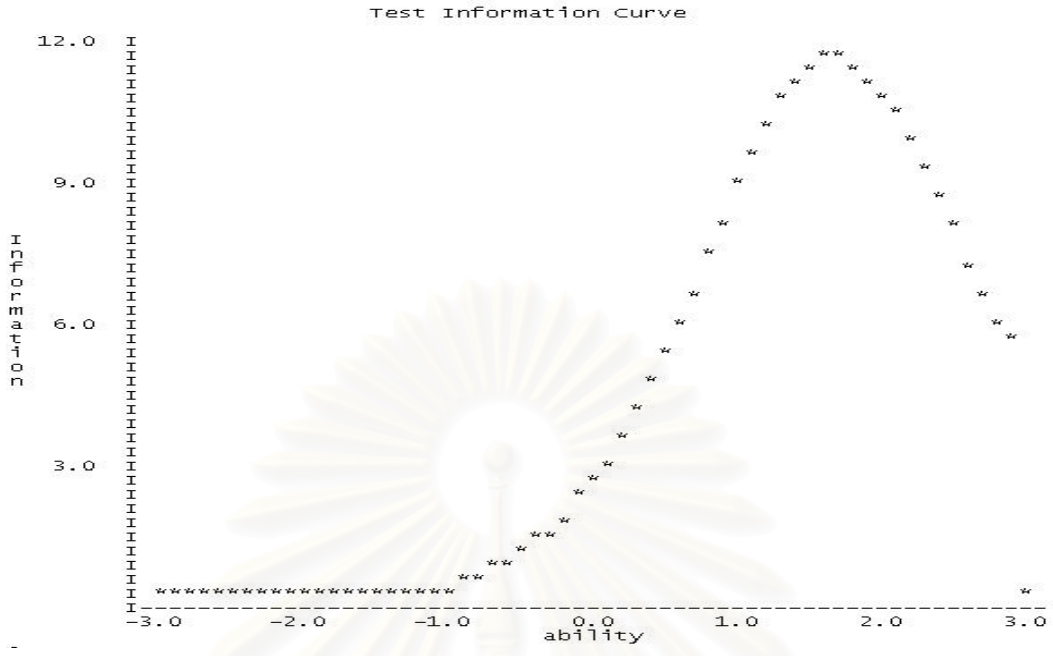


แผนภาพที่ 6 โค้งสารสนเทศของแบบสอบฉบับที่ 1 เรื่อง พื้นที่ผิวและปริมาตร

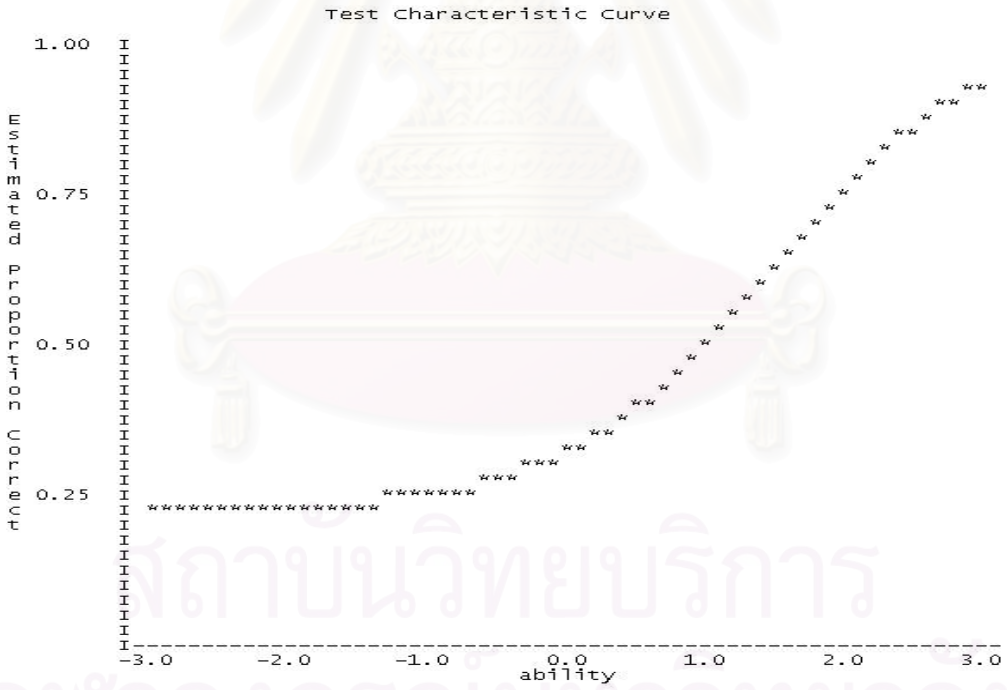


แผนภาพที่ 7 โค้งลักษณะแบบสอบฉบับที่ 1 เรื่อง พื้นที่ผิวและปริมาตร

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

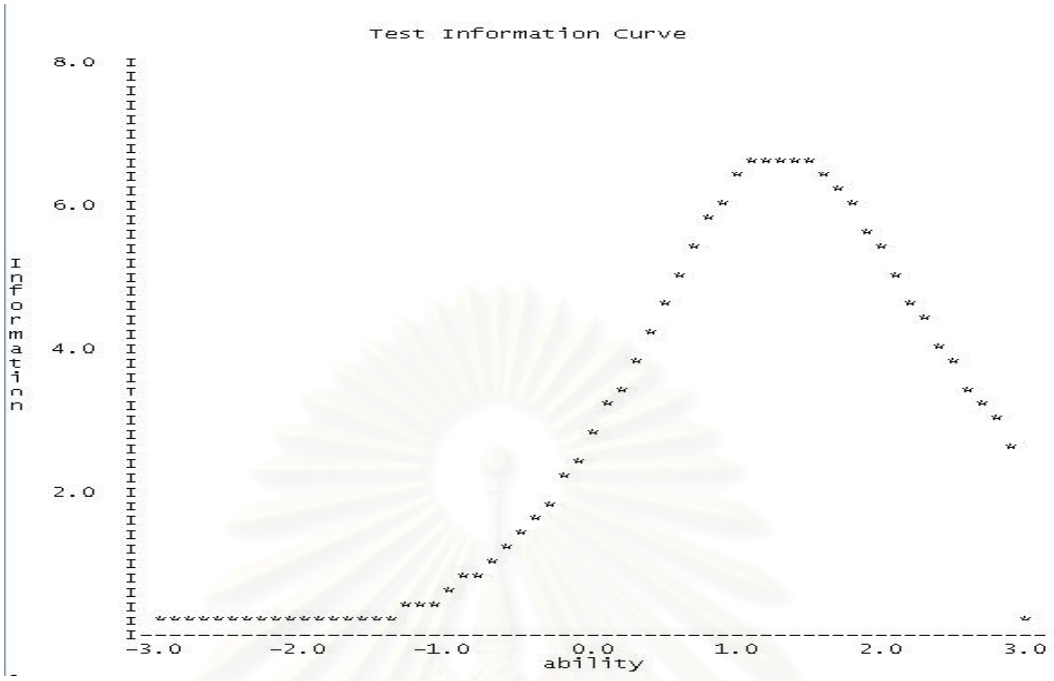


แผนภาพที่ 8 โค้งสารสนเทศของแบบสอบฉบับที่ 2 เรื่อง กราฟ

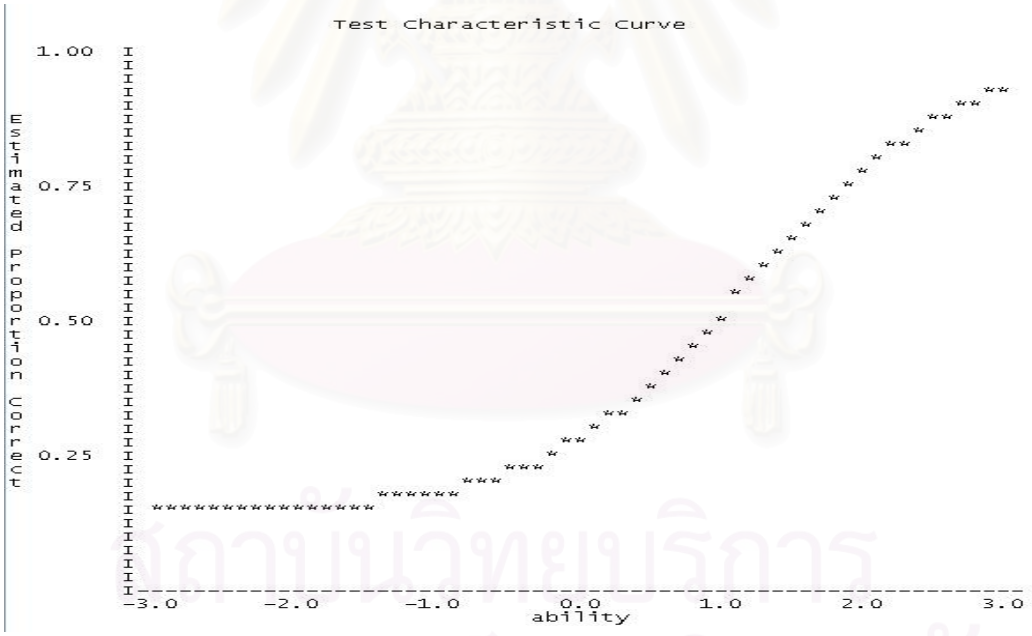


แผนภาพที่ 9 โค้งลักษณะแบบสอบฉบับที่ 2 เรื่อง กราฟ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

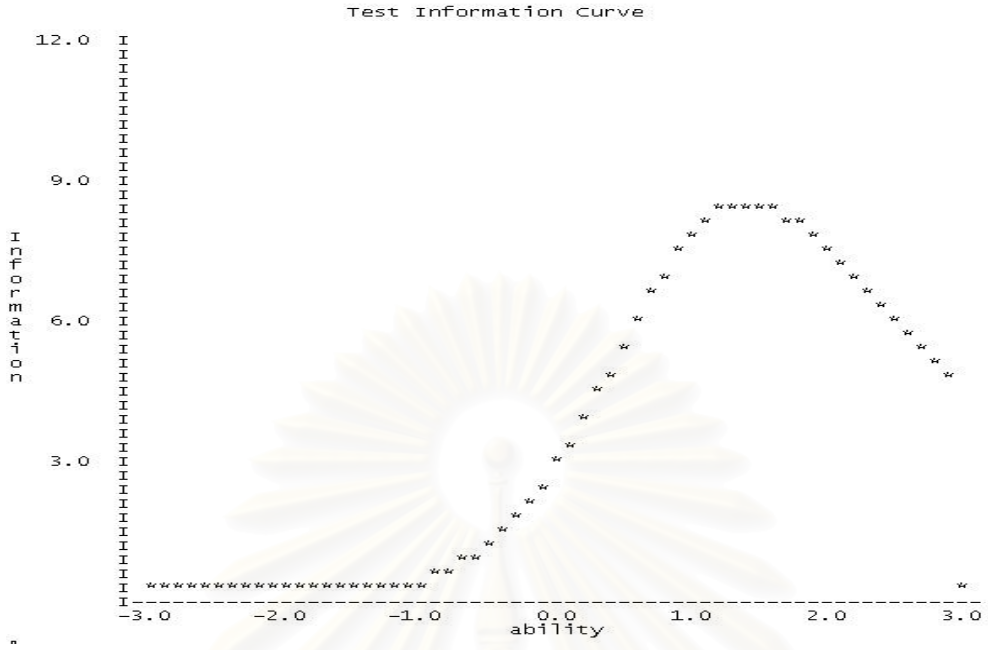


แผนภาพที่ 10 โค้งสารสนเทศของแบบสอบ ฉบับที่3 ระบบสมการเชิงเส้น

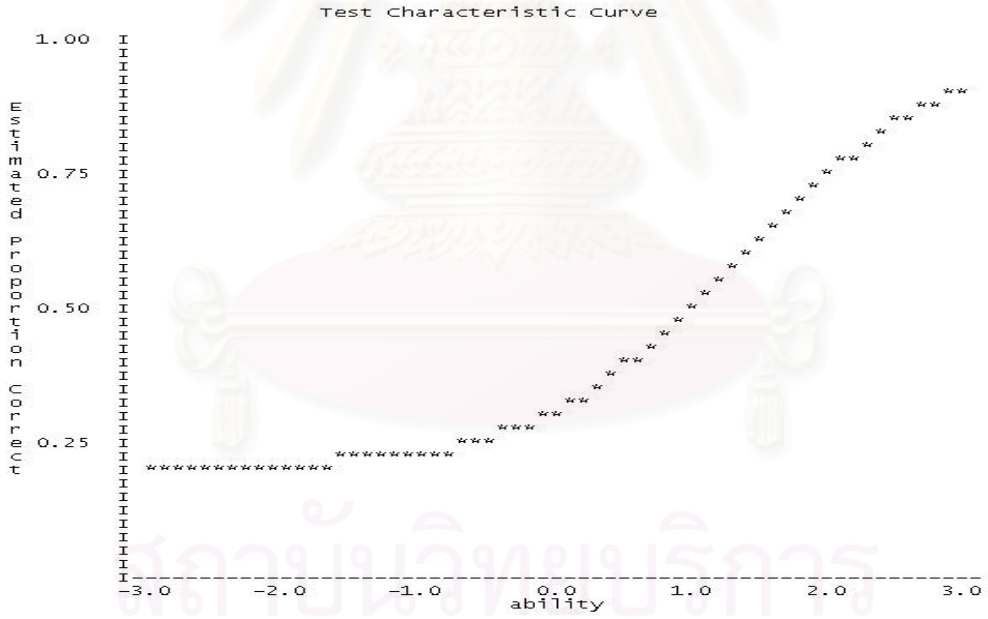


แผนภาพที่ 11 โค้งลักษณะแบบสอบ ฉบับที่3 ระบบสมการเชิงเส้น

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

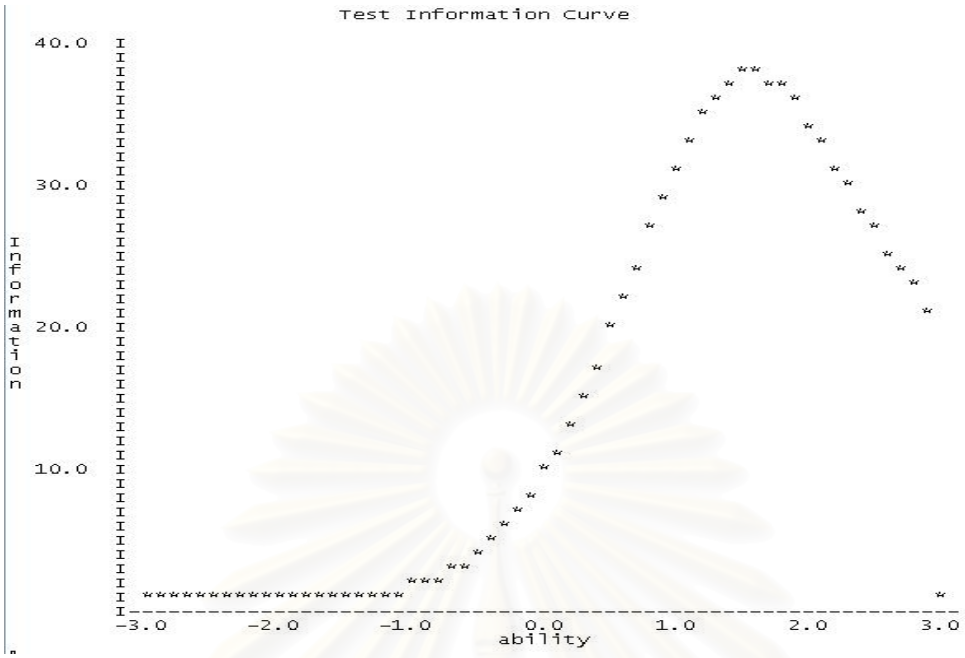


แผนภาพที่ 12 โค้งสารสนเทศของแบบสอบ ฉบับที่ 4 ความคล้าย

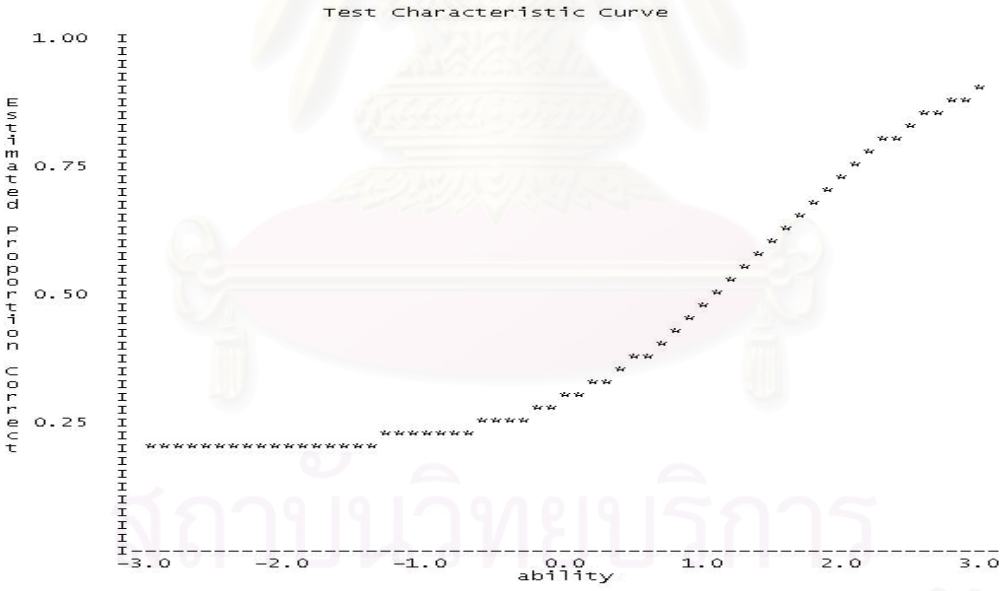


แผนภาพที่ 13 โค้งลักษณะแบบสอบ ฉบับที่ 4 ความคล้าย

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย



แผนภาพที่ 14 โค้งสารสนเทศของแบบสอบรวม 4 ฉบับ 100 ข้อ



แผนภาพที่ 15 โค้งลักษณะแบบสอบรวม 4 ฉบับ 100 ข้อ

2. เครื่องมือเกี่ยวกับความเหมาะสมและความเป็นไปได้ในการนำไปใช้ของการกำหนดมาตรฐานแองกอฟที่ได้รับการปรับปรุง และวิธีการบุ๊คมาร์ค

ขั้นตอนการสร้างและพัฒนาแบบสอบถามเกี่ยวกับความเหมาะสม และความเป็นไปได้ของการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และวิธีบุ๊คมาร์ค มีดังนี้

1. ศึกษาแนวคิดเกณฑ์ประเมินกระบวนการกำหนดมาตรฐานของ Pitoniak (Cizek, Bunch, Koons, 2004) และ Jaeger, Kane, Norcini และ Mehrens (Kane, 1998) ประกอบด้วย

1) **ความชัดเจน** (Explicitness) คือ ระดับความชัดเจนเกี่ยวกับจุดประสงค์ของการกำหนดมาตรฐาน และกระบวนการ

2) **การนำไปปฏิบัติได้** (Practicability) คือ ความง่ายในการนำกระบวนการไปใช้ และการวิเคราะห์ข้อมูลแสดงถึงระดับกระบวนการที่มีความน่าเชื่อถือ และแปลผลได้ตรงกัน

3) **การดำเนินการ** (Implementation) คือ ระดับแสดงถึงกระบวนการที่มีความสมเหตุสมผล เป็นระบบ และได้ผลแม่นยำ เช่น การเลือก และการฝึกผู้ตัดสิน การกำหนดมาตรฐานการปฏิบัติ และการรวบรวมข้อมูล

4) **ผลย้อนกลับ** (feedback)คือ ผู้ตัดสินมีความเชื่อมั่นในกระบวนการการตัดสินคะแนนจุดตัด

5) **เอกสารที่ใช้ในกระบวนการ** (Documentation) คือ จุดเด่นในการศึกษาที่เกิดจากการทบทวนเอกสารต่างๆ สำหรับการประเมิน และจุดประสงค์ของการสื่อสาร

นอกจากนี้ Jaeger, Kane, Norcini และ Mehrens (Kane, 1998) ได้กล่าวถึงเกณฑ์กระบวนการในการกำหนดมาตรฐานไว้ว่า เป็นเรื่องเกี่ยวกับการเลือกและการฝึกผู้ตัดสิน การกำหนดนิยาม มาตรฐานการปฏิบัติ และการออกแบบการเก็บรวบรวมข้อมูล

จะเห็นได้ว่า แนวคิดเกณฑ์ประเมินกระบวนการกำหนดมาตรฐานของ Jaeger, Kane, Norcini และ Mehrens เป็นส่วนหนึ่งของแนวคิดเกณฑ์ประเมินกระบวนการกำหนดมาตรฐานของ Pitoniak ดังนั้นผู้วิจัยจึงนำแนวคิดของ Pitoniak มาสร้างแบบสอบถามเกี่ยวกับความเหมาะสม และความเป็นไปได้ในกระบวนการการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และ วิธีการบุ๊คมาร์ค

นอกจากนี้เพื่อให้รายละเอียดเพิ่มเติม ผู้วิจัยยังนำเกณฑ์ประเมินกระบวนการกำหนดมาตรฐานของ Pitoniak มาเป็นกรอบในการสร้างแบบสอบถามความคิดเห็นเกี่ยวกับกระบวนการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และวิธีบุ๊คมาร์ค ส่วนข้อมูลความคิดเห็นเชิงลึกผู้วิจัยได้ทำการสัมภาษณ์ผู้ตัดสินเกี่ยวกับกระบวนการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และวิธีบุ๊คมาร์ค

2. ผู้วิจัยกำหนดโครงสร้างแบบสอบถามความเหมาะสม และความเป็นไปได้ในกระบวนการการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และวิธีการบุ๊คมาร์ค และโครงสร้างแบบสอบถามกระบวนการการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และ วิธีการบุ๊คมาร์ค ดังตารางที่ 17 ถึง 18

ตารางที่ 17 โครงสร้างแบบสอบถามความเหมาะสม และความเป็นไปได้ในกระบวนการการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และ วิธีการบุ๊คมาร์ค

เนื้อหา	จำนวนข้อ
ความเหมาะสม และความเป็นไปได้ในกระบวนการกำหนดมาตรฐาน	
1) ความชัดเจน	1
2) การนำไปปฏิบัติได้	1
3) การดำเนินการ	5
4) ผลย้อนกลับ	1
5) เอกสารที่ใช้ในกระบวนการ	5
รวม	13

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 18 โครงสร้างแบบสอบถามกระบวนการการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และ วิธีการบุ๊คมาร์ค

เนื้อหา	จำนวนข้อ	
	วิธีแองกอฟที่ได้รับการปรับปรุง	วิธีบุ๊คมาร์ค
กระบวนการกำหนดมาตรฐาน		
1) ความชัดเจน	2	2
2) การนำไปปฏิบัติได้	4	4
3) การดำเนินการ	7	7
4) ผลย้อนกลับ	4	4
5) เอกสารที่ใช้ในกระบวนการ	8	8
รวม	25	25

3. สร้างแบบสอบถามแบบมาตรประมาณค่า 5 ระดับ และสร้างแบบสัมภาษณ์แบบมีโครงสร้าง

4. นำแบบสอบถาม ไปให้ผู้ทรงคุณวุฒิด้านการวัด และประเมินผลการศึกษา ที่มีประสบการณ์ด้านการวัด และประเมินผลการศึกษามาไม่ต่ำกว่า 3 ปี จำนวน 3 คน พิจารณาความเหมาะสม และความตรงเชิงเนื้อหา (content validity) ของเครื่องมือ ผู้วิจัยพิจารณาค่า IOC ซึ่งแต่ละข้อควรมีค่าตั้งแต่ .50 ขึ้นไป จากนั้นดำเนินการปรับปรุงแก้ไขเครื่องมือต่อไป

ผลการตรวจสอบความตรงเชิงเนื้อหาของแบบสอบถามความเหมาะสม และความเป็นไปได้ในกระบวนการการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และวิธีการบุ๊คมาร์ค พบว่าค่า IOC เท่ากับ 1.0 ทุกข้อ

ผลการตรวจสอบความตรงเชิงเนื้อหาของแบบสอบถามกระบวนการการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และ วิธีการบุ๊คมาร์ค พบว่าค่า IOC เท่ากับ 1.0 ทุกข้อ และผู้เชี่ยวชาญได้แนะนำปรับคำถามเพื่อให้เกิดความเข้าใจยิ่งขึ้น ดังนี้

"วิธีการนี้ทำให้ท่านเข้าใจวัตถุประสงค์ของการประชุมกลุ่มได้อย่างชัดเจน" **แก้ไขเป็น** *"วิธีการนี้ทำให้ท่านสามารถดำเนินการให้เป็นไปตามวัตถุประสงค์ได้"*

5. ดำเนินการเก็บรวบรวมข้อมูลกับผู้ตัดสินจำนวน 12 คน คือ หลังจากที่ผู้ตัดสินกำหนดมาตรฐานด้วยวิธีการบุ๊คมาร์คและวิธีการแองกอฟที่ได้รับการปรับปรุงเป็นที่

เรียบร้อยแล้ว ผู้วิจัยแจกแบบสอบถามให้ผู้ตัดสินทำ จากนั้นผู้วิจัย และผู้สัมภาษณ์อีก 1 คน ดำเนินการสัมภาษณ์ผู้ตัดสินเหล่านั้น

6. วิเคราะห์คุณภาพด้านความเที่ยงของแบบสอบถามด้วยวิธีสัมประสิทธิ์แอลฟาของครอนบาค ผลการวิเคราะห์ค่าความเที่ยงของแบบสอบถามความเหมาะสม และความเป็นไปได้ในกระบวนการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และวิธีบุ๊คมาร์คคือ 0.74 และค่าความเที่ยงของแบบสอบถามกระบวนการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุงคือ 0.81 และค่าความเที่ยงของแบบสอบถามกระบวนการกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์คคือ 0.84

3. คู่มือวิธีการกำหนดมาตรฐานวิธีแองกอฟที่ได้รับการปรับปรุง และวิธีการบุ๊คมาร์ค

1. ผู้วิจัยดำเนินการสร้างคู่มือวิธีการกำหนดมาตรฐานวิธีแองกอฟที่ได้รับการปรับปรุง และวิธีบุ๊คมาร์ค จำนวน 2 ฉบับ แต่ละฉบับมีรายละเอียดหัวข้อดังนี้

- บทนำ
- วัตถุประสงค์ของกระบวนการกำหนดมาตรฐาน
- นิยามคำศัพท์
- ขั้นตอนการกำหนดมาตรฐาน

2. ผู้วิจัยนำคู่มือที่สร้างขึ้น มาทดลองใช้กับกลุ่มตัวอย่างผู้ตัดสิน
3. ปรับปรุงคู่มือตามข้อเสนอแนะของผู้ตัดสิน

การเก็บรวบรวมข้อมูล

การเก็บรวบรวมข้อมูลแบ่งออกเป็น 2 ระยะคือ ระยะที่ 1 ดำเนินการสอบกับนักเรียน และระยะที่ 2 ดำเนินการเก็บข้อมูลกับกลุ่มผู้ตัดสิน มีขั้นตอนการดำเนินงานเก็บรวบรวมข้อมูลแต่ละระยะดังนี้

การเก็บข้อมูลระยะที่ 1 (ดำเนินการเก็บข้อมูลกับผู้สอบ)

1. ผู้วิจัยนำหนังสือขอความร่วมมือในการทำวิจัยจากบัณฑิตวิทยาลัยไปติดต่อกับโรงเรียนที่เป็นกลุ่มตัวอย่างเพื่อขอความร่วมมือในการเก็บข้อมูล เพื่อทำการนัดวัน เวลา สถานที่ในการเก็บข้อมูล

2. ผู้วิจัยดำเนินการสอบนักเรียนจำนวน 1,287 คนเพื่อพัฒนาแบบสอบวิชาคณิตศาสตร์ชั้นมัธยมศึกษาปีที่ 3 ทั้งหมด 4 ฉบับ รวม 100 ข้อ และดำเนินการเก็บข้อมูลจริง โดยดำเนินการทดลองใช้แบบสอบครั้งแรกกับนักเรียนจำนวน 10 คน ทดลองใช้แบบสอบครั้งที่ 2 จำนวน 290 คน และเก็บข้อมูลจริง จำนวน 987 คน ในโรงเรียนสังกัดการศึกษาขั้นพื้นฐาน ชลบุรีเขต 1 การเก็บข้อมูลระยะที่ 1 ใช้เวลาทั้งสิ้นหนึ่งภาคเรียนคือ ปีการศึกษา 2549 ภาคเรียนที่ 1 วิธีการเก็บข้อมูลคือ เมื่อนักเรียนเรียนเนื้อหาแต่ละบทเรียนจบ ก็จะสอบบทเรียนนั้น เริ่มจากแบบสอบฉบับที่ 1 พื้นที่ผิวและปริมาตร ฉบับที่ 2 กราฟ ฉบับที่ 3 สมการเชิงเส้นสองตัวแปร และฉบับที่ 4 ความคล้าย ตามลำดับ ปัญหาอุปสรรคในการเก็บรวบรวมข้อมูลระยะที่ 1 คือ นักเรียนบางคนไม่สนใจทำแบบสอบเท่าที่ควรเพราะเห็นว่า ไม่มีผลต่อตัวนักเรียนเอง และบางโรงเรียนต้องการให้สอบทีเดียวเนื่องจากไม่อยากรบกวนเวลาเรียนของนักเรียนมากเกินไป จึงจัดให้สอบพร้อมกันที่โรงอาหารทำให้การควบคุมการสอบเป็นไปด้วยความยากลำบาก นักเรียนลอกกัน และไม่สนใจทำข้อสอบเท่าที่ควร

3. นำผลการสอบจากการทดลองใช้แบบสอบมาวิเคราะห์ข้อสอบเพื่อปรับปรุงข้อสอบ และนำผลการสอบจากการเก็บข้อมูลจริงมาหาค่าพารามิเตอร์ของข้อสอบ

การเก็บข้อมูลระยะที่ 2 (ดำเนินการเก็บข้อมูลกับผู้ตัดสิน)

1. ผู้วิจัยจัดทำโครงการอบรมเชิงปฏิบัติการเรื่อง การกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และวิธีการบุ๊คมาร์ค โดยทำโครงการอบรมในนามคณะศึกษาศาสตร์ มหาวิทยาลัยบูรพา

2. ส่งหนังสือขอเชิญอบรมเชิงปฏิบัติการไปยังโรงเรียนต่างๆ ในเขตพื้นที่การการศึกษาชลบุรี เขต 1 เพื่อให้ทางโรงเรียนประชาสัมพันธ์ให้แก่อาจารย์ที่สนใจเข้าร่วมอบรม ณ คณะศึกษาศาสตร์ มหาวิทยาลัยบูรพา และอาจารย์ที่เข้าร่วมอบรมต้องมีคุณสมบัติครบถ้วน ตามที่ผู้วิจัยระบุไว้ คือ เป็นอาจารย์สอนวิชาคณิตศาสตร์ มีประสบการณ์สอนมาไม่น้อยกว่า 5 ปี และเป็นครูชำนาญการ หรือชำนาญการพิเศษ

3. ผู้วิจัยจัดเตรียมเอกสารการอบรม เช่น คู่มือวิธีการกำหนดมาตรฐานแองกอฟที่ได้รับการปรับปรุง และวิธีการบุ๊คมาร์ค แบบฟอร์มสำหรับกำหนดคะแนนจุดตัด แบบสอบทั้ง 4 ฉบับ และได้นำข้อมูลผลการสอบ และค่าพารามิเตอร์ข้อสอบ เป็นต้น จากการเก็บข้อมูลระยะที่ 1 มาเป็นข้อมูลสร้างกราฟแจกแจงความถี่คะแนนสอบ คู่มือจัดเรียงข้อสอบ และค่าความยากของข้อสอบ เพื่อใช้เป็นข้อมูลสารสนเทศสำหรับผู้ตัดสินใช้ในการกำหนดคะแนนจุดตัด

4. ดำเนินการจัดอบรมเชิงปฏิบัติการระหว่างวันที่ 3 – 5 เมษายน 2550 รวม 3 วัน และเนื่องจากผู้วิจัยไม่ต้องการให้ลำดับของวิธีการกำหนดมาตรฐานมีผลทำให้เกิดความคลาดเคลื่อนที่เรียกว่า carry-over effect (Federer, 1955) ดังนั้นผู้วิจัยจึงออกแบบการวิจัยเป็นแบบ double-reversal design โดยการกำหนด treatment 2 ตัว คือ วิธีการกำหนดมาตรฐานแองกอฟที่ได้รับการปรับปรุง และวิธีการกำหนดมาตรฐานบุ๊คมาร์ค กำหนดช่วงเวลาของ treament เป็น 3 ระยะ ดังนี้

ลำดับของ treatment

ครั้งที่	1	วิธีการแองกอฟที่ได้รับการปรับปรุง	วิธีการบุ๊คมาร์ค
ครั้งที่	2	วิธีการบุ๊คมาร์ค	วิธีการแองกอฟที่ได้รับการปรับปรุง
ครั้งที่	3	วิธีการแองกอฟที่ได้รับการปรับปรุง	วิธีการบุ๊คมาร์ค

5. ในการอบรมเชิงปฏิบัติการครั้งนี้มีอาจารย์ตอบรับเข้าร่วมอบรมทั้งหมด 12 คน ผู้อบรบซึ่งเป็นผู้ตัดสินในงานวิจัยนี้ ได้มีหนังสือจากทางโรงเรียนผ่านผู้อำนวยการโรงเรียนเพื่อส่งอาจารย์เหล่านี้เข้าร่วมอบรม การดำเนินงานเก็บข้อมูลวันแรก ช่วงเช้า ผู้วิจัยบรรยายวิธีการกำหนดมาตรฐานทั้งสองวิธี พร้อมลองฝึกปฏิบัติ ช่วงบ่ายดำเนินการเก็บข้อมูลจริงโดยให้ผู้ตัดสินกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง

วันที่สองดำเนินการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง (ต่อ) และดำเนินการกำหนดมาตรฐานบุ๊คมาร์ค ซึ่งจากการสังเกตพบว่า ผู้ตัดสินใช้เวลาในการกำหนดมาตรฐานวิธีแองกอฟที่ได้รับการปรับปรุงนานกว่าวิธีบุ๊คมาร์คประมาณ 2 เท่า จากนั้นช่วงบ่ายดำเนินการกำหนดมาตรฐานครั้งที่ 2 ด้วยวิธีบุ๊คมาร์ค และตามด้วยวิธีการแองกอฟที่ได้รับการปรับปรุง การดำเนินงานครั้งที่สองไม่ใช้เวลาเหมือนครั้งแรกเนื่องจากผู้ตัดสินมีความคุ้นเคยกับแบบสอบแล้ว

วันที่สาม ช่วงเช้าดำเนินการกำหนดมาตรฐานครั้งที่ 3 ด้วยวิธีแองกอฟที่ได้รับการปรับปรุงและวิธีการบุ๊คมาร์ค การดำเนินงานครั้งที่ 3 ใช้เวลาในการกำหนดมาตรฐานไม่มากเหมือนครั้งที่ 1 และ 2 ส่วนช่วงบ่ายดำเนินการระดมสมองเขียนบรรยายลักษณะนักเรียนในแต่ละระดับ จากการสังเกตในวันที่สาม ผู้ตัดสินมีความเหนื่อยล้าอย่างเห็นได้ชัด แต่ทุกคนได้ให้ความร่วมมือเป็นอย่างดี ผู้วิจัยคาดว่า ผู้ตัดสินเหล่านี้มาเข้าร่วมอบรมด้วยความสมัครใจ และมาในราชการ นอกจากนี้ผู้วิจัยยังมอบวุฒิบัตรผ่านการอบรมในครั้งนี้แก่ผู้ตัดสิน เพื่อใช้เป็นหลักฐานในการเข้าอบรม และนำเข้าแฟ้มสะสมงานของผู้ตัดสินต่อไป

6. ผลการกำหนดมาตรฐานจากทั้งสองวิธี ผู้วิจัยนำผลดังกล่าวมาวิเคราะห์หาค่าความเที่ยงและความตรงของการกำหนดมาตรฐานทั้งสองวิธีต่อไป

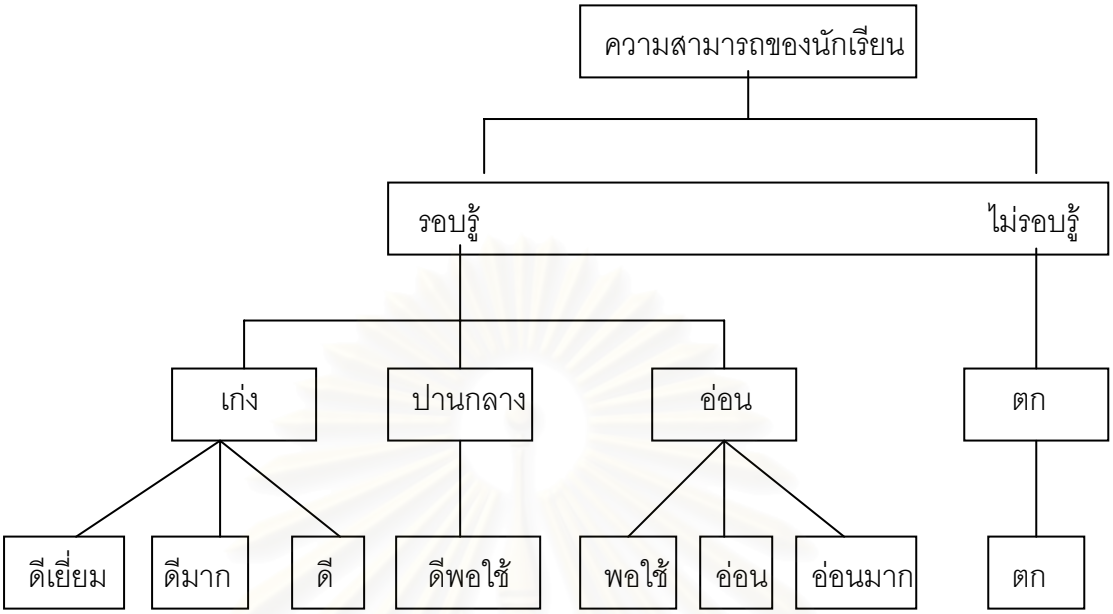
7. หลังจากผู้ตัดสินดำเนินการกำหนดมาตรฐานทั้งสองวิธีเรียบร้อยแล้ว ผู้วิจัยแจกแบบสอบถามความเหมาะสม และความเป็นไปได้ในกระบวนการการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และ วิธีการบุ๊คมาร์ค และแบบสอบถามกระบวนการการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และ วิธีการบุ๊คมาร์คให้ผู้ตัดสินทำ แล้วนำผลการประเมินมาหาค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน มัธยฐาน และพิสัยคลอไทล์

8. ผู้วิจัย และผู้สัมภาษณ์ รวม 2 คน ดำเนินการสัมภาษณ์ผู้ตัดสิน แล้วนำผลการสัมภาษณ์มาวิเคราะห์เนื้อหา

สำหรับขั้นตอนการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และวิธีบุ๊คมาร์ค มีรายละเอียดวิธีดำเนินการดังนี้

ก่อนการกำหนดมาตรฐาน ผู้วิจัยแนะนำลำดับการพิจารณาคะแนนจุดตัดแก่ผู้ตัดสิน มีรายละเอียดคือ ขั้นที่ 1 พิจารณานักเรียน 2 กลุ่มคือ รอบรู้ และไม่รอบรู้ ขั้นที่ 2 กลุ่มรอบรู้แยกเป็น กลุ่มเก่ง กลุ่มปานกลาง และกลุ่มอ่อน ส่วนกลุ่มไม่รอบรู้ แยกเป็นกลุ่มตก และขั้นที่ 3 แยกกลุ่มเก่ง เป็น กลุ่มดีเยี่ยม กลุ่มดีมาก และกลุ่มดี กลุ่มปานกลาง คือ กลุ่มดีพอใช้ กลุ่มอ่อน แยกเป็น กลุ่มพอใช้ กลุ่มอ่อน และ กลุ่มอ่อนมาก ดังแผนภาพที่ 16 ลำดับขั้นการพิจารณาความสามารถของนักเรียน 8 ระดับ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย



แผนภาพที่ 16 ลำดับขั้นการพิจารณาความสามารถของนักเรียน 8 ระดับ

จากนั้นผู้ตัดสิน 12 คน ร่วมอบรมเชิงปฏิบัติการเพื่อรับฟังวิธีการกำหนดมาตรฐานด้วยวิธีการแองกอฟที่ได้รับการปรับปรุง และการกำหนดมาตรฐานด้วยวิธีการบุ๊คมาร์ค พร้อมทั้งปฏิบัติการกำหนดมาตรฐานทั้งสองวิธี เพื่อให้เกิดความเข้าใจในวิธีการกำหนดมาตรฐาน

วิธีกำหนดมาตรฐานแองกอฟที่ได้รับการปรับปรุง(Modified Angoff)

1. ผู้วิจัยแจกคำอธิบายลักษณะความสามารถของนักเรียนแต่ละระดับ เพื่อให้ผู้ตัดสินเข้าใจตรงกัน พร้อมทั้งแจกค่าความยากของข้อสอบแต่ละข้อ แบบสอบทั้ง 4 ฉบับ โครงสร้างเนื้อหาของแบบสอบทั้ง 4 ฉบับ นอกจากนี้ผู้วิจัยยังได้เพิ่มข้อมูลประกอบการตัดสินแก่ผู้ตัดสินอีกอย่างคือ กราฟแจกแจงความถี่ของคะแนนสอบ

2. ผู้ตัดสินแต่ละคนพิจารณาข้อสอบทีละข้อเพื่อกำหนดคะแนนจุดตัด โดยข้อสอบที่พิจารณาจะมีลำดับข้อสอบเหมือนกับแบบสอบ ทำการตัดสินว่า *1) กลุ่มผู้สอบที่มีความสามารถ<u>ระดับอ่อนมาก</u>มีโอกาสตอบข้อสอบข้อนั้นถูกกี่เปอร์เซนต์ และ 2) กลุ่มผู้สอบที่มีความสามารถ<u>ระดับอ่อน</u>มีโอกาสตอบข้อสอบข้อนั้นถูกกี่เปอร์เซนต์ 3) กลุ่มผู้สอบที่มีความสามารถ<u>ระดับพอใช้</u>มีโอกาสตอบข้อสอบข้อนั้นถูกกี่เปอร์เซนต์ 4) กลุ่มผู้สอบที่มีความสามารถ<u>ระดับดีพอใช้</u>มีโอกาสตอบข้อสอบข้อนั้นถูกกี่เปอร์เซนต์ 5) กลุ่มผู้สอบที่มีความสามารถ<u>ระดับดี</u>มีโอกาสตอบข้อสอบข้อนั้นถูกกี่เปอร์เซนต์ 6) กลุ่มผู้สอบที่มีความสามารถ<u>ระดับดีมาก</u>มีโอกาสตอบข้อสอบข้อนั้นถูกกี่*

เปอร์เซนต์ 7) กลุ่มผู้สอบที่มีความสามารถ<u>ระดับดีเยี่ยม</u>มีโอกาสตอบข้อสอบข้อนั้นถูกกี่เปอร์เซนต์

3. จากนั้นแบ่งกลุ่มผู้ตัดสิน 12 คน ออกเป็นกลุ่มย่อย 2 กลุ่ม กลุ่มละ 6 คน ในกลุ่มย่อยแต่ละกลุ่ม ผู้ตัดสินร่วมกันอภิปรายผลการตัดสินของแต่ละคน โดยเฉพาะผู้ตัดสินที่มีผลตัดสินต่ำสุด และสูงสุดให้อธิบายแสดงความคิดเห็นถึงเหตุผลที่ให้ผลการตัดสินต่ำสุด และสูงสุด และระหว่างการอภิปรายผู้อำนวยความสะดวกแสดงรูปการแจกแจงผลการตัดสินของผู้ตัดสินแต่ละคน เมื่อผู้ตัดสินอภิปรายเสร็จ ก็เปิดโอกาสให้ตัดสินคะแนนอีกครั้งหนึ่ง การให้คะแนนอีกครั้งเป็นการเปิดโอกาสให้ผู้ตัดสินปรับเปลี่ยนคะแนน หรือยังคงคะแนนเดิมก็ได้ จากนั้นนำผลการตัดสินของผู้ตัดสินทุกคน โดยแต่ละคนจะมีคะแนนจุดตัด 7 ระดับ คือ 1) ระดับดีเยี่ยม 2) ระดับดีมาก 3) ระดับดี 4) ระดับดีพอใช้ 5)ระดับพอใช้ 6) ระดับอ่อน 7) ระดับอ่อนมาก มาหาค่าเฉลี่ยในแต่ละระดับ

4. นำผู้ตัดสินทั้งสองกลุ่มรวมเป็นกลุ่มเดียวกัน ผู้อำนวยความสะดวกแสดงค่าเฉลี่ยคะแนนจุดตัดของแต่ละกลุ่มพิจารณา เปิดโอกาสให้ผู้ตัดสินอภิปรายร่วมกัน จากนั้นให้โอกาสผู้ตัดสินทำการตัดสินคะแนนจุดตัดทั้ง 7 ระดับอีกครั้งหนึ่งอย่างอิสระจากกัน แล้วนำผลการตัดสินของผู้ตัดสินแต่ละคนมาหาค่าเฉลี่ยของแต่ละระดับ เปลี่ยนคะแนนจุดตัดทั้ง 7 ระดับให้เป็นค่าความสามารถ (θ) โดยอาศัยโค้งลักษณะแบบสอบ (test characteristic curve) ที่ได้จากโค้งลักษณะข้อสอบ (item characteristic curve) ของข้อสอบแต่ละข้อ จะทำให้ได้คะแนนจุดตัดบนสเกลความสามารถ

วิธีกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์ค (Bookmark)

ผู้วิจัยสรุปขั้นตอนการกำหนดมาตรฐานบุ๊คมาร์คจากการสังเคราะห์เอกสารงานวิจัยที่เกี่ยวข้องดังนี้ (Buckendahl, 2002; Wang, 2003; Beretvas, 2004; Lewis, Green, Mitzel, Baum & Patz, 1998; Kiplinger, 1997; Lin, 2005; Dawber and Lewis, 2005; Wisconsin Department of Public Instruction, 2004; Cizek, Bunch & Koons, 2004; Ferrara, Johnson & Chen, 2005)

1. สร้างคู่มือการจัดเรียงข้อสอบ (ordered item booklet : OIB) คู่มือการจัดเรียงข้อสอบดังกล่าว มีการจัดเรียงข้อสอบตามค่าความยาก โดยจัดเรียงข้อสอบจากข้อสอบง่ายที่สุดไปยังข้อสอบยากที่สุด มีรูปแบบการจัดเรียงข้อสอบ 1 ข้อ ต่อ 1 หน้า ในงานวิจัยนี้ ผู้วิจัยได้เพิ่มสารสนเทศสำหรับผู้ตัดสินลงในคู่มือจัดเรียงข้อสอบแต่ละหน้าคือ เปอร์เซ็นต์ของจำนวนนักเรียนที่

เลือกตัวเลือกแต่ละตัวเลือก ดังแผนภาพที่ 17 ตัวอย่างรูปแบบหน้าในคู่มือจัดเรียงข้อสอบ และแผนภาพที่ 18 ตัวอย่างการเรียงข้อสอบจากข้อง่ายสุดไปยังข้อยากสุดในคู่มือจัดเรียงข้อสอบ

ข้อที่ 22 35

ระดับความสามารถที่โอกาสความน่าจะเป็นในการตอบถูก 0.67 **คือ** 1.725

(นักเรียนค่อนข้างเก่ง)

การหาคำตอบของระบบสมการเชิงเส้นสองตัวแปร

คำตอบของระบบสมการ $x + y = 2$ และ $2x - 3y = 1$ คือคู่อันดับในข้อใด

ก. (1, 1)

ข. (2, 0)

ค. $\left(\frac{5}{3}, \frac{1}{3}\right)$

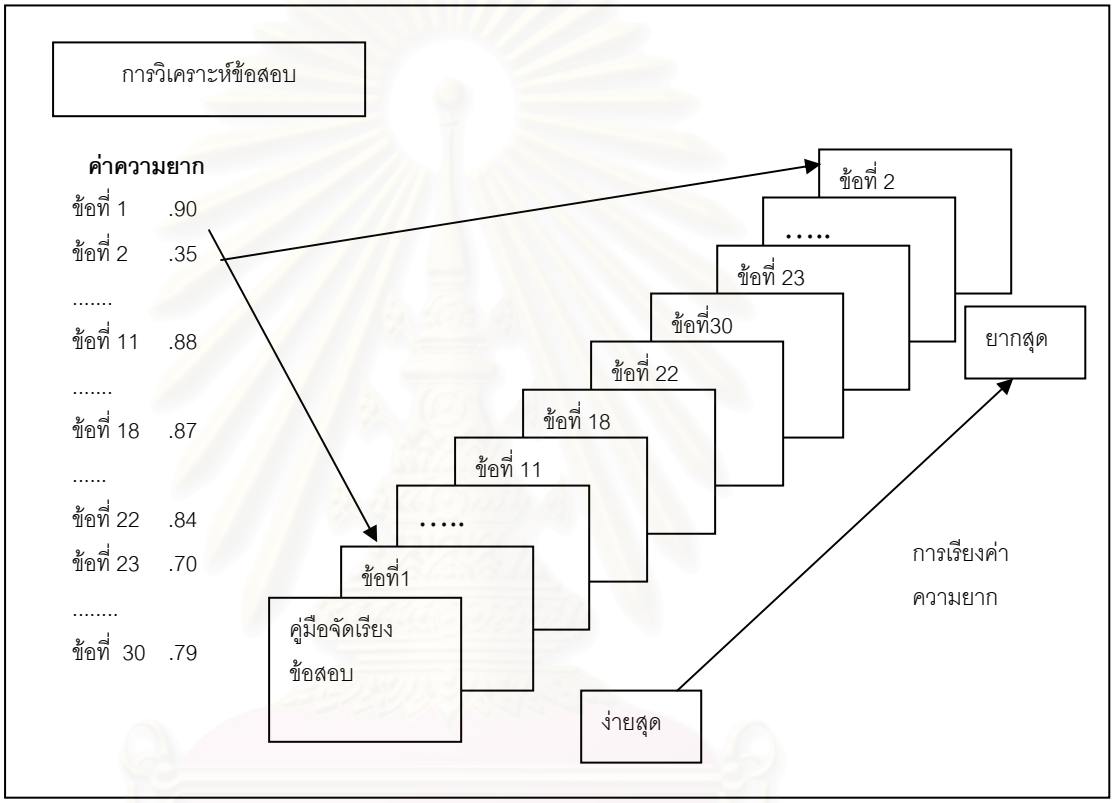
(ง.) $\left(\frac{7}{5}, \frac{3}{5}\right)$

ตัวเลือก	จำนวนนักเรียนที่เลือก (%)	กลุ่มสูง (%)	กลุ่มต่ำ (%)
ก	4	0	4
ข	8	1	4
ค	12	3	7
(ง)	76	96	85

แผนภาพที่ 17 ตัวอย่างรูปแบบหน้าในคู่มือจัดเรียงข้อสอบ

จากแผนภาพที่ 17 ตัวเข้มบนขวา (35) เป็นตัวเลขที่แสดงเลขที่หน้าของคู่มือจัดเรียงข้อสอบ ตัวเข้มบนซ้าย (ข้อที่ 22) เป็นเลขที่ข้อสอบ ถัดลงมาคือ ระดับความสามารถของผู้สอบ (θ=1.725) จากทฤษฎีการตอบสนองข้อสอบ (item response theory) ที่มีความน่าจะเป็นในการตอบข้อสอบข้อนี้ถูก .67 ข้อมูลถัดลงมาคือ ข้อมูลของข้อสอบที่แสดงข้อสอบ

ข้อนี้เป็นเนื้อหาเรื่อง การหาคำตอบของระบบสมการเชิงเส้นสองตัวแปร ถัดลงมาเป็นตารางแสดงเปอร์เซ็นต์จำนวนนักเรียนที่เลือกแต่ละตัวเลือก และเปอร์เซ็นต์นักเรียนกลุ่มสูง และกลุ่มต่ำที่เลือกตัวเลือกแต่ละตัวเลือก ตารางดังกล่าวผู้วิจัยใส่เพิ่มลงไปในคู่มือจัดเรียงข้อสอบ เพื่อให้ผู้ตัดสินมีสารสนเทศเพิ่มเติมประกอบการตัดสินคะแนนจุดตัด ส่วนเครื่องหมาย ( ) ตรงตัวเลือก ง. แสดงถึงคำตอบที่ถูก



แผนภาพที่ 18 ตัวอย่างการเรียงข้อสอบจากข้อง่ายสุดไปยังข้อยากสุดในคู่มือจัดเรียงข้อสอบ

2. ผู้ตัดสินทั้ง 12 คน ร่วมสร้างนิยามความสามารถของนักเรียนใน 1) ระดับดีเยี่ยม 2) ระดับดีมาก 3) ระดับดี 4) ระดับดีพอใช้ 5) ระดับพอใช้ 6) ระดับอ่อน 7) ระดับอ่อนมาก และ 8) ระดับตก ว่านักเรียนแต่ละระดับควรมีความรอบรู้อย่างไร เพื่อสร้างให้เกิดความเข้าใจตรงกันในการพิจารณาตัดสินคะแนนจุดตัดในแต่ละระดับ จากนั้นแบ่งผู้ตัดสินออกเป็น 2 กลุ่มย่อย กลุ่มย่อยละ 6 คน

3. ผู้ตัดสินพิจารณากำหนดมาตรฐาน โดยดำเนินการพิจารณา 3 รอบ ดังนี้

รอบที่ 1 เป้าหมายหลัก คือเพื่อให้ผู้ตัดสินมีความคุ้นเคยกับคู่มือการจัดเรียงข้อสอบ ในรอบนี้จะทำงานเป็นกลุ่มย่อย ผู้ตัดสินจะทำงานอิสระจากกันนั้นคือผู้ตัดสินแต่ละคนทำการ

กำหนดคะแนนจุดตัดที่ระดับพอใจเป็นจุดแรก โดยให้ผู้ตัดสินแต่ละคนพิจารณาข้อสอบทีละข้อในคู่มือจัดเรียงข้อสอบซึ่งพิจารณาตั้งแต่หน้าแรกซึ่งเป็นข้อสอบที่ง่ายที่สุด โดยตอบคำถาม 2 คำถามในแต่ละข้อต่อไปนี้

(1) ในระดับชั้นมัธยมศึกษาปีที่ 3 นักเรียนต้องมีความรู้ความสามารถอะไรเพื่อให้ตอบข้อสอบข้อนี้ถูก ?

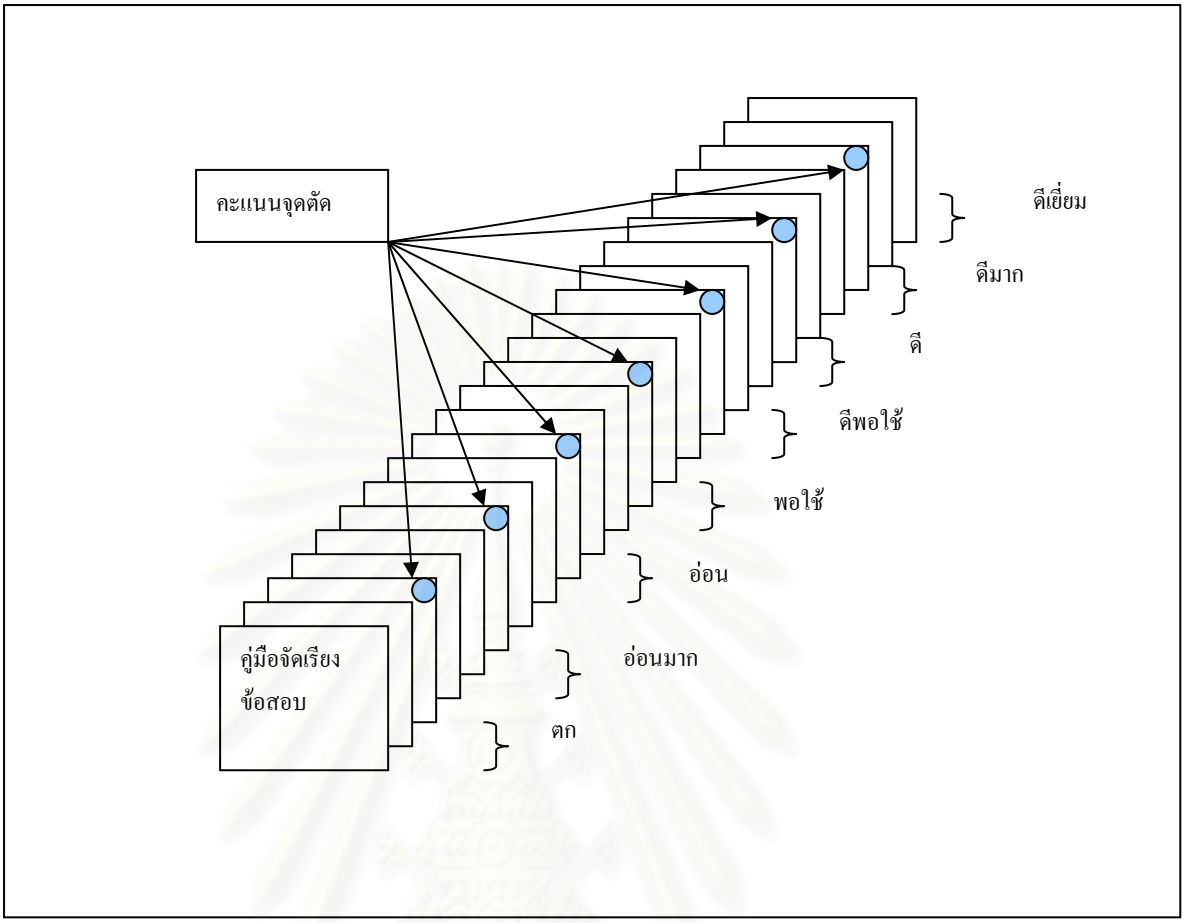
(2) อะไรที่ทำให้ข้อสอบข้อนี้ยากกว่าข้อสอบข้อที่แล้วในคู่มือจัดเรียงข้อสอบ ?

การตอบคำถามเหล่านี้เป็นการเตรียมตัวผู้ตัดสินในการทำการตัดสินหาคะแนนจุดตัดโดยผู้ตัดสินทำการกำหนดบุ๊คมาร์ค(คั่นหนังสือ) บนหน้าที่ผู้ตัดสินคิดว่า *"ผู้ที่คาบเส้นระดับอ่อนมากจะมีโอกาสตอบข้อสอบข้อนั้นถูก 67%"* หรือมีผู้คาบเส้นระดับพอใช้ 100 คน จะมีผู้ตอบถูก 67 คน โดยผู้ตัดสินถูกฝึกให้เข้าใจว่านักเรียนที่คาบเส้นในระดับพอใช้ควรมีโอกาสตอบถูกอย่างน้อย 67% ในข้อสอบหน้าถัดไป และมีโอกาสตอบถูกสูงกว่า 67% ในหน้าก่อนหน้าที่คั่นหนังสือ

หลังจากดำเนินการกำหนดคะแนนจุดตัดที่ระดับอ่อนมากแล้ว ผู้ตัดสินดำเนินการหาคะแนนจุดตัดระดับอ่อน โดยใช้กระบวนการคล้ายกับที่กล่าวมา ในระดับนี้ผู้ตัดสินกำหนดบุ๊คมาร์ค(คั่นหนังสือ) บนหน้าที่คิดว่า *"ผู้ที่คาบเส้นระดับอ่อนมีโอกาสตอบข้อสอบข้อนั้นถูก 67%"* จากนั้นผู้ตัดสินดำเนินการหาคะแนนจุดตัดระดับพอใช้ ดีพอใช้ ระดับดีมาก ระดับดีเยี่ยม ตามลำดับเช่นเดียวกับข้างต้น ผลที่ได้จากการหาคะแนนจุดตัดจะได้ระดับคะแนนจุดตัด 7 ระดับ สำหรับวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 คือ 1) ระดับดีเยี่ยม 2) ระดับดีมาก 3) ระดับดี 4) ระดับดีพอใช้ 5) ระดับพอใช้ 6) ระดับอ่อน 7) ระดับอ่อนมาก

เพื่อให้เห็นภาพที่ชัดเจนขึ้น ผู้วิจัยขอเสนอภาพการกำหนดบุ๊คมาร์ค (การคั่นหนังสือ) ดังแผนภาพที่ 19

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย



แผนภาพที่ 19 ตัวอย่างการกำหนดบุ๊คมาร์ค (การคั่นหนังสือ) 7 ระดับ

รอบที่ 2 จัดให้มีการอภิปรายในกลุ่มย่อย ก่อนการอภิปรายผู้ตัดสินได้รับข้อมูลเลขที่หน้าในคู่มือจัดเรียงข้อสอบที่ผู้ตัดสินคนอื่นกำหนดบุ๊คมาร์ค(คั่นหนังสือ) เมื่อผู้ตัดสินพิจารณาข้อมูลนั้นแล้ว ผู้ตัดสินร่วมกันอภิปรายถึงเหตุผล และความเหมาะสมของคะแนนจุดตัดที่ผู้ตัดสินแต่ละคนตัดสินไว้ โดยเฉพาะให้ผู้ตัดสินคนที่ให้คะแนนจุดตัดสูงสุด และต่ำสุดอภิปรายเหตุผลที่ใช้คะแนนจุดตัดตรงนี้ ซึ่งการอภิปรายอาจอภิปรายในประเด็นตัวข้อสอบเกี่ยวกับระดับทักษะ และเนื้อหาทางวิชาการที่ควรจะรอบรู้สำหรับนักเรียนในแต่ละคะแนนจุดตัด เหตุผลในการอภิปรายก็เพื่อให้ผู้ตัดสินได้รับฟังเหตุผลของผู้ตัดสินคนอื่น การอภิปรายในกลุ่มย่อยนั้น ผู้ตัดสินแสดงความคิดเห็นและประสบการณ์ที่มีความหลากหลาย หลังจากนั้นเปิดโอกาสให้ผู้ตัดสินพิจารณาคู่มือจัดเรียงข้อสอบอีกครั้งหนึ่งโดยนำความคิดเห็นจากผู้ตัดสินคนอื่นมาพิจารณาประกอบด้วย เพื่อหาคะแนนจุดตัดในรอบที่ 2 การพิจารณาคะแนนจุดตัดดำเนินการพิจารณาทีละระดับครบทั้ง 7 ระดับ และให้ผู้ตัดสินแต่ละคนพิจารณาความสามารถของนักเรียนในระดับเขต

พื้นที่การศึกษา ชลบุรี เขต 1 ไม่ให้มองระดับห้องเรียนเหมือนตอนรอบแรก เมื่อได้เลขที่หน้าที่ของผู้ตัดสินแต่ละคนกำหนดขึ้นมาแล้ว จากนั้นเลขที่หน้าจะถูกแปลงเป็นค่าความสามารถ(θ) (เลขที่หน้าที่คั่นคู่มือจัดเรียงข้อสอบไว้นี้ตรงกับค่า θ เท่าไร นำค่าθ เหล่านี้มาหาค่าเฉลี่ยเพื่อนำมาประมาณความสามารถขั้นต่ำของนักเรียน)

รอบที่ 3 นำผู้ตัดสินทั้ง 2 กลุ่มย่อยมารวมกันเป็นกลุ่มใหญ่เพียงกลุ่มเดียว ผู้อำนวยความสะดวกนำข้อมูลผลกระทบ (impact data) คือ เปอร์เซ็นต์ของนักเรียนที่อยู่ในแต่ละระดับ ค่าเฉลี่ยคะแนนจุดตัดจากกลุ่มย่อยทั้งสองกลุ่ม มาแสดงให้ผู้ตัดสินพิจารณา จากนั้นผู้ตัดสินอภิปรายร่วมกันโดยพิจารณาข้อสอบที่อยู่ระหว่างค่าเฉลี่ยคะแนนจุดตัดทั้งสองกลุ่ม แล้วเปิดโอกาสให้ผู้ตัดสินทำการกำหนดคะแนนจุดตัดอีกครั้งอย่างอิสระจากกัน โดยการกำหนดหน้าที่คั่นหนังสือในคู่มือจัดเรียงข้อสอบ เลขที่หน้าที่ของผู้ตัดสินแต่ละคนที่กำหนดขึ้นมานั้นถูกแปลงเป็นค่าความสามารถ(θ) (เลขที่หน้าที่คั่นคู่มือจัดเรียงข้อสอบไว้นี้ตรงกับค่าθ เท่าไร แล้วนำค่าθ เหล่านี้มาหาค่าเฉลี่ยเพื่อนำมาประมาณความสามารถขั้นต่ำของนักเรียน

4. เมื่อได้มาตรฐานที่มีคะแนนจุดตัดทั้ง 7 ระดับแล้ว จากนั้นผู้ตัดสินระดมพลังสมองเพื่อเขียนบรรยายระดับการปฏิบัติของนักเรียนในประเด็นความรู้ ทักษะ และความสามารถของนักเรียน ณ ระดับคะแนนจุดตัดของระดับนั้น รวมทั้งสิ้น 8 ระดับ

การวิเคราะห์ข้อมูล

การกำหนดมาตรฐานหรือคะแนนจุดตัดด้วยโมเดลการตอบสนองข้อสอบ

1. **วิธีการแองกอฟที่ได้รับการปรับปรุง**

ผู้วิจัยประมาณค่าθ^* ของข้อสอบแต่ละข้อโดยอาศัยโค้งลักษณะข้อสอบรายข้อ แล้วจึงคำนวณหา θ^* เฉลี่ยของข้อสอบทุกข้อในแบบสอบ มี 5 ขั้นตอน ดังนี้

1. คำนวณหาค่าเฉลี่ยของความน่าจะเป็นในการตอบข้อสอบ i ได้ถูกโดยใช้ดุลพินิจของผู้เชี่ยวชาญตัดสิน (M_{iR})

2. ให้ M_{iR} เป็นค่าประมาณของความน่าจะเป็นที่ผู้สอบซึ่งมีระดับความสามารถขั้นต่ำสุดที่ยอมให้ผ่าน สามารถตอบข้อสอบได้ถูกในรูปแบบโลจิสติก ซึ่งแทนได้ด้วยโค้งลักษณะข้อสอบที่ระดับความสามารถ θ^*_{iR} ดังสมการ

$$M_{iR} = P_i(\theta^*_{iR})$$

3. ประมาณค่าความแปรปรวนของ θ^*_{iR} ในข้อ i

$$\text{กำหนดให้} \quad P_i(\theta^*_{iR}) = \frac{\Delta M_{iR}}{\Delta \theta_{iR}} \qquad (6)$$

เป็นความชันของโค้งลักษณะข้อสอบ ข้อ i ที่ความสามารถ (θ^*_{iR}) แทนด้วยความเบี่ยงเบนมาตรฐาน

$$\frac{P_i'(\theta^*_{iR})}{\sigma_i(M_{iR})} \cong \frac{1}{\sigma_i(\theta^*_{iR})} \qquad (7)$$

$$\frac{[P_i'(\theta^*_{iR})]^2}{\sigma_i^2(M_{iR})} \cong \frac{1}{\sigma_i^2(\theta^*_{iR})} \qquad (8)$$

4. ประมาณค่า θ^* จากค่าเฉลี่ยของ θ^*_{iR}

$$\theta^* = \frac{1}{n}\sum_i \theta^*_{iR} \qquad (9)$$

ซึ่งมีความแปรปรวน $\sigma^2(\theta^*) = \frac{1}{n^2}\sum_i \sigma_i^2(\theta^*_{iR})$ (10)

5. ดำเนินการประมาณค่าความสามารถ (θ) ในระดับอ่อน ระดับพอใช้ ระดับดีพอใช้ ระดับดี ระดับดีมาก และระดับดีเยี่ยม ทีละระดับตามขั้นตอนที่ 1 - 4

2. วิธีการบุ๊คมาร์ค

นำผลการสอบของนักเรียนมาหาค่าพารามิเตอร์ความยาก และอำนาจจำแนกโดยการวิเคราะห์ตามทฤษฎี IRT ค่าพารามิเตอร์เหล่านี้จะถูกนำมาแปลงเข้าสูตรหาค่า θ จากนั้นนำค่า θ ที่คำนวณได้ไปจัดทำคู่มือจัดเรียงข้อสอบต่อไป

สำหรับสูตรการหาค่า θ ในงานวิจัยนี้ มาจาก Mitzel et al.(2001, cited in Cizek, Bunch & Koons, 2004) ได้เสนอสูตรดั้งเดิมแบบโมเดล 3 พารามิเตอร์ สำหรับข้อสอบแบบเลือกตอบ ดังนี้

$$P_j(\theta) = \frac{c_j + (1 - c_j)}{1 + \exp[-1.7a_j(\theta - b_j)]} \quad (1)$$

เมื่อ

P_j = ความน่าจะเป็นของการตอบถูก

θ = ฟังก์ชันความสามารถของผู้สอบ

b_j = ความยากของข้อสอบ

a_j = อำนาจจำแนกข้อสอบ

c_j = threshold หรือ ตัวแปรการเดา (chance variable)

exp = natural logarithm e (2.71828…)

อย่างไรก็ตาม Mitzel etal (2001 , cited in Cizek, Bunch & Koons, 2004) ได้กำหนดให้ $c_j = 0$ จากสมการ (1) จะได้

$$P_j(\theta) = \frac{1}{1 + \exp[-1.7a_j(\theta - b_j)]} \quad (2)$$

สำหรับงานวิจัยนี้ แบบสอบที่ใช้เป็นการให้คะแนนแบบ 2 ค่า(เช่น แบบเลือกตอบ) และเนื่องจากว่างานวิจัยนี้ได้ตั้งคำถามว่า นักเรียนที่คาบเส้นมีโอกาสตอบข้อสอบถูกที่ความน่าจะเป็น .67 ซึ่งความน่าจะเป็น .67 นี้นำมาแก้สมการในสมการ (2) สำหรับค่า θ ที่ต้องการให้ตอบข้อสอบถูก จะได้

$$\theta = b_j + \frac{.708}{1.7a_j} \quad (3)$$

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

การหาคุณภาพของการกำหนดมาตรฐานด้วยวิธีการแองกอฟที่ได้รับการปรับปรุง และวิธีการบุ๊คมาร์ค คุณภาพของการกำหนดมาตรฐานที่ต้องพิจารณาได้แก่ ความเที่ยงของการกำหนดมาตรฐาน และความตรงของการกำหนดมาตรฐาน

ความเที่ยงของการกำหนดมาตรฐาน

ผู้วิจัยทำการวิเคราะห์เพื่อหาค่าความสอดคล้องหรือสัมประสิทธิ์ความเที่ยงในการตัดสินมาตรฐานหรือ คะแนนจุดตัดจากค่าสัมประสิทธิ์การสรุปอ้างอิง (Generalizability coefficient) โดยใช้โปรแกรม GENOVA โดยมีขั้นตอนดังนี้

1) การออกแบบการวัด แบบ One-facet crossed design

1.1) สิ่งที่ต้องการวัด คือ ผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3

1.2) เอกภพของการสังเกต คือ ฟาเซทของผู้ตัดสิน (R)

1.3) ฟาเซท (facets) ของการวัด มี 1 ฟาเซท คือ ผู้ตัดสิน (r) ฟาเซทของจำนวนผู้ตัดสิน ผู้วิจัยกำหนดจำนวนระดับเป็น 8, 12, 14, 16, 18, 20, 22, 24 คน

1.4) รูปแบบของความสัมพันธ์ระหว่างฟาเซทเป็น i x r นั้นคือ ให้ผู้ตัดสินทุกคน ตัดสินข้อสอบทุกข้อ

2) ประมาณค่าความแปรปรวนจากแหล่งต่างๆ ของ i x r design ขั้นนี้เรียกว่า G-Study สามารถคำนวณได้ดังนี้

$$\sigma_i^2 = \frac{(MS_I - \sigma_e^2)}{n_r}$$

$$\sigma_r^2 = \frac{(MS_R - \sigma_e^2)}{n_i}$$

เมื่อ

σ_i^2 = ค่าประมาณแหล่งความแปรปรวนของข้อสอบ

σ_r^2 = ค่าประมาณแหล่งความแปรปรวนของผู้ตัดสิน

σ_e^2 = ค่าประมาณแหล่งความแปรปรวนของส่วนที่เหลือ

n_r = จำนวนผู้ตัดสิน

n_i = จำนวนข้อสอบ

MS_e = ค่าเฉลี่ยกำลังสอง (Mean square) ของส่วนที่เหลือ

MS_I = ค่าเฉลี่ยกำลังสองที่คาดหมาย(Expected mean square) ของข้อสอบ มีค่าเท่ากับ $\sigma_e^2 + n_r\sigma_i^2$

MS_R = ค่าเฉลี่ยกำลังสองที่คาดหมาย(Expected mean square) ของผู้ตัดสิน มีค่าเท่ากับ $\sigma_e^2 + n_i\sigma_r^2$

3) คำนวณสัมประสิทธิ์การสรุปอ้างอิง ขั้นนี้เรียกว่า D-Study โดยมีขั้นตอนการคำนวณดังนี้

3.1) คำนวณความแปรปรวนของความคลาดเคลื่อน แบ่งเป็น 2 ประเภทคือ

3.1.1) คำนวณแหล่งความแปรปรวนของความคลาดเคลื่อนสัมพัทธ์ (Relative error variance : $\sigma^2_{\mathrm{Re}l}$)

$$\sigma^2_{\mathrm{Re}l} = \frac{\sigma^2_{ir,e}}{n_r}$$

3.1.2) คำนวณแหล่งความแปรปรวนของความคลาดเคลื่อนสัมบูรณ์ (Absolute error variance : $\sigma_{\mathrm{Abs}}{}^2$)

$$\sigma^2_{Abs} = \frac{\sigma^2_r}{n_r} + \frac{\sigma^2_{ir,e}}{n_r}$$

3.2) คำนวณสัมประสิทธิ์การสรุปอ้างอิง แบ่งเป็น 2 ประเภทคือ

3.2.1) คำนวณสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมพัทธ์ (G-Coefficient for relative decision : $\rho^2_{\mathrm{Re}l}$)

$$\rho^2_{\mathrm{Re}l} = \frac{\sigma^2_i}{\sigma^2_i + \sigma^2_{\mathrm{Re}l}}$$

3.2.2) คำนวณสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์ (G-Coefficient for absolute decision : ρ^2_{Abs})

$$\rho^2_{Abs} = \frac{\sigma^2_i}{\sigma^2_i + \sigma^2_{Abs}}$$

เมื่อ $\sigma^2_{ir,e}$ = ค่าประมาณแหล่งความแปรปรวนของส่วนที่เหลือ

$\sigma^2_{\mathrm{Re}l}$ = ผลรวมค่าประมาณแหล่งความแปรปรวนเชิงสัมพัทธ์

σ^2_{Abs} = ผลรวมค่าประมาณแหล่งความแปรปรวนเชิงสัมบูรณ์

ρ^2_{Rel} = สัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมพัทธ์

ρ^2_{Abs} = สัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์

n_r = จำนวนผู้ตัดสิน

นอกจากนี้ผู้วิจัยยังได้ศึกษาผลของจำนวนผู้ตัดสิน และจำนวนครั้ง ที่มีต่อความเที่ยงในการกำหนดมาตรฐานด้วยวิธีการแองกอฟที่ได้รับการปรับปรุง และวิธีการบุ๊คมาร์ค ผู้วิจัยออกแบบการวัดเป็น แบบ Two-facet crossed design ฟาเซท (facets) ของการวัด มี 2 ฟาเซท คือ ผู้ตัดสิน (r) โดยผู้วิจัยกำหนดจำนวนระดับเป็น 8, 12, 14, 16, 18, 20, 22, 24 คน และจำนวนครั้ง (o) กำหนดจำนวนระดับเป็น 1, 2, 3 ครั้ง รูปแบบของความสัมพันธ์ระหว่างฟาเซทเป็น i x rxo นั้นคือ ให้ผู้ตัดสินทุกคนตัดสินข้อสอบทุกข้อ และทำการตัดสินซ้ำ 3 ครั้ง .ในการกำหนดมาตรฐานทั้งสองวิธี

ความตรงของการกำหนดมาตรฐาน

ผู้วิจัยหาความตรงตามสภาพของการกำหนดมาตรฐาน โดยการหาความสัมพันธ์ระหว่างระดับผลการเรียนที่นักเรียนได้รับจากโรงเรียน กับระดับผลการตัดสินที่กำหนดด้วยวิธีการแองกอฟที่ได้รับการปรับปรุง และวิธีบุ๊คมาร์ค คำนวณจากสูตรสหสัมพันธ์แบบ Spearman's Rho

การทดสอบความแตกต่างความเที่ยงระหว่างการกำหนดมาตรฐานด้วยวิธีการแองกอฟที่ได้รับการปรับปรุง และวิธีการบุ๊คมาร์ค

การเปรียบเทียบความแตกต่างของค่าความเที่ยงระหว่างการกำหนดมาตรฐานวิธีการแองกอฟที่ได้รับการปรับปรุง และวิธีการบุ๊คมาร์ค โดยสถิติทดสอบ UX_1 ของวูดรัฟฟ์และ เฟลดท์ (Woodruff and Feldt, 1986) ซึ่งเป็นการทดสอบความแตกต่างของความเที่ยง m ค่า ที่มาจากสูตรเดียวกัน และค่าความเที่ยงเหล่านั้นมาจากกลุ่มตัวอย่างเดียวกัน แต่กำหนดคะแนนจุดตัดต่างวิธีกัน สำหรับสถิติทดสอบสูตรของวูดรัฟฟ์และเฟลดท์ เป็นดังนี้

$$UX_1 = \sum_{i=1}^{m} \frac{(u_i - \bar{u})^2}{s_U^2 - c_U}$$

เมื่อ $$u_i = \left(\frac{1}{1-a_i}\right)^{\frac{1}{3}}$$

$$\bar{u} = \sum_{i=1}^{m} \frac{u_i}{m}$$

$$s_U^2 = \frac{2}{9m(N_C - 1)} \sum_{i=1}^{m} u_i^2$$

$$C_U = \frac{4}{9m(m-1)(N_C - 1)} \sum_{i=2}^{m} \sum_{j=1}^{i=1} \left[r(x_i, x_j)\right]^2 u_i u_j$$

$$N_C = N\left(\frac{n-1}{n+1}\right)$$

เมื่อ UX_i คือ สถิติทดสอบมีการแจกแจงคล้ายกับ χ^2 ที่ df = m-1

N คือ จำนวนผู้ตัดสิน

n คือ จำนวนข้อสอบ

m คือ จำนวนค่าสัมประสิทธิ์การสรุปอ้างอิง

a_i คือ ค่าสัมประสิทธิ์การสรุปอ้างอิง หรือ ค่าสัมประสิทธิ์ความเที่ยง

s_U^2 คือ ความแปรปรวนของ u_i

C_U คือ ค่าปรับแก้ความแปรปรวนของ u_i ในกรณีค่าสัมประสิทธิ์ความเที่ยงคำนวณจากกลุ่มตัวอย่างกลุ่มเดียวกัน

$r(x_i, x_j)$ คือ ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างการกำหนดคะแนนจุดตัดแต่ละวิธีคำนวณโดยใช้สูตร phi correlation

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

การทดสอบความแตกต่างความตรงระหว่างการกำหนดมาตรฐานด้วยวิธีการแองกอฟที่ได้รับการปรับปรุง และวิธีการบุ๊คมาร์ค

การเปรียบเทียบความแตกต่างของค่าความความตรงระหว่างการกำหนดมาตรฐานวิธีการแองกอฟที่ได้รับการปรับปรุง และวิธีการบุ๊คมาร์ค โดยสถิติทดสอบ t ของโฮเทลลิ่ง (Hotelling, 1931, cited in Howell, 1997) ซึ่งเป็นการทดสอบความแตกต่างของค่าสัมประสิทธิ์สหสัมพันธ์ที่มาจากกลุ่มตัวอย่างเดียวกัน สำหรับสถิติทดสอบ t เป็นดังนี้

$$t = (r_{12} - r_{13}) \sqrt{\frac{(N-1)(1+r_{23})}{2\left(\frac{N-1}{N-3}\right)|R| + \frac{(r_{12}+r_{13})^2}{4}(1-r_{23})^3}}$$

เมื่อ

$$|R| = (1 - r_{12}^2 - r_{13}^2 - r_{23}^2) + (2r_{12}r_{13}r_{23})$$

r_{12} = สัมประสิทธิ์สหสัมพันธ์ระหว่างระดับผลการเรียนที่นักเรียนได้รับจากโรงเรียน กับระดับผลการตัดสินที่กำหนดด้วยวิธีการแองกอฟที่ได้รับการปรับปรุง

r_{13} = สัมประสิทธิ์สหสัมพันธ์ระหว่างระดับผลการเรียนที่นักเรียนได้รับจากโรงเรียน กับระดับผลการตัดสินที่กำหนดด้วยวิธีบุ๊คมาร์ค

r_{23} = สัมประสิทธิ์สหสัมพันธ์ระหว่างระดับผลการตัดสินที่กำหนดด้วยวิธีการแองกอฟที่ได้รับการปรับปรุงกับระดับผลการตัดสินที่กำหนดด้วยวิธีบุ๊คมาร์ค

N = จำนวนกลุ่มตัวอย่าง

df = N - 3

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 4

ผลการวิเคราะห์ข้อมูล

การวิจัยนี้มีวัตถุประสงค์ของการวิจัยคือ 1) เพื่อเปรียบเทียบคุณภาพของการกำหนดมาตรฐานด้านความเที่ยง ความตรง และความเหมาะสมและความเป็นไปได้ระหว่างวิธีแองกอฟที่ได้รับการปรับปรุงกับวิธีการบุ๊คมาร์ค 2) เพื่อศึกษาผลของจำนวนผู้ตัดสิน และจำนวนครั้งที่มีต่อการกำหนดมาตรฐานด้วยวิธีการแองกอฟที่ได้รับการปรับปรุง และวิธีการบุ๊คมาร์ค ผู้วิจัยได้นำเสนอผลการวิเคราะห์ข้อมูล 6 ตอน ดังนี้

ตอนที่ 1 ผลการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และวิธีบุ๊คมาร์ค

ตอนที่ 2 ผลการตรวจสอบคุณภาพของการกำหนดมาตรฐานด้านความเที่ยง ด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และวิธีบุ๊คมาร์ค จำแนกตามจำนวนผู้ตัดสิน

ตอนที่ 3 ผลการตรวจสอบคุณภาพของการกำหนดมาตรฐานด้านความตรง ด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และวิธีบุ๊คมาร์ค

ตอนที่ 4 ผลการตรวจสอบความเหมาะสมและความเป็นไปได้ในการนำไปใช้ ด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และวิธีบุ๊คมาร์ค

ตอนที่ 5 ผลการเปรียบเทียบความเที่ยงของการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และวิธีบุ๊คมาร์ค จำแนกตามจำนวนผู้ตัดสิน

ตอนที่ 6 ผลของจำนวนผู้ตัดสิน และจำนวนครั้งที่มีต่อการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และวิธีการบุ๊คมาร์ค

และเพื่อความสะดวกในการนำเสนอผลการวิเคราะห์ข้อมูล ผู้วิจัยได้นำเอาสัญลักษณ์และอักษรย่อมาใช้ดังนี้

SS	ผลรวมความเบี่ยงเบนกำลังสอง (Mean of Squares)
MS	กำลังสองเฉลี่ย
df	ชั้นแห่งความเป็นอิสระ
r	ผู้ตัดสิน (rater)
i	ข้อสอบ (item)
m	วิธีการตัดสินคะแนนจุดตัด (method)
G-study	ขั้นการศึกษาเพื่อการสรุปอ้างอิง (Generalizability Study)
D-study	ขั้นการศึกษาเพื่อการตัดสินใจ (Decision Study)

σ^2_{Rel} ความแปรปรวนของคะแนนความคลาดเคลื่อนสัมพันธ์ (Relative error variance)

σ^2_{Abs} ความแปรปรวนของคะแนนความคลาดเคลื่อนสัมบูรณ์ (Absolute error variance)

$E\rho^2$ สัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจสัมพัทธ์ หรือค่าความสอดคล้องในการตัดสินคะแนนจุดตัด (G-coefficient for relative decision)

ϕ สัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจสัมบูรณ์หรือค่าความสอดคล้องในการตัดสินคะแนนจุดตัด (G-coefficient for absolute decision)

Ω สัมประสิทธิ์การอ้างอิงสรุปสำหรับการตัดสินแบบสัมพัทธ์ (G-coefficient for relative decision) ในลักษณะความตรงลู่เข้า

β สัมประสิทธิ์การอ้างอิงสรุปสำหรับการตัดสินแบบสัมบูรณ์ (G-coefficient for absolute decision) ในลักษณะของความตรงลู่เข้า

N จำนวนผู้ตัดสินในแต่ละกลุ่ม

K จำนวนข้อสอบในแบบสอบ

Z คะแนนมาตรฐานของฟิชเชอร์ ซี

ตอนที่ 1 ผลการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และวิธีบุ๊คมาร์ค

ในการกำหนดมาตรฐาน 8 ระดับผู้ตัดสินทั้งหมด 12 คน ทำการตัดสินหาคะแนนจุดตัดแต่ละมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุงและวิธีบุ๊คมาร์ค โดยกำหนดคะแนนจุดตัดทั้งหมด 7 ระดับคือ ระดับดีเยี่ยม (A) ระดับดีมาก (B+) ระดับดี (B) ระดับดีพอใช้ (C+) ระดับพอใช้ (C) ระดับอ่อน (D+) และระดับอ่อนมาก (D) ผู้ตัดสินทำการตัดสินทั้งสองวิธีรวม 3 ครั้ง แต่ละครั้งทำกำหนดมาตรฐานทั้งสองวิธีโดยมีรูปแบบดังนี้

ครั้งที่ 1 วิธีแองกอฟที่ได้รับการปรับปรุงก่อน แล้วตามด้วยวิธีบุ๊คมาร์ค

ครั้งที่ 2 วิธีบุ๊คมาร์คก่อน แล้วตามด้วยวิธีแองกอฟที่ได้รับการปรับปรุง

ครั้งที่ 3 วิธีแองกอฟที่ได้รับการปรับปรุงก่อน แล้วตามด้วยวิธีบุ๊คมาร์ค

ผู้วิจัยได้นำเสนอผลการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และวิธีบุ๊คมาร์ค ตามหัวข้อต่อไปนี้

1. คะแนนจุดตัดด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และบุ๊คมาร์ค
2. จำนวนนักเรียนผ่านแต่ละระดับมาตรฐาน
3. มาตรฐานการปฏิบัติวิชาคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 3 ภาคเรียนที่ 1 ทั้งหมด 8 ระดับ

1. คะแนนจุดตัดด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และวิธีบุ๊คมาร์ค

ตารางที่ 19 คะแนนจุดตัดของผู้ตัดสินระดับดีเยี่ยม (A) ด้วยวิธีแองกอฟที่ได้รับการปรับปรุงและวิธีบุ๊คมาร์ค

ผู้ตัดสินคนที่	คะแนนจุดตัดระดับดีเยี่ยม(A)					
	การตัดสินครั้งที่ 1		การตัดสินครั้งที่ 2		การตัดสินครั้งที่ 3	
	แองกอฟ	บุ๊คมาร์ค	แองกอฟ	บุ๊คมาร์ค	แองกอฟ	บุ๊คมาร์ค
1	88.69	82	76.85	82	84.85	82
2	54.36	79	56.18	79	56.23	79
3	54.36	81	68.45	81	68.94	81
4	84.85	81	81.03	80	76.89	82
5	85.38	79	59.18	81	69.31	82
6	54.36	81	76.36	81	67.96	81
7	76.36	79	74.80	81	74.80	81
8	95.11	81	67.16	80	69.20	81
9	54.36	81	90.05	81	90.35	81
10	91.20	81	80.88	81	80.88	81
11	80.93	81	88.69	81	85.62	81
12	54.36	79	59.74	79	56.80	79
คะแนนจุดตัดเฉลี่ย	72.860	80.417	73.281	80.583	73.486	80.917
S.D.	16.256	1.037	10.787	0.862	10.376	0.954
(2,22) F-ratio	.012	2.03				
คะแนนจุดตัด	72.86	80.42	73.28	80.58	73.49	80.92
คะแนนความสามารถ	2.1351	2.4043*	2.1538	2.3816*	2.1783	2.4168*

* Theta @ RP=.67

จากตารางที่ 19 การกำหนดคะแนนจุดตัดระดับดีเยี่ยม (A) ด้วยวิธีแองกอฟที่ได้รับการปรับปรุงครั้งที่ 1 ของผู้ตัดสินแต่ละคน พบว่า มีคะแนนจุดตัดอยู่ในช่วง 54.36 ถึง 95.11 คิดเป็นคะแนนจุดตัดเฉลี่ย 72.86 ส่วนเบี่ยงเบนมาตรฐาน 16.256 การตัดสินครั้งที่ 2 มีคะแนนจุดตัดอยู่ในช่วง 56.18 ถึง 90.05 คิดเป็นคะแนนจุดตัดเฉลี่ย 73.281 ส่วนเบี่ยงเบนมาตรฐาน 10.787 และการตัดสินครั้งที่ 3 มีคะแนนจุดตัดอยู่ในช่วง 56.23 ถึง 90.35 คิดเป็นคะแนนจุดตัดเฉลี่ย 73.486 ส่วนเบี่ยงเบนมาตรฐาน 10.376 ส่วนวิธีบุ๊คมาร์คครั้งที่ 1 พบว่า มีคะแนนจุดตัดอยู่ในช่วง 79 ถึง 82 คิดเป็นคะแนนจุดตัดเฉลี่ย 80.417 ส่วนเบี่ยงเบนมาตรฐาน 1.037 การตัดสินครั้งที่ 2 มีคะแนนจุดตัดอยู่ในช่วง 79 ถึง 82 คิดเป็นคะแนนจุดตัดเฉลี่ย 80.583 ส่วนเบี่ยงเบนมาตรฐาน 0.862 และการตัดสินครั้งที่ 3 มีคะแนนจุดตัดอยู่ในช่วง 79 ถึง 82 คิดเป็นคะแนนจุดตัดเฉลี่ย 80.917 ส่วนเบี่ยงเบนมาตรฐาน 0.954 สำหรับการเปรียบเทียบผลการกำหนดคะแนนจุดตัดระดับดีเยี่ยม (A) 3 ครั้ง พบว่า ทั้งสองวิธีมีคะแนนจุดตัดระดับดีเยี่ยม (A) ทั้ง 3 ครั้ง ไม่แตกต่างกัน สำหรับคะแนนความสามารถระดับดีเยี่ยม (A) ที่กำหนดด้วยวิธีแองกอฟที่ได้รับการปรับปรุงทั้ง 3 ครั้งอยู่ในช่วง 2.1351 - 2.4168 ส่วนวิธีบุ๊คมาร์คอยู่ในช่วง 2.3816 - 2.4168

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 20 คะแนนจุดตัดของผู้ตัดสินระดับดีมาก (B+) ด้วยวิธีแองกอฟที่ได้รับการปรับปรุงและวิธีบุ๊คมาร์ค

ผู้ตัดสินคนที่	คะแนนจุดตัดระดับดีมาก (B+)					
	การตัดสินครั้งที่ 1		การตัดสินครั้งที่ 2		การตัดสินครั้งที่ 3	
	แองกอฟ	บุ๊คมาร์ค	แองกอฟ	บุ๊คมาร์ค	แองกอฟ	บุ๊คมาร์ค
1	79.77	66	72.65	66	79.77	66
2	50.25	65	50.68	65	50.95	65
3	50.25	66	58.69	66	58.96	64
4	79.75	66	76.36	65	77.34	66
5	50.25	65	54.18	66	59.20	66
6	69.31	66	69.31	67	54.20	66
7	79.08	66	68.85	66	69.05	66
8	50.25	66	53.65	62	57.80	65
9	73.65	66	73.75	65	73.65	66
10	71.20	66	71.20	66	71.20	66
11	67.67	66	67.67	66	59.32	66
12	50.25	65	47.88	65	47.55	65
คะแนนจุดตัดเฉลี่ย	64.307	65.750	63.739	65.417	63.249	65.583
S.D.	12.433	0.433	9.609	1.187	10.149	0.640
(2,22) F-ratio	.175	.550				
คะแนนจุดตัด	64.31	65.75	63.74	65.42	63.25	65.58
คะแนนความสามารถ	1.6711	2.0453*	1.6550	2.0392*	1.6394	2.9042*

* Theta @ RP=.67

จากตารางที่ 20 การกำหนดคะแนนจุดตัดระดับดีมาก (B+) ด้วยวิธีแองกอฟที่ได้รับการปรับปรุงครั้งที่ 1 ของผู้ตัดสินแต่ละคน พบว่า มีคะแนนจุดตัดอยู่ในช่วง 50.25 ถึง 79.75 คิดเป็นคะแนนจุดตัดเฉลี่ย 64.307 ส่วนเบี่ยงเบนมาตรฐาน 12.433 การตัดสินครั้งที่ 2 มีคะแนนจุดตัดอยู่ในช่วง 47.88 ถึง 76.36 คิดเป็นคะแนนจุดตัดเฉลี่ย 63.739 ส่วนเบี่ยงเบนมาตรฐาน 9.609 และการตัดสินครั้งที่ 3 มีคะแนนจุดตัดอยู่ในช่วง 47.55 ถึง 79.77 คิดเป็นคะแนนจุดตัดเฉลี่ย 63.249 ส่วนเบี่ยงเบนมาตรฐาน 10.149 ส่วนวิธีบุ๊คมาร์คครั้งที่ 1 พบว่า มีคะแนนจุดตัดอยู่ในช่วง 65 ถึง 66 คิดเป็นคะแนนจุดตัดเฉลี่ย 65.750 ส่วนเบี่ยงเบนมาตรฐาน 0.433 การตัดสินครั้งที่ 2 มีคะแนนจุดตัดอยู่ในช่วง 62 ถึง 67 คิดเป็นคะแนนจุดตัดเฉลี่ย 65.417 ส่วน

เบี่ยงเบนมาตรฐาน 1.187 และการตัดสินครั้งที่ 3 มีคะแนนจุดตัดอยู่ในช่วง 64 ถึง 66 คิดเป็นคะแนนจุดตัดเฉลี่ย 65.583 ส่วนเบี่ยงเบนมาตรฐาน 0.640 สำหรับการเปรียบเทียบผลการกำหนดคะแนนจุดตัด 3 ครั้ง ระดับดีมาก (B+) พบว่า ทั้งสองวิธีมีคะแนนจุดตัดระดับดีเยี่ยม (A) ทั้ง 3 ครั้ง ไม่แตกต่างกัน สำหรับคะแนนความสามารถระดับดีมาก (B+) ที่กำหนดด้วยวิธีแองกอฟที่ได้รับการปรับปรุงทั้ง 3 ครั้งอยู่ในช่วง 1.6394-1.6711 ส่วนวิธีบุ๊คมาร์คอยู่ในช่วง 2.0392-2.0453

ตารางที่ 21 คะแนนจุดตัดของผู้ตัดสินระดับดี (B) ด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และวิธีบุ๊คมาร์ค

ผู้ตัดสินคนที่	คะแนนจุดตัดระดับดี (B)					
	การตัดสินครั้งที่ 1		การตัดสินครั้งที่ 2		การตัดสินครั้งที่ 3	
	แองกอฟ	บุ๊คมาร์ค	แองกอฟ	บุ๊คมาร์ค	แองกอฟ	บุ๊คมาร์ค
1	69.79	47	64.65	47	69.79	47
2	47.60	48	45.59	48	47.60	48
3	47.60	47	48.84	47	47.60	47
4	74.26	48	70.56	48	74.26	48
5	47.60	48	49.24	48	47.60	48
6	63.10	48	63.10	49	63.10	48
7	55.00	48	58.85	48	55.00	48
8	47.60	48	45.51	48	47.60	48
9	60.00	46	60.00	45	60.00	46
10	57.67	46	59.48	46	57.67	46
11	56.29	48	56.29	48	56.29	48
12	47.60	48	41.24	48	47.60	48
คะแนนจุดตัดเฉลี่ย	56.176	47.500	55.279	47.500	56.176	47.500
S.D.	8.880	0.764	8.663	1.041	8.880	0.764
(2,22) F-ratio	1.044	.000				
คะแนนจุดตัด	56.18	47.50	55.28	47.50	56.18	47.50
คะแนนความสามารถ	1.3070	2.0453*	1.2702	2.0392*	1.1643	2.0429*

* Theta @ RP=.67

จากตารางที่ 21 การกำหนดคะแนนจุดตัดระดับดี (B) ด้วยวิธีแองกอฟที่ได้รับการปรับปรุงครั้งที่ 1 ของผู้ตัดสินแต่ละคน พบว่า มีคะแนนจุดตัดอยู่ในช่วง 47.60 ถึง 69.79 คิดเป็นคะแนนจุดตัดเฉลี่ย 56.176 ส่วนเบี่ยงเบนมาตรฐาน 8.880 การตัดสินครั้งที่ 2 มีคะแนนจุดตัดอยู่ในช่วง 41.24 ถึง 70.56 คิดเป็นคะแนนจุดตัดเฉลี่ย 55.279 ส่วนเบี่ยงเบนมาตรฐาน 8.663 และการตัดสินครั้งที่ 3 มีคะแนนจุดตัดอยู่ในช่วง 47.60 ถึง 69.79 คิดเป็นคะแนนจุดตัดเฉลี่ย 56.176 ส่วนเบี่ยงเบนมาตรฐาน 8.880 ส่วนวิธีบุ๊คมาร์คครั้งที่ 1 พบว่า มีคะแนนจุดตัดอยู่ในช่วง 46 ถึง 48 คิดเป็นคะแนนจุดตัดเฉลี่ย 47.500 ส่วนเบี่ยงเบนมาตรฐาน 0.764 การตัดสินครั้งที่ 2 มีคะแนนจุดตัดอยู่ในช่วง 45 ถึง 48 คิดเป็นคะแนนจุดตัดเฉลี่ย 47.500 ส่วนเบี่ยงเบนมาตรฐาน 1.041 และการตัดสินครั้งที่ 3 มีคะแนนจุดตัดอยู่ในช่วง 46 ถึง 48 คิดเป็นคะแนนจุดตัดเฉลี่ย 47.500 ส่วนเบี่ยงเบนมาตรฐาน 0.764 สำหรับการเปรียบเทียบผลการกำหนดคะแนนจุดตัด 3 ครั้ง ระดับดี (B) พบว่า ทั้งสองวิธีมีคะแนนจุดตัดทั้ง 3 ครั้ง ไม่แตกต่างกัน สำหรับคะแนนความสามารถระดับดีมาก (B+) ที่กำหนดด้วยวิธีแองกอฟที่ได้รับการปรับปรุงทั้ง 3 ครั้งอยู่ในช่วง 1.1643-1.3070 ส่วนวิธีบุ๊คมาร์คอยู่ในช่วง 1.7434-1.7460

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 22 คะแนนจุดตัดของผู้ตัดสินระดับดีพอใช้ (C+) ด้วยวิธีแองกอฟที่ได้รับการปรับปรุงและวิธีบุ๊คมาร์ค

ผู้ตัดสินคนที่	คะแนนจุดตัดดีพอใช้ (C+)					
	การตัดสินครั้งที่ 1		การตัดสินครั้งที่ 2		การตัดสินครั้งที่ 3	
	แองกอฟ	บุ๊คมาร์ค	แองกอฟ	บุ๊คมาร์ค	แองกอฟ	บุ๊คมาร์ค
1	43.07	35	39.1	35	43.07	35
2	34.36	25	35.68	34	36.12	34
3	34.36	33	39.03	33	38.95	33
4	64.32	35	62.85	34	56.27	34
5	34.36	34	40.18	34	38.85	34
6	56.10	35	56.10	35	28.39	35
7	45.12	34	57.10	34	57.01	34
8	34.36	34	37.10	34	37.10	34
9	44.53	33	44.53	32	44.53	32
10	48.52	33	48.52	33	48.47	33
11	47.01	33	47.01	33	35.03	33
12	34.36	34	34.08	34	33.77	34
คะแนนจุดตัดเฉลี่ย	43.373	33.167	45.107	33.750	41.463	33.750
S.D.	9.359	2.577	8.977	0.829	8.437	0.829
(2,22) F-ratio	1.235	.569				
คะแนนจุดตัด	43.37	33.17	45.11	33.75	41.46	33.75
คะแนนความสามารถ	0.7448	1.5054*	0.8248	1.5148*	0.6617	1.5148*

** Theta @ RP=.67

จากตารางที่ 22 การกำหนดคะแนนจุดตัดระดับดีพอใช้ (C+) ด้วยวิธีแองกอฟที่ได้รับการปรับปรุงครั้งที่ 1 ของผู้ตัดสินแต่ละคน พบว่า มีคะแนนจุดตัดอยู่ในช่วง 34.36 ถึง 64.32 คิดเป็นคะแนนจุดตัดเฉลี่ย 43.373 ส่วนเบี่ยงเบนมาตรฐาน 9.359 การตัดสินครั้งที่ 2 มีคะแนนจุดตัดอยู่ในช่วง 34.08 ถึง 62.85 คิดเป็นคะแนนจุดตัดเฉลี่ย 45.107 ส่วนเบี่ยงเบนมาตรฐาน 8.977 และการตัดสินครั้งที่ 3 มีคะแนนจุดตัดอยู่ในช่วง 33.77 ถึง 57.01 คิดเป็นคะแนนจุดตัดเฉลี่ย 41.463 ส่วนเบี่ยงเบนมาตรฐาน 8.437 ส่วนวิธีบุ๊คมาร์คครั้งที่ 1 พบว่า มีคะแนนจุดตัดอยู่ในช่วง 25 ถึง 35 คิดเป็นคะแนนจุดตัดเฉลี่ย 33.167 ส่วนเบี่ยงเบนมาตรฐาน 2.577 การตัดสินครั้งที่ 2 มีคะแนนจุดตัดอยู่ในช่วง 32 ถึง 35 คิดเป็นคะแนนจุดตัดเฉลี่ย 33.750 ส่วน

เบี่ยงเบนมาตรฐาน 0.829 และการตัดสินครั้งที่ 3 มีคะแนนจุดตัดอยู่ในช่วง 33 ถึง 35 คิดเป็นคะแนนจุดตัดเฉลี่ย 33.750 ส่วนเบี่ยงเบนมาตรฐาน 0.829 สำหรับการเปรียบเทียบผลการกำหนดคะแนนจุดตัด 3 ครั้ง ระดับดีพอใช้ (C+) พบว่า ทั้งสองวิธีมีคะแนนจุดตัดทั้ง 3 ครั้ง ไม่แตกต่างกัน สำหรับคะแนนความสามารถระดับดีพอใช้ (C+) ที่กำหนดด้วยวิธีแองกอฟที่ได้รับการปรับปรุงทั้ง 3 ครั้งอยู่ในช่วง 0.6617-0.8248 ส่วนวิธีบุ๊คมาร์คอยู่ในช่วง 1.5148-1.5054

ตารางที่ 23 คะแนนจุดตัดของผู้ตัดสินระดับพอใช้ (C) ด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และวิธีบุ๊คมาร์ค

ผู้ตัดสินคนที่	คะแนนจุดตัดพอใช้ (C)					
	การตัดสินครั้งที่ 1		การตัดสินครั้งที่ 2		การตัดสินครั้งที่ 3	
	แองกอฟ	บุ๊คมาร์ค	แองกอฟ	บุ๊คมาร์ค	แองกอฟ	บุ๊คมาร์ค
1	33.27	24	30.78	24	33.27	24
2	25.15	24	26.44	24	26.80	24
3	25.15	24	29.65	24	29.63	24
4	54.60	24	54.13	24	54.90	24
5	25.15	24	40.18	26	20.66	25
6	51.15	24	51.15	24	22.15	24
7	36.90	24	37.25	24	37.45	24
8	25.15	24	30.81	29	30.84	24
9	30.73	24	30.56	24	30.61	30
10	34.73	24	34.73	24	34.7	24
11	35.90	24	35.90	24	24.83	24
12	25.15	24	26.50	24	26.16	28
คะแนนจุดตัดเฉลี่ย	33.586	24.000	35.673	24.583	31.000	24.917
S.D.	9.712	0.000	8.580	1.441	8.651	1.891
(2,22) F-ratio	1.899	1.180				
คะแนนจุดตัด	33.59	24.00	35.67	24.58	31.00	24.92
คะแนนความสามารถ	0.2897	1.3030*	0.3929	1.3161*	0.1611	1.3267*

* Theta @ RP=.67

จากตารางที่ 23 การกำหนดคะแนนจุดตัดระดับพอใช้ (C) ด้วยวิธีแองกอฟที่ได้รับการปรับปรุงครั้งที่ 1 ของผู้ตัดสินแต่ละคน พบว่า มีคะแนนจุดตัดอยู่ในช่วง 25.15 ถึง 54.60 คิดเป็นคะแนนจุดตัดเฉลี่ย 33.586 ส่วนเบี่ยงเบนมาตรฐาน 9.712 การตัดสินครั้งที่ 2 มีคะแนนจุดตัดอยู่ในช่วง 26.44 ถึง 54.13 คิดเป็นคะแนนจุดตัดเฉลี่ย 35.673 ส่วนเบี่ยงเบนมาตรฐาน 8.580 และการตัดสินครั้งที่ 3 มีคะแนนจุดตัดอยู่ในช่วง 20.66 ถึง 54.90 คิดเป็นคะแนนจุดตัดเฉลี่ย 31.000 ส่วนเบี่ยงเบนมาตรฐาน 8.651 ส่วนวิธีบุ๊คมาร์คครั้งที่ 1 ของผู้ตัดสินแต่ละคน พบว่า มีคะแนนจุดตัดอยู่ที่ 24 คิดเป็นคะแนนจุดตัดเฉลี่ย 24.000 ส่วนเบี่ยงเบนมาตรฐาน 0.000 การตัดสินครั้งที่ 2 มีคะแนนจุดตัดอยู่ในช่วง 24 ถึง 29 คิดเป็นคะแนนจุดตัดเฉลี่ย 24.583 ส่วนเบี่ยงเบนมาตรฐาน 1.441 และการตัดสินครั้งที่ 3 มีคะแนนจุดตัดอยู่ในช่วง 24 ถึง 30 คิดเป็นคะแนนจุดตัดเฉลี่ย 24.917 ส่วนเบี่ยงเบนมาตรฐาน 1.891 สำหรับการเปรียบเทียบผลการกำหนดคะแนนจุดตัด 3 ครั้ง ระดับพอใช้ (C) พบว่า ทั้งสองวิธีมีคะแนนจุดตัดทั้ง 3 ครั้ง ไม่แตกต่างกัน สำหรับคะแนนความสามารถระดับพอใช้ (C) ที่กำหนดด้วยวิธีแองกอฟที่ได้รับการปรับปรุงทั้ง 3 ครั้งอยู่ในช่วง 0.1611-0.3929 ส่วนวิธีบุ๊คมาร์คอยู่ในช่วง 1.3030-1.3267

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 24 คะแนนจุดตัดของผู้ตัดสินระดับอ่อน (D+) ด้วยวิธีแองกอฟที่ได้รับการปรับปรุงและวิธีบุ๊คมาร์ค

ผู้ตัดสินคนที่	คะแนนจุดตัดระดับอ่อน (D+)					
	การตัดสินครั้งที่ 1		การตัดสินครั้งที่ 2		การตัดสินครั้งที่ 3	
	แองกอฟ	บุ๊คมาร์ค	แองกอฟ	บุ๊คมาร์ค	แองกอฟ	บุ๊คมาร์ค
1	23.35	14	22.40	14	23.37	14
2	20.62	15	21.88	15	21.92	15
3	20.62	15	25.13	15	25.08	15
4	17.30	15	17.18	15	18.47	16
5	20.62	15	30.38	15	12.19	15
6	46.34	15	46.34	15	16.90	15
7	14.59	15	19.95	15	17.2	15
8	20.62	15	21.76	15	21.82	15
9	17.96	14	17.87	14	17.96	14
10	20.20	14	20.18	14	20.50	14
11	13.29	15	13.29	15	14.40	15
12	20.62	15	18.89	15	19.02	15
คะแนนจุดตัดเฉลี่ย	21.344	14.750	22.938	14.750	19.069	14.833
S.D.	8.016	0.433	8.142	0.433	3.551	0.553
(2,22) F-ratio	1.452	1.000				
คะแนนจุดตัด	21.34	14.75	22.94	14.75	19.07	14.83
คะแนนความสามารถ	-0.4222	0.8590*	-0.3114	0.8590*	-0.5878	0.8601*

* Theta @ RP=.67

จากตารางที่ 24 การกำหนดคะแนนจุดตัดระดับอ่อน (D+) ด้วยวิธีแองกอฟที่ได้รับการปรับปรุงครั้งที่ 1 ของผู้ตัดสินแต่ละคน พบว่า มีคะแนนจุดตัดอยู่ในช่วง 13.29 ถึง 46.34 คิดเป็นคะแนนจุดตัดเฉลี่ย 21.344 ส่วนเบี่ยงเบนมาตรฐาน 8.016 การตัดสินครั้งที่ 2 มีคะแนนจุดตัดอยู่ในช่วง 13.29 ถึง 46.34 คิดเป็นคะแนนจุดตัดเฉลี่ย 22.938 ส่วนเบี่ยงเบนมาตรฐาน 8.142 และการตัดสินครั้งที่ 3 มีคะแนนจุดตัดอยู่ในช่วง 12.19 ถึง 25.08 คิดเป็นคะแนนจุดตัดเฉลี่ย 19.069 ส่วนเบี่ยงเบนมาตรฐาน 3.551 ส่วนวิธีบุ๊คมาร์คครั้งที่ 1 พบว่า มีคะแนนจุดตัดอยู่ในช่วง 14 ถึง 15 คิดเป็นคะแนนจุดตัดเฉลี่ย 14.750 ส่วนเบี่ยงเบนมาตรฐาน 0.433 การตัดสินครั้งที่ 2 มีคะแนนจุดตัดอยู่ในช่วง 14 ถึง 15 คิดเป็นคะแนนจุดตัดเฉลี่ย 14.750 ส่วนเบี่ยงเบนมาตรฐาน

0.433 และการตัดสินครั้งที่ 3 มีคะแนนจุดตัดอยู่ในช่วง 14 ถึง 15 คิดเป็นคะแนนจุดตัดเฉลี่ย 14.833 ส่วนเบี่ยงเบนมาตรฐาน 0.553 สำหรับการเปรียบเทียบผลการกำหนดคะแนนจุดตัด 3 ครั้ง ระดับอ่อน (D+) พบว่า ทั้งสองวิธีมีคะแนนจุดตัดทั้ง 3 ครั้ง ไม่แตกต่างกัน สำหรับคะแนนความสามารถระดับอ่อน (D+) ที่กำหนดด้วยวิธีแองกอฟที่ได้รับการปรับปรุงทั้ง 3 ครั้งอยู่ในช่วง -0.3114 ถึง -0.5878 ส่วนวิธีบุ๊คมาร์คอยู่ในช่วง 0.8590 ถึง 0.8601

ตารางที่ 25 คะแนนจุดตัดของผู้ตัดสินระดับอ่อนมาก (D) ด้วยวิธีแองกอฟที่ได้รับการปรับปรุงและวิธีบุ๊คมาร์ค

ผู้ตัดสินคนที่	คะแนนจุดตัดระดับอ่อนมาก (D)					
	การตัดสินครั้งที่ 1		การตัดสินครั้งที่ 2		การตัดสินครั้งที่ 3	
	แองกอฟ	บุ๊คมาร์ค	แองกอฟ	บุ๊คมาร์ค	แองกอฟ	บุ๊คมาร์ค
1	13.39	8	12.56	8	13.57	8
2	17.03	8	17.86	8	17.42	8
3	17.03	7	20.62	7	20.62	7
4	11.84	8	10.28	8	14.16	8
5	17.03	8	20.93	8	6.27	9
6	11.20	9	11.04	9	12.14	9
7	10.14	8	7.43	8	7.60	8
8	17.03	8	14.13	8	14.15	8
9	9.42	7	9.24	7	9.44	7
10	6.05	7	6.04	7	7.66	7
11	8.54	8	8.54	8	6.16	8
12	17.03	8	11.44	8	11.42	8
คะแนนจุดตัดเฉลี่ย	12.978	7.833	12.509	7.833	11.718	7.917
S.D.	3.820	0.553	4.746	0.553	4.345	0.640
(2,22) F-ratio	.669	1.000				
คะแนนจุดตัด	12.98	7.83	12.51	7.83	11.72	7.92
คะแนนความสามารถ	-1.2770	0.2060*	-1.2193	0.2060*	-1.3065	0.2218*

* Theta @ RP=.67

จากตารางที่ 25 การกำหนดคะแนนจุดตัดระดับอ่อนมาก (D) ด้วยวิธีแองกอฟที่ได้รับการปรับปรุงครั้งที่ 1 ของผู้ตัดสินแต่ละคน พบว่า มีคะแนนจุดตัดอยู่ในช่วง 6.05 ถึง 17.03 คิดเป็นคะแนนจุดตัดเฉลี่ย 12.978 ส่วนเบี่ยงเบนมาตรฐาน 3.820 การตัดสินครั้งที่ 2 มีคะแนนจุดตัดอยู่ในช่วง 6.04 ถึง 20.93 คิดเป็นคะแนนจุดตัดเฉลี่ย 12.509 ส่วนเบี่ยงเบนมาตรฐาน 4.746 และการตัดสินครั้งที่ 3 มีคะแนนจุดตัดอยู่ในช่วง 6.16 ถึง 20.62 คิดเป็นคะแนนจุดตัดเฉลี่ย 11.718 ส่วนเบี่ยงเบนมาตรฐาน 4.345 ส่วนวิธีบุ๊คมาร์คครั้งที่ 1 ของผู้ตัดสินแต่ละคน พบว่า มีคะแนนจุดตัดอยู่ในช่วง 7 ถึง 8 คิดเป็นคะแนนจุดตัดเฉลี่ย 7.833 ส่วนเบี่ยงเบนมาตรฐาน 0.553 การตัดสินครั้งที่ 2 มีคะแนนจุดตัดอยู่ในช่วง 7 ถึง 8 คิดเป็นคะแนนจุดตัดเฉลี่ย 7.833 ส่วนเบี่ยงเบนมาตรฐาน 0.553 และการตัดสินครั้งที่ 3 มีคะแนนจุดตัดอยู่ในช่วง 7 ถึง 9 คิดเป็นคะแนนจุดตัดเฉลี่ย 7.917 ส่วนเบี่ยงเบนมาตรฐาน 0.640 สำหรับการเปรียบเทียบผลการกำหนดคะแนนจุดตัด 3 ครั้ง ระดับอ่อนมาก (D) พบว่า ทั้งสองวิธีมีคะแนนจุดตัดทั้ง 3 ครั้ง ไม่แตกต่างกัน สำหรับคะแนนความสามารถระดับอ่อนมาก (D) ที่กำหนดด้วยวิธีแองกอฟที่ได้รับการปรับปรุงทั้ง 3 ครั้งอยู่ในช่วง -1.2193 ถึง -1.3065 ส่วนวิธีบุ๊คมาร์คอยู่ในช่วง 0.2060 ถึง 0.2218

2. จำนวนนักเรียนผ่านแต่ละระดับมาตรฐาน

ตารางที่ 26 จำนวนนักเรียนผ่านแต่ละระดับมาตรฐานครั้งที่ 1 ครั้งที่ 2 และครั้งที่ 3 ด้วยการกำหนดมาตรฐานวิธีการแองกอฟที่ได้รับการปรับปรุง

ระดับมาตรฐาน	ครั้งที่ 1		ครั้งที่ 2		ครั้งที่ 3	
	จำนวนนักเรียน	ร้อยละ	จำนวนนักเรียน	ร้อยละ	จำนวนนักเรียน	ร้อยละ
ตก (F)	4	0.41	4	0.41	4	0.41
อ่อนมาก(D)	167	16.92	207	20.97	104	10.54
อ่อน(D+)	419	42.45	425	43.06	393	39.82
พอใช้(C)	179	18.14	160	16.21	235	23.81
ดีพอใช้(C+)	125	12.66	92	9.32	130	13.17
ดี(B)	37	3.75	40	4.05	62	6.28
ดีมาก(B+)	26	2.63	33	3.34	33	3.34
ดีเยี่ยม(A)	30	3.04	26	2.63	26	2.63
รวม	987	100	987	100	987	100

ตารางที่ 27 จำนวนนักเรียนผ่านแต่ละระดับมาตรฐานครั้งที่ 1 ครั้งที่ 2 และครั้งที่ 3 ด้วยการกำหนดมาตรฐานวิธีการบุ๊คมาร์ค

ระดับมาตรฐาน	ครั้งที่ 1		ครั้งที่ 2		ครั้งที่ 3	
	จำนวนนักเรียน	ร้อยละ	จำนวนนักเรียน	ร้อยละ	จำนวนนักเรียน	ร้อยละ
ตก (F)	0	0	0	0	0	0
อ่อนมาก(D)	14	1.418	14	1.4184	14	1.418
อ่อน(D+)	233	23.61	275	27.86	275	27.86
พอใช้(C)	343	34.75	301	30.5	301	30.5
ดีพอใช้(C+)	232	23.51	232	23.51	232	23.51
ดี(B)	112	11.35	115	11.65	109	11.04
ดีมาก(B+)	48	4.863	45	4.559	51	5.167
ดีเยี่ยม(A)	5	0.507	5	0.507	5	0.507
รวม	987	100	987	100	987	100

3. มาตรฐานในการปฏิบัติวิชาคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 3 ภาคเรียนที่ 1 ทั้งหมด 8 ระดับ

เมื่อผู้ตัดสินกำหนดคะแนนจุดตัดทั้ง 7 ระดับเรียบร้อยแล้ว ผู้ตัดสินทั้ง 12 คน ได้ระดมพลังสมอง เพื่อเขียนบรรยายมาตรฐานการปฏิบัติของนักเรียนชั้นมัธยมศึกษาปีที่ 3 ภาคเรียนที่ 1 ได้มาตรฐานทั้งหมด 8 ระดับ ดังต่อไปนี้

1. **<u>ระดับตก</u>** (F) หมายถึง นักเรียนที่มีความสามารถไม่ถึงระดับที่คาดหวังขั้นต่ำสำหรับการปฏิบัติของนักเรียนในวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ตามหลักสูตรกระทรวงศึกษาธิการ โดยทำเนื้อหาในบทเรียนวิชานี้**ไม่ได้** กล่าวคือ ไม่สามารถอ่านโจทย์ และตีความหมายได้ ไม่สามารถคิดวิเคราะห์ได้ ไม่สามารถเขียนสื่อความได้ ขาดทักษะในการคิดคำนวณ และขาดความรอบคอบในการคิดคำนวณ นักเรียนระดับนี้ไม่มีแรงจูงใจ ไม่มีความสนใจ และไม่มีความรับผิดชอบในการทำงานวิชานี้

2. **<u>ระดับอ่อนมาก</u>** (D) หมายถึง นักเรียนที่มีความสามารถ**ถึงระดับที่คาดหวังขั้นต่ำ** สำหรับการปฏิบัติของนักเรียนในวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ตามหลักสูตรกระทรวงศึกษาธิการนั้นคือ นักเรียนมีความรู้ที่จำเป็นต้องมี และมีทักษะขั้นพื้นฐานในการปฏิบัติงานในระดับคาบเส้น โดยทำเนื้อหาในบทเรียนวิชานี้**ได้บ้างเล็กน้อย** กล่าวคือ สามารถ

คำนวณโจทย์พื้นฐานง่ายๆ ที่ไม่พลิกแพลงได้ สามารถลอกเลียนแบบจากตัวอย่างได้โดยมีเพื่อนหรือครูแนะนำ แต่ไม่สามารถเชื่อมโยงความสัมพันธ์ และไม่สามารถประยุกต์ใช้ได้ คิดคำนวณช้ามาก ขาดความรอบคอบทำให้คำนวณผิดพลาด นักเรียนระดับนี้มีแรงจูงใจ ความสนใจ และความรับผิดชอบที่จะทำงานในวิชานี้น้อย

3. **<u>ระดับอ่อน</u>** (D+) หมายถึง นักเรียนที่มีความสามารถ**ถึงระดับที่คาดหวังขั้นต่ำขึ้นมาเล็กน้อย** สำหรับการปฏิบัติของนักเรียนในวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ตามหลักสูตรกระทรวงศึกษาธิการ นั้นคือ นักเรียนมีความรู้ที่จำเป็นต้องมี และมีทักษะขั้นพื้นฐานในการปฏิบัติงานเกินระดับคาบเส้นขึ้นมาไม่มาก โดยทำเนื้อหาในบทเรียนวิชานี้**ได้น้อย** กล่าวคือ สามารถอ่าน และเข้าใจโจทย์พื้นฐานง่ายๆ สามารถลอกเลียนแบบจากตัวอย่างได้ แต่ต้องใช้เวลานานกว่าปกติ คำนวณหรือคิดได้บ้างจากโจทย์ที่ไม่ได้พลิกแพลง ต้องได้รับการแนะนำชี้แจงบ้างเพื่อกระตุ้นเพื่อให้เกิดความมั่นใจ แต่คิดคำนวณช้า นักเรียนระดับนี้ไม่ค่อยมีความสนใจที่จะทำงานในวิชานี้ อย่างไรก็ตามก็มีความรับผิดชอบบ้าง ถ้ามีแรงจูงใจ

4. **<u>ระดับพอใช้</u>** (C) หมายถึง นักเรียนที่มีความสามารถ**ถึงระดับที่คาดหวัง**สำหรับการปฏิบัติของนักเรียนในวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ตามหลักสูตรกระทรวงศึกษาธิการ สามารถทำเนื้อหาในบทเรียนวิชานี้ได้บ้าง กล่าวคือ อ่านและเข้าใจโจทย์ เรื่องราว สถานการณ์ง่ายๆ ที่ไม่ซับซ้อนในระดับพื้นฐานโดยเป็นโจทย์ที่ไม่พลิกแพลงมากนักได้ สามารถนำเนื้อหาไปประยุกต์ใช้ในชีวิตประจำวันได้บ้าง การคิดคำนวณค่อนข้างช้า มีความรอบคอบน้อย แต่นักเรียนระดับนี้มีแรงจูงใจ ความสนใจ และความรับผิดชอบที่จะทำงานในวิชานี้บ้าง

5. **<u>ระดับดีพอใช้</u>** (C+) หมายถึง นักเรียนที่มีความสามารถ**ถึงระดับที่ความคาดหวังไว้**สำหรับการปฏิบัติของนักเรียนในวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ตามหลักสูตรกระทรวงศึกษาธิการ ระดับนี้เป็นเป้าหมายระยะยาวสำหรับการปฏิบัติของนักเรียน นั้นคือ นักเรียนควรมีความรู้ที่แน่นพอสมควร สามารถประยุกต์ความรู้ในสถานการณ์อื่นที่ไม่ซับซ้อนและมีทักษะการวิเคราะห์ในเนื้อหานั้น มีความคิดสร้างสรรค์บ้าง สามารถทำเนื้อหาในบทเรียนวิชานี้พอได้โดยเป็นโจทย์พื้นฐานที่ไม่พลิกแพลง แต่ไม่สามารถทำโจทย์ปัญหาที่ซับซ้อนนอกเหนือจากบทเรียนได้ คิดคำนวณไม่เร็วไม่ช้านัก คำตอบอาจผิดพลาดบ้างถ้าให้เวลาจำกัด นักเรียนระดับนี้มีแรงจูงใจ และความรับผิดชอบที่จะทำงานในวิชานี้ดี

6. **<u>ระดับดี</u>** (B) หมายถึง นักเรียนที่มีความสามารถถึงระดับที่**เหนือความคาดหวัง**สำหรับการปฏิบัติของนักเรียนในวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ตามหลักสูตรกระทรวงศึกษาธิการ นั้นคือ นักเรียนมีความรู้แน่น สามารถทำเนื้อหาในบทเรียนวิชานี้ได้กล่าวคือ สามารถประยุกต์ความรู้ในสถานการณ์อื่นได้เป็นอย่างดี มีความสามารถเข้าใจโจทย์

เรื่องราว สถานการณ์ในระดับซับซ้อนได้บ้าง สามารถแก้โจทย์ปัญหา เรื่องราว สถานการณ์ในบทเรียนได้อย่างสร้างสรรค์ มีลำดับขั้นตอน พร้อมแสดงเหตุผลได้ถูกต้อง และมีทักษะการวิเคราะห์ สังเคราะห์ในเนื้อหานั้น มีความคิดสร้างสรรค์ คิดคำนวณได้เร็วแต่ยังผิดพลาดเนื่องจากขาดความรอบคอบอยู่บ้าง นักเรียนระดับนี้มีแรงจูงใจ ความสนใจ และความรับผิดชอบที่จะทำงานในวิชานี้ค่อนข้างสูง

7. **<u>ระดับดีมาก</u>** (B+) หมายถึง นักเรียนที่มีความสามารถอยู่**เหนือกว่าระดับที่คาดหวังไว้มาก**สำหรับการปฏิบัติของนักเรียนในวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ตามหลักสูตรกระทรวงศึกษาธิการ นั้นคือ นักเรียนมีความรู้แน่น กล่าวคือ สามารถเข้าใจโจทย์เรื่องราว สถานการณ์ที่ซับซ้อนได้ มีลำดับขั้นตอนในการแก้ปัญหา ชัดเจน พร้อมแสดงเหตุผลประกอบได้ถูกต้อง สามารถประยุกต์ความรู้ในสถานการณ์อื่นได้ดีมาก และคล่องแคล่ว มีทักษะการคิดวิเคราะห์ สังเคราะห์ในเนื้อหานั้นได้อย่างรวดเร็ว มีความคิดสร้างสรรค์ นักเรียนระดับนี้มีความสนใจศึกษาหาความรู้เพิ่มเติมจากข้างนอก มีความรู้ในเนื้อหาวิชาแน่น คิดคำนวณได้เร็วคล่องแคล่ว ถูกต้องแม่นยำพอสมควร แต่ขาดความรอบคอบเป็นบ้างครั้ง นอกจากนี้นักเรียนมีแรงจูงใจ ชอบคิดเลข และมีความรับผิดชอบต่อวิชานี้สูง

8. **<u>ระดับดีเยี่ยม</u>** (A) หมายถึง นักเรียนที่มีความสามารถอยู่**เหนือกว่าระดับที่คาดหวังไว้สูงมาก**สำหรับการปฏิบัติของนักเรียนในวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ตามหลักสูตรกระทรวงศึกษาธิการ นั้นคือ นักเรียนมีความรู้ในเนื้อหาที่ลุ่มลึก กล่าวคือ สามารถวิเคราะห์โจทย์ในทุกสถานการณ์ แยกแยะจับประเด็นสำคัญ เปรียบเทียบหาข้อสรุปได้ มีความชำนาญแก้โจทย์ปัญหาที่ซับซ้อนได้ สามารถแสดงวิธีทำ การแก้ปัญหาที่มีลำดับขั้นตอนที่ชัดเจน พร้อมแสดงเหตุผลประกอบอย่างถูกต้อง ชัดเจน ตรงประเด็น สามารถสร้างแบบจำลองทางคณิตศาสตร์ได้อย่างดีเยี่ยม นำความรู้ในเนื้อหาไปสร้างวิธีการใหม่ๆ ที่ช่วยในการคิดคำนวณเพื่อให้ได้คำตอบถูกต้องรวดเร็ว สมเหตุสมผล สามารถประยุกต์ความรู้ในสถานการณ์อื่นได้ดีเยี่ยม คล่องแคล่ว และมีทักษะการวิเคราะห์ สังเคราะห์ในเนื้อหานั้น มีความคิดสร้างสรรค์ คาดเดา และมองแนวโน้มที่จะเกิดขึ้นได้อย่างดีเยี่ยม มีการเชื่อมโยงความรู้ทางคณิตศาสตร์เข้ากับศาสตร์สาขาอื่นเพื่อเกิดวิธีการใหม่ๆขึ้นมา คิดคำนวณอย่างรอบคอบ และคิดได้เร็วถูกต้องแม่นยำ มีความรับผิดชอบสูงมาก สนใจศึกษาหาความรู้เพิ่มเติมจากข้างนอกตลอดเวลา

ตอนที่ 2 ผลการตรวจสอบคุณภาพของการกำหนดมาตรฐานด้านความเที่ยงด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และวิธีบุ๊คมาร์ค จำแนกตามจำนวนผู้ตัดสิน

การกำหนดมาตรฐานทั้งสองวิธีคือ วิธีแองกอฟที่ได้รับการปรับปรุงและวิธีบุ๊คมาร์ค โดยผู้ตัดสินทำการตัดสินคะแนนจุดตัดทั้งหมด 7 ระดับคือ ระดับดีเยี่ยม (A) ระดับดีมาก (B+) ระดับดี (B) ระดับดีพอใช้ (C+) ระดับพอใช้ (C) ระดับอ่อน (D+) และระดับอ่อนมาก (D) โดยทำการกำหนดมาตรฐานทั้งสองวิธีรวม 3 ครั้ง แต่ละครั้งมีรูปแบบดังนี้ ครั้งที่ 1 กำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุงก่อน แล้วตามด้วยวิธีบุ๊คมาร์ค ครั้งที่ 2 กำหนดมาตรฐานด้วยวิธีบุ๊คมาร์คก่อนแล้วตามด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และครั้งที่ 3 กำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุงก่อน แล้วตามด้วยวิธีบุ๊คมาร์ค สำหรับรายละเอียดของผลการวิเคราะห์ ตอนที่ 2 ผู้วิจัยนำเสนอเป็น 2 หัวข้อดังนี้

1. การประมาณค่าความเที่ยงของการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง
2. การประมาณค่าความเที่ยงของการกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์ค

1. การประมาณค่าความเที่ยงของการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง

ในการประมาณค่าความเที่ยงของการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง ทั้ง 3 ครั้ง เป็นการคำนวณหาค่าสัมประสิทธิ์การสรุปอ้างอิงโดยใช้ทฤษฎีการสรุปอ้างอิง (Generalizability Theory) ลักษณะ cross design ที่มีรูปแบบ i x r design นั้นคือ ผู้ตัดสินทั้งหมด 12 คน ตัดสินข้อสอบทั้งหมด 100 ข้อ รายละเอียดผลการวิเคราะห์มีดังนี้

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 28 การวิเคราะห์ความแปรปรวนและการประมาณค่าความแปรปรวนของการกำหนดมาตรฐาน ด้วยวิธีแองกอฟที่ได้รับการปรับปรุงที่คะแนนจุดตัดระดับดีเยี่ยม (A) ในขั้นสรุปอ้างอิง (G-study)

Source of Variation	df	Sum of Squares	Mean Squares	Estimated Variance Component	Percentage of Total Variance
ครั้งที่ 1					
Item (i)	99	85453.3133	863.1648	64.3362	14.5281
Rater (r)	11	317114.2400	28828.5673	287.3744	64.8934
Residual (ir,e)	1089	99240.9267	91.1303	91.1303	20.5786
Total	1199	501808.4800		442.8409	100
ครั้งที่ 2					
Item (i)	99	110393.0625	1115.0814	83.5577	25.5607
Rater (r)	11	145283.2092	13207.5645	130.9518	40.0588
Residual (ir,e)	1089	122391.7075	112.3891	112.3891	34.3804
Total	1199	378067.9792		326.8986	100
ครั้งที่ 3					
Item (i)	99	101253.0092	1022.7577	76.4697	25.6618
Rater (r)	11	129195.7692	11745.0699	116.3995	39.0615
Residual (ir,e)	1089	114476.9808	105.1212	105.1212	35.2767
Total	1199	344925.7592		297.9904	100

จากตารางที่ 28 เมื่อผู้ตัดสินกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุงที่คะแนนจุดตัดระดับดีเยี่ยม (A) ครั้งที่ 1 โดยที่เอกภพของการสังเกตประกอบด้วย ข้อสอบ และผู้ตัดสิน ซึ่งมีขนาดไม่จำกัด โดยเป็นการประมาณค่าความแปรปรวนของแหล่งต่าง ๆ เมื่อเทียบกับความแปรปรวนรวม มีค่าดังนี้ ความแปรปรวนของข้อสอบ (σ_i^2) เป็นความแตกต่างของความยากของข้อสอบมีค่า

64.3362 คิดเป็นร้อยละ 15 ของความแปรปรวนรวม ความแปรปรวนของผู้ตัดสิน(σ_r^2) เป็นการประมาณค่าความแปรปรวนของคะแนนเอกภพที่แสดงถึงความแตกต่างในการตัดสินคะแนนจุดตัดของผู้ตัดสิน มีค่า 287.3744 คิดเป็นร้อยละ 65 ของความแปรปรวนรวม และความแปรปรวนของส่วนที่เหลือ (σ_{ir}^2) เป็นความแปรปรวนของผลร่วมระหว่างข้อสอบกับผู้ตัดสิน แสดงให้เห็นถึงความไม่คงเส้นคงวาของผู้ตัดสินแต่ละคนในการตัดสินข้อสอบแต่ละข้อ มีค่า 91.1303 คิดเป็นร้อยละ 21 ของความแปรปรวนรวม

เมื่อผู้ตัดสินกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุงที่คะแนนจุดตัดระดับดีเยี่ยม (A) ครั้งที่ 2 มีค่าดังนี้ ความแปรปรวนของข้อสอบ (σ_i^2) เป็นความแตกต่างของความยากของข้อสอบมีค่า 83.5577 คิดเป็นร้อยละ 26 ของความแปรปรวนรวม ความแปรปรวนของผู้ตัดสิน (σ_r^2) เป็นการประมาณค่าความแปรปรวนของคะแนนเอกภพที่แสดงถึงความแตกต่างในการตัดสินคะแนนจุดตัดของผู้ตัดสิน มีค่า 130.9518 คิดเป็นร้อยละ 40 ของความแปรปรวนรวม และความแปรปรวนของส่วนที่เหลือ (σ_{ir}^2) เป็นความแปรปรวนของผลร่วมระหว่างข้อสอบกับผู้ตัดสิน แสดงให้เห็นถึงความไม่คงเส้นคงวาของผู้ตัดสินแต่ละคนในการตัดสินข้อสอบแต่ละข้อมีค่า 112.3891 คิดเป็นร้อยละ 34 ของความแปรปรวนรวม

เมื่อผู้ตัดสินกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุงที่คะแนนจุดตัดระดับดีเยี่ยม (A) ครั้งที่ 3 มีค่าดังนี้ ความแปรปรวนของข้อสอบ (σ_i^2) เป็นความแตกต่างของความยากของข้อสอบมีค่า 76.4697 คิดเป็นร้อยละ 26 ของความแปรปรวนรวม ความแปรปรวนของผู้ตัดสิน (σ_r^2) เป็นการประมาณค่าความแปรปรวนของคะแนนเอกภพที่แสดงถึงความแตกต่างในการตัดสินคะแนนจุดตัดของผู้ตัดสิน มีค่า 116.3995 คิดเป็นร้อยละ 39 ของความแปรปรวนรวม และความแปรปรวนของส่วนที่เหลือ (σ_{ir}^2) เป็นความแปรปรวนของผลร่วมระหว่างข้อสอบกับผู้ตัดสิน แสดงให้เห็นถึงความไม่คงเส้นคงวาของผู้ตัดสินแต่ละคนในการตัดสินข้อสอบแต่ละข้อมีค่า 105.1212 คิดเป็นร้อยละ 35.2767 ของความแปรปรวนรวม

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 29 การประมาณค่าความแปรปรวนในขั้นการสรุปอ้างอิง (G-study) ขั้นการตัดสินใจ (D-study) และสัมประสิทธิ์การอ้างอิงสรุปของการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุงที่คะแนนจุดตัดระดับดีเยี่ยม (A) จำแนกตามจำนวนผู้ตัดสิน

Source of Variation	Estimated G Study Variance Components	Alternative D Studies Design Variance Component 8	12	14	16	18	20	22	24
Item (i)									
ครั้งที่ 1	64.3362	64.3362	64.3362	64.3362	64.3362	64.3362	64.3362	64.3362	64.3362
ครั้งที่ 2	83.5577	83.5577	83.5577	83.5577	83.5577	83.5577	83.5577	83.5577	83.5577
ครั้งที่ 3	76.4697	76.4697	76.4697	76.4697	76.4697	76.4697	76.4697	76.4697	76.4697
Rater (r)									
ครั้งที่ 1	287.3744	35.9218	23.9479	20.5267	17.9609	15.9652	14.3687	13.0625	11.9739
ครั้งที่ 2	130.9581	16.3690	10.9126	9.3537	8.1845	7.2751	6.5476	5.9524	5.4563
ครั้งที่ 3	116.3995	14.5499	9.7000	8.3143	7.2750	6.4666	5.8200	5.2909	4.8500
Residual (ir,e)									
ครั้งที่ 1	91.1303	11.3913	7.5942	6.5093	5.6957	5.0628	4.5565	4.14229	3.7971
ครั้งที่ 2	112.3891	14.0486	9.3658	8.0278	7.0243	6.2438	5.6195	5.1086	4.6829
ครั้งที่ 3	105.1212	13.1402	8.7601	7.5087	6.5700	5.8401	5.2561	4.7782	4.3801
σ^2_{Rel}	ครั้งที่ 1	11.3913	7.5942	6.5093	5.6957	5.0628	4.5565	4.14229	3.7971
	ครั้งที่ 2	14.0486	9.3658	8.0278	7.0243	6.2438	5.6195	5.1086	4.6829
	ครั้งที่ 3	13.1402	8.7601	7.5087	6.5700	5.8401	5.2561	4.7782	4.3801
σ^2_{Abs}	ครั้งที่ 1	47.3131	31.5421	27.0360	23.6566	21.0280	18.9252	17.2048	15.771
	ครั้งที่ 2	30.4176	20.2784	17.3815	15.2088	13.5189	12.1671	11.0610	10.1392
	ครั้งที่ 3	27.6901	18.4601	15.8230	13.8450	12.3067	11.0761	10.0691	9.2301
$E\rho^2$	ครั้งที่ 1	0.8496	0.8944	0.9081	0.9187	0.9271	0.9339	0.9395	0.9443
	ครั้งที่ 2	0.8561	0.8992	0.9124	0.9225	0.9305	0.9370	0.9424	0.9469
	ครั้งที่ 3	0.8534	0.8972	0.9106	0.9209	0.9291	0.9357	0.9412	0.9458
ϕ	ครั้งที่ 1	0.5762	0.6710	0.7041	0.7311	0.7537	0.7727	0.7890	0.8031
	ครั้งที่ 2	0.7331	0.8047	0.8278	0.8460	0.8607	0.8729	0.8831	0.8918
	ครั้งที่ 3	0.7342	0.8055	0.8286	0.8467	0.8614	0.8735	0.8837	0.8923

จากตารางที่ 29 ผลการศึกษา D ประกอบด้วยการประมาณค่าคามแปรปรวนของผู้ตัดสิน จำนวน 8 , 12, 14, 16, 18, 20, 22, 24 คน จากการตัดสินคะแนนจุดตัดของผู้ตัดสินจำนวน 12 คน ด้วยวิธี แองกอฟที่ได้รับการปรับปรุง ที่ระดับดีเยี่ยม (A) ครั้งที่ 1 ความแปรปรวนของคะแนนความคลาดเคลื่อนสัมพันธ์ ($\sigma^2_{\mathrm{Re}l}$) ซึ่งเป็นความคลาดเคลื่อนที่เกิดจากผลร่วมของข้อสอบ และผู้ตัดสิน มีค่าลดลง เมื่อจำนวนผู้ติดสินเพิ่มขึ้น มีค่า 11.3913 7.5942 6.5093 5.6957 5.0628 4.5565 4.14229 3.7971 ครั้งที่ 2 มีค่า ดังนี้ 14.0486 9.3658 8.0278 7.0243 6.2438 5.6195 5.1086 4.6829 ครั้งที่ 3 มีค่า ดังนี้ 13.1402 8.7601 7.5087 6.5700 5.8401 5.2561 4.7782 4.3801 เมื่อมีผู้ตัดสิน จำนวน 8 , 12, 14, 16, 18, 20, 22, 24 คน ตามลำดับ

ส่วนความแปรปรวนของคะแนนความคลาดเคลื่อนสัมบูรณ์ (σ^2_{Abs})ซึ่งเป็นผลรวมของความคลาดเคลื่อนที่เกิดจากผลร่วมของข้อสอบและผู้ตัดสิน มีค่าลดลงเมื่อจำนวนผู้ตัดสินเพิ่มขึ้น ครั้งที่ 1 มีค่า 47.3131 31.5421 27.0360 23.6566 21.0280 18.9252 17.2048 ครั้งที่ 2 มีค่าดังนี้ 30.4176 20.2784 17.3815 15.2088 13.5189 12.1671 11.0610 10.1392 ครั้งที่ 3 มีค่าดังนี้ 27.6901 18.4601 15.8230 13.8450 12.3067 11.0761 10.0691 9.2301 เมื่อมีผู้ตัดสินจำนวน 8 , 12, 14, 16, 18, 20, 22, 24 คน ตามลำดับ

ค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินแบบสัมพัทธ์ ($\mathrm{E}\rho^2$) มีค่าเพิ่มขึ้นเมื่อจำนวนผู้ตัดสินเพิ่มขึ้น ครั้งที่ 1 มีค่าดังนี้ 0.8496 0.8944 0.9081 0.9187 0.9271 0.9339 0.9395 0.9443 ครั้งที่ 2 มีค่าดังนี้ 0.8561 0.8992 0.9124 0.9225 0.9305 0.9370 0.9424 0.9469 ครั้งที่ 3 มีค่าดังนี้ 0.8534 0.8972 0.9106 0.9209 0.9291 0.9357 0.9412 0.9458 เมื่อมีผู้ตัดสินจำนวน 8 , 12, 14, 16, 18, 20, 22, 24 คน ตามลำดับ

ส่วนค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินแบบสัมบูรณ์ (ϕ) มีค่าเพิ่มขึ้นเมื่อจำนวนผู้ตัดสินเพิ่มขึ้น ครั้งที่ 1 มีค่าดังนี้ 0.5762 0.6710 0.7041 0.7311 0.7537 0.7727 0.7890 0.8031 ครั้งที่ 2 มีค่าดังนี้ 0.7331 0.8047 0.8278 0.8460 0.8607 0.8729 0.8831 0.8918 ครั้งที่ 3 มีค่าดังนี้ 0.7342 0.8055 0.8286 0.8467 0.8614 0.8735 0.8837 0.8923 เมื่อมีผู้ตัดสินจำนวน 8 , 12, 14, 16, 18, 20, 22, 24 คน ตามลำดับ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 30 การวิเคราะห์ความแปรปรวนและการประมาณค่าความแปรปรวนของการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุงที่คะแนนจุดตัดระดับดีมาก (B+) ในขั้นสรุปอ้างอิง (G-study)

Source of Variation	df	Sum of Squares	Mean Squares	Estimated Variance Component	Percentage of Total Variance
ครั้งที่ 1					
Item (i)	99	48308.6467	487.9661	37.1318	14.9890
Rater (r)	11	185497.3267	16863.3933	168.2101	67.9015
Residual (ir,e)	1089	46157.1733	42.3849	42.3849	17.1095
Total	1199	279963.1467		247.7268	100
ครั้งที่ 2					
Item (i)	99	123972.9367	1252.2519	94.4182	28.0827
Rater (r)	11	136131.0367	12375.5488	122.5632	36.4538
Residual (ir,e)	1089	129845.4633	119.2337	119.2337	35.4635
Total	1199			336.2151	100
ครั้งที่ 3					
Item (i)	99	112133.9158	1132.6659	85.0499	27.5808
Rater (r)	11	123607.7492	11237.0681	111.2500	36.0772
Residual (ir,e)	1089	122040.8342	112.0669	112.0669	36.3421
Total	1199	357782.4992		308.3668	100

จากตารางที่ 30 เมื่อผู้ตัดสินกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุงที่คะแนนจุดตัดระดับดีมาก (B+) ครั้งที่ 1 โดยที่เอกภพของการสังเกตประกอบด้วย ข้อสอบ และผู้ตัดสิน ซึ่งมีขนาดไม่จำกัด โดยเป็นการประมาณค่าความแปรปรวนของแหล่งต่างๆ เมื่อเทียบกับความแปรปรวนรวม มีค่าดังนี้ ความแปรปรวนของข้อสอบ (σ_i^2) เป็นความแตกต่างของความยากของข้อสอบมีค่า 37.1318 คิดเป็นร้อยละ 15 ของความแปรปรวนรวม ความแปรปรวนของผู้ตัดสิน (σ_r^2) เป็นการประมาณค่าความแปรปรวนของคะแนนเอกภพที่แสดงถึงความแตกต่างในการตัดสินคะแนนจุดตัดของผู้ตัดสิน มีค่า 168.2101

คิดเป็นร้อยละ 68 ของความแปรปรวนรวม และความแปรปรวนของส่วนที่เหลือ (σ_{ir}^2) เป็นความแปรปรวนของผลร่วมระหว่างข้อสอบกับผู้ตัดสิน แสดงให้เห็นถึง ความไม่คงเส้นคงวาของผู้ตัดสินแต่ละคนในการตัดสินข้อสอบแต่ละข้อมีค่า 42.3849 คิดเป็นร้อยละ 17 ของความแปรปรวนรวม

เมื่อผู้ตัดสินกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุงที่คะแนนจุดตัดระดับดีมาก (B+) ครั้งที่ 2 มีค่าดังนี้ ความแปรปรวนของข้อสอบ (σ_i^2) เป็นความแตกต่างของความยากของข้อสอบมีค่า 94.4182 คิดเป็นร้อยละ 28 ของความแปรปรวนรวม ความแปรปรวนของผู้ตัดสิน (σ_r^2) เป็นการประมาณค่าความแปรปรวนของคะแนนเอกภพที่แสดงถึงความแตกต่างในการตัดสินคะแนนจุดตัดของผู้ตัดสิน มีค่า 122.5632 คิดเป็นร้อยละ 37 ของความแปรปรวนรวม และความแปรปรวนของส่วนที่เหลือ (σ_{ir}^2) เป็นความแปรปรวนของผลร่วมระหว่างข้อสอบกับผู้ตัดสิน แสดงให้เห็นถึงความไม่คงเส้นคงวาของผู้ตัดสินแต่ละคนในการตัดสินข้อสอบแต่ละข้อมีค่า 19.2337 คิดเป็นร้อยละ 36 ของความแปรปรวนรวม

เมื่อผู้ตัดสินกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุงที่คะแนนจุดตัดระดับดีมาก (B+) ครั้งที่ 3 ความแปรปรวนของข้อสอบ (σ_i^2) เป็นความแตกต่างของความยากของข้อสอบมีค่า 85.0499 คิดเป็นร้อยละ 28 ของความแปรปรวนรวม ความแปรปรวนของผู้ตัดสิน (σ_r^2) เป็นการประมาณค่าความแปรปรวนของคะแนนเอกภพที่แสดงถึงความแตกต่างในการตัดสินคะแนนจุดตัดของผู้ตัดสิน มีค่า 111.2500 คิดเป็นร้อยละ 36 ของความแปรปรวนรวม และความแปรปรวนของส่วนที่เหลือ (σ_{ir}^2) เป็นความแปรปรวนของผลร่วมระหว่างข้อสอบกับผู้ตัดสิน แสดงให้เห็นถึง ความไม่คงเส้นคงวาของผู้ตัดสินแต่ละคนในการตัดสินข้อสอบแต่ละข้อ มีค่า 112.0669 คิดเป็นร้อยละ 36 ของความแปรปรวนรวม

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 31 การประมาณค่าความแปรปรวนในขั้นการสรุปอ้างอิง (G-study) ขั้นการตัดสินใจ (D-study) และสัมประสิทธิ์การอ้างอิงสรุปของการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุงที่คะแนนจุดตัดระดับดีมาก (B+) จำแนกตามจำนวนผู้ตัดสิน

Source of Variation	Estimated G Study Variance Components	Alternative Estimated D Studies Design Variance Component 8	12	14	16	18	20	22	24
Item (i)									
ครั้งที่ 1	37.1318	37.1318	37.1318	37.1318	37.1318	37.1318	37.1318	37.1318	37.1318
ครั้งที่ 2	94.4182	94.4182	94.4182	94.4182	94.4182	94.4182	94.4182	94.4182	94.4182
ครั้งที่ 3	85.0499	85.0499	85.0499	85.0499	85.0499	85.0499	85.0499	85.0499	85.0499
Rater (r)									
ครั้งที่ 1	168.2101	21.0263	14.0175	12.0150	10.5131	9.3450	8.4105	7.6459	7.0088
ครั้งที่ 2	122.5632	15.3204	10.2136	8.7545	7.6602	6.8091	6.1282	5.5711	5.1068
ครั้งที่ 3	111.2500	13.9063	9.2708	7.9464	6.9531	6.1806	5.5625	5.0568	4.6354
Residual (ir,e)									
ครั้งที่ 1	42.3849	5.2981	3.5321	3.0275	2.6491	2.3547	2.1193	1.9266	1.7660
ครั้งที่ 2	199.2337	14.9042	9.9361	8.5167	7.4521	6.6241	5.9617	5.4197	4.9681
ครั้งที่ 3	112.0669	14.0084	9.3389	8.0048	7.0042	6.2259	5.6033	5.0904	4.6695
σ^2_{Rel}	ครั้งที่ 1	5.2981	3.5321	3.0275	2.6491	2.3547	2.1193	1.9266	1.7660
	ครั้งที่ 2	14.9042	9.9361	8.5167	7.4521	6.6241	5.9617	5.4197	4.9681
	ครั้งที่ 3	14.0084	9.3389	8.0048	7.0042	6.2259	5.6033	5.0904	4.6695
σ^2_{Abs}	ครั้งที่ 1	35.0438	17.5496	15.0425	13.1622	11.6997	10.5298	9.5725	8.7748
	ครั้งที่ 2	30.2246	20.1497	17.2712	15.1123	13.4332	12.0899	10.9908	10.0749
	ครั้งที่ 3	27.9147	18.6097	15.9512	13.9573	12.4065	11.1658	10.1508	9.3049
$E\rho^2$	ครั้งที่ 1	0.8751	0.9131	0.9246	0.9334	0.9404	0.9460	0.9507	0.9546
	ครั้งที่ 2	0.8637	0.9048	0.9173	0.9269	0.9344	0.9406	0.9457	0.9500
	ครั้งที่ 3	0.8586	0.9011	0.9140	0.9240	0.9318	0.9382	0.9435	0.9480
ϕ	ครั้งที่ 1	0.5852	0.6791	0.7117	0.7383	0.7604	0.7791	0.7950	0.8089
	ครั้งที่ 2	0.7575	0.8241	0.8454	0.8620	0.8755	0.8865	0.8957	0.9036
	ครั้งที่ 3	0.7529	0.8205	0.8421	0.8590	0.872 7	0.8840	0.8834	0.9014

จากตารางที่ 31 ผลการศึกษา D ประกอบด้วยการประมาณค่าความแปรปรวนของผู้ตัดสิน จำนวน 8, 12, 14, 16, 18, 20, 22, 24 คน จากการตัดสินคะแนนจุดตัดของผู้ตัดสินจำนวน 12 คน ด้วยวิธีแองกอฟที่ได้รับการปรับปรุง ที่ระดับดีมาก (B+) ครั้งที่ 1 ความแปรปรวนของคะแนนความคลาดเคลื่อนสัมพันธ์($\sigma^2_{\mathrm{Re}l}$) ซึ่งเป็นความคลาดเคลื่อนที่เกิดจากผลร่วมของข้อสอบ และผู้ตัดสิน มีค่าลดลง เมื่อจำนวนผู้ตัดสินเพิ่มขึ้น มีค่า ดังนี้ 5.2981 3.5321 3.0275 2.6491 2.3547 2.1193 1.9266 1.7660 ครั้งที่ 2 มีค่า ดังนี้ 14.9042 9.9361 8.5167 7.4521 6.6241 5.9617 5.4197 4.9681 ครั้งที่ 3 มีค่า ดังนี้ 14.0084 9.3389 8.0048 7.0042 6.2259 5.6033 5.0904 4.6695 เมื่อมีผู้ตัดสินจำนวน 8, 12, 14, 16, 18, 20, 22, 24 คน ตามลำดับ

ส่วนความแปรปรวนของคะแนนความคลาดเคลื่อนสัมบูรณ์(σ^2_{Abs})ซึ่งเป็นผลรวมของความคลาดเคลื่อนที่เกิดจากผลร่วมของข้อสอบและผู้ตัดสิน มีค่าลดลงเมื่อจำนวนผู้ตัดสินเพิ่มขึ้น ครั้งที่ 1 มีค่าดังนี้ 35.0438 17.5496 15.0425 13.1622 11.6997 10.5298 9.5725 8.7748 ครั้งที่ 2 มีค่าดังนี้ 30.2246 20.1497 17.2712 15.1123 13.4332 12.0899 10.9908 10.0749 ครั้งที่ 3 มีค่าดังนี้ 27.9147 18.6097 15.9512 13.9573 12.4065 11.1658 10.1508 9.3049 เมื่อมีผู้ตัดสินจำนวน 8, 12, 14, 16, 18, 20, 22, 24 คน ตามลำดับ

ค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินแบบสัมพัทธ์ ($\mathrm{E}\rho^2$) มีค่าเพิ่มขึ้นเมื่อจำนวนผู้ตัดสินเพิ่มขึ้น ครั้งที่ 1 มีค่าดังนี้ 0.8751 0.9131 0.9246 0.9334 0.9404 0.9460 0.9507 0.9546 ครั้งที่ 2 มีค่าดังนี้ 0.8637 0.9048 0.9173 0.9269 0.9344 0.9406 0.9457 0.9500 ครั้งที่ 3 มีค่าดังนี้ 0.8586 0.9011 0.9140 0.9240 0.9318 0.9382 0.9435 0.9480 เมื่อมีผู้ตัดสินจำนวน 8, 12, 14, 16, 18, 20, 22, 24 คน ตามลำดับ

ส่วนค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินแบบสัมบูรณ์ (ϕ) มีค่าเพิ่มขึ้นเมื่อจำนวนผู้ตัดสินเพิ่มขึ้น ครั้งที่ 1 มีค่าดังนี้ 0.5852 0.6791 0.7117 0.7383 0.7604 0.7791 0.7950 0.8089 ครั้งที่ 2 มีค่าดังนี้ 0.7575 0.8241 0.8454 0.8620 0.8755 0.8865 0.8957 0.9036 ครั้งที่ 3 มีค่าดังนี้ 0.7529 0.8205 0.8421 0.8590 0.872 7 0.8840 0.8834 0.9014 เมื่อมีผู้ตัดสินจำนวน 8, 12, 14, 16, 18, 20, 22, 24 คน ตามลำดับ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 32 การวิเคราะห์ความแปรปรวนและการประมาณค่าความแปรปรวนของการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุงที่คะแนนจุดตัดระดับดี (B) ในขั้นสรุปอ้างอิง (G-study)

Source of Variation	df	Sum of Squares	Mean Squares	Estimated Variance Component	Percentage of Total Variance
ครั้งที่ 1					
Item (i)	99	44453.8158	449.0284	32.9158	19.0881
Rater (r)	11	94630.3692	8602.7608	85.4872	49.5747
Residual (ir,e)	1089	58847.7142	54.0383	54.0383	31.3372
Total	1199	197931.8992		172.4413	100
ครั้งที่ 2					
Item (i)	99	139662.9492	1410.7369	83.2155	13.5516
Rater (r)	11	135102.5692	12282.0517	118.6990	19.3300
Residual (ir,e)	1089	448832.1808	412.1508	412.1508	67.1184
Total	1199	723597.6992		614.0653	100
ครั้งที่ 3					
Item (i)	99	108972.7300	1100.7347	82.6834	29.1223
Rater (r)	11	103164.5967	9378.5997	92.7007	32.6506
Residual (ir,e)	1089	118193.0700	108.5336	108.5336	38.2271
Total	1199	330330.3967		283.9177	100

จากตารางที่ 32 เมื่อผู้ตัดสินกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุงที่คะแนนจุดตัดระดับดี (B) ครั้งที่ 1 โดยที่เอกภพของการสังเกตประกอบด้วย ข้อสอบ และผู้ตัดสิน ซึ่งมีขนาดไม่จำกัด โดยเป็นการประมาณค่าความแปรปรวนของแหล่งต่างๆ เมื่อเทียบกับความแปรปรวนรวม มีค่าดังนี้ ความแปรปรวนของข้อสอบ (σ_i^2) เป็นความแตกต่างของความยากของข้อสอบมีค่า 32.9158 คิดเป็นร้อยละ 19 ของความแปรปรวนรวม ความแปรปรวนของผู้ตัดสิน (σ_r^2) เป็นการประมาณค่าความแปรปรวนของคะแนนเอกภพที่แสดงถึงความแตกต่างในการตัดสินคะแนนจุดตัดของผู้ตัดสิน มีค่า 85.4872 คิดเป็นร้อยละ 50 ของความแปรปรวนรวม และความแปรปรวนของส่วนที่เหลือ (σ_{ir}^2) เป็นความแปรปรวนของ

ผลร่วมระหว่างข้อสอบกับผู้ตัดสิน แสดงให้เห็นถึงความไม่คงเส้นคงวาของผู้ตัดสินแต่ละคนในการตัดสินข้อสอบแต่ละข้อมีค่า 54.0383 คิดเป็นร้อยละ 31 ของความแปรปรวนรวม

เมื่อผู้ตัดสินกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุงที่คะแนนจุดตัดระดับดี (B) ครั้งที่ 2 มีค่าดังนี้ ความแปรปรวนของข้อสอบ (σ_i^2) เป็นความแตกต่างของความยากของข้อสอบมีค่า 83.2155 คิดเป็นร้อยละ 14 ของความแปรปรวนรวม ความแปรปรวนของผู้ตัดสิน (σ_r^2) เป็นการประมาณค่าความแปรปรวนของคะแนนเอกภพที่แสดงถึงความแตกต่างในการตัดสินคะแนนจุดตัดของผู้ตัดสินมีค่า 118.6990 คิดเป็นร้อยละ 19 ของความแปรปรวนรวม และความแปรปรวนของส่วนที่เหลือ (σ_{ir}^2) เป็นความแปรปรวนของผลร่วมระหว่างข้อสอบกับผู้ตัดสิน แสดงให้เห็นถึงความไม่คงเส้นคงวาของผู้ตัดสินแต่ละคนในการตัดสินข้อสอบแต่ละข้อมีค่า 412.1508 คิดเป็นร้อยละ 67 ของความแปรปรวนรวม

เมื่อผู้ตัดสินกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุงที่คะแนนจุดตัดระดับดี (B) ครั้งที่ 3 มีค่าดังนี้ ความแปรปรวนของข้อสอบ (σ_i^2) เป็นความแตกต่างของความยากของข้อสอบมีค่า 82.6834 คิดเป็นร้อยละ 29 ของความแปรปรวนรวม ความแปรปรวนของผู้ตัดสิน (σ_r^2) เป็นการประมาณค่าความแปรปรวนของคะแนนเอกภพที่แสดงถึงความแตกต่างในการตัดสินคะแนนจุดตัดของผู้ตัดสิน มีค่า 92.7007 คิดเป็นร้อยละ 33 ของความแปรปรวนรวม และความแปรปรวนของส่วนที่เหลือ (σ_{ir}^2) เป็นความแปรปรวนของผลร่วมระหว่างข้อสอบกับผู้ตัดสิน แสดงให้เห็นถึงความไม่คงเส้นคงวาของผู้ตัดสินแต่ละคนในการตัดสินข้อสอบแต่ละข้อมีค่า 108.5336 คิดเป็นร้อยละ 38 ของความแปรปรวนรวม

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 33 การประมาณค่าความแปรปรวนในขั้นการสรุปอ้างอิง (G-study) ขั้นการตัดสินใจ (D-study) และสัมประสิทธิ์การอ้างอิงสรุปของการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุงที่คะแนนจุดตัดระดับดี (B) จำแนกตามจำนวนผู้ตัดสิน

Source of Variation		Estimated G Study Variance Components	Alternative Estimated D Studies Design Variance Component 8	12	14	16	18	20	22	24
Item (i)										
ครั้งที่ 1		32.9158	32.9158	32.9158	32.9158	32.9158	32.9158	32.9158	32.9158	32.9158
ครั้งที่ 2		83.2155	83.2155	83.2155	83.2155	83.2155	83.2155	83.2155	83.2155	83.2155
ครั้งที่ 3		82.6834	82.6834	82.6834	82.6834	82.6834	82.6834	82.6834	82.6834	82.6834
Rater (r)										
ครั้งที่ 1		85.4872	10.6859	7.1239	6.1062	5.3430	4.7493	4.2744	3.8858	3.5620
ครั้งที่ 2		118.6990	14.8374	9.8916	8.4785	7.4187	6.5944	5.9350	5.3954	4.9458
ครั้งที่ 3		92.7007	11.5876	7.7251	6.6215	5.7938	5.1500	4.6350	4.2137	3.8625
Residual (ir,e)										
ครั้งที่ 1		54.0383	6.7548	4.5032	3.8599	3.3774	3.0021	2.7019	2.4563	2.2516
ครั้งที่ 2		412.1508	51.5189	34.3459	29.4393	25.7594	22.8973	20.6075	18.7341	17.1730
ครั้งที่ 3		108.5336	13.5667	9.0445	7.7524	6.7836	6.0296	5.4267	4.9333	4.5222
σ^2_{Rel}	ครั้งที่ 1		6.7548	4.5032	3.8599	3.3774	3.0021	2.7019	2.4563	2.2516
	ครั้งที่ 2		51.5189	34.3459	29.4393	25.7594	22.8973	20.6075	18.7341	17.1730
	ครั้งที่ 3		13.5667	9.0445	7.7524	6.7836	6.0296	5.4267	4.9333	4.5222
σ^2_{Abs}	ครั้งที่ 1		17.4408	11.6271	9.9661	8.7204	7.7514	6.9763	6.3421	5.8136
	ครั้งที่ 2		66.3563	44.2375	37.9178	33.1781	29.4917	26.5425	24.1295	22.1188
	ครั้งที่ 3		25.1543	16.7696	14.3739	12.5774	11.1796	10.0617	9.1470	8.3847
$E\rho^2$	ครั้งที่ 1		0.8297	0.8797	0.8950	0.9069	0.9164	0.9241	0.9306	0.9360
	ครั้งที่ 2		0.6176	0.7079	0.7387	0.7636	0.7842	0.8015	0.8162	0.8289
	ครั้งที่ 3		0.8591	0.9014	0.9143	0.9242	0.9320	0.9384	0.9437	0.9481
ϕ	ครั้งที่ 1		0.6537	0.7390	0.7676	0.7906	0.8094	0.8251	0.8385	0.8499
	ครั้งที่ 2		0.5564	0.6529	0.6870	0.7150	0.7383	0.7582	0.7752	0.7900
	ครั้งที่ 3		0.7667	0.8314	0.8519	0.8680	0.8809	0.8915	0.9004	0.9079

จากตารางที่ 33 ผลการศึกษา D ประกอบด้วยการประมาณค่าความแปรปรวนของผู้ตัดสิน จำนวน 8, 12, 14, 16, 18, 20, 22, 24 คน จากการตัดสินคะแนนจุดตัดของผู้ตัดสินจำนวน 12 คน ด้วยวิธีแองกอฟที่ได้รับการปรับปรุง ที่ระดับดี (B) ครั้งที่ 1 ความแปรปรวนของคะแนนความคลาดเคลื่อนสัมพันธ์(σ^2_{Rel}) ซึ่งเป็นความคลาดเคลื่อนที่เกิดจากผลร่วมของข้อสอบ และผู้ตัดสิน มีค่าลดลง เมื่อจำนวนผู้ตัดสินเพิ่มขึ้น มีค่าดังนี้ 6.7548 4.5032 3.8599 3.3774 3.0021 2.7019 2.4563 2.2516 ครั้งที่ 2 51.5189 34.3459 29.4393 25.7594 22.8973 20.6075 18.7341 17.1730 ครั้งที่ 3 มีค่า ดังนี้ 13.5667 9.0445 7.7524 6.7836 6.0296 5.4267 4.9333 4.5222 เมื่อมีผู้ตัดสินจำนวน 8, 12, 14, 16, 18, 20, 22, 24 คน ตามลำดับ

ส่วนความแปรปรวนของคะแนนความคลาดเคลื่อนสัมบูรณ์(σ^2_{Abs})ซึ่งเป็นผลรวมของความคลาดเคลื่อนที่เกิดจากผลร่วมของข้อสอบและผู้ตัดสิน มีค่าลดลงเมื่อจำนวนผู้ตัดสินเพิ่มขึ้น ครั้งที่ 1 มีค่าดังนี้ 17.4408 11.6271 9.9661 8.7204 7.7514 6.9763 6.3421 5.8136 ครั้งที่ 2 มีค่าดังนี้ 66.3563 44.2375 37.9178 33.1781 29.4917 26.5425 24.1295 22.1188 ครั้งที่ 3 มีค่าดังนี้ 25.1543 16.7696 14.3739 12.5774 11.1796 10.0617 9.1470 8.3847 เมื่อมีผู้ตัดสินจำนวน 8, 12, 14, 16, 18, 20, 22, 24 คน ตามลำดับ

ค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินแบบสัมพัทธ์ ($E\rho^2$) มีค่าเพิ่มขึ้นเมื่อจำนวนผู้ตัดสินเพิ่มขึ้น ครั้งที่ 1 มีค่าดังนี้ 0.8297 0.8797 0.8950 0.9069 0.9164 0.9241 0.9306 0.9360 ครั้งที่ 2 มีค่าดังนี้ 0.6176 0.7079 0.7387 0.7636 0.7842 0.8015 0.8162 0.8289 ครั้งที่ 3 มีค่าดังนี้ 0.8591 0.9014 0.9143 0.9242 0.9320 0.9384 0.9437 0.9481 เมื่อมีผู้ตัดสินจำนวน 8, 12, 14, 16, 18, 20, 22, 24 คน ตามลำดับ

ส่วนค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินแบบสัมบูรณ์ (ϕ) มีค่าเพิ่มขึ้นเมื่อจำนวนผู้ตัดสินเพิ่มขึ้น ครั้งที่ 1 มีค่าดังนี้ 0.6537 0.7390 0.7676 0.7906 0.8094 0.8251 0.8385 0.8499 ครั้งที่ 2 มีค่าดังนี้ 0.5564 0.6529 0.6870 0.7150 0.7383 0.7582 0.7752 0.7900 ครั้งที่ 3 มีค่าดังนี้ 0.7667 0.8314 0.8519 0.8680 0.8809 0.8915 0.9004 0.9079 เมื่อมีผู้ตัดสินจำนวน 8, 12, 14, 16, 18, 20, 22, 24 คน ตามลำดับ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 34 การวิเคราะห์ความแปรปรวนและการประมาณค่าความแปรปรวนของการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุงที่คะแนนจุดตัดระดับดีพอใช้ (C+) ในขั้นสรุปอ้างอิง (G-study)

Source of Variation	df	Sum of Squares	Mean Squares	Estimated Variance Component	Percentage of Total Variance
ครั้งที่ 1					
Item (i)	99	100791.4082	1018.0950	76.9083	28.8360
Rater (r)	11	105112.6025	9555.6911	94.6050	35.4712
Residual (ir,e)	1089	103668.4808	95.1960	95.1960	35.6928
Total	1199	309572.4925		266.7093	100
ครั้งที่ 2					
Item (i)	99	125767.5133	1270.3789	95.7704	32.0185
Rater (r)	11	91757.3867	8341.5806	82.2045	27.4831
Residual (ir,e)	1089	131915.4467	121.1345	121.1345	40.4984
Total	1199	349440.3467		299.1094	100
ครั้งที่ 3					
Item (i)	99	120978.3867	1222.0039	91.9122	31.9764
Rater (r)	11	85423.8867	7765.8079	76.4675	26.6032
Residual (ir,e)	1089	129654.1133	119.0580	119.0580	41.4205
Total	1199	336056.3867		287.4377	100

จากตารางที่ 34 เมื่อผู้ตัดสินกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุงที่คะแนนจุดตัดระดับดีพอใช้ (C+) ครั้งที่ 1 โดยที่เอกภพของการสังเกตประกอบด้วย ข้อสอบ และผู้ตัดสิน ซึ่งมีขนาดไม่จำกัด โดยเป็นการประมาณค่าความแปรปรวนของแหล่งต่างๆ เมื่อเทียบกับความแปรปรวนรวม มีค่าดังนี้ ความแปรปรวนของข้อสอบ (σ_i^2) เป็นความแตกต่างของความยากของข้อสอบมีค่า 76.9083 คิดเป็นร้อยละ 29 ของความแปรปรวนรวม ความแปรปรวนของผู้ตัดสิน (σ_r^2) เป็นการประมาณค่าความแปรปรวนของคะแนนเอกภพที่แสดงถึงความแตกต่างในการตัดสินคะแนนจุดตัดของผู้ตัดสินมีค่า 94.6050 คิด

เป็นร้อยละ 35 ของความแปรปรวนรวม และความแปรปรวนของส่วนที่เหลือ (σ_{ir}^2) เป็นความแปรปรวนของผลร่วมระหว่างข้อสอบกับผู้ตัดสิน แสดงให้เห็นถึงความไม่คงเส้นคงวาของผู้ตัดสินแต่ละคนในการตัดสินข้อสอบแต่ละข้อ มีค่า 95.1960 คิดเป็นร้อยละ 36 ของความแปรปรวนรวม

เมื่อผู้ตัดสินกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุงที่คะแนนจุดตัดระดับดีพอใช้ (C+) ครั้งที่ 2 มีค่าดังนี้ ความแปรปรวนของข้อสอบ (σ_i^2) เป็นความแตกต่างของความยากของข้อสอบมีค่า 95.7704 คิดเป็นร้อยละ 32 ของความแปรปรวนรวม ความแปรปรวนของผู้ตัดสิน (σ_r^2) เป็นการประมาณค่าความแปรปรวนของคะแนนเอกภพที่แสดงถึงความแตกต่างในการตัดสินคะแนนจุดตัดของผู้ตัดสินมีค่า 82.2045 คิดเป็นร้อยละ 27 ของความแปรปรวนรวม และความแปรปรวนของส่วนที่เหลือ (σ_{ir}^2) เป็นความแปรปรวนของผลร่วมระหว่างข้อสอบกับผู้ตัดสิน แสดงให้เห็นถึงความไม่คงเส้นคงวาของผู้ตัดสินแต่ละคนในการตัดสินข้อสอบแต่ละข้อ มีค่า 121.1345 คิดเป็นร้อยละ 41 ของความแปรปรวนรวม

เมื่อผู้ตัดสินกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุงที่คะแนนจุดตัดระดับดีพอใช้ (C+) ครั้งที่ 3 มีค่าดังนี้ ความแปรปรวนของข้อสอบ (σ_i^2) เป็นความแตกต่างของความยากของข้อสอบมีค่า 91.9122 คิดเป็นร้อยละ 32 ของความแปรปรวนรวม ความแปรปรวนของผู้ตัดสิน (σ_r^2) เป็นการประมาณค่าความแปรปรวนของคะแนนเอกภพที่แสดงถึงความแตกต่างในการตัดสินคะแนนจุดตัดของผู้ตัดสิน มีค่า 76.4675 คิดเป็นร้อยละ 27 ของความแปรปรวนรวม และความแปรปรวนของส่วนที่เหลือ (σ_{ir}^2) เป็นความแปรปรวนของผลร่วมระหว่างข้อสอบกับผู้ตัดสิน แสดงให้เห็นถึงความไม่คงเส้นคงวาของผู้ตัดสินแต่ละคนในการตัดสินข้อสอบแต่ละข้อมีค่า 119.0580 คิดเป็นร้อยละ 41.4205 ของความแปรปรวนรวม

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 35 การประมาณค่าความแปรปรวนในขั้นการสรุปอ้างอิง (G-study) ขั้นการตัดสินใจ (D-study) และสัมประสิทธิ์การอ้างอิงสรุปของการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุงที่คะแนนจุดตัดระดับดีพอใช้ (C+) จำแนกตามจำนวนผู้ตัดสิน

Source of Variation		Estimated G Study Variance Components	Alternative D Studies Design Variance Component 8	Estimated 12	14	16	18	20	22	24
Item (i)										
ครั้งที่ 1		76.9083	76.9083	76.9083	76.9083	76.9083	76.9083	76.9083	76.9083	76.9083
ครั้งที่ 2		95.7704	95.7704	95.7704	95.7704	95.7704	95.7704	95.7704	95.7704	95.7704
ครั้งที่ 3		91.9122	91.9122	91.9122	91.9122	91.9122	91.9122	91.9122	91.9122	91.9122
Rater (r)										
ครั้งที่ 1		94.6050	11.8256	7.8838	6.7575	5.9128	5.2558	4.7303	4.3002	3.9419
ครั้งที่ 2		82.2045	10.2756	6.8504	5.8718	5.1378	4.5669	4.1102	3.7366	3.4252
ครั้งที่ 3		76.4675	9.5584	6.3723	5.4620	4.7792	4.2482	3.8234	3.4758	3.1862
Residual (ir,e)										
ครั้งที่ 1		95.1960	11.8995	7.9330	6.7997	5.9498	5.2887	4.7598	4.3271	3.9665
ครั้งที่ 2		121.1345	15.1418	10.0945	8.6525	7.5709	6.7297	6.0567	5.5061	5.0473
ครั้งที่ 3		119.0580	14.8822	9.9215	8.5041	7.4411	6.6143	5.9529	5.4117	4.9608
σ^2_{Rel}	ครั้งที่ 1		11.8995	7.9330	6.7997	5.9498	5.2887	4.7598	4.3271	3.9665
	ครั้งที่ 2		15.1418	10.0945	8.6525	7.5709	6.7297	6.0567	5.5061	5.0473
	ครั้งที่ 3		14.8822	9.9215	8.5041	7.4411	6.6143	5.9529	5.4117	4.9608
σ^2_{Abs}	ครั้งที่ 1		23.7251	15.9168	13.5572	11.8626	10.5445	9.4901	8.6273	7.9084
	ครั้งที่ 2		25.4174	16.9449	14.5243	12.7087	11.2966	10.1669	9.2427	8.4725
	ครั้งที่ 3		24.4406	16.2938	13.9661	12.2203	10.8625	9.7763	8.8875	8.1470
$E\rho^2$	ครั้งที่ 1		0.8660	0.9065	0.9188	0.9282	0.9357	0.9417	0.9467	0.9510
	ครั้งที่ 2		0.8635	0.9047	0.9171	0.9267	0.9343	0.9405	0.9456	0.9499
	ครั้งที่ 3		0.8607	0.9026	0.9153	0.9251	0.9329	0.9392	0.9444	0.9488
ϕ	ครั้งที่ 1		0.7642	0.8294	0.8501	0.8664	0.8794	0.8902	0.8991	0.9068
	ครั้งที่ 2		0.7903	0.8497	0.8683	0.8829	0.8945	0.9040	0.9120	0.9187
	ครั้งที่ 3		0.7899	0.8494	0.8681	0.8827	0.8943	0.9039	0.9118	0.9186

จากตารางที่ 35 ผลการศึกษา D ประกอบด้วยการประมาณค่าความแปรปรวนของผู้ตัดสิน จำนวน 8, 12, 14, 16, 18, 20, 22, 24 คน จากการตัดสินคะแนนจุดตัดของผู้ตัดสินจำนวน 12 คน ด้วยวิธีแองกอฟที่ได้รับการปรับปรุง ที่ระดับดีพอใช้ (C+) ครั้งที่ 1 ความแปรปรวนของคะแนนความคลาดเคลื่อนสัมพันธ์ ($\sigma^2_{\mathrm{Re}l}$) ซึ่งเป็นความคลาดเคลื่อนที่เกิดจากผลร่วมของข้อสอบและผู้ตัดสินมีค่าลดลง เมื่อจำนวนผู้ตัดสินเพิ่มขึ้น มีค่า ดังนี้ 11.8995 7.9330 6.7997 5.9498 5.2887 4.7598 4.3271 3.9665 ครั้งที่ 2 มีค่า ดังนี้ 15.1418 10.0945 8.6525 7.5709 6.7297 6.0567 5.5061 5.0473 ครั้งที่ 3 มีค่า ดังนี้ 14.8822 9.9215 8.5041 7.4411 6.6143 5.9529 5.4117 4.9608 เมื่อมีผู้ตัดสินจำนวน 8, 12, 14, 16, 18, 20, 22, 24 คน ตามลำดับ

ส่วนความแปรปรวนของคะแนนความคลาดเคลื่อนสัมบูรณ์(σ^2_{Abs})ซึ่งเป็นผลรวมของความคลาดเคลื่อนที่เกิดจากผลร่วมของข้อสอบและผู้ตัดสิน มีค่าลดลงเมื่อจำนวนผู้ตัดสินเพิ่มขึ้น ครั้งที่ 1 มีค่าดังนี้ 23.7251 15.9168 13.5572 11.8626 10.5445 9.4901 8.6273 ครั้งที่ 2 มีค่าดังนี้ 25.4174 16.9449 14.5243 12.7087 11.2966 10.1669 9.2427 8.4725 ครั้งที่ 3 มีค่าดังนี้ 24.4406 16.2938 13.9661 12.2203 10.8625 9.7763 8.8875 8.1470 เมื่อมีผู้ตัดสินจำนวน 8, 12, 14, 16, 18, 20, 22, 24 คน ตามลำดับ

ค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินแบบสัมพัทธ์ ($\mathrm{E}\rho^2$) มีค่าเพิ่มขึ้นเมื่อจำนวนผู้ตัดสินเพิ่มขึ้น ครั้งที่ 1 มีค่าดังนี้ 0.8660 0.9065 0.9188 0.9282 0.9357 0.9417 0.9467 0.9510 ครั้งที่ 2 มีค่าดังนี้ 0.8635 0.9047 0.9171 0.9267 0.9343 0.9405 0.9456 0.9499 ครั้งที่ 3 มีค่าดังนี้ 0.8607 0.9026 0.9153 0.9251 0.9329 0.9392 0.9444 0.9488 เมื่อมีผู้ตัดสินจำนวน 8, 12, 14, 16, 18, 20, 22, 24 คน ตามลำดับ

ส่วนค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินแบบสัมบูรณ์ (ϕ) มีค่าเพิ่มขึ้นเมื่อจำนวนผู้ตัดสินเพิ่มขึ้น ครั้งที่ 1 มีค่าดังนี้ 0.7642 0.8294 0.8501 0.8664 0.8794 0.8902 0.8991 0.9068 ครั้งที่ 2 มีค่าดังนี้ 0.7903 0.8497 0.8683 0.8829 0.8945 0.9040 0.9120 0.9187 ครั้งที่ 3 มีค่าดังนี้ 0.7899 0.8494 0.8681 0.8827 0.8943 0.9039 0.9118 0.9186 เมื่อมีผู้ตัดสินจำนวน 8, 12, 14, 16, 18, 20, 22, 24 คน ตามลำดับ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 36 การวิเคราะห์ความแปรปรวนและการประมาณค่าความแปรปรวนของการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุงที่ระดับพอใช้ (C) ในขั้นสรุปอ้างอิง

Source of Variation	df	Sum of Squares	Mean Squares	Estimated Variance Component	Percentage of Total Variance
ครั้งที่ 1					
Item (i)	99	97202.7425	981.8459	75.6039	29.9600
Rater (r)	11	113181.5292	10289.2299	102.1463	40.4781
Residual (ir,e)	1089	81238.8875	74.5995	74.5995	29.5620
Total	1199	291623.1592		252.3497	100
ครั้งที่ 2					
Item (i)	99	99443.8092	1004.4829	76.6632	30.4144
Rater (r)	11	100891.3225	9171.9384	90.8741	36.0523
Residual (ir,e)	1089	92047.2608	84.5246	84.5246	33.5333
Total	1199	292382.3925		252.0619	100

จากตารางที่ 36 เมื่อผู้ตัดสินกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุงที่คะแนนจุดตัดระดับพอใช้ (C) ครั้งที่ 1 โดยที่เอกภพของการสังเกตประกอบด้วย ข้อสอบและผู้ตัดสิน ซึ่งมีขนาดไม่จำกัด โดยเป็นการประมาณค่าความแปรปรวนของแหล่งต่างๆ เมื่อเทียบกับความแปรปรวนรวม มีค่าดังนี้ ความแปรปรวนของข้อสอบ (σ_i^2) เป็นความแตกต่างของความยากของข้อสอบมีค่า 75.6039 คิดเป็นร้อยละ 30 ของความแปรปรวนรวม ความแปรปรวนของผู้ตัดสิน (σ_r^2) เป็นการประมาณค่าความแปรปรวนของคะแนนเอกภพที่แสดงถึงความแตกต่างในการตัดสินคะแนนจุดตัดของผู้ตัดสินมีค่า 102.1463 คิดเป็นร้อยละ 40 ของความแปรปรวนรวม และความแปรปรวนของส่วนที่เหลือ (σ_{ir}^2) เป็นความแปรปรวนของผลร่วมระหว่างข้อสอบกับผู้ตัดสิน แสดงให้เห็นถึง ความไม่คงเส้นคงวาของผู้ตัดสินแต่ละคนในการตัดสินข้อสอบแต่ละข้อ มีค่า 74.5995 คิดเป็นร้อยละ 30 ของความแปรปรวนรวม

เมื่อผู้ตัดสินกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุงที่คะแนนจุดตัดระดับพอใช้ (C) ครั้งที่ 2 มีค่าดังนี้ ความแปรปรวนของข้อสอบ (σ_i^2) เป็นความแตกต่างของความยากของข้อสอบมีค่า 76.6632 คิดเป็นร้อยละ 30 ของความแปรปรวนรวม ความแปรปรวนของผู้ตัดสิน (σ_r^2) เป็นการประมาณค่าความแปรปรวนของคะแนนเอกภพที่แสดงถึงความแตกต่างในการตัดสินคะแนนจุดตัดของผู้ตัดสิน มีค่า

90.8741 คิดเป็นร้อยละ 36 ของความแปรปรวนรวม และความแปรปรวนของส่วนที่เหลือ (σ_{ir}^2) เป็นความแปรปรวนของผลร่วมระหว่างข้อสอบกับผู้ตัดสิน แสดงให้เห็นถึงความไม่คงเส้นคงวาของผู้ตัดสินแต่ละคนในการตัดสินข้อสอบแต่ละข้อมีค่า 84.5246 คิดเป็นร้อยละ 34 ของความแปรปรวนรวม

เมื่อผู้ตัดสินกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุงที่คะแนนจุดตัดระดับพอใช้ (C) ครั้งที่ 3 มีค่าดังนี้ ความแปรปรวนของข้อสอบ (σ_i^2) เป็นความแตกต่างของความยากของข้อสอบมีค่า 67.9771 คิดเป็นร้อยละ 29 ของความแปรปรวนรวม ความแปรปรวนของผู้ตัดสิน (σ_r^2) เป็นการประมาณค่าความแปรปรวนของคะแนนเอกภพที่แสดงถึงความแตกต่างในการตัดสินคะแนนจุดตัดของผู้ตัดสิน มีค่า 80.7514 คิดเป็นร้อยละ 34 ของความแปรปรวนรวม และความแปรปรวนของส่วนที่เหลือ (σ_{ir}^2) เป็นความแปรปรวนของผลร่วมระหว่างข้อสอบกับผู้ตัดสิน แสดงให้เห็นถึง ความไม่คงเส้นคงวาของผู้ตัดสินแต่ละคนในการตัดสินข้อสอบแต่ละข้อมีค่า 89.2517 คิดเป็นร้อยละ 38 ของความแปรปรวนรวม

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 37 การประมาณค่าความแปรปรวนในขั้นการสรุปอ้างอิง (G-study) ขั้นการตัดสินใจ (D-study) และสัมประสิทธิ์การอ้างอิงสรุปของการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุงที่คะแนนจุดตัดระดับพอใช้ (C) จำแนกตามจำนวนผู้ตัดสิน

Source of Variation		Estimated G Study Variance Components	Alternative D Studies Design Variance Component 8	Estimated 12	14	16	18	20	22	24
Item (i)										
ครั้งที่ 1		75.6039	75.6039	75.6039	75.6039	75.6039	75.6039	75.6039	75.6039	75.6039
ครั้งที่ 2		76.6632	76.6632	76.6632	76.6632	76.6632	76.6632	76.6632	76.6632	76.6632
ครั้งที่ 3		67.9771	67.9771	67.9771	67.9771	67.9771	67.9771	67.9771	67.9771	67.9771
Rater (r)										
ครั้งที่ 1		102.1463	12.7683	8.5122	7.2962	6.3841	5.6748	5.1073	4.6430	4.2561
ครั้งที่ 2		90.8741	11.3593	7.5728	6.4910	5.6796	5.0486	4.5437	4.1306	3.7864
ครั้งที่ 3		80.7514	10.0939	6.7293	5.7680	5.0470	4.4862	4.0376	3.6705	3.3646
Residual (ir,e)										
ครั้งที่ 1		74.5995	9.3249	6.2166	5.3285	4.6625	4.1444	3.7300	3.3909	3.1083
ครั้งที่ 2		84.5246	10.5656	7.0437	6.0375	5.2828	4.6958	4.2262	3.8420	3.5219
ครั้งที่ 3		89.2517	11.1565	7.4376	6.3751	5.5782	4.9584	4.4626	4.0569	3.7188
σ^2_{Rel}	ครั้งที่ 1		9.3249	6.2166	5.3285	4.6625	4.1444	3.7300	3.3909	3.1083
	ครั้งที่ 2		10.5656	7.0437	6.0375	5.2828	4.6958	4.2262	3.8420	3.5219
	ครั้งที่ 3		11.1565	7.4376	6.3751	5.5782	4.9584	4.4626	4.0569	3.7188
σ^2_{Abs}	ครั้งที่ 1		22.0932	14.7288	12.6247	11.04666	9.8192	8.8373	8.0339	7.3644
	ครั้งที่ 2		21.9249	14.6165	12.5285	10.9624	9.7444	8.7699	7.9726	7.3083
	ครั้งที่ 3		21.2504	14.1669	12.1431	10.6252	9.4446	8.5002	7.7274	7.0834
$E\rho^2$	ครั้งที่ 1		0.8902	0.9240	0.9342	0.9419	0.9480	0.9530	0.9571	0.9605
	ครั้งที่ 2		0.8789	0.9159	0.9270	0.9355	0.9423	0.9478	0.9523	0.9561
	ครั้งที่ 3		0.8590	0.9014	0.9143	0.9242	0.9320	0.9384	0.9437	0.9481
ϕ	ครั้งที่ 1		0.7739	0.8370	0.8569	0.8725	0.8851	0.8953	0.9039	0.9112
	ครั้งที่ 2		0.7776	0.8399	0.8595	0.8749	0.8872	0.8974	0.9058	0.9130
	ครั้งที่ 3		0.7618	0.8275	0.8484	0.8648	0.8780	0.8889	0.8979	0.9056

จากตารางที่ 37 ผลการศึกษา D ประกอบด้วยการประมาณค่าความแปรปรวนของผู้ตัดสิน จำนวน 8, 12, 14, 16, 18, 20, 22, 24 คน จากการตัดสินคะแนนจุดตัดของผู้ตัดสินจำนวน 12 คน ด้วยวิธีแองกอฟที่ได้รับการปรับปรุง ที่ระดับ พอใช้ (C) ครั้งที่ 1 ความแปรปรวนของคะแนนความคลาดเคลื่อนสัมพันธ์($\sigma^2_{\mathrm{Re}l}$) ซึ่งเป็นความคลาดเคลื่อนที่เกิดจากผลร่วมของข้อสอบ และผู้ตัดสินมีค่าลดลง เมื่อจำนวนผู้ตัดสินเพิ่มขึ้น มีค่าดังนี้ 9.3249 6.2166 5.3285 4.6625 4.1444 3.7300 3.3909 3.1083 ครั้งที่ 2 มีค่า ดังนี้ 10.5656 7.0437 6.0375 5.2828 4.6958 4.2262 3.8420 3.5219 ครั้งที่ 3 มีค่า ดังนี้ 11.1565 7.4376 6.3751 5.5782 4.9584 4.4626 4.0569 3.7188 เมื่อมีผู้ตัดสินจำนวน 8, 12, 14, 16, 18, 20, 22, 24 คน ตามลำดับ

ส่วนความแปรปรวนของคะแนนความคลาดเคลื่อนสัมบูรณ์(σ^2_{Abs})ซึ่งเป็นผลรวมของความคลาดเคลื่อนที่เกิดจากผลร่วมของข้อสอบและผู้ตัดสิน มีค่าลดลงเมื่อจำนวนผู้ตัดสินเพิ่มขึ้น ครั้งที่ 1 มีค่าดังนี้ 22.0932 14.7288 12.6247 11.04666 9.8192 8.8373 8.0339 7.3644 ครั้งที่ 2 มีค่าดังนี้ 21.9249 14.6165 12.5285 10.9624 9.7444 8.7699 7.9726 7.3083 ครั้งที่ 3 มีค่าดังนี้ 21.2504 14.1669 12.1431 10.6252 9.4446 8.5002 7.7274 7.0834 เมื่อมีผู้ตัดสินจำนวน 8, 12, 14, 16, 18, 20, 22, 24 คน ตามลำดับ

ค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินแบบสัมพัทธ์ ($\mathrm{E}\rho^2$) มีค่าเพิ่มขึ้นเมื่อจำนวนผู้ตัดสินเพิ่มขึ้น ครั้งที่ 1 มีค่าดังนี้ 0.8902 0.9240 0.9342 0.9419 0.9480 0.9530 0.9571 0.9605 ครั้งที่ 2 มีค่าดังนี้ 0.8789 0.9159 0.9270 0.9355 0.9423 0.9478 0.9523 0.9561 ครั้งที่ 3 มีค่าดังนี้ 0.8590 0.9014 0.9143 0.9242 0.9320 0.9384 0.9437 0.9481 เมื่อมีผู้ตัดสินจำนวน 8, 12, 14, 16, 18, 20, 22, 24 คน ตามลำดับ

ส่วนค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินแบบสัมบูรณ์ (ϕ) มีค่าเพิ่มขึ้นเมื่อจำนวนผู้ตัดสินเพิ่มขึ้น ครั้งที่ 1 มีค่าดังนี้ 0.7739 0.8370 0.8569 0.8725 0.8851 0.8953 0.9039 0.9112 ครั้งที่ 2 มีค่าดังนี้ 0.7776 0.8399 0.8595 0.8749 0.8872 0.8974 0.9058 0.9130 ครั้งที่ 3 มีค่าดังนี้ 0.7618 0.8275 0.8484 0.8648 0.8780 0.8889 0.8979 0.9056 เมื่อมีผู้ตัดสินจำนวน 8, 12, 14, 16, 18, 20, 22, 24 คน ตามลำดับ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 38 การวิเคราะห์ความแปรปรวนและการประมาณค่าความแปรปรวนการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุงที่คะแนนจุดตัดระดับอ่อน (D+) ในขั้นสรุปอ้างอิง (G-study)

Source of Variation	df	Sum of Squares	Mean Squares	Estimated Variance Component	Percentage of Total Variance
ครั้งที่ 1					
Item (i)	99	99029.7758	1000.30077	77.8421	36.4651
Rater (r)	11	77104.2492	7009.4772	69.4328	32.5257
Residual (ir,e)	1089	72086.8342	66.1954	66.1954	31.0092
Total	1199	248220.8592		213.4703	100
ครั้งที่ 2					
Item (i)	99	108382.2292	1094.7700	85.0890	50.5040
Rater (r)	11	11467.7492	1042.5227	9.6882	5.7504
Residual (ir,e)	1089	80282.0008	73.7025	73.7025	43.7456
Total	1199	200111.9792		168.4797	100
ครั้งที่ 3					
Item (i)	99	91883.8425	928.1196	71.5609	46.4653
Rater (r)	11	15128.3692	1375.3063	13.0592	8.4795
Residual (ir,e)	1089	75565.0475	69.3894	69.3894	45.0553
Total	1199	182577.2592		154.0095	100

จากตารางที่ 38 เมื่อผู้ตัดสินกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุงที่คะแนนจุดตัดระดับอ่อน (D+) ครั้งที่ 1 โดยที่เอกภพของการสังเกตประกอบด้วย ข้อสอบและผู้ตัดสินซึ่งมีขนาดไม่จำกัด โดยเป็นการประมาณค่าความแปรปรวนของแหล่งต่างๆ เมื่อเทียบกับความแปรปรวนรวมมีค่าดังนี้ ความแปรปรวนของข้อสอบ (σ_i^2) เป็นความแตกต่างของความยากของข้อสอบมีค่า 77.8421 คิดเป็นร้อยละ 36 ของความแปรปรวนรวม ความแปรปรวนของผู้ตัดสิน (σ_r^2) เป็นการประมาณค่าความแปรปรวนของคะแนนเอกภพที่แสดงถึงความแตกต่างในการตัดสินคะแนนจุดตัดของผู้ตัดสิน มีค่า 69.4328 คิดเป็นร้อยละ 33 ของความแปรปรวนรวม และความแปรปรวนของส่วนที่เหลือ (σ_{ir}^2) เป็นความแปรปรวนของ

ผลร่วมระหว่างข้อสอบกับผู้ตัดสิน แสดงให้เห็นถึง ความไม่คงเส้นคงวาของผู้ตัดสินแต่ละคนในการตัดสินข้อสอบแต่ละข้อ มีค่า 66.1954 คิดเป็นร้อยละ 31 ของความแปรปรวนรวม

เมื่อผู้ตัดสินกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุงที่คะแนนจุดตัดระดับอ่อน (D+) ครั้งที่ 2 มีค่าดังนี้ ความแปรปรวนของข้อสอบ (σ_i^2) เป็นความแตกต่างของความยากของข้อสอบมีค่า 85.0890 คิดเป็นร้อยละ 51 ของความแปรปรวนรวม ความแปรปรวนของผู้ตัดสิน (σ_r^2) เป็นการประมาณค่าความแปรปรวนของคะแนนเอกภพที่แสดงถึงความแตกต่างในการตัดสินคะแนนจุดตัดของผู้ตัดสิน มีค่า 9.6882 คิดเป็นร้อยละ 6 ของความแปรปรวนรวม และความแปรปรวนของส่วนที่เหลือ (σ_{ir}^2) เป็นความแปรปรวนของผลร่วมระหว่างข้อสอบกับผู้ตัดสิน แสดงให้เห็นถึงความไม่คงเส้นคงวาของผู้ตัดสินแต่ละคนในการตัดสินข้อสอบแต่ละข้อ มีค่า 73.7025 คิดเป็นร้อยละ 44 ของความแปรปรวนรวม

เมื่อผู้ตัดสินกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุงที่คะแนนจุดตัดระดับอ่อน (D+) ครั้งที่ 3 มีค่าดังนี้ ความแปรปรวนของข้อสอบ (σ_i^2) เป็นความแตกต่างของความยากของข้อสอบมีค่า 71.5609 คิดเป็นร้อยละ 46 ของความแปรปรวนรวม ความแปรปรวนของผู้ตัดสิน (σ_r^2) เป็นการประมาณค่าความแปรปรวนของคะแนนเอกภพที่แสดงถึงความแตกต่างในการตัดสินคะแนนจุดตัดของผู้ตัดสิน มีค่า 13.0592 คิดเป็นร้อยละ 8 ของความแปรปรวนรวม และความแปรปรวนของส่วนที่เหลือ (σ_{ir}^2) เป็นความแปรปรวนของผลร่วมระหว่างข้อสอบกับผู้ตัดสิน แสดงให้เห็นถึงความไม่คงเส้นคงวาของผู้ตัดสินแต่ละคนในการตัดสินข้อสอบแต่ละข้อมีค่า 69.3894 คิดเป็นร้อยละ 45 ของความแปรปรวนรวม

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 39 การประมาณค่าความแปรปรวนในขั้นการสรุปอ้างอิง (G-study) ขั้นการตัดสินใจ (D-study) และสัมประสิทธิ์การอ้างอิงสรุปของการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุงที่คะแนนจุดตัดระดับอ่อน (D+) จำแนกตามจำนวนผู้ตัดสิน

Source of Variation		Estimated G Study Variance Components	Alternative Estimated D Studies Design Variance Component 8	12	14	16	18	20	22	24
Item (i)										
ครั้งที่ 1		77.8421	77.8421	77.8421	77.8421	77.8421	77.8421	77.8421	77.8421	77.8421
ครั้งที่ 2		85.0890	85.0890	85.0890	85.0890	85.0890	85.0890	85.0890	85.0890	85.0890
ครั้งที่ 3		71.5609	71.5609	71.5609	71.5609	71.5609	71.5609	71.5609	71.5609	71.5609
Rater (r)										
ครั้งที่ 1		69.4328	8.6791	5.7861	4.9595	4.3396	3.8574	3.4716	3.1560	2.8930
ครั้งที่ 2		9.6882	1.2110	0.8074	0.6920	0.6055	0.5382	0.4844	0.4404	0.4037
ครั้งที่ 3		13.0592	1.6324	1.0883	0.9328	0.8162	0.7255	0.6530	0.5936	0.5441
Residual (ir,e)										
ครั้งที่ 1		66.1954	8.2744	5.5163	4.7283	4.1372	3.6775	3.3098	3.0089	2.7581
ครั้งที่ 2		73.7025	9.2128	6.1419	5.2645	4.6064	4.0946	3.6851	3.3501	3.0709
ครั้งที่ 3		69.3894	8.6737	5.7825	4.9564	4.3368	3.8550	3.4695	3.1541	2.8912
σ^2_{Rel}	ครั้งที่ 1		8.2744	5.5163	4.7283	4.1372	3.6775	3.3098	3.0089	2.7581
	ครั้งที่ 2		9.2128	6.1419	5.2645	4.6064	4.0946	3.6851	3.3501	3.0709
	ครั้งที่ 3		8.6737	5.7825	4.9564	4.3368	3.8550	3.4695	3.1541	2.8912
σ^2_{Abs}	ครั้งที่ 1		16.9535	11.3024	9.6878	8.4768	7.5349	6.7814	6.1649	5.6511
	ครั้งที่ 2		10.4238	6.9493	5.9565	5.2119	4.6328	4.1695	3.7905	3.4746
	ครั้งที่ 3		10.3061	6.8708	5.8892	5.1530	4.5805	4.1225	3.7504	3.4353
$E\rho^2$	ครั้งที่ 1		0.9039	0.9338	0.9427	0.9495	0.9549	0.9592	0.9628	0.9658
	ครั้งที่ 2		0.9023	0.9327	0.9417	0.9486	0.9541	0.9585	0.9621	0.9652
	ครั้งที่ 3		0.8919	0.9252	0.9352	0.9429	0.9489	0.9538	0.9578	0.9612
ϕ	ครั้งที่ 1		0.8212	0.8732	0.8893	0.9018	0.9118	0.9199	0.9266	0.9323
	ครั้งที่ 2		0.8909	0.9245	0.9346	0.9423	0.9484	0.9533	0.9574	0.9608
	ครั้งที่ 3		0.8741	0.9124	0.9240	0.9328	0.9398	0.9455	0.9502	0.9542

จากตารางที่ 39 ผลการศึกษา D ประกอบด้วยการประมาณค่าความแปรปรวนของผู้ตัดสิน จำนวน 8, 12, 14, 16, 18, 20, 22, 24 คน จากการตัดสินคะแนนจุดตัดของผู้ตัดสินจำนวน 12 คน ด้วยวิธีแองกอฟที่ได้รับการปรับปรุง ที่ระดับอ่อน (D+) ครั้งที่ 1 ความแปรปรวนของคะแนนความคลาดเคลื่อนสัมพันธ์(σ^2_{Rel}) ซึ่งเป็นความคลาดเคลื่อนที่เกิดจากผลร่วมของข้อสอบ และผู้ตัดสิน มีค่าลดลง เมื่อจำนวนผู้ตัดสินเพิ่มขึ้น มีค่า ดังนี้ 8.2744 5.5163 4.7283 4.1372 3.6775 3.3098 3.0089 2.7581 ครั้งที่ 2 มีค่า ดังนี้ 9.2128 6.1419 5.2645 4.6064 4.0946 3.6851 3.3501 3.0709 ครั้งที่ 3 มีค่า ดังนี้ 8.6737 5.7825 4.9564 4.3368 3.8550 3.4695 3.1541 2.8912 เมื่อมีผู้ตัดสินจำนวน 8, 12, 14, 16, 18, 20, 22, 24 คน ตามลำดับ

ส่วนความแปรปรวนของคะแนนความคลาดเคลื่อนสัมบูรณ์(σ^2_{Abs})ซึ่งเป็นผลรวมของความคลาดเคลื่อนที่เกิดจากผลร่วมของข้อสอบและผู้ตัดสิน มีค่าลดลงเมื่อจำนวนผู้ตัดสินเพิ่มขึ้น ครั้งที่ 1 มีค่าดังนี้ 16.9535 11.3024 9.6878 8.4768 7.5349 6.7814 6.1649 5.6511 ครั้งที่ 2 มีค่าดังนี้ 10.4238 6.9493 5.9565 5.2119 4.6328 4.1695 3.7905 3.4746 ครั้งที่ 3 มีค่าดังนี้ 10.3061 6.8708 5.8892 5.1530 4.5805 4.1225 3.7504 3.4353 เมื่อมีผู้ตัดสินจำนวน 8, 12, 14, 16, 18, 20, 22, 24 คน ตามลำดับ

ค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินแบบสัมพัทธ์ ($E\rho^2$) มีค่าเพิ่มขึ้นเมื่อจำนวนผู้ตัดสินเพิ่มขึ้น ครั้งที่ 1 มีค่าดังนี้ 0.9039 0.9338 0.9427 0.9495 0.9549 0.9592 0.9628 0.9658 ครั้งที่ 2 มีค่าดังนี้ 0.9023 0.9327 0.9417 0.9486 0.9541 0.9585 0.9621 0.9652 ครั้งที่ 3 มีค่าดังนี้ 0.8919 0.9252 0.9352 0.9429 0.9489 0.9538 0.9578 0.9612 เมื่อมีผู้ตัดสินจำนวน 8, 12, 14, 16, 18, 20, 22, 24 คน ตามลำดับ

ส่วนค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินแบบสัมบูรณ์ (ϕ) มีค่าเพิ่มขึ้นเมื่อจำนวนผู้ตัดสินเพิ่มขึ้น ครั้งที่ 1 มีค่าดังนี้ 0.8212 0.8732 0.8893 0.9018 0.9118 0.9199 0.9266 0.9323 ครั้งที่ 2 มีค่าดังนี้ 0.8909 0.9245 0.9346 0.9423 0.9484 0.9533 0.9574 0.9608 ครั้งที่ 3 มีค่าดังนี้ 0.8741 0.9124 0.9240 0.9328 0.9398 0.9455 0.9502 0.9542 เมื่อมีผู้ตัดสินจำนวน 8, 12, 14, 16, 18, 20, 22, 24 คน ตามลำดับ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 40 การวิเคราะห์ความแปรปรวนและการประมาณค่าความแปรปรวนของการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุงที่คะแนนจุดตัดระดับอ่อนมาก (D) ในขั้นสรุปอ้างอิง (G-study)

Source of Variation	df	Sum of Squares	Mean Squares	Estimated Variance Component	Percentage of Total Variance
ครั้งที่ 1					
Item (i)	99	112197.3092	1133.3062	90.4685	58.9000
Rater (r)	11	17512.6225	1592.0566	15.4437	10.0547
Residual (ir,e)	1089	51928.4608	47.6845	47.6845	31.0453
Total	1199	181638.3925		153.5967	100
ครั้งที่ 2					
Item (i)	99	74122.1200	748.7083	57.5302	40.9307
Rater (r)	11	27789.1867	2526.2897	24.6794	17.5585
Residual (ir,e)	1089	63538.4800	58.3457	58.3457	41.5109
Total	1199	165449.7867		140.5553	100
ครั้งที่ 3					
Item (i)	99	67241.8158	679.2103	51.9154	40.5047
Rater (r)	11	22651.9825	2059.2711	20.0305	15.6279
Residual (ir,e)	1089	61229.4342	56.2254	56.2254	43.8674
Total	1199	151123.2325		128.1713	100

จากตารางที่ 40 เมื่อผู้ตัดสินกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุงที่คะแนนจุดตัดระดับอ่อนมาก (D) ครั้งที่ 1 โดยที่เอกภพของการสังเกตประกอบด้วย ข้อสอบและผู้ตัดสิน ซึ่งมีขนาดไม่จำกัด โดยเป็นการประมาณค่าความแปรปรวนของแหล่งต่างๆ เมื่อเทียบกับความแปรปรวนรวมมีค่าดังนี้ ความแปรปรวนของข้อสอบ (σ_i^2) เป็นความแตกต่างของความยากของข้อสอบมีค่า 90.4685 คิดเป็นร้อยละ 59 ของความแปรปรวนรวม ความแปรปรวนของผู้ตัดสิน (σ_r^2) เป็นการประมาณค่าความแปรปรวนของคะแนนเอกภพที่แสดงถึงความแตกต่างในการตัดสินคะแนนจุดตัดของผู้ตัดสิน มีค่า 15.4437 คิดเป็นร้อยละ 10 ของความแปรปรวนรวม และความแปรปรวนของส่วนที่เหลือ (σ_{ir}^2) เป็นความแปรปรวนของผล

ร่วมระหว่างข้อสอบกับผู้ตัดสิน แสดงให้เห็นถึงความไม่คงเส้นคงวาของผู้ตัดสินแต่ละคนในการตัดสินข้อสอบแต่ละข้อ มีค่า 47.6845 คิดเป็นร้อยละ 31 ของความแปรปรวนรวม

เมื่อผู้ตัดสินกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุงที่คะแนนจุดตัดระดับอ่อนมาก (D) ครั้งที่ 2 มีค่าดังนี้ ความแปรปรวนของข้อสอบ (σ_i^2) เป็นความแตกต่างของความยากของข้อสอบ มีค่า 57.5302 คิดเป็นร้อยละ 41 ของความแปรปรวนรวม ความแปรปรวนของผู้ตัดสิน (σ_r^2) เป็นการประมาณค่าความแปรปรวนของคะแนนเอกภพที่แสดงถึงความแตกต่างในการตัดสินคะแนนจุดตัดของผู้ตัดสิน มีค่า 17.5585 คิดเป็นร้อยละ 25 ของความแปรปรวนรวม และความแปรปรวนของส่วนที่เหลือ (σ_{ir}^2) เป็นความแปรปรวนของผลร่วมระหว่างข้อสอบกับผู้ตัดสินแสดงให้เห็นถึง ความไม่คงเส้นคงวาของผู้ตัดสินแต่ละคนในการตัดสินข้อสอบแต่ละข้อมีค่า 58.3457 คิดเป็นร้อยละ 42 ของความแปรปรวนรวม

เมื่อผู้ตัดสินกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุงที่คะแนนจุดตัดระดับอ่อนมาก (D) ครั้งที่ 3 มีค่าดังนี้ ความแปรปรวนของข้อสอบ (σ_i^2) เป็นความแตกต่างของความยากของข้อสอบ มีค่า 51.9154 คิดเป็นร้อยละ 41 ของความแปรปรวนรวม ความแปรปรวนของผู้ตัดสิน (σ_r^2) เป็นการประมาณค่าความแปรปรวนของคะแนนเอกภพที่แสดงถึงความแตกต่างในการตัดสินคะแนนจุดตัดของผู้ตัดสิน มีค่า 20.0305 คิดเป็นร้อยละ 16 ของความแปรปรวนรวม และความแปรปรวนของส่วนที่เหลือ (σ_{ir}^2) เป็นความแปรปรวนของผลร่วมระหว่างข้อสอบกับผู้ตัดสิน แสดงให้เห็นถึง ความไม่คงเส้นคงวาของผู้ตัดสินแต่ละคนในการตัดสินข้อสอบแต่ละข้อมีค่า 56.2254 คิดเป็นร้อยละ 44 ของความแปรปรวนรวม

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 41 การประมาณค่าความแปรปรวนในขั้นการสรุปอ้างอิง (G-study) ขั้นการตัดสินใจ (D-study)และสัมประสิทธิ์การอ้างอิงสรุปของการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุงที่คะแนนจุดตัดระดับอ่อนมาก (D) จำแนกตามจำนวนผู้ตัดสิน

Source of Variation		Estimated G Study Variance Components	Alternative D Studies Design Estimated Variance Component							
			8	12	14	16	18	20	22	24
Item (i)										
ครั้งที่ 1		90.4685	90.4685	90.4685	90.4685	90.4685	90.4685	90.4685	90.4685	90.4685
ครั้งที่ 2		57.5302	57.5302	57.5302	57.5302	57.5302	57.5302	57.5302	57.5302	57.5302
ครั้งที่ 3		51.9154	51.9154	51.9154	51.9154	51.9154	51.9154	51.9154	51.9154	51.9154
Rater (r)										
ครั้งที่ 1		15.4437	1.9305	1.2870	1.1031	0.9652	0.8580	0.7722	0.7020	0.6435
ครั้งที่ 2		24.6794	3.0849	2.0566	1.7628	1.5425	1.3711	1.2340	1.1218	1.0283
ครั้งที่ 3		20.0305	2.5038	1.6692	1.4308	1.2519	1.1128	1.0015	0.9105	0.8346
Residual (ir,e)										
ครั้งที่ 1		47.6845	5.9606	3.9737	3.4060	2.9803	2.6491	2.3842	2.1675	1.9869
ครั้งที่ 2		58.3457	7.2932	4.8621	4.1676	3.6466	3.2414	2.9173	2.6521	2.4311
ครั้งที่ 3		56.2254	7.0282	4.6855	4.0161	3.5141	3.1236	2.8113	2.5557	2.3430
σ^2_{Rel}	ครั้งที่ 1		5.9606	3.9737	3.4060	2.9803	2.6491	2.3842	2.1675	1.9869
	ครั้งที่ 2		7.2932	4.8621	4.1676	3.6466	3.2414	2.9173	2.6521	2.4311
	ครั้งที่ 3		7.0282	4.6855	4.0161	3.5141	3.1236	2.8113	2.5557	2.3430
σ^2_{Abs}	ครั้งที่ 1		7.8911	5.2607	4.5091	3.9455	3.5071	3.1564	2.8695	2.6304
	ครั้งที่ 2		10.3781	6.9187	5.9304	5.1891	4.6125	4.1513	3.7739	3.4594
	ครั้งที่ 3		9.5320	6.3547	5.4469	4.7660	4.2364	3.8128	3.4662	3.1776
$E\rho^2$	ครั้งที่ 1		0.9382	0.9579	0.9637	0.9681	0.9716	0.9743	0.9766	0.9785
	ครั้งที่ 2		0.8875	0.9221	0.9325	0.9404	0.9467	0.9517	0.9559	0.9595
	ครั้งที่ 3		0.8808	0.9172	0.9282	0.9366	0.9433	0.9486	0.9531	0.9568
ϕ	ครั้งที่ 1		0.9198	0.9451	0.9525	0.9582	0.9627	0.9663	0.9693	0.9718
	ครั้งที่ 2		0.8472	0.8927	0.9066	0.9173	0.9258	0.9327	0.9385	0.9433
	ครั้งที่ 3		0.8449	0.8909	0.9050	0.9159	0.9246	0.9316	0.9374	0.9423

จากตารางที่ 41 ผลการศึกษา D ประกอบด้วยการประมาณค่าความแปรปรวนของผู้ตัดสิน จำนวน 8, 12, 14, 16, 18, 20, 22, 24 คน จากการตัดสินคะแนนจุดตัดของผู้ตัดสินจำนวน 12 คน ด้วยวิธีแองกอฟที่ได้รับการปรับปรุง ที่ระดับอ่อนมาก (D) ครั้งที่ 1 ความแปรปรวนของคะแนนความคลาดเคลื่อนสัมพันธ์

($\sigma^2_{\mathrm{Re}l}$) ซึ่งเป็นความคลาดเคลื่อนที่เกิดจากผลร่วมของข้อสอบ และผู้ตัดสิน มีค่าลดลง เมื่อจำนวนผู้ตัดสินเพิ่มขึ้น มีค่า ดังนี้ 5.9606 3.9737 3.4060 2.9803 2.6491 2.3842 2.1675 1.9869 ครั้งที่ 2 มีค่า ดังนี้ 7.2932 4.8621 4.1676 3.6466 3.2414 2.9173 2.6521 2.4311 ครั้งที่ 3 มีค่า ดังนี้ 7.0282 4.6855 4.0161 3.5141 3.1236 2.8113 2.5557 2.3430 เมื่อมีผู้ตัดสินจำนวน 8, 12, 14, 16, 18, 20, 22, 24 คน ตามลำดับ

ส่วนความแปรปรวนของคะแนนความคลาดเคลื่อนสัมบูรณ์(σ^2_{Abs})ซึ่งเป็นผลรวมของความคลาดเคลื่อนที่เกิดจากผลร่วมของข้อสอบและผู้ตัดสิน มีค่าลดลงเมื่อจำนวนผู้ตัดสินเพิ่มขึ้น ครั้งที่ 1 มีค่าดังนี้ 7.8911 5.2607 4.5091 3.9455 3.5071 3.1564 2.8695 2.6304 ครั้งที่ 2 มีค่าดังนี้ 10.3781 6.9187 5.9304 5.1891 4.6125 4.1513 3.7739 3.4594 ครั้งที่ 3 มีค่าดังนี้ 9.5320 6.3547 5.4469 4.7660 4.2364 3.8128 3.4662 3.1776 เมื่อมีผู้ตัดสินจำนวน 8, 12, 14, 16, 18, 20, 22, 24 คน ตามลำดับ

ค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินแบบสัมพัทธ์ ($\mathrm{E}\rho^2$) มีค่าเพิ่มขึ้นเมื่อจำนวนผู้ตัดสินเพิ่มขึ้น ครั้งที่ 1 มีค่าดังนี้ 0.9382 0.9579 0.9637 0.9681 0.9716 0.9743 0.9766 0.9785 ครั้งที่ 2 มีค่าดังนี้ 0.8875 0.9221 0.9325 0.9404 0.9467 0.9517 0.9559 0.9595 ครั้งที่ 3 มีค่าดังนี้ 0.8808 0.9172 0.9282 0.9366 0.9433 0.9486 0.9531 0.9568 เมื่อมีผู้ตัดสินจำนวน 8, 12, 14, 16, 18, 20, 22, 24 คน ตามลำดับ

ส่วนค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินแบบสัมบูรณ์ (ϕ) มีค่าเพิ่มขึ้นเมื่อจำนวนผู้ตัดสินเพิ่มขึ้น ครั้งที่ 1 มีค่าดังนี้ 0.9198 0.9451 0.9525 0.9582 0.9627 0.9663 0.9693 0.9718 ครั้งที่ 2 มีค่าดังนี้ 0.8472 0.8927 0.9066 0.9173 0.9258 0.9327 0.9385 0.9433 ครั้งที่ 3 มีค่าดังนี้ 0.8449 0.8909 0.9050 0.9159 0.9246 0.9316 0.9374 0.9423 เมื่อมีผู้ตัดสินจำนวน 8, 12, 14, 16, 18, 20, 22, 24 คน ตามลำดับ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

2. การประมาณค่าความเที่ยงของการกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์ค

ในการประมาณค่าความเที่ยงของการกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์ค ทั้ง 3 ครั้ง เป็นการคำนวณหาค่าสัมประสิทธิ์การสรุปอ้างอิง โดยใช้ทฤษฎีการสรุปอ้างอิง (Generalizability Theory) ลักษณะ cross design ที่มีรูปแบบ I x R design นั้นคือ ผู้ตัดสินทั้งหมด 12 คน ตัดสินข้อสอบทั้งหมด 100 ข้อ รายละเอียดผลการวิเคราะห์ดังต่อไปนี้

ตารางที่ 42 การวิเคราะห์ความแปรปรวนและการประมาณค่าความแปรปรวนของการกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์คที่คะแนนจุดตัดระดับดีเยี่ยม (A) ในขั้นสรุปอ้างอิง (G-study)

Source of Variation	df	Sum of Squares	Mean Squares	Estimated Variance Component	Percentage of Total Variance
ครั้งที่ 1					
Item (i)	99	182.7292	1.8458	0.1533	96.4151
Rater (r)	11	0.1292	0.0117	0.0001	0.0629
Residual (ir,e)	1089	6.1208	0.0056	0.0056	3.5220
Total	1199	188.9792		0.1590	100
ครั้งที่ 2					
Item (i)	99	182.8700	1.8472	0.1535	96.8454
Rater (r)	11	0.0900	0.0082	0.0000	0.0000
Residual (ir,e)	1089	5.4100	0.0050	0.0050	3.1546
Total	1199	188.3700		0.1585	100
ครั้งที่ 3					
Item (i)	99	179.7158	1.8153	0.1509	96.7928
Rater (r)	11	0.1092	0.0099	0.0000	0.0000
Residual (ir,e)	1089	5.4742	0.0050	0.0050	3.2072
Total	1199	185.2992		0.1559	100

จากตารางที่ 42 เมื่อผู้ตัดสินกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์คที่คะแนนจุดตัดระดับดีเยี่ยม (A) ครั้งที่ 1 โดยที่เอกภพของการสังเกตประกอบด้วย ข้อสอบและผู้ตัดสินซึ่งมีขนาดไม่จำกัด โดยเป็นการประมาณค่าความแปรปรวนของแหล่งต่างๆ เมื่อเทียบกับความแปรปรวนรวมมีค่าดังนี้ ความ

แปรปรวนของข้อสอบ (σ_i^2) เป็นความแตกต่างของความยากของข้อสอบมีค่า 0.1533 คิดเป็นร้อยละ 96 ของความแปรปรวนรวม ความแปรปรวนของผู้ตัดสิน (σ_r^2) เป็นการประมาณค่าความแปรปรวนของคะแนนเอกภพที่แสดงถึงความแตกต่างในการตัดสินคะแนนจุดตัดของผู้ตัดสินมีค่า 0.0001 คิดเป็นร้อยละ 0 ของความแปรปรวนรวม และความแปรปรวนของส่วนที่เหลือ (σ_{ir}^2) เป็นความแปรปรวนของผลร่วมระหว่างข้อสอบกับผู้ตัดสิน แสดงให้เห็นถึง ความไม่คงเส้นคงวาของผู้ตัดสินแต่ละคนในการตัดสินข้อสอบแต่ละข้อมีค่า 0.0056 คิดเป็นร้อยละ 4 ของความแปรปรวนรวม

เมื่อผู้ตัดสินกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์คที่คะแนนจุดตัดระดับดีเยี่ยม (A) ครั้งที่ 2 มีค่าดังนี้ ความแปรปรวนของข้อสอบ (σ_i^2) เป็นความแตกต่างของความยากของข้อสอบมีค่า 0.1535 คิดเป็นร้อยละ 97 ของความแปรปรวนรวม ความแปรปรวนของผู้ตัดสิน (σ_r^2) เป็นการประมาณค่าความแปรปรวนของคะแนนเอกภพที่แสดงถึงความแตกต่างในการตัดสินคะแนนจุดตัดของผู้ตัดสิน มีค่า 0.0000 คิดเป็นร้อยละ 0 ของความแปรปรวนรวม และความแปรปรวนของส่วนที่เหลือ (σ_{ir}^2) เป็นความแปรปรวนของผลร่วมระหว่างข้อสอบกับผู้ตัดสิน แสดงให้เห็นถึงความไม่คงเส้นคงวาของผู้ตัดสินแต่ละคนในการตัดสินข้อสอบแต่ละข้อมีค่า 0.0050 คิดเป็นร้อยละ 3 ของความแปรปรวนรวม

เมื่อผู้ตัดสินกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์คที่คะแนนจุดตัดระดับดีเยี่ยม (A) ครั้งที่ 3 มีค่าดังนี้ ความแปรปรวนของข้อสอบ (σ_i^2) เป็นความแตกต่างของความยากของข้อสอบมีค่า 0.1509 คิดเป็นร้อยละ 97 ของความแปรปรวนรวม ความแปรปรวนของผู้ตัดสิน (σ_r^2) เป็นการประมาณค่าความแปรปรวนของคะแนนเอกภพที่แสดงถึงความแตกต่างในการตัดสินคะแนนจุดตัดของผู้ตัดสิน มีค่า 0.0000 คิดเป็นร้อยละ 0 ของความแปรปรวนรวม และความแปรปรวนของส่วนที่เหลือ (σ_{ir}^2) เป็นความแปรปรวนของผลร่วมระหว่างข้อสอบกับผู้ตัดสิน แสดงให้เห็นถึงความไม่คงเส้นคงวาของผู้ตัดสินแต่ละคนในการตัดสินข้อสอบแต่ละข้อ มีค่า 0.0050 คิดเป็นร้อยละ 3 ของความแปรปรวนรวม

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 43 การประมาณค่าความแปรปรวนในขั้นการสรุปอ้างอิง (G-study) ขั้นการตัดสินใจ (D-study)และสัมประสิทธิ์การอ้างอิงสรุปของการกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์คที่คะแนนจุดตัดระดับดีเยี่ยม (A) จำแนกตามจำนวนผู้ตัดสิน

Source of Variation		Estimated G Study Variance Components	Alternative D Studies Design Estimated Variance Component 8	12	14	16	18	20	22	24
Item (i)										
ครั้งที่ 1		0.1533	0.1533	0.1533	0.1533	0.1533	0.1533	0.1533	0.1533	0.1533
ครั้งที่ 2		0.1535	0.1535	0.1535	0.1535	0.1535	0.1535	0.1535	0.1535	0.1535
ครั้งที่ 3		0.1509	0.1509	0.1509	0.1509	0.1509	0.1509	0.1509	0.1509	0.1509
Rater (r)										
ครั้งที่ 1		0.0001	0.0001	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
ครั้งที่ 2		0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
ครั้งที่ 3		0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Residual (ir,e)										
ครั้งที่ 1		0.0056	0.0007	0.0005	0.0004	0.0004	0.0003	0.0003	0.0003	0.0002
ครั้งที่ 2		0.0050	0.0006	0.0004	0.0004	0.0003	0.0003	0.0003	0.0002	0.0002
ครั้งที่ 3		0.0050	0.0006	0.0004	0.0004	0.0003	0.0003	0.0003	0.0002	0.0002
σ^2_{Rel}	ครั้งที่ 1		0.0007	0.0005	0.0004	0.0004	0.0003	0.0003	0.0003	0.0002
	ครั้งที่ 2		0.0006	0.0004	0.0004	0.0003	0.0003	0.0003	0.0002	0.0002
	ครั้งที่ 3		0.0006	0.0004	0.0004	0.0003	0.0003	0.0003	0.0002	0.0002
σ^2_{Abs}	ครั้งที่ 1		0.0006	0.0006	0.0004	0.0004	0.0003	0.0003	0.0003	0.0002
	ครั้งที่ 2		0.0006	0.0004	0.0004	0.0003	0.0003	0.0003	0.0002	0.0002
	ครั้งที่ 3		0.0006	0.0004	0.0004	0.0003	0.0003	0.0003	0.0002	0.0002
$E\rho^2$	ครั้งที่ 1		0.9954	0.9970	0.9974	0.9977	0.9980	0.9982	0.9983	0.9985
	ครั้งที่ 2		0.9960	0.9973	0.9977	0.9980	0.9982	0.9984	0.9985	0.9987
	ครั้งที่ 3		0.9959	0.9972	0.9976	0.9979	0.9982	0.9983	0.9985	0.9986
ϕ	ครั้งที่ 1		0.9954	0.9969	0.9974	0.9977	0.9980	0.9982	0.9983	0.9985
	ครั้งที่ 2		0.9960	0.9973	0.9977	0.9980	0.9982	0.9984	0.9985	0.9986
	ครั้งที่ 3		0.9958	0.9972	0.9976	0.9979	0.9981	0.9983	0.9985	0.9986

จากตารางที่ 43 ผลการศึกษา D ประกอบด้วยการประมาณค่าความแปรปรวนของผู้ตัดสิน จำนวน 8, 12, 14, 16, 18, 20, 22, 24 คน จากการตัดสินคะแนนจุดตัดของผู้ตัดสินจำนวน 12 คน ด้วยวิธีบุ๊คมาร์ค ที่ระดับดีเยี่ยม (A) ครั้งที่ 1 ความแปรปรวนของคะแนนความคลาดเคลื่อนสัมพันธ์(σ^2_{Rel}) ซึ่งเป็นความคลาดเคลื่อนที่เกิดจากผลร่วมของข้อสอบ และผู้ตัดสิน มีค่าลดลง เมื่อจำนวนผู้ตัดสินเพิ่มขึ้น ครั้งที่ 1 มีค่า ดังนี้

0.0007 0.0005 0.0004 0.0004 0.0003 0.0003 0.0003 0.0002 ครั้งที่ 2 มีค่า ดังนี้ 0.0006 0.0004 0.0004 0.0003 0.0003 0.0003 0.0002 0.0002 ครั้งที่ 3 มีค่า ดังนี้ 0.0006 0.0004 0.0004 0.0003 0.0003 0.0003 0.0002 0.0002 เมื่อมีผู้ตัดสินจำนวน 8, 12, 14, 16, 18, 20, 22, 24 คน ตามลำดับ

ส่วนความแปรปรวนของคะแนนความคลาดเคลื่อนสัมบูรณ์(σ^2_{Abs})ซึ่งเป็นผลรวมของความคลาดเคลื่อนที่เกิดจากผลร่วมของข้อสอบและผู้ตัดสิน มีค่าลดลงเมื่อจำนวนผู้ตัดสินเพิ่มขึ้น ครั้งที่ 1 มีค่าดังนี้ 0.0006 0.0006 0.0004 0.0004 0.0003 0.0003 0.0003 0.0002 ครั้งที่ 2 มีค่าดังนี้ 0.0006 0.0004 0.0004 0.0003 0.0003 0.0003 0.0002 0.0002 ครั้งที่ 3 มีค่าดังนี้ 0.0006 0.0004 0.0004 0.0003 0.0003 0.0003 0.0002 0.0002 เมื่อมีผู้ตัดสินจำนวน 8, 12, 14, 16, 18, 20, 22, 24 คน ตามลำดับ

ค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินแบบสัมพัทธ์ ($E\rho^2$) มีค่าเพิ่มขึ้นเมื่อจำนวนผู้ตัดสินเพิ่มขึ้น ครั้งที่ 1 มีค่าดังนี้ 0.9954 0.9970 0.9974 0.9977 0.9980 0.9982 0.9983 0.9985 ครั้งที่ 2 มีค่าดังนี้ 0.9960 0.9973 0.9977 0.9980 0.9982 0.9984 0.9985 0.9987 ครั้งที่ 3 มีค่าดังนี้ 0.9959 0.9972 0.9976 0.9979 0.9982 0.9983 0.9985 0.9986 เมื่อมีผู้ตัดสินจำนวน 8, 12, 14, 16, 18, 20, 22, 24 คน ตามลำดับ

ส่วนค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินแบบสัมบูรณ์ (ϕ) มีค่าเพิ่มขึ้นเมื่อจำนวนผู้ตัดสินเพิ่มขึ้น ครั้งที่ 1 มีค่าดังนี้ 0.9954 0.9969 0.9974 0.9977 0.9980 0.9982 0.9983 0.9985 ครั้งที่ 2 มีค่าดังนี้ 0.9960 0.9973 0.9977 0.9980 0.9982 0.9984 0.9985 0.9986 ครั้งที่ 3 มีค่าดังนี้ 0.9958 0.9972 0.9976 0.9979 0.9981 0.9983 0.9985 0.9986 เมื่อมีผู้ตัดสินจำนวน 8, 12, 14, 16, 18, 20, 22, 24 คน ตามลำดับ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 44 การวิเคราะห์ความแปรปรวนและการประมาณค่าความแปรปรวนของการกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์คที่คะแนนจุดตัดระดับดีมาก (B+) ในขั้นสรุปอ้างอิง (G-study)

Source of Variation	df	Sum of Squares	Mean Squares	Estimated Variance Component	Percentage of Total Variance
ครั้งที่ 1					
Item (i)	99	267.9825	2.7069	0.2254	99.1205
Rater (r)	11	0.0225	0.0021	0.0000	0.0000
Residual (ir,e)	1089	2.2275	0.0021	0.0020	0.8795
Total	1199	270.2325		0.2274	100
ครั้งที่ 2					
Item (i)	99	246.5000	2.4899	0.2061	91.8450
Rater (r)	11	2.4067	0.2188	0.0020	0.8913
Residual (ir,e)	1089	17.7600	0.0163	0.0163	7.2638
Total	1199	266.6667		0.2244	100
ครั้งที่ 3					
Item (i)	99	256.1158	2.5870	0.2143	92.1720
Rater (r)	11	3.5892	0.3263	0.0031	1.3333
Residual (ir,e)	1089	16.4942	0.0152	0.0151	6.4946
Total	1199	276.1992		0.2325	100

จากตารางที่ 44 เมื่อผู้ตัดสินกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์คที่คะแนนจุดตัดระดับดีมาก (B+) ครั้งที่ 1 โดยที่เอกภพของการสังเกตประกอบด้วย ข้อสอบ และผู้ตัดสิน ซึ่งมีขนาดไม่จำกัด โดยเป็นการประมาณค่าความแปรปรวนของแหล่งต่างๆ เมื่อเทียบกับความแปรปรวนรวม มีค่าดังนี้ ความแปรปรวนของข้อสอบ (σ_i^2) เป็นความแตกต่างของความยากของข้อสอบมีค่า 0.2254 คิดเป็นร้อยละ 99 ของความแปรปรวนรวม ความแปรปรวนของผู้ตัดสิน (σ_r^2) เป็นการประมาณค่าความแปรปรวนของคะแนนเอกภพที่แสดงถึงความแตกต่างในการตัดสินคะแนนจุดตัดของผู้ตัดสิน มีค่า 0.0000 คิดเป็นร้อยละ 0 ของความแปรปรวนรวม และความแปรปรวนของส่วนที่เหลือ (σ_{ir}^2) เป็นความแปรปรวนของผลร่วมระหว่างข้อสอบ

กับผู้ตัดสิน แสดงให้เห็นถึงความไม่คงเส้นคงวาของผู้ตัดสินแต่ละคนในการตัดสินข้อสอบแต่ละข้อ มีค่า 0.0020 คิดเป็นร้อยละ 1 ของความแปรปรวนรวม

เมื่อผู้ตัดสินกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์คที่คะแนนจุดตัดระดับดีมาก (B+) ครั้งที่ 2 มีค่าดังนี้ ความแปรปรวนของข้อสอบ (σ_i^2) เป็นความแตกต่างของความยากของข้อสอบมีค่า 0.2061 คิดเป็นร้อยละ 92 ของความแปรปรวนรวม ความแปรปรวนของผู้ตัดสิน (σ_r^2) เป็นการประมาณค่าความแปรปรวนของคะแนนเอกภพที่แสดงถึงความแตกต่างในการตัดสินคะแนนจุดตัดของผู้ตัดสิน มีค่า 0.0020 คิดเป็นร้อยละ 1 ของความแปรปรวนรวม และความแปรปรวนของส่วนที่เหลือ (σ_{ir}^2) เป็นความแปรปรวนของผลร่วมระหว่างข้อสอบกับผู้ตัดสิน แสดงให้เห็นถึงความไม่คงเส้นคงวาของผู้ตัดสินแต่ละคนในการตัดสินข้อสอบแต่ละข้อมีค่า 0.0163 คิดเป็นร้อยละ 7 ของความแปรปรวนรวม

เมื่อผู้ตัดสินกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์คที่คะแนนจุดตัดระดับดีมาก (B+) ครั้งที่ 3 มีค่าดังนี้ ความแปรปรวนของข้อสอบ (σ_i^2) เป็นความแตกต่างของความยากของข้อสอบมีค่า 0.2143 คิดเป็นร้อยละ 92 ของความแปรปรวนรวม ความแปรปรวนของผู้ตัดสิน (σ_r^2) เป็นการประมาณค่าความแปรปรวนของคะแนนเอกภพที่แสดงถึงความแตกต่างในการตัดสินคะแนนจุดตัดของผู้ตัดสิน มีค่า 0.0031 คิดเป็นร้อยละ 1.3333 ของความแปรปรวนรวม และความแปรปรวนของส่วนที่เหลือ (σ_{ir}^2) เป็นความแปรปรวนของผลร่วมระหว่างข้อสอบกับผู้ตัดสิน แสดงให้เห็นถึงความไม่คงเส้นคงวาของผู้ตัดสินแต่ละคนในการตัดสินข้อสอบแต่ละข้อมีค่า 0.0151 คิดเป็นร้อยละ 6.4946 ของความแปรปรวนรวม

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 45 การประมาณค่าความแปรปรวนในขั้นการสรุปอ้างอิง (G-study) ขั้นการตัดสินใจ (D-study)และสัมประสิทธิ์การอ้างอิงสรุปของการกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์คที่คะแนนจุดตัดระดับที่ระดับดีมาก (B+) จำแนกตามจำนวนผู้ตัดสิน

Source of Variation		Estimated G Study Variance Components	Alternative D Studies Design	Estimated Variance Component					
			8	12	14	16	18	20	22	24
Item (i)										
ครั้งที่ 1		0.2254	0.2254	0.2254	0.2254	0.2254	0.2254	0.2254	0.2254	0.2254
ครั้งที่ 2		0.2061	0.2061	0.2061	0.2061	0.2061	0.2061	0.2061	0.2061	0.2061
ครั้งที่ 3		0.2143	0.2143	0.2143	0.2143	0.2143	0.2143	0.2143	0.2143	0.2143
Rater (r)										
ครั้งที่ 1		0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
ครั้งที่ 2		0.0020	0.0003	0.0002	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
ครั้งที่ 3		0.0031	0.0004	0.0003	0.0002	0.0002	0.0002	0.0002	0.0001	0.0001
Residual (ir,e)										
ครั้งที่ 1		0.0020	0.0002	0.0002	0.0002	0.0001	0.0001	0.0001	0.0001	0.0001
ครั้งที่ 2		0.0163	0.0020	0.0014	0.0012	0.0010	0.0009	0.0008	0.0007	0.0007
ครั้งที่ 3		0.0151	0.0019	0.0013	0.0011	0.0010	0.0008	0.0008	0.0007	0.0006
σ^2_{Rel}	ครั้งที่ 1		0.0002	0.0002	0.0002	0.0001	0.0001	0.0001	0.0001	0.0001
	ครั้งที่ 2		0.0020	0.0014	0.0012	0.0010	0.0009	0.0008	0.0007	0.0007
	ครั้งที่ 3		0.0019	0.0013	0.0011	0.0010	0.0008	0.0008	0.0007	0.0006
σ^2_{Abs}	ครั้งที่ 1		0.0002	0.0002	0.0002	0.0001	0.0001	0.0001	0.0001	0.0001
	ครั้งที่ 2		0.0023	0.0016	0.0013	0.0011	0.0010	0.0009	0.0008	0.0008
	ครั้งที่ 3		0.0023	0.0016	0.0013	0.0012	0.0010	0.0010	0.0008	0.0007
$E\rho^2$	ครั้งที่ 1		0.9989	0.9992	0.9994	0.9994	0.9995	0.9996	0.9996	0.9996
	ครั้งที่ 2		0.9902	0.9935	0.9944	0.9951	0.9956	0.9961	0.9964	0.9967
	ครั้งที่ 3		0.9912	0.9942	0.9950	0.9956	0.9961	0.9965	0.9968	0.9971
ϕ	ครั้งที่ 1		0.9989	0.9992	0.9994	0.9994	0.9995	0.9996	0.9996	0.9996
	ครั้งที่ 2		0.9890	0.9926	0.9937	0.9945	0.9951	0.9956	0.9960	0.9963
	ครั้งที่ 3		0.9895	0.9930	0.9940	0.9947	0.9953	0.9958	0.9961	0.9965

จากตารางที่ 45 ผลการศึกษา D ประกอบด้วยการประมาณค่าความแปรปรวนของผู้ตัดสิน จำนวน 8, 12, 14, 16, 18, 20, 22, 24 คน จากการตัดสินคะแนนจุดตัดของผู้ตัดสินจำนวน 12 คน ด้วยวิธีบุ๊คมาร์คที่ระดับดีมาก (B+) ครั้งที่ 1 ความแปรปรวนของคะแนนความคลาดเคลื่อนสัมพันธ์($\sigma^2_{\mathrm{Re}l}$) ซึ่งเป็นความคลาดเคลื่อนที่เกิดจากผลร่วมของข้อสอบ และผู้ตัดสิน มีค่าลดลง เมื่อจำนวนผู้ตัดสินเพิ่มขึ้น ครั้งที่ 1 มีค่า ดังนี้ 0.0002 0.0002 0.0002 0.0001 0.0001 0.0001 0.0001 0.0001 ครั้งที่ 2 มีค่า ดังนี้ 0.0020 0.0014 0.0012 0.0010 0.0009 0.0008 0.0007 0.0007 ครั้งที่ 3 มีค่า ดังนี้ 0.0019 0.0013 0.0011 0.0010 0.0008 0.0008 0.0007 0.0006 เมื่อมีผู้ตัดสินจำนวน 8, 12, 14, 16, 18, 20, 22, 24 คน ตามลำดับ

ส่วนความแปรปรวนของคะแนนความคลาดเคลื่อนสัมบูรณ์(σ^2_{Abs})ซึ่งเป็นผลรวมของความคลาดเคลื่อนที่เกิดจากผลร่วมของข้อสอบและผู้ตัดสิน มีค่าลดลงเมื่อจำนวนผู้ตัดสินเพิ่มขึ้น ครั้งที่ 1 มีค่าดังนี้ 0.0002 0.0002 0.0002 0.0001 0.0001 0.0001 0.0001 0.0001 ครั้งที่ 2 มีค่าดังนี้ 0.0023 0.0016 0.0013 0.0011 0.0010 0.0009 0.0008 0.0008 ครั้งที่ 3 0.0023 0.0016 0.0013 0.0012 0.0010 0.0010 0.0008 0.0007 เมื่อมีผู้ตัดสินจำนวน 8, 12, 14, 16, 18, 20, 22, 24 คน ตามลำดับ

ค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินแบบสัมพัทธ์ ($\mathrm{E}\rho^2$) มีค่าเพิ่มขึ้นเมื่อจำนวนผู้ตัดสินเพิ่มขึ้น ครั้งที่ 1 มีค่าดังนี้ 0.9989 0.9992 0.9994 0.9994 0.9995 0.9996 0.9996 0.9996 0.9996 ครั้งที่ 2 มีค่าดังนี้ 0.9902 0.9935 0.9944 0.9951 0.9956 0.9961 0.9964 0.9967 ครั้งที่ 3 มีค่าดังนี้ 0.9912 0.9942 0.9950 0.9956 0.9961 0.9965 0.9968 0.9971 เมื่อมี ผู้ตัดสินจำนวน 8, 12, 14, 16, 18, 20, 22, 24 คน ตามลำดับ

ส่วนค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินแบบสัมบูรณ์ (ϕ) มีค่าเพิ่มขึ้นเมื่อจำนวนผู้ตัดสินเพิ่มขึ้น ครั้งที่ 1 มีค่าดังนี้ 0.9989 0.9992 0.9994 0.9994 0.9995 0.9996 0.9996 0.9996 ครั้งที่ 2 มีค่าดังนี้ 0.9890 0.9926 0.9937 0.9945 0.9951 0.9956 0.9960 0.9963 ครั้งที่ 3 มีค่าดังนี้ 0.9895 0.9930 0.9940 0.9947 0.9953 0.9958 0.9961 0.9965 เมื่อมีผู้ตัดสินจำนวน 8, 12, 14, 16, 18, 20, 22, 24 คน ตามลำดับ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 46 การวิเคราะห์ความแปรปรวนและการประมาณค่าความแปรปรวนของการกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์คที่คะแนนจุดตัดระดับดี (B)ในขั้นสรุปอ้างอิง (G-study)

Source of Variation	df	Sum of Squares	Mean Squares	Estimated Variance Component	Percentage of Total Variance
ครั้งที่ 1					
Item (i)	99	294.1967	2.9717	0.2473	98.4475
Rater (r)	11	0.0700	0.0064	0.0000	0.0000
Residual (ir,e)	1089	4.2633	0.0039	0.0039	1.5525
Total	1199	298.5300		0.2512	100
ครั้งที่ 2					
Item (i)	99	293.0833	2.9604	0.2462	97.8149
Rater (r)	11	0.1300	0.0118	0.0000	0.0000
Residual (ir,e)	1089	6.0367	0.0055	0.0055	2.1851
Total	1199	299.2500		0.2517	100
ครั้งที่ 3					
Item (i)	99	291.1467	2.9409	0.2445	97.1009
Rater (r)	11	0.1867	0.0170	0.0001	0.0397
Residual (ir,e)	1089	7.8133	0.0072	0.0072	2.8594
Total	1199	299.1467		0.2518	100

จากตารางที่ 46 เมื่อผู้ตัดสินกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์คที่คะแนนจุดตัดระดับดี (B) ครั้งที่ 1 โดยที่เอกภพของการสังเกตประกอบด้วย ข้อสอบ และผู้ตัดสิน ซึ่งมีขนาดไม่จำกัด โดยเป็นการประมาณค่าความแปรปรวนของแหล่งต่างๆ เมื่อเทียบกับความแปรปรวนรวม มีค่าดังนี้ ความแปรปรวนของข้อสอบ (σ_i^2) เป็นความแตกต่างของความยากของข้อสอบมีค่า 0.2473 คิดเป็นร้อยละ 98 ของความแปรปรวนรวม ความแปรปรวนของผู้ตัดสิน (σ_r^2) เป็นการประมาณค่าความแปรปรวนของคะแนนเอกภพที่แสดงถึงความแตกต่างในการตัดสินคะแนนจุดตัดของผู้ตัดสิน มีค่า 0.0000 คิดเป็นร้อยละ 0 ของความแปรปรวนรวม และความแปรปรวนของส่วนที่เหลือ (σ_{ir}^2) เป็นความแปรปรวนของผลร่วมระหว่างข้อสอบกับผู้

ตัดสิน แสดงให้เห็นถึง ความไม่คงเส้นคงวาของผู้ตัดสินแต่ละคนในการตัดสินข้อสอบแต่ละข้อมีค่า 0.0039 คิดเป็นร้อยละ 2 ของความแปรปรวนรวม

เมื่อผู้ตัดสินกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์คที่คะแนนจุดตัดระดับดี (B) ครั้งที่ 2 มีค่าดังนี้ ความแปรปรวนของข้อสอบ (σ_i^2) เป็นความแตกต่างของความยากของข้อสอบมีค่า 0.2462 คิดเป็นร้อยละ 98 ของความแปรปรวนรวม ความแปรปรวนของผู้ตัดสิน (σ_r^2) เป็นการประมาณค่าความแปรปรวนของคะแนนเอกภาพที่แสดงถึงความแตกต่างในการตัดสินคะแนนจุดตัดของผู้ตัดสิน มีค่า 0.0000 คิดเป็นร้อยละ 0 ของความแปรปรวนรวม และความแปรปรวนของส่วนที่เหลือ (σ_{ir}^2) เป็นความแปรปรวนของผลร่วมระหว่างข้อสอบกับผู้ตัดสิน แสดงให้เห็นถึง ความไม่คงเส้นคงวาของผู้ตัดสินแต่ละคนในการตัดสินข้อสอบแต่ละข้อ มีค่า 0.0055 คิดเป็นร้อยละ 2 ของความแปรปรวนรวม

เมื่อผู้ตัดสินกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์คที่คะแนนจุดตัดระดับดี (B) ครั้งที่ 3 มีค่าดังนี้ ความแปรปรวนของข้อสอบ (σ_i^2) เป็นความแตกต่างของความยากของข้อสอบมีค่า .2445 คิดเป็นร้อยละ 97 ของความแปรปรวนรวม ความแปรปรวนของผู้ตัดสิน (σ_r^2) เป็นการประมาณค่าความแปรปรวนของคะแนนเอกภาพที่แสดงถึงความแตกต่างในการตัดสินคะแนนจุดตัดของผู้ตัดสิน มีค่า .0001 คิดเป็นร้อยละ 0 ของความแปรปรวนรวม และความแปรปรวนของส่วนที่เหลือ (σ_{ir}^2) เป็นความแปรปรวนของผลร่วมระหว่างข้อสอบกับผู้ตัดสิน แสดงให้เห็นถึงความไม่คงเส้นคงวาของผู้ตัดสินแต่ละคนในการตัดสินข้อสอบแต่ละข้อ มีค่า .0072 คิดเป็นร้อยละ 3 ของความแปรปรวนรวม

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 47 การประมาณค่าความแปรปรวนในขั้นการสรุปอ้างอิง (G-study) ขั้นการตัดสินใจ (D-study)และสัมประสิทธิ์การอ้างอิงสรุปของการกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์คที่คะแนนจุดตัดระดับระดับดี (B) จำแนกตามจำนวนผู้ตัดสิน

Source of Variation	Estimated G Study Variance Components	Alternative D Studies Design	Estimated Variance Component						
		8	12	14	16	18	20	22	24
Item (i)									
ครั้งที่ 1	0.2473	0.2473	0.2473	0.2473	0.2473	0.2473	0.2473	0.2473	0.2473
ครั้งที่ 2	0.2462	0.2462	0.2462	0.2462	0.2462	0.2462	0.2462	0.2462	0.2462
ครั้งที่ 3	0.2445	0.2445	0.2445	0.2445	0.2445	0.2445	0.2445	0.2445	0.2445
Rater (r)									
ครั้งที่ 1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
ครั้งที่ 2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
ครั้งที่ 3	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Residual (ir,e)									
ครั้งที่ 1	0.0039	0.0005	0.0003	0.0003	0.0002	0.0002	0.0002	0.0002	0.0002
ครั้งที่ 2	0.0055	0.0007	0.0005	0.0004	0.0004	0.0003	0.0003	0.0003	0.0002
ครั้งที่ 3	0.0072	0.0009	0.0006	0.0005	0.0005	0.0004	0.0004	0.0003	0.0003
σ^2_{Rel} ครั้งที่ 1		0.0005	0.0003	0.0003	0.0002	0.0002	0.0002	0.0002	0.0002
ครั้งที่ 2		0.0007	0.0005	0.0004	0.0004	0.0003	0.0003	0.0003	0.0002
ครั้งที่ 3		0.0009	0.0006	0.0005	0.0005	0.0004	0.0004	0.0003	0.0003
σ^2_{Abs} ครั้งที่ 1		0.0005	0.0003	0.0003	0.0002	0.0002	0.0002	0.0002	0.0002
ครั้งที่ 2		0.0007	0.0005	0.0004	0.0004	0.0003	0.0003	0.0003	0.0002
ครั้งที่ 3		0.0009	0.0006	0.0005	0.0005	0.0004	0.0004	0.0003	0.0003
$E\rho^2$ ครั้งที่ 1		0.9980	0.9987	0.9989	0.9990	0.9991	0.9992	0.9993	0.9993
ครั้งที่ 2		0.9972	0.9981	0.9984	0.9986	0.9988	0.9989	0.9990	0.9991
ครั้งที่ 3		0.9963	0.9976	0.9979	0.9982	0.9984	0.9985	0.9987	0.9988
ϕ ครั้งที่ 1		0.9980	0.9987	0.9987	0.9990	0.9991	0.9992	0.9993	0.9993
ครั้งที่ 2		0.9972	0.9981	0.9984	0.9986	0.9987	0.9989	0.9990	0.9991
ครั้งที่ 3		0.9963	0.9975	0.9979	0.9981	0.9984	0.9985	0.9987	0.9988

จากตารางที่ 47 ผลการศึกษา D ประกอบด้วยการประมาณค่าความแปรปรวนของผู้ตัดสิน จำนวน 8, 12, 14, 16, 18, 20, 22, 24 คน จากการตัดสินคะแนนจุดตัดของผู้ตัดสินจำนวน 12 คน ด้วยวิธีบุ๊คมาร์คที่ระดับดี (B) ครั้งที่ 1 ความแปรปรวนของคะแนนความคลาดเคลื่อนสัมพันธ์(σ^2_{Rel}) ซึ่งเป็นความคลาดเคลื่อนที่เกิดจากผลร่วมของข้อสอบ และผู้ตัดสิน มีค่าลดลง เมื่อจำนวนผู้ตัดสินเพิ่มขึ้น ครั้งที่ 1 มีค่า ดังนี้ 0.0005 0.0003 0.0003 0.0002 0.0002 0.0002 0.0002 0.0002 ครั้งที่ 2 มีค่า ดังนี้ 0.0007 0.0005 0.0004 0.0004 0.0003 0.0003 0.0003 0.0002 ครั้งที่ 3 มีค่า ดังนี้ 0.0009 0.0006 0.0005 0.0005 0.0004 0.0004 0.0003 0.0003 เมื่อมีผู้ตัดสินจำนวน 8, 12, 14, 16, 18, 20, 22, 24 คน ตามลำดับ

ส่วนความแปรปรวนของคะแนนความคลาดเคลื่อนสัมบูรณ์(σ^2_{Abs})ซึ่งเป็นผลรวมของความคลาดเคลื่อนที่เกิดจากผลร่วมของข้อสอบและผู้ตัดสิน มีค่าลดลงเมื่อจำนวนผู้ตัดสินเพิ่มขึ้น ครั้งที่ 1 มีค่าดังนี้ 0.0005 0.0003 0.0003 0.0002 0.0002 0.0002 0.0002 0.0002 ครั้งที่ 2 มีค่าดังนี้ 0.0007 0.0005 0.0004 0.0004 0.0003 0.0003 0.0003 0.0002 ครั้งที่ 3 มีค่าดังนี้ 0.0009 0.0006 0.0005 0.0005 0.0004 0.0004 0.0003 0.0003 เมื่อมีผู้ตัดสินจำนวน 8, 12, 14, 16, 18, 20, 22, 24 คน ตามลำดับ

ค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินแบบสัมพัทธ์ ($E\rho^2$) มีค่าเพิ่มขึ้นเมื่อจำนวนผู้ตัดสินเพิ่มขึ้น ครั้งที่ 1 มีค่าดังนี้ 0.9980 0.9987 0.9989 0.9990 0.9991 0.9992 0.9993 0.9993 ครั้งที่ 2 มีค่าดังนี้ 0.9972 0.9981 0.9984 0.9986 0.9988 0.9989 0.9990 0.9991 ครั้งที่ 3 มีค่าดังนี้ 0.9963 0.9976 0.9979 0.9982 0.9984 0.9985 0.9987 0.9988 เมื่อมีผู้ตัดสินจำนวน 8, 12, 14, 16, 18, 20, 22, 24 คน ตามลำดับ

ส่วนค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินแบบสัมบูรณ์ (ϕ) มีค่าเพิ่มขึ้นเมื่อจำนวนผู้ตัดสินเพิ่มขึ้น ครั้งที่ 1 มีค่าดังนี้ 0.9980 0.9987 0.9987 0.9990 0.9991 0.9992 0.9993 0.9993 ครั้งที่ 2 มีค่าดังนี้ 0.9972 0.9981 0.9984 0.9986 0.9987 0.9989 0.9990 0.9991 ครั้งที่ 3 มีค่าดังนี้ 0.9963 0.9975 0.9979 0.9981 0.9984 0.9985 0.9987 0.9988 เมื่อมีผู้ตัดสินจำนวน 8, 12, 14, 16, 18, 20, 22, 24 คน ตามลำดับ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 48 การวิเคราะห์ความแปรปรวนและการประมาณค่าความแปรปรวนของการกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์คที่คะแนนจุดตัดระดับดีพอใช้ (C+) ในขั้นสรุปอ้างอิง (G-study)

Source of Variation	df	Sum of Squares	Mean Squares	Estimated Variance Component	Percentage of Total Variance
ครั้งที่ 1					
Item (i)	99	253.4158	2.5598	0.2124	94.7791
Rater (r)	11	0.8225	0.0748	0.0006	0.2677
Residual (ir,e)	1089	12.0942	0.0111	0.0111	4.9531
Total	1199	266.3325		0.2241	100
ครั้งที่ 2					
Item (i)	99	262.8033	2.6548	0.2208	97.6559
Rater (r)	11	0.0967	0.0088	0.0000	0.0000
Residual (ir,e)	1089	5.7367	0.0053	0.0053	2.3441
Total	1199	268.6367		0.2261	100
ครั้งที่ 3					
Item (i)	99	263.9467	2.6661	0.2218	97.8817
Rater (r)	11	0.0800	0.0073	0.0000	0.0000
Residual (ir,e)	1089	5.2533	0.0048	0.0048	2.1183
Total	1199	269.2800		0.2266	100

จากตารางที่ 48 เมื่อผู้ตัดสินกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์คที่คะแนนจุดตัดระดับดีพอใช้ (C+) ครั้งที่ 1 โดยที่เอกภพของการสังเกตประกอบด้วย ข้อสอบ และผู้ตัดสิน ซึ่งมีขนาดไม่จำกัด โดยเป็นการประมาณค่าความแปรปรวนของแหล่งต่างๆ เมื่อเทียบกับความแปรปรวนรวม มีค่าดังนี้ ความแปรปรวนของข้อสอบ (σ_i^2) เป็นความแตกต่างของความยากของข้อสอบมีค่า 0.2124 คิดเป็นร้อยละ 95 ของความแปรปรวนรวม ความแปรปรวนของผู้ตัดสิน (σ_r^2) เป็นการประมาณค่าความแปรปรวนของคะแนนเอกภพที่แสดงถึงความแตกต่างในการตัดสินคะแนนจุดตัดของผู้ตัดสิน มีค่า 0.0006 คิดเป็นร้อยละ 0 ของความแปรปรวนรวม และความแปรปรวนของส่วนที่เหลือ (σ_{ir}^2) เป็นความแปรปรวนของผลร่วมระหว่างข้อสอบ

กับผู้ตัดสิน แสดงให้เห็นถึง ความไม่คงเส้นคงวาของผู้ตัดสินแต่ละคนในการตัดสินข้อสอบแต่ละข้อมีค่า 0.0111 คิดเป็นร้อยละ 5 ของความแปรปรวนรวม

เมื่อผู้ตัดสินกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์คที่คะแนนจุดตัดระดับดีพอใช้ (C+) ครั้งที่ 2 มีค่าดังนี้ ความแปรปรวนของข้อสอบ (σ_i^2) เป็นความแตกต่างของความยากของข้อสอบมีค่า 0.2208 คิดเป็นร้อยละ 98 ของความแปรปรวนรวม ความแปรปรวนของผู้ตัดสิน (σ_r^2) เป็นการประมาณค่าความแปรปรวนของคะแนนเอกภพที่แสดงถึงความแตกต่างในการตัดสินคะแนนจุดตัดของผู้ตัดสิน มีค่า 0.0000 คิดเป็นร้อยละ 0 ของความแปรปรวนรวม และความแปรปรวนของส่วนที่เหลือ (σ_{ir}^2) เป็นความแปรปรวนของผลร่วมระหว่างข้อสอบกับผู้ตัดสิน แสดงให้เห็นถึง ความไม่คงเส้นคงวาของผู้ตัดสินแต่ละคนในการตัดสินข้อสอบแต่ละข้อ มีค่า 0.0053 คิดเป็นร้อยละ 2 ของความแปรปรวนรวม

เมื่อผู้ตัดสินกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์คที่คะแนนจุดตัดระดับดีพอใช้ (C+) ครั้งที่ 3 มีค่าดังนี้ ความแปรปรวนของข้อสอบ (σ_i^2) เป็นความแตกต่างของความยากของข้อสอบมีค่า 0.2218 คิดเป็นร้อยละ 98 ของความแปรปรวนรวม ความแปรปรวนของผู้ตัดสิน (σ_r^2) เป็นการประมาณค่าความแปรปรวนของคะแนนเอกภพที่แสดงถึงความแตกต่างในการตัดสินคะแนนจุดตัดของผู้ตัดสิน มีค่า 0.0000 คิดเป็นร้อยละ 0 ของความแปรปรวนรวม และความแปรปรวนของส่วนที่เหลือ (σ_{ir}^2) เป็นความแปรปรวนของผลร่วมระหว่างข้อสอบกับผู้ตัดสินแสดงให้เห็นถึง ความไม่คงเส้นคงวาของผู้ตัดสินแต่ละคนในการตัดสินข้อสอบแต่ละข้อ มีค่า 0.0048 คิดเป็นร้อยละ 2.1183 ของความแปรปรวนรวม

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 49 การประมาณค่าความแปรปรวนในขั้นการสรุปอ้างอิง (G-study) ขั้นการตัดสินใจ (D-study) และสัมประสิทธิ์การอ้างอิงสรุปของการกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์คที่คะแนนจุดตัดระดับดีพอใช้ (C+) จำแนกตามจำนวนผู้ตัดสิน

Source of Variation		Estimated G Study Variance Components	Alternative Estimated D Studies Design Variance Component 8	12	14	16	18	20	22	24
Item (i)										
ครั้งที่ 1		0.2124	0.2124	0.2124	0.2124	0.2124	0.2124	0.2124	0.2124	0.2124
ครั้งที่ 2		0.2208	0.2208	0.2208	0.2208	0.2208	0.2208	0.2208	0.2208	0.2208
ครั้งที่ 3		0.2218	0.2218	0.2218	0.2218	0.2218	0.2218	0.2218	0.2218	0.2218
Rater (r)										
ครั้งที่ 1		0.0006	0.0001	0.0001	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
ครั้งที่ 2		0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
ครั้งที่ 3		0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Residual (ir,e)										
ครั้งที่ 1		0.0111	0.0014	0.0009	0.0008	0.0007	0.0006	0.0006	0.0005	0.0005
ครั้งที่ 2		0.0053	0.0007	0.0004	0.0004	0.0003	0.0003	0.0003	0.0002	0.0002
ครั้งที่ 3		0.0048	0.0006	0.0004	0.0003	0.0003	0.0003	0.0002	0.0002	0.0002
$\sigma^2_{\mathrm{Re}l}$	ครั้งที่ 1		0.0014	0.0009	0.0008	0.0007	0.0006	0.0006	0.0005	0.0005
	ครั้งที่ 2		0.0007	0.0004	0.0004	0.0003	0.0003	0.0003	0.0002	0.0002
	ครั้งที่ 3		0.0006	0.0004	0.0003	0.0003	0.0003	0.0002	0.0002	0.0002
σ^2_{Abs}	ครั้งที่ 1		0.0015	0.0010	0.0009	0.0007	0.0006	0.0006	0.0005	0.0005
	ครั้งที่ 2		0.0007	0.0004	0.0004	0.0003	0.0003	0.0003	0.0002	0.0002
	ครั้งที่ 3		0.0006	0.0004	0.0003	0.0003	0.0003	0.0002	0.0002	0.0002
$\mathrm{E}\rho^2$	ครั้งที่ 1		0.9935	0.9957	0.9963	0.9967	0.9971	0.9974	0.9976	0.9978
	ครั้งที่ 2		0.9970	0.9980	0.9983	0.9985	0.9987	0.9988	0.9989	0.9990
	ครั้งที่ 3		0.9973	0.9982	0.9985	0.9986	0.9988	0.9989	0.9990	0.9991
ϕ	ครั้งที่ 1		0.9931	0.9954	0.9961	0.9966	0.9969	0.9972	0.9975	0.9977
	ครั้งที่ 2		0.9970	0.9980	0.9983	0.9985	0.9987	0.9988	0.9989	0.9990
	ครั้งที่ 3		0.9973	0.9982	0.9984	0.9986	0.9988	0.9989	0.9990	0.9991

จากตารางที่ 49 ผลการศึกษา D ประกอบด้วยการประมาณค่าความแปรปรวนของผู้ตัดสิน จำนวน 8, 12, 14, 16, 18, 20, 22, 24 คน จากการตัดสินคะแนนจุดตัดของผู้ตัดสินจำนวน 12 คน ด้วยวิธีบุ๊คมาร์ค ที่ระดับดีพอใช้ (C+) ครั้งที่ 1 ความแปรปรวนของคะแนนความคลาดเคลื่อนสัมพันธ์(σ^2_{Rel}) ซึ่งเป็นความคลาดเคลื่อนที่เกิดจากผลร่วมของข้อสอบ และผู้ตัดสิน มีค่าลดลง เมื่อจำนวนผู้ตัดสินเพิ่มขึ้น มีค่า ดังนี้ 0.0014 0.0009 0.0008 0.0007 0.0006 0.0006 0.0005 0.0005 ครั้งที่ 2 มีค่า ดังนี้ 0.0007 0.0004 0.0004 0.0003 0.0003 0.0003 0.0002 0.0002 ครั้งที่ 3 มีค่า ดังนี้ 0.0006 0.0004 0.0003 0.0003 0.0003 0.0002 0.0002 0.0002 เมื่อมีผู้ตัดสินจำนวน 8, 12, 14, 16, 18, 20, 22, 24 คน ตามลำดับ

ส่วนความแปรปรวนของคะแนนความคลาดเคลื่อนสัมบูรณ์(σ^2_{Abs})ซึ่งเป็นผลรวมของความคลาดเคลื่อนที่เกิดจากผลร่วมของข้อสอบและผู้ตัดสิน มีค่าลดลงเมื่อจำนวนผู้ตัดสินเพิ่มขึ้น ครั้งที่ 1 มีค่าดังนี้ 0.0015 0.0010 0.0009 0.0007 0.0006 0.0006 0.0005 0.0005 ครั้งที่ 2 มีค่าดังนี้ 0.0007 0.0004 0.0004 0.0003 0.0003 0.0003 0.0002 0.0002 ครั้งที่ 3 มีค่าดังนี้ 0.0006 0.0004 0.0003 0.0003 0.0003 0.0002 0.0002 0.0002 เมื่อมีผู้ตัดสินจำนวน 8, 12, 14, 16, 18, 20, 22, 24 คน ตามลำดับ

ค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินแบบสัมพัทธ์ ($E\rho^2$) มีค่าเพิ่มขึ้นเมื่อจำนวนผู้ตัดสินเพิ่มขึ้น ครั้งที่ 1 มีค่าดังนี้ 0.9935 0.9957 0.9963 0.9967 0.9971 0.9974 0.9976 0.9978 ครั้งที่ 2 มีค่าดังนี้ 0.9970 0.9980 0.9983 0.9985 0.9987 0.9988 0.9989 0.9990 ครั้งที่ 3 มีค่าดังนี้ 0.9973 0.9982 0.9985 0.9986 0.9988 0.9989 0.9990 0.9991 เมื่อมีผู้ตัดสินจำนวน 8, 12, 14, 16, 18, 20, 22, 24 คน ตามลำดับ

ส่วนค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินแบบสัมบูรณ์ (ϕ) มีค่าเพิ่มขึ้นเมื่อจำนวนผู้ตัดสินเพิ่มขึ้น ครั้งที่ 1 มีค่าดังนี้ 0.9931 0.9954 0.9961 0.9966 0.9969 0.9972 0.9975 0.9977 ครั้งที่ 2 มีค่าดังนี้ 0.9970 0.9980 0.9983 0.9985 0.9987 0.9988 0.9989 0.9990 ครั้งที่ 3 มีค่าดังนี้ 0.9973 0.9982 0.9984 0.9986 0.9988 0.9989 0.9990 0.9991 เมื่อมีผู้ตัดสินจำนวน 8, 12, 14, 16, 18, 20, 22, 24 คน ตามลำดับ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 50 การวิเคราะห์ความแปรปรวนและการประมาณค่าความแปรปรวนของการกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์คที่คะแนนจุดตัดพอใช้ (C) ในขั้นสรุปอ้างอิง (G-study)

Source of Variation	df	Sum of Squares	Mean Squares	Estimated Variance Component	Percentage of Total Variance
ครั้งที่ 1					
Item (i)	99	218.8800	2.2109	0.1842	100
Rater (r)	11	0.0000	0.0000	0.0000	0.0000
Residual (ir,e)	1089	0.0000	0.0000	0.0000	0.0000
Total	1199	218.8800		0.1842	100
ครั้งที่ 2					
Item (i)	99	216.3958	2.1858	0.1817	97.0101
Rater (r)	11	0.2492	0.0227	0.0002	0.1068
Residual (ir,e)	1089	5.8342	0.0054	0.0054	2.8831
Total	1199	222.4792		0.1873	100
ครั้งที่ 3					
Item (i)	99	218.4825	2.2069	0.1838	99.5667
Rater (r)	11	0.0092	0.0008	0.0000	0.0000
Residual (ir,e)	1089	0.9075	0.0008	0.0008	0.4334
Total	1199	219.3992		0.1846	100

จากตารางที่ 50 เมื่อผู้ตัดสินกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์คที่คะแนนจุดตัดระดับพอใช้ (C) ครั้งที่ 1 โดยที่เอกภพของการสังเกตประกอบด้วย ข้อสอบ และผู้ตัดสิน ซึ่งมีขนาดไม่จำกัด โดยเป็นการประมาณค่าความแปรปรวนของแหล่งต่างๆ เมื่อเทียบกับความแปรปรวนรวม มีค่าดังนี้ ความแปรปรวนของข้อสอบ (σ_i^2) เป็นความแตกต่างของความยากของข้อสอบมีค่า 0.1842 คิดเป็นร้อยละ 100 ของความแปรปรวนรวม ความแปรปรวนของผู้ตัดสิน (σ_r^2) เป็นการประมาณค่าความแปรปรวนของคะแนนเอกภพที่แสดงถึงความแตกต่างในการตัดสินคะแนนจุดตัดของผู้ตัดสิน มีค่า 0.0000 คิดเป็นร้อยละ 0 ของความแปรปรวนรวม และความแปรปรวนของส่วนที่เหลือ (σ_{ir}^2) เป็นความแปรปรวนของผลร่วมระหว่างข้อสอบ

กับผู้ตัดสิน แสดงให้เห็นถึง ความไม่คงเส้นคงวาของผู้ตัดสินแต่ละคนในการตัดสินข้อสอบแต่ละข้อมีค่า 0.0000 คิดเป็นร้อยละ 0 ของความแปรปรวนรวม

เมื่อผู้ตัดสินกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์คที่คะแนนจุดตัดระดับพอใช้ (C) ครั้งที่ 2 มีค่าดังนี้ ความแปรปรวนของข้อสอบ (σ_i^2) เป็นความแตกต่างของความยากของข้อสอบมีค่า 0.1817 คิดเป็นร้อยละ 97 ของความแปรปรวนรวม ความแปรปรวนของผู้ตัดสิน (σ_r^2) เป็นการประมาณค่าความแปรปรวนของคะแนนเอกภพที่แสดงถึงความแตกต่างในการตัดสินคะแนนจุดตัดของผู้ตัดสิน มีค่า 0.0002 คิดเป็นร้อยละ 0 ของความแปรปรวนรวม และความแปรปรวนของส่วนที่เหลือ (σ_{ir}^2) เป็นความแปรปรวนของผลร่วมระหว่างข้อสอบกับผู้ตัดสิน แสดงให้เห็นถึงความไม่คงเส้นคงวาของผู้ตัดสินแต่ละคนในการตัดสินข้อสอบแต่ละข้อ มีค่า 0.0054 คิดเป็นร้อยละ 3 ของความแปรปรวนรวม

เมื่อผู้ตัดสินกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์คที่คะแนนจุดตัดระดับพอใช้ (C) ครั้งที่ 3 มีค่าดังนี้ ความแปรปรวนของข้อสอบ (σ_i^2) เป็นความแตกต่างของความยากของข้อสอบมีค่า 0.1838 คิดเป็นร้อยละ 100 ของความแปรปรวนรวม ความแปรปรวนของผู้ตัดสิน (σ_r^2) เป็นการประมาณค่าความแปรปรวนของคะแนนเอกภพที่แสดงถึงความแตกต่างในการตัดสินคะแนนจุดตัดของผู้ตัดสิน มีค่า 0.0000 คิดเป็นร้อยละ 0 ของความแปรปรวนรวม และความแปรปรวนของส่วนที่เหลือ (σ_{ir}^2) เป็นความแปรปรวนของผลร่วมระหว่างข้อสอบกับผู้ตัดสิน แสดงให้เห็นถึงความไม่คงเส้นคงวาของผู้ตัดสินแต่ละคนในการตัดสินข้อสอบแต่ละข้อมีค่า 0.0008 คิดเป็นร้อยละ 0 ของความแปรปรวนรวม

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 51 การประมาณค่าความแปรปรวนในขั้นการสรุปอ้างอิง (G-study) ขั้นการตัดสินใจ (D-study) และสัมประสิทธิ์การอ้างอิงสรุปของการกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์คที่คะแนนจุดตัดระดับระดับพอใช้ (C) จำแนกตามจำนวนผู้ตัดสิน

Source of Variation		Estimated G Study Variance Components	Alternative D Studies Design	Estimated Variance Component					
			8	12	14	16	18	20	22	24
Item (i)										
ครั้งที่ 1		0.1842	0.0259	0.0259	0.0259	0.0259	0.0259	0.0259	0.0259	0.0259
ครั้งที่ 2		0.1817	0.1817	0.1817	0.1817	0.1817	0.1817	0.1817	0.1817	0.1817
ครั้งที่ 3		0.1838	0.1838	0.1838	0.1838	0.1838	0.1838	0.1838	0.1838	0.1838
Rater (r)										
ครั้งที่ 1		0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
ครั้งที่ 2		0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
ครั้งที่ 3		0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Residual (ir,e)										
ครั้งที่ 1		0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
ครั้งที่ 2		0.0054	0.0007	0.0005	0.0004	0.0003	0.0003	0.0003	0.0003	0.0002
ครั้งที่ 3		0.0008	0.0001	0.0001	0.0001	0.0001	0.0001	0.0000	0.0000	0.0000
σ^2_{Rel}	ครั้งที่ 1		0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	ครั้งที่ 2		0.0007	0.0005	0.0004	0.0003	0.0003	0.0003	0.0003	0.0002
	ครั้งที่ 3		0.0001	0.0001	0.0001	0.0001	0.0001	0.0000	0.0000	0.0000
σ^2_{Abs}	ครั้งที่ 1		0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	ครั้งที่ 2		0.0007	0.0005	0.0004	0.0003	0.0003	0.0003	0.0003	0.0002
	ครั้งที่ 3		0.0001	0.0001	0.0001	0.0001	0.0001	0.0000	0.0000	0.0000
$E\rho^2$	ครั้งที่ 1		1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	ครั้งที่ 2		0.9963	0.9976	0.9979	0.9982	0.9984	0.9985	0.9987	0.9988
	ครั้งที่ 3		0.9994	0.9996	0.9997	0.9997	0.9998	0.9998	0.9998	0.9998
ϕ	ครั้งที่ 1		1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	ครั้งที่ 2		0.9962	0.9975	0.9978	0.9981	0.9983	0.9985	0.9986	0.9987
	ครั้งที่ 3		0.9994	0.9996	0.9997	0.9997	0.9998	0.9998	0.9998	0.9998

จากตารางที่ 51 ผลการศึกษา D ประกอบด้วยการประมาณค่าความแปรปรวนของผู้ตัดสิน จำนวน 8, 12, 14, 16, 18, 20, 22, 24 คน จากการตัดสินคะแนนจุดตัดของผู้ตัดสินจำนวน 12 คน ด้วยวิธีบุ๊คมาร์คที่ระดับพอใช้ (C) ครั้งที่ 1 ความแปรปรวนของคะแนนความคลาดเคลื่อนสัมพันธ์(σ^2_{Rel}) ซึ่งเป็นความคลาดเคลื่อนที่เกิดจากผลร่วมของข้อสอบ และผู้ตัดสิน มีค่า 0.0000 ครั้งที่ 2 มีค่า ดังนี้ 0.0007 0.0005 0.0004 0.0003 0.0003 0.0003 0.0003 0.0002 ครั้งที่ 3 มีค่า ดังนี้ 0.0001 0.0001 0.0001 0.0001 0.0001 0.0000 0.0000 0.0000 เมื่อมีผู้ตัดสินจำนวน 8, 12, 14, 16, 18, 20, 22, 24 คน

ส่วนความแปรปรวนของคะแนนความคลาดเคลื่อนสัมบูรณ์(σ^2_{Abs})ซึ่งเป็นผลรวมของความคลาดเคลื่อนที่เกิดจากผลร่วมของข้อสอบและผู้ตัดสิน ครั้งที่ 1 มีค่า 0.0000 ครั้งที่ 2 มีค่า ดังนี้ 0.0007 0.0005 0.0004 0.0003 0.0003 0.0003 0.0003 0.0002 ครั้งที่ 3 มีค่าดังนี้ 0.0001 0.0001 0.0001 0.0001 0.0001 0.0000 0.0000 0.0000 เมื่อมีผู้ตัดสินจำนวน 8, 12, 14, 16, 18, 20, 22, 24 คน

ค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินแบบสัมพัทธ์ ($E\rho^2$) ครั้งที่ 1 มีค่า 1.0000 ครั้งที่ 2 มีค่าดังนี้ 0.9963 0.9976 0.9979 0.9982 0.9984 0.9985 0.9987 0.9988 ครั้งที่ 3 มีค่าดังนี้ 0.9994 0.9996 0.9997 0.9997 0.9998 0.9998 0.9998 0.9998 เมื่อมีผู้ตัดสินจำนวน 8, 12, 14, 16, 18, 20, 22, 24 คน

ส่วนค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินแบบสัมบูรณ์ ครั้งที่ 1 มีค่า 1.0000 ครั้งที่ 2 มีค่าดังนี้ 0.9962 0.9975 0.9978 0.9981 0.9983 0.9985 0.9986 0.9987 ครั้งที่ 3 มีค่าดังนี้ 0.9994 0.9996 0.9997 0.9997 0.9998 0.9998 0.9998 0.9998 เมื่อมีผู้ตัดสินจำนวน 8, 12, 14, 16, 18, 20, 22, 24 คน

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 52 การวิเคราะห์ความแปรปรวนและการประมาณค่าความแปรปรวนของการกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์คที่คะแนนจุดตัดระดับอ่อน (D+) ในขั้นสรุปอ้างอิง (G-study)

Source of Variation	df	Sum of Squares	Mean Squares	Estimated Variance Component	Percentage of Total Variance
ครั้งที่ 1					
Item (i)	99	148.6425	1.5014	0.1249	98.4240
Rater (r)	11	0.0225	0.0021	0.0000	0.0000
Residual (ir,e)	1089	2.2275	0.0021	0.0020	1.5760
Total	1199	150.8925		0.1269	100
ครั้งที่ 2					
Item (i)	99	148.6425	1.5014	0.1250	98.4250
Rater (r)	11	0.0225	0.0021	0.0000	0.0000
Residual (ir,e)	1089	2.2275	0.0021	0.0020	1.5748
Total	1199	150.8925		0.1270	100
ครั้งที่ 3					
Item (i)	99	148.4300	1.4993	0.1247	97.7273
Rater (r)	11	0.0367	0.0033	0.0000	0.0000
Residual (ir,e)	1089	3.1300	0.0029	0.0029	2.2727
Total	1199	151.5967		0.1276	100

จากตารางที่ 52 เมื่อผู้ตัดสินกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์คที่คะแนนจุดตัดระดับอ่อน (D+) ครั้งที่ 1 โดยที่เอกภพของการสังเกตประกอบด้วย ข้อสอบ และผู้ตัดสิน ซึ่งมีขนาดไม่จำกัด โดยเป็นการประมาณค่าความแปรปรวนของแหล่งต่างๆ เมื่อเทียบกับความแปรปรวนรวม มีค่าดังนี้ ความแปรปรวนของข้อสอบ (σ_i^2) เป็นความแตกต่างของความยากของข้อสอบมีค่า 0.1249 คิดเป็นร้อยละ 98 ของความแปรปรวนรวม ความแปรปรวนของผู้ตัดสิน (σ_r^2) เป็นการประมาณค่าความแปรปรวนของคะแนนเอกภพที่แสดงถึงความแตกต่างในการตัดสินคะแนนจุดตัดของผู้ตัดสิน มีค่า 0.0000 คิดเป็นร้อยละ 0 ของความแปรปรวนรวม และความแปรปรวนของส่วนที่เหลือ (σ_{ir}^2) เป็นความแปรปรวนของผลร่วมระหว่างข้อสอบ

กับผู้ตัดสินแสดงให้เห็นถึง ความไม่คงเส้นคงวาของผู้ตัดสินแต่ละคนในการตัดสินข้อสอบแต่ละข้อมีค่า 0.0020 คิดเป็นร้อยละ 2 ของความแปรปรวนรวม

เมื่อผู้ตัดสินกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์คที่คะแนนจุดตัดระดับอ่อน (D+) ครั้งที่ 2 มีค่าดังนี้ ความแปรปรวนของข้อสอบ (σ_i^2) เป็นความแตกต่างของความยากของข้อสอบมีค่า 0.1250 คิดเป็นร้อยละ 98 ของความแปรปรวนรวม ความแปรปรวนของผู้ตัดสิน (σ_r^2) เป็นการประมาณค่าความแปรปรวนของคะแนนเอกภาพที่แสดงถึงความแตกต่างในการตัดสินคะแนนจุดตัดของผู้ตัดสิน มีค่า 0.0000 คิดเป็นร้อยละ 0 ของความแปรปรวนรวม และความแปรปรวนของส่วนที่เหลือ (σ_{ir}^2) เป็นความแปรปรวนของผลร่วมระหว่างข้อสอบกับผู้ตัดสิน แสดงให้เห็นถึง ความไม่คงเส้นคงวาของผู้ตัดสินแต่ละคนในการตัดสินข้อสอบแต่ละข้อมีค่า 0.0020 คิดเป็นร้อยละ 2 ของความแปรปรวนรวม

เมื่อผู้ตัดสินกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์คที่คะแนนจุดตัดระดับอ่อน (D+) ครั้งที่ 3 มีค่าดังนี้ ความแปรปรวนของข้อสอบ (σ_i^2) เป็นความแตกต่างของความยากของข้อสอบมีค่า .1247 คิดเป็นร้อยละ 98 ของความแปรปรวนรวม ความแปรปรวนของผู้ตัดสิน (σ_r^2) เป็นการประมาณค่าความแปรปรวนของคะแนนเอกภาพที่แสดงถึงความแตกต่างในการตัดสินคะแนนจุดตัดของผู้ตัดสิน มีค่า .0000 คิดเป็นร้อยละ 0 ของความแปรปรวนรวม และความแปรปรวนของส่วนที่เหลือ (σ_{ir}^2) เป็นความแปรปรวนของผลร่วมระหว่างข้อสอบกับผู้ตัดสินแสดงให้เห็นถึง ความไม่คงเส้นคงวาของผู้ตัดสินแต่ละคนในการตัดสินข้อสอบแต่ละข้อมีค่า .0029 คิดเป็นร้อยละ 2 ของความแปรปรวนรวม

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 53 การประมาณค่าความแปรปรวนในขั้นการสรุปอ้างอิง (G-study) ขั้นการตัดสินใจ (D-study) และสัมประสิทธิ์การอ้างอิงสรุปของการกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์คที่คะแนนจุดตัดระดับอ่อน (D+) จำแนกตามจำนวนผู้ตัดสิน

Source of Variation	Estimated G Study Variance Components	Alternative Estimated D Studies Design Variance Component							
		8	12	14	16	18	20	22	24
Item (i)									
ครั้งที่ 1	0.1250	0.1250	0.1250	0.1250	0.1250	0.1250	0.1250	0.1250	0.1250
ครั้งที่ 2	0.1250	0.1250	0.1250	0.1250	0.1250	0.1250	0.1250	0.1250	0.1250
ครั้งที่ 3	0.1247	0.1247	0.1247	0.1247	0.1247	0.1247	0.1247	0.1247	0.1247
Rater (r)									
ครั้งที่ 1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
ครั้งที่ 2	0.7806	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
ครั้งที่ 3	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Residual (ir,e)									
ครั้งที่ 1	0.0020	0.0003	0.0002	0.0002	0.0001	0.0001	0.0001	0.0001	0.0001
ครั้งที่ 2	0.0020	0.0003	0.0002	0.0002	0.0001	0.0001	0.0001	0.0001	0.0001
ครั้งที่ 3	0.0029	0.0004	0.0002	0.0002	0.0002	0.0002	0.0001	0.0001	0.0001
σ^2_{Rel} ครั้งที่ 1		0.0003	0.0002	0.0002	0.0001	0.0001	0.0001	0.0001	0.0001
ครั้งที่ 2		0.0003	0.0002	0.0002	0.0001	0.0001	0.0001	0.0001	0.0001
ครั้งที่ 3		0.0004	0.0002	0.0002	0.0002	0.0002	0.0001	0.0001	0.0001
σ^2_{Abs} ครั้งที่ 1		0.0003	0.0002	0.0002	0.0001	0.0001	0.0001	0.0001	0.0001
ครั้งที่ 2		0.0003	0.0002	0.0002	0.0001	0.0001	0.0001	0.0001	0.0001
ครั้งที่ 3		0.0004	0.0002	0.0002	0.0002	0.0002	0.0001	0.0001	0.0001
$E\rho^2$ ครั้งที่ 1		0.9980	0.9986	0.9988	0.9990	0.9991	0.9992	0.9993	0.9993
ครั้งที่ 2		0.9980	0.9986	0.9988	0.9990	0.9991	0.9992	0.9993	0.9993
ครั้งที่ 3		0.9971	0.9981	0.9984	0.9986	0.9987	0.9989	0.9990	0.9990
ϕ ครั้งที่ 1		0.9980	0.9986	0.9988	0.9990	0.9991	0.9992	0.9993	0.9993
ครั้งที่ 2		0.9980	0.9986	0.9988	0.9990	0.9991	0.9992	0.9993	0.9993
ครั้งที่ 3		0.9971	0.9981	0.9984	0.9986	0.9987	0.9989	0.9990	0.9990

จากตารางที่ 53 ผลการศึกษา D ประกอบด้วยการประมาณค่าความแปรปรวนของผู้ตัดสิน จำนวน 8, 12, 14, 16, 18, 20, 22, 24 คน จากการตัดสินคะแนนจุดตัดของผู้ตัดสินจำนวน 12 คน ด้วยวิธีบุ๊คมาร์ค ที่ระดับอ่อน (D+) ครั้งที่ 1 ความแปรปรวนของคะแนนความคลาดเคลื่อนสัมพันธ์($\sigma^2_{\mathrm{Re}l}$) ซึ่งเป็นความคลาดเคลื่อนที่เกิดจากผลร่วมของข้อสอบ และผู้ตัดสิน มีค่าลดลง เมื่อจำนวนผู้ตัดสินเพิ่มขึ้น มีค่า ดังนี้ 0.0003 0.0002 0.0002 0.0001 0.0001 0.0001 0.0001 0.0001 ครั้งที่ 2 มีค่า ดังนี้ 0.0003 0.0002 0.0002 0.0001 0.0001 0.0001 0.0001 0.0001 ครั้งที่ 3 มีค่าดังนี้ 0.0004 0.0002 0.0002 0.0002 0.0002 0.0001 0.0001 0.0001 เมื่อมีผู้ตัดสินจำนวน 8, 12, 14, 16, 18, 20, 22, 24 คน ตามลำดับ

ส่วนความแปรปรวนของคะแนนความคลาดเคลื่อนสัมบูรณ์(σ^2_{Abs})ซึ่งเป็นผลรวมของความคลาดเคลื่อนที่เกิดจากผลร่วมของข้อสอบและผู้ตัดสิน มีค่าลดลงเมื่อจำนวนผู้ตัดสินเพิ่มขึ้น ครั้งที่ 1 มีค่าดังนี้ 0.0003 0.0002 0.0002 0.0001 0.0001 0.0001 0.0001 0.0001 ครั้งที่ 2 มีค่าดังนี้ 0.0003 0.0002 0.0002 0.0001 0.0001 0.0001 0.0001 0.0001 ครั้งที่ 3 มีค่าดังนี้ 0.0004 0.0002 0.0002 0.0002 0.0002 0.0001 0.0001 0.0001 เมื่อมีผู้ตัดสินจำนวน 8, 12, 14, 16, 18, 20, 22, 24 คน ตามลำดับ

ค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินแบบสัมพัทธ์ ($\mathrm{E}\rho^2$) มีค่าเพิ่มขึ้นเมื่อจำนวนผู้ตัดสินเพิ่มขึ้น ครั้งที่ 1 มีค่าดังนี้ 0.9980 0.9986 0.9988 0.9990 0.9991 0.9992 0.9993 0.9993 ครั้งที่ 2 มีค่าดังนี้ 0.9980 0.9986 0.9988 0.9990 0.9991 0.9992 0.9993 0.9993 ครั้งที่ 3 มีค่าดังนี้ 0.9971 0.9981 0.9984 0.9986 0.9987 0.9989 0.9990 0.9990 เมื่อมีผู้ตัดสินจำนวน 8, 12, 14, 16, 18, 20, 22, 24 คน ตามลำดับ

ส่วนค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินแบบสัมบูรณ์ (ϕ) มีค่าเพิ่มขึ้นเมื่อจำนวนผู้ตัดสินเพิ่มขึ้น ครั้งที่ 1 มีค่าดังนี้ 0.9980 0.9986 0.9988 0.9990 0.9991 0.9992 0.9993 0.9993 ครั้งที่ มีค่าดังนี้ มีค่าดังนี้ 0.9980 0.9986 0.9988 0.9990 0.9991 0.9992 0.9993 0.9993 ครั้งที่ 3 มีค่าดังนี้ 0.9971 0.9981 0.9984 0.9986 0.9987 0.9989 0.9990 0.9990 เมื่อมีผู้ตัดสินจำนวน 8, 12, 14, 16, 18, 20, 22, 24 คน ตามลำดับ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 54 การวิเคราะห์ความแปรปรวนและการประมาณค่าความแปรปรวนของการกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์คที่คะแนนจุดตัดระดับอ่อนมาก (D)ในขั้นสรุปอ้างอิง(G-study)

Source of Variation	df	Sum of Squares	Mean Squares	Estimated Variance Component	Percentage of Total Variance
ครั้งที่ 1					
Item (i)	99	83.4700	0.8431	0.0700	96.0220
Rater (r)	11	0.0367	0.0033	0.0000	0.0000
Residual (ir,e)	1089	3.1300	0.0021	0.0029	3.9781
Total	1199	86.6367		0.0729	100
ครั้งที่ 2					
Item (i)	99	83.4700	0.8431	0.0700	96.0220
Rater (r)	11	0.0367	0.0033	0.0000	0.0000
Residual (ir,e)	1089	3.1300	0.0029	0.0029	3.9781
Total	1199	86.6367		0.0729	100
ครั้งที่ 3					
Item (i)	99	83.5625	0.8441	0.0700	95.1087
Rater (r)	11	0.0492	0.0045	0.0000	0.0000
Residual (ir,e)	1089	3.8675	0.0035	0.0036	4.8913
Total	1199	87.4792		0.0736	100

จากตารางที่ 54 เมื่อผู้ตัดสินกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์คที่คะแนนจุดตัดระดับอ่อนมาก (D) ครั้งที่ 1 โดยที่เอกภพของการสังเกตประกอบด้วย ข้อสอบ และผู้ตัดสิน ซึ่งมีขนาดไม่จำกัด โดยเป็นการประมาณค่าความแปรปรวนของแหล่งต่างๆ เมื่อเทียบกับความแปรปรวนรวมมีค่าดังนี้ ความแปรปรวนของข้อสอบ (σ_i^2) เป็นความแตกต่างของความยากของข้อสอบมีค่า 0.0700 คิดเป็นร้อยละ 96 ของความแปรปรวนรวม ความแปรปรวนของผู้ตัดสิน (σ_r^2) เป็นการประมาณค่าความแปรปรวนของคะแนนเอกภพที่แสดงถึงความแตกต่างในการตัดสินคะแนนจุดตัดของผู้ตัดสิน มีค่า 0.0000 คิดเป็นร้อยละ 0 ของความแปรปรวนรวม และความแปรปรวนของส่วนที่เหลือ (σ_{ir}^2) เป็นความแปรปรวนของผลร่วมระหว่างข้อสอบกับผู้

ตัดสิน แสดงให้เห็นถึง ความไม่คงเส้นคงวาของผู้ตัดสินแต่ละคนในการตัดสินข้อสอบแต่ละข้อมีค่า 0.0029 คิดเป็นร้อยละ 4 ของความแปรปรวนรวม

เมื่อผู้ตัดสินกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์คที่คะแนนจุดตัดระดับอ่อนมาก (D) ครั้งที่ 2 มีค่าดังนี้ ความแปรปรวนของข้อสอบ (σ_i^2) เป็นความแตกต่างของความยากของข้อสอบมีค่า 0.0700 คิดเป็นร้อยละ 96 ของความแปรปรวนรวม ความแปรปรวนของผู้ตัดสิน (σ_r^2) เป็นการประมาณค่าความแปรปรวนของคะแนนเอกภพที่แสดงถึงความแตกต่างในการตัดสินคะแนนจุดตัดของผู้ตัดสิน มีค่า 0.0000 คิดเป็นร้อยละ 0 ของความแปรปรวนรวม และความแปรปรวนของส่วนที่เหลือ (σ_{ir}^2) เป็นความแปรปรวนของผลร่วมระหว่างข้อสอบกับผู้ตัดสิน แสดงให้เห็นถึง ความไม่คงเส้นคงวาของผู้ตัดสินแต่ละคนในการตัดสินข้อสอบแต่ละข้อ มีค่า 0.0029 คิดเป็นร้อยละ 4 ของความแปรปรวนรวม

เมื่อผู้ตัดสินกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์คที่คะแนนจุดตัดระดับอ่อนมาก (D) ครั้งที่ 3 มีค่าดังนี้ ความแปรปรวนของข้อสอบ (σ_i^2) เป็นความแตกต่างของความยากของข้อสอบมีค่า 0.0700 คิดเป็นร้อยละ 95 ของความแปรปรวนรวม ความแปรปรวนของผู้ตัดสิน (σ_r^2) เป็นการประมาณค่าความแปรปรวนของคะแนนเอกภพที่แสดงถึงความแตกต่างในการตัดคะแนนจุดตัดของผู้ตัดสินมีค่า 0.0000 คิดเป็นร้อยละ 0 ของความแปรปรวนรวม และความแปรปรวนของส่วนที่เหลือ (σ_{ir}^2) เป็นความแปรปรวนของผลร่วมระหว่างข้อสอบกับผู้ตัดสิน แสดงให้เห็นถึง ความไม่คงเส้นคงวาของผู้ตัดสินแต่ละคนในการตัดสินข้อสอบแต่ละข้อมีค่า 0.0036 คิดเป็นร้อยละ 4.8913 ของความแปรปรวนรวม

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 55 การประมาณค่าความแปรปรวนในขั้นการสรุปอ้างอิง (G-study) ขั้นการตัดสินใจ (D-study) และสัมประสิทธิ์การอ้างอิงสรุปของการกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์คที่คะแนนจุดตัดระดับอ่อนมาก (D) จำแนกตามจำนวนผู้ตัดสิน

Source of Variation		Estimated G Study Variance Components	Alternative D Studies Design Variance Component Estimated							
			8	12	14	16	18	20	22	24
Item (i)										
ครั้งที่ 1		0.0700	0.0700	0.0700	0.0700	0.0700	0.0700	0.0700	0.0700	0.0700
ครั้งที่ 2		0.0700	0.0700	0.0700	0.0700	0.0700	0.0700	0.0700	0.0700	0.0700
ครั้งที่ 3		0.0700	0.0700	0.0700	0.0700	0.0700	0.0700	0.0700	0.0700	0.0700
Rater (r)										
ครั้งที่ 1		0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
ครั้งที่ 2		0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
ครั้งที่ 3		0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Residual (ir,e)										
ครั้งที่ 1		0.0029	0.0004	0.0002	0.0002	0.0002	0.0002	0.0001	0.0001	0.0001
ครั้งที่ 2		0.0029	0.0004	0.0002	0.0002	0.0002	0.0002	0.0001	0.0001	0.0001
ครั้งที่ 3		0.0036	0.0004	0.0003	0.0003	0.0002	0.0002	0.0002	0.0002	0.0002
σ^2_{Rel}	ครั้งที่ 1		0.0004	0.0002	0.0002	0.0002	0.0002	0.0001	0.0001	0.0001
	ครั้งที่ 2		0.0004	0.0002	0.0002	0.0002	0.0002	0.0001	0.0001	0.0001
	ครั้งที่ 3		0.0004	0.0003	0.0003	0.0002	0.0002	0.0002	0.0002	0.0002
σ^2_{Abs}	ครั้งที่ 1		0.0004	0.0002	0.0002	0.0002	0.0002	0.0001	0.0001	0.0001
	ครั้งที่ 2		0.0004	0.0002	0.0002	0.0002	0.0002	0.0001	0.0001	0.0001
	ครั้งที่ 3		0.0004	0.0003	0.0003	0.0002	0.0002	0.0002	0.0002	0.0002
$E\rho^2$	ครั้งที่ 1		0.9949	0.9966	0.9971	0.9974	0.9977	0.9980	0.9981	0.9983
	ครั้งที่ 2		0.9949	0.9966	0.9971	0.9974	0.9977	0.9980	0.9981	0.9983
	ครั้งที่ 3		0.9937	0.9958	0.9964	0.9968	0.9972	0.9975	0.9977	0.9979
ϕ	ครั้งที่ 1		0.9950	0.9966	0.9971	0.9974	0.9977	0.9980	0.9981	0.9983
	ครั้งที่ 2		0.9949	0.9966	0.9971	0.9974	0.9977	0.9980	0.9981	0.9983
	ครั้งที่ 3		0.9937	0.9958	0.9964	0.9968	0.9972	0.9975	0.9977	0.9979

จากตารางที่ 55 ผลการศึกษา D ประกอบด้วยการประมาณค่าความแปรปรวนของผู้ตัดสิน จำนวน 8, 12, 14, 16, 18, 20, 22, 24 คน จากการตัดสินคะแนนจุดตัดของผู้ตัดสินจำนวน 12 คน ด้วยวิธีบุ๊คมาร์คที่ระดับอ่อนมาก (D) ครั้งที่ 1 ความแปรปรวนของคะแนนความคลาดเคลื่อนสัมพันธ์($\sigma^2_{\mathrm{Re}l}$) ซึ่งเป็นความคลาดเคลื่อนที่เกิดจากผลร่วมของข้อสอบ และผู้ตัดสิน มีค่าลดลง เมื่อจำนวนผู้ตัดสินเพิ่มขึ้น มีค่า ดังนี้ 0.0004 0.0002 0.0002 0.0002 0.0002 0.0001 0.0001 0.0001 ครั้งที่ 2 มีค่าดังนี้ 0.0004 0.0002 0.0002 0.0002 0.0002 0.0001 0.0001 0.0001 ครั้งที่ 3 มีค่า ดังนี้ 0.0004 0.0003 0.0003 0.0002 0.0002 0.0002 0.0002 0.0002 เมื่อมีผู้ตัดสินจำนวน 8, 12, 14, 16, 18, 20, 22, 24 คน ตามลำดับ

ส่วนความแปรปรวนของคะแนนความคลาดเคลื่อนสัมบูรณ์(σ^2_{Abs})ซึ่งเป็นผลรวมของความคลาดเคลื่อนที่เกิดจากผลร่วมของข้อสอบและผู้ตัดสิน มีค่าลดลงเมื่อจำนวนผู้ตัดสินเพิ่มขึ้นมีค่าดังนี้ 0.0004 0.0002 0.0002 0.0002 0.0002 0.0001 0.0001 0.0001 ครั้งที่ 2 มีค่าดังนี้ 0.0004 0.0002 0.0002 0.0002 0.0002 0.0001 0.0001 0.0001 ครั้งที่ 3 มีค่าดังนี้ 0.0004 0.0003 0.0003 0.0002 0.0002 0.0002 0.0002 0.0002 เมื่อมีผู้ตัดสินจำนวน 8, 12, 14, 16, 18, 20, 22, 24 คน ตามลำดับ

ค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินแบบสัมพัทธ์ ($\mathrm{E}\rho^2$) มีค่าเพิ่มขึ้นเมื่อจำนวนผู้ตัดสินเพิ่มขึ้น มีค่าดังนี้ 0.9949 0.9966 0.9971 0.9974 0.9977 0.9980 0.9981 0.9983 ครั้งที่ 2 มีค่าดังนี้ 0.9949 0.9966 0.9971 0.9974 0.9977 0.9980 0.9981 0.9983 ครั้งที่ 3 มีค่าดังนี้ 0.9937 0.9958 0.9964 0.9968 0.9972 0.9975 0.9977 0.9979 เมื่อมีผู้ตัดสินจำนวน 8, 12, 14, 16, 18, 20, 22, 24 คน ตามลำดับ

ส่วนค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินแบบสัมบูรณ์ (ϕ) มีค่าเพิ่มขึ้นเมื่อจำนวนผู้ตัดสินเพิ่มขึ้น มีค่าดังนี้ 0.9950 0.9966 0.9971 0.9974 0.9977 0.9980 0.9981 0.9983 ครั้งที่ 2 มีค่าดังนี้ 0.9949 0.9966 0.9971 0.9974 0.9977 0.9980 0.9981 0.9983 ครั้งที่ 3 มีค่าดังนี้ 0.9937 0.9958 0.9964 0.9968 0.9972 0.9975 0.9977 0.9979 เมื่อมีผู้ตัดสินจำนวน 8, 12, 14, 16, 18, 20, 22, 24 คน ตามลำดับ

ผู้วิจัยได้สรุปค่าความเที่ยงหรือสัมประสิทธิ์การสรุปอ้างอิงที่กำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และวิธีบุ๊คมาร์ค จำแนกตามจำนวนผู้ตัดสินคือ 8, 12, 14, 16,18, 20, 22 และ 24 โดยผู้ตัดสินทำการกำหนดคะแนนจุดตัดทั้งหมด 7 ระดับคือ ระดับดีเยี่ยม (A) ระดับดีมาก (B+) ระดับดี (B) ระดับดีพอใช้ (C+) ระดับพอใช้ (C) ระดับอ่อน (D+) และระดับอ่อนมาก (D) โดยทำการกำหนดมาตรฐานทั้งสองวิธีรวม 3 ครั้ง แต่ละครั้งมีรูปแบบดังนี้ ครั้งที่ 1 กำหนดมาตรฐานด้วยวิธี แองกอฟที่ได้รับการปรับปรุงก่อน แล้วตามด้วยวิธีบุ๊คมาร์ค ครั้งที่ 2 กำหนดมาตรฐานด้วยวิธี

บุ๊คมาร์คก่อน แล้วตามด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และครั้งที่ 3 กำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุงก่อน แล้วตามด้วยวิธีบุ๊คมาร์ค ดังตารางที่ 56

ตารางที่ 56 ค่าความเที่ยงหรือสัมประสิทธิ์การสรุปอ้างอิงที่กำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และวิธีบุ๊คมาร์ค จำแนกตามจำนวนผู้ตัดสิน

จำนวนผู้ตัดสิน	ค่าความเที่ยง หรือสัมประสิทธิ์สรุปอ้างอิง					
	การตัดสินครั้งที่ 1		การตัดสินครั้งที่ 2		การตัดสินครั้งที่ 3	
	แองกอฟ	บุ๊คมาร์ค	แองกอฟ	บุ๊คมาร์ค	แองกอฟ	บุ๊คมาร์ค
คะแนนจุดตัดระดับดีเยี่ยม (A)						
8	0.5762	0.9954	0.7331	0.9960	0.7342	0.9958
12	0.6710	0.9969	0.8047	0.9973	0.8055	0.9972
14	0.7041	0.9974	0.8278	0.9977	0.8286	0.9976
16	0.7311	0.9977	0.8460	0.9980	0.8467	0.9979
18	0.7537	0.9980	0.8607	0.9982	0.8614	0.9981
20	0.7727	0.9982	0.8729	0.9984	0.8735	0.9983
22	0.7890	0.9983	0.8831	0.9985	0.8837	0.9985
24	0.8031	0.9985	0.8918	0.9986	0.8923	0.9986
คะแนนจุดตัดระดับดีมาก (B+)						
8	0.5852	0.9989	0.7575	0.9890	0.7529	0.9895
12	0.6791	0.9992	0.8241	0.9926	0.8205	0.9930
14	0.7117	0.9994	0.8454	0.9937	0.8421	0.9940
16	0.7383	0.9994	0.8620	0.9945	0.8590	0.9947
18	0.7604	0.9995	0.8755	0.9951	0.8727	0.9953
20	0.7791	0.9996	0.8865	0.9956	0.8840	0.9958
22	0.7950	0.9996	0.8957	0.996	0.8934	0.9961
24	0.8089	0.9996	0.9036	0.9963	0.9014	0.9965

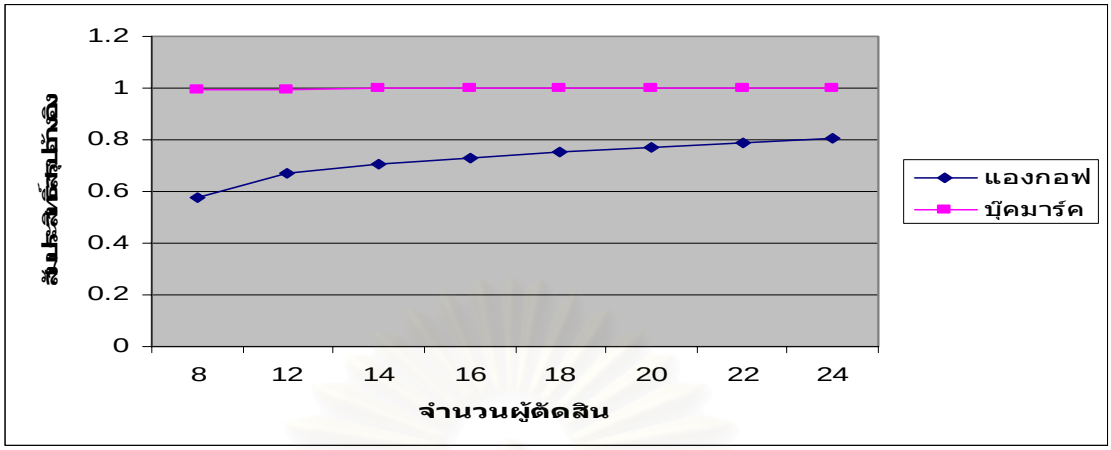
ตารางที่ 56 (ต่อ)

จำนวนผู้ตัดสิน	ค่าความเที่ยง หรือสัมประสิทธิ์สรุปอ้างอิง					
	การตัดสินครั้งที่ 1		การตัดสินครั้งที่ 2		การตัดสินครั้งที่ 3	
	แองกอฟ	บุ๊คมาร์ค	แองกอฟ	บุ๊คมาร์ค	แองกอฟ	บุ๊คมาร์ค
คะแนนจุดตัดระดับดี (B)						
8	0.6537	0.9980	0.5564	0.9972	0.7667	0.9963
12	0.7390	0.9987	0.6529	0.9981	0.8314	0.9975
14	0.7676	0.9987	0.6870	0.9984	0.8519	0.9979
16	0.7906	0.9990	0.7150	0.9986	0.8680	0.9981
18	0.8094	0.9991	0.7383	0.9987	0.8809	0.9984
20	0.8251	0.9992	0.7582	0.9989	0.8915	0.9985
22	0.8385	0.9993	0.7752	0.9990	0.9004	0.9987
24	0.8499	0.9993	0.7900	0.9991	0.9079	0.9988
คะแนนจุดตัดระดับดีพอใช้ (C+)						
8	0.7642	0.9931	0.7903	0.9970	0.7899	0.9973
12	0.8294	0.9954	0.8497	0.9980	0.8494	0.9982
14	0.8501	0.9961	0.8683	0.9983	0.8681	0.9984
16	0.8664	0.9966	0.8829	0.9985	0.8827	0.9986
18	0.8794	0.9969	0.8945	0.9987	0.8943	0.9988
20	0.8902	0.9972	0.9040	0.9988	0.9039	0.9989
22	0.8991	0.9975	0.9120	0.9989	0.9118	0.9990
24	0.9068	0.9977	0.9187	0.9990	0.9186	0.9991
คะแนนจุดตัดระดับพอใช้ (C)						
8	0.7739	1.0000	0.7776	0.9962	0.7618	0.9994
12	0.8370	1.0000	0.8399	0.9975	0.8275	0.9996
14	0.8569	1.0000	0.8595	0.9978	0.8484	0.9997
16	0.8725	1.0000	0.8749	0.9981	0.8648	0.9997
18	0.8851	1.0000	0.8872	0.9983	0.8780	0.9998
20	0.8953	1.0000	0.8974	0.9985	0.8889	0.9998
22	0.9039	1.0000	0.9058	0.9986	0.8979	0.9998
24	0.9112	1.0000	0.9130	0.9987	0.9056	0.9998

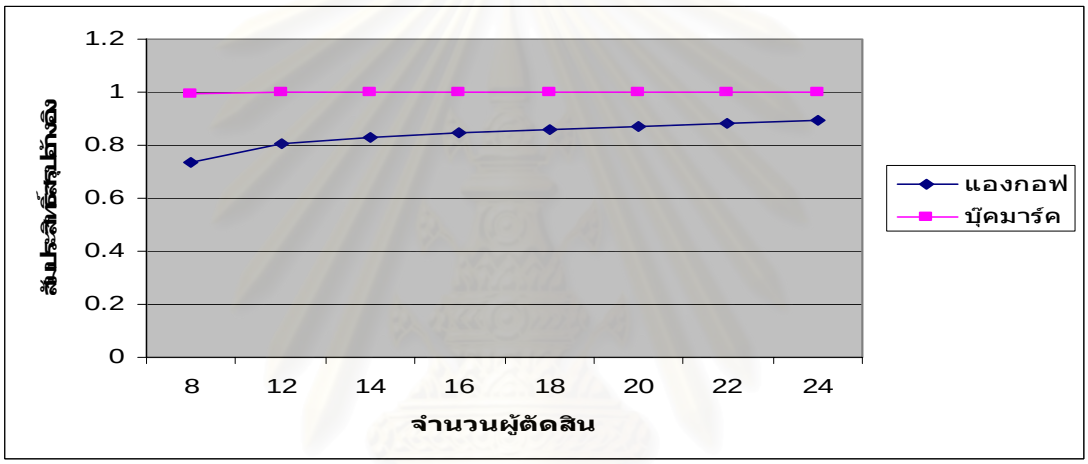
ตารางที่ 56 (ต่อ)

จำนวนผู้ตัดสิน	ค่าความเที่ยง หรือสัมประสิทธิ์สรุปอ้างอิง					
	การตัดสินครั้งที่ 1		การตัดสินครั้งที่ 2		การตัดสินครั้งที่ 3	
	แองกอฟ	บุ๊คมาร์ค	แองกอฟ	บุ๊คมาร์ค	แองกอฟ	บุ๊คมาร์ค
คะแนนจุดตัดระดับอ่อน (D+)						
8	0.8212	0.9980	0.8909	0.9980	0.8741	0.9971
12	0.8732	0.9986	0.9245	0.9986	0.9124	0.9981
14	0.8893	0.9988	0.9346	0.9988	0.9240	0.9984
16	0.9018	0.9990	0.9423	0.9990	0.9328	0.9986
18	0.9118	0.9991	0.9484	0.9991	0.9398	0.9987
20	0.9199	0.9992	0.9533	0.9992	0.9455	0.9989
22	0.9266	0.9993	0.9574	0.9993	0.9502	0.9990
24	0.9323	0.9993	0.9608	0.9993	0.9542	0.9990
คะแนนจุดตัดระดับอ่อนมาก (D)						
8	0.9198	0.9950	0.8472	0.9949	0.8449	0.9937
12	0.9451	0.9966	0.8927	0.9966	0.8909	0.9958
14	0.9525	0.9971	0.9066	0.9971	0.9050	0.9964
16	0.9582	0.9974	0.9173	0.9974	0.9159	0.9968
18	0.9627	0.9977	0.9258	0.9977	0.9246	0.9972
20	0.9663	0.9980	0.9327	0.9980	0.9316	0.9975
22	0.9693	0.9981	0.9385	0.9981	0.9374	0.9977
24	0.9718	0.9983	0.9435	0.9983	0.9423	0.9979

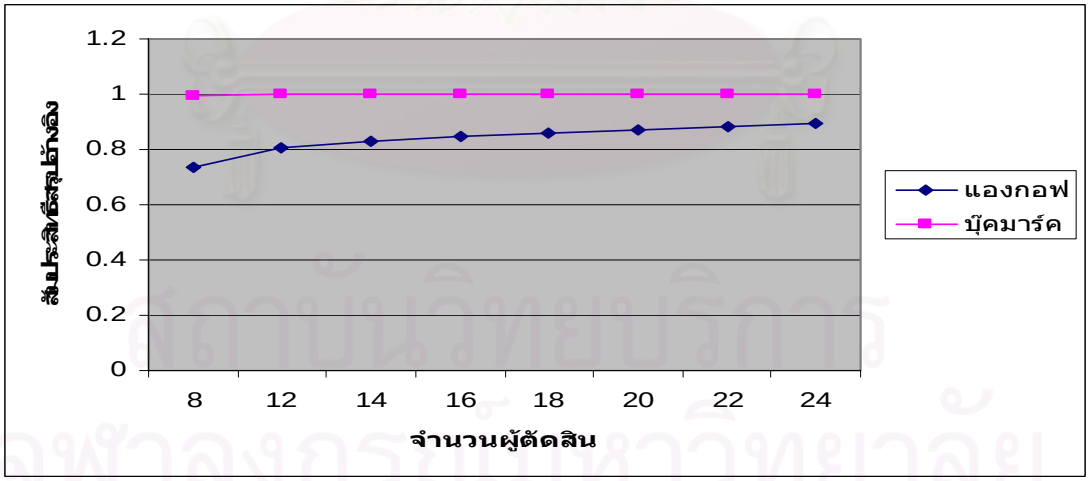
เพื่อแสดงให้เห็นการเปรียบเทียบระหว่างค่าความเที่ยง หรือสัมประสิทธิ์สรุปอ้างอิงของการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์ค จำแนกตามจำนวนผู้ตัดสิน ผู้วิจัยจึงนำเสนอการเปรียบเทียบเป็นกราฟเส้น ดังแสดงในแผนภาพ



การตัดสินครั้งที่ 1

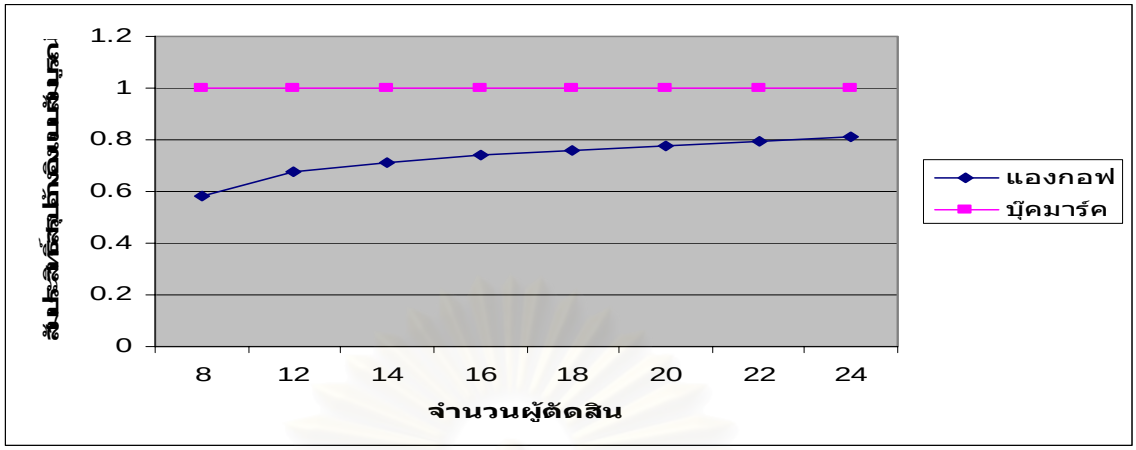


การตัดสินครั้งที่ 2

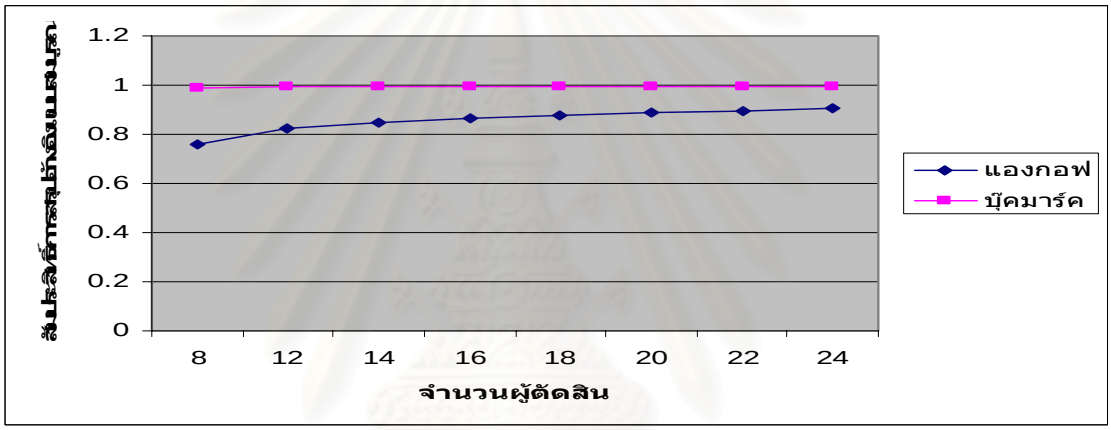


การตัดสินครั้งที่ 3

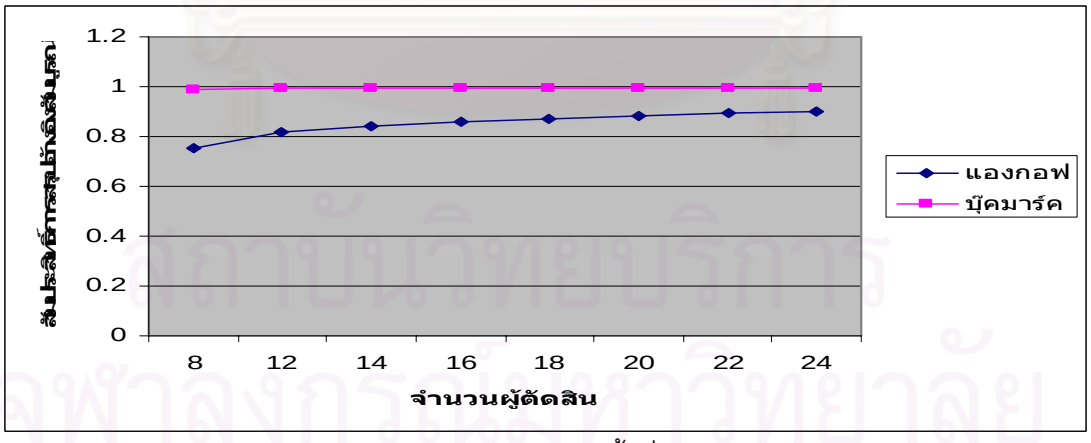
แผนภาพที่ 20 การเปรียบเทียบค่าความเที่ยงของการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์ค จำแนกตามจำนวนผู้ตัดสินที่คะแนนจุดตัดระดับดีเยี่ยม (A)



การตัดสินครั้งที่ 1

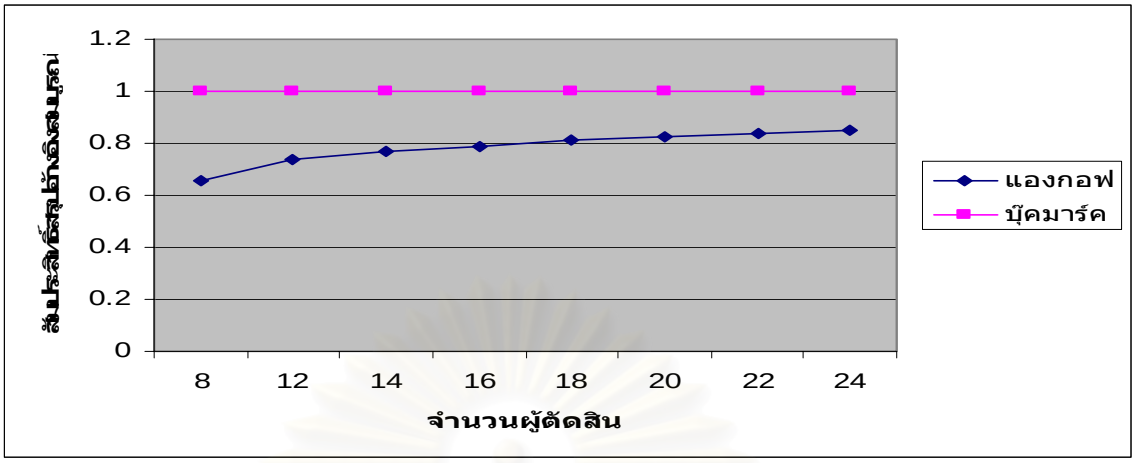


การตัดสินครั้งที่ 2

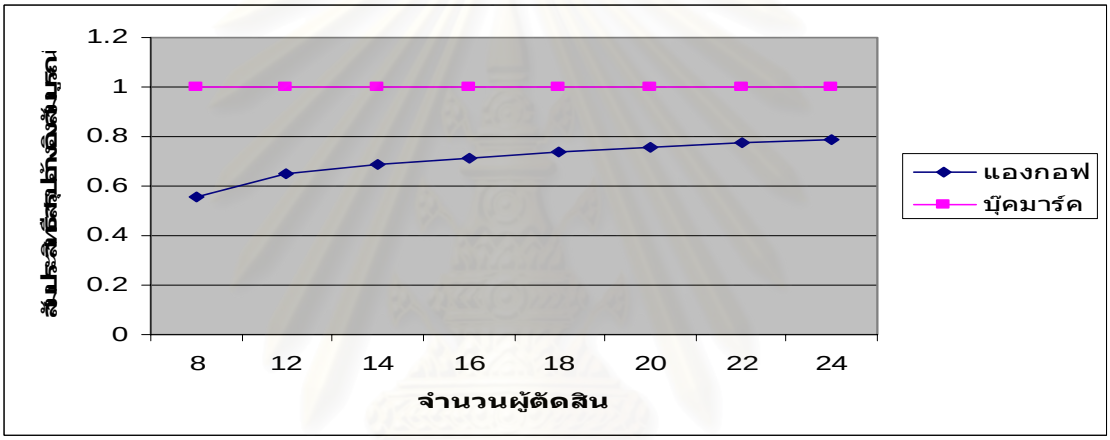


การตัดสินครั้งที่ 3

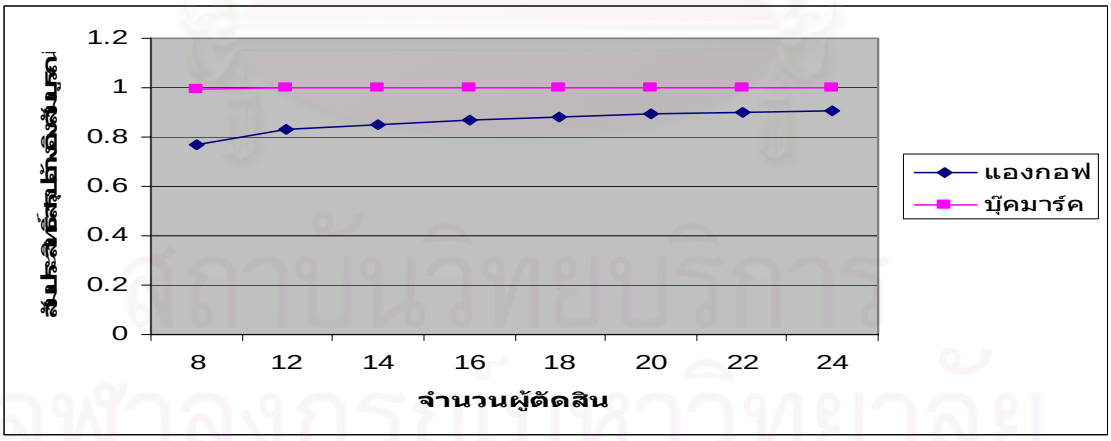
แผนภาพที่ 21 การเปรียบเทียบค่าความเที่ยงของการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์ค จำแนกตามจำนวนผู้ตัดสินที่คะแนนจุดตัดระดับดีมาก (B+)



การตัดสินครั้งที่ 1

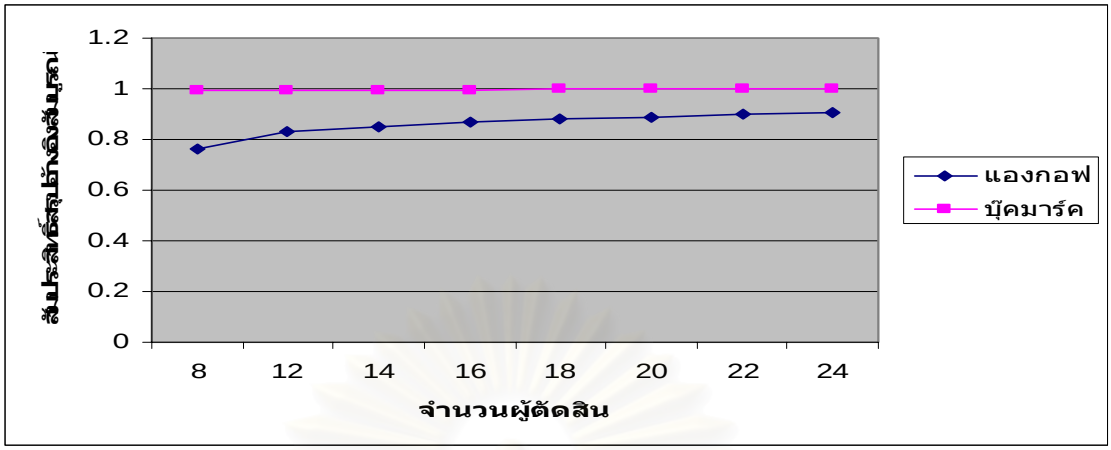


การตัดสินครั้งที่ 2

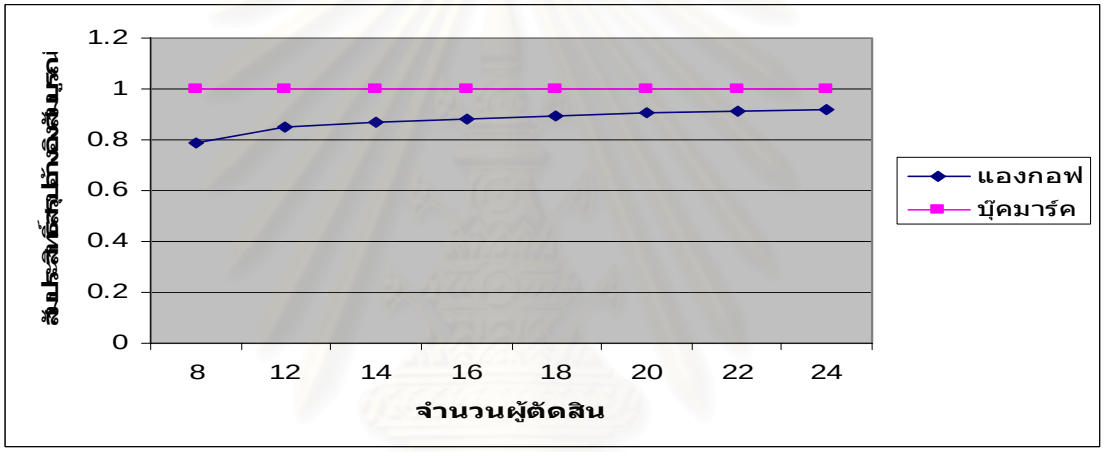


การตัดสินครั้งที่ 3

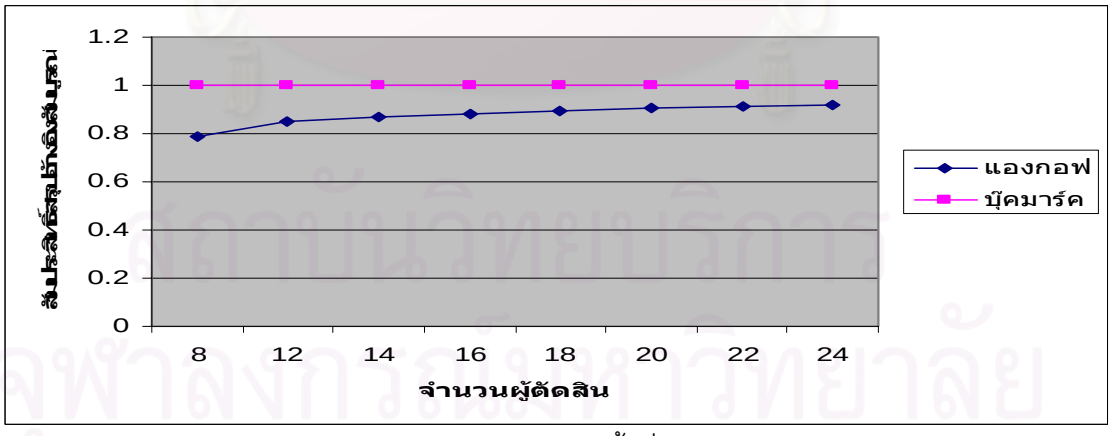
แผนภาพที่ 22 การเปรียบเทียบค่าความเที่ยงของการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์ค จำแนกตามจำนวนผู้ตัดสินที่คะแนนจุดตัดระดับดี (B)



การตัดสินครั้งที่ 1

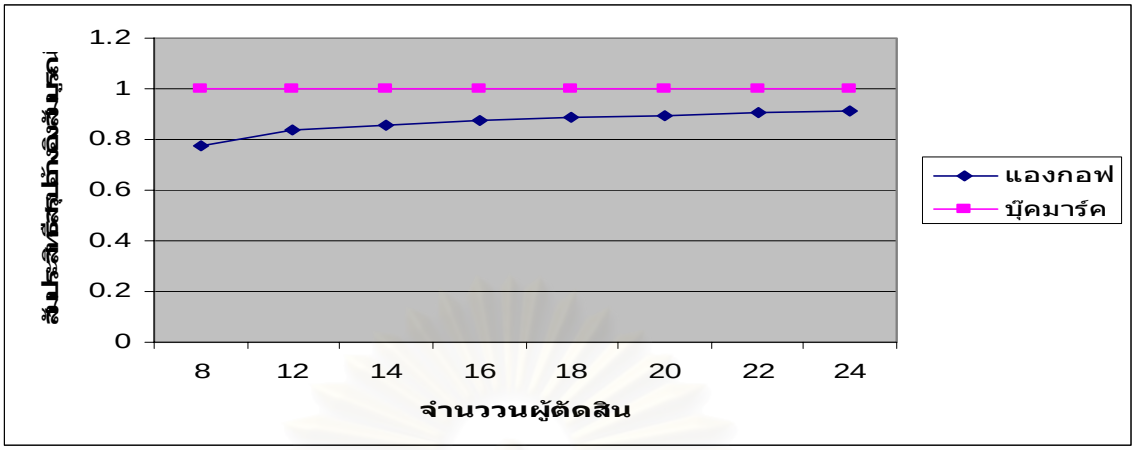


การตัดสินครั้งที่ 2

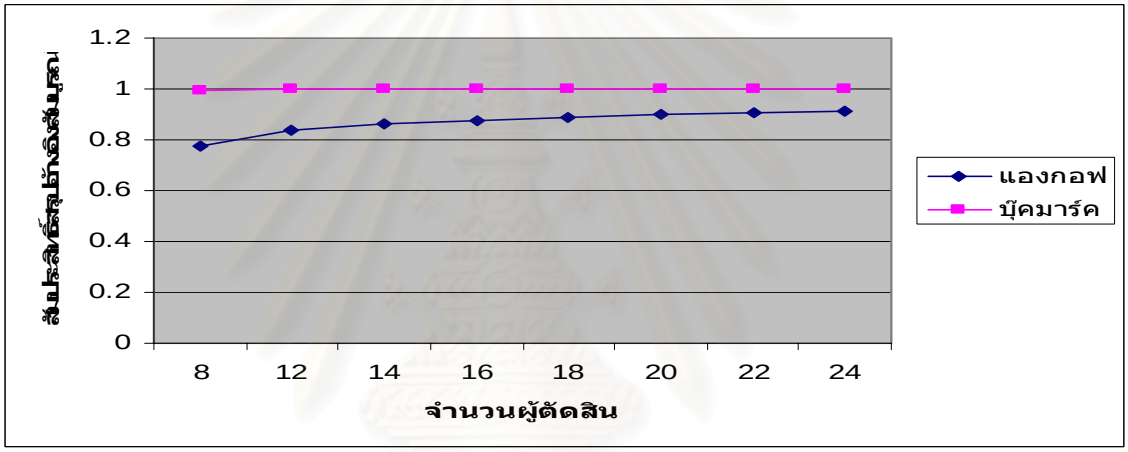


การตัดสินครั้งที่ 3

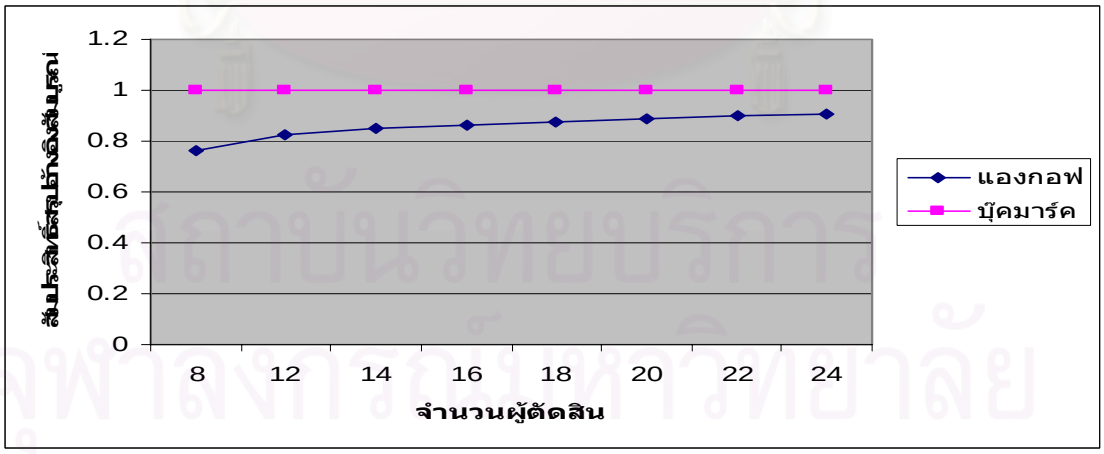
แผนภาพที่ 23 การเปรียบเทียบค่าความเที่ยงของการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์ค จำแนกตามจำนวนผู้ตัดสินที่คะแนนจุดตัดระดับดีพอใช้ (C+)



การตัดสินครั้งที่ 1

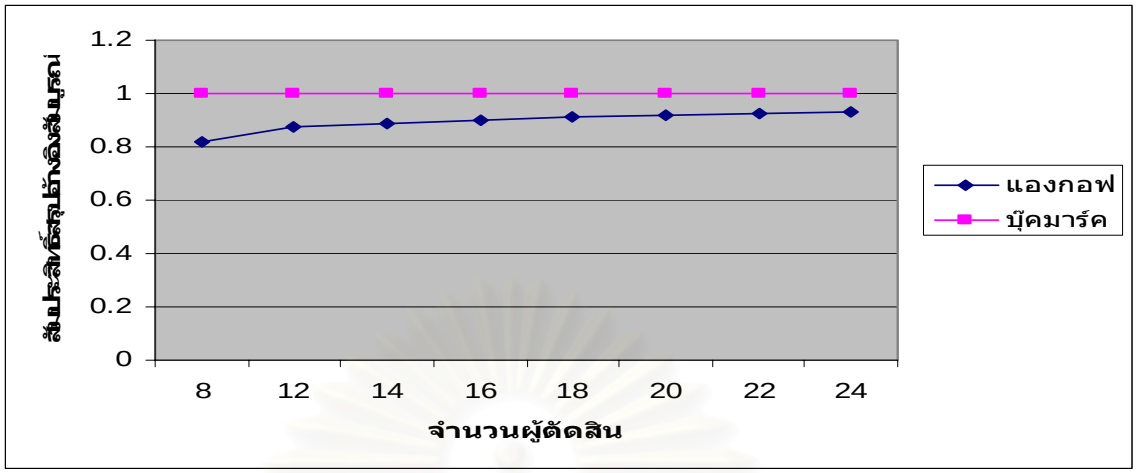


การตัดสินครั้งที่ 2

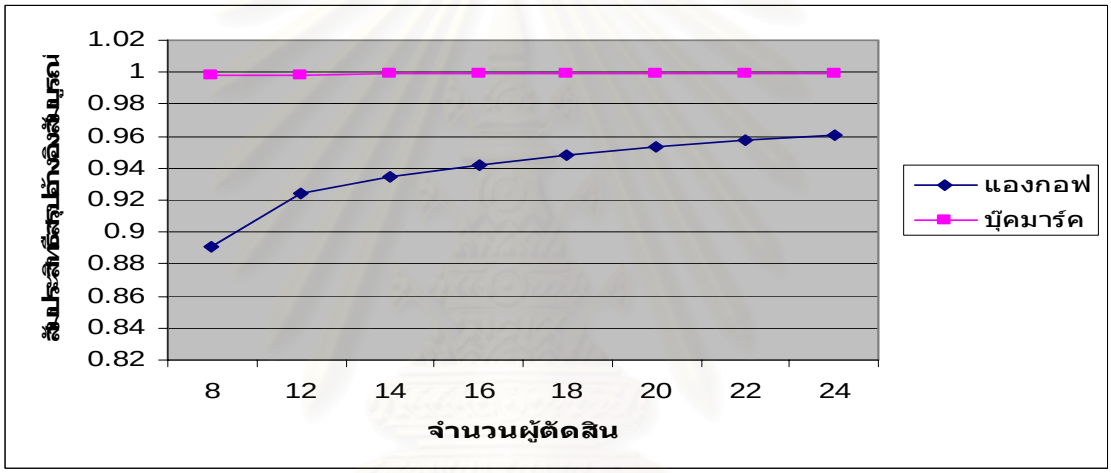


การตัดสินครั้งที่ 3

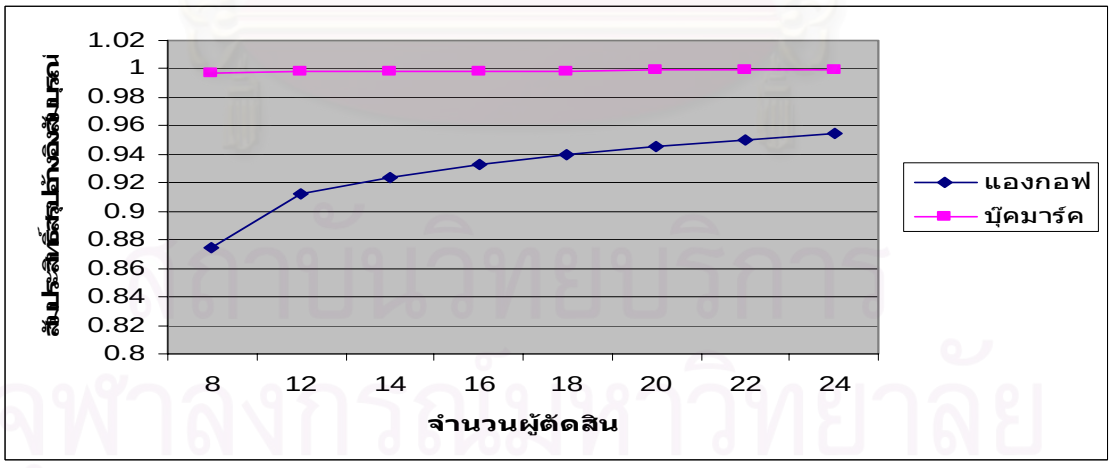
แผนภาพที่ 24 การเปรียบเทียบค่าความเที่ยงของการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์ค จำแนกตามจำนวนผู้ตัดสินที่คะแนนจุดตัดระดับพอใช้ (C)



การตัดสินครั้งที่ 1

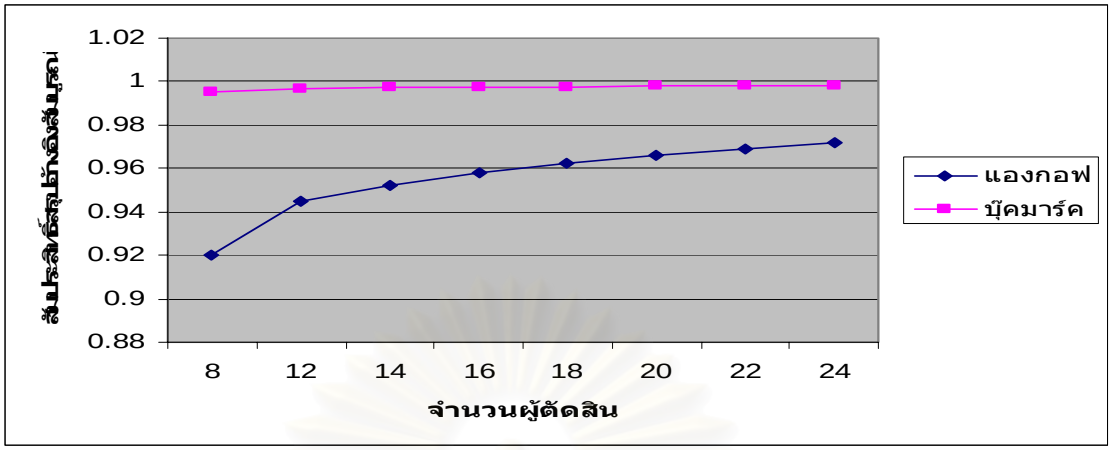


การตัดสินครั้งที่ 2

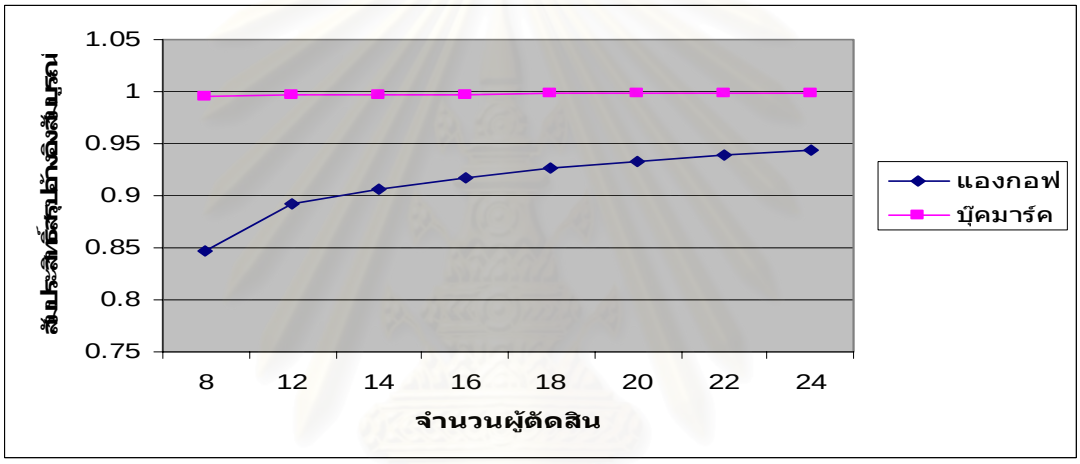


การตัดสินครั้งที่ 3

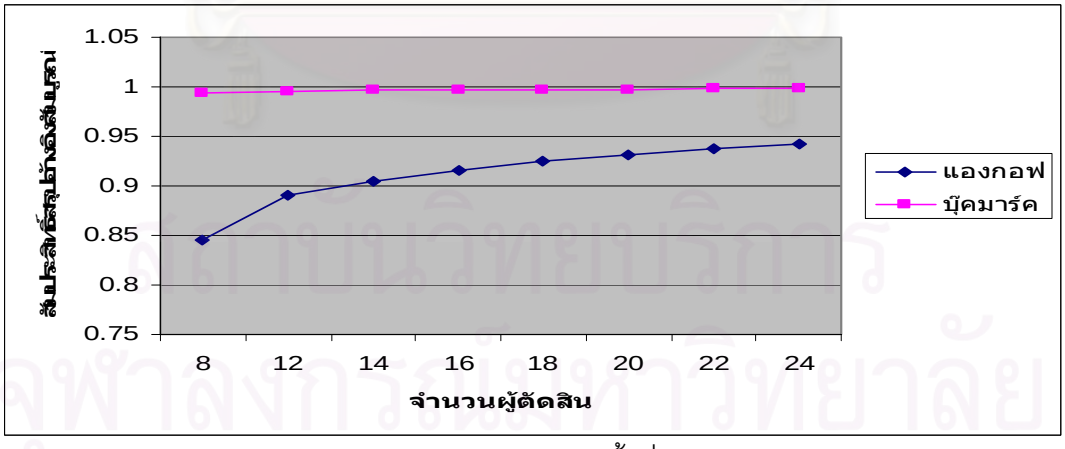
แผนภาพที่ 25 การเปรียบเทียบค่าความเที่ยงของการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์ค จำแนกตามจำนวนผู้ตัดสินที่คะแนนจุดตัดระดับอ่อน (D+)



การตัดสินครั้งที่ 1



การตัดสินครั้งที่ 2



การตัดสินครั้งที่ 3

แผนภาพที่ 26 การเปรียบเทียบค่าความเที่ยงของการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์ค จำแนกตามจำนวนผู้ตัดสินที่คะแนนจุดตัดระดับอ่อนมาก (D)

จากตารางที่ 56 และแผนภาพที่ 20 – 26 พบว่า เมื่อกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์ค ช่วงการตัดสินใจที่จำนวนผู้ตัดสิน 8 ถึง 24 คน ให้ค่าความเที่ยง หรือสัมประสิทธิ์สรุปอ้างอิงสูงกว่า การกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุงทุกช่วงค่า และคะแนนจุดตัดทุกระดับ โดยการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุงมีค่าความเที่ยงประมาณ 0.6 ถึง 0.9 ในขณะที่วิธีบุ๊คมาร์คมีค่าความเที่ยงสูงถึง 0.9 ในคะแนนจุดตัดทุกระดับ

และเมื่อพิจารณากราฟเส้นจะเห็นว่า เมื่อกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุงความชันเริ่มเปลี่ยนแปลงน้อยลง เมื่อจำนวนผู้ตัดสินตั้งแต่ 12 คนขึ้นไป ส่วนการกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์คความชันของกราฟแทบจะเป็นเส้นตรง แสดงว่าจำนวนผู้ตัดสินไม่มีผลต่อค่าความเที่ยง

ตอนที่ 3 ผลการตรวจสอบคุณภาพของการกำหนดมาตรฐานด้านความตรงด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และวิธีบุ๊คมาร์ค

การกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุงและวิธีบุ๊คมาร์ค ผู้ตัดสินกำหนดคะแนนจุดตัด 7 ระดับคือ ระดับดีเยี่ยม (A) ระดับดีมาก (B+) ระดับดี (B) ระดับดีพอใช้ (C+) ระดับพอใช้ (C) ระดับอ่อน (D+) และระดับอ่อนมาก (D) โดยทำการกำหนดมาตรฐานทั้งสองวิธีรวม 3 ครั้ง แต่ละครั้งมีรูปแบบดังนี้ **ครั้งที่** 1 กำหนดมาตรฐานด้วยวิธี แองกอฟที่ได้รับการปรับปรุงก่อน แล้วตามด้วยวิธีบุ๊คมาร์ค **ครั้งที่** 2 กำหนดมาตรฐานด้วยวิธีบุ๊คมาร์คก่อน แล้วตามด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และ **ครั้งที่** 3 กำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุงก่อน แล้วตามด้วยวิธีบุ๊คมาร์ค

สำหรับความตรงตามสภาพ ผู้วิจัยหาจากความสัมพันธ์ระหว่างระดับผลการเรียนที่นักเรียนได้รับจากโรงเรียน กับระดับผลการตัดสินที่กำหนดด้วยวิธีการแองกอฟที่ได้รับการปรับปรุง และวิธีบุ๊คมาร์ค โดยผู้วิจัยดำเนินการขอระดับผลการเรียนวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 โรงเรียนบ้านสวน (จั่นอนุสรณ์) ปีการศึกษา 2549 ซึ่งเป็นนักเรียนที่ได้รับการสอบด้วยแบบสอบวัดผลสัมฤทธิ์ทางการเรียนที่ผู้วิจัยพัฒนาขึ้น สำหรับเหตุผลที่ผู้วิจัยเลือกนักเรียนโรงเรียนดังกล่าว เนื่องจากโรงเรียนมีการจัดการเรียนการสอนที่เน้นการเรียนการสอนตามสภาพจริง โดยมีการประเมินนักเรียนด้วยวิธีการที่หลากหลายอย่างต่อเนื่อง จากเหตุผลดังกล่าวน่าจะได้ผลการเรียนที่ตรงกับความสามารถที่เป็นจริงของนักเรียน ข้อมูลที่ได้ผู้วิจัยนำมาคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์แบบ Spearman's Rho จากนักเรียนจำนวน 67 คน รายละเอียดดังตารางที่ 57

ตารางที่ 57 ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างระดับผลการเรียนที่นักเรียนได้รับจากโรงเรียน กับระดับมาตรฐานที่กำหนดด้วยวิธีการแองกอฟที่ได้รับการปรับปรุง และวิธีบุ๊คมาร์ค จำนวน 67 คน

ตัวแปร	ค่าสัมประสิทธิ์สหสัมพันธ์		
	ครั้งที่ 1	ครั้งที่ 2	ครั้งที่ 3
1. ระดับผลการเรียนที่นักเรียนได้รับจากโรงเรียน กับระดับผลการตัดสินที่กำหนดด้วยวิธีการแองกอฟที่ได้รับการปรับปรุง	.661**	.675**	.678**
2. ระดับผลการเรียนที่นักเรียนได้รับจากโรงเรียน กับระดับผลการตัดสินที่กำหนดด้วยวิธีบุ๊คมาร์ค	.533**	.533**	.533**
t	0.936	1.150	0.954

** มีระดับนัยสำคัญที่ระดับ .01 , t (.05 , df=9) = 2.262

จากตารางที่ 57 ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างระดับผลการเรียนที่นักเรียนได้รับจากโรงเรียน กับระดับมาตรฐานที่กำหนดมาตรฐานด้วยวิธีการแองกอฟที่ได้รับการปรับปรุง ครั้งที่ 1 ครั้งที่ 2 และครั้งที่ 3 เท่ากับ .661 .675 .678 อย่างมีนัยสำคัญที่ระดับ .01 ตามลำดับ ส่วนค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างระดับผลการเรียนที่นักเรียนได้รับจากโรงเรียน กับระดับมาตรฐานที่กำหนดด้วยวิธีการบุ๊คมาร์ค ครั้งที่ 1 ครั้งที่ 2 และครั้งที่ 3 มีค่าเท่ากันคือ .533 อย่างมีนัยสำคัญที่ระดับ .01 ตามลำดับ จะเห็นว่าทั้ง 3 ครั้ง ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างระดับผลการเรียนที่นักเรียนได้รับจากโรงเรียน กับระดับมาตรฐานที่กำหนดด้วยวิธีการแองกอฟที่ได้รับการปรับปรุง สูงกว่าวิธีบุ๊คมาร์คเล็กน้อย และเมื่อทดสอบความแตกต่างที่ระดับนัยสำคัญทางสถิติ.05 ด้วยสถิติทดสอบ t แบบ dependent พบว่า ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างระดับผลการเรียนที่นักเรียนได้รับจากโรงเรียน กับระดับมาตรฐานที่กำหนดด้วยวิธีการแองกอฟที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์คทั้ง 3 ครั้งไม่แตกต่างกัน

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตอนที่ 4 ผลการวิเคราะห์ความเหมาะสม และความเป็นไปได้ของการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และวิธีบุ๊คมาร์ค

ผู้วิจัยสร้างเครื่องมือเกี่ยวกับความเหมาะสม และความเป็นไปได้ของการกำหนดมาตรฐานวิธีการบุ๊คมาร์ค และวิธีการแองกอฟที่ได้รับการปรับปรุงประกอบด้วย แบบประเมินตนเองแบบมาตรประมาณค่า 5 ระดับ แบบสัมภาษณ์ โดยนำเกณฑ์ประเมินกระบวนการกำหนดมาตรฐานของ Pitoniak (2003, cited in Cizek., Bunch; Koons, 2004) และ Jaeger ,1989 ; Kane, 1994 ; Norcini, 1994 และ Mehrens, 1995 (Kane, 1998) มาเป็นกรอบในการสร้างเครื่องมือ โดยผู้วิจัยดำเนินการเก็บข้อมูลหลังจากที่ผู้ตัดสินทั้ง 12 คนได้กำหนดมาตรฐานวิธีแองกอฟที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์คเป็นที่เรียบร้อยแล้ว ผลการวิเคราะห์ตอนที่ 4 แบ่งเป็น 4 หัวข้อคือ

1. ผลประเมินของผู้ตัดสินเกี่ยวกับกระบวนการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และวิธีบุ๊คมาร์ค
2. ผลการพิจารณาของผู้ตัดสินเกี่ยวกับความเหมาะสม และความเป็นไปได้ของการใช้วิธีการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์ค
3. ผลการเปรียบเทียบความเหมาะสมและความเป็นไปได้ของผู้ตัดสินที่ใช้การกำหนดมาตรฐานวิธีแองกอฟที่ได้รับการปรับปรุง และวิธีบุ๊คมาร์ค
4. ข้อมูลความคิดเห็นเกี่ยวกับการกำหนดมาตรฐานวิธีแองกอฟที่ได้รับการปรับปรุง และวิธีบุ๊คมาร์คของผู้ตัดสิน

ผลการวิเคราะห์แต่ละหัวข้อมีรายละเอียดดังต่อไปนี้

1. ผลประเมินของผู้ตัดสินเกี่ยวกับกระบวนการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และวิธีบุ๊คมาร์ค

1) ข้อมูลทั่วไปเกี่ยวกับผู้ตัดสิน

ผู้ตัดสินทั้งหมด 12 คน ส่วนใหญ่เป็นครูชำนาญการ 8 คน คิดเป็นร้อยละ 66.67 รองลงมาเป็นครูชำนาญการพิเศษ 4 คน คิดเป็นร้อยละ 33.33 ส่วนใหญ่มีประสบการณ์สอนวิชาคณิตศาสตร์ มากกว่า 15 ปี ขึ้นไปมีจำนวน7 คน คิดเป็นร้อยละ 58.33 รองลงมา 11-15 ปี จำนวน 4 คน คิดเป็นร้อยละ 33.33 และ 6-10 ปี จำนวน 1 คน คิดเป็นร้อยละ 8.33 สำหรับด้านการศึกษา ส่วนใหญ่จบการศึกษาระดับปริญญาตรีจำนวน 8 คน คิดเป็นร้อยละ 66.67 รองลงมา

มีจำนวนเท่ากันคือ จบการศึกษาระดับปริญญาโทจำนวน 2 คน และกำลังศึกษาระดับปริญญาเอกจำนวน 2 คน คิดเป็นร้อยละ 16.67

2) ผลประเมินของผู้ตัดสินเกี่ยวกับกระบวนการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง

ผู้วิจัยกำหนดเกณฑ์การแปลความหมายคะแนน 5 ระดับคือ

ค่าเฉลี่ยอยู่ในช่วง 1.00 – 1.49 หมายถึง ความคิดเห็นอยู่ในระดับต่ำที่สุด
ค่าเฉลี่ยอยู่ในช่วง 1.50 – 2.49 หมายถึง ความคิดเห็นอยู่ในระดับต่ำ
ค่าเฉลี่ยอยู่ในช่วง 2.50 – 3.49 หมายถึง ความคิดเห็นอยู่ในระดับปานกลาง
ค่าเฉลี่ยอยู่ในช่วง 3.50 – 4.49 หมายถึง ความคิดเห็นอยู่ในระดับสูง
ค่าเฉลี่ยอยู่ในช่วง 4.50 – 5.00 หมายถึง ความคิดเห็นอยู่ในระดับสูงที่สุด

ผลประเมินของผู้ตัดสินเกี่ยวกับกระบวนการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง ดังตารางที่ 58

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 58 ผลประเมินของผู้ตัดสินเกี่ยวกับกระบวนการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง

รายการ	ความคิดเห็น (N=12)		
	$\bar{X}$	$S.D.$	แปลผล
1. วิธีการนี้ทำให้ท่านสามารถดำเนินการให้เป็นไปตามวัตถุประสงค์ได้	4.33	0.651	สูง
2. ผู้บรรยายอธิบายขั้นตอนการกำหนดคะแนนจุดตัดได้อย่างชัดเจน	4.58	0.515	สูงที่สุด
3. การฝึกปฏิบัติช่วยให้ท่านเข้าใจวิธีการ กำหนดคะแนนจุดตัดมากยิ่งขึ้น	4.50	0.522	สูงที่สุด
4. วิธีการนี้ง่ายต่อการกำหนดคะแนนจุดตัด	3.64	0.924	สูง
5. วิธีการนี้เป็นรูปแบบการกำหนดคะแนนจุดตัดที่ยอมรับได้	3.82	0.603	สูง
6. การบรรยายระดับการปฏิบัติมีความชัดเจน และใช้ประโยชน์ได้	4.27	0.647	สูง
7. การอภิปรายกลุ่มย่อย และกลุ่มใหญ่ช่วยให้เกิดความเข้าใจกระบวนการกำหนดคะแนนจุดตัดมากยิ่งขึ้น	4.64	0.505	สูงที่สุด
8. เวลาที่ใช้ในการอภิปรายมีความเหมาะสม	4.36	0.505	สูง
9. มีการเปิดโอกาสให้ผู้ตัดสินภายในกลุ่มได้แสดงความคิดเห็นทุกคน.	4.67	0.492	สูงที่สุด
10. วิธีการนี้ทำให้ท่านกำหนดคะแนนจุดตัดได้อย่างคล่องแคล่ว	4.09	0.701	สูง
11. การอภิปรายหลังจากกำหนดคะแนนจุดตัดรอบแรกช่วยให้ท่านกำหนดคะแนนจุดตัดได้แม่นยำขึ้น	4.08	0.515	สูง
12. การอภิปรายหลังจากกำหนดคะแนนจุดตัดรอบสองช่วยให้ท่านกำหนดคะแนนจุดตัดได้แม่นยำขึ้น	4.25	0.452	สูง
13. การอภิปรายหลังจากกำหนดคะแนนจุดตัดรอบสามช่วยให้ท่านกำหนดคะแนนจุดตัดได้แม่นยำขึ้น	4.75	0.452	สูงที่สุด
14. วิธีการนี้ทำให้ท่านเสียเวลาในการกำหนดคะแนนจุดตัด	3.25	1.215	ปานกลาง
15. ท่านมั่นใจคะแนนจุดตัดที่ได้จากการกำหนดคะแนนจุดตัดด้วยวิธีนี้	4.00	0.739	สูง
16. วิธีการนี้ทำให้ได้คะแนนจุดตัดที่มีความแม่นยำในการจำแนกความสามารถของนักเรียน	4.00	0.426	สูง
17. วิธีการนี้เหมาะสมสำหรับกำหนดคะแนนจุดตัดที่น่าเชื่อถือมากน้อยเพียงใด	4.00	0.603	สูง
18. วิธีการนี้มีความเป็นไปได้ในการนำไปใช้ในสถานการณ์ที่เป็นจริงมากน้อยเพียงใด	3.42	1.084	ปานกลาง

ตารางที่ 58 (ต่อ)

รายการ	ความคิดเห็น (N=12)		
	$\bar{X}$	$S.D.$	แปลผล
19. คำอธิบายลักษณะความสามารถของนักเรียนแต่ละระดับ ช่วยท่านในการกำหนดคะแนนจุดตัด	4.25	0.622	สูง
20. ค่าความยากของข้อสอบแต่ละข้อ ช่วยท่านกำหนดคะแนนจุดตัด	4.33	0.651	สูง
21. แบบทดสอบ ช่วยท่านในการกำหนดคะแนนจุดตัด	4.25	0.622	สูง
22. กราฟแท่งแจกแจงคะแนนของนักเรียน ช่วยท่านในการกำหนดคะแนนจุดตัด	3.92	0.669	สูง
23. แบบฟอร์มสำหรับกรอกเปอร์เซ็นต์การตอบถูกของผู้สอบคาบเส้นในแต่ละระดับ ง่ายต่อการกรอกตัวเลข	4.08	0.515	สูง
24. คู่มือการใช้อ่านเข้าใจมากน้อยเพียงใด	4.33	0.492	สูง
25. เมื่ออ่านคู่มือการใช้แล้วท่านสามารถนำไปใช้ปฏิบัติได้	4.25	0.452	สูง
รวม	**4.16**	**0.298**	**สูง**

จากตารางที่ 58 เมื่อพิจารณาภาพรวมพบว่า ค่าเฉลี่ยความคิดเห็นของผู้ตัดสินเกี่ยวกับกระบวนการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุงเท่ากับ 4.16 ส่วนเบี่ยงเบนมาตรฐาน 0.298 แสดงว่า ผู้ตัดสินมีความคิดเห็นในทางบวกต่อการกำหนดคะแนนจุดตัดด้วยวิธีแองกอฟที่ได้รับการปรับปรุงในระดับสูง นอกจากนี้ยังพบอีกว่าข้อรายการที่มีค่าเฉลี่ยสูงสุดได้แก่ การอภิปรายหลังจากกำหนดคะแนนจุดตัดรอบสามช่วยให้ผู้ตัดสินกำหนดคะแนนจุดตัดได้แม่นยำขึ้น ($\bar{X}$ = 4.75, $S.D.$ = 0.452) รองลงมาได้แก่ มีการเปิดโอกาสให้ผู้ตัดสินภายในกลุ่มได้แสดงความคิดเห็นทุกคน ($\bar{X}$ = 4.67 , $S.D.$ = 0.492) และการอภิปรายกลุ่มย่อย และกลุ่มใหญ่ช่วยให้เกิดความเข้าใจในกระบวนการกำหนดคะแนนจุดตัดมากยิ่งขึ้น.($\bar{X}$ = 4.64 , $S.D.$ = 0.505) ตามลำดับ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

3) ผลประเมินของผู้ตัดสินเกี่ยวกับกระบวนการกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์ค

ผู้วิจัยได้กำหนดเกณฑ์การแปลความหมายคะแนน 5 ระดับคือ

ค่าเฉลี่ยอยู่ในช่วง 1.00 – 1.49 หมายถึง ความคิดเห็นอยู่ในระดับน้อยที่สุด
ค่าเฉลี่ยอยู่ในช่วง 1.50 – 2.49 หมายถึง ความคิดเห็นอยู่ในระดับน้อย
ค่าเฉลี่ยอยู่ในช่วง 2.50 – 3.49 หมายถึง ความคิดเห็นอยู่ในระดับปานกลาง
ค่าเฉลี่ยอยู่ในช่วง 3.50 – 4.49 หมายถึง ความคิดเห็นอยู่ในระดับสูง
ค่าเฉลี่ยอยู่ในช่วง 4.50 – 5.00 หมายถึง ความคิดเห็นอยู่ในระดับสูงที่สุด

ผลประเมินของผู้ตัดสินเกี่ยวกับกระบวนการกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์ค ดังตารางที่ 59

ตารางที่ 59 ผลประเมินของผู้ตัดสินเกี่ยวกับกระบวนการกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์ค

รายการ	**ความคิดเห็น** (N=12)		
	$\bar{X}$	$S.D.$	**แปลผล**
1. วิธีการนี้ทำให้ท่านสามารถดำเนินการให้เป็นไปตามวัตถุประสงค์ได้	4.42	.515	สูง
2. ผู้บรรยายอธิบายขั้นตอนการกำหนดคะแนนจุดตัดได้อย่างชัดเจน	4.58	.515	สูงที่สุด
3. การฝึกปฏิบัติช่วยให้ท่านเข้าใจวิธีการกำหนดคะแนนจุดตัดมากยิ่งขึ้น	4.75	0.452	สูงที่สุด
4. วิธีการนี้ง่ายต่อการกำหนดคะแนนจุดตัด	4.67	0.651	สูงที่สุด
5. วิธีการนี้เป็นรูปแบบการกำหนดคะแนนจุดตัดที่ยอมรับได้	4.08	0.669	สูง
6. การบรรยายระดับการปฏิบัติมีความชัดเจน และใช้ประโยชน์ได้	4.42	0.515	สูง
7. การอภิปรายกลุ่มย่อย และกลุ่มใหญ่ช่วยให้เกิดความเข้าใจกระบวนการกำหนดคะแนนจุดตัดมากยิ่งขึ้น	4.83	0.389	สูงที่สุด
8. เวลาที่ใช้ในการอภิปรายมีความเหมาะสม	4.58	0.515	สูงที่สุด
9. มีการเปิดโอกาสให้ผู้ตัดสินภายในกลุ่มได้แสดงความคิดเห็นทุกคน	4.75	0.452	สูงที่สุด
10. วิธีการนี้ทำให้ท่านกำหนดคะแนนจุดตัดได้อย่างคล่องแคล่ว	4.58	0.669	สูงที่สุด
11. การอภิปรายหลังจากกำหนดคะแนนจุดตัดรอบแรกช่วยให้ท่านกำหนดคะแนนจุดตัดได้แม่นยำขึ้น	4.58	0.515	สูงที่สุด
12. การอภิปรายหลังจากกำหนดคะแนนจุดตัดรอบสองช่วยให้ท่านกำหนดคะแนนจุดตัดได้แม่นยำขึ้น	4.55	0.522	สูงที่สุด
13. การอภิปรายหลังจากกำหนดคะแนนจุดตัดรอบสามช่วยให้ท่านกำหนดคะแนนจุดตัดได้แม่นยำขึ้น	4.73	0.467	สูงที่สุด

ตารางที่ 59 (ต่อ)

รายการ	ความคิดเห็น (N=12)		
	$\bar{X}$	$S.D.$	แปลผล
14. วิธีการนี้ทำให้ท่านเสียเวลาในการกำหนดคะแนนจุดตัด	2.84	0.505	ปานกลาง
15. ท่านมั่นใจคะแนนจุดตัดที่ได้จากการกำหนดคะแนนจุดตัดด้วยวิธีนี้	4.36	0.505	สูง
16. วิธีการนี้ทำให้ได้คะแนนจุดตัดที่มีความแม่นยำในการจำแนกความสามารถของนักเรียน	4.27	0.467	สูง
17. วิธีการนี้มีความเป็นไปได้ในการนำไปใช้ในสถานการณ์ที่เป็นจริงมากน้อยเพียงใด	4.18	0.751	สูง
18. วิธีการนี้เหมาะสมสำหรับกำหนดคะแนนจุดตัดที่น่าเชื่อถือมากน้อยเพียงใด	4.00	0.447	สูง
19. คำอธิบายลักษณะความสามารถของนักเรียนแต่ละระดับ ช่วยท่านในการกำหนดคะแนนจุดตัด	4.55	0.522	สูงที่สุด
20. คู่มือจัดเรียงข้อสอบ ทำให้เกิดความสะดวกในการกำหนดคะแนนจุดตัด	4.55	0.522	สูงที่สุด
21. แบบฟอร์มสำหรับใส่เลขที่หน้ามีความง่ายต่อการกรอก	4.45	0.688	สูง
22. เปอร์เซนต์ของนักเรียนที่อยู่ในแต่ละระดับช่วยท่านในการกำหนดคะแนนจุดตัด	4.55	0.522	สูงที่สุด
23. ค่าเฉลี่ยคะแนนจุดตัดของแต่ละกลุ่มย่อย ช่วยท่านในการกำหนดคะแนนจุดตัด	4.36	0.505	สูง
24. คู่มือการใช้อ่านเข้าใจมากน้อยเพียงใด	4.27	0.467	สูง
25 เมื่อท่านอ่านคู่มือการใช้แล้วท่านสามารถนำไปใช้ปฏิบัติได้	4.27	0.467	สูง
รวม	**4.46**	**0.313**	**สูง**

จากตารางที่ 59 เมื่อพิจารณาภาพรวมพบว่า ค่าเฉลี่ยความคิดเห็นของผู้ตัดสินเกี่ยวกับกระบวนการกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์ค 4.46 ส่วนเบี่ยงเบนมาตรฐาน 0.313 แสดงว่าผู้ตัดสินมีความคิดเห็นในทางบวกต่อการตัดสินคะแนนจุดตัดด้วยวิธีบุ๊คมาร์คในระดับสูง นอกจากนี้ยังพบอีกว่าข้อรายการที่มีค่าเฉลี่ยสูงสุดได้แก่ การอภิปรายกลุ่มย่อย และกลุ่มใหญ่ช่วยให้เกิดความเข้าใจในกระบวนการกำหนดคะแนนจุดตัดมากยิ่งขึ้น ($\bar{X}$ = 4.83, $S.D.$ = 0.389) รองลงมาได้แก่ การฝึกปฏิบัติการกำหนดคะแนนจุดตัดช่วยให้ผู้ตัดสินเข้าใจวิธีการกำหนดคะแนนจุดตัดมากยิ่งขึ้น และมีการเปิดโอกาสให้ผู้ตัดสินภายในกลุ่มได้แสดงความคิดเห็นทุกคน ($\bar{X}$ = 4.75,

$S.D.$ = 0.452) การอภิปรายหลังจากกำหนดคะแนนจุดตัดรอบสามช่วยให้ผู้ตัดสินกำหนดคะแนนจุดตัดได้แม่นยำขึ้น ($\overline{X}$ = 4.73, $S.D.$ = 0.467) ตามลำดับ

2. ผลการพิจารณาของผู้ตัดสินเกี่ยวกับความเหมาะสม และความเป็นไปได้ของการใช้วิธีการกำหนดคะแนนจุดตัดด้วยวิธีแองกอฟที่ได้รับการปรับปรุงกับวิธีบุ๊คมาร์ค

ผู้วิจัยได้จัดสัมมนาเชิงปฏิบัติการวิธีการกำหนดคะแนนจุดตัดด้วยวิธีแองกอฟที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์ค และได้เชิญผู้ตัดสินจำนวน 12 คน เข้าร่วมการอบรมและดำเนินการปฏิบัติจริง หลังจากดำเนินการเป็นที่เรียบร้อยแล้ว ผู้วิจัยได้แจกแบบสอบถามเกี่ยวกับความเหมาะสม และความเป็นไปได้ของการใช้วิธีการกำหนดคะแนนจุดตัดด้วยวิธีแองกอฟที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์ค โดยพิจารณาค่ามัธยฐาน(Median : Mdn) และพิสัยระหว่างควอไทล์ (Inter quartile range : IR)

เกณฑ์ค่ามัธยฐาน (Mdn) ที่ใช้ในการแปลความหมายเป็น 5 ระดับ คือ

ค่ามัธยฐานอยู่ในช่วง 1.00–1.49 หมายถึง วิธีการแองกอฟที่ได้รับการปรับปรุงและวิธีการบุ๊คมาร์คมีความเหมาะสม และความเป็นไปได้น้อยที่สุด

ค่ามัธยฐานอยู่ในช่วง 1.50–2.49 หมายถึง วิธีการแองกอฟที่ได้รับการปรับปรุงและวิธีการบุ๊คมาร์คมีความเหมาะสม และความเป็นไปได้น้อย

ค่ามัธยฐานอยู่ในช่วง 2.50–3.49 หมายถึง วิธีการแองกอฟที่ได้รับการปรับปรุงและวิธีการบุ๊คมาร์ค มีความเหมาะสม และความเป็นไปได้ปานกลาง

ค่ามัธยฐานอยู่ในช่วง 3.50–4.49 หมายถึง วิธีการแองกอฟที่ได้รับการปรับปรุงและวิธีการบุ๊คมาร์ค มีความเหมาะสม และความเป็นไปได้มาก

ค่ามัธยฐานอยู่ในช่วง 4.50–5.00 หมายถึง วิธีการแองกอฟที่ได้รับการปรับปรุงและวิธีการบุ๊คมาร์ค มีความเหมาะสม และความเป็นไปได้มากที่สุด

เกณฑ์พิสัยระหว่างควอไทล์ (IR) ที่ใช้แปลความหมายคือ

ค่าพิสัยระหว่างควอไทล์ไม่เกิน 1.5 หมายถึง ความคิดเห็นของผู้ตัดสินสอดคล้องกัน

ค่าพิสัยระหว่างควอไทล์เกิน 1.5 หมายถึง ความคิดเห็นของผู้ตัดสินไม่สอดคล้องกัน

ผลการพิจารณาของผู้ตัดสินเกี่ยวกับความเหมาะสม และความเป็นไปได้ของการใช้วิธีการกำหนดคะแนนจุดตัดด้วยวิธีแองกอฟที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์ค มีดังนี้

ตารางที่ 60 ค่ามัธยฐาน (Median: Mdn) และพิสัยระหว่างควอไทล์ (Inter quartile range: IR) จากการประเมินความเหมาะสมและความเป็นไปได้ในการใช้วิธีการกำหนดคะแนนจุดตัดด้วยวิธีแองกอฟที่ได้รับการปรับปรุง

รายการ	Mdn	แปลผล	IR	แปลผล
ด้านความเหมาะสม				
1. วัตถุประสงค์การกำหนดคะแนนจุดตัด	4.00	มาก	0.75	สอดคล้อง
2. กระบวนการกำหนดคะแนนจุดตัดทำได้ง่าย ไม่ยุ่งยาก	3.00	ปานกลาง	1.00	สอดคล้อง
3. การดำเนินการกำหนดคะแนนจุดตัด				
3.1 ความสมเหตุสมผลของกระบวนการ	4.00	มาก	1.00	สอดคล้อง
3.2 การจัดให้มีการฝึกปฏิบัติหลังฟังบรรยาย	4.50	มากที่สุด	1.00	สอดคล้อง
3.3 การจัดให้มีการอภิปราย และเปิดโอกาสให้เปลี่ยนคะแนนจุดตัด 3 รอบ	4.50	มากที่สุด	1.00	สอดคล้อง
3.4 การจัดให้มีการกำหนดมาตรฐานการปฏิบัติ	4.00	มาก	1.00	สอดคล้อง
3.5 คุณสมบัติผู้ตัดสิน คือ มีประสบการณ์การสอนคณิตศาสตร์ไม่น้อยกว่า 5 ปี และเป็นครูระดับชำนาญการขึ้นไป	5.00	มากที่สุด	1.00	สอดคล้อง
4. วิธีการนี้มีความน่าเชื่อถือ	4.00	มาก	1.00	สอดคล้อง
5. เอกสารประกอบการพิจารณาคะแนนจุดตัด				
5.1 แบบทดสอบ	4.00	มาก	1.00	สอดคล้อง
5.2 ค่าความยากของข้อสอบ	4.00	มาก	2.00	ไม่สอดคล้อง
5.3 กราฟแท่งแจกแจงคะแนนนักเรียน	4.00	มาก	0.75	สอดคล้อง
5.4 คะแนนจุดตัดของผู้ตัดสินแต่ละคน	4.00	มาก	0.75	สอดคล้อง
5.5 ค่าเฉลี่ยคะแนนจุดตัดของแต่ละกลุ่มย่อย	4.00	มาก	1.00	สอดคล้อง
รวม	**4.08**	**มาก**	**1.02**	**สอดคล้อง**

ตารางที่ 60 (ต่อ)

รายการ	Mdn	แปลผล	IR	แปลผล
ด้านความเป็นไปได้				
1. วัตถุประสงค์การกำหนดคะแนนจุดตัด	4.00	มาก	0.75	สอดคล้อง
2. กระบวนการกำหนดคะแนนจุดตัดทำได้ง่าย ไม่ยุ่งยาก	3.00	ปานกลาง	2.00	ไม่สอดคล้อง
3. การดำเนินการกำหนดคะแนนจุดตัด				
3.1 ความสมเหตุสมผลของกระบวนการ	4.00	มาก	1.00	สอดคล้อง
3.2 การจัดให้มีการฝึกปฏิบัติหลังฟังบรรยาย	4.50	มากที่สุด	1.00	สอดคล้อง
3.3 การจัดให้มีการอภิปราย และเปิดโอกาสให้เปลี่ยนคะแนนจุดตัด 3 รอบ	5.00	มากที่สุด	1.00	สอดคล้อง
3.4 การจัดให้มีการกำหนดมาตรฐานการปฏิบัติ	4.00	มาก	1.00	สอดคล้อง
3.5 คุณสมบัติผู้ตัดสิน คือ มีประสบการณ์การสอนคณิตศาสตร์ไม่น้อยกว่า 5 ปี และเป็นครูระดับชำนาญการขึ้นไป	5.00	มากที่สุด	1.00	สอดคล้อง
4. วิธีการนี้มีความน่าเชื่อถือ	4.00	มาก	2.00	ไม่สอดคล้อง
5. เอกสารประกอบการพิจารณาคะแนนจุดตัด				
5.1 แบบทดสอบ	5.00	มากที่สุด	1.00	สอดคล้อง
5.2 ค่าความยากของข้อสอบ	4.00	มาก	1.25	สอดคล้อง
5.3 กราฟแท่งแจกแจงคะแนนนักเรียน	4.00	มาก	0.75	สอดคล้อง
5.4 คะแนนจุดตัดของผู้ตัดสินแต่ละคน	4.00	มาก	1.00	สอดคล้อง
5.5 ค่าเฉลี่ยคะแนนจุดตัดของแต่ละกลุ่มย่อย	4.00	มาก	1.00	สอดคล้อง
รวม	**4.19**	**มาก**	**1.13**	**สอดคล้อง**

จากตารางที่ 60 เมื่อพิจารณาด้านความเหมาะสม ในภาพรวมผู้ตัดสินมีความคิดเห็นว่าวิธีการกำหนดคะแนนจุดตัดด้วยวิธีแองกอฟที่ได้รับการปรับปรุงมีความเหมาะสมระดับมาก (Mdn= 4.08) เมื่อพิจารณาแยกทีละประเด็น ผู้ตัดสินส่วนใหญ่มีความคิดเห็นสอดคล้องกันว่ามี

ความเหมาะสมมากที่สุดในประเด็นต่อไปนี้ 1) การจัดให้มีการฝึกปฏิบัติหลังฟังบรรยาย 2) การจัดให้มีการอภิปราย และเปิดโอกาสให้เปลี่ยนคะแนนจุดตัด 3 รอบ และ 3) คุณสมบัติผู้ตัดสิน คือ มีประสบการณ์การสอนคณิตศาสตร์ไม่น้อยกว่า 5 ปี และเป็นครูระดับชำนาญการขึ้นไป ส่วนผู้ตัดสินมีความคิดเห็นสอดคล้องกันว่ามีความเหมาะสมปานกลางมีประเด็นเดียวคือ กระบวนการกำหนดคะแนนจุดตัดทำได้ง่าย ไม่ยุ่งยาก นอกจากนี้ยังพบว่า มีประเด็นหนึ่งที่ผู้ตัดสินมีความคิดเห็นไม่สอดคล้องกันคือ เอกสารประกอบการพิจารณาคะแนนจุดตัดคือ ค่าความยากของข้อสอบ

เมื่อพิจารณาด้านความเป็นไปได้ในภาพรวมผู้ตัดสินมีความคิดเห็นว่าวิธีการกำหนดคะแนนจุดตัดด้วยวิธีแองกอฟที่ได้รับการปรับปรุงมีความเป็นไปได้ระดับมาก (Mdn= 4.19) เมื่อพิจารณาแยกทีละประเด็น ผู้ตัดสินส่วนใหญ่มีความคิดเห็นสอดคล้องกันว่ามีความเป็นไปได้มากที่สุดในประเด็นต่อไปนี้ 1) การจัดให้มีการฝึกปฏิบัติหลังฟังบรรยาย 2) การจัดให้มีการอภิปรายและเปิดโอกาสให้เปลี่ยนคะแนนจุดตัด 3 รอบ 3) คุณสมบัติผู้ตัดสิน คือ มีประสบการณ์การสอนคณิตศาสตร์ไม่น้อยกว่า 5 ปี และเป็นครูระดับชำนาญการขึ้นไป 4) เอกสารประกอบการพิจารณาคะแนนจุดตัดคือแบบทดสอบ นอกจากนี้ยังพบว่า มีประเด็นหนึ่งที่ผู้ตัดสินมีความคิดเห็นไม่สอดคล้องกันคือ กระบวนการกำหนดคะแนนจุดตัดทำได้ง่าย ไม่ยุ่งยาก และวิธีการที่มีความน่าเชื่อถือ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 61 ค่ามัธยฐาน (Median: Mdn) และพิสัยระหว่างควอไทล์ (Inter quartile range: IR) จากการประเมินความเหมาะสมและความเป็นไปได้ในการใช้วิธีการกำหนดคะแนนจุดตัดด้วยวิธีบุ๊คมาร์ค

รายการ	Mdn	แปลผล	IR	แปลผล
ด้านความเหมาะสม				
1. วัตถุประสงค์การกำหนดคะแนนจุดตัด	4.50	มากที่สุด	1.00	สอดคล้อง
2. กระบวนการกำหนดคะแนนจุดตัดทำได้ง่าย ไม่ยุ่งยาก	5.00	มากที่สุด	0.00	สอดคล้อง
3. การดำเนินการกำหนดคะแนนจุดตัด				
3.1 ความสมเหตุสมผลของกระบวนการ	4.00	มาก	1.00	สอดคล้อง
3.2 การจัดให้มีการฝึกปฏิบัติหลังฟังบรรยาย	5.00	มากที่สุด	1.00	สอดคล้อง
3.3 การจัดให้มีการอภิปราย และเปิดโอกาสให้เปลี่ยนคะแนนจุดตัด 3 รอบ	5.00	มากที่สุด	0.75	สอดคล้อง
3.4 การจัดให้มีการกำหนดมาตรฐานการปฏิบัติ	4.50	มากที่สุด	1.00	สอดคล้อง
3.5 คุณสมบัติผู้ตัดสิน คือ มีประสบการณ์การสอนคณิตศาสตร์ไม่น้อยกว่า5 ปี และเป็นครูระดับชำนาญการขึ้นไป	5.00	มากที่สุด	1.00	สอดคล้อง
4. วิธีการนี้มีความน่าเชื่อถือ	4.00	มาก	1.75	ไม่สอดคล้อง
5. เอกสารประกอบการพิจารณาคะแนนจุดตัด				
5.1 คำอธิบายลักษณะความสามารถของนักเรียนแต่ละระดับ	4.00	มาก	1.00	สอดคล้อง
5.2 คู่มือจัดเรียงข้อสอบ	5.00	มากที่สุด	1.00	สอดคล้อง
5.3 เปอร์เซนต์ของนักเรียนที่อยู่ในแต่ละระดับ	5.00	มากที่สุด	1.00	สอดคล้อง
5.4 ค่าเฉลี่ยของคะแนนจุดตัดจากกลุ่มย่อย	5.00	มากที่สุด	1.00	สอดคล้อง
5.5 แบบฟอร์มกรอกเลขที่หน้าจากในคู่มือจัดเรียงข้อสอบ	5.00	มากที่สุด	0.75	สอดคล้อง
รวม	**4.69**	**มากที่สุด**	**0.94**	**สอดคล้อง**

ตารางที่ 61 (ต่อ)

รายการ	Mdn	แปลผล	IR	แปลผล
ด้านความเป็นไปได้				
1. วัตถุประสงค์การกำหนดคะแนนจุดตัด	4.00	มาก	1.00	สอดคล้อง
2. กระบวนการกำหนดคะแนนจุดตัดทำได้ง่าย ไม่ยุ่งยาก	5.00	มากที่สุด	0.00	สอดคล้อง
3. การดำเนินการกำหนดคะแนนจุดตัด				
3.1 ความสมเหตุสมผลของกระบวนการ	4.00	มาก	1.00	สอดคล้อง
3.2 การจัดให้มีการฝึกปฏิบัติหลังฟังบรรยาย	5.00	มากที่สุด	1.00	สอดคล้อง
3.3 การจัดให้มีการอภิปราย และเปิดโอกาสให้เปลี่ยนคะแนนจุดตัด 3 รอบ	5.00	มากที่สุด	1.00	สอดคล้อง
3.4 การจัดให้มีการกำหนดมาตรฐานการปฏิบัติ	4.50	มากที่สุด	0.75	สอดคล้อง
3.5 คุณสมบัติผู้ตัดสิน คือ มีประสบการณ์การสอนคณิตศาสตร์ไม่น้อยกว่า5 ปี และเป็นครูระดับชำนาญการขึ้นไป	5.00	มากที่สุด	1.00	สอดคล้อง
4. วิธีการนี้มีความน่าเชื่อถือ	4.00	มาก	1.00	สอดคล้อง
5. เอกสารประกอบการพิจารณาคะแนนจุดตัด				
5.1 คำอธิบายลักษณะความสามารถของนักเรียนแต่ละระดับ	4.00	มาก	1.50	สอดคล้อง
5.2 คู่มือจัดเรียงข้อสอบ	5.00	มากที่สุด	1.00	สอดคล้อง
5.3 เปอร์เซนต์ของนักเรียนที่อยู่ในแต่ละระดับ	4.50	มากที่สุด	1.00	สอดคล้อง
5.4 ค่าเฉลี่ยของคะแนนจุดตัดจากกลุ่มย่อย	4.50	มากที่สุด	1.00	สอดคล้อง
5.5 แบบฟอร์มกรอกเลขที่หน้าจากในคู่มือจัดเรียงข้อสอบ	5.00	มากที่สุด	0.75	สอดคล้อง
รวม	**4.58**	**มากที่สุด**	**0.85**	**สอดคล้อง**

จากตารางที่ 61 เมื่อพิจารณาด้านความเหมาะสม ในภาพรวมผู้ตัดสินมีความคิดเห็นว่าวิธีการกำหนดคะแนนจุดตัดด้วยวิธีบุ๊คมาร์คมีความเหมาะสมระดับมากที่สุด (Mdn= 4.69) เมื่อพิจารณาแยกเป็นประเด็น ผู้ตัดสินส่วนใหญ่มีความคิดเห็นสอดคล้องกันว่ามีความเหมาะสมมาก

ที่สุดในประเด็นต่อไปนี้ 1) วัตถุประสงค์การกำหนดคะแนนจุดตัด 2) กระบวนการกำหนดคะแนนจุดตัดทำได้ง่าย ไม่ยุ่งยาก 3) การดำเนินการกำหนดคะแนนจุดตัดในด้านการจัดให้มีการฝึกปฏิบัติหลังฟังบรรยาย ด้านการจัดให้มีการอภิปราย และด้านเปิดโอกาสให้เปลี่ยนคะแนนจุดตัด 3 รอบ การจัดให้มีการกำหนดมาตรฐานการปฏิบัติ และคุณสมบัติผู้ตัดสินคือ มีประสบการณ์การสอนคณิตศาสตร์ไม่น้อยกว่า 5 ปี และเป็นครูระดับชำนาญการขึ้นไป 4) เอกสารประกอบการพิจารณาคะแนนจุดตัดประกอบด้วย คู่มือจัดเรียงข้อสอบ เปอร์เซนต์ของนักเรียนที่อยู่ในแต่ละระดับ ค่าเฉลี่ยของคะแนนจุดตัดจากกลุ่มย่อย และแบบฟอร์มกรอกเลขที่หน้าจากในคู่มือจัดเรียงข้อสอบ นอกจากนี้ยังพบว่า มีประเด็นหนึ่งที่ผู้ตัดสินมีความคิดเห็นไม่สอดคล้องกันคือ วิธีการนี้มีความน่าเชื่อถือ

เมื่อพิจารณาด้านความเป็นไปได้ ภาพรวมผู้ตัดสินมีความคิดเห็นว่าวิธีการกำหนดคะแนนจุดตัดด้วยวิธีบุ๊คมาร์คมีความเป็นไปได้ระดับมากที่สุด (Mdn= 4.58) เมื่อพิจารณาแยกทีละประเด็น ผู้ตัดสินส่วนใหญ่มีความคิดเห็นสอดคล้องกันว่ามีความเป็นไปได้มากที่สุดในประเด็นต่อไปนี้ 1) กระบวนการกำหนดคะแนนจุดตัดทำได้ง่าย ไม่ยุ่งยาก 2) การดำเนินการกำหนดคะแนนจุดตัดในด้านการจัดให้มีการฝึกปฏิบัติหลังฟังบรรยาย ด้านการจัดให้มีการอภิปราย และเปิดโอกาสให้เปลี่ยนคะแนนจุดตัด 3 รอบ ด้านการจัดให้มีการกำหนดมาตรฐานการปฏิบัติ และด้านคุณสมบัติผู้ตัดสิน คือ มีประสบการณ์การสอนคณิตศาสตร์ไม่น้อยกว่า 5 ปี และเป็นครูระดับชำนาญการขึ้นไป 3) เอกสารประกอบการพิจารณาคะแนนจุดตัดประกอบด้วย คู่มือจัดเรียงข้อสอบ เปอร์เซนต์ของนักเรียนที่อยู่ในแต่ละระดับ ค่าเฉลี่ยของคะแนนจุดตัดจากกลุ่มย่อย และแบบฟอร์มกรอกเลขที่หน้าจากในคู่มือจัดเรียงข้อสอบ

3. ผลการเปรียบเทียบความเหมาะสมและความเป็นไปได้ของผู้ตัดสินที่ใช้การกำหนดมาตรฐานวิธีแองกอฟที่ได้รับการปรับปรุง และวิธีบุ๊คมาร์ค

ผู้ตัดสินจำนวน 12 คน ทำการตัดสินคะแนนจุดตัดทั้งสองวิธีคือ วิธีแองกอฟที่ได้รับการปรับปรุง และวิธีบุ๊คมาร์ค เมื่อนำผลคะแนนจุดตัดมาเปรียบเทียบความเหมาะสมและความเป็นไปได้ของทั้งสองวิธีด้วยสถิติทดสอบ t แบบ depentdent

ตารางที่ 62 การเปรียบเทียบความเหมาะสมระหว่างวิธีแองกอฟที่ได้รับการปรับปรุง และวิธีบุ๊คมาร์ค

	$\bar{X}$	$\bar{D}$	S_d	t	p
วิธีแองกอฟที่ได้รับการปรับปรุง	4.20	-0.31	0.21	-5.146	.000*
วิธีบุ๊คมาร์ค	4.51				

* P<.05

จากตารางที่ 62 พบว่า ค่าเฉลี่ยความเหมาะสมของวิธีบุ๊คมาร์คมากกว่าค่าเฉลี่ยความเหมาะสมของวิธีแองกอฟที่ได้รับการปรับปรุงอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

ตารางที่ 63 การเปรียบเทียบความเป็นไปได้ระหว่างวิธีแองกอฟที่ได้รับการปรับปรุง และวิธีบุ๊คมาร์ค

	$\bar{X}$	$\bar{D}$	S_d	t	p
วิธีแองกอฟที่ได้รับการปรับปรุง	4.19	-0.31	0.24	-4.560	.001*
วิธีบุ๊คมาร์ค	4.50				

* P<.05

จากตารางที่ 63 พบว่า ค่าเฉลี่ยความเป็นไปได้ของวิธีบุ๊คมาร์คมากกว่าค่าเฉลี่ยความเป็นไปได้ของวิธีแองกอฟที่ได้รับการปรับปรุงอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

4. ข้อมูลความคิดเห็นของผู้ตัดสินเกี่ยวกับการกำหนดมาตรฐานด้วยวิธีการแองกอฟที่ได้รับการปรับปรุง และวิธีการบุ๊คมาร์ค

4.1 ความคิดเห็นของผู้ตัดสินเกี่ยวกับจุดเด่นวิธีการแองกอฟที่ได้รับการปรับปรุง

ผู้ตัดสินส่วนใหญ่มีความเห็นว่า วิธีการแองกอฟที่ได้รับการปรับปรุงเป็นวิธีที่มีจุดเด่นคือ มีการอภิปรายกลุ่มถึง 3 ครั้ง การอภิปรายกลุ่มทำให้เห็นมุมมองอีกด้านจากผู้ตัดสินคนอื่นมาประกอบการพิจารณาใหม่อีกครั้ง มุมมองดังกล่าวเช่น ประสบการณ์เกี่ยวกับผู้เรียน ลักษณะข้อสอบ คะแนนจุดตัด เป็นต้น และจุดเด่นอีกประการหนึ่งคือ การจัดเอกสารประกอบการพิจารณา เช่น ค่าความยาก มีส่วนช่วยประกอบการพิจารณาตัดสินคะแนนจุดตัดได้เป็นอย่างดี

นอกจากนี้มีผู้ตัดสินบางส่วนเห็นว่า การที่ให้ผู้ตัดสินพิจารณาข้อสอบทีละข้อนั้นทำให้คะแนนจุดตัดมีความน่าเชื่อถือ

"พี่คิดว่า วิธีนี้น่าจะทำให้ได้คะแนนจุดตัดที่น่าเชื่อถือนะ เหมาะกับสภาพความเป็นจริงของเด็ก เพราะ ได้วิเคราะห์ข้อสอบทีละข้อ เพื่อหาจุดตัดตามความสามารถของเด็กกลุ่มต่างๆ ตามความเป็นจริง แล้วพี่ว่าการที่มีการอภิปรายผลการตัดสินถึง 3 ครั้ง เป็นเหมือนการตรวจสอบตั้ง 3 หนแนะ"

ผู้ตัดสิน A

" อืม ! วิธีการแองกอฟ ฯ มีจุดเด่นตรงไหนหรือ..... ข้อมูลได้รับการกลั่นกรองจากผู้ที่มีประสบการณ์ มีความชำนาญในการสอนคณิตศาสตร์ ที่เรียกว่า การระดมสมอง ทำให้ผลที่ได้ออกมาค่อนข้างสมบูรณ์ ตรงกับความเป็นจริง แล้วมีขั้นตอนมากมาย น่าเชื่อถือ น่าจะนำไปใช้ประโยชน์ในการกำหนดระดับคะแนนได้ดี"

ผู้ตัดสิน L

4.2 ความคิดเห็นของผู้ตัดสินเกี่ยวกับจุดอ่อนวิธีการแองกอฟที่ได้รับการปรับปรุง

ผู้ตัดสินทั้งหมดมีความคิดเห็นตรงกันว่า วิธีการแองกอฟที่ได้รับการปรับปรุง มีขั้นตอนมาก ทำให้ใช้เวลาในการตัดสินนาน เสียเวลาในการกำหนดคะแนนจุดตัด สาเหตุหนึ่งเนื่องมาจากข้อสอบไม่ได้จัดเรียงตามความยากข้อสอบ

"เป็นวิธีการที่ต้องใช้เวลามาก เพราะจะต้องพิจารณาข้อสอบทีละข้อ และต้องทำอีกถึง 3 ครั้ง แล้วยังต้องอภิปรายถึงความยากง่ายของข้อสอบในแต่ละข้ออีก ข้อสอบก็ไม่ได้จัดเรียงไว้ให้เสียเวลามากเลย"

ผู้ตัดสิน B

"ใช้เวลาในการหาจุดตัดนาน เพราะข้อมูลมีมาก และทำหลายขั้นตอน อาจทำให้เกิดข้อผิดพลาดได้ง่ายในการหาจุดตัดของแต่ละระดับ"

ผู้ตัดสิน D

4.3 ความคิดเห็นของผู้ตัดสินเกี่ยวกับจุดเด่นวิธีการบุ๊คมาร์ค

ผู้ตัดสินทุกคนมีความคิดเห็นตรงกันว่าวิธีบุ๊คมาร์คเป็นวิธีที่ง่ายต่อการนำไปใช้ ขั้นตอนไม่ยุ่งยาก สะดวก ประหยัดเวลา เนื่องจากการพิจารณาข้อสอบทำได้รวดเร็วเพราะมีการจัดเรียงข้อสอบตามความยากของข้อสอบจากง่ายไปหายาก ทำให้การพิจารณาตัดสินคะแนนจุดตัดเป็นไปอย่างรวดเร็ว ผู้ตัดสินมีความคล่องตัวในการดำเนินงานตัดสิน

"พี่ว่า วิธีนี้ง่าย สะดวก ประหยัดเวลาดีนะ ลำดับขั้นตอนการทำงานไม่สลับซับซ้อน เนื่องจากข้อสอบได้เรียงลำดับความยากง่ายไว้เรียบร้อยแล้ว"

ผู้ตัดสิน A

"วิธีนี้ก็ดีเหมือนกัน มีการจัดเรียงข้อสอบจากง่ายไปยากทำให้ผู้ตัดสินมีความคล่องตัว ตัดสินได้รวดเร็วดี ไม่เสียเวลามากเหมือนวิธีแองกอฟ"

ผู้ตัดสิน F

4.4 ความคิดเห็นของผู้ตัดสินเกี่ยวกับจุดอ่อนวิธีการบุ๊คมาร์ค

ผู้ตัดสินบางส่วนมีความคิดเห็นว่า วิธีการบุ๊คมาร์คเป็นวิธีค่อนข้างหยาบ ถึงแม้จะใช้เวลารวดเร็วในการตัดสิน และผู้ตัดสินอีกบางส่วนเห็นว่า ถ้าจัดเรียงข้อสอบผิดจะทำให้คะแนนจุดตัดผิดได้ และยังให้ความคิดเห็นเพิ่มเติมอีกว่า การเรียงข้อสอบตามค่าความยากง่ายบางข้อไม่ตรงกับความเป็นจริง เช่น ข้อสอบง่ายที่ทุกคนน่าจะทำได้ แต่จริงๆ แล้วนักเรียนบางคนอาจทำไม่ได้ก็ได้

"อันที่จริงวิธีนี้ก็ใช้ได้ แต่พี่ว่าเป็นวิธีที่ค่อนข้างหยาบ มีค่าความเที่ยงน้อย ถ้าจัดเรียงข้อสอบผิดก็จะทำให้ค่ามาตรฐานผิดได้"

ผู้ตัดสิน E

"การนำข้อสอบมาเรียงลำดับ ตามค่าความยากง่าย แล้วหาจุดตัด วิธีการนี้จะไม่ตรงกับความจริง เช่น ข้อสอบที่ง่ายที่สุด มักจะไม่ถูกมาร์ค จะอยู่ในลักษณะที่คาดว่าผู้สอบผ่านทุกคนจะทำได้ แต่ความจริงอาจมีผู้สอบผ่านบางคนที่ทำไม่ได้ก็ได้"

ผู้ตัดสิน J

4.5 ความคิดเห็นของผู้ตัดสินเกี่ยวกับความเชื่อมั่นในคะแนนจุดตัดระหว่างวิธีการแองกอฟที่ได้รับการปรับปรุง และวิธีการบุ๊คมาร์ค

ผู้ตัดสินทั้งหมด 12 คน มีความคิดเห็นเกี่ยวกับ ความเชื่อมั่นในคะแนนจุดตัดด้วยวิธีการกำหนดมาตรฐานสองวิธีดังนี้ ผู้ตัดสินจำนวน 7 คน มีความเชื่อมั่นในคะแนนจุดตัดด้วยวิธีการแองกอฟที่ได้รับการปรับปรุง ส่วนอีก 3 คน มีความเชื่อมั่นในคะแนนจุดตัดด้วยวิธีบุ๊คมาร์ค และอีก 2 คน มีความคิดเห็นว่าทั้งสองมีความเชื่อมั่นพอๆ กัน สำหรับเหตุผลที่ผู้ตัดสินส่วนใหญ่เลือกวิธีแองกอฟที่ได้รับการปรับปรุง คือ วิธีนี้มีการพิจารณาข้อสอบทีละข้อว่าเปอร์เซ็นต์ที่นักเรียนระดับนั้นจะตอบถูกเป็นเท่าไร และทำการตัดสินคะแนนจุดตัด 3 ครั้ง โดยมีการอภิปรายทั้งกลุ่มย่อย และกลุ่มใหญ่ ซึ่งเป็นการระดมความรู้ การแลกเปลี่ยนความคิดเห็นจากผู้ตัดสินที่มีประสบการณ์ในการสอนมาไม่น้อยกว่า 5 ปี ส่วนเหตุผลของผู้ตัดสินส่วนใหญ่ที่เลือกวิธีบุ๊คมาร์คคือ วิธีนี้มีการจัดเรียงข้อสอบจากง่ายที่สุดไปยากที่สุด ทำให้การพิจารณาตัดสินคะแนนจุดตัดทำให้ง่ายขึ้น นอกจากนี้ยังมีการพิจารณาคะแนนจุดตัด 3 ครั้ง และมีการอภิปรายทั้งกลุ่มใหญ่ กลุ่มย่อย

"*แองกอฟ เพราะมีความละเอียดกว่า ต้องพิจารณาข้อสอบในแต่ละข้อ และทำถึง 3 ครั้ง มีการอภิปรายทั้งกลุ่มย่อย และกลุ่มใหญ่ พี่เชื่อมั่นในวิธีนี้คะ*"

ผู้ตัดสิน A

"*วิธีแองกอฟ เพราะต้องผ่านการกลั่นกรองจากผู้ที่มีความชำนาญ มีประสบการณ์ในด้านการสอนมาไม่น้อยกว่า 5 ปี และมีขั้นตอน 3 ขั้นตอน แต่ละขั้นต้องผ่านการระดมความรู้ ความคิดการแลกเปลี่ยนความคิดเห็นจนได้ข้อมูลออกมา น่าเชื่อถือ ควรที่จะนำไปใช้ประโยชน์ได้*"

ผู้ตัดสิน L

"*วิธีการบุ๊คมาร์คคะ มันมีการจัดเรียงลำดับข้อทดสอบจากง่ายสุดไปยากสุด ทำให้ง่ายในการพิจารณา แล้วยังมีข้อมูลจากการวิเคราะห์ข้อสอบที่บอกถึงจำนวนผู้ตอบในแต่ละกลุ่มด้วย ทำให้ได้ข้อมูลมาช่วยพิจารณาจุดตัดได้เหมาะสมมากยิ่งขึ้น*"

ผู้ตัดสิน D

4.6 ความคิดเห็นของผู้ตัดสินเกี่ยวกับความง่ายในการปฏิบัติระหว่างวิธีการแองกอฟที่ได้รับการปรับปรุง และวิธีการบุ๊คมาร์ค

ผู้ตัดสินทั้ง 12 คนมีความคิดเห็นตรงกันว่า วิธีการบุ๊คมาร์คเป็นวิธีที่ง่ายในการปฏิบัติ พราะ ใช้เวลาในการกำหนดคะแนนจุดตัดไม่มากเนื่องจากมีการจัดเรียงข้อสอบจากข้อสอบง่ายสุดไปยังข้อสอบที่ยากสุด และแต่ละหน้ามีข้อสอบให้พิจารณาประกอบพร้อมบอกถึงระดับความสามารถของนักเรียนที่จะตอบข้อสอบข้อนั้นถูก และ ตารางแสดงเปอร์เซ็นต์การตอบของนักเรียนแต่ละตัวเลือก ซึ่งวิธีนี้ครูสามารถนำไปใช้ในโรงเรียนได้

“บุ๊คมาร์ค ทำง่ายใช้เวลาน้อย ครูสามารถนำไปใช้ได้ในโรงเรียน กลุ่มโรงเรียนได้”

ผู้ตัดสิน H

“สำหรับพี่คิดว่าวิธีบุ๊คมาร์ค เพราะ เราสามารถพิจารณาได้จากตัวข้อสอบ และดูจากค่าคะแนนความยากง่าย และมีการเรียงลำดับข้อสอบจากง่ายไปยาก สะดวกในการพิจารณา”

ผู้ตัดสิน I

“วิธีบุ๊คมาร์ค ผมว่าทำได้สะดวก รวดเร็ว มีข้อมูลให้พิจารณารวมอยู่ในหน้าเดียวกัน เช่น มีระดับความสามารถของนักเรียนที่จะตอบถูกในข้อนั้น มีตารางแสดงเปอร์เซ็นต์การตอบของเด็กทำให้ง่ายในการพิจารณา”

ผู้ตัดสิน G

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตอนที่ 5 ผลการเปรียบเทียบความเที่ยงของการกำหนดคะแนนจุดตัดด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และวิธีบุ๊คมาร์ค

ผู้วิจัยทำการเปรียบเทียบค่าสัมประสิทธิ์ความเที่ยงจากค่าสัมประสิทธิ์การสรุปอ้างอิง (Generalizability coefficient) ที่ผู้ตัดสินได้ดำเนินการตัดสินหาคะแนนจุดตัดด้วยวิธีแองกอฟที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์ค จำนวน 7 ระดับ คือ ระดับดีเยี่ยม (A) ระดับดีมาก (B+) ระดับดี (B) ระดับดีพอใช้ (C+) ระดับพอใช้ (C) ระดับอ่อน (D+) ระดับอ่อนมาก (D) โดยแต่ละระดับทำการเปรียบเทียบจำแนกตามจำนวนผู้ตัดสิน รายละเอียดดังต่อไปนี้

ตารางที่ 64 การเปรียบเทียบค่าความเที่ยงตามทฤษฎีสรุปอ้างอิง ระหว่างวิธีแองกอฟที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์คที่ระดับดีเยี่ยม (A)

จำนวนผู้ตัดสิน	**ครั้งที่**	**ค่าความเที่ยง (แองกอฟฯ และ บุ๊คมาร์ค ตามลำดับ)**	**สถิติทดสอบ UX_1**
8	1	0.8496 0.9954	22.79661
	2	0.8561 0.9973	29.05213
	3	0.8534 0.9959	28.41437
12	1	0.8944 0.9970	23.35959
	2	0.8992 0.9973	26.54327
	3	0.8972 0.9972	28.60733

χ^2 (.01 , df=1)=6.635

ตารางที่ 64 (ต่อ)

จำนวนผู้ตัดสิน	ครั้งที่	ค่าความเที่ยง (แองกอฟฯ และ บุ๊คมาร์ค ตามลำดับ)	สถิติทดสอบ UX_1
14	1	0.9081 0.9974	23.39095
	2	0.9124 0.9977	26.6365
	3	0.9106 0.9976	28.71231
16	1	0.9187 0.9977	23.39131
	2	0.9225 0.9980	26.7229
	3	0.9209 0.9979	23.62289
18	1	0.9271 0.9980	23.62289
	2	0.9305 0.9982	26.80898
	3	0.9291 0.9982	29.11218
20	1	0.9339 0.9982	23.67879
	2	0.9370 0.9984	26.81869
	3	0.9357 0.9983	28.82249

χ^2 (.01 , df=1)=6.635

ตารางที่ 64 (ต่อ)

จำนวนผู้ตัดสิน	ครั้งที่	ค่าความเที่ยง (แองกอฟฯ และ บุ๊คมาร์ค ตามลำดับ)	สถิติทดสอบ UX_1
22	1	0.9395 0.9984	23.89775
	2	0.9424 0.9985	26.87538
	3	0.9412 0.9985	29.07805
24	1	0.9443 0.9985	23.76226
	2	0.9469 0.9987	26.89658
	3	0.9458 0.9986	28.98926

χ^2 (.01 , df=1)=6.635

จากตารางที่ 64 ค่าความเที่ยงตามทฤษฎีสรุปอ้างอิงระหว่างวิธีแองกอฟที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์คที่ระดับดีเยี่ยม (A) เมื่อทำการตัดสินครั้งที่ 1 ที่มีจำนวนผู้ตัดสิน 8 12 14 16 18 20 22 24 พบว่า มีค่าแตกต่างกันที่ระดับนัยสำคัญ .01 (UX_1 = 22.79661, 23.35959, 23.39095, 23.62289, 23.39131, 23.67879, 23.89775, 23.76226 ตามลำดับ)

ค่าความเที่ยงตามทฤษฎีสรุปอ้างอิงระหว่างวิธีแองกอฟที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์คที่ระดับดีเยี่ยม (A) เมื่อทำการตัดสินครั้งที่ 2 ที่มีจำนวนผู้ตัดสิน 8 12 14 16 18 20 22 24 พบว่า มีค่าแตกต่างกันที่ระดับนัยสำคัญ .01 (UX_1 = 29.05213, 26.54327, 26.6365, 26.7229, 23.62289, 26.81869, 26.87538, 26.89658 ตามลำดับ)

ค่าความเที่ยงตามทฤษฎีสรุปอ้างอิงระหว่างวิธีแองกอฟที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์คที่ระดับดีเยี่ยม (A) เมื่อทำการตัดสินครั้งที่ 3 ที่มีจำนวนผู้ตัดสิน 8 12 14 16 18 20 22 24 พบว่า มีค่าแตกต่างกันที่ระดับนัยสำคัญ .01 (UX_1 = 28.41437, 28.60733, 28.71231, 28.79257, 29.11218, 28.82249, 29.07805, 28.98926 ตามลำดับ)

ตารางที่ 65 การเปรียบเทียบค่าความเที่ยงตามทฤษฎีสรุปอ้างอิง ระหว่างวิธีแองกอฟที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์คที่ระดับดีมาก (B+)

จำนวนผู้ตัดสิน	**ครั้งที่**	**ค่าความเที่ยง (แองกอฟฯ และ บุ๊คมาร์ค ตามลำดับ)**	**สถิติทดสอบ UX_1**
8	1	0.8751 0.9989	33.22276
	2	0.8637 0.9902	18.09774
	3	0.8586 0.9912	15.63891
12	1	0.9131 0.9992	33.40060
	2	0.9048 0.9935	16.08943
	3	0.9011 0.9942	16.01788
14	1	0.9246 0.9994	33.47772
	2	0.9173 0.9944	16.19757
	3	0.9140 0.9950	16.12424
16	1	0.9334 0.9994	33.51537
	2	0.9269 0.9951	16.28159
	3	0.9239 0.9956	16.19871

χ^2 (.01 , df=1)=6.635

ตารางที่ 65 (ต่อ)

จำนวนผู้ตัดสิน	ครั้งที่	ค่าความเที่ยง (แองกอฟฯ และ บุ๊คมาร์ค ตามลำดับ)	สถิติทดสอบ UX_1
18	1	0.9404 0.9995	33.62254
	2	0.9344 0.9956	16.33892
	3	0.9318 0.9961	16.26973
20	1	0.9460 0.9996	33.65357
	2	0.9406 0.9961	16.39891
	3	0.9382 0.9965	16.32292
22	1	0.9507 0.9996	33.66815
	2	0.9457 0.9964	16.45038
	3	0.9435 0.9968	16.36889
24	1	0.9546 0.9996	33.63181
	2	0.9500 0.9967	16.46879
	3	0.9480 0.9971	16.38941

χ^2 (.01 , df=1)=6.635

จากตารางที่ 65 ค่าความเที่ยงตามทฤษฎีสรุปอ้างอิงระหว่างวิธีแองกอฟที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์ค ที่ระดับดีมาก (B+) เมื่อทำการตัดสินครั้งที่ 1 ที่มีจำนวนผู้ตัดสิน 8 12 14 16 18 20 22 24 พบว่า มีค่าแตกต่างกันที่ระดับนัยสำคัญ .01 (UX_1 = 33.22276, 33.40060, 33.47772, 33.51537, 33.62254, 33.65357, 33.66815, 33.63181 ตามลำดับ)

ค่าความเที่ยงตามทฤษฎีสรุปอ้างอิงระหว่างวิธีแองกอฟที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์คที่ระดับดีมาก (B+) เมื่อทำการตัดสินครั้งที่ 2 ที่มีจำนวนผู้ตัดสิน 8 12 14 16 18 20 22 24 พบว่า มีค่าแตกต่างกันที่ระดับนัยสำคัญ .01 (UX_1 = 18.09774, 16.08943, 16.19757, 16.28159, 16.33892, 16.39891, 16.45038, 16.46879 ตามลำดับ)

ค่าความเที่ยงตามทฤษฎีสรุปอ้างอิงระหว่างวิธีแองกอฟที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์ค ที่ระดับดีมาก (B+) เมื่อทำการตัดสินครั้งที่ 3 ที่มีจำนวนผู้ตัดสิน 8 12 14 16 18 20 22 24 พบว่า มีค่าแตกต่างกันที่ระดับนัยสำคัญ .01 (UX_1 = 15.63891, 16.01788, 16.12424, 16.19871, 16.26973, 16.32292, 16.36889, 16.38941 ตามลำดับ)

ตารางที่ 66 การเปรียบเทียบค่าความเที่ยงตามทฤษฎีสรุปอ้างอิง ระหว่างวิธีแองกอฟที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์คที่ระดับดี (B)

จำนวนผู้ตัดสิน	**ครั้งที่**	**ค่าความเที่ยง (แองกอฟฯ และ บุ๊คมาร์ค ตามลำดับ)**	**สถิติทดสอบ UX_1**
8	1	0.8297 0.9980	27.95017
	2	0.6176 0.9972	30.59450
	3	0.8591 0.9963	22.52072
12	1	0.8797 0.9987	28.37400
	2	0.7079 0.9981	31.34988
	3	0.9014 0.9976	22.87522

χ^2 (.01 , df=1)=6.635

ตารางที่ 66 (ต่อ)

จำนวนผู้ตัดสิน	**ครั้งที่**	**ค่าความเที่ยง (แองกอฟฯ และ บุ๊คมาร์ค ตามลำดับ)**	**สถิติทดสอบ UX_1**
14	1	0.8950 0.9989	28.48869
	2	0.7387 0.9984	31.55378
	3	0.9143 0.9979	22.98398
16	1	0.9069 0.9990	28.56170
	2	0.7636 0.9986	31.72483
	3	0.9242 0.9982	23.05796
18	1	0.9164 0.9991	28.62481
	2	0.7842 0.9988	31.87831
	3	0.9320 0.9984	23.10496
20	1	0.9241 0.9992	28.69160
	2	0.8015 0.9989	32.01516
	3	0.9384 0.9985	23.13959

χ^2 (.01 , df=1)=6.635

ตารางที่ 66 (ต่อ)

จำนวนผู้ตัดสิน	ครั้งที่	ค่าความเที่ยง (แองกอฟฯ และ บุ๊คมาร์ค ตามลำดับ)	สถิติทดสอบ UX_1
22	1	0.9306 0.9993	28.71806
	2	0.8162 0.9990	32.10011
	3	0.9437 0.9987	23.21570
24	1	0.9360 0.9993	28.75385
	2	0.8290 0.9991	32.15198
	3	0.9481 0.9988	23.24480

χ^2 (.01 , df=1)=6.635

จากตารางที่ 66 ค่าความเที่ยงตามทฤษฎีสรุปอ้างอิงระหว่างวิธีแองกอฟที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์คที่ระดับดี (B) เมื่อทำการตัดสินครั้งที่ 1 ที่มีจำนวนผู้ตัดสิน 8 12 14 16 18 20 22 24 พบว่า มีค่าแตกต่างกันที่ระดับนัยสำคัญ .01 (UX_1 = 27.95017, 28.37400, 28.48869, 28.56170, 28.62481, 28.69160, 28.71806, 28.75385 ตามลำดับ)

ค่าความเที่ยงตามทฤษฎีสรุปอ้างอิงระหว่างวิธีแองกอฟที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์คที่ระดับดี (B) เมื่อทำการตัดสินครั้งที่ 2 ที่มีจำนวนผู้ตัดสิน 8 12 14 16 18 20 22 24 พบว่า มีค่าแตกต่างกันที่ระดับนัยสำคัญ .01 (UX_1 = 30.59450, 31.34988, 31.55378, 31.72483, 31.87831, 32.01516, 32.10011, 32.15198 ตามลำดับ)

ค่าความเที่ยงตามทฤษฎีสรุปอ้างอิงระหว่างวิธีแองกอฟที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์คที่ระดับดี (B) เมื่อทำการตัดสินครั้งที่ 3 ที่มีจำนวนผู้ตัดสิน 8 12 14 16 18 20 22 24 พบว่า มีค่าแตกต่างกันที่ระดับนัยสำคัญ .01 (UX_1 = 22.52072, 22.87522, 22.98398, 23.05796, 23.10496, 23.13959, 23.21570, 23.24480 ตามลำดับ)

ตารางที่ 67 การเปรียบเทียบค่าความเที่ยงตามทฤษฎีสรุปอ้างอิง ระหว่างวิธีแองกอฟที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์คที่ระดับดีพอใช้ (C+)

จำนวนผู้ตัดสิน	**ครั้งที่**	**ค่าความเที่ยง (แองกอฟฯ และ บุ๊คมาร์ค ตามลำดับ)**	**สถิติทดสอบ UX_1**
8	1	0.8660 0.9935	19.16168
	2	0.8635 0.9970	23.42162
	3	0.8607 0.9973	24.57465
12	1	0.9065 0.9957	19.51368
	2	0.9047 0.9980	23.75326
	3	0.9026 0.9981	24.89375
14	1	0.9188 0.9963	19.62464
	2	0.9171 0.9983	23.83877
	3	0.9153 0.9985	24.99723
16	1	0.9282 0.9967	19.69646
	2	0.9267 0.9985	23.90040
	3	0.9251 0.9986	25.05203

χ^2 (.01 , df=1)=6.635

ตารางที่ 67 (ต่อ)

จำนวนผู้ตัดสิน	**ครั้งที่**	**ค่าความเที่ยง (แองกอฟฯ และ บุ๊คมาร์ค ตามลำดับ)**	**สถิติทดสอบ UX_1**
18	1	0.9357 0.9971	19.75538
	2	0.9343 0.9987	23.98242
	3	0.9329 0.9988	25.10269
20	1	0.9417 0.9974	19.80820
	2	0.9405 0.9988	24.01645
	3	0.9392 0.9989	25.14332
22	1	0.9467 0.9976	19.86209
	2	0.9456 0.9989	24.06689
	3	0.9444 0.9990	25.18812
24	1	0.9510 0.9978	19.90653
	2	0.9499 0.9990	24.09797
	3	0.9488 0.9991	25.20072

χ^2 (.01 , df=1)=6.635

จากตารางที่ 67 ค่าความเที่ยงตามทฤษฎีสรุปอ้างอิงระหว่างวิธีแองกอฟที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์ค ที่ระดับดีพอใช้ (C+) เมื่อทำการตัดสินครั้งที่ 1 ที่มีจำนวนผู้ตัดสิน 8 12 14 16 18 20 22 24 พบว่า มีค่าแตกต่างกันที่ระดับนัยสำคัญ .01 (UX_1 = 19.16168, 19.51368, 19.62464, 19.69646, 19.75538, 19.80820, 19.86209, 19.90653 ตามลำดับ)

ค่าความเที่ยงตามทฤษฎีสรุปอ้างอิงระหว่างวิธีแองกอฟที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์ค ที่ระดับดีพอใช้ (C+) เมื่อทำการตัดสินครั้งที่ 2 ที่มีจำนวนผู้ตัดสิน 8 12 14 16 18 20 22 24 พบว่า มีค่าแตกต่างกันที่ระดับนัยสำคัญ .01 (UX_1 = 23.42162, 23.75326, 23.83877, 23.90040, 23.98242, 24.01645, 24.06689, 24.09797 ตามลำดับ)

ค่าความเที่ยงตามทฤษฎีสรุปอ้างอิงระหว่างวิธีแองกอฟที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์ค ที่ระดับดีพอใช้ (C+) เมื่อทำการตัดสินครั้งที่ 3 ที่มีจำนวนผู้ตัดสิน 8 12 14 16 18 20 22 24 พบว่า มีค่าแตกต่างกันที่ระดับนัยสำคัญ .01 (UX_1 = 24.57465, 24.89375, 24.99723, 25.05203, 25.10269, 25.14332, 25.18812, 25.20072 ตามลำดับ)

ตารางที่ 68 การเปรียบเทียบค่าความเที่ยงตามทฤษฎีสรุปอ้างอิง ระหว่างวิธีแองกอฟที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์คที่ระดับพอใช้ (C)

จำนวนผู้ตัดสิน	**ครั้งที่**	**ค่าความเที่ยง (แองกอฟฯ และ บุ๊คมาร์ค ตามลำดับ)**	**สถิติทดสอบ UX_1**
8	1	0.8902 1.0000	10.97586
	2	0.8789 0.9963	21.03879
	3	0.8590 0.9994	33.66537
12	1	0.9240 1.0000	13.84282
	2	0.9159 0.9976	21.33975
	3	0.9014 0.9996	33.89091

χ^2 (.01 , df=1)=6.635

ตารางที่ 68 (ต่อ)

จำนวนผู้ตัดสิน	**ครั้งที่**	**ค่าความเที่ยง (แองกอฟฯ และ บุ๊คมาร์ค ตามลำดับ)**	**สถิติทดสอบ UX_1**
14	1	0.9342 1.0000	15.00840
	2	0.9270 0.9979	21.42992
	3	0.9143 0.9997	34.03869
16	1	0.9419 1.0000	16.02449
	2	0.9355 0.9982	21.48925
	3	0.9242 0.9997	34.08859
18	1	0.9480 1.0000	16.9227
	2	0.9423 0.9983	21.52275
	3	0.9320 0.9998	34.10658
20	1	0.9530 1.0000	17.72447
	2	0.9478 0.9985	21.5966
	3	0.9384 0.9998	34.03677

χ^2 (.01 , df=1)=6.635

ตารางที่ 68 (ต่อ)

จำนวนผู้ตัดสิน	ครั้งที่	ค่าความเที่ยง (แองกอฟฯ และ บุ๊คมาร์ค ตามลำดับ)	สถิติทดสอบ UX_1
22	1	0.9571 1.0000	18.44748
	2	0.9523 0.9987	21.61083
	3	0.9437 0.9998	34.04304
24	1	0.96051 1.00000	19.10527
	2	0.9561 0.9988	21.63082
	3	0.9481 0.9998	34.12477

χ^2 (.01 , df=1)=6.635

จากตารางที่ 68 ค่าความเที่ยงตามทฤษฎีสรุปอ้างอิงระหว่างวิธีแองกอฟที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์คที่ระดับพอใช้ (C) เมื่อทำการตัดสินครั้งที่ 1 ที่มีจำนวนผู้ตัดสิน 8 12 14 16 18 20 22 24 พบว่า มีค่าแตกต่างกันที่ระดับนัยสำคัญ .01 (UX_1 = 10.97586, 13.84282, 15.00840, 16.02449, 16.9227, 17.72447, 18.44748, 19.10527 ตามลำดับ)

ค่าความเที่ยงตามทฤษฎีสรุปอ้างอิงระหว่างวิธีแองกอฟที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์คที่ระดับพอใช้ (C) เมื่อทำการตัดสินครั้งที่ 2 ที่มีจำนวนผู้ตัดสิน 8 12 14 16 18 20 22 24 พบว่า มีค่าแตกต่างกันที่ระดับนัยสำคัญ .01 (UX_1 = 21.03879, 21.33975, 21.42992, 21.48925, 21.52275, 21.5966, 21.61083, 21.63082 ตามลำดับ)

ค่าความเที่ยงตามทฤษฎีสรุปอ้างอิงระหว่างวิธีแองกอฟที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์คที่ระดับพอใช้ (C) เมื่อทำการตัดสินครั้งที่ 3 ที่มีจำนวนผู้ตัดสิน 8 12 14 16 18 20 22 24 พบว่า มีค่าแตกต่างกันที่ระดับนัยสำคัญ .01 (UX_1 = 33.66537, 33.89091, 21.42992, 21.48925, 34.10658, 34.03677, 34.04304, 34.12477 ตามลำดับ)

ตารางที่ 69 การเปรียบเทียบค่าความเที่ยงตามทฤษฎีสรุปอ้างอิง ระหว่างวิธีแองกอฟที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์คที่ระดับอ่อน (D+)

จำนวนผู้ตัดสิน	**ครั้งที่**	**ค่าความเที่ยง (แองกอฟฯ และ บุ๊คมาร์ค ตามลำดับ)**	**สถิติทดสอบ UX_1**
8	1	0.9039 0.9980	23.56842
	2	0.9023 0.9980	24.12568
	3	0.8919 0.9971	22.50773
12	1	0.9338 0.9986	23.80037
	2	0.9327 0.9986	24.36024
	3	0.9252 0.9981	22.75383
14	1	0.9427 0.9988	23.84074
	2	0.9417 0.9988	24.40334
	3	0.9352 0.9984	22.85853
16	1	0.9495 0.9990	23.91830
	2	0.9486 0.9990	24.48061
	3	0.9429 0.9986	22.89326

χ^2 (.01 , df=1)=6.635

ตารางที่ 69 (ต่อ)

จำนวนผู้ตัดสิน	**ครั้งที่**	**ค่าความเที่ยง (แองกอฟฯ และ บุ๊คมาร์ค ตามลำดับ)**	**สถิติทดสอบ UX_1**
18	1	0.9549 0.9991	23.93127
	2	0.9541 0.9991	24.49423
	3	0.9489 0.9987	22.94076
20	1	0.9592 0.9992	23.95403
	2	0.9585 0.9991	24.51807
	3	0.9538 0.9989	22.9905
22	1	0.9628 0.9993	24.03369
	2	0.9621 0.9993	24.59628
	3	0.9578 0.9990	22.98891
24	1	0.9658 0.9993	24.03735
	2	0.9652 0.9993	24.60058
	3	0.9612 0.9990	23.03405

χ^2 (.01 , df=1)=6.635

จากตารางที่ 69 ค่าความเที่ยงตามทฤษฎีสรุปอ้างอิงระหว่างวิธีแองกอฟที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์คที่ระดับอ่อน (D+) เมื่อทำการตัดสินครั้งที่ 1 ที่มีจำนวนผู้ตัดสิน 8 12 14 16 18 20 22 24 พบว่า มีค่าแตกต่างกันที่ระดับนัยสำคัญ .01 (UX_1 = 23.56842, 23.80037, 23.84074, 23.91830, 23.93127, 23.95403, 24.03369, 24.03735 ตามลำดับ)

ค่าความเที่ยงตามทฤษฎีสรุปอ้างอิงระหว่างวิธีแองกอฟที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์ค ที่ระดับ อ่อน (D+) เมื่อทำการตัดสินครั้งที่ 2 ที่มีจำนวนผู้ตัดสิน 8 12 14 16 18 20 22 24 พบว่า มีค่าแตกต่างกันที่ระดับนัยสำคัญ .01 (UX_1 = 24.12568, 24.36024, 24.40334, 24.48061, 24.49423, 24.51807, 24.59628, 24.60058 ตามลำดับ)

ค่าความเที่ยงตามทฤษฎีสรุปอ้างอิงระหว่างวิธีแองกอฟที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์คที่ระดับอ่อน (D+) เมื่อทำการตัดสินครั้งที่ 3 ที่มีจำนวนผู้ตัดสิน 8 12 14 16 18 20 22 24 พบว่า มีค่าแตกต่างกันที่ระดับนัยสำคัญ .01 (UX_1 = 22.50773, 22.75383, 22.85853, 22.89326, 22.94076, 22.9905, 22.98891, 23.03405 ตามลำดับ)

ตารางที่ 70 การเปรียบเทียบค่าความเที่ยงตามทฤษฎีสรุปอ้างอิง ระหว่างวิธีแองกอฟที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์คที่ระดับอ่อนมาก(D)

จำนวนผู้ตัดสิน	**ครั้งที่**	**ค่าความเที่ยง (แองกอฟฯ และ บุ๊คมาร์ค ตามลำดับ)**	**สถิติทดสอบ UX_1**
8	1	0.9382 0.9949	13.31307
	2	0.8875 0.9949	17.79302
	3	0.8808 0.9937	17.16031
12	1	0.9575 0.9966	13.53807
	2	0.9221 0.9966	18.07182
	3	0.9172 0.9958	17.4699

χ^2 (.01 , df=1)=6.635

ตารางที่ 70 (ต่อ)

จำนวนผู้ตัดสิน	ครั้งที่	ค่าความเที่ยง (แองกอฟฯ และ บุ๊คมาร์ค ตามลำดับ)	สถิติทดสอบ UX_1
14	1	0.9637 0.9971	13.51648
	2	0.9325 0.9971	18.16787
	3	0.9282 0.9964	17.56269
16	1	0.9681 0.9974	13.53777
	2	0.9404 0.9974	18.21931
	3	0.9366 0.9968	17.63324
18	1	0.9716 0.9977	13.55159
	2	0.9467 0.9972	16.68287
	3	0.9433 0.9972	17.68646
20	1	0.9743 0.9980	13.58351
	2	0.9517 0.9980	18.30544
	3	0.9486 0.9975	17.72982

χ^2 (.01 , df=1)=6.635

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 70 (ต่อ)

จำนวนผู้ตัดสิน	ครั้งที่	ค่าความเที่ยง (แองกอฟฯ และ บุ๊คมาร์ค ตามลำดับ)	สถิติทดสอบ UX_1
22	1	0.9766 0.9981	13.6186
	2	0.9559 0.9981	18.35607
	3	0.9531 0.9977	17.76781
24	1	0.9785 0.9983	13.60988
	2	0.9595 0.9983	18.36073
	3	0.9568 0.9979	17.79329

χ^2 (.01 , df=1)=6.635

จากตารางที่ 70 ค่าความเที่ยงตามทฤษฎีสรุปอ้างอิงระหว่างวิธีแองกอฟที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์คที่ระดับอ่อนมาก(D) เมื่อทำการตัดสินครั้งที่ 1 ที่มีจำนวนผู้ตัดสิน 8 12 14 16 18 20 22 24 พบว่า มีค่าแตกต่างกันที่ระดับนัยสำคัญ .01 (UX_1 = 13.31307, 13.53807, 13.51648, 13.53777, 13.55159, 13.58351, 13.6186, 13.60988 ตามลำดับ)

ค่าความเที่ยงตามทฤษฎีสรุปอ้างอิงระหว่างวิธีแองกอฟที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์คที่ระดับอ่อนมาก(D)เมื่อทำการตัดสินครั้งที่ 2 ที่มีจำนวนผู้ตัดสิน 8 12 14 16 18 20 22 24 พบว่า มีค่าแตกต่างกันที่ระดับนัยสำคัญ .01 (UX_1 = 17.79302, 18.07182, 18.16787, 18.21931, 16.68287, 18.30544, 18.35607, 18.36073 ตามลำดับ)

ค่าความเที่ยงตามทฤษฎีสรุปอ้างอิงระหว่างวิธีแองกอฟที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์คที่ระดับอ่อนมาก(D) เมื่อทำการตัดสินครั้งที่ 3 ที่มีจำนวนผู้ตัดสิน 8 12 14 16 18 20 22 24 พบว่า มีค่าแตกต่างกันที่ระดับนัยสำคัญ .01 (UX_1 = 17.16031, 17.4699, 17.56269, 17.63324, 17.68646, 17.72982, 17.76781, 17.79329 ตามลำดับ)

ตอนที่ 6 ผลของจำนวนผู้ตัดสิน และจำนวนครั้งที่มีต่อการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และวิธีการบุ๊คมาร์ค

ผู้วิจัยได้ศึกษาผลของจำนวนผู้ตัดสิน และจำนวนครั้ง ที่มีต่อความเที่ยงในการกำหนดมาตรฐานด้วยวิธีการแองกอฟที่ได้รับการปรับปรุง และวิธีการบุ๊คมาร์ค ผู้วิจัยออกแบบการวัดเป็นแบบ Two-facet crossed design ฟาเซท (facets) ของการวัด มี 2 ฟาเซท คือ ผู้ตัดสิน (r) กำหนดจำนวนระดับเป็น 8, 12, 14, 16, 18, 20, 22, 24 คน และจำนวนครั้ง (o) กำหนดจำนวนระดับเป็น 1, 2, 3 ครั้ง รูปแบบของความสัมพันธ์ระหว่างฟาเซทเป็น ixrxo นั้นคือ ให้ผู้ตัดสินทุกคนตัดสินข้อสอบทุกข้อ และทำการตัดสินซ้ำ 3 ครั้ง ในการกำหนดมาตรฐานทั้งสองวิธี

1. ผลของจำนวนผู้ตัดสิน และจำนวนครั้งที่มีต่อการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง

ตาราง 71 การวิเคราะห์ความแปรปรวนและการประมาณค่าความแปรปรวน ในขั้นศึกษา G-study โดยทำการกำหนดมาตรฐานวิธีแองกอฟที่ได้รับการปรับปรุง ที่คะแนนจุดตัดระดับดีเยี่ยม (A)

Source of variance	DF	SS	MS	Estimated Variance component	% of total variance
I	99	254093.3611	2566.59961	67.2526	17.4811
O	2	61844.77722	30922.38861	8.7747	2.2808
R	11	72642.13889	6603.83081	0.0000	0.0000
IO	198	36456.50056	184.12374	7.0474	1.8318
IR	1089	66361.80556	60.9383	0.0000	0.0000
OR	22	446780.9494	20308.22497	202.0867	52.5287
Residual (IOR,e)	2178	216831.7728	99.55545	99.5555	25.8776
Total	3599	1155011.306		384.7168	100

เมื่อ I แทน ข้อสอบ
R แทน ผู้ตัดสิน
O แทน จำนวนครั้งที่ทำการตัดสิน
Residual (IOR,e) ส่วนที่เหลือ
IO แทน ข้อสอบและจำนวนครั้ง
IR แทน ข้อสอบและผู้ตัดสิน
OR แทน จำนวนครั้งและผู้ตัดสิน

จากตารางที่ 71 พบว่า การกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง ค่าความแปรปรวนของแหล่งต่างๆ เมื่อเทียบกับความแปรปรวนรวมมีค่าเรียงจากมากไปหาน้อยดังนี้ ความแปรปรวนของจำนวนครั้งและผู้ตัดสินมีค่า 202.0867 คิดเป็นร้อยละ 52.5287 รองลงมาคือ ความแปรปรวนของส่วนที่เหลือมีค่า 99.5555 คิดเป็นร้อยละ 25.8776 รองลงมาคือ ความแปรปรวนของข้อสอบ มีค่า 67.2526 คิดเป็นร้อยละ 17.4811 ส่วนความแปรปรวนของผู้ตัดสินและความแปรปรวนของข้อสอบและผู้ตัดสินมีค่าน้อยที่สุดคือ 0.0000

ตาราง 72 การเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิง (Genaralizability coefficient) ในขั้นศึกษา D-study โดยทำการกำหนดมาตรฐานวิธีแองกอฟที่ได้รับการปรับปรุง มีรูปแบบ i x r x o design ที่คะแนนจุดตัดระดับดีเยี่ยม (A)

Alternative Estimated D studies Design Variance Component								
n_o	1	1	1	1	1	1	1	1
n_r	8	12	14	16	18	20	22	24
$E\rho^2$	0.7724	0.8118	0.8238	0.8330	0.8403	0.8462	0.8512	0.8553
ϕ	0.5529	0.6177	0.6391	0.6562	0.6701	0.6817	0.6914	0.6998
n_o	2	2	2	2	2	2	2	2
n_r	8	12	14	16	18	20	22	24
$E\rho^2$	0.8716	0.8961	0.9034	0.9089	0.9132	0.9167	0.9196	0.9220
ϕ	0.7120	0.7637	0.7798	0.7924	0.8025	0.8107	0.8176	0.8234
n_o	3	3	3	3	3	3	3	3
n_r	8	12	14	16	18	20	22	24
$E\rho^2$	0.9106	0.9283	0.9334	0.9374	0.9404	0.9429	0.9449	0.9466
ϕ	0.7877	0.8290	0.8416	0.8513	0.8590	0.8653	0.8705	0.8749

จากตารางที่ 72 พบว่า การเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิงมีรูปแบบ i x r x o design ที่คะแนนจุดตัดระดับดีเยี่ยม (A) ถ้ากำหนดมาตรฐานวิธีแองกอฟที่ได้รับการปรับปรุง 1 ครั้ง จำนวนผู้ตัดสิน 8, 12, 14, 16, 18, 20, 22 และ 24 คน ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมพันธ์ มีค่าเพิ่มขึ้นเมื่อผู้ตัดสินเพิ่มขึ้น มีค่าดังนี้ 0.7724, 0.8118 0.8238, 0.8330, 0.8403, 0.8462, 0.8512, 0.8553 ตามลำดับ ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมบูรณ์ มีค่าดังนี้ 0.5529, 0.6177, 0.6391, 0.6562, 0.6701, 0.6817, 0.6914, 0.6998 ตามลำดับ

ถ้ากำหนดมาตรฐานวิธีแองกอฟที่ได้รับการปรับปรุง ซ้ำ 2 ครั้ง จำนวนผู้ตัดสิน 8, 12, 14, 16, 18, 20, 22 และ 24 คน ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมพันธ์ มีค่าเพิ่มขึ้นเมื่อผู้ตัดสินเพิ่มขึ้น และเมื่อจำนวนครั้งเพิ่มขึ้น โดยมีค่าดังนี้ 0.8716, 0.8961, 0.9034, 0.9089, 0.9132, 0.9167, 0.9196, 0.9220 ตามลำดับ ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมบูรณ์ มีค่าดังนี้ 0.7120, 0.7637, 0.7798, 0.7924, 0.8025, 0.8107, 0.8176, 0.8234 ตามลำดับ

ถ้ากำหนดมาตรฐานวิธีแองกอฟที่ได้รับการปรับปรุง ซ้ำ 3 ครั้ง จำนวนผู้ตัดสิน 8, 12, 14, 16, 18, 20, 22 และ 24 คน ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมพันธ์ มีค่าเพิ่มขึ้นเมื่อผู้ตัดสินเพิ่มขึ้น และเมื่อจำนวนครั้งเพิ่มขึ้น โดยมีค่าดังนี้ 0.9106, 0.9283, 0.9334, 0.9374, 0.9404, 0.9429, 0.9449, 0.9466 ตามลำดับ ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมบูรณ์ มีค่าดังนี้ 0.7877, 0.8290, 0.8416, 0.8513, 0.8590, 0.8653, 0.8705, 0.8749 ตามลำดับ จะเห็นได้ว่าเมื่อจำนวนผู้ตัดสิน และจำนวนครั้งเพิ่มขึ้น ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมพันธ์ และค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมบูรณ์ก็มีค่าเพิ่มขึ้นตามด้วย

ตาราง 73 การวิเคราะห์ความแปรปรวนและการประมาณค่าความแปรปรวน ในขั้นศึกษา G-study โดยทำการกำหนดมาตรฐานวิธีแองกอฟที่ได้รับการปรับปรุง ที่คะแนนจุดตัดระดับดีมาก (B+)

Source of variance	DF	SS	MS	Estimated Variance component	% of total variance
I	99	223764.9608	2260.2521	60.2828	19.4847
O	2	4990.7272	2495.3636	0.0000	0.0000
R	11	28392.0364	2581.0942	0.0000	0.0000
IO	198	23908.1617	120.7483	2.9906	0.9666
IR	1089	59005.2692	54.1830	0.0000	0.0000
OR	22	356618.5061	16209.9321	161.2507	52.1197
Residual (IOR,e)	2178	184827.9383	84.8613	84.8613	27.4290
Total	3599	881507.5997		309.3854	100

เมื่อ I แทน ข้อสอบ
R แทน ผู้ตัดสิน
O แทน จำนวนครั้งที่ทำการตัดสิน
Residual (IOR,e) ส่วนที่เหลือ
IO แทน ข้อสอบและจำนวนครั้ง
IR แทน ข้อสอบและผู้ตัดสิน
OR แทน จำนวนครั้งและผู้ตัดสิน

จากตารางที่ 73 พบว่า การกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง ค่าความแปรปรวนของแหล่งต่างๆ เมื่อเทียบกับความแปรปรวนรวมมีค่าเรียงจากมากไปหาน้อยดังนี้ ความแปรปรวนของจำนวนครั้งและผู้ตัดสิน มีค่า 161.2507 คิดเป็นร้อยละ 52.1197 รองลงมาคือ ความแปรปรวนของส่วนที่เหลือ มีค่า 84.8613 คิดเป็นร้อยละ 27.4290 รองลงมาคือ ความแปรปรวนของข้อสอบ มีค่า 60.2828 คิดเป็นร้อยละ 19.4847 ส่วนความแปรปรวนของจำนวนครั้ง ความแปรปรวนของผู้ตัดสิน และความแปรปรวนของข้อสอบและผู้ตัดสินมีค่าน้อยที่สุดคือ 0.0000

ตาราง 74 การเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิง (Genaralizability coefficient) ในขั้นศึกษา D-study โดยทำการกำหนดมาตรฐานวิธีแองกอฟที่ได้รับการปรับปรุง มีรูปแบบ i x r x o design ที่คะแนนจุดตัดระดับดีมาก (B+)

Alternative Estimated D studies Design Variance Component								
n_o	1	1	1	1	1	1	1	1
n_r	8	12	14	16	18	20	22	24
$E\rho^2$	0.8138	0.8552	0.8678	0.8775	0.8852	0.8915	0.8967	0.9011
ϕ	0.6378	0.7166	0.7429	0.7639	0.7810	0.7953	0.8074	0.8178
n_o	2	2	2	2	2	2	2	2
n_r	8	12	14	16	18	20	22	24
$E\rho^2$	0.8973	0.9220	0.9292	0.9348	0.9391	0.9426	0.9455	0.9480
ϕ	0.7788	0.8349	0.8525	0.8661	0.8770	0.8860	0.8934	0.8997
n_o	3	3	3	3	3	3	3	3
n_r	8	12	14	16	18	20	22	24
$E\rho^2$	0.9291	0.9466	0.9517	0.9556	0.9586	0.9610	0.9630	0.9647
ϕ	0.8408	0.8835	0.8966	0.9066	0.9145	0.9210	0.9263	0.9309

จากตารางที่ 74 พบว่า การเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิง มีรูปแบบ i x r x o design ที่คะแนนจุดตัดระดับดีมาก (B+) ถ้ากำหนดมาตรฐานวิธีแองกอฟที่ได้รับการปรับปรุง 1 ครั้ง จำนวนผู้ตัดสิน 8, 12, 14, 16, 18, 20, 22 และ 24 คน ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมพันธ์ มีค่าเพิ่มขึ้นเมื่อผู้ตัดสินเพิ่มขึ้น มีค่าดังนี้ 0.8138, 0.8552, 0.8678, 0.8775, 0.8852, 0.8915, 0.8967, 0.9011 ตามลำดับ ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมบูรณ์ มีค่าดังนี้ 0.6378, 0.7166, 0.7429, 0.7639, 0.7810, 0.7953, 0.8074, 0.8178 ตามลำดับ

ถ้ากำหนดมาตรฐานวิธีแองกอฟที่ได้รับการปรับปรุงซ้ำ 2 ครั้ง จำนวนผู้ตัดสิน 8, 12, 14, 16, 18, 20, 22 และ 24 คน ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมพันธ์ มีค่าเพิ่มขึ้นเมื่อผู้ตัดสินเพิ่มขึ้น และเมื่อจำนวนครั้งเพิ่มขึ้น โดยมีค่าดังนี้ 0.8973, 0.9220, 0.9292, 0.9348, 0.9391, 0.9426, 0.9455, 0.9480 ตามลำดับ ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมบูรณ์ มีค่าดังนี้ 0.7788, 0.8349, 0.8525, 0.8661, 0.8770, 0.8860, 0.8934, 0.8997 ตามลำดับ

ถ้ากำหนดมาตรฐานวิธีแองกอฟที่ได้รับการปรับปรุง ซ้ำ 3 ครั้ง จำนวนผู้ตัดสิน 8, 12, 14, 16, 18, 20, 22 และ 24 คน ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมพันธ์ มีค่าเพิ่มขึ้นเมื่อผู้ตัดสินเพิ่มขึ้น และเมื่อจำนวนครั้งเพิ่มขึ้น โดยมีค่าดังนี้ 0.9291, 0.9466, 0.9517, 0.9556, 0.9586, 0.9610, 0.9630, 0.9647 ตามลำดับ ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมบูรณ์ มีค่าดังนี้ 0.8408, 0.8835, 0.8966, 0.9066, 0.9145, 0.9210, 0.9263, 0.9309 ตามลำดับ จะเห็นได้ว่าเมื่อจำนวนผู้ตัดสิน และจำนวนครั้งเพิ่มขึ้น ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมพันธ์ และค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมบูรณ์ก็มีค่าเพิ่มขึ้นตามด้วย

ตาราง 75 การวิเคราะห์ความแปรปรวนและการประมาณค่าความแปรปรวน ในขั้นศึกษา G-study โดยทำการกำหนดมาตรฐานวิธีแองกอฟที่ได้รับการปรับปรุง ที่คะแนนจุดตัดระดับดี (B)

Source of variance	DF	SS	MS	Estimated Variance component	% of total variance
I	99	209659.6319	2117.7741	55.3561	21.6476
O	2	10890.4539	5445.2269	0.0000	0.0000
R	11	14573.7831	1324.8894	0.0000	0.0000
IO	198	30667.9906	154.8888	5.7397	2.2446
IR	1089	61068.0781	56.0772	0.0000	0.0000
OR	22	240826.4261	10946.6557	108.6064	42.4717
Residual (IOR,e)	2178	187335.1294	86.0125	86.0125	33.6361
Total	3599	755021.4931		255.7147	100

เมื่อ I แทน ข้อสอบ
R แทน ผู้ตัดสิน
O แทน จำนวนครั้งที่ทำการตัดสิน
Residual (IOR,e) ส่วนที่เหลือ
IO แทน ข้อสอบและจำนวนครั้ง
IR แทน ข้อสอบและผู้ตัดสิน
OR แทน จำนวนครั้งและผู้ตัดสิน

จากตารางที่ 75 พบว่า การกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง ค่าความแปรปรวนของแหล่งต่างๆ เมื่อเทียบกับความแปรปรวนรวมมีค่าเรียงจากมากไปหาน้อยดังนี้ ความแปรปรวนของจำนวนครั้งและผู้ตัดสิน มีค่า 108.6064 คิดเป็นร้อยละ 42.4717 รองลงมาคือ ความแปรปรวนของส่วนที่เหลือ มีค่า 86.0125 คิดเป็นร้อยละ 33.6361 รองลงมาคือ ความแปรปรวนของข้อสอบ มีค่า 55.3561 คิดเป็นร้อยละ 21.6476 ส่วนความแปรปรวนของจำนวนครั้ง ความแปรปรวนของผู้ตัดสิน และความแปรปรวนของข้อสอบและผู้ตัดสินมีค่าน้อยที่สุดคือ 0.0000

ตาราง 76 การเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิง (Genaralizability coefficient) ในขั้นศึกษา D-study โดยทำการกำหนดมาตรฐานวิธีแองกอฟที่ได้รับการปรับปรุง มีรูปแบบ i x r x o design ที่คะแนนจุดตัดระดับดี (B)

Alternative Estimated D studies Design Variance Component								
n_o	1	1	1	1	1	1	1	1
n_r	8	12	14	16	18	20	22	24
$E\rho^2$	0.7678	0.8086	0.8211	0.8307	0.8383	0.8445	0.8496	0.8540
ϕ	0.6446	0.7129	0.7352	0.7528	0.7671	0.7790	0.7890	0.7975
n_o	2	2	2	2	2	2	2	2
n_r	8	12	14	16	18	20	22	24
$E\rho^2$	0.8686	0.8942	0.9017	0.9075	0.9120	0.9157	0.9187	0.9212
ϕ	0.7839	0.8324	0.8474	0.8590	0.8682	0.8758	0.8820	0.8873
n_o	3	3	3	3	3	3	3	3
n_r	8	12	14	16	18	20	22	24
$E\rho^2$	0.9084	0.9269	0.9323	0.9364	0.9396	0.9422	0.9443	0.9461
ϕ	0.8447	0.8817	0.8928	0.9014	0.9081	0.9136	0.9181	0.9219

จากตารางที่ 76 พบว่า การเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิง มีรูปแบบ i x r x o design ที่คะแนนจุดตัดระดับดี (B) ถ้ากำหนดมาตรฐานวิธีแองกอฟที่ได้รับการปรับปรุง 1 ครั้ง จำนวนผู้ตัดสิน 8, 12, 14, 16, 18, 20, 22 และ 24 คน ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมพันธ์ มีค่าเพิ่มขึ้นเมื่อผู้ตัดสินเพิ่มขึ้น มีค่าดังนี้ 0.7678, 0.8086, 0.8211, 0.8307, 0.8383, 0.8445, 0.8496, 0.8540 ตามลำดับ ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมบูรณ์ มีค่าดังนี้ 0.6446, 0.7129, 0.7352, 0.7528, 0.7671, 0.7790, 0.7890, 0.7975 ตามลำดับ

ถ้ากำหนดมาตรฐานวิธีแองกอฟที่ได้รับการปรับปรุง ซ้ำ 2 ครั้ง จำนวนผู้ตัดสิน 8, 12, 14, 16, 18, 20, 22 และ 24 คน ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมพันธ์ มีค่าเพิ่มขึ้นเมื่อผู้ตัดสินเพิ่มขึ้น และเมื่อจำนวนครั้งเพิ่มขึ้น โดยมีค่าดังนี้ 0.8686, 0.8942, 0.9017, 0.9075, 0.9120, 0.9157, 0.9187, 0.9212 ตามลำดับ ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมบูรณ์ มีค่าดังนี้ 0.7839, 0.8324, 0.8474, 0.8590, 0.8682, 0.8758, 0.8820, 0.8873 ตามลำดับ

ถ้ากำหนดมาตรฐานวิธีแองกอฟที่ได้รับการปรับปรุง ซ้ำ 3 ครั้ง จำนวนผู้ตัดสิน 8, 12, 14, 16, 18, 20, 22 และ 24 คน ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมพันธ์ มีค่าเพิ่มขึ้นเมื่อผู้ตัดสินเพิ่มขึ้น และเมื่อจำนวนครั้งเพิ่มขึ้น โดยมีค่าดังนี้ 0.9084, 0.9269, 0.9323, 0.9364, 0.9396, 0.9422, 0.9443, 0.9461 ตามลำดับ ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมบูรณ์ มีค่าดังนี้ 0.8447, 0.8817, 0.8928, 0.9014, 0.9081, 0.9136, 0.9181, 0.9219 ตามลำดับ จะเห็นได้ว่าเมื่อจำนวนผู้ตัดสิน และจำนวนครั้งเพิ่มขึ้น ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมพันธ์ และค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมบูรณ์ก็มีค่าเพิ่มขึ้นตามด้วย

ตาราง 77 การวิเคราะห์ความแปรปรวนและการประมาณค่าความแปรปรวน ในขั้นศึกษา G-study โดยทำการกำหนดมาตรฐานวิธีแองกอฟที่ได้รับการปรับปรุง ที่คะแนนจุดตัดระดับดีพอใช้ (C+)

Source of variance	DF	SS	MS	Estimated Variance component	% of total variance
I	99	293416.9233	2963.80731	77.6463	25.3625
O	2	3323.435	1661.7175	0.0000	0.0000
R	11	23296.01	2117.81909	0.0000	0.0000
IO	198	42124.73167	212.75117	9.6190	3.1420
IR	1089	57837.99	53.1111	0.0000	0.0000
OR	22	269569.365	12253.15295	121.5583	39.7059
Residual (IOR,e)	2178	211969.135	97.32284	97.3228	31.7896
Total	3599	901537.59		306.1465	100

เมื่อ I แทน ข้อสอบ
R แทน ผู้ตัดสิน
O แทน จำนวนครั้งที่ทำการตัดสิน
Residual (IOR,e) ส่วนที่เหลือ
IO แทน ข้อสอบและจำนวนครั้ง
IR แทน ข้อสอบและผู้ตัดสิน
OR แทน จำนวนครั้งและผู้ตัดสิน

จากตารางที่ 77 พบว่า การกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง ค่าความแปรปรวนของแหล่งต่างๆ เมื่อเทียบกับความแปรปรวนรวมมีค่าเรียงจากมากไปหาน้อยดังนี้ ความแปรปรวนของจำนวนครั้งและผู้ตัดสิน มีค่า 121.5583 คิดเป็นร้อยละ 39.7059 รองลงมาคือ ความแปรปรวนของส่วนที่เหลือมีค่า 97.3228 คิดเป็นร้อยละ 31.7896 รองลงมาคือ ความแปรปรวนของข้อสอบ มีค่า 77.6463 คิดเป็นร้อยละ 25.3625 ส่วนความแปรปรวนของจำนวนครั้ง ความแปรปรวนของผู้ตัดสิน และความแปรปรวนของข้อสอบและผู้ตัดสินมีค่าน้อยที่สุดคือ 0.0000

ตาราง 78 การเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิง (Genaralizability coefficient) ในขั้นศึกษา D-study โดยทำการกำหนดมาตรฐานวิธีแองกอฟที่ได้รับการปรับปรุง มีรูปแบบ i x r x o design ที่คะแนนจุดตัดระดับดีพอใช้ (C+)

Alternative Estimated D studies Design Variance Component								
n_o	1	1	1	1	1	1	1	1
n_r	8	12	14	16	18	20	22	24
$E\rho^2$	0.7782	0.8117	0.8218	0.8296	0.8357	0.8407	0.8448	0.8482
ϕ	0.6739	0.7328	0.7516	0.7664	0.7782	0.7880	0.7961	0.8031
n_o	2	2	2	2	2	2	2	2
n_r	8	12	14	16	18	20	22	24
$E\rho^2$	0.87525	0.89606	0.90218	0.90684	0.91049	0.91343	0.91585	0.91788
ϕ	0.80518	0.84582	0.8582	0.86772	0.87527	0.88141	0.8865	0.89078
n_o	3	3	3	3	3	3	3	3
n_r	8	12	14	16	18	20	22	24
$E\rho^2$	0.9132	0.9282	0.9326	0.9359	0.9385	0.9406	0.9423	0.9437
ϕ	0.8611	0.8917	0.9008	0.9078	0.9132	0.9177	0.9214	0.9244

จากตารางที่ 78 พบว่า การเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิง มีรูปแบบ i x r x o design ที่คะแนนจุดตัดระดับดีพอใช้ (C+) ถ้ากำหนดมาตรฐานวิธีแองกอฟที่ได้รับการปรับปรุง 1 ครั้ง จำนวนผู้ตัดสิน 8, 12, 14, 16, 18, 20, 22 และ 24 คน ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมพันธ์ มีค่าเพิ่มขึ้นเมื่อผู้ตัดสินเพิ่มขึ้น มีค่าดังนี้ 0.7782, 0.8117, 0.8218, 0.8296, 0.8357, 0.8407, 0.8448, 0.8482 ตามลำดับ ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมบูรณ์ มีค่าดังนี้ 0.6739, 0.7328, 0.7516, 0.7664, 0.7782, 0.7880, 0.7961, 0.8031 ตามลำดับ

ถ้ากำหนดมาตรฐานวิธีแองกอฟที่ได้รับการปรับปรุงซ้ำ 2 ครั้ง จำนวนผู้ตัดสิน 8, 12, 14, 16, 18, 20, 22 และ 24 คน ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมพันธ์ มีค่าเพิ่มขึ้นเมื่อผู้ตัดสินเพิ่มขึ้น และเมื่อจำนวนครั้งเพิ่มขึ้น โดยมีค่าดังนี้ 0.87525, 0.89606, 0.90218, 0.90684, 0.91049, 0.91343, 0.91585, 0.91788 ตามลำดับ ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมบูรณ์ มีค่าดังนี้ 0.80518, 0.84582, 0.8582, 0.86772, 0.87527, 0.88141, 0.8865, 0.89078 ตามลำดับ

ถ้ากำหนดมาตรฐานวิธีแองกอฟที่ได้รับการปรับปรุงซ้ำ 3 ครั้ง จำนวนผู้ตัดสิน 8, 12, 14, 16, 18, 20, 22 และ 24 คน ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมพันธ์ มีค่าเพิ่มขึ้นเมื่อผู้ตัดสินเพิ่มขึ้น และเมื่อจำนวนครั้งเพิ่มขึ้น โดยมีค่าดังนี้ 0.9132, 0.9282, 0.9326, 0.9359, 0.9385, 0.9406, 0.9423, 0.9437 ตามลำดับ ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมบูรณ์มีค่าดังนี้ 0.8611, 0.8917, 0.9008, 0.9078, 0.9132, 0.9177, 0.9214, 0.9244 ตามลำดับ จะเห็นได้ว่าเมื่อจำนวนผู้ตัดสิน และจำนวนครั้งเพิ่มขึ้น ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมพันธ์ และค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมบูรณ์ก็มีค่าเพิ่มขึ้นตามด้วย

ตาราง 79 การวิเคราะห์ความแปรปรวนและการประมาณค่าความแปรปรวน ในขั้นศึกษา G-study โดยทำการกำหนดมาตรฐานวิธีแองกอฟที่ได้รับการปรับปรุง ที่คะแนนจุดตัดระดับพอใช้ (C)

Source of variance	DF	SS	MS	Estimated Variance component	% of total variance
I	99	243485.8211	2459.4527	63.9165	24.1629
O	2	19312.6439	9656.3219	0.0000	0.0000
R	11	51757.5722	4705.2338	0.0000	0.0000
IO	198	34740.9672	175.4594	8.7173	3.2955
IR	1089	58644.8722	53.8520	0.0000	0.0000
OR	22	267841.1494	12174.5977	121.0375	45.7569
Residual (IOR,e)	2178	154315.2394	70.8518	70.8518	26.7847
Total	3599	830098.2656		264.5230	100

เมื่อ I แทน ข้อสอบ
R แทน ผู้ตัดสิน
O แทน จำนวนครั้งที่ทำการตัดสิน
IO แทน ข้อสอบและจำนวนครั้ง
IR แทน ข้อสอบและผู้ตัดสิน
OR แทน จำนวนครั้งและผู้ตัดสิน

Residual (IOR,e) ส่วนที่เหลือ

จากตารางที่ 79 พบว่า การกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง ค่าความแปรปรวนของแหล่งต่างๆ เมื่อเทียบกับความแปรปรวนรวมมีค่าเรียงจากมากไปหาน้อยดังนี้ ความแปรปรวนของจำนวนครั้งและผู้ตัดสิน มีค่า 121.0375 คิดเป็นร้อยละ 45.7569 รองลงมาคือ ความแปรปรวนของส่วนที่เหลือ มีค่า 70.8518 คิดเป็นร้อยละ 26.7847 รองลงมาคือ ความแปรปรวนของข้อสอบ มีค่า 63.9165 คิดเป็นร้อยละ 24.1629 ส่วนความแปรปรวนของจำนวนครั้ง ความแปรปรวนของผู้ตัดสิน และความแปรปรวนของข้อสอบและผู้ตัดสินมีค่าน้อยที่สุดคือ 0.0000

ตาราง 80 การเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิง (Genaralizability coefficient) ในขั้นศึกษา D-study โดยทำการกำหนดมาตรฐานวิธีแองกอฟที่ได้รับการปรับปรุง มีรูปแบบ i x r x o design ที่คะแนนจุดตัดระดับพอใช้ (C)

Alternative Estimated D studies Design Variance Component								
n_o	1	1	1	1	1	1	1	1
n_r	8	12	14	16	18	20	22	24
$E\rho^2$	0.7831	0.8127	0.8216	0.8284	0.8337	0.8381	0.8416	0.8446
ϕ	0.6599	0.7197	0.7389	0.7539	0.7660	0.7760	0.7844	0.7915
n_o	2	2	2	2	2	2	2	2
n_r	8	12	14	16	18	20	22	24
$E\rho^2$	0.8784	0.8967	0.9021	0.9061	0.9093	0.9119	0.9140	0.9158
ϕ	0.7951	0.8370	0.8498	0.8597	0.8675	0.8739	0.8792	0.8836
n_o	3	3	3	3	3	3	3	3
n_r	8	12	14	16	18	20	22	24
$E\rho^2$	0.9155	0.9287	0.9325	0.9354	0.9377	0.9395	0.9410	0.9422
ϕ	0.8534	0.8851	0.8946	0.9019	0.9076	0.9122	0.9161	0.9193

จากตารางที่ 80 พบว่า การเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิง มีรูปแบบ i x r x o design ที่คะแนนจุดตัดระดับพอใช้ (C) ถ้ากำหนดมาตรฐานวิธีแองกอฟที่ได้รับการปรับปรุง 1 ครั้ง จำนวนผู้ตัดสิน 8, 12, 14, 16, 18, 20, 22 และ 24 คน ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมพันธ์ มีค่าเพิ่มขึ้นเมื่อผู้ตัดสินเพิ่มขึ้น มีค่าดังนี้ 0.7831, 0.8127, 0.8216, 0.8284, 0.8337,

0.8381, 0.8416, 0.8446 ตามลำดับ ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมบูรณ์ มีค่าดังนี้ 0.6599, 0.7197, 0.7389, 0.7539, 0.7660, 0.7760, 0.7844, 0.7915 ตามลำดับ

ถ้ากำหนดมาตรฐานวิธีแองกอฟที่ได้รับการปรับปรุงซ้ำ 2 ครั้ง จำนวนผู้ตัดสิน 8, 12, 14, 16, 18, 20, 22 และ 24 คน ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมพันธ์ มีค่าเพิ่มขึ้นเมื่อผู้ตัดสินเพิ่มขึ้น และเมื่อจำนวนครั้งเพิ่มขึ้น โดยมีค่าดังนี้ 0.8784, 0.8967, 0.9021, 0.9061, 0.9093, 0.9119, 0.9140, 0.9158 ตามลำดับ ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมบูรณ์ มีค่าดังนี้ 0.7951, 0.8370, 0.8498, 0.8597, 0.8675, 0.8739, 0.8792, 0.8836 ตามลำดับ

ถ้ากำหนดมาตรฐานวิธีแองกอฟที่ได้รับการปรับปรุง ซ้ำ 3 ครั้ง จำนวนผู้ตัดสิน 8, 12, 14, 16, 18, 20, 22 และ 24 คน ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมพันธ์มีค่าเพิ่มขึ้นเมื่อผู้ตัดสินเพิ่มขึ้น และเมื่อจำนวนครั้งเพิ่มขึ้น โดยมีค่าดังนี้ 0.9155, 0.9287, 0.9325, 0.9354, 0.9377, 0.9395, 0.9410, 0.9422 ตามลำดับ ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมบูรณ์ มีค่าดังนี้ 0.8534, 0.8851, 0.8946, 0.9019, 0.9076, 0.9122, 0.9161, 0.9193 ตามลำดับ จะเห็นได้ว่าเมื่อจำนวนผู้ตัดสิน และจำนวนครั้งเพิ่มขึ้น ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมพันธ์ และค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมบูรณ์ก็มีค่าเพิ่มขึ้นตามด้วย

ตาราง 81 การวิเคราะห์ความแปรปรวนและการประมาณค่าความแปรปรวน ในขั้นศึกษา G-study โดยทำการกำหนดมาตรฐานวิธีแองกอฟที่ได้รับการปรับปรุง ที่คะแนนจุดตัดระดับอ่อน (D+)

Source of variance	DF	SS	MS	Estimated Variance component	% of total variance
I	99	274406.0253	2771.7780	72.0052	33.3882
O	2	47512.185	23756.0925	15.2892	7.0895
R	11	78905.1542	7173.1958	6.3649	2.9513
IO	198	37981.5372	191.8260	11.0906	5.1426
IR	1089	50642.8181	46.5040	0.0000	0.0000
OR	22	116071.2083	5275.9640	52.1723	24.1918
Residual (IOR,e)	2178	127933.0694	58.7388	58.7388	27.2366
Total	3599	733451.9975		215.6609	100

เมื่อ I แทน ข้อสอบ
R แทน ผู้ตัดสิน
O แทน จำนวนครั้งที่ทำการตัดสิน
Residual (IOR,e) ส่วนที่เหลือ
IO แทน ข้อสอบและจำนวนครั้ง
IR แทน ข้อสอบและผู้ตัดสิน
OR แทน จำนวนครั้งและผู้ตัดสิน

จากตารางที่ 81 พบว่า การกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง ค่าความแปรปรวนของแหล่งต่างๆ เมื่อเทียบกับความแปรปรวนรวมมีค่าเรียงจากมากไปหาน้อยดังนี้ ความแปรปรวนของข้อสอบมีค่า 72.0052 คิดเป็นร้อยละ 33.3882 รองลงมาคือ ความแปรปรวนของส่วนที่เหลือมีค่า 58.7388 คิดเป็นร้อยละ 27.2366 รองลงมาคือ ความแปรปรวนของจำนวนครั้งและผู้ตัดสิน มีค่า 52.1723 คิดเป็นร้อยละ 24.1918 ส่วนความแปรปรวนของข้อสอบและผู้ตัดสิน มีค่าน้อยที่สุดคือ 0.0000

ตาราง 82 การเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิง (Genaralizability coefficient) ในขั้นศึกษา D-study โดยทำการกำหนดมาตรฐานวิธีแองกอฟที่ได้รับการปรับปรุง มีรูปแบบ i x r x o design ที่คะแนนจุดตัดระดับอ่อน (D+)

Alternative Estimated D studies Design Variance Component								
n_o	1	1	1	1	1	1	1	1
n_r	8	12	14	16	18	20	22	24
$E\rho^2$	0.7954	0.8176	0.8242	0.8292	0.8331	0.8363	0.8389	0.8411
ϕ	0.6359	0.6647	0.6734	0.6801	0.6854	0.6897	0.6933	0.6963
n_o	2	2	2	2	2	2	2	2
n_r	8	12	14	16	18	20	22	24
$E\rho^2$	0.8861	0.8997	0.9036	0.9066	0.9090	0.9109	0.9124	0.9137
ϕ	0.7741	0.7963	0.8028	0.8078	0.8117	0.8149	0.8175	0.8197
n_o	3	3	3	3	3	3	3	3
n_r	8	12	14	16	18	20	22	24
$E\rho^2$	0.9210	0.9308	0.9336	0.9358	0.9374	0.9388	0.9399	0.9408
ϕ	0.8346	0.8525	0.8577	0.8617	0.8649	0.8674	0.8695	0.8712

จากตารางที่ 82 พบว่า การเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิง มีรูปแบบ i x r x o design ที่คะแนนจุดตัดระดับอ่อน (D+) ถ้ากำหนดมาตรฐานวิธีแองกอฟที่ได้รับการปรับปรุง 1 ครั้ง จำนวนผู้ตัดสิน 8, 12, 14, 16, 18, 20, 22 และ 24 คน ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสิน

แบบสัมพันธ์ มีค่าเพิ่มขึ้นเมื่อผู้ตัดสินเพิ่มขึ้น มีค่าดังนี้ 0.7954, 0.8176, 0.8242, 0.8292, 0.8331, 0.8363, 0.8389, 0.8411 ตามลำดับ ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมบูรณ์ มีค่าดังนี้ 0.6359, 0.6647, 0.6734, 0.6801, 0.6854, 0.6897, 0.6933, 0.6963 ตามลำดับ

ถ้ากำหนดมาตรฐานวิธีแองกอฟที่ได้รับการปรับปรุง ซ้ำ 2 ครั้ง จำนวนผู้ตัดสิน 8, 12, 14, 16, 18, 20, 22 และ 24 คน ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมพันธ์ มีค่าเพิ่มขึ้นเมื่อผู้ตัดสินเพิ่มขึ้น และเมื่อจำนวนครั้งเพิ่มขึ้น โดยมีค่าดังนี้ 0.8861, 0.8997, 0.9036, 0.9066, 0.9090, 0.9109, 0.9124, 0.9137 ตามลำดับ ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมบูรณ์ มีค่าดังนี้ 0.7741, 0.7963, 0.8028, 0.8078, 0.8117, 0.8149, 0.8175, 0.8197 ตามลำดับ

ถ้ากำหนดมาตรฐานวิธีแองกอฟที่ได้รับการปรับปรุง ซ้ำ 3 ครั้ง จำนวนผู้ตัดสิน 8, 12, 14, 16, 18, 20, 22 และ 24 คน ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมพันธ์ มีค่าเพิ่มขึ้นเมื่อผู้ตัดสินเพิ่มขึ้น และเมื่อจำนวนครั้งเพิ่มขึ้น โดยมีค่าดังนี้ 0.9210, 0.9308, 0.9336, 0.9358, 0.9374, 0.9388, 0.9399, 0.9408 ตามลำดับ ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมบูรณ์ มีค่าดังนี้ 0.8346, 0.8525, 0.8577, 0.8617, 0.8649, 0.8674, 0.8695, 0.8712 ตามลำดับ จะเห็นได้ว่าเมื่อจำนวนผู้ตัดสิน และจำนวนครั้งเพิ่มขึ้น ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมพันธ์ และค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมบูรณ์ก็มีค่าเพิ่มขึ้นตามด้วย

ตาราง 83 การวิเคราะห์ความแปรปรวนและการประมาณค่าความแปรปรวน ในขั้นศึกษา G-study โดยทำการกำหนดมาตรฐานวิธีแองกอฟที่ได้รับการปรับปรุงที่คะแนนจุดตัดระดับอ่อนมาก (D)

Source of variance	DF	SS	MS	Estimated Variance component	% of total variance
I	99	261260.4456	2638.9944	69.6297	47.3290
O	2	22656.6600	11328.3300	7.9778	5.4227
R	11	7031.4500	639.2227	0.0000	0.0000
IO	198	29043.2844	146.6833	8.4778	5.7625
IR	1089	33315.1611	30.5924	0.0000	0.0000
OR	22	36372.1400	1653.2791	16.0833	10.9322
Residual (IOR,e)	2178	97901.2489	44.9501	44.9501	30.5536
Total	3599	487580.3900		147.1186	100

เมื่อ I แทน ข้อสอบ
IO แทน ข้อสอบและจำนวนครั้ง
R แทน ผู้ตัดสิน
IR แทน ข้อสอบและผู้ตัดสิน
O แทน จำนวนครั้งที่ทำการตัดสิน
OR แทน จำนวนครั้งและผู้ตัดสิน
Residual (IOR,e) ส่วนที่เหลือ

จากตารางที่ 83 พบว่า การกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง ค่าความแปรปรวนของแหล่งต่างๆ เมื่อเทียบกับความแปรปรวนรวมมีค่าเรียงจากมากไปหาน้อยดังนี้ ความแปรปรวนของข้อสอบมีค่า 69.6297 คิดเป็นร้อยละ 47.3290 รองลงมาคือความแปรปรวนของส่วนที่เหลือ มีค่า 44.9501 คิดเป็นร้อยละ 30.5536 ส่วนความแปรปรวนของข้อสอบและผู้ตัดสิน มีค่าน้อยที่สุดคือ 0.0000

ตาราง 84 การเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิง (Genaralizability coefficient) ในขั้นศึกษา D-study โดยทำการกำหนดมาตรฐานวิธีแองกอฟที่ได้รับการปรับปรุง มีรูปแบบ i x r x o design ที่คะแนนจุดตัดระดับอ่อนมาก (D)

Alternative Estimated D studies Design Variance Component								
n_o	1	1	1	1	1	1	1	1
n_r	8	12	14	16	18	20	22	24
$E\rho^2$	0.8308	0.8499	0.8556	0.8598	0.8632	0.8659	0.8681	0.8699
ϕ	0.7419	0.7627	0.7688	0.7735	0.7772	0.7802	0.7826	0.7847
n_o	2	2	2	2	2	2	2	2
n_r	8	12	14	16	18	20	22	24
$E\rho^2$	0.9076	0.9189	0.9222	0.9246	0.9266	0.9281	0.9294	0.9304
ϕ	0.8518	0.8654	0.8693	0.8723	0.8746	0.8765	0.8781	0.8793
n_o	3	3	3	3	3	3	3	3
n_r	8	12	14	16	18	20	22	24
$E\rho^2$	0.9364	0.9444	0.9467	0.9485	0.9498	0.9509	0.9518	0.9525
ϕ	0.8961	0.9060	0.9089	0.9111	0.9128	0.9141	0.9153	0.9162

จากตารางที่ 84 พบว่า การเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิง มีรูปแบบ i x r x o design ที่คะแนนจุดตัดระดับอ่อนมาก (D) ถ้ากำหนดมาตรฐานวิธีแองกอฟที่ได้รับการปรับปรุง 1 ครั้ง จำนวนผู้ตัดสิน 8, 12, 14, 16, 18, 20, 22 และ 24 คน ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับ

การตัดสินแบบสัมพันธ์ มีค่าเพิ่มขึ้นเมื่อผู้ตัดสินเพิ่มขึ้น มีค่าดังนี้ 0.8308, 0.8499, 0.8556, 0.8598, 0.8632, 0.8659, 0.8681, 0.8699 ตามลำดับ ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมบูรณ์ มีค่าดังนี้ 0.7419, 0.7627, 0.7688, 0.7735, 0.7772, 0.7802, 0.7826, 0.7847 ตามลำดับ

ถ้ากำหนดมาตรฐานวิธีแองกอฟที่ได้รับการปรับปรุง ซ้ำ 2 ครั้ง จำนวนผู้ตัดสิน 8, 12, 14, 16, 18, 20, 22 และ 24 คน ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมพันธ์ มีค่าเพิ่มขึ้นเมื่อผู้ตัดสินเพิ่มขึ้น และเมื่อจำนวนครั้งเพิ่มขึ้น โดยมีค่าดังนี้ 0.9076, 0.9189, 0.9222, 0.9246, 0.9266, 0.9281, 0.9294, 0.9304 ตามลำดับ ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมบูรณ์ มีค่าดังนี้ 0.8518, 0.8654, 0.8693, 0.8723, 0.8746, 0.8765, 0.8781, 0.8793 ตามลำดับ

ถ้ากำหนดมาตรฐานวิธีแองกอฟที่ได้รับการปรับปรุง ซ้ำ 3 ครั้ง จำนวนผู้ตัดสิน 8, 12, 14, 16, 18, 20, 22 และ 24 คน ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมพันธ์ มีค่าเพิ่มขึ้นเมื่อผู้ตัดสินเพิ่มขึ้น และเมื่อจำนวนครั้งเพิ่มขึ้น โดยมีค่าดังนี้ 0.9364, 0.9444, 0.9467, 0.9485, 0.9498, 0.9509, 0.9518, 0.9525 ตามลำดับ ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมบูรณ์ มีค่าดังนี้ 0.8961, 0.9060, 0.9089, 0.9111, 0.9128, 0.9141, 0.9153, 0.9162 ตามลำดับ จะเห็นได้ว่าเมื่อจำนวนผู้ตัดสิน และจำนวนครั้งเพิ่มขึ้น ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมพันธ์ และค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมบูรณ์ก็มีค่าเพิ่มขึ้นตามด้วย

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

2. ผลของจำนวนผู้ตัดสิน และจำนวนครั้งที่มีต่อการกำหนดมาตรฐานด้วยวิธีการบุ๊คมาร์ค

ตาราง 85 การวิเคราะห์ความแปรปรวนและการประมาณค่าความแปรปรวน ในขั้นศึกษา G-study โดยทำการกำหนดมาตรฐานวิธีบุ๊คมาร์ค ที่คะแนนจุดตัดระดับดีเยี่ยม (A)

Source of variance	DF	SS	MS	Estimated Variance component	% of total variance
I	99	545.6656	5.5118	0.1530	96.6501
O	2	0.0039	0.0019	0.0000	0.0000
R	11	0.0789	0.0072	0.0000	0.0000
IO	198	0.9961	0.0050	0.0000	0.0000
IR	1089	4.2544	0.0039	0.0000	0.0000
OR	22	0.2428	0.0110	0.0001	0.0366
Residual (IOR,e)	2178	11.4239	0.0053	0.0052	3.3133
Total	3599	562.6656		0.1583	100

เมื่อ I แทน ข้อสอบ
R แทน ผู้ตัดสิน
O แทน จำนวนครั้งที่ทำการตัดสิน
Residual (IOR,e) ส่วนที่เหลือ
IO แทน ข้อสอบและจำนวนครั้ง
IR แทน ข้อสอบและผู้ตัดสิน
OR แทน จำนวนครั้งและผู้ตัดสิน

จากตารางที่ 85 พบว่า การกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์ค ค่าความแปรปรวนของแหล่งต่างๆ เมื่อเทียบกับความแปรปรวนรวมมีค่าเรียงจากมากไปหาน้อยดังนี้ ความแปรปรวนของข้อสอบ มีค่า 0.1530 คิดเป็นร้อยละ 96.6501 รองลงมาคือความแปรปรวนของส่วนที่เหลือ มีค่า 0.0052 คิดเป็นร้อยละ 3.3133 ส่วนความแปรปรวนของจำนวนครั้ง ความแปรปรวนของจำนวนผู้ตัดสิน ความแปรปรวนของข้อสอบและจำนวนครั้ง ความแปรปรวนของข้อสอบและจำนวนผู้ตัดสิน มีค่าน้อยที่สุดคือ 0.0000

ตาราง 86 การเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิง (Genaralizability coefficient) .ในขั้นศึกษา D-study โดยทำการกำหนดมาตรฐานวิธีบุ๊คมาร์ค มีรูปแบบ i x r x o design ที่คะแนนจุดตัดระดับ ดีเยี่ยม (A)

Alternative Estimated D studies Design Variance Component								
n_o	1	1	1	1	1	1	1	1
n_r	8	12	14	16	18	20	22	24
$E\rho^2$	0.9957	0.9972	0.9976	0.9979	0.9981	0.9983	0.9984	0.9986
ϕ	0.9957	0.9971	0.9975	0.9978	0.9981	0.9983	0.9984	0.9986
n_o	2	2	2	2	2	2	2	2
n_r	8	12	14	16	18	20	22	24
$E\rho^2$	0.9979	0.9986	0.9988	0.9989	0.9991	0.9991	0.9992	0.9993
ϕ	0.9978	0.9986	0.9988	0.9989	0.9990	0.9991	0.9992	0.9993
n_o	3	3	3	3	3	3	3	3
n_r	8	12	14	16	18	20	22	24
$E\rho^2$	0.9985	0.9990	0.9991	0.9992	0.9993	0.9994	0.9994	0.9995
ϕ	0.9985	0.9990	0.9991	0.9992	0.9993	0.9994	0.9994	0.9995

จากตารางที่ 86 พบว่า การเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิง มีรูปแบบ i x r x o design ที่คะแนนจุดตัดระดับดีเยี่ยม (A) ถ้ากำหนดมาตรฐานวิธีบุ๊คมาร์ค 1 ครั้ง จำนวนผู้ตัดสิน 8, 12, 14, 16, 18, 20, 22 และ 24 คน ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมพันธ์ มีค่าเพิ่มขึ้นเมื่อผู้ตัดสินเพิ่มขึ้น มีค่าดังนี้ 0.9957, 0.9972, 0.9976, 0.9979, 0.9981, 0.9983, 0.9984, 0.9986 ตามลำดับ ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมบูรณ์ มีค่าดังนี้ 0.9957, 0.9971, 0.9975, 0.9978, 0.9981, 0.9983, 0.9984, 0.9986 ตามลำดับ

ถ้ากำหนดมาตรฐานวิธีบุ๊คมาร์ค ซ้ำ 2 ครั้ง จำนวนผู้ตัดสิน 8, 12, 14, 16, 18, 20, 22 และ 24 คน ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมพันธ์ มีค่าเพิ่มขึ้นเมื่อผู้ตัดสินเพิ่มขึ้น และเมื่อจำนวนครั้งเพิ่มขึ้น โดยมีค่าดังนี้ 0.9979, 0.9986, 0.9988, 0.9989, 0.9991, 0.9991, 0.9992, 0.9993 ตามลำดับ ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมบูรณ์ มีค่าดังนี้ 0.9978, 0.9986, 0.9988, 0.9989, 0.9990, 0.9991, 0.9992, 0.9993 ตามลำดับ

ถ้ากำหนดมาตรฐานวิธีบุ๊คมาร์ค ซ้ำ 3 ครั้ง จำนวนผู้ตัดสิน 8, 12, 14, 16, 18, 20, 22 และ 24 คน ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมพันธ์ มีค่าเพิ่มขึ้นเมื่อผู้ตัดสินเพิ่มขึ้น และเมื่อจำนวนครั้งเพิ่มขึ้น โดยมีค่าดังนี้ 0.9985, 0.9990, 0.9991, 0.9992, 0.9993,

0.9994, 0.9994, 0.9995 ตามลำดับ ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมบูรณ์มีค่าดังนี้ 0.9985, 0.9990, 0.9991, 0.9992, 0.9993, 0.9994, 0.9994, 0.9995 ตามลำดับ จะเห็นได้ว่าเมื่อจำนวนผู้ตัดสิน และจำนวนครั้งเพิ่มขึ้น ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมพันธ์ และค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมบูรณ์ก็มีค่าเพิ่มขึ้นตามด้วย

ตาราง 87 การวิเคราะห์ความแปรปรวนและการประมาณค่าความแปรปรวน ในขั้นศึกษา G-study โดยทำการกำหนดมาตรฐานวิธีบุ๊คมาร์ค ที่คะแนนจุดตัดระดับดีมาก(B+)

Source of variance	DF	SS	MS	Estimated Variance component	% of total variance
I	99	801.7100	8.0981	0.2249	98.6268
O	2	0.0022	0.0011	0.0000	0.0000
R	11	0.0989	0.0090	0.0000	0.0000
IO	198	0.2200	0.0011	0.0000	0.0000
IR	1089	3.7900	0.0035	0.0002	0.0832
OR	22	0.1044	0.0048	0.0000	0.0000
Residual (IOR,e)	2178	6.3400	0.0029	0.0029	1.2765
Total	3599	812.2656		0.2280	100

เมื่อ I แทน ข้อสอบ
IO แทน ข้อสอบและจำนวนครั้ง
R แทน ผู้ตัดสิน
IR แทน ข้อสอบและผู้ตัดสิน
O แทน จำนวนครั้งที่ทำการตัดสิน
OR แทน จำนวนครั้งและผู้ตัดสิน
Residual (IOR,e) ส่วนที่เหลือ

จากตารางที่ 87 พบว่า การกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์ค ค่าความแปรปรวนของแหล่งต่างๆ เมื่อเทียบกับความแปรปรวนรวมมีค่าเรียงจากมากไปหาน้อยดังนี้ ความแปรปรวนของข้อสอบ มีค่า 0.2249 คิดเป็นร้อยละ 98.6268 รองลงมาคือความแปรปรวนของส่วนที่เหลือ มีค่า 0.0029 คิดเป็นร้อยละ 1.2765 ส่วนความแปรปรวนของจำนวนครั้ง ความแปรปรวนของจำนวนผู้ตัดสิน ความแปรปรวนของข้อสอบและจำนวนครั้ง ความแปรปรวนของข้อสอบและจำนวนผู้ตัดสิน มีค่าน้อยที่สุดคือ 0.0000

ตาราง 88 การเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิง (Genaralizability coefficient) .ในขั้นศึกษา D-study โดยทำการกำหนดมาตรฐานวิธีบุ๊คมาร์ค มีรูปแบบ i x r x o design ที่คะแนนจุดตัดระดับดีมาก (B+)

Alternative Estimated D studies Design Variance Component								
n_o	1	1	1	1	1	1	1	1
n_r	8	12	14	16	18	20	22	24
$E\rho^2$	0.9983	0.9989	0.9990	0.9991	0.9992	0.9993	0.9994	0.9994
ϕ	0.9983	0.9988	0.9990	0.9991	0.9992	0.9993	0.9994	0.9994
n_o	2	2	2	2	2	2	2	2
n_r	8	12	14	16	18	20	22	24
$E\rho^2$	0.9991	0.9994	0.9995	0.9995	0.9996	0.9996	0.9997	0.9997
ϕ	0.9991	0.9994	0.9995	0.9995	0.9996	0.9996	0.9997	0.9997
n_o	3	3	3	3	3	3	3	3
n_r	8	12	14	16	18	20	22	24
$E\rho^2$	0.9994	0.9996	0.9996	0.9997	0.9997	0.9997	0.9998	0.9998
ϕ	0.9994	0.9996	0.9996	0.9997	0.9997	0.9997	0.9998	0.9998

จากตารางที่ 88 พบว่า การเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิง มีรูปแบบ i x r x o design ที่คะแนนจุดตัดระดับดีมาก (B+) ถ้ากำหนดมาตรฐานวิธีบุ๊คมาร์ค 1 ครั้ง จำนวนผู้ตัดสิน 8, 12, 14, 16, 18, 20, 22 และ 24 คน ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมพันธ์ มีค่าเพิ่มขึ้นเมื่อผู้ตัดสินเพิ่มขึ้น มีค่าดังนี้ 0.9983, 0.9989, 0.9990, 0.9991, 0.9992, 0.9993, 0.9994, 0.9994 ตามลำดับ ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมบูรณ์ มีค่าดังนี้ 0.9983, 0.9988, 0.9990, 0.9991, 0.9992, 0.9993, 0.9994, 0.9994 ตามลำดับ

ถ้ากำหนดมาตรฐานวิธีบุ๊คมาร์ค ซ้ำ 2 ครั้ง จำนวนผู้ตัดสิน 8, 12, 14, 16, 18, 20, 22 และ 24 คน ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมพันธ์ มีค่าเพิ่มขึ้นเมื่อผู้ตัดสินเพิ่มขึ้น และเมื่อจำนวนครั้งเพิ่มขึ้น โดยมีค่าดังนี้ 0.9991, 0.9994, 0.9995, 0.9995, 0.9996, 0.9996, 0.9997, 0.9997 ตามลำดับ ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมบูรณ์ มีค่าดังนี้ 0.9991, 0.9994, 0.9995, 0.9995, 0.9996, 0.9996, 0.9997, 0.9997 ตามลำดับ

ถ้ากำหนดมาตรฐานวิธีบุ๊คมาร์ค ซ้ำ 3 ครั้ง จำนวนผู้ตัดสิน 8, 12, 14, 16, 18, 20, 22 และ 24 คน ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมพันธ์ มีค่าเพิ่มขึ้นเมื่อผู้ตัดสินเพิ่มขึ้น และเมื่อจำนวนครั้งเพิ่มขึ้น โดยมีค่าดังนี้ 0.9994, 0.9996, 0.9996, 0.9997, 0.9997,

0.9997, 0.9998, 0.9998 ตามลำดับ ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมบูรณ์มีค่าดังนี้ 0.9994, 0.9996, 0.9996, 0.9997, 0.9997, 0.9997, 0.9998, 0.9998 ตามลำดับ จะเห็นได้ว่าเมื่อจำนวนผู้ตัดสิน และจำนวนครั้งเพิ่มขึ้น ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมพันธ์ และค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมบูรณ์ก็มีค่าเพิ่มขึ้นตามด้วย

ตาราง 89 การวิเคราะห์ความแปรปรวนและการประมาณค่าความแปรปรวน ในขั้นศึกษา G-study โดยทำการกำหนดมาตรฐานวิธีบุ๊คมาร์ค ที่คะแนนจุดตัดระดับดี(B)

Source of variance	DF	SS	MS	Estimated Variance component	% of total variance
I	99	878.70444	8.8758	0.2460	97.7013
O	2	0.0339	0.0169	0.0000	0.0000
R	11	0.1356	0.0123	0.0000	0.0000
IO	198	3.57722	0.0181	0.0011	0.4535
IR	1089	5.4756	0.0050	0.0002	0.0880
OR	22	0.2194	0.0010	0.0001	0.0223
Residual (IOR,e)	2178	9.5028	0.0044	0.0044	1.7326
Total	3599	897.6489		0.2518	100

เมื่อ I แทน ข้อสอบ
IO แทน ข้อสอบและจำนวนครั้ง
R แทน ผู้ตัดสิน
IR แทน ข้อสอบและผู้ตัดสิน
O แทน จำนวนครั้งที่ทำการตัดสิน
OR แทน จำนวนครั้งและผู้ตัดสิน
Residual (IOR,e) ส่วนที่เหลือ

จากตารางที่ 89 พบว่า การกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์ค ค่าความแปรปรวนของแหล่งต่างๆ เมื่อเทียบกับความแปรปรวนรวมมีค่าเรียงจากมากไปหาน้อยดังนี้ ความแปรปรวนของข้อสอบ มีค่า 0.2460 คิดเป็นร้อยละ 97.7013 รองลงมาคือความแปรปรวนของส่วนที่เหลือ มีค่า 0.0044 คิดเป็นร้อยละ 1.7326 ส่วนความแปรปรวนของจำนวนครั้ง ความแปรปรวนของจำนวนผู้ตัดสิน มีค่าน้อยที่สุดคือ 0.0000

ตาราง 90 การเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิง (Genaralizability coefficient) .ในขั้นศึกษา D-study โดยทำการกำหนดมาตรฐานวิธีบุ๊คมาร์ค มีรูปแบบ i x r x o design ที่คะแนนจุดตัดระดับดี (B)

Alternative Estimated D studies Design Variance Component								
n_o	1	1	1	1	1	1	1	1
n_r	8	12	14	16	18	20	22	24
$E\rho^2$	0.9931	0.9938	0.9941	0.9942	0.9944	0.9945	0.9945	0.9946
ϕ	0.9931	0.9938	0.9941	0.9942	0.9943	0.9945	0.9945	0.9946
n_o	2	2	2	2	2	2	2	2
n_r	8	12	14	16	18	20	22	24
$E\rho^2$	0.9965	0.9969	0.9970	0.9971	0.9971	0.9972	0.9972	0.9973
ϕ	0.9965	0.9969	0.9970	0.9971	0.9971	0.9972	0.9972	0.9973
n_o	3	3	3	3	3	3	3	3
n_r	8	12	14	16	18	20	22	24
$E\rho^2$	0.9976	0.9979	0.9980	0.9980	0.9981	0.9981	0.9982	0.9982
ϕ	0.9976	0.9979	0.9980	0.9980	0.9981	0.9981	0.9981	0.9982

จากตารางที่ 90 พบว่า การเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิง มีรูปแบบ i x r x o design ที่คะแนนจุดตัดระดับดี (B) ถ้ากำหนดมาตรฐานวิธีบุ๊คมาร์ค 1 ครั้ง จำนวนผู้ตัดสิน 8, 12, 14, 16, 18, 20, 22 และ 24 คน ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมพันธ์มีค่าเพิ่มขึ้นเมื่อผู้ตัดสินเพิ่มขึ้น มีค่าดังนี้ 0.9931, 0.9938, 0.9941, 0.9942, 0.9944, 0.9945, 0.9945, 0.9946 ตามลำดับ ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมบูรณ์ มีค่าดังนี้ 0.9931, 0.9938, 0.9941, 0.9942, 0.9943, 0.9945, 0.9945, 0.9946 ตามลำดับ

ถ้ากำหนดมาตรฐานวิธีบุ๊คมาร์ค ซ้ำ 2 ครั้ง จำนวนผู้ตัดสิน 8, 12, 14, 16, 18, 20, 22 และ 24 คน ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมพันธ์ มีค่าเพิ่มขึ้นเมื่อผู้ตัดสินเพิ่มขึ้น และเมื่อจำนวนครั้งเพิ่มขึ้น โดยมีค่าดังนี้ 0.9965, 0.9969, 0.9970, 0.9971, 0.9971, 0.9972, 0.9972, 0.9973 ตามลำดับ ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมบูรณ์ มีค่าดังนี้ 0.9965, 0.9969, 0.9970, 0.9971, 0.9971, 0.9972, 0.9972, 0.9973 ตามลำดับ

ถ้ากำหนดมาตรฐานวิธีบุ๊คมาร์ค ซ้ำ 3 ครั้ง จำนวนผู้ตัดสิน 8, 12, 14, 16, 18, 20, 22 และ 24 คน ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมพันธ์ มีค่าเพิ่มขึ้นเมื่อผู้ตัดสินเพิ่มขึ้น และเมื่อจำนวนครั้งเพิ่มขึ้น โดยมีค่าดังนี้ 0.9976, 0.9979, 0.9980, 0.9980,

0.9981, 0.9981, 0.9982, 0.9982 ตามลำดับ ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมบูรณ์ มีค่าดังนี้ 0.9976, 0.9979, 0.9980, 0.9980, 0.9981, 0.9981, 0.9981, 0.9982 ตามลำดับ จะเห็นได้ว่าเมื่อจำนวนผู้ตัดสิน และจำนวนครั้งเพิ่มขึ้น ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมพันธ์ และค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมบูรณ์ก็มีค่าเพิ่มขึ้นตามด้วย

ตาราง 91 การวิเคราะห์ความแปรปรวนและการประมาณค่าความแปรปรวน ในขั้นศึกษา G-study โดยทำการกำหนดมาตรฐานวิธีบุ๊คมาร์ค ที่คะแนนจุดตัดระดับดีพอใช้(C+)

Source of variance	DF	SS	MS	Estimated Variance component	% of total variance
I	99	293416.9233	2963.8073	0.2181	96.7557
O	2	3323.4350	1661.7175	0.0000	0.0000
R	11	23296.0100	2117.8191	0.0000	0.0000
IO	198	42124.7317	212.7512	0.0016	0.7180
IR	1089	57837.9900	53.1111	0.0001	0.0396
OR	22	269569.3650	12253.1530	0.0002	0.1044
Residual (IOR,e)	2178	211969.1350	97.3228	0.0054	2.3823
Total	3599	901537.5900		0.2254	100

เมื่อ I แทน ข้อสอบ
IO แทน ข้อสอบและจำนวนครั้ง
R แทน ผู้ตัดสิน
IR แทน ข้อสอบและผู้ตัดสิน
O แทน จำนวนครั้งที่ทำการตัดสิน
OR แทน จำนวนครั้งและผู้ตัดสิน
Residual (IOR,e) ส่วนที่เหลือ

จากตารางที่ 91 พบว่า การกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์ค ค่าความแปรปรวนของแหล่งต่างๆ เมื่อเทียบกับความแปรปรวนรวมมีค่าเรียงจากมากไปหาน้อยดังนี้ ความแปรปรวนของข้อสอบ มีค่า 0.2181 คิดเป็นร้อยละ96.7557 รองลงมาคือความแปรปรวนของส่วนที่เหลือ มีค่า 0.0054 คิดเป็นร้อยละ 2.3823 ส่วนความแปรปรวนของจำนวนครั้ง ความแปรปรวนของจำนวนผู้ตัดสิน มีค่าน้อยที่สุดคือ 0.0000

ตาราง 92 การเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิง (Genaralizability coefficient) .ในขั้นศึกษา D-study โดยทำการกำหนดมาตรฐานวิธีบุ๊คมาร์ค มีรูปแบบ i x r x o design ที่คะแนนจุดตัดระดับดีพอใช้ (C+)

Alternative Estimated D studies Design Variance Component								
n_o	1	1	1	1	1	1	1	1
n_r	8	12	14	16	18	20	22	24
$E\rho^2$	0.9896	0.9906	0.9909	0.9911	0.9913	0.9914	0.9915	0.9916
ϕ	0.9894	0.9905	0.9908	0.9910	0.9912	0.9914	0.9915	0.9916
n_o	2	2	2	2	2	2	2	2
n_r	8	12	14	16	18	20	22	24
$E\rho^2$	0.9947	0.9953	0.9954	0.9955	0.9956	0.9957	0.9957	0.9958
ϕ	0.9947	0.9952	0.9954	0.9955	0.9956	0.9957	0.9957	0.9958
n_o	3	3	3	3	3	3	3	3
n_r	8	12	14	16	18	20	22	24
$E\rho^2$	0.9965	0.9968	0.9969	0.9970	0.9971	0.9971	0.9971	0.9972
ϕ	0.9964	0.9968	0.9969	0.9970	0.9970	0.9971	0.9971	0.9972

จากตารางที่ 92 พบว่า การเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิง มีรูปแบบ i x r x o design ที่คะแนนจุดตัดระดับดีพอใช้ (C+) ถ้ากำหนดมาตรฐานวิธีบุ๊คมาร์ค 1 ครั้ง จำนวนผู้ตัดสิน 8, 12, 14, 16, 18, 20, 22 และ 24 คน ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมพันธ์ มีค่าเพิ่มขึ้นเมื่อผู้ตัดสินเพิ่มขึ้น มีค่าดังนี้ 0.9896, 0.9906, 0.9909, 0.9911, 0.9913, 0.9914, 0.9915, 0.9916 ตามลำดับ ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมบูรณ์ มีค่าดังนี้ 0.9894, 0.9905, 0.9908, 0.9910, 0.9912, 0.9914, 0.9915, 0.9916 ตามลำดับ

ถ้ากำหนดมาตรฐานวิธีบุ๊คมาร์ค ซ้ำ 2 ครั้ง จำนวนผู้ตัดสิน 8, 12, 14, 16, 18, 20, 22 และ 24 คน ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมพันธ์ มีค่าเพิ่มขึ้นเมื่อผู้ตัดสินเพิ่มขึ้น และเมื่อจำนวนครั้งเพิ่มขึ้น โดยมีค่าดังนี้ 0.9947, 0.9953, 0.9954, 0.9955, 0.9956, 0.9957, 0.9957, 0.9958 ตามลำดับ ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมบูรณ์ มีค่าดังนี้ 0.9947, 0.9952, 0.9954, 0.9955, 0.9956, 0.9957, 0.9957, 0.9958 ตามลำดับ

ถ้ากำหนดมาตรฐานวิธีบุ๊คมาร์ค ซ้ำ 3 ครั้ง จำนวนผู้ตัดสิน 8, 12, 14, 16, 18, 20, 22 และ 24 คน ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมพันธ์ มีค่าเพิ่มขึ้นเมื่อผู้ตัดสินเพิ่มขึ้น และเมื่อจำนวนครั้งเพิ่มขึ้น โดยมีค่าดังนี้ 0.9965, 0.9968, 0.9969, 0.9970, 0.9971,

0.9971, 0.9971, 0.9972 ตามลำดับ ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมบูรณ์มีค่าดังนี้ 0.9964, 0.9968, 0.9969, 0.9970, 0.9970, 0.9971, 0.9971, 0.9972 ตามลำดับ จะเห็นได้ว่าเมื่อจำนวนผู้ตัดสิน และจำนวนครั้งเพิ่มขึ้น ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมพันธ์ และค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมบูรณ์ก็มีค่าเพิ่มขึ้นตามด้วย

ตาราง 93 การวิเคราะห์ความแปรปรวนและการประมาณค่าความแปรปรวน ในขั้นศึกษา G-study โดยทำการกำหนดมาตรฐานวิธีบุ๊คมาร์ค ที่คะแนนจุดตัดระดับพอใช้(C)

Source of variance	DF	SS	MS	Estimated Variance component	% of total variance
I	99	653.2267	6.5983	0.1832	98.8013
O	2	0.0356	0.0178	0.0000	0.0000
R	11	0.0822	0.0075	0.0000	0.0000
IO	198	0.8533	0.0043	0.0002	0.1054
IR	1089	2.1400	0.0020	0.0000	0.0000
OR	22	0.1644	0.0075	0.0001	0.0297
Residual (IOR,e)	2178	4.2800	0.0020	0.0020	1.0600
Total	3599	660.7822		0.1854	100

เมื่อ I แทน ข้อสอบ
R แทน ผู้ตัดสิน
O แทน จำนวนครั้งที่ทำการตัดสิน
Residual (IOR,e) ส่วนที่เหลือ
IO แทน ข้อสอบและจำนวนครั้ง
IR แทน ข้อสอบและผู้ตัดสิน
OR แทน จำนวนครั้งและผู้ตัดสิน

จากตารางที่ 93 พบว่า การกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์ค ค่าความแปรปรวนของแหล่งต่างๆ เมื่อเทียบกับความแปรปรวนรวมมีค่าเรียงจากมากไปหาน้อยดังนี้ ความแปรปรวนของข้อสอบ มีค่า 0.1832 คิดเป็นร้อยละ 98.8013 รองลงมาคือความแปรปรวนของส่วนที่เหลือ มีค่า 0.0020 คิดเป็นร้อยละ 1.0600 ส่วนความแปรปรวนของจำนวนครั้ง ความแปรปรวนของจำนวนผู้ตัดสิน ความแปรปรวนของข้อสอบและผู้ตัดสิน มีค่าน้อยที่สุดคือ 0.0000

ตาราง 94 การเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิง (Genaralizability coefficient) .ในขั้นศึกษา D-study โดยทำการกำหนดมาตรฐานวิธีบุ๊คมาร์ค มีรูปแบบ i x r x o design ที่คะแนนจุดตัดระดับพอใช้ (C)

Alternative Estimated D studies Design Variance Component								
n_o	1	1	1	1	1	1	1	1
n_r	8	12	14	16	18	20	22	24
$E\rho^2$	0.9976	0.9980	0.9982	0.9983	0.9983	0.9984	0.9985	0.9985
ϕ	0.9975	0.9980	0.9981	0.9982	0.9983	0.9984	0.9984	0.9984
n_o	2	2	2	2	2	2	2	2
n_r	8	12	14	16	18	20	22	24
$E\rho^2$	0.9988	0.9990	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992
ϕ	0.9988	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992
n_o	3	3	3	3	3	3	3	3
n_r	8	12	14	16	18	20	22	24
$E\rho^2$	0.9992	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995	0.9995
ϕ	0.9992	0.9993	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995

จากตารางที่ 94 พบว่า การเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิง มีรูปแบบ i x r x o design ที่คะแนนจุดตัดระดับพอใช้ (C) ถ้ากำหนดมาตรฐานวิธีบุ๊คมาร์ค 1 ครั้ง จำนวนผู้ตัดสิน 8, 12, 14, 16, 18, 20, 22 และ 24 คน ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมพันธ์ มีค่าเพิ่มขึ้นเมื่อผู้ตัดสินเพิ่มขึ้น มีค่าดังนี้ 0.9976, 0.9980, 0.9982, 0.9983, 0.9983, 0.9984, 0.9985, 0.9985 ตามลำดับ ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมบูรณ์ มีค่าดังนี้ 0.9975, 0.9980, 0.9981, 0.9982, 0.9983, 0.9984, 0.9984, 0.9984 ตามลำดับ

ถ้ากำหนดมาตรฐานวิธีบุ๊คมาร์ค ซ้ำ 2 ครั้ง จำนวนผู้ตัดสิน 8, 12, 14, 16, 18, 20, 22 และ 24 คน ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมพันธ์ มีค่าเพิ่มขึ้นเมื่อผู้ตัดสินเพิ่มขึ้น และเมื่อจำนวนครั้งเพิ่มขึ้น โดยมีค่าดังนี้ 0.9988, 0.9990, 0.9991, 0.9991, 0.9992, 0.9992, 0.9992, 0.9992 ตามลำดับ ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมบูรณ์ มีค่าดังนี้ 0.9988, 0.9990, 0.9991, 0.9991, 0.9991, 0.9992, 0.9992, 0.9992 ตามลำดับ

ถ้ากำหนดมาตรฐานวิธีบุ๊คมาร์ค ซ้ำ 3 ครั้ง จำนวนผู้ตัดสิน 8, 12, 14, 16, 18, 20, 22 และ 24 คน ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมพันธ์ มีค่าเพิ่มขึ้นเมื่อผู้ตัดสินเพิ่มขึ้น และเมื่อจำนวนครั้งเพิ่มขึ้น โดยมีค่าดังนี้ 0.9992, 0.9994, 0.9994, 0.9994, 0.9995,

0.9995, 0.9995, 0.9995 ตามลำดับ ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมบูรณ์มีค่าดังนี้ 0.9992, 0.9993, 0.9994, 0.9994, 0.9994, 0.9995, 0.9995, 0.9995 ตามลำดับ จะเห็นได้ว่าเมื่อจำนวนผู้ตัดสิน และจำนวนครั้งเพิ่มขึ้น ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมพันธ์ และค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมบูรณ์ก็มีค่าเพิ่มขึ้นตามด้วย

ตาราง 95 การวิเคราะห์ความแปรปรวนและการประมาณค่าความแปรปรวน ในขั้นศึกษา G-study โดยทำการกำหนดมาตรฐานวิธีบุ๊คมาร์ค ที่คะแนนจุดตัดระดับอ่อน(D+)

Source of variance	DF	SS	MS	Estimated Variance component	% of total variance
I	99	274406.0253	2771.7780	0.1249	98.3297
O	2	47512.1850	23756.0925	0.0000	0.0000
R	11	78905.1542	7173.1958	0.0000	0.0000
IO	198	37981.5372	191.8259	0.0005	0.4176
IR	1089	50642.8181	46.5039	0.0005	0.3579
OR	22	116071.2083	5275.9640	0.0000	0.0000
Residual (IOR,e)	2178	127933.0694	58.7387	0.0011	0.8948
Total	3599	733451.9975		0.1270	100

เมื่อ I แทน ข้อสอบ
R แทน ผู้ตัดสิน
O แทน จำนวนครั้งที่ทำการตัดสิน
Residual (IOR,e) ส่วนที่เหลือ
IO แทน ข้อสอบและจำนวนครั้ง
IR แทน ข้อสอบและผู้ตัดสิน
OR แทน จำนวนครั้งและผู้ตัดสิน

จากตารางที่ 95 พบว่า การกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์ค ค่าความแปรปรวนของแหล่งต่างๆ เมื่อเทียบกับความแปรปรวนรวมมีค่าเรียงจากมากไปหาน้อยดังนี้ ความแปรปรวนของข้อสอบ มีค่า 0.1249 คิดเป็นร้อยละ 98.3297 รองลงมาคือความแปรปรวนของส่วนที่เหลือ มีค่า 0.0011 คิดเป็นร้อยละ 0.8948 ส่วนความแปรปรวนของจำนวนครั้ง ความแปรปรวนของจำนวนผู้ตัดสิน ความแปรปรวนของจำนวนครั้งและผู้ตัดสิน มีค่าน้อยที่สุดคือ 0.0000

ตาราง 96 การเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิง (Genaralizability coefficient) .ในขั้นศึกษา D-study โดยทำการกำหนดมาตรฐานวิธีบุ๊คมาร์ค มีรูปแบบ i x r x o design ที่คะแนนจุดตัดระดับอ่อน (D+)

Alternative Estimated D studies Design Variance Component								
n_o	1	1	1	1	1	1	1	1
n_r	8	12	14	16	18	20	22	24
$E\rho^2$	0.7954	0.8176	0.8242	0.8292	0.8331	0.8363	0.8389	0.8411
ϕ	0.9942	0.9947	0.9949	0.9950	0.9951	0.9951	0.9952	0.9953
n_o	2	2	2	2	2	2	2	2
n_r	8	12	14	16	18	20	22	24
$E\rho^2$	0.9969	0.9972	0.9973	0.9974	0.9974	0.9975	0.9975	0.9975
ϕ	0.9969	0.9972	0.9973	0.9974	0.9974	0.9975	0.9975	0.9975
n_o	3	3	3	3	3	3	3	3
n_r	8	12	14	16	18	20	22	24
$E\rho^2$	0.9978	0.9980	0.9981	0.9982	0.9982	0.9983	0.9983	0.9983
ϕ	0.9978	0.9980	0.9981	0.9982	0.9982	0.9983	0.9983	0.9983

จากตารางที่ 96 พบว่า การเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิง มีรูปแบบ i x r x o design ที่คะแนนจุดตัดระดับอ่อน (D+) ถ้ากำหนดมาตรฐานวิธีบุ๊คมาร์ค 1 ครั้ง จำนวนผู้ตัดสิน 8, 12, 14, 16, 18, 20, 22 และ 24 คน ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมพันธ์ มีค่าเพิ่มขึ้นเมื่อผู้ตัดสินเพิ่มขึ้น มีค่าดังนี้ 0.7954, 0.8176, 0.8242, 0.8292, 0.8331, 0.8363, 0.8389, 0.8411 ตามลำดับ ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมบูรณ์ มีค่าดังนี้ 0.9942, 0.9947, 0.9949, 0.9950, 0.9951, 0.9951, 0.9952, 0.9953 ตามลำดับ

ถ้ากำหนดมาตรฐานวิธีบุ๊คมาร์ค ซ้ำ 2 ครั้ง จำนวนผู้ตัดสิน 8, 12, 14, 16, 18, 20, 22 และ 24 คน ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมพันธ์ มีค่าเพิ่มขึ้นเมื่อผู้ตัดสินเพิ่มขึ้น และเมื่อจำนวนครั้งเพิ่มขึ้น โดยมีค่าดังนี้ 0.9969, 0.9972, 0.9973, 0.9974, 0.9974, 0.9975, 0.9975, 0.9975 ตามลำดับ ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมบูรณ์ มีค่าดังนี้ 0.9969, 0.9972, 0.9973, 0.9974, 0.9974, 0.9975, 0.9975, 0.9975 ตามลำดับ

ถ้ากำหนดมาตรฐานวิธีบุ๊คมาร์ค ซ้ำ 3 ครั้ง จำนวนผู้ตัดสิน 8, 12, 14, 16, 18, 20, 22 และ 24 คน ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมพันธ์ มีค่าเพิ่มขึ้นเมื่อผู้ตัดสิน

เพิ่มขึ้น และเมื่อจำนวนครั้งเพิ่มขึ้น โดยมีค่าดังนี้ 0.9978, 0.9980, 0.9981, 0.9982, 0.9982, 0.9983, 0.9983, 0.9983 ตามลำดับ ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมบูรณ์ มีค่าดังนี้ 0.9978, 0.9980, 0.9981, 0.9982, 0.9982, 0.9983, 0.9983, 0.9983 ตามลำดับ จะเห็นได้ว่าเมื่อจำนวนผู้ตัดสิน และจำนวนครั้งเพิ่มขึ้น ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมพันธ์ และค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมบูรณ์ก็มีค่าเพิ่มขึ้นตามด้วย

ตาราง 97 การวิเคราะห์ความแปรปรวนและการประมาณค่าความแปรปรวน ในขั้นศึกษา G-study โดยทำการกำหนดมาตรฐานวิธีบุ๊คมาร์ค ที่คะแนนจุดตัดระดับอ่อนมาก (D)

Source of variance	DF	SS	MS	Estimated Variance component	% of total variance
I	99	250.4475	2.52977	0.0700	95.0929
O	2	0.04389	0.02194	0.0000	0.0000
R	11	0.00972	0.00088	0.0000	0.0000
IO	198	2.345	0.01184	0.0007	1.0183
IR	1089	1.62917	0.0015	0.0000	0.0000
OR	22	0.06944	0.00316	0.0000	0.0000
Residual (IOR,e)	2178	6.20833	0.00285	0.0029	3.8734
Total	3599	260.75306		0.0736	100

เมื่อ I แทน ข้อสอบ
IO แทน ข้อสอบและจำนวนครั้ง
R แทน ผู้ตัดสิน
IR แทน ข้อสอบและผู้ตัดสิน
O แทน จำนวนครั้งที่ทำการตัดสิน
OR แทน จำนวนครั้งและผู้ตัดสิน
Residual (IOR,e) ส่วนที่เหลือ

จากตารางที่ 97 พบว่า การกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์ค ค่าความแปรปรวนของแหล่งต่างๆ เมื่อเทียบกับความแปรปรวนรวมมีค่าเรียงจากมากไปหาน้อยดังนี้ ความแปรปรวนของข้อสอบ มีค่า 0.0700 คิดเป็นร้อยละ 95.0929 รองลงมาคือความแปรปรวนของส่วนที่เหลือ มีค่า 0.0029 คิดเป็นร้อยละ 3.8734 ส่วนความแปรปรวนของจำนวนครั้ง ความแปรปรวนของจำนวนผู้ตัดสิน ความแปรปรวนของจำนวนครั้งและผู้ตัดสิน มีค่าน้อยที่สุดคือ 0.0000

ตาราง 98 การเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิง (Genaralizability coefficient) .ในขั้นศึกษา D-study โดยทำการกำหนดมาตรฐานวิธีบุ๊คมาร์ค มีรูปแบบ i x r x o design ที่คะแนนจุดตัดระดับอ่อนมาก (D)

Alternative Estimated D studies Design Variance Component								
n_o	1	1	1	1	1	1	1	1
n_r	8	12	14	16	18	20	22	24
$E\rho^2$	0.9844	0.9861	0.9866	0.9869	0.9872	0.9874	0.9876	0.9877
ϕ	0.9843	0.9860	0.9864	0.9868	0.9871	0.9873	0.9875	0.9876
n_o	2	2	2	2	2	2	2	2
n_r	8	12	14	16	18	20	22	24
$E\rho^2$	0.9922	0.9930	0.9932	0.9934	0.9936	0.9937	0.9938	0.9938
ϕ	0.9921	0.9929	0.9932	0.9934	0.9935	0.9936	0.9937	0.9938
n_o	3	3	3	3	3	3	3	3
n_r	8	12	14	16	18	20	22	24
$E\rho^2$	0.9948	0.9953	0.9955	0.9956	0.9957	0.9958	0.9958	0.9959
ϕ	0.9947	0.9953	0.9954	0.9956	0.9957	0.9957	0.9958	0.9958

จากตารางที่ 98 พบว่า การเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิง มีรูปแบบ i x r x o design ที่คะแนนจุดตัดระดับอ่อนมาก (D) ถ้ากำหนดมาตรฐานวิธีบุ๊คมาร์ค 1 ครั้ง จำนวนผู้ตัดสิน 8, 12, 14, 16, 18, 20, 22 และ 24 คน ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมพันธ์ มีค่าเพิ่มขึ้นเมื่อผู้ตัดสินเพิ่มขึ้น มีค่าดังนี้ 0.9844, 0.9861, 0.9866, 0.9869, 0.9872, 0.9874, 0.9876, 0.9877 ตามลำดับ ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมบูรณ์ มีค่าดังนี้ 0.9843, 0.9860, 0.9864, 0.9868, 0.9871, 0.9873, 0.9875, 0.9876 ตามลำดับ

ถ้ากำหนดมาตรฐานวิธีบุ๊คมาร์ค ซ้ำ 2 ครั้ง จำนวนผู้ตัดสิน 8, 12, 14, 16, 18, 20, 22 และ 24 คน ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมพันธ์ มีค่าเพิ่มขึ้นเมื่อผู้ตัดสินเพิ่มขึ้น และเมื่อจำนวนครั้งเพิ่มขึ้น โดยมีค่าดังนี้ 0.9922, 0.9930, 0.9932, 0.9934, 0.9936, 0.9937, 0.9938, 0.9938 ตามลำดับ ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมบูรณ์ มีค่าดังนี้ 0.9921, 0.9929, 0.9932, 0.9934, 0.9935, 0.9936, 0.9937, 0.9938 ตามลำดับ

ถ้ากำหนดมาตรฐานวิธีบุ๊คมาร์ค ซ้ำ 3 ครั้ง จำนวนผู้ตัดสิน 8, 12, 14, 16, 18, 20, 22 และ 24 คน ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมพันธ์ มีค่าเพิ่มขึ้นเมื่อผู้ตัดสินเพิ่มขึ้น และเมื่อจำนวนครั้งเพิ่มขึ้น โดยมีค่าดังนี้ 0.9948, 0.9953, 0.9955, 0.9956, 0.9957,

0.9958, 0.9958, 0.9959 ตามลำดับ ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมบูรณ์มีค่าดังนี้ 0.9947, 0.9953, 0.9954, 0.9956, 0.9957, 0.9957, 0.9958, 0.9958 ตามลำดับ จะเห็นได้ว่าเมื่อจำนวนผู้ตัดสิน และจำนวนครั้งเพิ่มขึ้น ค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมพันธ์ และค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการตัดสินแบบสัมบูรณ์ก็มีค่าเพิ่มขึ้นตามด้วย

ผู้วิจัยได้สรุปค่าความเที่ยงหรือสัมประสิทธิ์การสรุปอ้างอิงที่กำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และวิธีบุ๊คมาร์ค ผู้วิจัยออกแบบการวัดเป็น แบบ Two-facet crossed design ฟาเซท (facets) ของการวัด มี 2 ฟาเซท คือ ผู้ตัดสิน (r) กำหนดจำนวนระดับเป็น 8, 12, 14, 16, 18, 20, 22, 24 คน และจำนวนครั้ง (o) กำหนดจำนวนระดับเป็น 1, 2, 3 ครั้ง รูปแบบของความสัมพันธ์ระหว่างฟาเซทเป็น i x r x o นั้นคือ ให้ผู้ตัดสินทุกคนตัดสินข้อสอบทุกข้อ และทำการตัดสินซ้ำ 3 ครั้ง .ในการกำหนดมาตรฐานทั้งสอง ดังตารางที่ 99

ตารางที่ 99 ค่าความเที่ยงหรือสัมประสิทธิ์การสรุปอ้างอิงที่กำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และวิธีบุ๊คมาร์ค มีรูปแบบ i x r x o design

จำนวนผู้ตัดสิน	ค่าความเที่ยง หรือสัมประสิทธิ์สรุปอ้างอิง					
	การตัดสินครั้งที่ 1		การตัดสินครั้งที่ 2		การตัดสินครั้งที่ 3	
	แองกอฟ	บุ๊คมาร์ค	แองกอฟ	บุ๊คมาร์ค	แองกอฟ	บุ๊คมาร์ค
คะแนนจุดตัดระดับดีเยี่ยม (A)						
8	0.5529	0.9957	0.7120	0.9978	0.7877	0.9985
12	0.6177	0.9971	0.7637	0.9986	0.8290	0.9990
14	0.6391	0.9975	0.7798	0.9988	0.8416	0.9991
16	0.6562	0.9978	0.7924	0.9989	0.8513	0.9992
18	0.6701	0.9981	0.8025	0.9990	0.8590	0.9993
20	0.6817	0.9983	0.8107	0.9991	0.8653	0.9994
22	0.6914	0.9984	0.8176	0.9992	0.8705	0.9994
24	0.6998	0.9986	0.8234	0.9993	0.8749	0.9995

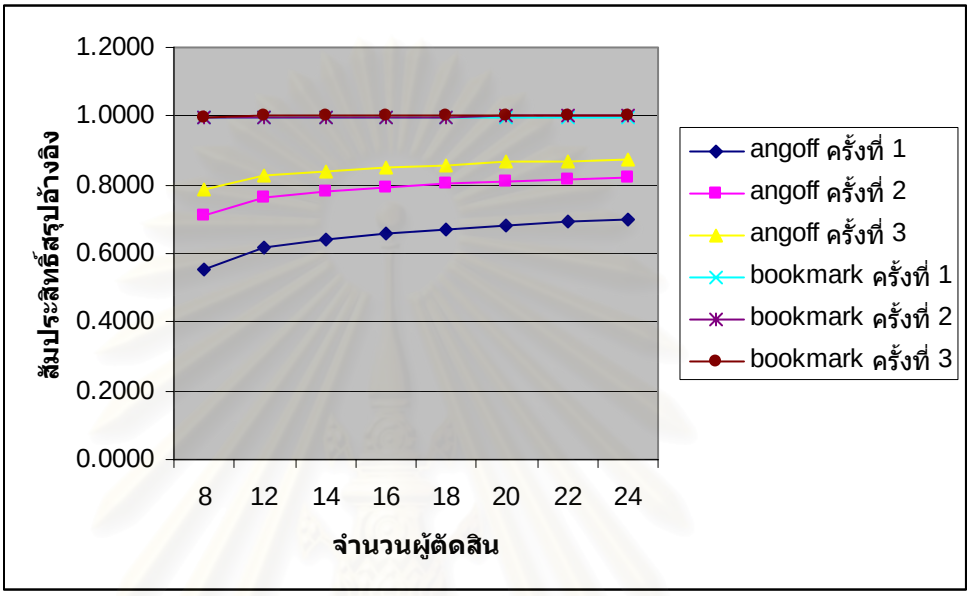
ตารางที่ 99 (ต่อ)

จำนวนผู้ตัดสิน	ค่าความเที่ยง หรือสัมประสิทธิ์สรุปอ้างอิง					
	การตัดสินครั้งที่ 1		การตัดสินครั้งที่ 2		การตัดสินครั้งที่ 3	
	แองกอฟ	บุ๊คมาร์ค	แองกอฟ	บุ๊คมาร์ค	แองกอฟ	บุ๊คมาร์ค
คะแนนจุดตัดระดับดีมาก (B+)						
8	0.6378	0.9983	0.7788	0.9991	0.8408	0.9994
12	0.7166	0.9988	0.8349	0.9994	0.8835	0.9996
14	0.7429	0.9990	0.8525	0.9995	0.8966	0.9996
16	0.7639	0.9991	0.8661	0.9995	0.9066	0.9997
18	0.7810	0.9992	0.8770	0.9996	0.9145	0.9997
20	0.7953	0.9993	0.8860	0.9996	0.9210	0.9997
22	0.8074	0.9994	0.8934	0.9997	0.9263	0.9998
24	0.8178	0.9994	0.8997	0.9997	0.9309	0.9998
คะแนนจุดตัดระดับดี (B)						
8	0.6446	0.9931	0.7839	0.9965	0.8447	0.9976
12	0.7129	0.9938	0.8324	0.9969	0.8817	0.9979
14	0.7352	0.9941	0.8474	0.9970	0.8928	0.9980
16	0.7528	0.9942	0.8590	0.9971	0.9014	0.9980
18	0.7671	0.9943	0.8682	0.9971	0.9081	0.9981
20	0.7790	0.9945	0.8758	0.9972	0.9136	0.9981
22	0.7890	0.9945	0.8820	0.9972	0.9181	0.9981
24	0.7975	0.9946	0.8873	0.9973	0.9219	0.9982
คะแนนจุดตัดระดับดีพอใช้ (C+)						
8	0.6739	0.9894	0.8052	0.9947	0.8611	0.9964
12	0.7328	0.9905	0.8458	0.9952	0.8917	0.9968
14	0.7516	0.9908	0.8582	0.9954	0.9008	0.9969
16	0.7664	0.9910	0.8677	0.9955	0.9078	0.9970
18	0.7782	0.9912	0.8753	0.9956	0.9132	0.9970
20	0.7880	0.9914	0.8814	0.9957	0.9177	0.9971
22	0.7961	0.9915	0.8865	0.9957	0.9214	0.9971
24	0.8031	0.9916	0.8908	0.9958	0.9244	0.9972

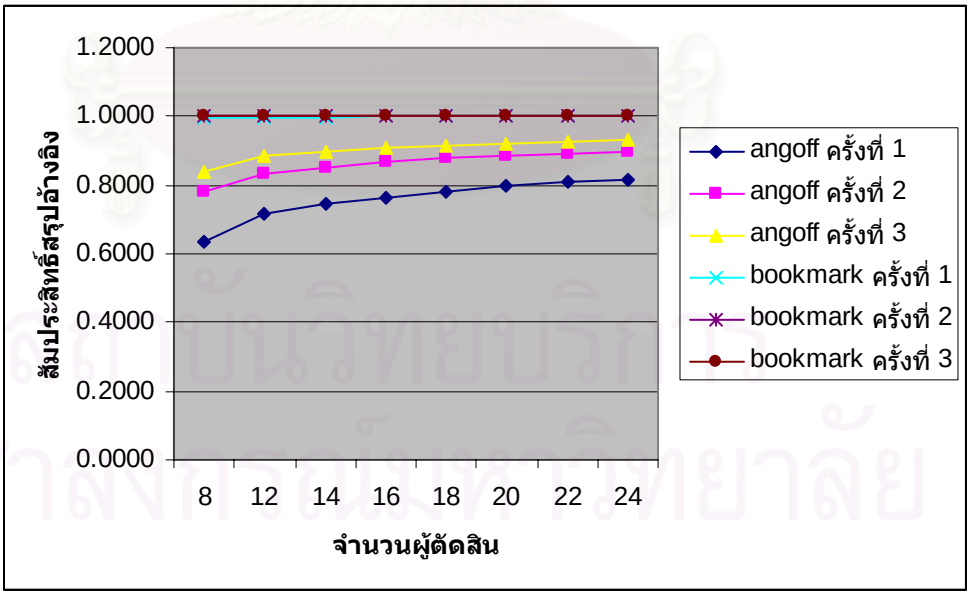
ตารางที่ 99 (ต่อ)

จำนวนผู้ตัดสิน	ค่าความเที่ยง หรือสัมประสิทธิ์สรุปอ้างอิง					
	การตัดสินครั้งที่ 1		การตัดสินครั้งที่ 2		การตัดสินครั้งที่ 3	
	แองกอฟ	บุ๊คมาร์ค	แองกอฟ	บุ๊คมาร์ค	แองกอฟ	บุ๊คมาร์ค
คะแนนจุดตัดระดับพอใช้ (C)						
8	0.6599	0.9975	0.7951	0.9988	0.8534	0.9992
12	0.7197	0.9980	0.8370	0.9990	0.8851	0.9993
14	0.7389	0.9981	0.8498	0.9991	0.8946	0.9994
16	0.7539	0.9982	0.8597	0.9991	0.9019	0.9994
18	0.7660	0.9983	0.8675	0.9991	0.9076	0.9994
20	0.7760	0.9984	0.8739	0.9992	0.9122	0.9995
22	0.7844	0.9984	0.8792	0.9992	0.9161	0.9995
24	0.7915	0.9984	0.8836	0.9992	0.9193	0.9995
คะแนนจุดตัดระดับอ่อน (D+)						
8	0.6359	0.9942	0.7741	0.9969	0.8346	0.9978
12	0.6647	0.9947	0.7963	0.9972	0.8525	0.9980
14	0.6734	0.9949	0.8028	0.9973	0.8577	0.9981
16	0.6801	0.9950	0.8078	0.9974	0.8617	0.9982
18	0.6854	0.9951	0.8117	0.9974	0.8649	0.9982
20	0.6897	0.9951	0.8149	0.9975	0.8674	0.9983
22	0.6933	0.9952	0.8175	0.9975	0.8695	0.9983
24	0.6963	0.9953	0.8197	0.9975	0.8712	0.9983
คะแนนจุดตัดระดับอ่อนมาก (D)						
8	0.7419	0.9843	0.8518	0.9921	0.8961	0.9947
12	0.7627	0.9860	0.8654	0.9929	0.9060	0.9953
14	0.7688	0.9864	0.8693	0.9932	0.9089	0.9954
16	0.7735	0.9868	0.8723	0.9934	0.9111	0.9956
18	0.7772	0.9871	0.8746	0.9935	0.9128	0.9957
20	0.7802	0.9873	0.8765	0.9936	0.9141	0.9957
22	0.7826	0.9875	0.8781	0.9937	0.9153	0.9958
24	0.7847	0.9876	0.8793	0.9938	0.9162	0.9958

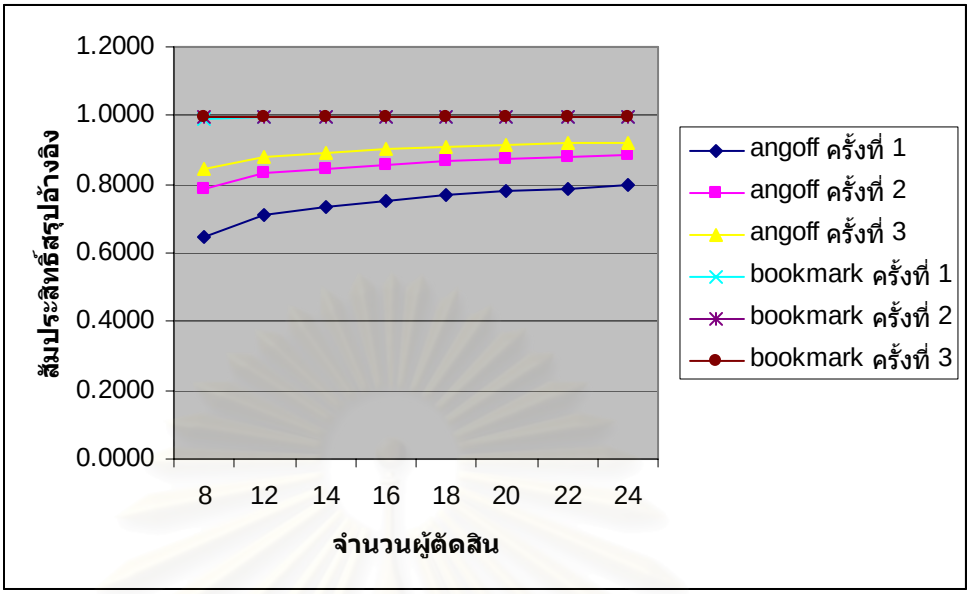
เพื่อแสดงให้เห็นการเปรียบเทียบระหว่างค่าความเที่ยง หรือสัมประสิทธิ์สรุปอ้างอิงของการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์ค ผู้วิจัยจึงนำเสนอการเปรียบเทียบเป็นกราฟเส้น ดังแสดงในแผนภาพต่อไปนี้



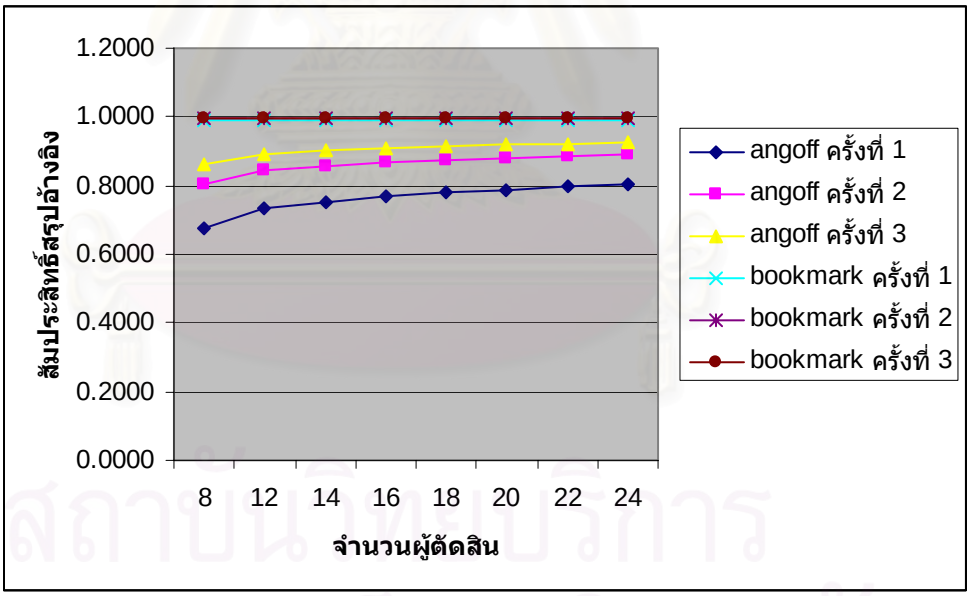
แผนภาพที่ 27 การเปรียบเทียบค่าสัมประสิทธิ์สรุปอ้างอิงของการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์ค ที่คะแนนจุดตัดระดับดีเยี่ยม (A)



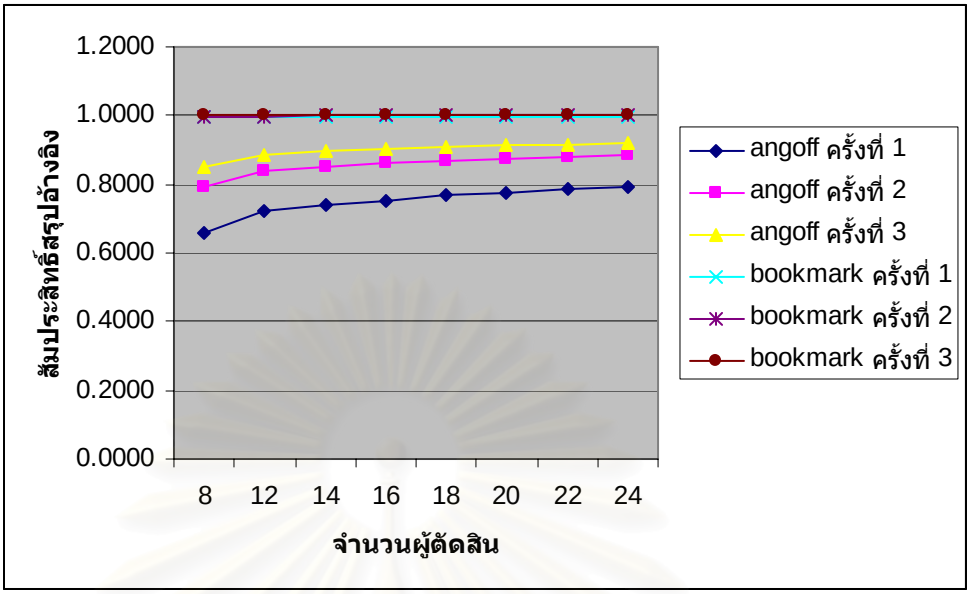
แผนภาพที่ 28 การเปรียบเทียบค่าสัมประสิทธิ์สรุปอ้างอิงของการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์ค ที่คะแนนจุดตัดระดับดีมาก (B+)



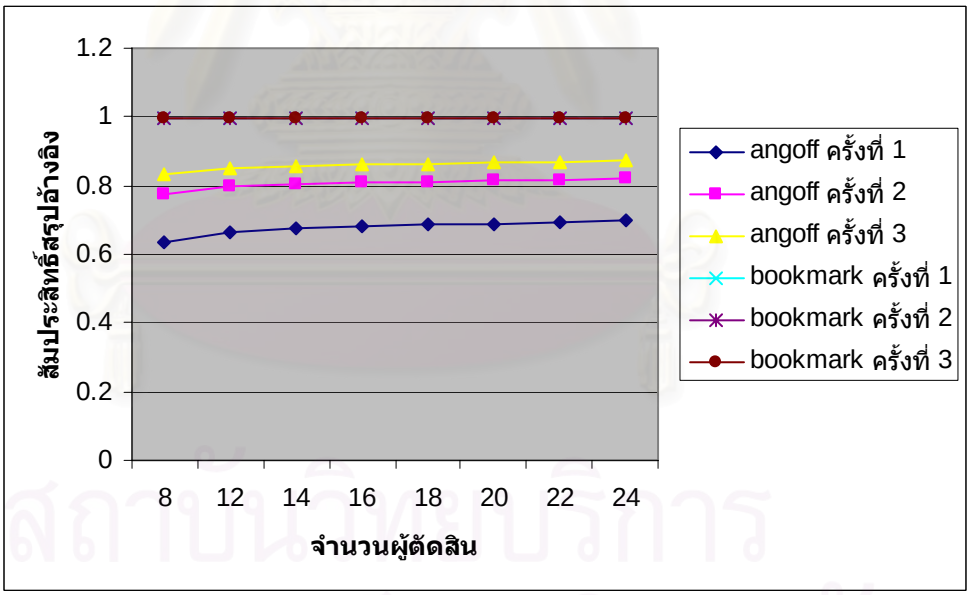
แผนภาพที่ 29 การเปรียบเทียบค่าสัมประสิทธิ์สรุปอ้างอิงของการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์ค ที่คะแนนจุดตัดระดับดี (B)



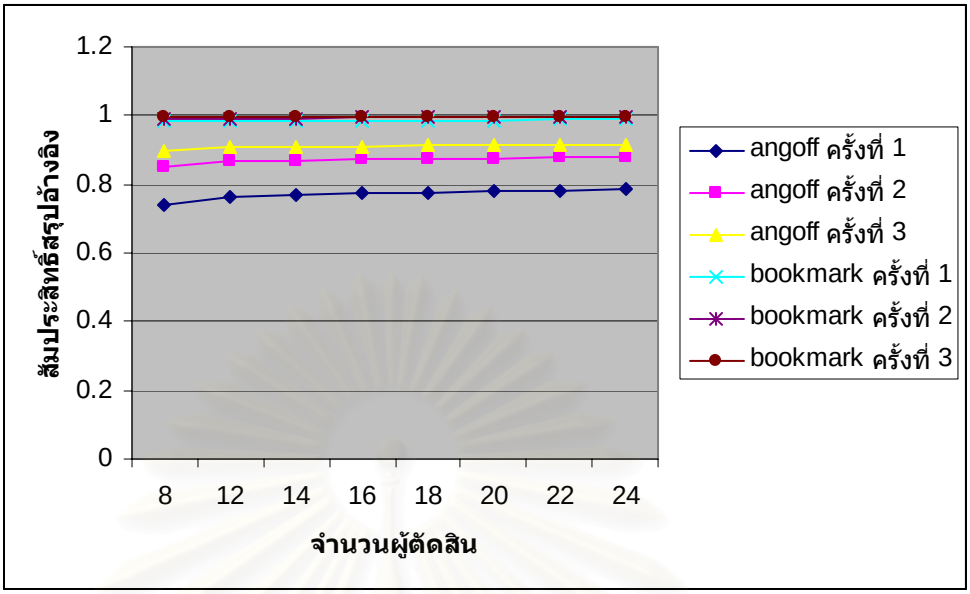
แผนภาพที่ 30 การเปรียบเทียบค่าสัมประสิทธิ์สรุปอ้างอิงของการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์ค ที่คะแนนจุดตัดระดับดีพอใช้ (C+)



แผนภาพที่ 31 การเปรียบเทียบค่าสัมประสิทธิ์สรุปอ้างอิงของการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์ค ที่คะแนนจุดตัดระดับพอใช้ (C)



แผนภาพที่ 32 การเปรียบเทียบค่าสัมประสิทธิ์สรุปอ้างอิงของการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์ค ที่คะแนนจุดตัดระดับอ่อน (D+)



แผนภาพที่ 33 การเปรียบเทียบค่าสัมประสิทธิ์สรุปอ้างอิงของการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์ค ที่คะแนนจุดตัดระดับอ่อนมาก (D)

จากตารางที่ 99 และแผนภาพที่ 27–33 พบว่า ในแต่ละระดับคะแนนจุดตัด เมื่อจำนวนผู้ตัดสินเพิ่มขึ้น และจำนวนครั้งเพิ่มขึ้นในการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และวิธีบุ๊คมาร์ค จะทำให้สัมประสิทธิ์สรุปอ้างอิง หรือค่าความเที่ยงเพิ่มขึ้นซึ่งมีค่าความเที่ยงระหว่าง 0.6–0.9

สำหรับการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง ถ้ากำหนดมาตรฐานซ้ำกันตั้งแต่ 2 ครั้งขึ้นไป และกำหนดจำนวนผู้ตัดสินตั้งแต่ 8 คนขึ้นไปจะได้ค่าความเที่ยงตั้งแต่ 0.8 ขึ้นไป แต่ถ้ากำหนดมาตรฐานเพียง 1 ครั้ง แต่ถ้าต้องการค่าความเที่ยงประมาณ 0.6–0.7 ก็ใช้จำนวนผู้ตัดสินประมาณ 8–18 คน โดยถ้าใช้ผู้ตัดสินตั้งแต่ 16 คนขึ้นไป จะได้ค่าความเที่ยง 0.7 ขึ้นไป และถ้าต้องการค่าความเที่ยงประมาณ 0.8 ควรใช้ผู้ตัดสินประมาณ 20 คนขึ้นไป

จะเห็นได้ว่า ถ้ากำหนดมาตรฐานเพียงครั้งเดียวจะใช้จำนวนผู้ตัดสินมากกว่า การกำหนดมาตรฐานตั้งแต่สองครั้งขึ้นไป จึงจะทำให้ได้ค่าความเที่ยงตั้งแต่ 0.7 ขึ้นไป

ส่วนการกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์ค การกำหนดมาตรฐานตั้งแต่ครั้งที่ 1 ขึ้นไป และใช้จำนวนผู้ตัดสินตั้งแต่ 8 คนขึ้นไป จะได้ค่าความเที่ยง 0.9 ดังนั้นการเพิ่มจำนวนผู้ตัดสินและจำนวนครั้งทำให้ค่าความเที่ยงเพิ่มเพียงเล็กน้อย นั่นคือในการกำหนดมาตรฐานเพียง 1 ครั้ง และใช้ผู้ตัดสินเพียง 8 คนก็ให้ค่าความเที่ยง 0.9 แล้ว

บทที่ 5

สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ

การวิจัยเรื่องการเปรียบเทียบคุณภาพของการกำหนดมาตรฐานระหว่างวิธีแองกอฟที่ได้รับการปรับปรุงกับวิธีบุ๊คมาร์ค มีวัตถุประสงค์การวิจัยคือ 1) เพื่อเปรียบเทียบคุณภาพของการกำหนดมาตรฐานระหว่างวิธีการแองกอฟที่ได้รับการปรับปรุงกับวิธีการบุ๊คมาร์ค โดยพิจารณาจากความตรง และความเที่ยง ความเหมาะสมและความเป็นไปได้ 2) เพื่อศึกษาผลของจำนวนผู้ตัดสินและจำนวนครั้งที่มีต่อการกำหนดมาตรฐานที่ตั้งขึ้นด้วยวิธีการแองกอฟที่ได้รับการปรับปรุง และวิธีการบุ๊คมาร์ค

กลุ่มตัวอย่างที่ใช้ในการศึกษาแบ่งออกเป็น 2 กลุ่ม คือ 1) **ผู้ตัดสิน** คือ ครูสอนวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 สังกัดสำนักงานการศึกษาขั้นพื้นฐานที่มีคุณสมบัติ 2 ประการดังนี้ ประการแรก มีความเชี่ยวชาญในเนื้อหาคณิตศาสตร์มัธยมศึกษาตอนต้น คือ จบการศึกษาระดับปริญญาตรีขึ้นไปในสาขาทางคณิตศาสตร์ ประการที่สอง มีความเชี่ยวชาญในการสอนและการวัดผล คือ มีประสบการณ์การสอนมาไม่น้อยกว่า 5 ปี และเป็นครูชำนาญการ หรือชำนาญการพิเศษ ผู้วิจัยใช้วิธีการสุ่มตัวอย่างอย่างง่ายได้ผู้ตัดสินจำนวน 12 คน 2) **ผู้สอบ** คือ นักเรียนชั้นมัธยมศึกษาปีที่ 3 ปีการศึกษา 2549 ในโรงเรียนสังกัดสำนักงานการศึกษาขั้นพื้นฐาน เขตพื้นที่การศึกษาชลบุรีเขต 1 โดยสุ่มตัวอย่างแบบหลายขั้นตอน (Multi-stage sampling) คือ การทดลองแบบสอบครั้งที่ 1 สุ่มตัวอย่างโรงเรียนโดยการสุ่มอย่างง่ายจำนวน 1 โรงเรียน จากนั้นสุ่มตัวอย่างห้องเรียนได้ห้องเรียน 1 ห้อง รวมกลุ่มตัวอย่างนักเรียนทั้งสิ้น 10 คน การทดลองแบบสอบครั้งที่ 2 สุ่มตัวอย่างโรงเรียนโดยการสุ่มแบบแบ่งชั้นตามขนาดของโรงเรียนคือขนาดเล็ก กลาง และใหญ่ จำนวน 4 โรงเรียน จากนั้นสุ่มตัวอย่างห้องเรียนในแต่ละโรงเรียนที่สุ่มมาโดยใช้วิธีการสุ่มตัวอย่างอย่างง่ายได้ห้องเรียนจำนวน 9 ห้องเรียน รวมกลุ่มตัวอย่างนักเรียนทั้งสิ้น 290 คน การทดลองแบบสอบครั้งที่ 3 สุ่มตัวอย่างโรงเรียนโดยการสุ่มแบบแบ่งชั้นตามขนาดของโรงเรียน จำนวน 8 โรงเรียน จากนั้นสุ่มตัวอย่างห้องเรียนโดยสุ่มตัวอย่างอย่างง่ายได้ห้องเรียนจำนวน 26 ห้องเรียน รวมกลุ่มตัวอย่างนักเรียนทั้งสิ้น 987 คน

เครื่องมือที่ใช้ในการวิจัยมีดังนี้ 1) แบบสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ภาคเรียนที่ 1 แบบเลือกตอบ 4 ตัวเลือก แบ่งเป็น 4 ฉบับ คือ ฉบับที่ 1 เรื่องพื้นที่และปริมาตร จำนวน 30 ข้อ ฉบับที่ 2 เรื่องกราฟ จำนวน 30 ข้อ ฉบับที่ 3 เรื่องระบบ

สมการเชิงเส้น จำนวน 16 ข้อ และฉบับที่ 4 ความคล้าย จำนวน 24 ข้อ รวมข้อสอบทั้งสิ้น 100 ข้อ 2) เครื่องมือเกี่ยวกับความเหมาะสมและความเป็นไปได้ในการกำหนดมาตรฐานด้วยวิธีการแองกอฟที่ได้รับการปรับปรุง และวิธีบุ๊คมาร์คจำนวน 3 ฉบับ คือ แบบสอบถามเกี่ยวกับความเหมาะสม และความเป็นไปได้ของการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และวิธีบุ๊คมาร์ค แบบสอบถามความคิดเห็นเกี่ยวกับกระบวนการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และวิธีบุ๊คมาร์ค และแบบสัมภาษณ์ผู้ตัดสินเกี่ยวกับกระบวนการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และวิธีบุ๊คมาร์ค 3). คู่มือวิธีการกำหนดมาตรฐานวิธีแองกอฟที่ได้รับการปรับปรุง และวิธีการบุ๊คมาร์ค สำหรับให้ผู้ตัดสินอ่าน

การเก็บรวบรวมข้อมูล ผู้วิจัยแบ่งการเก็บรวบรวมข้อมูลออกเป็น 2 ระยะ คือ **การเก็บข้อมูลระยะที่ 1** เป็นการดำเนินการเก็บข้อมูลกับผู้สอบ เพื่อพัฒนาแบบสอบวัดผลสัมฤทธิ์ทางการเรียนชั้นมัธยมศึกษาปีที่ 3 ดำเนินการเก็บรวบรวมข้อมูล 3 ครั้งดังนี้ ทดลองใช้แบบสอบครั้งที่ 1 เพื่อปรับปรุงภาษา คำชี้แจงในการสอบ และเวลาในการสอบ นำผลที่ได้ไปปรับปรุงข้อสอบ และกำหนดเวลาในการสอบที่เหมาะสม ทดลองใช้แบบสอบครั้งที่ 2 โดยใช้โปรแกรม CTIA เพื่อหาค่าสถิติพื้นฐาน ค่าความยาก อำนาจจำแนก และความเที่ยงแบบ KR-20 แล้วนำผลการวิเคราะห์มาปรับปรุงข้อสอบ ทดลองใช้แบบสอบครั้งที่ 3 โดยใช้โปรแกรม Xcalibre เพื่อหาคุณภาพของข้อสอบ และแบบสอบ คือค่าพารามิเตอร์ของข้อสอบคือ ค่าความยาก ค่าอำนาจจำแนก และค่าการเดา และฟังก์ชันสารสนเทศของแบบสอบ จากนั้นผู้วิจัยนำค่าพารามิเตอร์ของข้อสอบมาจัดทำเป็นสารสนเทศสำหรับผู้ตัดสินในการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และวิธีบุ๊คมาร์คต่อไป **การเก็บข้อมูลระยะที่ 2** เป็นการดำเนินการเก็บข้อมูลกับผู้ตัดสิน ผู้ตัดสินทั้ง 12 คน ดำเนินการกำหนดมาตรฐานทั้งสองวิธีคือ วิธีแองกอฟที่ได้รับการปรับปรุงและวิธีบุ๊คมาร์ค โดยดำเนินการกำหนดมาตรฐานทั้งหมด 3 ครั้ง แต่ละครั้งทำการตัดสินโดยมีรูปแบบดังนี้ ครั้งที่ 1 ตัดสินคะแนนจุดตัดด้วยวิธีแองกอฟที่ได้รับการปรับปรุงก่อน แล้วตามด้วยการตัดสินคะแนนจุดตัดด้วยวิธีบุ๊คมาร์ค ครั้งที่ 2 ตัดสินคะแนนจุดตัดด้วยวิธีบุ๊คมาร์คก่อน แล้วตามด้วยการตัดสินด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และ ครั้งที่ 3 ตัดสินคะแนนจุดตัดด้วยวิธีแองกอฟที่ได้รับการปรับปรุงก่อน แล้วตามด้วยการตัดสินด้วยวิธีบุ๊คมาร์ค

การวิเคราะห์ข้อมูล หาคุณภาพของการกำหนดมาตรฐานด้วยวิธีการแองกอฟที่ได้รับการปรับปรุง และวิธีการบุ๊คมาร์ค ดังนี้ ความเที่ยงของการกำหนดมาตรฐาน โดยการหาค่าสัมประสิทธิ์การสรุปอ้างอิง (Generalizability coefficient) ใช้โปรแกรม GENOVA แล้วทำการทดสอบความแตกต่างความเที่ยง โดยสถิติทดสอบ UX_1 ของวูดรัฟฟ์และ เฟลดท์ ความตรงของ

การกำหนดมาตรฐาน โดยการหาความสัมพันธ์ระหว่างระดับผลการเรียนที่นักเรียนได้รับจากโรงเรียน กับระดับผลการตัดสินที่กำหนดด้วยวิธีการแองกอฟที่ได้รับการปรับปรุง และวิธีบุ๊คมาร์ค คำนวณจากสูตรสหสัมพันธ์แบบ Spearman's Rho แล้วทำการทดสอบความแตกต่างความตรงด้วยสถิติทดสอบ t ของโฮเทลลิ่ง สำหรับความเหมาะสมและความเป็นไปได้ในการกำหนดมาตรฐานด้วยวิธีการแองกอฟที่ได้รับการปรับปรุง และวิธีบุ๊คมาร์ค โดยการนำผลการประเมินมาหาค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน มัธยฐาน พิสัยคลอไทล์ แล้วทำการทดสอบความแตกต่างความเหมาะสม และความเป็นไปได้ด้วยสถิติทดสอบ t แบบ dependent ส่วนผลการสัมภาษณ์มาวิเคราะห์เนื้อหา

สรุปผลการวิจัย

1. คะแนนจุดตัดของแบบสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ภาคเรียนที่ 1 เมื่อผู้ตัดสินกำหนดคะแนนจุดตัด 7 ระดับ ด้วยวิธีแองกอฟที่ได้รับการปรับปรุง ครั้งที่ 1 ระดับดีเยี่ยม (A) ระดับดีมาก (B+) ระดับดี (B) ระดับดีพอใช้ (C+) ระดับพอใช้ (C) ระดับอ่อน (D+) และระดับอ่อนมาก (D) เท่ากับ 72.86, 64.31, 56.18, 43.37, 33.59, 21.34 และ 12.98 ตามลำดับ ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 16.256, 12.433, 8.880, 9.359, 9.712, 8.016 และ 3.820 ตามลำดับ ครั้งที่ 2 เท่ากับ 73.28, 63.74, 55.28, 45.11, 35.67, 22.94, 12.51 ตามลำดับ ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 10.787, 9.609, 8.663, 8.977, 8.580, 8.142 และ 4.746 ตามลำดับ ครั้งที่ 3 เท่ากับ 73.49, 63.25, 52.82, 52.82, 41.46, 31, 19.07, 11.72 ตามลำดับ ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 10.376, 10.149, 8.880, 8.437, 8.651, 3.551, 4.345 ตามลำดับ สำหรับคะแนนความสามารถ ครั้งที่ 1 เท่ากับ 2.1351, 1.6711, 1.3070, 0.7448, 0.2897, -0.4222, -1.2770 ตามลำดับ ครั้งที่ 2 เท่ากับ 2.1538, 1.6550, 1.2702, 0.8248, 0.3929, -0.3114, -1.2193 ตามลำดับ ครั้งที่ 3 เท่ากับ 2.1783, 1.6394, 1.1643, 0.6617, 0.1611, -0.5878 ตามลำดับ

2. คะแนนจุดตัดของแบบสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ภาคเรียนที่ 1 เมื่อผู้ตัดสินกำหนดคะแนนจุดตัด 7 ระดับ ด้วยวิธีบุ๊คมาร์ค ครั้งที่ 1 ระดับดีเยี่ยม (A) ระดับดีมาก (B+) ระดับดี (B) ระดับดีพอใช้ (C+) ระดับพอใช้ (C) ระดับอ่อน (D+) และระดับอ่อนมาก (D) เท่ากับ 80.42, 65.75, 47.5, 33.17, 24, 14.75, 7.83 ตามลำดับ ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1.037, 0.433, 0.764, 2.577, 0.000, 0.433, 0.553 ตามลำดับ ครั้งที่ 2 เท่ากับ 80.58, 66.67, 47.5, 33.75, 24.58, 14.75, 7.83 ตามลำดับ ส่วน

เบี่ยงเบนมาตรฐานเท่ากับ 0.862, 1.187, 1.041, 0.829, 1.441, 0.433, 0.553 ตามลำดับ ครั้งที่ 3 เท่ากับ 80.92, 64.08, 47.33, 34, 24.08, 14.83, 7.92 ตามลำดับ ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.954, 0.640, 0.764, 0.829, 1.891, 0.553, 0.640 ตามลำดับ สำหรับคะแนนความสามารถ ครั้งที่ 1 เท่ากับ 2.4043, 2.0453, 1.7460, 1.5054, 1.3030, 0.8590, 0.2060 ตามลำดับ ครั้งที่ 2 เท่ากับ 2.3816, 2.0392, 1.7434, 1.5148, 1.3161, 0.8590, 0.2060 ตามลำดับ ครั้งที่ 3 เท่ากับ 2.4168, 2.0429, 1.7460, 1.5148, 1.3267, 0.8601, 0.2218 ตามลำดับ

3. ค่าสัมประสิทธิ์การสรุปอ้างอิงหรือความเที่ยงในการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง

3.1 ค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินแบบสมบูรณ์หรือความเที่ยงในการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง เมื่อมีผู้ตัดสินจำนวน 8 , 12, 14, 16, 18, 20, 22, 24 คน ครั้งที่ 1 ระดับดีเยี่ยม (A) คือ 0.5762 0.6710 0.7041 0.7311 0.7537 0.7727 0.7890 0.8031 ตามลำดับ ระดับดีมาก (B+) คือ 0.5852 0.6791 0.7117 0.7383 0.7604 0.7791 0.7950 0.8089 ตามลำดับ ระดับดี คือ 0.6537 0.7390 0.7676 0.7906 0.8094 0.8251 0.8385 0.8499 ตามลำดับ ระดับดีพอใช้ (C+) คือ 0.7642 0.8294 0.8501 0.8664 0.8794 0.8902 0.8991 0.9068 ตามลำดับ ระดับพอใช้ (C) คือ 0.7739 0.8370 0.8569 0.8725 0.8851 0.8953 0.9039 0.9112 ตามลำดับ ระดับอ่อน (D+) คือ 0.8212 0.8732 0.8893 0.9018 0.9118 0.9199 0.9266 0.9323 ตามลำดับ ระดับอ่อนมาก (D) คือ 0.9198 0.9451 0.9525 0.9582 0.9627 0.9663 0.9693 0.9718 ตามลำดับ

3.2 ค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินแบบสมบูรณ์หรือความเที่ยงในการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง ครั้งที่ 2 ระดับดีเยี่ยม (A) คือ 0.7331 0.8047 0.8278 0.8460 0.8607 0.8729 0.8831 0.8918 ตามลำดับ ระดับดีมาก (B+) คือ 0.7575 0.8241 0.8454 0.8620 0.8755 0.8865 0.8957 0.9036 ตามลำดับ ระดับดี (B) คือ 0.5564 0.6529 0.6870 0.7150 0.7383 0.7582 0.7752 0.7900 ตามลำดับ ระดับดีพอใช้ (C+) คือ 0.7903 0.8497 0.8683 0.8829 0.8945 0.9040 0.9120 0.9187 ตามลำดับ ระดับพอใช้ (C) คือ 0.7776 0.8399 0.8595 0.8749 0.8872 0.8974 0.9058 0.9130 ตามลำดับ ระดับอ่อน (D+) คือ 0.8909 0.9245 0.9346 0.9423 0.9484 0.9533 0.9574 0.9608 ตามลำดับ ระดับอ่อนมาก (D) คือ 0.8472 0.8927 0.9066 0.9173 0.9258 0.9327 0.9385 0.9433 ตามลำดับ

3.3 ค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินแบบสมบูรณ์หรือความเที่ยงในการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง เมื่อมีผู้ตัดสินจำนวน 8 , 12, 14, 16, 18, 20, 22, 24 คน ครั้งที่ 3 ระดับดีเยี่ยม (A คือ 0.7342 0.8055 0.8286 0.8467 0.8614 0.8735 0.8837 0.8923 ตามลำดับ ระดับดีมาก (B+) คือ 0.7529 0.8205 0.8421 0.8590 0.8727 0.8840 0.8834 0.9014 ตามลำดับ ระดับดี (B) คือ 0.7667 0.8314 0.8519 0.8680 0.8809 0.8915 0.9004 0.9079 ตามลำดับ ระดับดีพอใช้ (C+) คือ 0.7899 0.8494 0.8681 0.8827 0.8943 0.9039 0.9118 0.9186 ตามลำดับ ระดับพอใช้ (C) คือ 0.7618 0.8275 0.8484 0.8648 0.8780 0.8889 0.8979 0.9056 ตามลำดับ ระดับอ่อน (D+) คือ 0.8741 0.9124 0.9240 0.9328 0.9398 0.9455 0.9502 0.9542 ตามลำดับ ระดับอ่อนมาก (D) คือ 0.8449 0.8909 0.9050 0.9159 0.9246 0.9316 0.9374 0.9423 ตามลำดับ

4. ค่าสัมประสิทธิ์การสรุปอ้างอิงหรือความเที่ยงในการกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์ค

4.1 ค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินแบบสมบูรณ์หรือความเที่ยงในการกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์ค ครั้งที่ 1 ระดับดีเยี่ยม (A) คือ 0.9954 0.9969 0.9974 0.9977 0.9980 0.9982 0.9983 0.9985 ตามลำดับ ระดับดีมาก (B+) คือ 0.9989 0.9992 0.9994 0.9994 0.9995 0.9996 0.9996 0.9996 ตามลำดับ ระดับดี (B) คือ 0.9980 0.9987 0.9987 0.9990 0.9991 0.9992 0.9993 0.9993 ตามลำดับ ระดับดีพอใช้ (C+) คือ 0.9931 0.9954 0.9961 0.9966 0.9969 0.9972 0.9975 0.9977 ตามลำดับ ระดับพอใช้ (C) คือ 1.0000 ทุกค่า ระดับอ่อน (D+) คือ 0.9980 0.9986 0.9988 0.9990 0.9991 0.9992 0.9993 0.9993 ตามลำดับ ระดับอ่อนมาก (D) คือ 0.9950 0.9966 0.9971 0.9974 0.9977 0.9980 0.9981 0.9983 ตามลำดับ

4.2 ค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินแบบสมบูรณ์หรือความเที่ยงในการกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์ค เมื่อมีผู้ตัดสินจำนวน 8 , 12, 14, 16, 18, 20, 22, 24 คน ครั้งที่ 2 ระดับดีเยี่ยม (A) คือ 0.9960 0.9973 0.9977 0.9980 0.9982 0.9984 .9985 0.9986 ตามลำดับ ระดับดีมาก (B+) คือ 0.9890 0.9926 0.9937 0.9945 0.9951 0.9956 0.9960 0.9963 ตามลำดับ ระดับดี (B) คือ 0.9972 0.9981 0.9984 0.9986 0.9987 0.9989 0.9990 0.9991 ตามลำดับ ระดับดีพอใช้ (C+) คือ 0.9970 0.9980 0.9983 0.9985 0.9987 0.9988 0.9989 0.9990 ตามลำดับ ระดับพอใช้ (C) คือ 0.9962 0.9975 0.9978 0.9981 0.9983 0.9985 0.9986 0.9987 ตามลำดับ ระดับอ่อน (D+) คือ 0.9980 0.9986

0.9988 0.9990 0.9991 0.9992 0.9993 0.9993 ตามลำดับ ระดับอ่อนมาก (D) คือ 0.9949 0.9966 0.9971 0.9974 0.9977 0.9980 0.9981 0.9983 ตามลำดับ

4.3 ค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินแบบสมบูรณ์หรือความเที่ยงในการกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์ค เมื่อมีผู้ตัดสินจำนวน 8 , 12, 14, 16, 18, 20, 22, 24 คน ครั้งที่ 3 ระดับดีเยี่ยม (A) คือ 0.9958 0.9972 0.9976 0.9979 0.9981 0.9983 0.9985 0.9986 ตามลำดับ ระดับดีมาก (B+) คือ 0.9895 0.9930 0.9940 0.9947 0.9953 0.9958 0.9961 0.9965 ตามลำดับ ระดับดี (B) คือ 0.9963 0.9975 0.9979 0.9981 0.9984 0.9985 0.9987 0.9988 ตามลำดับ ระดับดีพอใช้ (C+) คือ 0.9973 0.9982 0.9984 0.9986 0.9988 0.9989 0.9990 0.9991 ตามลำดับ ระดับพอใช้ (C) คือ 0.9994 0.9996 0.9997 0.9997 0.9998 0.9998 0.9998 0.9998 ตามลำดับ ระดับอ่อน (D+) คือ 0.9971 0.9981 0.9984 0.9986 0.9987 0.9989 0.9990 0.9990 ตามลำดับ ระดับอ่อนมาก (D) คือ 0.9937 0.9958 0.9964 0.9968 0.9972 0.9975 0.9977 0.9979 ตามลำดับ

5 ค่าความตรงตามสภาพ โดยหาจากความสัมพันธ์ระหว่างระดับผลการเรียนที่นักเรียนได้รับจากโรงเรียน กับระดับผลการตัดสินที่กำหนดด้วยวิธีการแองกอฟที่ได้รับการปรับปรุง และวิธีบุ๊คมาร์ค พบว่า ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างระดับผลการเรียนที่นักเรียนได้รับจากโรงเรียน กับระดับผลการตัดสินที่กำหนดด้วยวิธีการแองกอฟที่ได้รับการปรับปรุง ครั้งที่ 1 ครั้งที่ 2 และครั้งที่ 3 เท่ากับ .661 .675 และ .678 อย่างมีนัยสำคัญที่ระดับ .01 ตามลำดับ ส่วนค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างระดับผลการเรียนที่นักเรียนได้รับจากโรงเรียน กับระดับผลการตัดสินที่กำหนดด้วยวิธีการบุ๊คมาร์ค ครั้งที่ 1 ครั้งที่ 2 และครั้งที่ 3 เท่ากับ .533 ทุกครั้ง อย่างมีนัยสำคัญที่ระดับ .01 นั้นคือ วิธีแองกอฟที่ได้รับการปรับปรุง มีค่าความตรงสูงกว่าวิธีการบุ๊คมาร์คเล็กน้อย และเมื่อทดสอบความแตกต่างที่ระดับนัยสำคัญทางสถิติ.05 ด้วยสถิติทดสอบ t แบบ dependent พบว่า ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างระดับผลการเรียนที่นักเรียนได้รับจากโรงเรียน กับระดับมาตรฐานที่กำหนดด้วยวิธีการแองกอฟที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์คทั้ง 3 ครั้งไม่แตกต่างกัน

6. ผลการวิเคราะห์ความเหมาะสม และความเป็นไปได้ของการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และวิธีบุ๊คมาร์ค

6.1 ค่าเฉลี่ยความคิดเห็นเกี่ยวกับกระบวนการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุงของผู้ตัดสินเท่ากับ 4.16 ส่วนเบี่ยงเบนมาตรฐาน 0.298 แสดงว่า ผู้ตัดสินมี

ความคิดเห็นในทางบวกต่อการตัดสินคะแนนจุดตัดด้วยวิธีแองกอฟที่ได้รับการปรับปรุง ในภาพรวมอยู่ในระดับสูง ส่วนค่าเฉลี่ยความคิดเห็นเกี่ยวกับกระบวนการกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์คเท่ากับ 4.46 ส่วนเบี่ยงเบนมาตรฐาน เท่ากับ 0.313 แสดงว่า ผู้ตัดสินมีความคิดเห็นในทางบวกต่อการตัดสินคะแนนจุดตัดด้วยวิธีบุ๊คมาร์คในภาพรวมระดับสูง

6.2 ด้านความเหมาะสม ในภาพรวมผู้ตัดสินมีความคิดเห็นว่าวิธีการกำหนดคะแนนจุดตัดด้วยวิธีแองกอฟที่ได้รับการปรับปรุงมีความเหมาะสมระดับมาก (Mdn= 4.08) ส่วนการกำหนดคะแนนจุดตัดด้วยวิธีบุ๊คมาร์ค ในภาพรวมผู้ตัดสินมีความคิดเห็นว่าเป็นวิธีที่มีความเหมาะสมระดับมากที่สุด (Mdn= 4.69) เมื่อเปรียบเทียบค่าเฉลี่ยความคิดเห็นด้านความเหมาะสมระหว่างสองวิธี พบว่า ค่าเฉลี่ยความเหมาะสมของวิธีบุ๊คมาร์คมากกว่าค่าเฉลี่ยความเหมาะสมของวิธีแองกอฟที่ได้รับการปรับปรุงอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

6.3 ด้านความเป็นไปได้ ในภาพรวมผู้ตัดสินมีความคิดเห็นว่าวิธีการกำหนดคะแนนจุดตัดด้วยวิธีแองกอฟที่ได้รับการปรับปรุงมีความเป็นไประดับมาก (Mdn= 4.19) ส่วนการกำหนดคะแนนจุดตัดด้วยวิธีบุ๊คมาร์ค ในภาพรวมผู้ตัดสินมีความคิดเห็นว่าเป็นวิธีที่มีความเป็นไปได้ระดับมากที่สุด (Mdn= 4.58) เมื่อเปรียบเทียบค่าเฉลี่ยความคิดเห็นด้านความเป็นไปได้ระหว่างสองวิธี พบว่า ค่าเฉลี่ยความเป็นไปได้ของวิธีบุ๊คมาร์คมากกว่าค่าเฉลี่ยความเป็นไปได้ของวิธีแองกอฟที่ได้รับการปรับปรุงอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

7. ผลการเปรียบเทียบสัมประสิทธิ์ความเที่ยงของการกำหนดคะแนนจุดตัดด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และวิธีบุ๊คมาร์ค จำแนกตามจำนวนผู้ตัดสิน พบว่า ค่าความเที่ยงตามทฤษฎีสรุปอ้างอิงระหว่างวิธีแองกอฟที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์ค แต่ละระดับคือ ระดับดีเยี่ยม (A) ระดับดีมาก (B+) ระดับดี (B) ระดับดีพอใช้ (C+) ระดับพอใช้ (C) ระดับอ่อน (D+) และระดับอ่อนมาก (D) เมื่อทำการตัดสินครั้งที่ 1 ครั้งที่ 2 และครั้ง 3 ที่มีจำนวนผู้ตัดสิน 8 12 14 16 18 20 22 24 พบว่า มีค่าแตกต่างกันที่ระดับนัยสำคัญ .01 ทุกค่า

8. ผลของจำนวนผู้ตัดสิน และจำนวนครั้งที่มีต่อการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และวิธีบุ๊คมาร์ค ผู้วิจัยออกแบบการวัดเป็นแบบ Two-facet crossed design ฟาเซท (facets) ของการวัด มี 2 ฟาเซท คือ ผู้ตัดสิน (r) กำหนดจำนวนระดับเป็น 8, 12, 14, 16, 18, 20, 22, 24 คน และจำนวนครั้ง (o) กำหนดจำนวนระดับเป็น 1, 2, 3 ครั้ง รูปแบบของความสัมพันธ์ระหว่างฟาเซทเป็น ixrxo ผลการวิเคราะห์พบว่า การกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับ

การปรับปรุงมีค่าความเที่ยงระหว่าง 0.6–0.9 ส่วนวิธีการบุ๊คมาร์คมีค่าความเที่ยงประมาณ 0.9 . ในทุกระดับคะแนนจุดตัด และทั้ง 3 ครั้ง

สำหรับการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง ถ้ากำหนดมาตรฐานซ้ำกันตั้งแต่ 2 ครั้งขึ้นไป และกำหนดจำนวนผู้ตัดสินตั้งแต่ 8 คนขึ้นไปจะได้ค่าความเที่ยงตั้งแต่ 0.8 ขึ้นไป แต่ถ้ากำหนดมาตรฐานเพียง 1 ครั้งควรใช้ผู้ตัดสินตั้งแต่ 16 คนขึ้นไป จึงจะได้ค่าความเที่ยง 0.7 ขึ้นไป จะเห็นได้ว่า ถ้ากำหนดมาตรฐานเพียงครั้งเดียวจะใช้จำนวนผู้ตัดสินมากกว่าการกำหนดมาตรฐานตั้งแต่สองครั้งขึ้นไป จึงจะทำให้ได้ค่าความเที่ยงตั้งแต่ 0.7 ขึ้นไป

ส่วนการกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์ค การกำหนดมาตรฐานตั้งแต่ครั้งที่ 1 ขึ้นไป และใช้จำนวนผู้ตัดสินตั้งแต่ 8 คนขึ้นไป จะได้ค่าความเที่ยง 0.9 ดังนั้นการเพิ่มจำนวนผู้ตัดสินและจำนวนครั้งทำให้ค่าความเที่ยงเพิ่มเพียงเล็กน้อย นั้นคือในการกำหนดมาตรฐานเพียง 1 ครั้ง และใช้ผู้ตัดสินเพียง 8 คนก็ให้ค่าความเที่ยง 0.9 แล้ว

อภิปรายผลการวิจัย

1. การกำหนดคะแนนจุดตัดด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และวิธีบุ๊คมาร์คโดยประยุกต์ทฤษฎีการสรุปอ้างอิง

1.1 ผลคะแนนจุดตัดของแบบสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ชั้นมัธยมศึกษาปีที่ 3 ภาคเรียนที่ 1 เมื่อผู้ตัดสินกำหนดคะแนนจุดตัดด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และวิธีการบุ๊คมาร์ค จำนวน 3 ครั้ง 7 ระดับคือ ระดับดีเยี่ยม (A) ระดับดีมาก (B+) ระดับดี (B) ระดับดีพอใช้ (C+) ระดับพอใช้ (C) ระดับอ่อน (D+) และระดับอ่อนมาก (D) พบว่า คะแนนจุดตัดที่กำหนดด้วยวิธีบุ๊คมาร์ค ระดับ A และ B+ สูงกว่าวิธีแองกอฟที่ได้รับการปรับปรุง โดยระดับ A คะแนนจุดตัดที่กำหนดด้วยวิธีบุ๊คมาร์คสูงกว่าคะแนนจุดตัดที่กำหนดด้วยวิธีแองกอฟประมาณ 7.3% – 7.56% และระดับ B คะแนนจุดตัดที่กำหนดด้วยวิธีบุ๊คมาร์คสูงกว่าคะแนนจุดตัดที่กำหนดด้วยวิธีแองกอฟประมาณ 0.83% – 2.93%

และในทางตรงข้ามพบว่า คะแนนจุดตัดที่กำหนดด้วยวิธีแองกอฟที่ได้รับการปรับปรุง ระดับดี (B) ระดับดีพอใช้ (C+) ระดับพอใช้ (C) ระดับอ่อน (D+) และระดับอ่อนมาก (D) สูงกว่าคะแนนจุดตัดที่กำหนดด้วยวิธีบุ๊คมาร์ค โดยระดับดี (B) คะแนนจุดตัดที่กำหนดด้วยวิธีแองกอฟที่ได้รับการปรับปรุงสูงกว่าคะแนนจุดตัดที่กำหนดด้วยวิธีบุ๊คมาร์ค ประมาณ 4.13%–7.78% ระดับดีพอใช้ (C+) คะแนนจุดตัดสูงกว่าประมาณ 7.46%–11.36% ระดับพอใช้ (C)

คะแนนจุดตัดสูงกว่าประมาณ 6.92% – 11.09% ระดับอ่อน (D+) คะแนนจุดตัดสูงกว่าประมาณ 4.24% – 8.19% และระดับอ่อนมาก (D) คะแนนจุดตัดสูงกว่าประมาณ 3.8% – 5.15%

จะเห็นได้ว่าการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และวิธีบุ๊คมาร์คให้คะแนนจุดตัดที่แตกต่างกัน ซึ่งเป็นไปตาม มิทเซล (Mitzel etal, 2001) กล่าวว่า ข้อค้นพบส่วนใหญ่ในเอกสารงานวิจัยที่เกี่ยวกับการกำหนดมาตรฐาน ที่มีกระบวนการกำหนดมาตรฐานแตกต่างกัน จะได้คะแนนจุดตัดต่างกันเช่น งานวิจัยของแวง (Wang, 2003) พบว่า การกำหนดคะแนนจุดตัดด้วยวิธีแองกอฟมีคะแนนจุดตัดสูงกว่าวิธีผังข้อสอบและมีคะแนนจุดตัดแตกต่างกัน 9% ถึง 19% จากการสอบ 4 ครั้ง ซึ่งสอดคล้องกับผลการวิจัยนี้ซึ่งพบว่า คะแนนจุดตัดที่กำหนดด้วยวิธีแองกอฟที่ได้รับการปรับปรุงสูงกว่าวิธีบุ๊คมาร์คในระดับดี (B) ระดับดีพอใช้ (C+) ระดับพอใช้ (C) ระดับอ่อน (D+) และระดับอ่อนมาก (D) แต่เปอร์เซ็นต์ความแตกต่างของคะแนนจุดตัดในงานวิจัยนี้แตกต่างน้อยกว่างานวิจัยของแวง อาจเป็นเพราะว่าในการกำหนดคะแนนจุดตัดในงานวิจัยของแวง การตัดสินการสอบทั้ง 4 ครั้ง ใช้ผู้ตัดสินคนละกลุ่มกัน และแบบสอบคนละฉบับที่มีความยาวแตกต่างกัน แต่งานวิจัยนี้ใช้ผู้ตัดสินกลุ่มเดียวกัน และแบบสอบฉบับเดียวกัน

1.2 เมื่อพิจารณาส่วนเบี่ยงเบนมาตรฐานของคะแนนจุดตัดของผู้ตัดสินทั้ง 12 คน เมื่อกำหนดคะแนนจุดตัด 7 ระดับ พบว่าส่วนเบี่ยงเบนมาตรฐานของคะแนนจุดตัดที่กำหนดด้วยวิธีแองกอฟที่ได้รับการปรับปรุงมีค่าระหว่าง 3.551-16.256 ในขณะที่ส่วนเบี่ยงเบนมาตรฐานของคะแนนจุดตัดที่กำหนดด้วยวิธีบุ๊คมาร์คมีค่าระหว่าง 0.000–2.577 จะเห็นได้ว่าวิธีการกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์คมีส่วนเบี่ยงเบนมาตรฐานต่ำกว่าวิธีการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุงค่อนข้างมาก แสดงให้เห็นว่า ผู้ตัดสินมีความสอดคล้องกันสูงในการกำหนดคะแนนจุดตัดด้วยวิธีบุ๊คมาร์ค ดังที่เบอร์ก (Berk, 1995) กล่าวว่า ความแปรผันระหว่างผู้ตัดสินต่ำเป็นผลทางบวก

1.3 เปอร์เซนต์การผ่านของนักเรียนในแต่ละระดับมาตรฐานที่กำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และการกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์ค พบว่า การกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุงทั้ง 3 ครั้ง มีเปอร์เซ็นต์การผ่านของนักเรียนแต่ละระดับดังนี้ ระดับดีเยี่ยม (A) ทั้ง 3 ครั้งมีจำนวน 3% ระดับดีมาก (B+) ทั้ง 3 ครั้งมีจำนวน 3% ระดับดี (B) อยู่ในช่วง 4% - 6% ระดับดีพอใช้ (C+) อยู่ในช่วง 9% - 13% ระดับพอใช้ (C) อยู่ในช่วง 16% - 24% ระดับอ่อน (D+) อยู่ในช่วง 40% - 43% ระดับอ่อนมาก (D) อยู่ในช่วง 11% - 21% และระดับตก (F) ทุกครั้งมีจำนวน 0.4% ส่วนการกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์คทั้ง 3 ครั้ง มีเปอร์เซ็นต์การผ่านของนักเรียนแต่ละระดับดังนี้ ระดับดีเยี่ยม (A) ทั้ง 3 ครั้งมีจำนวน 1%

ระดับดีมาก (B+) ทั้ง 3 ครั้งมีจำนวน 5% ระดับดี (B) อยู่ในช่วง 11% - 12% ระดับดีพอใช้ (C+) ทั้ง 3 ครั้งมีจำนวน 23% ระดับพอใช้ (C) อยู่ในช่วง 30% - 35% ระดับอ่อน (D+) อยู่ในช่วง 24% - 28% ระดับอ่อนมาก (D) ทั้ง 3 ครั้งมีจำนวน 1% และ ระดับตก (F) ไม่มีนักเรียนตก

เมื่อพิจารณาจะเห็นว่า เปอร์เซนต์การผ่านของนักเรียนที่กำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง ส่วนใหญ่นักเรียนเกือบครึ่งหนึ่งคือ ประมาณร้อยละ 40 อยู่ระดับอ่อน (D+) ส่วนการกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์ค นักเรียนส่วนใหญ่อยู่ระดับพอใช้ (C) ประมาณร้อยละ 30 และยังมีข้อสังเกตอีกว่า เมื่อรวมเปอร์เซ็นต์ของนักเรียนที่อยู่ระดับพอใช้ (C) ระดับอ่อน (D+) และระดับอ่อนมาก (D) มีประมาณ 60% ในขณะที่การกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์ค เมื่อรวมเปอร์เซ็นต์ของนักเรียนที่อยู่ระดับดีพอใช้ (C+) ระดับพอใช้ (C) และระดับอ่อน (D+) มีเกือบประมาณ 80% จากที่กล่าวมาจะเห็นว่า การกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์ค นักเรียนส่วนใหญ่อยู่ในระดับมาตรฐานที่ดีกว่าการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง

2. **ความเที่ยงของการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุงและวิธีบุ๊คมาร์ค**

ผู้วิจัยจำนวน 12 คนทำการกำหนดคะแนนจุดตัด 7 ระดับ ทำให้ได้มาตรฐาน 8 ระดับขึ้นมาโดยใช้วิธีการแองกอฟที่ได้รับการปรับปรุง และวิธีการบุ๊คมาร์ค จากผลการวิเคราะห์ความแปรปรวนของคะแนนจุดตัดแต่ละระดับที่กำหนดด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และวิธีบุ๊คมาร์ค ทั้ง 3 ครั้ง พบว่า วิธีแองกอฟที่ได้รับการปรับปรุง **ระดับดีเยี่ยม** (A) ความแปรปรวนของข้อสอบประมาณ 14% - 25% ของความแปรปรวนรวม ความแปรปรวนของผู้ตัดสินประมาณ 39% - 65% ของความแปรปรวนรวม และความแปรปรวนของผลร่วมระหว่างข้อสอบกับผู้ตัดสินประมาณ 20% - 35% ของความแปรปรวนรวม **ระดับดีมาก** (B+) ความแปรปรวนของข้อสอบประมาณ 15%-28% ของความแปรปรวนรวม ความแปรปรวนของผู้ตัดสินประมาณ 36% - 68% ของความแปรปรวนรวม และความแปรปรวนของผลร่วมระหว่างข้อสอบกับผู้ตัดสินประมาณ 17% - 36% ของความแปรปรวนรวม **ระดับดี** (B) ความแปรปรวนของข้อสอบประมาณ 14% - 29% ของความแปรปรวนรวม ความแปรปรวนของผู้ตัดสินประมาณ 19% - 50% ของความแปรปรวนรวม และความแปรปรวนของผลร่วมระหว่างข้อสอบกับผู้ตัดสินประมาณ 31% - 67% ของความแปรปรวนรวม **ระดับดีพอใช้** (C+) ความแปรปรวนของข้อสอบประมาณ 29% - 32%- ของความแปรปรวนรวม ความแปรปรวนของผู้ตัดสินประมาณ 27% - 35% ของความแปรปรวนรวม และความแปรปรวนของผลร่วมระหว่างข้อสอบกับผู้ตัดสินประมาณ 36% - 41% ของความแปรปรวนรวม **ระดับพอใช้** (C) ความแปรปรวนของข้อสอบประมาณ 29% - 30% ของความแปรปรวนรวม ความแปรปรวนของผู้ตัดสินประมาณ 34% - 40% ของความแปรปรวนรวม และ

ความแปรปรวนของผลร่วมระหว่างข้อสอบกับผู้ตัดสินประมาณ 30% - 38% ของความแปรปรวนรวม **ระดับอ่อน** (D+) ความแปรปรวนของข้อสอบประมาณ 36% - 51% ของความแปรปรวนรวม ความแปรปรวนของผู้ตัดสินประมาณ 5% - 32% ของความแปรปรวนรวม และความแปรปรวนของผลร่วมระหว่างข้อสอบกับผู้ตัดสินประมาณ 31% - 44% ของความแปรปรวนรวม **ระดับอ่อนมาก** (D) ความแปรปรวนของข้อสอบประมาณ 41%-59% ของความแปรปรวนรวม ความแปรปรวนของผู้ตัดสินประมาณ 10% - 18% ของความแปรปรวนรวม และความแปรปรวนของผลร่วมระหว่างข้อสอบกับผู้ตัดสินประมาณ 31% - 44% ของความแปรปรวนรวม

ส่วนการกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์คทั้ง 3 ครั้ง **ระดับดีเยี่ยม** (A) ความแปรปรวนของข้อสอบประมาณ 96% ของความแปรปรวนรวม ความแปรปรวนของผู้ตัดสินประมาณ 0% ของความแปรปรวนรวม และความแปรปรวนของผลร่วมระหว่างข้อสอบกับผู้ตัดสินประมาณ 3% ของความแปรปรวนรวม **ระดับดีมาก** (B+) ความแปรปรวนของข้อสอบประมาณ 91% - 99% ของความแปรปรวนรวม ความแปรปรวนของผู้ตัดสินประมาณ 0% - 1% ของความแปรปรวนรวม และความแปรปรวนของผลร่วมระหว่างข้อสอบกับผู้ตัดสินประมาณ 1% - 7% ของความแปรปรวนรวม **ระดับดี** (B) ความแปรปรวนของข้อสอบประมาณ 97% - 98% ของความแปรปรวนรวม ความแปรปรวนของผู้ตัดสินประมาณ 0% ของความแปรปรวนรวม และความแปรปรวนของผลร่วมระหว่างข้อสอบกับผู้ตัดสินประมาณ 2% - 3% ของความแปรปรวนรวม **ระดับดีพอใช้** (C+) ความแปรปรวนของข้อสอบประมาณ 95% - 98% ของความแปรปรวนรวม ความแปรปรวนของผู้ตัดสินประมาณ 0% ของความแปรปรวนรวม และความแปรปรวนของผลร่วมระหว่างข้อสอบกับผู้ตัดสินประมาณ 2% - 5% ของความแปรปรวนรวม **ระดับพอใช้** (C) ความแปรปรวนของข้อสอบประมาณ 97% - 100% ของความแปรปรวนรวม ความแปรปรวนของผู้ตัดสินประมาณ 0% ของความแปรปรวนรวม และความแปรปรวนของผลร่วมระหว่างข้อสอบกับผู้ตัดสินประมาณ 0% - 3% ของความแปรปรวนรวม **ระดับอ่อน** (D+) ความแปรปรวนของข้อสอบประมาณ 98% ของความแปรปรวนรวม ความแปรปรวนของผู้ตัดสินประมาณ 0% ของความแปรปรวนรวม และความแปรปรวนของผลร่วมระหว่างข้อสอบกับผู้ตัดสินประมาณ 2% ของความแปรปรวนรวม **ระดับอ่อนมาก** (D) ความแปรปรวนของข้อสอบประมาณ 95% - 96% ของความแปรปรวนรวม ความแปรปรวนของผู้ตัดสินประมาณ 0 % ของความแปรปรวนรวม และความแปรปรวนของผลร่วมระหว่างข้อสอบกับผู้ตัดสินประมาณ 4% - 5% ของความแปรปรวนรวม

จากที่กล่าวมา การกำหนดคะแนนจุดตัดด้วยวิธีบุ๊คมาร์ค มีความแปรปรวนของข้อสอบเกือบทั้งหมดของความแปรปรวนรวม ในขณะที่ความแปรปรวนของผู้ตัดสินมีค่าเพียง

เล็กน้อย ส่วนการกำหนดคะแนนจุดตัดด้วยวิธีแองกอฟที่ได้รับการปรับปรุง ความแปรปรวนของข้อสอบ และ ความแปรปรวนของผลร่วมระหว่างข้อสอบกับผู้ตัดสินมีค่าสูงกว่าความแปรปรวนของผู้ตัดสิน แต่การกำหนดคะแนนจุดตัดด้วยวิธีบุ๊คมาร์ค มีความแปรปรวนของผู้ตัดสินต่ำกว่ามากเมื่อเทียบกับความแปรปรวนของผู้ตัดสินที่กำหนดด้วยวิธีแองกอฟที่ได้รับการปรับปรุง สิ่งนี้ที่เป็นตัวชี้ว่า วิธีการบุ๊คมาร์คเป็นวิธีที่ผู้ตัดสินมีการกำหนดคะแนนจุดตัดสอดคล้องกันมากกว่าวิธีการแองกอฟที่ได้รับการปรับปรุง ซึ่งผลการวิเคราะห์ดังกล่าวสอดคล้องกับผลการวิจัยของแวง (Wang, 2003) เพียงแต่งานวิจัยของแวงเป็นการเปรียบเทียบการกำหนดมาตรฐานระหว่างวิธีแองกอฟ และวิธีผังข้อสอบ

สำหรับค่าความเที่ยงของการกำหนดคะแนนจุดตัด หรือค่าสัมประสิทธิ์การสรุปอ้างอิงแบบสัมบูรณ์ (ϕ) เป็นค่าที่ให้ข้อมูลเกี่ยวกับค่าเฉลี่ยการกำหนดคะแนนจุดตัดของผู้ตัดสิน ซึ่งแสดงในรูปดัชนีการตัดสินใจของผู้ตัดสิน(index of rater agreement) (Brennan& Lockwood, 1980) ผลการเปรียบเทียบค่าความเที่ยงทั้ง 7 ระดับ จำนวน 3 ครั้ง พบว่า การกำหนดคะแนนจุดตัดด้วยวิธีบุ๊คมาร์คมีความเที่ยงสูงกว่าวิธีการแองกอฟที่ได้รับการปรับปรุง หรือกล่าวว่า วิธีบุ๊คมาร์คผู้ตัดสินตัดสินได้สอดคล้องกว่าวิธีการแองกอฟที่ได้รับการปรับปรุง เมื่อพิจารณาวิธีการบุ๊คมาร์คจะเห็นว่าค่าเที่ยงประมาณ 0.9 ในทุกระดับ และในผู้ตัดสินจำนวน 8, 12, 14, 16, 18, 20, 22, และ 24 คน

ส่วนวิธีแองกอฟที่ได้รับการปรับปรุง ค่าความเที่ยงหรือค่าสัมประสิทธิ์การสรุปอ้างอิงแบบสัมบูรณ์ทั้ง 7 ระดับ จำนวน 3 ครั้ง **ระดับดีเยี่ยม** (A) เมื่อผู้ตัดสิน 8 คน มีค่า 0.57 - 0.73 เมื่อผู้ตัดสิน 12 คน มีค่า 0.67 - 0.81 เมื่อผู้ตัดสิน 14 คน มีค่า 0.70 - 0.82 เมื่อผู้ตัดสิน 16 คน มีค่า 0.73 - 0.84 เมื่อผู้ตัดสิน 18 คน มีค่า 0.75-0.86 เมื่อผู้ตัดสิน 20 คน มีค่า 0.77 - 0.87 เมื่อผู้ตัดสิน 22 คน มีค่า 0.79 - 0.88 และเมื่อผู้ตัดสิน 24 คน มีค่า 0.80 - 0.89 **ระดับดีมาก** (B+) เมื่อผู้ตัดสิน 8 คน มีค่า 0.58 - 0.75 เมื่อผู้ตัดสิน 12 คน มีค่า 0.67 - 0.82 เมื่อผู้ตัดสิน 14 คน มีค่า 0.71 - 0.84 เมื่อผู้ตัดสิน 16 คน มีค่า 0.74 - 0.86 เมื่อผู้ตัดสิน 18 คน มีค่า 0.76 - 0.88 เมื่อผู้ตัดสิน 20 คน มีค่า 0.78 - 0.89 เมื่อผู้ตัดสิน 22 คน มีค่า 0.80 - 0.90 และเมื่อผู้ตัดสิน 24 คน มีค่า 0.81 - 0.90 **ระดับดี** (B) เมื่อผู้ตัดสิน 8 คน มีค่า 0.55 – 0.77 เมื่อผู้ตัดสิน 12 คน มีค่า 0.65 – 0.83 เมื่อผู้ตัดสิน 14 คน มีค่า 0.68 – 0.85 เมื่อผู้ตัดสิน 16 คน มีค่า 0.71 – 0.87 เมื่อผู้ตัดสิน 18 คน มีค่า 0.74 – 0.88 เมื่อผู้ตัดสิน 20 คน มีค่า 0.76 – 0.89 เมื่อผู้ตัดสิน 22 คน มีค่า 0.78 – 0.90 และเมื่อผู้ตัดสิน 24 คน มีค่า 0.79 – 0.91 **ระดับดีพอใช้** (C+) เมื่อผู้ตัดสิน 8 คน มีค่า 0.76 – 0.79 เมื่อผู้ตัดสิน 12 คน มีค่า 0.82 – 0.84 เมื่อผู้ตัดสิน 14 คน มีค่า 0.85 – 0.89 เมื่อผู้ตัดสิน 16 คน มีค่า 0.87 – 0.88 เมื่อผู้ตัดสิน 18 คน มีค่า 0.88 – 0.89

เมื่อผู้ตัดสิน 20 คน มีค่า 0.89 – 0.90 เมื่อผู้ตัดสิน 22 คน มีค่า 0.90 – 0.91 และเมื่อผู้ตัดสิน 24 คน มีค่า 0.91 – 0.92 **ระดับพอใช้ (C)** เมื่อผู้ตัดสิน 8 คน มีค่า 0.76 – 0.77 เมื่อผู้ตัดสิน 12 คน มีค่า 0.83 – 0.84 เมื่อผู้ตัดสิน 14 คน มีค่า 0.85 – 0.86 เมื่อผู้ตัดสิน 16 คน มีค่า 0.86 – 0.87 เมื่อผู้ตัดสิน 18 คน มีค่า 0.88 – 0.89 เมื่อผู้ตัดสิน 20 คน มีค่า 0.89 – 0.90 เมื่อผู้ตัดสิน 22 คน มีค่า 0.90 – 0.91 และเมื่อผู้ตัดสิน 24 คน มีค่า 0.91 **ระดับอ่อน (D+)** เมื่อผู้ตัดสิน 8 คน มีค่า 0.82 – 0.89 เมื่อผู้ตัดสิน 12 คน มีค่า 0.87 – 0.92 เมื่อผู้ตัดสิน 14 คน มีค่า 0.89 – 0.93 เมื่อผู้ตัดสิน 16 คน มีค่า 0.90 – 0.94 เมื่อผู้ตัดสิน 18 คน มีค่า 0.91 – 0.95 เมื่อผู้ตัดสิน 20 คน มีค่า 0.92 – 0.95 เมื่อผู้ตัดสิน 22 คน มีค่า 0.93 – 0.96 และเมื่อผู้ตัดสิน 24 คน มีค่า 0.93 – 0.96 **ระดับอ่อนมาก (D)** เมื่อผู้ตัดสิน 8 คน มีค่า 0.84 – 0.92 เมื่อผู้ตัดสิน 12 คน มีค่า 0.89 – 0.95 เมื่อผู้ตัดสิน 14 คน มีค่า 0.91 – 0.95 เมื่อผู้ตัดสิน 16 คน มีค่า 0.92 – 0.96 เมื่อผู้ตัดสิน 18 คน มีค่า 0.93 – 0.96 เมื่อผู้ตัดสิน 20 คน มีค่า 0.93 – 0.97 เมื่อผู้ตัดสิน 22 คน มีค่า 0.94 – 0.97 และเมื่อผู้ตัดสิน 24 คน มีค่า 0.94 – 0.97

เมื่อทำการเปรียบเทียบค่าความเที่ยงของการกำหนดมาตรฐาน หรือค่าสัมประสิทธิ์การสรุปอ้างอิงแบบสัมบูรณ์ (ϕ) ระหว่างวิธีแองกอฟที่ได้รับการปรับปรุง กับวิธีบุ๊คมาร์คแต่ละระดับคือ ระดับดีเยี่ยม (A) ระดับดีมาก (B+) ระดับดี (B) ระดับดีพอใช้ (C+) ระดับพอใช้ (C) ระดับอ่อน (D+) และระดับอ่อนมาก (D) เมื่อทำการตัดสินครั้งที่ 1 ครั้งที่ 2 และครั้ง 3 ที่มีจำนวนผู้ตัดสิน 8 12 14 16 18 20 22 24 พบว่า มีค่าแตกต่างกันที่ระดับนัยสำคัญทางสถิติที่ระดับ .01 ทุกค่า โดยค่าความเที่ยงของการกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์คมีค่าสูงกว่าวิธีแองกอฟที่ได้รับการปรับปรุง นั้นคือ ผลการวิจัยสอดคล้องกับสมมุติฐานการวิจัย และสอดคล้องกับผลการวิจัยของแวง (Wang, 2003) และ บาเคนดัลส (Buckendahl, 2002) เมื่อพิจารณาอีกจะเห็นว่าค่าความเที่ยงจากการกำหนดมาตรฐานทั้งสองวิธีต่างมีค่าสูงทั้งคู่ เนื่องจากในกระบวนการกำหนดคะแนนจุดตัดทั้งสองวิธีนั้น มีการเปิดโอกาสให้ผู้ตัดสินเปลี่ยนคะแนนจุดตัดได้หลังจากผู้ตัดสินได้อภิปรายร่วมกัน 2-3 รอบ ซึ่ง Berk(1996) ได้เสนอแนะกระบวนการกำหนดมาตรฐานไว้ 10 ข้อ หนึ่งในนั้นระบุว่า การเปิดโอกาสให้ผู้ตัดสินเปลี่ยนคะแนนจุดตัด 1 - 2 รอบ หลังจากได้ข้อมูลเพิ่มเติมใหม่ เป็นการเพิ่มความสอดคล้องระหว่างผู้ตัดสิน (interjudge consistency)

นอกจากนี้ผู้วิจัยยังคาดว่าค่าความเที่ยงที่กำหนดมาตรฐานด้วยวิธีบุ๊คมาร์คสูงกว่าวิธีแองกอฟ เนื่องจากการกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์คทำได้ง่าย ใช้เวลาไม่มาก เนื่องจากมีคู่มือการจัดเรียงข้อสอบมาให้ผู้ตัดสินพิจารณาโดยข้อสอบมีการจัดเรียงค่าความยากของข้อสอบจากง่ายสุดไปยังข้อสอบที่ยากสุด จากนั้นก็เพียงคั่นหนังสือในหน้าที่ผู้ตัดสินคิดว่า

นักเรียนคาบเส้นในระดับนั้นไม่สามารถทำข้อสอบได้ถูกต้องในหน้าถัดไป จะเห็นว่าวิธีนี้ไม่ต้องระบุความน่าจะเป็นที่จะตอบถูกของนักเรียนคาบเส้นในแต่ละระดับที่จะทำข้อสอบแต่ละข้อถูกต้องเหมือนในวิธีแองกอฟที่ได้รับการปรับปรุง อย่างไรก็ตามจากประสบการณ์ของผู้วิจัยพบว่าวิธีบุ๊คมาร์คมีจุดอ่อนคือ การจัดเตรียมสารสนเทศในคู่มือจัดเรียงข้อสอบเสียเวลาค่อนข้างมาก

3. ความตรงของการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุงและวิธีบุ๊คมาร์ค

ผู้วิจัยหาความตรงตามสภาพของการกำหนดมาตรฐาน เป็นการหาความสัมพันธ์ระหว่างระดับผลการเรียนที่นักเรียนได้รับจากโรงเรียน กับระดับผลการตัดสินที่กำหนดด้วยวิธีการแองกอฟที่ได้รับการปรับปรุง และวิธีบุ๊คมาร์ค พบว่า ค่าความตรงของทั้งสองวิธีไม่แตกต่างกัน โดยทั้งสองวิธีมีค่าความตรงประมาณ 0.5 – 0.6 ซึ่งมีค่าความตรงอยู่ในระดับปานกลาง แสดงว่าทั้งสองวิธีต่างก็มีความตรงระดับปานกลางพอๆ กัน อย่างไรก็ตามการศึกษาความตรงด้วยหลักฐานเพียงอย่างเดียวอาจไม่หนักแน่นเพียงพอ ควรมีการนำหลักฐานอย่างอื่นมาประกอบการพิจารณาด้วยเช่น แฟ้มสะสมงานของนักเรียน การสัมภาษณ์นักเรียน เป็นต้น

4. ความเหมาะสม และความเป็นไปได้ของการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และวิธีบุ๊คมาร์ค

ด้านความเหมาะสม พบว่า ค่าเฉลี่ยความเหมาะสมของวิธีบุ๊คมาร์คมากกว่าค่าเฉลี่ยความเหมาะสมของวิธีแองกอฟที่ได้รับการปรับปรุงอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนด้านความเป็นไปได้ พบว่า ค่าเฉลี่ยความเป็นไปได้ของวิธีบุ๊คมาร์คมากกว่าค่าเฉลี่ยความเป็นไปได้ของวิธีแองกอฟที่ได้รับการปรับปรุงอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 เช่นเดียวกันซึ่งสอดคล้องกับสมมุติฐานการวิจัย และตรงกับผลการวิจัยของบาเคนดัลส (Buckendahl, 2002) ที่นำแบบสอบถามมาตรประมาณค่า 4 ระดับมาให้ผู้ตัดสินทำ พบว่า ผู้ตัดสินมีความคิดเห็นว่าการกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์คมีความสะดวกสบายกว่าวิธีแองกอฟที่ได้รับการปรับปรุง (3.50 สำหรับวิธีแองกอฟ และ 3.64 สำหรับวิธีบุ๊คมาร์ค) และขณะเดียวกันก็สอดคล้องกับผลสัมภาษณ์และความคิดเห็นจากแบบสอบถามเกี่ยวกับการกำหนดมาตรฐานทั้งสองวิธีโดยพบว่าวิธีบุ๊คมาร์คเป็นวิธีที่ง่ายต่อการนำไปใช้ ขั้นตอนไม่ยุ่งยาก สะดวก ประหยัดเวลา นำไปใช้ในสถานการณ์จริงได้ดีกว่าวิธีการ แองกอฟที่ได้รับการปรับปรุง เนื่องจากการพิจารณาข้อสอบทำได้รวดเร็วเพราะมีการจัดเรียงข้อสอบตามความยากของข้อสอบจากง่ายไปหายาก ทำให้การพิจารณาตัดสินคะแนนจุดตัดเป็นไปอย่างรวดเร็ว ผู้ตัดสินมีความคล่องตัวในการดำเนินงานตัดสิน อย่างไรก็ตามผู้ตัดสินบางคนมีความคิดเห็นว่าวิธีการบุ๊คมาร์คมีความน่าเชื่อถือน้อย เนื่องจาก

เป็นวิธีการที่หยาบเนื่องจากไม่มีการกำหนดว่าผู้สอบที่มีคาบเส้นแต่ละระดับมีโอกาสตอบข้อสอบข้อนั้นถูกเท่าไร

จากการเตรียมงานเพื่อจัดสัมมนาอบรมเชิงปฏิบัติการพบว่า ขั้นตอนการเตรียมคู่มือจัดเรียงข้อสอบใช้เวลานานมากเกือบ 1 เดือนจึงแล้วเสร็จเนื่องจากต้องจัดเตรียมสารสนเทศทั้งค่าความยาก อำนาจจำแนก คำนวณค่าความสามารถรายข้อ จัดข้อสอบให้อยู่ทีละหน้า คำนวณสัดส่วนนักเรียนที่เลือกตอบแต่ละตัวเลือก ซึ่งสัดส่วนดังกล่าวผู้วิจัยได้เพิ่มเติมเข้าไปในคู่มือจัดเรียงข้อสอบ และพบว่าผู้ตัดสินเห็นว่าเป็นข้อมูลที่ดีในการใช้ประกอบการกำหนดคะแนนจุดตัด จากการเตรียมงานดังกล่าวพบว่า การทำคู่มือจัดเรียงข้อสอบเสียเวลา และค่าใช้จ่ายมากกว่าวิธีการแองกอฟที่ได้รับการปรับปรุง ส่วนวิธีแองกอฟที่ได้รับการปรับปรุงใช้เวลาในการเตรียมข้อมูลไม่มากเพราะแบบสอบได้ทำไว้อยู่แล้ว เพียงแต่มีแบบฟอร์มสำหรับผู้ตัดสินกรอกโอกาสตอบถูกของนักเรียนในแต่ละข้อ อาจจะเสียเวลาบ้างในการเตรียมสารสนเทศสำหรับผู้ตัดสิน แต่ไม่มากเท่าวิธีบุ๊คมาร์ค นั้นคือเตรียมค่าความยากของข้อสอบ กราฟแท่งแจกแจงคะแนนผู้สอบซึ่งผู้วิจัยเพิ่มกราฟดังกล่าวเข้าไป จากการสังเกตพบว่าผู้ตัดสินส่วนใหญ่ไม่ค่อยได้ใช้ประกอบการพิจารณาเท่าไร ส่วนใหญ่ใช้ค่าความยากประกอบการพิจารณาตัดสิน อาจเนื่องจากผู้ตัดสินพิจารณาเพื่อหาโอกาสการตอบถูกของนักเรียนคาบเส้นระดับนั้นของข้อสอบแต่ละข้อ ซึ่งค่าความยากน่าจะมีประโยชน์ในการพิจารณาประกอบในแต่ละข้อมากกว่า

และจากการสังเกตของผู้วิจัยขณะที่ผู้ตัดสินทำการกำหนดมาตรฐานทั้งสองวิธี พบว่าวิธีแองกอฟที่ได้รับการปรับปรุง ผู้ตัดสินใช้เวลาในการกำหนดคะแนนจุดตัดมากกว่าวิธีบุ๊คมาร์คนานมาก เนื่องจากต้องกำหนดโอกาสการตอบแต่ละข้อถูกของนักเรียนที่คาบเส้นแต่ละระดับ ซึ่งมีถึง 7 ระดับ และมีข้อสอบถึง 100 ข้อ ทำให้ผู้ตัดสินเกิดอาการล้าอย่างเห็นได้ชัด ในขณะที่วิธีบุ๊คมาร์คผู้ตัดสินทำการเปิดคู่มือจัดเรียงข้อสอบทีละหน้าพอถึงระดับที่นักเรียนคาบเส้นระดับนั้นไม่สามารถทำได้ก็คั่นหนังสือไว้ จากนั้นก็ทำต่อไปจนครบ 7 ระดับ ซึ่งวิธีนี้ผู้ตัดสินใช้เวลาไม่มาก จากการสังเกตดังกล่าวจะเห็นว่าวิธีบุ๊คมาร์คเหมาะสำหรับการประเมินที่มีคะแนนจุดตัดหลายระดับ ส่วนแองกอฟที่ได้รับการปรับปรุงเหมาะกับการกำหนดคะแนนจุดตัดที่มีระดับไม่มาก 1-2 ระดับ และมีจำนวนข้อสอบไม่มากนัก

5. ผลของจำนวนผู้ตัดสิน และจำนวนครั้งที่มีต่อการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และวิธีการบุ๊คมาร์ค

การกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง ถ้ากำหนดมาตรฐาน 1 ครั้งควรใช้ผู้ตัดสินตั้งแต่ 16 คนขึ้นไป จึงจะได้ค่าความเที่ยง 0.7 ขึ้นไป ซึ่งสอดคล้องกับ เรย์มอน และรีสด์

(Raymond and Reid, 2001) ที่กล่าวว่า ในการประเมินทางการศึกษา จำนวนผู้ตัดสินระหว่าง 15 - 30 คน เป็นจำนวนที่ยอมรับได้เพราะว่า เป็นจำนวนที่เป็นตัวแทนของผู้ตัดสินทั้งหมดในการกำหนดมาตรฐาน และทำให้ผลที่ได้มีความคงที่

จากผลการวิเคราะห์จะเห็นว่า ยิ่งเพิ่มจำนวนครั้งในการตัดสิน และเพิ่มจำนวนผู้ตัดสินก็จะทำให้ค่าความเที่ยงของการกำหนดมาตรฐานเวิธีแองกอฟที่ได้รับการปรับปรุงพิ่มขึ้น แต่ในสภาพความเป็นจริงการจัดให้มีการกำหนดมาตรฐานหลายๆ ครั้ง จะเสียเวลา และค่าใช้จ่ายสูง การนำผู้ตัดสินมาประชุมกลุ่มหลายๆ วันก็เป็นไปได้ยาก ถึงแม้การกำหนดมาตรฐานตั้งแต่ 2 ครั้งขึ้นไป สามารถลดจำนวนผู้ตัดสินลงเหลือ 8 คนได้ก็ตาม ดังนั้นทางเลือกอีกทางคือ การกำหนดมาตรฐานครั้งเดียวโดยเพิ่มจำนวนผู้ตัดสินในระดับที่เหมาะสม เช่นถ้าต้องการค่าความเที่ยงประมาณ 0.6–0.7 ก็ใช้จำนวนผู้ตัดสินประมาณ 8–18 คน และถ้าต้องการค่าความเที่ยงประมาณ 0.8 ควรใช้ผู้ตัดสินประมาณ 20 คนขึ้นไป นอกจากนี้เจเกอร์ (Jaeger, 1991, cited in Berk, 1996) กล่าวว่า ใน 15 ปีที่ผ่านมาขนาดผู้ตัดสินอยู่ระหว่าง 5 – 20 คน โดยกล่าวว่าฉันทมติของผู้ตัดสินเพิ่มขึ้นยากเมื่อมีจำนวนผู้ตัดสินเกิน 20 คน ดังนั้นผู้วิจัยมีความคิดเห็นว่า ในการพิจารณาขนาดผู้ตัดสินจึงควรพิจารณาค่าความเที่ยงในขนาดที่ยอมรับได้ โดยไม่ใช้จำนวนผู้ตัดสินที่มากจนเกินไปนักเพื่อเป็นการประหยัดค่าใช้จ่าย ตลอดจนฉันทมติที่จะเกิดขึ้นในการกำหนดคะแนนจุดตัด และในการคัดเลือกผู้ตัดสินนอกจากจำนวนผู้ตัดสินที่ต้องพิจารณาประกอบการแล้ว อีกสิ่งหนึ่งที่ต้องพิจารณาคือ คุณสมบัติที่เหมาะสมของผู้ตัดสิน เช่น ประสบการณ์ในวิชาชีพ ประสบการณ์ในการกำหนดมาตรฐาน เป็นต้น

ส่วนการกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์ค การกำหนดมาตรฐานตั้งแต่ครั้งที่ 1 ขึ้นไปและใช้จำนวนผู้ตัดสินตั้งแต่ 8 คนขึ้นไป จะได้ค่าความเที่ยง 0.9 ขึ้นไป ดังนั้นการเพิ่มจำนวนผู้ตัดสิน และจำนวนครั้งไม่ค่อยมีผลต่อค่าความเที่ยงเท่าไร เนื่องจากทำให้ค่าความเที่ยงเพิ่มเพียงเล็กน้อยเท่านั้น อาจเป็นเพราะว่า วิธีนี้มีคู่มือจัดเรียงข้อสอบที่ช่วยผู้ตัดสินพิจารณาในการกำหนดคะแนนจุดตัด ซึ่งคู่มือดังกล่าวจะมีการจัดเรียงข้อสอบจากข้อง่ายที่สุด ไปยังข้อที่ยากที่สุด 1 ข้อต่อ 1 หน้า พร้อมทั้งมีสารสนเทศให้ผู้ตัดสินพิจารณาในแต่ละหน้าเช่น ความสามารถของผู้สอบที่จะทำข้อสอบนั้นได้ เปอร์เซนต์นักเรียนเก่ง และอ่อนที่เลือกแต่ละตัวเลือก พร้อมทั้งตัวข้อสอบให้ผู้ตัดสินพิจารณาอยู่ในหน้าเดียวกัน ผู้ตัดสินเพียงทำการกำหนดคะแนนจุดตัดโดยเปิดไปทีละข้อจนถึงหน้าที่ผู้ตัดสินคิดว่านักเรียนระดับนั้นไม่สามารถทำข้อสอบได้ ก็ทำการคั่นหนังสือ หรือกำหนดคะแนนจุดตัดตรงหน้านั้น จะเห็นได้ว่าวิธีบุ๊คมาร์คเป็นวิธีการที่ง่ายต่อการตัดสิน และช่วยจัดระบบการคิดของผู้ตัดสิน จึงทำให้ผลการกำหนดคะแนนจุดตัดของผู้ตัดสินสอดคล้องใกล้เคียงกัน หรือบางคนตรงกัน ทำให้ค่าความแปรปรวนของผู้ตัดสินต่ำ ส่งผลให้ค่าความเที่ยงของการกำหนดมาตรฐานสูงนั้นเอง

จากที่กล่าวมาจะเห็นว่า วิธีการบุ๊คมาร์คเป็นวิธีการที่มีค่าความเที่ยงในการกำหนดมาตรฐาน และความเหมาะสม และความเป็นไปได้สูงกว่าวิธีการแองกอฟที่ได้รับการปรับปรุง ในขณะที่ความตรงตามสภาพของการกำหนดมาตรฐานทั้งสองวิธี พบว่าทั้งสองวิธีมีค่าความตรงระดับปานกลาง และทั้งสองวิธีมีความตรงของการกำหนดมาตรฐานไม่แตกต่างกัน

สำหรับวิธีแองกอฟที่ได้รับการปรับปรุงนั้น น่าจะเหมาะสำหรับการประเมินในห้องเรียนที่มีการใช้แบบสอบแบบเลือกหลายคำตอบซึ่งประกอบด้วยแบบสอบย่อยหลายๆ ชุด แต่ละชุดมีจำนวนข้อสอบไม่มากนัก ซึ่งวิธีแองกอฟที่ได้รับการปรับปรุงมีข้อดีไม่ต้องเสียเวลาเตรียมสารสนเทศสำหรับผู้ตัดสินมากเท่าวิธีบุ๊คมาร์ค

ส่วนวิธีบุ๊คมาร์คเหมาะกับการสอบที่ต้องนำผลการสอบไปตัดสินอนาคตของผู้สอบ เนื่องจากวิธีนี้ให้ค่าความเที่ยงในการกำหนดมาตรฐานสูง และต้องใช้แบบสอบจำนวนมาก แต่มีจุดอ่อนที่ต้องเสียเวลาเตรียมคู่มือจัดเรียงข้อสอบค่อนข้างนานและใช้งบประมาณสูง

ข้อเสนอแนะ

ข้อเสนอแนะในการนำผลการวิจัยไปใช้

1. การกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง เหมาะสำหรับการประเมินที่ใช้แบบสอบที่ประกอบด้วยแบบสอบย่อยหลายแบบสอบที่มีลักษณะแบบเลือกหลายคำตอบ เพราะวิธีการนี้เหมาะสำหรับแบบสอบที่มีจำนวนข้อสอบไม่มาก เนื่องจากต้องพิจารณาโอกาสตอบถูกทีละข้อซึ่งต้องใช้ระยะเวลาพอควร และวิธีนี้เหมาะสำหรับการกำหนดมาตรฐานเป็นรายบุคคล เนื่องจากครูสอนนักเรียนเหล่านั้นทำให้ทราบว่านักเรียนแต่ละคนมีความสามารถเป็นอย่างไร วิธีนี้ยังเหมาะสำหรับการกำหนดคะแนนจุดตัดที่มีจำนวนระดับไม่มาก เพราะการกำหนดคะแนนจุดตัดแต่ละระดับต้องกำหนดโอกาสในการตอบถูกของนักเรียนระดับนั้นทุกข้อ ทำให้เสียเวลา และอาจเกิดการสับสนขึ้นมาได้

2. การกำหนดมาตรฐานด้วยวิธีการบุ๊คมาร์ค เหมาะสำหรับการกำหนดมาตรฐานที่ผู้สอบต้องนำผลการสอบไปตัดสินอนาคตของผู้สอบ เนื่องจากต้องการคะแนนจุดตัดที่เป็นมาตรฐานเดียวเหมือนกันหมด และเป็นคะแนนจุดตัดที่มีคุณภาพสูง วิธีนี้ให้ค่าความแปรปรวนของผู้ตัดสินไม่มาก จึงทำให้มีคะแนนจุดตัดที่คงเส้นคงวา นอกจากนี้ยังเหมาะสำหรับแบบสอบที่มีข้อสอบจำนวนมากเช่น100 ข้อขึ้นไป มีลักษณะการให้คะแนนคะแนนแบบหลายค่า และให้คะแนนแบบสองค่าอยู่ในฉบับเดียวกัน และเหมาะสำหรับคะแนนจุดตัดหลายระดับเพราะสะดวก

ในการกำหนดคะแนนจุดตัดเพียงคั่นหนังสือในหน้าที่คิดว่าผู้สอบคาบเส้นระดับนั้นสามารถทำได้และหน้าถัดไปไม่สามารถทำได้

สำหรับการสอบที่ตัดสินอนาคตผู้สอบ อาทิเช่น การทดสอบระดับชาติเพื่อให้ผลย้อนกลับถึงผู้เรียนว่า ผู้เรียนอยู่ในระดับใด เช่น อาจแบ่งเป็น 4 ระดับคือ ระดับดีมาก ระดับดี ระดับพอใช้ และระดับตก หรือ การสอบใบประกอบวิชาชีพต่างๆ เช่น อาจแบ่งเป็น 4 ระดับคือ ระดับเชี่ยวชาญ ระดับชำนาญการ ระดับพื้นฐาน และระดับตก หรือการทดสอบในห้องเรียนที่ต้องนำผลการเรียนมาใช้พิจารณาการเข้าเรียนต่อระดับอุดมศึกษา หรือเพื่อสอบชิงทุน

ข้อเสนอแนะในการวิจัย

1. เนื่องจากปัจจุบันนี้มีการประเมินการปฏิบัติที่ให้คะแนนแบบหลายค่า ดังนั้นควรมีการเปรียบเทียบคุณภาพของการกำหนดคะแนนจุดตัดด้วยวิธีแองกอฟแบบปรับขยาย (Extented Angoff method) กับวิธีบุ๊คมาร์ค (Bookmark method) ในแบบสอบที่มีทั้งแบบเลือกตอบ และแบบอัตนัย

2. ควรมีการศึกษาการกำหนดมาตรฐานแนวใหม่ เช่น วิธีแองกอฟแบบปรับขยาย (Extented Angoff) (Hambleton and Plake, 1995; Impara and Plake ,1997; Goodwin, 1999; Brandon, 2004) วิธีการเส้นภาพเด่น (Dominant Profile) (Putnam, Pence and Jaeger, 1995; Plake, Hambleton and Jaeger, 1997) วิธีการใช้การตัดสินเชิงนโยบาย (Judgmental Policy Capturing) (Jaeger, 1995)

3. ควรทำวิจัยเชิงคุณภาพกับผู้ตัดสินหลังจากกำหนดมาตรฐานด้วยแองกอฟที่ได้รับการปรับปรุง และวิธีบุ๊คมาร์ค เพื่อตรวจสอบประสบการณ์ และกระบวนการคิดในการกำหนดคะแนนจุดตัด ซึ่งเป็นการตรวจสอบความตรงภายนอก และขณะเดียวกันผลที่ได้นำไปใช้ปรับปรุงกระบวนการกำหนดคะแนนจุดตัดต่อไป

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

รายการอ้างอิง

ภาษาไทย

กัญจนา ลินทรัตนศิริกุล. 2534. การประยุกต์ทฤษฎีการตอบสนองต่อข้อสอบเพื่อกำหนดคะแนนจุดตัดตามทฤษฎีของนีเดลสกีกับวิธีการใช้กลุ่มคาบเส้น. วิทยานิพนธ์ปริญญาการศึกษาดุษฎีบัณฑิต สาขาการวิจัยและพัฒนาหลักสูตร บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร.

โกวิท ประวาลพฤกษ์ และสมศักดิ์ สินธุระเวชญ์. 2523. การประเมินผลในชั้นเรียน. กรุงเทพมหานคร: วัฒนาพานิช.

นภา นิ่มนาค. 2538. การศึกษาคะแนนจุดตัดและความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ เมื่อใช้วิธีหาคะแนนจุดตัด และประสบการณ์ของผู้เชี่ยวชาญต่างกัน. ปริญญาการศึกษามหาบัณฑิต สาขาการวัดผลการศึกษา มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร.

พวงแก้ว ปุณยกนก. 2546. การประเมินพุทธิพิสัย. ใน สุวิมล ว่องวาณิช(บรรณาธิการ), การประเมินผลการเรียนรู้แนวใหม่, หน้า 167-194. กรุงเทพมหานคร: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

พร้อมพรรณ อุดมสิน. การวัดและการประเมินผลการเรียนการสอนคณิตศาสตร์. 2544. พิมพ์ครั้งที่ 3. กรุงเทพมหานคร: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

ฟาฏินา วงค์เลขา. 2546. การประเมินคุณภาพการศึกษาระดับชาติ. วารสารวิชาการ. 6(มกราคม): 43-55.

บุญเชิด ภิญโญอนันตพงษ์. 2526. การทดสอบแบบอิงเกณฑ์ : แนวคิดและวิธีการ. เอกสารประกอบคำบรรยายลำดับที่ 42 ภาควิชาพื้นฐานการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ.

เยาวดี วิบูลย์ศรี. 2539. การวัดและการสร้างแบบสอบผลสัมฤทธิ์. กรุงเทพมหานคร: สำนักพิมพ์จุฬาลงกรณ์มหาวิทยาลัย.

วิภารัตน์ ศรีบุตรตา. 2541. การเปรียบเทียบความสอดคล้องในการตัดสินคะแนนจุดตัดระหว่างวิธีของแองกอฟ วิธีนีเดลสกี และวิธีอิมพาราและเพลค. วิทยานิพนธ์ครุศาสตร์มหาบัณฑิต ภาควิชาวิจัยการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.

วิมลมาศ ยิ้มละมัย. 2541. การกำหนดจุดตัดสำหรับแบบอิงโดเมนประเภทเลือกตอบ และตอบสั้น : การเปรียบเทียบคุณภาพระหว่างวิธีวิลคอกซ์กับวิธีแองกอฟตามทฤษฎีการตอบสนองข้อสอบ. วิทยานิพนธ์ครุศาสตร์มหาบัณฑิต ภาควิชาวิจัยการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.

ศิริชัย กาญจนวาสี. 2545. ทฤษฎีการประเมิน. พิมพ์ครั้งที่ 3. กรุงเทพมหานคร: สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

ศิริชัย กาญจนวาสี. 2545. ทฤษฎีการทดสอบแนวใหม่. พิมพ์ครั้งที่ 2. กรุงเทพมหานคร: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

ศิริชัย กาญจนวาสี. 2544. ทฤษฎีการทดสอบแบบดั้งเดิม. พิมพ์ครั้งที่ 4. กรุงเทพมหานคร: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

สมหวัง พิธิยานุวัฒน์. 2544. วิธีวิทยาการประเมิน: ศาสตร์แห่งคุณค่า. พิมพ์ครั้งที่ 2. กรุงเทพมหานคร: สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

สาลิกา เมธนาวิน. 2537. การประมาณค่าความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ ตามวิธีของฮวน วิธีของเบรนนอนและเคน และวิธีของราชู เมื่อกำหนดคะแนนจุดตัดต่างกัน. ปริญญาการศึกษามหาบัณฑิต สาขาการวัดผลการศึกษา มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร.

สุชาติ ประสิทธิ์รัฐสินธุ. 2545. การใช้สถิติในงานวิจัยอย่างถูกต้องและได้มาตรฐานสากล. กรุงเทพมหานคร: บริษัทเฟื่องฟ้า พริ้นติ้ง จำกัด.

สุพิมพ์ ศรีพันธ์วรสกุล. 2539. การวิเคราะห์ความตรงลู่เข้าของการวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ตามทฤษฎีการอ้างอิงสรุป. วิทยานิพนธ์ครุศาสตร์ดุษฎีบัณฑิต ภาควิชาวิจัยการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.

สุรินทร์ แพ่งจันทึก. 2528. การเปรียบเทียบคะแนนจุดตัด และความเที่ยงของแบบสอบอิงเกณฑ์ระหว่างแบบสอบประเภทเลือกตอบกับตอบสั้นจากการใช้ทฤษฎีการตัดสินใจของแกลส. วิทยานิพนธ์ครุศาสตร์มหาบัณฑิต ภาควิชาวิจัยการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.

เสมอ ตรีมงคล. 2539. สัมประสิทธิ์การสรุปอ้างอิงของวิธีหาคะแนนจุดตัดตามเทคนิคของนีเดลสกี และแองกอฟ. ปริญญาการศึกษามหาบัณฑิต สาขาการวัดผลการศึกษา มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร.

อรนุช อยู่ทิม และคณะ. 2546. การประเมินผลการเรียนรู้ของผู้เรียนในชั้นเรียน ตามหลักสูตรการศึกษาขั้นพื้นฐาน พุทธศักราช 2544. วารสารวิชาการ. 6(มกราคม): 9-19.

อัจฉรยา ปราบอริพ่าย. 2532. ผลของการให้สารสนเทศแก่ผู้เชี่ยวชาญต่อการกำหนดคะแนนจุดตัดของแบบสอบผลสัมฤทธิ์วิชาคณิตสาสตร์. วิทยานิพนธ์ครุศาสตร์มหาบัณฑิต ภาควิชาวิจัยการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.

อุทุมพร จามรมาน. 2540. การตีค่าความสามารถที่แท้จริงของผู้เรียนเพื่อการปฏิรูปการศึกษา. กรุงเทพมหานคร: หจก. ฟันนี่พับบลิชชิ่ง.

ภาษาอังกฤษ

Behuniak, P; Archambault, F.X.; Gable, R.K. 1982. Angoff and Nedelsky standard setting procedures: Implications for the validity of proficiency test score interpretation. Educational and Psychological Measurement. 42(1): 247-255.

Beretvas, S. N. 2004. Comparison of bookmark difficulty locations under different item response models. Applied Psychologcal Measurement. 28(1): 25-47.

Ben-David, M. F. 2000. AMEE guild no. 18: Standard setting in student assessment. Medical Teacher. 22(2): 120-130.

Berk, R.A. 1976. Determination of optimal cutting scores in criterion-referenced measurement. Journal of Experimental Education. 45(2): 4-9.

Berk, R.A. 1986. A comsumer' s guide to setting performance standards on criterion-referenced tests. Review of Educational Research. 56(1): 137-172.

Berk, R.A. 1995. Something old, something new, something borrowed, a lot to do. Applied Measurement in Education, 8(1), 99-109.

Berk, R.A. 1996. Standard setting: the next generation (where few psychometricians have gone before!). Applied Measurement in Education 9(3): 215-235.

Brandon, P. R. 2004. Conclusions about frequently studied modified angoff standard-setting topics. Applied Measurement in Education. 17(1): 59-88.

Brandon, P. R. 2002. Two versions of the contrasting-groups standard-setting method: A review. Measurement and Evaluation in Counseling and Development. 35: 167-181.

Brennan, R. L. & Lockwood, R.E. 1980. A comparison of the Nedelsky and Angoff cutting score procedures using generalizability theory. Applied Psychological Measurement. 4: 219-240.

Buckendahl, C. W., Smith, R.W., Impara, J.C. & Plake, B.S. 2002. A comparison of angoff and bookmark standard setting methods. Journal of Educational measurement. 33(3): 253-263.

Cascallar, A. S. & Cascallar, E. C. 2003. Setting standards in the assessment of complex performance: The optimized extended-response standard setting method. In M. Segers ; F. Dochy ; E. Cascallar (ed.), Optimising new modes of assessment: in search of qualities and standards. pp.247-266. Norwell: Kluwer Academic Publishers.

Cizek, G. J. ; Bunch, M. B. & Koons, H. 2004. Setting performance standards: Contemporary methods. Educational Measurement: Issues and Practice. 23(4): 31-50.

Cohen, A. S. ; Kane, M. T. & Crooks, T. J. 1999. A generalized examinee-centered method for setting standards on achivement tests. Applied Measurement in Education 12(4): 343-366.

Cross, L. H.; Impara, J. C.; Frary, R. B. & Jaeger, R. M. 1984. A comparison of three methods for establishing minimum standard on the national teacher examinations. Journal of Educational Measurement. 21: 113-129.

Egan, K. 2002. Bookmark standard setting overview. Retrieved 24/7/2004 from http://www.dpi.state.wi.us/oea/ctbbkmrk03.html

Fererer, W. T. 1955. Experimental design. NY: The Macmillan Company.

Ferrara, S. ; Johnson, E. & Chen, L. 2005. Vertically articulated performance standards: Logic, procedures, and likely classification accuracy. Applied Measurement in Education. 18(1): 35-59.

Friedman, C. B. & Fisbie, D.A. 1990. Interjudge consensus and intrajudge consistency: Is it possible to have both in standard setting. Paper presented at the Annual Meeting of the National council on Measurement in Education, Boston.

Howell, D. C. 1997. Statistical methods for psychology. 4th ed. CA: Wadsworth Publishing Company.

Hurtz, G. M. and Auerbach, M. A. 2003. A meta-analysis of the effects of modifications to the angoff method on cutoff scores and judgment consensus. Educational and Psychological Measurement. 63(4): 584-601.

Glass, G.A. 1978. Standard and criteria. Journal of Educational Measurement. 15(4): 237-261.

Green, B. F. 2000. Setting performance standards. meeting of MAPAC. Retrieved 12/9/2004 from http://www.ipmaac.org/mapac/meetings/2000/berrtgre.pdf

Green, D. R. 2003. Interpreting the results of three different standard-setting procedures. Educational Measurement: Issues and Practice. Spring: 22-32.

Gullickson, A.R. 2003. The student evaluation standards: How to improve evaluations of students. Thousand Oaks, CA: Corwin Press, Inc.

Haladyna and Roid. 1981. The role of instructional sensitivity in the empirical: Review of criterion-referenced test items. Journal of Educatioal Measurement. 18: 39 –41.

Hambleton, R. K. ; Jaeger, R. M. ; Plake, B. S. ; Mills, C. 2000. Setting performance standards on complex educational assessments. Applied Psychological Measurent 24(4): 355-366.

Hambleton, R. K. and others. 1978. Criterion-referenced testing and measurement: A review of technical issues and developments. Review of Educational Research. 48(1): 1-47.

Hambleton, R. K. ; Plake, B. S. 1995. Using an extended angoff procedure to set standards on complex performant assessments. Applied Measurement in Education 8(1): 41-55.

Huynh, H. 1976. On the reliability of decision in domain-referenced testing. Journal of Educational Measurement. 13(4): 253-263.

Huynh, H. 1998,Spring. On score locations of binary and partial credit items and their applications to item mapping and criterion – referenced interpretation. Journal of Educational and Behavioral Statistics. 23(1): 35-56.

Imprara, J. C. 1997. Standard setting: An alternative approach. Journal of Educational Measurement. 34(4): 353-366.

Imprara, J. C. & Plake, B. S. 1997. Standard setting: An alternative approach. Journal of Educational Measurement. 34: 69-81.

Jaeger, R. M. 1995. Setting performance standards through two-stage judgmental policy capturing. Applied Measurement in Education. 8(1): 15-40.

Kane, M. 1994. Validating the performance standards associated with passing scores. Review of Educational Research. 64(3): 425-461.

Kane, M. 1997. On the use of IRT models with judgemental standard setting procedures. Journal of Educational Measurement. 24: 333-345.

Kane, M. 1998. Choosing between examinee-centered and test-centered standard-setting methods. Educational Assessment 5(3): 129-145.

Kiplinger, V. L. 1997. Standard-setting procedures for the specification of performance levels on a standards-based assessment. Retrieved 5/7/2004 from http://www.cde.state.co.us/cdeassess/csap/asperf.htm

Kline, Theresa. 2005. Psychological testing : A practical approach to design and evaluation. Thousand Oaks: Sage publications, Inc.

Kunnan, A.J. 1992. An investigation of a criterion-referenced test using G-thoory, and factor and cluster. Languge testing. 9(1): 30-49.

Lewis, D. M., Green, D. R., Mitzel, H.C., Baum, K.,& Patz, R.J. 1998. The bookmark standard setting procedure: Methodology and Recent Implementations. Paper presented at the National Council for Mearsurement in Education annual meeting, San Diego, CA.

Lewis, D. M., Mitzel, H.C., Green, D.R.,& Patz, R. J. 1999. The bookmark standard setting procedure. Monterey, CA: McGraw-Hill.

Linn, R.L. 1978. Demands, cautions and suggestions for setting standards. Journal of Educational Measurement. 15(4): 301-308.

Livingston, S. A. and Zieky, M. J. 1982. Passing scores: A manual for setting standards of performance on educational and occupational tests. NJ: Educatioanl Testing Service.

Longman Group Ltd. 1999. Longman dictionary of contemporary English: International students edition. 6 th ed. Harlow: Pearson Education Limited.

Mills, C.N. 1983. A comparison of three methods of establishing cut-off scores on criterion-referenced tests. Journal of Educational Measurement 20(3): 283-292.

Mitzel, H.C., Lewis, D.M., Patz, R.J., Green, D.R. 2001. The bookmark procedure: Psychological perspectives. In G.J. Cizek (Ed.), Setting performance standards: Concepts, methods, and perspectives (pp. 249-281). Mahwah, NJ: Erlbaum.

Norcini, J. J. and others. 1987. A comparison of three variations on standard setting method. Journal of Educational Measurement. 24(1): 56-64.

Patricia, W. 1991. The relatonship between modified angoff knowledge estimation judgments and item difficulty values for seven NTE specialty area tests. Paper Present at the Annual Meeting of the California Educational Research Association, San Diego.

Plake, B.S. and Melican, G.J. 1989. Effects of item context on intrajudge consistency of expert judgment via the Nedelsky standard setting method. Educational and Psychological Measurement. 49(1): 45-51.

Plake, B. S. 1995. The performance domain and the structure of decision space. Applied Measurement in Education. 8(1): 3-13.

Plake, B. S. & Impara, J. C. 1996. Intrajudge consistency using the angoff standard setting method. Paper presented at the Annual Meetting of the National Council on Measurement in Education, New York.

Plake, B. S. ; Hambleton, R. K. ; Jaeger, R. M. 1997. A new standard-setting method for performance assessment: The dominant profile judgment method and some field-test results. Educational and Psychological Measurement. 57(3): 355-366.

Putnam, S. E.; Pence, P. & Jaeger, R. M. 1995. A multi-stage dominant profile method for setting standards on complex performance assessments. Applied Measurement in Education. 8(1): 57-83.

Raju, N.S. , Drasgow, F. , and Slinde, J.A. 1993. An empirical comparison of the area methods, Lord's Chi-square test, and the Mantel-Haenszel technique for assessing differential item functioning. Educational and Psychological Measurement. 53.

Raymond, M. R., & Reid, J. B. 2001. Who made thee a judge? Selecting and training participants for standard setting. In G. J. Cizek (Ed.), Standard setting: Concepts, methods, and perspectives(pp. 119 -157). Mahwah, NJ: Erlbaum.

Reckase, M. D. 2000. Survey and evaluation of recently developed procedures for setting standards on education tests. In Student performance standards on the National Assessment of Educational Progress: Affirmation and improvements. Washington, DC: National Assessment Governing Board.

Rowley, G.L. 1982. Historical antecedents of the standard-setting debate: An inside account of the minimal-beardness controversy. Journal of Educational Measurement. 19(2): 87-95.

Sireci, S.G. and Robin, F. 1999. Using Cluster Analysis to Facilitate Standard Setting. Applied Mesurement in Education. 12(3): 301-325.

Skaggs, G.,& Tessema,A. 2001. Item disordinality with the bookmark standard-setting procedure. Paper presented at the annual meeting of the American Educational Research Association, Seattle, WA.

Subkoviak, M.J. 1976. Estimation reliability from a single admistration of a criterion-referenced test. Journal of Educational Measurement. 13(4): 265-276.

Subkoviak, M.J. 1978. Empirical investigation of procedures for estimating reliability of mastery tests. Journal of Educational Measurement. 15(2): 111-116.

Van der Schoot, F. C. 2002, April. The application of an IRT-based method for standard setting in a three-stage procedure. Paper presented at the annual meeting of the National Council on Measurement in Education. New Orleans.

Wang, N. 2003. Use of the rasch IRT model in standard setting: An item-mapping method. Journal of Educational Measurement. 40(3): 231-253.

Woodruff, D.J. and Feldt, L. S. 1986. Tests for eqaulity of several alpha coefficients when their sample estimates are dependent. Psychometrika. 51(3): 393-413.

Zieky, M. 1987. Method of setting standards of performance on criterion referenced tests. Paper presented at the 13 th International Conference of International Association for Educational Assessment, Bangkok, Thailand.

Zieky, M. J. 2001. So much has changed: How the setting of cutscores has evolved since the 1980's. In G. J. Cizek (ed.) Setting performance standards: Concepts, method, and perspectives. Mahwah, NJ: Lawrence Erlbaum Publishers.

Zwick, R., Senturk, D., Wang, J., & Loomis, S. C. 2001. An investigation of alternative methods for item mapping in the national assessment of educational progress. Educational Measurement: Issues and Practice, 20(2): 15-25.

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก ก

รายนามผู้เชี่ยวชาญตรวจสอบเครื่องมือที่ใช้ในการวิจัย

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

รายนามผู้เชี่ยวชาญตรวจสอบความครอบคลุมของวัตถุประสงค์การเรียนรู้
ในขอบเขตเนื้อหา และการกำหนดน้ำหนักความสำคัญของวัตถุประสงค์การเรียนรู้
วิชาคณิตศาสตร์ชั้นมัธยมศึกษาปีที่ 3 ภาคเรียนที่ 1

1. **รองศาสตราจารย์ศักดา บุญยไวโรจน์ โรงเรียนสาธิตจุฬาลงกรณ์มหาวิทยาลัย**
   ปริญญาตรี ครุศาสตรบัณฑิต สาขามัธยมวิทยาศาสตร์-คณิตศาสตร์
   ปริญญาโท ครุศาสตรมหาบัณฑิต สาขาสถิติศึกษา
   ประสบการณ์การสอนวิชาคณิตศาสตร์ชั้นมัธยมศึกษาปีที่ 3 ทั้งหมด 20 ปี
2. **อาจารย์วิมลมาศ อำพลพงษ์ โรงเรียนสาธิตจุฬาลงกรณ์มหาวิทยาลัย**
   ปริญญาตรี ครุศาสตรบัณฑิต สาขามัธยมวิทยาศาสตร์-คณิตศาสตร์
   ปริญญาโท ครุศาสตรมหาบัณฑิต สาขาการวัดและประเมินผลการศึกษา
   ประสบการณ์การสอนวิชาคณิตศาสตร์ชั้นมัธยมศึกษาปีที่ 3 ทั้งหมด 5 ปี
3. **อาจารย์จิตสุดา ธราพร โรงเรียนบ้านคลองแห้ง จ.กระบี่**
   ปริญญาตรี ศึกษาศาสตรบัณฑิต สาขาคณิตศาสตร์
   ปริญญาโท ศึกษาศาสตรมหาบัณฑิต สาขาการวัดผลการศึกษา
   ประสบการณ์การสอนวิชาคณิตศาสตร์ชั้นมัธยมศึกษาปีที่ 3 ทั้งหมด 17 ปี

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

รายนามผู้เชี่ยวชาญตรวจสอบความตรงเชิงเนื้อหาของแบบสอบ
วิชาคณิตศาสตร์ชั้นมัธยมศึกษาปีที่ 3 ภาคเรียนที่ 1

1. **ดร. ผลาดร สุวรรณโพธิ์ คณะศึกษาศาสตร์ มหาวิทยาลัยบูรพา**
   ปริญญาตรี วิทยาศาสตรศาสตร์บัณฑิต สาขาคณิตศาสตร์
   ปริญญาโท ศึกษาศาสตรมหาบัณฑิต สาขาการสอนคณิตศาสตร์
   ปริญญาเอก วิทยาศาสตรดุษฎีบัณฑิต สาขาคณิตศาสตร์ประยุกต์
   ประสบการณ์การสอนวิชาคณิตศาสตร์ 23 ปี

2. **ดร. อุทัยวรรณ สายพัฒนะ โรงเรียนธรรมศาสตร์คลองหลวงวิทยาคม จ. ปทุมธานี**
   ปริญญาตรี ศึกษาศาสตรบัณฑิต สาขาคณิตศาสตร์
   ปริญญาโท ศึกษาศาสตรมหาบัณฑิต สาขาการวัดผลการศึกษา
   ปริญญาเอก ศึกษาศาสตรดุษฎีบัณฑิต
   สาขาการทดสอบและวัดผลการศึกษา
   ประสบการณ์การสอนวิชาคณิตศาสตร์ 17 ปี

3. **อาจารย์จิตสุดา ธราพร โรงเรียนบ้านคลองแห้ง จ.กระบี่**
   ปริญญาตรี ศึกษาศาสตรบัณฑิต สาขาคณิตศาสตร์
   ปริญญาโท ศึกษาศาสตรมหาบัณฑิต สาขาการวัดผลการศึกษา
   ประสบการณ์การสอนวิชาคณิตศาสตร์ 17 ปี

4. **อาจารย์วิวัฒน์ ลีวงศ์วัฒน์ โรงเรียนสาธิตจุฬาลงกรณ์มหาวิทยาลัย**
   ปริญญาตรี ครุศาสตรบัณฑิต
   สาขาคณิตศาสตร์-จิตวิทยาให้คำปรึกษาและแนะแนว
   ปริญญาโท ครุศาสตรมหาบัณฑิต สาขาการศึกษาคณิตศาสตร์
   ประสบการณ์การสอนวิชาคณิตศาสตร์ 9 ปี

5. **อาจารย์คมสัน ตรีไพบูลย์ คณะศึกษาศาสตร์ มหาวิทยาลัยบูรพา**
   ปริญญาตรี ศึกษาศาสตรบัณฑิต สาขาคณิตศาสตร์
   ปริญญาโท ศึกษาศาสตรมหาบัณฑิต สาขาการศึกษาคณิตศาสตร์
   ประสบการณ์การสอนวิชาคณิตศาสตร์ 4 ปี

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

รายนามผู้เชี่ยวชาญตรวจสอบความตรงเชิงเนื้อหาของ
แบบสอบถามความคิดเห็นเกี่ยวกับกระบวนการกำหนดมาตรฐานด้วยวิธีการ
แองกอฟที่ได้รับการปรับปรุง และวิธีบุ๊คมาร์ค
แบบประเมินความเหมาะสม และความเป็นไปได้ในการกำหนดมาตรฐานด้วยวิธีการ
แองกอฟที่ได้รับการปรับปรุง และวิธีบุ๊คมาร์ค

1. **รองศาสตราจารย์ ดร.ไพรัตน์ วงศ์นาม คณะศึกษาศาสตร์ มหาวิทยาลัยบูรพา**
   ปริญญาโท ครุศาสตรมหาบัณฑิต สาขาการวัดและประเมินผลการศึกษา
   ปริญญาเอก ครุศาสตรดุษฎีบัณฑิต สาขาการวัดและประเมินผลการศึกษา

2. **ดร. สมพงษ์ ปั้นหุ่น โรงเรียนบ้านทุ่งหลุก จ.เชียงใหม่**
   ปริญญาโท ศึกษาศาสตรมหาบัณฑิต สาขาการวัดและประเมินผลการศึกษา
   ปริญญาเอก ครุศาสตรดุษฎีบัณฑิต สาขาการวัดและประเมินผลการศึกษา

3. **อาจารย์พิกุล เอกวรางกูร คณะศึกษาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์**
   ปริญญาโท ศึกษาศาสตรมหาบัณฑิต สาขาการวิจัยและประเมินผลการศึกษา
   กำลังศึกษาต่อปริญญาเอก สาขาการวัดและประเมินผลการศึกษา

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก ข

รายนามผู้ตัดสินมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และวิธีบุ๊คมาร์ค

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

รายนามผู้ตัดสินมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และวิธีบุ๊คมาร์ค

1. **อาจารย์นริสา รอดทยอย โรงเรียนบางละมุง**
   ตำแหน่ง ครูชำนาญการ
   ปริญญาตรี ศึกษาศาสตรบัณฑิต สาขาคณิตศาสตร์

2. **อาจารย์อาภรณ์ อินทรเกษม โรงเรียนบางละมุง**
   ตำแหน่ง ครูชำนาญการ
   ปริญญาตรี วิทยาศาสตรบัณฑิต สาขาสถิติประยุกต์
   ปริญญาโท วิทยาศาสตรมหาบัณฑิต สาขาคณิตศาสตร์

3. **อาจารย์อุทิศ ทองสดายุ โรงเรียนชลราษฎรอำรุง**
   ตำแหน่ง ครูชำนาญการ
   ปริญญาตรี ศึกษาศาสตรบัณฑิต สาขาคณิตศาสตร์

4. **อาจารย์วรรณะ ธีระวิทย์ โรงเรียนชลบุรี"สุขบท"**
   ตำแหน่ง ครูชำนาญการพิเศษ
   ปริญญาตรี ศึกษาศาสตรบัณฑิต สาขาคณิตศาสตร์

5. **อาจารย์สุนทร ทิพย์วงษ์ โรงเรียนชลบุรี"สุขบท"**
   ตำแหน่ง ครูชำนาญการพิเศษ
   ปริญญาตรี ศึกษาศาสตรบัณฑิต สาขาคณิตศาสตร์

6. **อาจารย์พรรณี ทองดี โรงเรียนชลบุรี"สุขบท"**
   ตำแหน่ง ครูชำนาญการพิเศษ
   ประกาศนียบัตรชั้นสูง คณิตศาสตร์
   ปริญญาตรี ศึกษาศาสตรบัณฑิต สาขาการแนะแนว

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

7. **อาจารย์วัชรา จรูญผล โรงเรียนชลบุรี"สุขบท"**

ตำแหน่ง ครูชำนาญการ

ปริญญาตรี วิทยาศาสตรบัณฑิต สาขาสถิติ

ปริญญาโท วิทยาศาสตรมหาบัณฑิต สาขาเทคโนโลยีวิจัยการศึกษา

8. **อาจารย์ทวีศักดิ์ เจริญเตีย โรงเรียนหนองใหญ่ศิริวรวาทวิทยา**

ตำแหน่ง ครูชำนาญการ

ปริญญาตรี ศึกษาศาสตรบัณฑิต สาขาวิทยาศาสตร์-คณิตศาสตร์

ปริญญาโท ศึกษาศาสตรมหาบัณฑิต สาขาหลักสูตรและการสอน

กำลังศึกษาต่อระดับปริญญาเอก สาขาหลักสูตรและการสอน

9. **อาจารย์อรุณี ตั้นงาม โรงเรียนบ้านสวน(จั่นอนุสรณ์)**

ตำแหน่ง ครูชำนาญการพิเศษ

ปริญญาตรี ศึกษาศาสตรบัณฑิต สาขาคณิตศาสตร์

10. **อาจารย์ศศิธร จันทวี โรงเรียนบ้านสวน(จั่นอนุสรณ์)**

ตำแหน่ง ครูชำนาญการ

ปริญญาตรี ศึกษาศาสตรบัณฑิต สาขาคณิตศาสตร์

11. **อาจารย์อาภรณ์ ดวงรัตน์ โรงเรียนบ้านสวน(จั่นอนุสรณ์)**

ตำแหน่ง ครูชำนาญการ

ปริญญาตรี ศึกษาศาสตรบัณฑิต สาขาคณิตศาสตร์

12. **อาจารย์เบญจา อ่วมนุช โรงเรียนราษฎรอำรุง2**

ตำแหน่ง ครูชำนาญการพิเศษ

ปริญญาตรี ศึกษาศาสตรบัณฑิต สาขาวิทยาศาสตร์-คณิตศาสตร์

ปริญญาโท ศึกษาศาสตรมหาบัณฑิต สาขาชีววิทยา

กำลังศึกษาต่อระดับปริญญาเอก สาขาหลักสูตรและการสอน

ภาคผนวก ค

คู่มือการใช้การกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

คู่มือการใช้
การกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง

โดย
นางสุรีพร อนุศาสนนันท์
อาจารย์ที่ปรึกษาวิทยานิพนธ์ รศ.ดร.ศิริเดช สุชีวะ
อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม ดร.ชูศักดิ์ ขัมภลิขิต
ภาควิชาวิจัยและจิตวิทยาการศึกษา คณะครุศาสตร์
จุฬาลงกรณ์มหาวิทยาลัย

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

วัตถุประสงค์ในการสร้างและพัฒนากำหนดมาตรฐานวิธีแองกอฟที่ได้รับการปรับปรุง

วิธีการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง ได้พัฒนาขึ้นเพื่อให้ผู้ตัดสินหรือครูที่มีประสบการณ์ในวิชานั้นๆ ใช้เป็นแนวทางในการกำหนดคะแนนมาตรฐานสำหรับแบบสอบในวิชาต่างๆ ผู้ตัดสินสามารถนำคู่มือดังกล่าวมาใช้ในการกำหนดมาตรฐานในสถานการณ์ต่างๆ เช่นการสอบที่มีลักษณะส่งผลกระทบต่ออนาคตนักเรียนสูง (high stake) กล่าวคือ การสอบเข้ามหาวิทยาลัยหรือสถาบันการศึกษาต่างๆ การสอบในแต่ละช่วงชั้น การสอบในชั้นเรียน สำหรับวิธีแองกอฟที่ได้รับการปรับปรุงเป็นวิธีการที่นิยมอย่างแพร่หลาย นอกจากนี้ผู้วิจัยได้เพิ่มเติมขั้นตอนบางประการเข้าไปในวิธีการนี้เพื่อให้มีประสิทธิภาพยิ่งขึ้น

คำนิยามปฏิบัติการ

1. **มาตรฐาน** (Standard) หมายถึง ระดับที่บ่งบอกคุณภาพของนักเรียนในวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ในการศึกษาครั้งนี้ได้แบ่งระดับของนักเรียนเป็น 8 ระดับ คือ 1) ระดับดีเยี่ยม (A) 2) ระดับดีมาก (B+) 3) ระดับดี (B) 4) ระดับดีพอใช้ (C+) 5) ระดับพอใช้ (C) 6) ระดับอ่อน (D+) 7) ระดับอ่อนมาก (D) และ 8) ระดับตก (F) และในแต่ละระดับจะมีคำอธิบายระดับการปฏิบัติ (performance level descriptor) ที่เป็นข้อความในแต่ละระดับเพื่อแสดงว่านักเรียนควรมีความรู้ ทักษะ หรือคุณลักษณะในแต่ละระดับอย่างไร

2. **คะแนนจุดตัด** (cut score) หมายถึง จุดคะแนนที่ใช้เป็นจุดแบ่งนักเรียนในวิชานั้นตามความสามารถของนักเรียน ซึ่งในงานวิจัยนี้แบ่งคะแนนจุดตัดออกเป็น 7 จุด คือ 1) จุดคะแนนที่แบ่งความสามารถนักเรียนที่มีความสามารถอ่อนมาก ออกจากนักเรียนระดับตก 2) จุดคะแนนที่แบ่งนักเรียนที่มีความสามารถระดับอ่อนมาก ออกจากนักเรียนระดับอ่อน 3) จุดคะแนนที่แบ่งนักเรียนที่มีความสามารถระดับอ่อน ออกจากนักเรียนระดับพอใช้ 4) จุดคะแนนที่แบ่งนักเรียนที่มีความสามารถระดับพอใช้ ออกจากนักเรียนระดับดีพอใช้ 5) จุดคะแนนที่แบ่งนักเรียนที่มีความสามารถระดับดีพอใช้ ออกจากนักเรียนระดับดี 6) จุดคะแนนที่แบ่งนักเรียนที่มีความสามารถระดับดีออกจากนักเรียนระดับดีมาก 7) จุดคะแนนที่แบ่งนักเรียนที่มีความสามารถระดับดีมาก ออกจากนักเรียนระดับดีเยี่ยม

3. **วิธีการกำหนดมาตรฐาน** (Standard setting) หมายถึง กระบวนการหามาตรฐานเพื่อแบ่งความสามารถของนักเรียนออกเป็น 8 ระดับ คือ 1) ระดับดีเยี่ยม(A) 2) ระดับดีมาก(B+) 3) ระดับดี(B) 4) ระดับดีพอใช้ (C+) 5)ระดับพอใช้ (C) 6) ระดับอ่อน (D+) 7) ระดับอ่อนมาก (D) และ 8) ระดับตก (F)

4. **ผู้ตัดสิน** หมายถึง ครูผู้สอนวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ในโรงเรียนสังกัดสำนักงานการศึกษาขั้นพื้นฐานทุกคนที่มีคุณสมบัติดังนี้ 1) มีความเชี่ยวชาญในเนื้อหาคณิตศาสตร์มัธยมศึกษาตอนต้น คือ จบการศึกษาระดับปริญญาตรีขึ้นไปในสาขาทางคณิตศาสตร์ และ 2) มีความเชี่ยวชาญในการสอน และการวัดผล คือ มีประสบการณ์การสอนมาไม่น้อยกว่า 5 ปี และ เป็นครูชำนาญการ หรือชำนาญการพิเศษ

5. **วิธีการแองกอฟที่ได้รับการปรับปรุง (Modified Angoff method)** หมายถึง วิธีการหามาตรฐาน โดยให้ผู้ตัดสินประมาณระดับความสามารถขั้นต่ำของผู้สอบใน ข้อสอบแต่ละข้อจนครบทุกข้อ หลังจากนั้นผู้ตัดสินแต่ละคนร่วมกันอภิปรายผลการตัดสินของแต่ละคน ซึ่งระหว่างการอภิปรายผู้อำนวยความสะดวกจะแสดงรูปการแจกแจงผลการตัดสินของผู้ตัดสินแต่ละคน เมื่อผู้ตัดสินอภิปรายเสร็จก็เปิดโอกาสให้ตัดสินคะแนนอีกครั้งหนึ่ง ผลการตัดสินที่ได้จากผู้ตัดสินแต่ละคนในรอบที่สองนำมาหาค่าเฉลี่ยซึ่งจะเป็นจุดตัดคะแนน นอกจากนี้ผู้วิจัยยังได้พัฒนาวิธีการแองกอฟที่ได้รับการปรับปรุง โดยเพิ่มสารสนเทศประกอบการตัดสินแก่ผู้ตัดสินอีกอย่างคือ กราฟแจกแจงความถี่ของคะแนนสอบ

การเตรียมการก่อนการสัมมนา

1. จัดเตรียมรายชื่อผู้ตัดสินที่มีคุณสมบัติตามที่ต้องการ อาทิเช่น เป็นผู้มีประสบการณ์การสอนและการวัดประเมินในชั้นเรียนในเนื้อหาวิชาที่ทำการสอนมาไม่น้อยกว่า 5 ปี
2. จัดทำหนังสือเรียนเชิญอาจารย์ที่มีคุณสมบัติดังกล่าวมาเป็นผู้ตัดสิน
3. จัดเตรียมเอกสารการจัดสัมมนาดังต่อไปนี้

3.1 คู่มือการใช้การกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง

3.2 ค่าความยากของข้อสอบในรูปฮิสโตแกรม กล่าวคือ แกนนอนเป็นค่าความยาก แกนตั้งเป็นจำนวนข้อสอบ

3.3 แบบฟอร์มประมาณค่าเปอร์เซ็นต์การตอบข้อสอบแต่ละข้อถูกของผู้สอบที่มีความสามารถระดับต่างๆ เช่น ถ้าต้องการแบ่งเกรดเป็น 8 ระดับ คือ ระดับดีเยี่ยม (A) ระดับดีมาก (B+) ระดับดี (B) ระดับดีพอใช้ (C+) ระดับพอใช้ (C) ระดับอ่อน (D+) ระดับอ่อนมาก (D) และระดับตก (F) ผู้ตัดสินก็ต้องทำการตัดสินคะแนนจุดตัด 7 ระดับคือ ระดับดีเยี่ยม ระดับดีมาก ระดับดี ระดับดีพอใช้ ระดับพอใช้ ระดับอ่อน ระดับอ่อนมาก ดังนั้นผู้ตัดสินแต่ละคนก็จะได้รับแบบฟอร์มประมาณค่าเปอร์เซ็นต์การตอบข้อสอบแต่ละข้อถูกของผู้สอบที่มีความสามารถทั้ง 7 ระดับ นั้นคือผู้ตัดสินแต่ละคนจะได้รับแบบบฟอร์มคนละ 7 ฉบับนั้นเอง

3.4 แบบฟอร์มเขียนคำบรรยายการปฏิบัติแต่ละระดับ

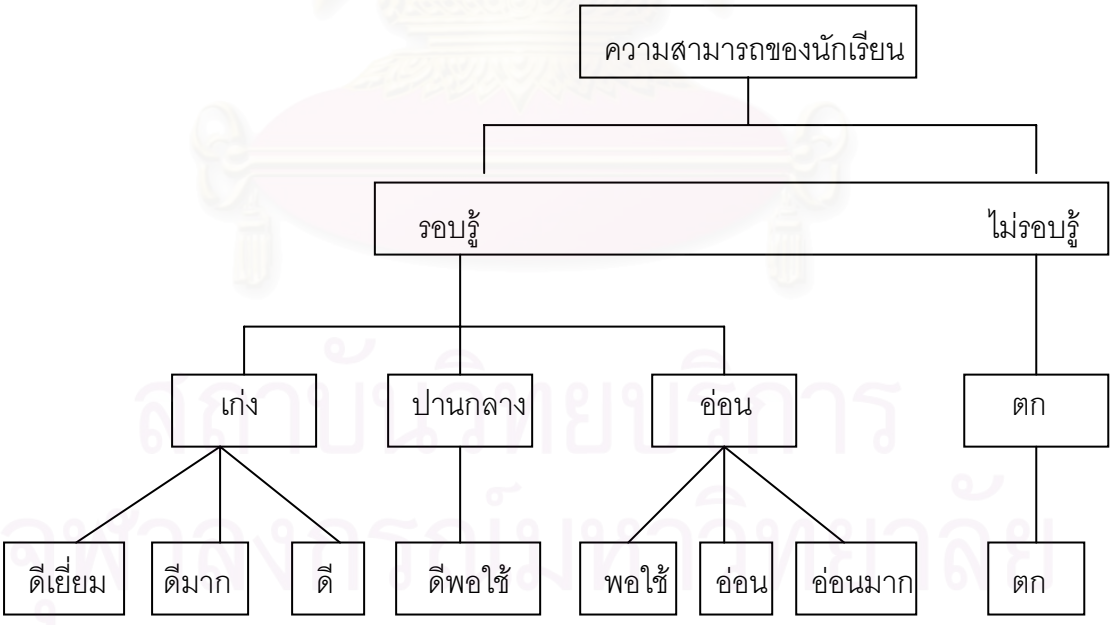
3.5 คำอธิบายลักษณะความสามารถในวิชานั้นของนักเรียนในแต่ละระดับ เช่น นักเรียนที่มีความสามารถระดับอ่อนมาก หมายถึง นักเรียนที่มีความสามารถถึงระดับที่คาดหวังขั้นต่ำสำหรับการปฏิบัติของนักเรียนในวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ตามหลักสูตรกระทรวงศึกษาธิการ นั้นคือ นักเรียนมีความรู้ที่จำเป็นต้องมี และมีทักษะขั้นพื้นฐานในการปฏิบัติงานในระดับคาบเส้น

3.6 ค่าความยากของข้อสอบแต่ละข้อ

3.7 แบบฟอร์มแสดงค่าเฉลี่ยผลการตัดสินของผู้ตัดสินแต่ละคนในแต่ละระดับ

ขั้นตอนการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง

1. ก่อนการกำหนดมาตรฐาน ผู้วิจัยแนะนำลำดับการพิจารณาคะแนนจุดตัดแก่ผู้ตัดสิน มีรายละเอียดคือ ขั้นที่ 1 พิจารณานักเรียน 2 กลุ่มคือ รอบรู้ และไม่รอบรู้ ขั้นที่ 2 กลุ่มรอบรู้แยกเป็น กลุ่มเก่ง กลุ่มปานกลาง และกลุ่มอ่อน ส่วนกลุ่มไม่รอบรู้ แยกเป็นกลุ่มตก และขั้นที่ 3 แยกกลุ่มเก่ง เป็น กลุ่มดีเยี่ยม กลุ่มดีมาก และกลุ่มดี กลุ่มปานกลาง คือ กลุ่มดีพอใช้ กลุ่มอ่อน แยกเป็น กลุ่มพอใช้ กลุ่มอ่อน และ กลุ่มอ่อนมาก ดังแผนภาพที่ 1 ลำดับขั้นการพิจารณาความสามารถของนักเรียน 8 ระดับ



แผนภาพที่ 1 ลำดับขั้นการพิจารณาความสามารถของนักเรียน 8 ระดับ

2. จากนั้นผู้ตัดสิน 12 คน ร่วมอบรมเชิงปฏิบัติการเพื่อรับฟังวิธีการกำหนดมาตรฐานด้วยวิธีการแองกอฟที่ได้รับการปรับปรุง พร้อมทั้งปฏิบัติการกำหนดมาตรฐานทั้งสองวิธีเพื่อให้เกิดความเข้าใจในวิธีการกำหนดมาตรฐาน

3. ผู้วิจัยแจกคำอธิบายลักษณะความสามารถของนักเรียนแต่ละระดับ เพื่อให้ผู้ตัดสินเข้าใจตรงกัน พร้อมทั้งแจกค่าความยากของข้อสอบแต่ละข้อ แบบสอบทั้ง 4 ฉบับ โครงสร้างเนื้อหาของแบบสอบทั้ง 4 ฉบับ นอกจากนี้ผู้วิจัยยังได้เพิ่มข้อมูลประกอบการตัดสินแก่ผู้ตัดสินอีกอย่างคือ กราฟแจกแจงความถี่ของคะแนนสอบ

4. ดำเนินการกำหนดเปอร์เซ็นต์การตอบข้อสอบแต่ละข้อถูก จากผู้สอบที่มีความสามารถแต่ละระดับ มี 3 ขั้นตอนดังนี้

<u>ขั้นที่ 1</u> รอบที่ 1 ดำเนินการตัดสินเป็นรายบุคคล ดังนี้

ผู้ตัดสินแต่ละคนพิจารณาข้อสอบทีละข้อตั้งแต่ข้อสอบข้อที่ 1 เป็นต้นไป เพื่อกำหนดคะแนนจุดตัด โดยข้อสอบที่พิจารณาจะมีลำดับข้อสอบเหมือนกับแบบสอบที่สอบกับนักเรียน โดยตัดสินว่า

1) กลุ่มผู้สอบที่มีความสามารถ<u>ระดับอ่อนมาก</u>มีโอกาสตอบข้อสอบข้อนั้นถูกกี่เปอร์เซนต์ นั้นหมายความว่า ผู้ตัดสินคาดว่า นักเรียนที่มีความสามารถระดับอ่อนมากประมาณกี่คนจากทั้งหมด 100 คนสามารถตอบข้อสอบข้อนั้นถูก จากนั้นกรอกตัวเลขลงในแบบประมาณเปอร์เซ็นต์การตอบข้อสอบแต่ละข้อถูกที่มีความสามารถระดับอ่อนมาก(D)ในช่องเปอร์เซ็นต์การตอบถูก รอบที่ 1 (รายบุคคล) จากนั้นพิจารณาข้อสอบข้อถัดไปเรื่อยๆ จนครบจำนวนข้อสอบทั้งหมด

2) กลุ่มผู้สอบที่มีความสามารถ<u>ระดับอ่อน</u>มีโอกาสตอบข้อสอบข้อนั้นถูกกี่เปอร์เซนต์ นั้นหมายความว่า ผู้ตัดสินคาดว่า นักเรียนที่มีความสามารถระดับอ่อนประมาณกี่คนจากทั้งหมด 100 คนสามารถตอบข้อสอบข้อนั้น ถูก จากนั้นกรอกตัวเลขลงในแบบประมาณเปอร์เซ็นต์การตอบข้อสอบแต่ละข้อถูกที่มีความสามารถระดับอ่อน(D+) ในช่องเปอร์เซ็นต์การตอบถูก รอบที่ 1 (รายบุคคล) จากนั้นพิจารณาข้อสอบข้อถัดไปเรื่อยๆ จนครบจำนวนข้อสอบทั้งหมด

3) กลุ่มผู้สอบที่มีความสามารถ<u>ระดับพอใช้</u>มีโอกาสตอบข้อสอบข้อนั้นถูกกี่เปอร์เซนต์ นั้นหมายความว่า ผู้ตัดสินคาดว่า นักเรียนที่มีความสามารถระดับพอใช้ประมาณกี่คนจากทั้งหมด 100 คนสามารถตอบข้อสอบข้อนั้นถูก จากนั้นกรอกตัวเลขลงในแบบประมาณเปอร์เซ็นต์การตอบข้อสอบแต่ละข้อถูกที่มีความสามารถระดับอ่อน (C) ในช่อง

เปอร์เซ็นต์การตอบถูก รอบที่ 1 (รายบุคคล) จากนั้นพิจารณาข้อสอบข้อถัดไปเรื่อยๆ จนครบจำนวนข้อสอบทั้งหมด

4) กลุ่มผู้สอบที่มีความสามารถ<u>ระดับดีพอใช้</u>มีโอกาสตอบข้อสอบข้อนั้นถูกกี่เปอร์เซนต์ นั้นหมายความว่า ผู้ตัดสินคาดว่า นักเรียนที่มีความสามารถระดับดีพอใช้ประมาณกี่คนจากทั้งหมด 100 คนสามารถตอบข้อสอบข้อนั้นถูก จากนั้นกรอกตัวเลขลงในแบบประมาณเปอร์เซ็นต์การตอบข้อสอบแต่ละข้อถูกที่มีความสามารถระดับดีพอใช้ (C+) ในช่องเปอร์เซ็นต์การตอบถูก รอบที่ 1 (รายบุคคล) จากนั้นพิจารณาข้อสอบข้อถัดไปเรื่อยๆ จนครบจำนวนข้อสอบทั้งหมด

5) กลุ่มผู้สอบที่มีความสามารถ<u>ระดับดี</u>มีโอกาสตอบข้อสอบข้อนั้นถูกกี่เปอร์เซนต์ นั้นหมายความว่า ผู้ตัดสินคาดว่า นักเรียนที่มีความสามารถระดับพอใช้ ประมาณกี่คนจากทั้งหมด 100 คนสามารถตอบข้อสอบข้อนั้นถูก จากนั้นกรอกตัวเลขลงในแบบประมาณเปอร์เซ็นต์การตอบข้อสอบแต่ละข้อถูกที่มีความสามารถระดับดี(B) ในช่องเปอร์เซ็นต์การตอบถูก รอบที่ 1 (รายบุคคล) จากนั้นพิจารณาข้อสอบข้อถัดไปเรื่อยๆ จนครบจำนวนข้อสอบทั้งหมด

6) กลุ่มผู้สอบที่มีความสามารถ<u>ระดับดีมาก</u>มีโอกาสตอบข้อสอบข้อนั้นถูกกี่เปอร์เซนต์ นั้นหมายความว่า ผู้ตัดสินคาดว่า นักเรียนที่มีความสามารถระดับพอใช้ประมาณกี่คนจากทั้งหมด 100 คนสามารถตอบข้อสอบข้อนั้นถูก จากนั้นกรอกตัวเลขลงในแบบประมาณเปอร์เซ็นต์การตอบข้อสอบแต่ละข้อถูกที่มีความสามารถระดับดีมาก(B+) ในช่องเปอร์เซ็นต์การตอบถูก รอบที่ 1 (รายบุคคล) จากนั้นพิจารณาข้อสอบข้อถัดไปเรื่อยๆ จนครบจำนวนข้อสอบทั้งหมด

7) กลุ่มผู้สอบที่มีความสามารถ<u>ระดับดีเยี่ยม</u>มีโอกาสตอบข้อสอบข้อนั้นถูกกี่เปอร์เซนต์ นั้นหมายความว่า ผู้ตัดสินคาดว่า นักเรียนที่มีความสามารถระดับพอใช้ ประมาณกี่คนจากทั้งหมด 100 คนสามารถตอบข้อสอบข้อนั้นถูก จากนั้นกรอกตัวเลขลงในแบบประมาณเปอร์เซ็นต์การตอบข้อสอบแต่ละข้อถูกที่มีความสามารถระดับดีเยี่ยม(A) ในช่องเปอร์เซ็นต์การตอบถูก รอบที่ 1 (รายบุคคล) จากนั้นพิจารณาข้อสอบข้อถัดไปเรื่อยๆ จนครบจำนวนข้อสอบทั้งหมด

<u>ขั้นที่ 2</u> รอบที่ 2 ดำเนินการแบ่งผู้ตัดสินเป็นกลุ่มย่อย 2 กลุ่ม ดังนี้

1) ในแต่ละกลุ่มย่อย ผู้ตัดสินร่วมกันอภิปรายผลการตัดสินของแต่ละคน โดยเฉพาะผู้ตัดสินที่มีผลตัดสินต่ำสุด และสูงสุดให้อธิบายแสดงความคิดเห็นถึงเหตุผลที่ให้ผลการ

ตัดสินต่ำสุด และสูงสุด และระหว่างการอภิปรายผู้อำนวยความสะดวกจะแสดงรูปการแจกแจงผลการตัดสินของผู้ตัดสินแต่ละคน

2) เมื่อผู้ตัดสินอภิปรายเสร็จ ก็เปิดโอกาสให้กำหนดคะแนนจุดตัดใหม่อีกครั้งหนึ่ง นั้นคือเป็นการเปิดโอกาสให้ผู้ตัดสินปรับเปลี่ยนคะแนนจุดตัด หรือยังคงคะแนนจุดตัดเดิมก็ได้

3) จากนั้นนำผลการตัดสินของผู้ตัดสินทุกคนจากแต่ละกลุ่ม โดยผู้ตัดสินทุกคนจะมีคะแนนจุดตัด 7 ระดับ คือ 1) ระดับดีเยี่ยม 2) ระดับดีมาก 3) ระดับดี 4) ระดับดีพอใช้ 5) ระดับพอใช้ 6) ระดับอ่อน 7) ระดับอ่อนมาก นำคะแนนจุดตัดของผู้ตัดสินแต่ละคนมาหาค่าเฉลี่ยในแต่ละระดับ จะได้ค่าเฉลี่ยของคะแนนจุดตัดทั้งหมด 7 ระดับ หรือมีมาตรฐานทั้งสิ้น 8 ระดับนั้นเอง

<u>ขั้นที่ 3</u> รอบที่ 3 ดำเนินการรวมผู้ตัดสินทั้งสองกลุ่มย่อยเป็นกลุ่มใหญ่เพียงกลุ่มเดียว แล้วดำเนินการดังนี้

1) นำผู้ตัดสินทั้งสองกลุ่มมารวมเป็นกลุ่มเดียวกัน ผู้อำนวยความสะดวกแสดงค่าเฉลี่ยคะแนนจุดตัดของแต่ละกลุ่มให้ผู้ตัดสินพิจารณา เปิดโอกาสให้ผู้ตัดสินอภิปรายร่วมกัน

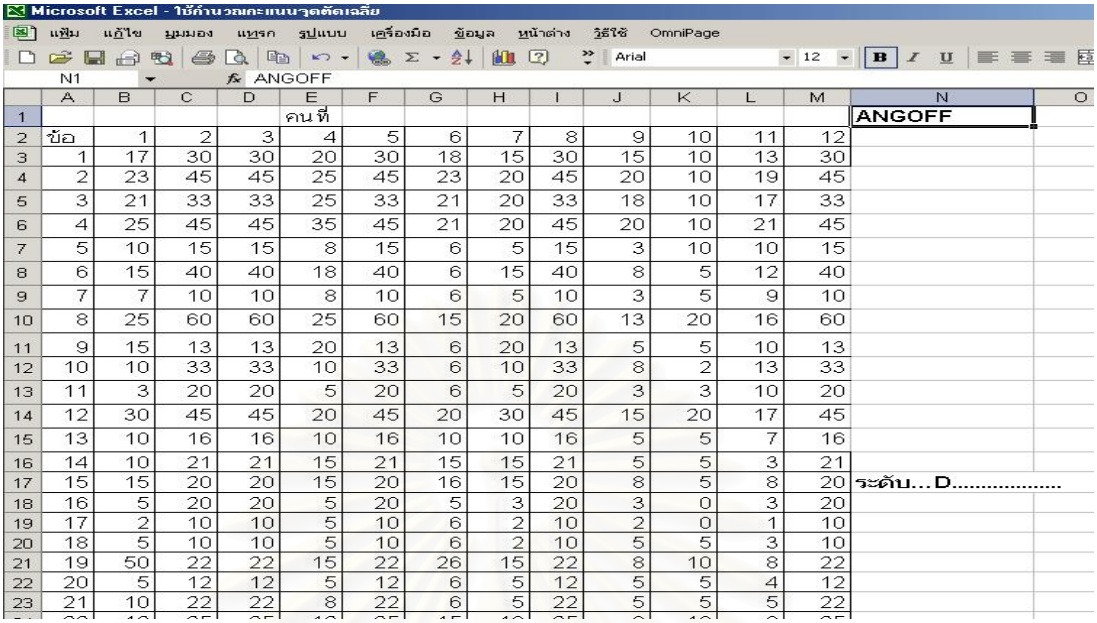
2) จากนั้นเปิดโอกาสให้ผู้ตัดสินทำการตัดสินคะแนนจุดตัดทั้ง 7 ระดับอีกครั้งหนึ่งอย่างอิสระจากกัน

3) นำผลการตัดสินของผู้ตัดสินแต่ละคนมาหาค่าเฉลี่ยของแต่ละระดับ

4) เมื่อได้คะแนนจุดตัดทั้ง 7 ระดับแล้ว ให้ผู้ตัดสินระดมพลังสมองเพื่อเขียนบรรยายระดับการปฏิบัติของนักเรียนในประเด็นความรู้ ทักษะ และความสามารถของนักเรียน ณ ระดับคะแนนจุดตัดของระดับนั้น รวมทั้งสิ้น 8 ระดับ

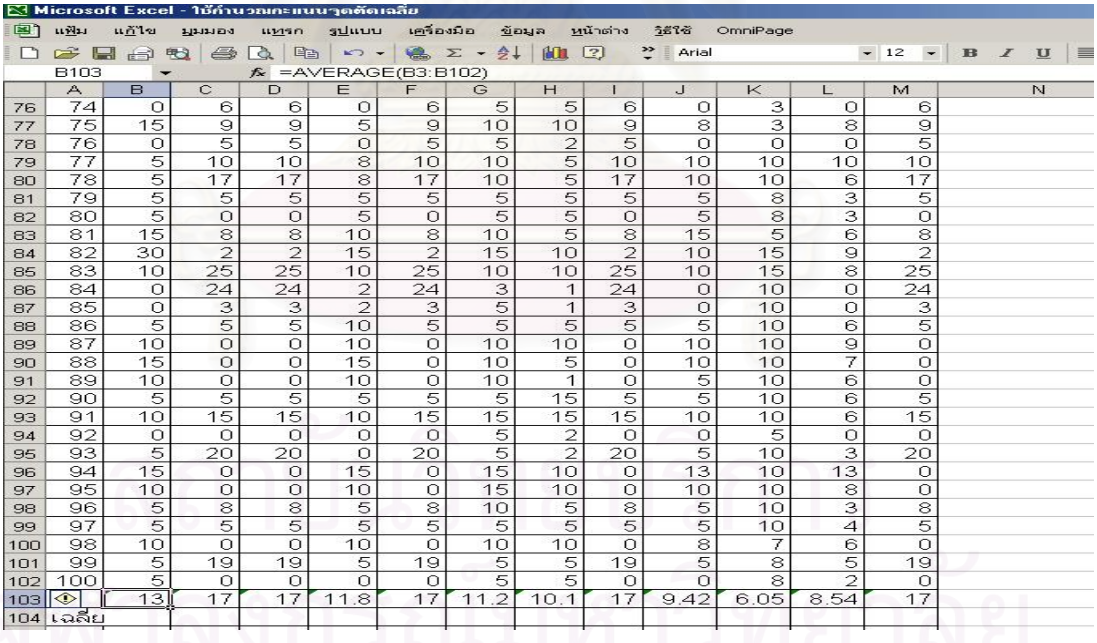
ตัวอย่างการวิเคราะห์เพื่อคำนวณหาคะแนนจุดตัดที่ระดับอ่อนมาก (D) ในโปรแกรม Excel

1. ผู้ตัดสิน 12 คนทำการกำหนดเปอร์เซ็นต์ลงในแบบฟอร์มโดยคิดว่า กลุ่มผู้สอบที่มีความสามารถ<u>ระดับอ่อนมาก</u>มีโอกาสตอบข้อสอบข้อนั้นถูกกี่เปอร์เซนต์ นั้นคือ ผู้ตัดสินคาดว่านักเรียนที่มีความสามารถระดับอ่อนมากประมาณกี่คนจากทั้งหมด 100 คนสามารถตอบข้อสอบข้อนั้นถูก จากตัวอย่างผู้ตัดสินดำเนินการกำหนดเปอร์เซ็นต์ทั้งหมด 100 ข้อ



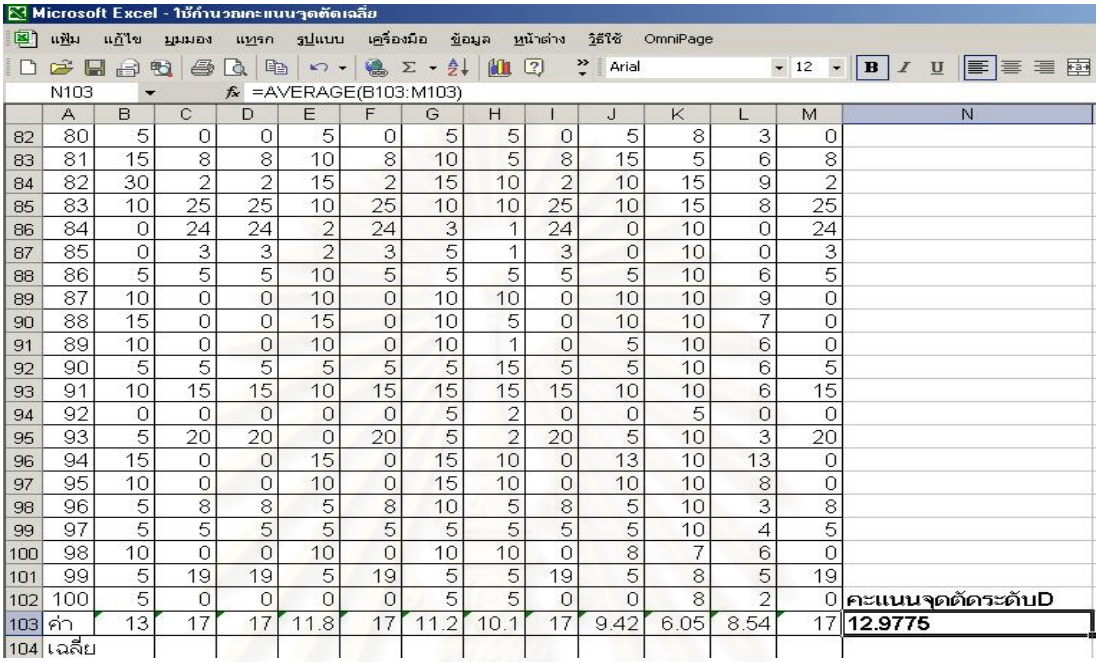
2. คำนวณค่าเฉลี่ยข้อสอบ 100 ข้อ ของผู้ตัดสินแต่ละคน

ตัวอย่างเช่น ผู้ตัดสินคนที่ 1 คำสั่งค่าเฉลี่ยคือ AVERAGE(B3:B102) จากนั้นหาค่าเฉลี่ยให้ครบทั้ง 12 คน



3. คำนวณคะแนนจุดตัดที่ระดับอ่อนมาก โดยนำค่าเฉลี่ยของผู้ตัดสินทั้งหมด 12 คนมาหาเฉลี่ย

ตัวอย่างเช่น คะแนนจุดตัดระดับอ่อนมาก คำสั่งค่าเฉลี่ยคือ AVERAGE(B103:M103)



	A	B	C	D	E	F	G	H	I	J	K	L	M	N
82	80	5	0	0	5	0	5	5	0	5	8	3	0	
83	81	15	8	8	10	8	10	5	8	15	5	6	8	
84	82	30	2	2	15	2	15	10	2	10	15	9	2	
85	83	10	25	25	10	25	10	10	25	10	15	8	25	
86	84	0	24	24	2	24	3	1	24	0	10	0	24	
87	85	0	3	3	2	3	5	1	3	0	10	0	3	
88	86	5	5	5	10	5	5	5	5	5	10	6	5	
89	87	10	0	0	10	0	10	10	0	10	10	9	0	
90	88	15	0	0	15	0	10	5	0	10	10	7	0	
91	89	10	0	0	10	0	10	1	0	5	10	6	0	
92	90	5	5	5	5	5	5	15	5	5	10	6	5	
93	91	10	15	15	10	15	15	15	15	10	10	6	15	
94	92	0	0	0	0	0	5	2	0	0	5	0	0	
95	93	5	20	20	0	20	5	2	20	5	10	3	20	
96	94	15	0	0	15	0	15	10	0	13	10	13	0	
97	95	10	0	0	10	0	15	10	0	10	10	8	0	
98	96	5	8	8	5	8	10	5	8	5	10	3	8	
99	97	5	5	5	5	5	5	5	5	5	10	4	5	
100	98	10	0	0	10	0	10	10	0	8	7	6	0	
101	99	5	19	19	5	19	5	5	19	5	8	5	19	
102	100	5	0	0	0	0	5	5	0	0	8	2	0	คะแนนจุดตัดระดับD
103	ค่า	13	17	17	11.8	17	11.2	10.1	17	9.42	6.05	8.54	17	12.9775
104	เฉลี่ย													

การหาค่าความยากจากทฤษฎีการตอบสนองข้อสอบโดยโปรแกรม Xcalibre for Windows95/NT(Version 1.10)

ขั้นที่ 1 การเตรียมข้อมูลลงใน Noteped

หมายเลข 1 มีดังนี้

30 คือ จำนวนข้อสอบ (อยู่ระหว่าง column 1-3) เช่น ถ้าแบบสอบมีข้อสอบ 30 ข้อ ให้พิมพ์ 30 ใน column 2-3 ถ้าแบบสอบมีข้อสอบ 100 ข้อให้พิมพ์ 100 ใน column 1-3

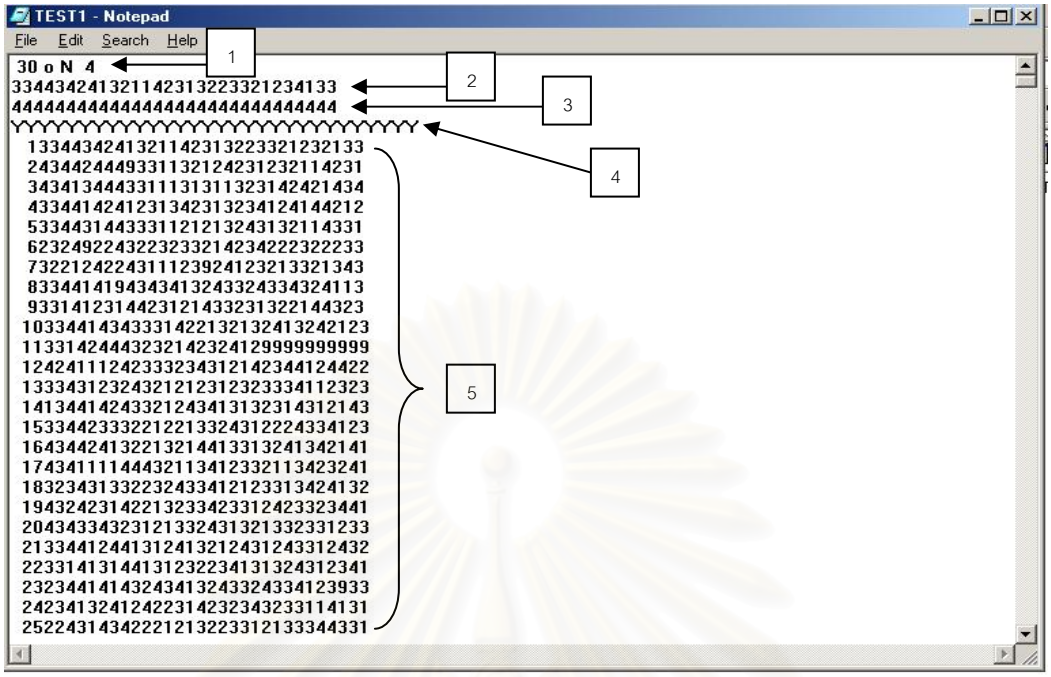
4 คือ จำนวนตัวเลือก

หมายเลข 2 คือ key คำตอบ

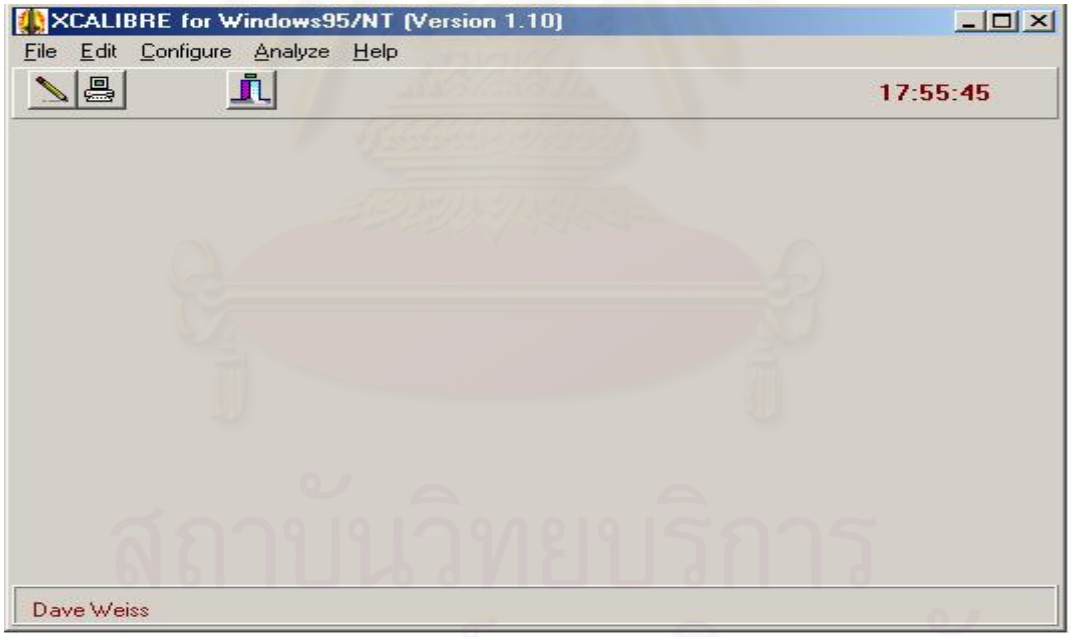
หมายเลข 3 คือ จำนวนตัวเลือก ในที่นี้เป็นข้อสอบแบบเลือกตอบ 4 ตัวเลือก

หมายเลข 4 คือ Y

หมายเลข 5 คือ ข้อมูลการตอบของนักเรียนแต่ละคน

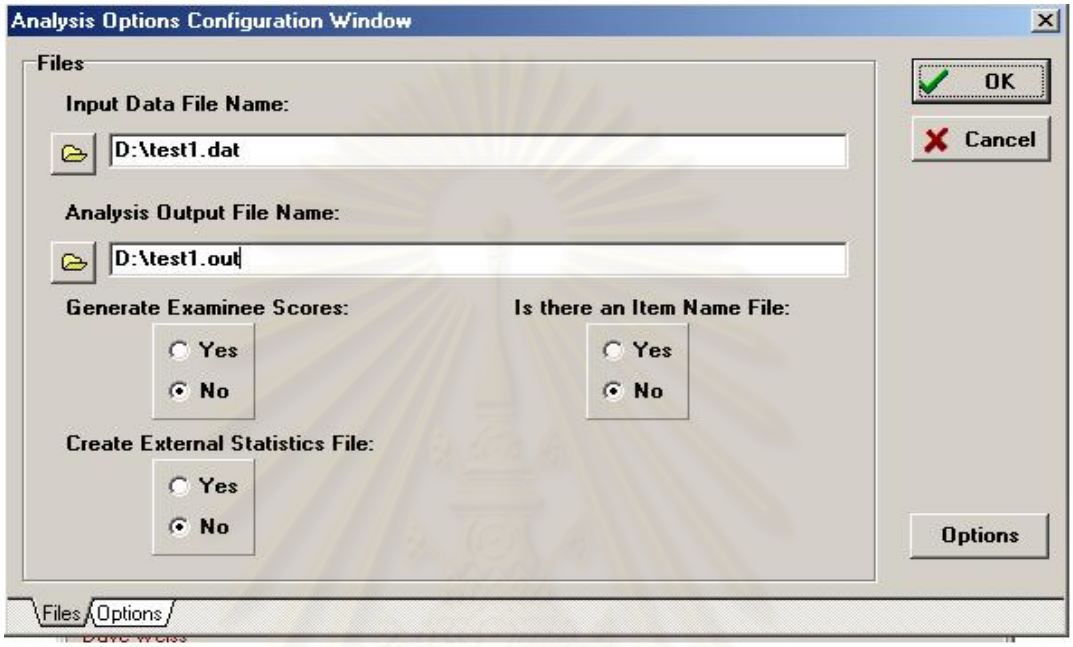


ขั้นที่ 2 เข้าสู่โปรแกรม Xcalibre for Windows95/NT(Version 1.10)



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

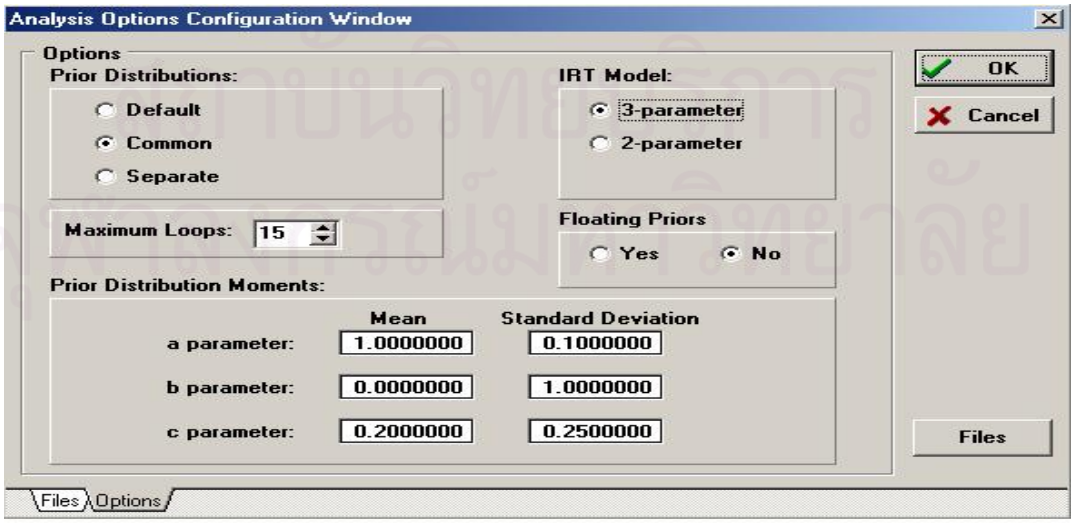
ขั้นที่ 3 พิมพ์ชื่อไฟล์ข้อมูลลงใน Input Data File Name และ
พิมพ์ชื่อไฟล์ผลการวิเคราะห์ลงใน Analysis Output File Name
จากนั้น กดปุ่ม Options



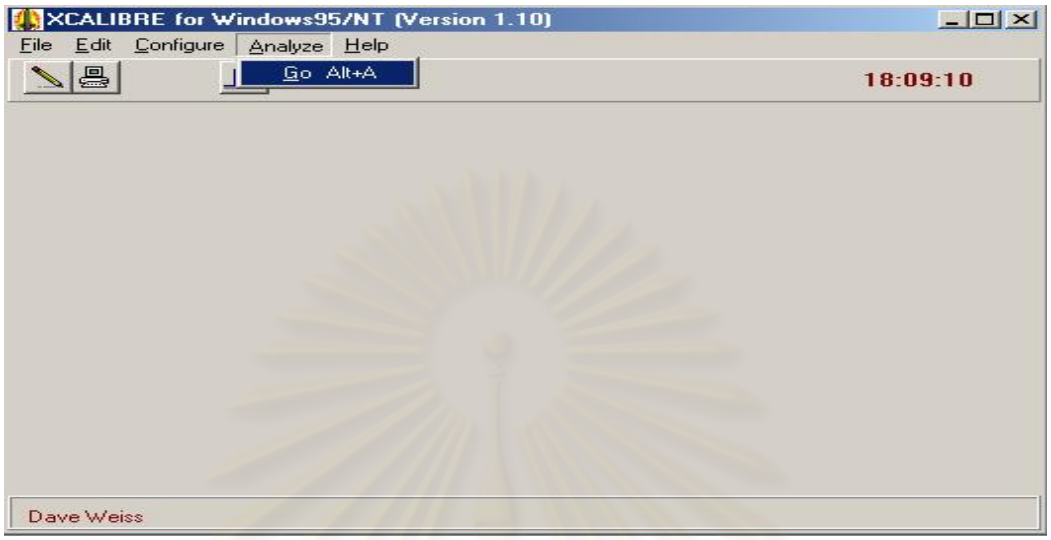
ขั้นที่ 4 เลือกจำนวนพารามิเตอร์ที่ต้องการวิเคราะห์ คือ

2 พารามิเตอร์ ประกอบด้วยค่าความยาก (b) และค่าอำนาจจำแนก (a)

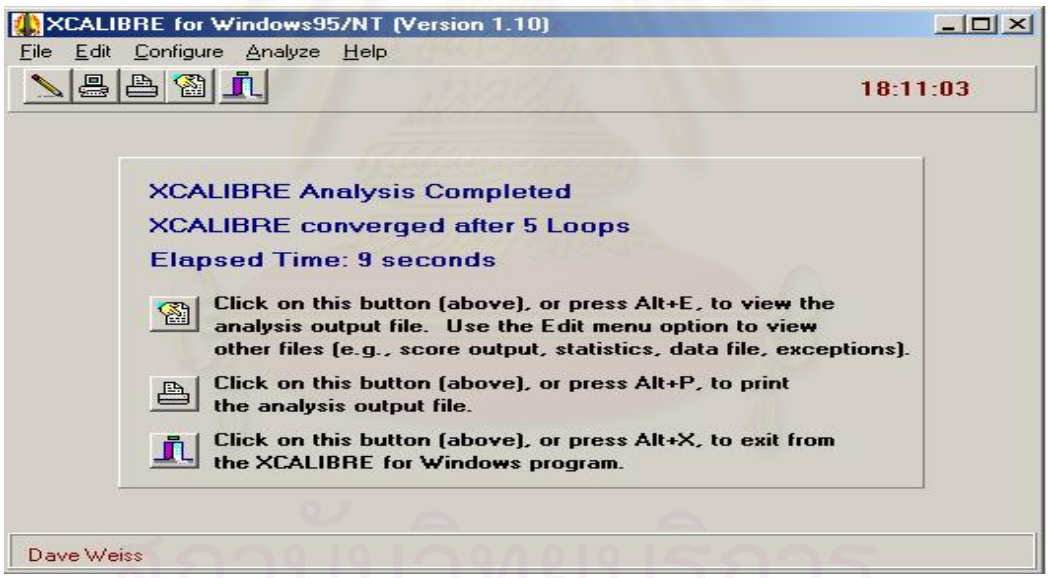
3 พารามิเตอร์ ประกอบด้วยค่าความยาก (b) และค่าอำนาจจำแนก (a) และค่าการเดา (c) จากนั้นกด OK



ขั้นที่ 5 RUN program โดย กดปุ่ม Analyze แล้วไปที่ Go

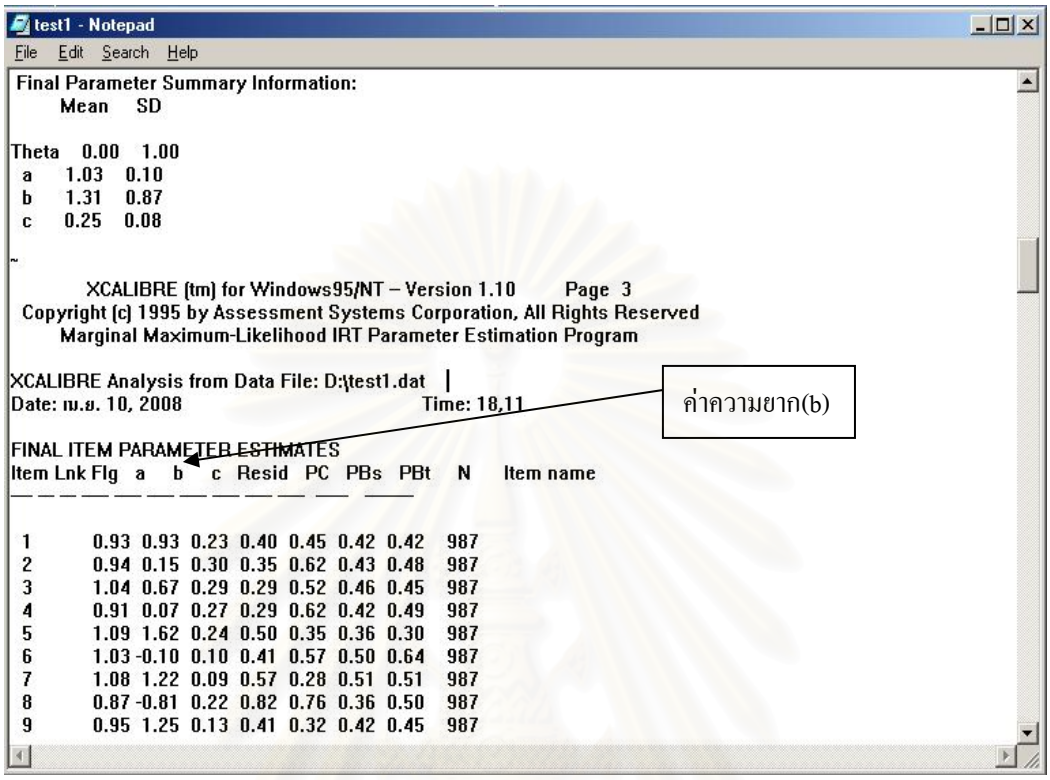


ขั้นที่ 6 เมื่อ Run Program สมบูรณ์จะปรากฏหน้าต่างดังข้างล่าง



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ขั้นที่ 7 ผลการวิเคราะห์ข้อมูลค่าความยาก (b) ปรากฏหน้าต่างข้างล่าง



รายละเอียดต่างๆ ของตัวอย่างแบบฟอร์ม ตลอดจนตัวอย่างสารสนเทศที่ให้ผู้ตัดสินพิจารณา แสดงไว้ในภาคผนวก

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

<u>แบบประมาณเปอร์เซนต์การตอบข้อสอบแต่ละข้อถูกจากผู้สอบที่มีความสามารถระดับอ่อนมาก (D)</u>

ชื่ออาจารย์.....**โรงเรียน**.....

<u>คำชี้แจง</u>

แบบสอบที่แนบมาพร้อมนี้เป็นแบบสอบเลือกตอบวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ภาคเรียนที่ 1 เรื่อง ปริมาตรและพื้นที่ผิว กราฟ ระบบสมการเชิงเส้น ความคล้าย ชนิด 4 ตัวเลือก จำนวน 100 ข้อ

ขอให้ท่านประมาณว่า นักเรียนชั้นมัธยมศึกษาปีที่ 3 ที่มีความสามารถ**ระดับอ่อนมาก**มีโอกาสตอบข้อสอบข้อนั้นถูกกี่เปอร์เซนต์ แล้วกรอกลงในช่องเปอร์เซ็นต์การตอบถูกที่เตรียมไว้ให้ สำหรับข้อสอบทุกข้อ เปอร์เซนต์การตอบถูกมีค่าตั้งแต่ 0 ถึง 100 ดังนี้

0 หมายถึง โอกาสที่นักเรียนจะตอบข้อสอบข้อนั้นถูกน้อยที่สุด

100 หมายถึง โอกาสที่นักเรียนจะตอบข้อสอบข้อนั้นถูกมากที่สุด

<u>ตัวอย่าง</u>

ข้อ(0) 8.3 + 11.5 เท่ากับเท่าไร

ก. 19.8 ข. 18

ค. 20.1 ง. 17.5

1) ถ้าข้อ (0) นี้ ท่านคาดว่านักเรียนที่มีความสามารถระดับ**อ่อนมาก**ประมาณ 60 คน จาก100 คน สามารถตอบข้อสอบข้อนี้ถูกต้อง ดังนั้นเปอร์เซ็นต์ที่จะตอบข้อสอบข้อนี้ถูกจะเท่ากับ 60 % ให้กรอกลงในช่องเปอร์เซ็นต์การตอบถูกที่เตรียมไว้ในช่องรอบที่ 1

2) จากนั้นดำเนินการร่วมกลุ่มย่อย เพื่ออภิปรายผลการตัดสินของแต่ละคนโดยเฉพาะผู้ตัดสินที่มีผลตัดสินสูงสุด และต่ำสุดให้อธิบายแสดงความคิดเห็น จากนั้นให้ตัดสินคะแนนอีกครั้งหนึ่งตามข้อที่ 1 ถ้าท่านต้องการเปลี่ยนเปอร์เซ็นต์การตอบถูกข้อ(0) เป็น 70 % ให้ท่านกรอกลงใน ช่องเปอร์เซ็นต์การตอบถูกที่เตรียมไว้ในช่องรอบที่ 2

3) นำผู้ตัดสินทั้งหมดร่วมเป็นกลุ่มใหญ่ อภิปรายผลการตัดสินร่วมกันอีกครั้ง ในครั้งนี้จะมีค่าเฉลี่ยคะแนนจุดตัดของแต่ละกลุ่มให้พิจารณาอีกครั้ง จากนั้นให้ตัดสินคะแนนเป็นครั้งสุดท้าย ถ้าท่านไม่ต้องการเปลี่ยนเปอร์เซ็นต์การตอบถูกข้อ(0) ให้ท่านกรอกลงในช่องเปอร์เซ็นต์การตอบถูกที่เตรียมไว้ในช่องรอบที่ 3 ดังนี้

ข้อที่	เปอร์เซ็นต์การตอบถูก		
	รอบที่ 1 (รายบุคคล)	รอบที่ 2 (กลุ่มย่อย)	รอบที่ 3 (กลุ่มใหญ่)
0	60	70	70

ชื่ออาจารย์……**โรงเรียน**……

ข้อที่	เปอร์เซ็นต์การตอบถูก		
	รอบที่ 1 (รายบุคคล)	**รอบที่ 2 (กลุ่มย่อย)**	**รอบที่ 3 (กลุ่มใหญ่)**
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			
17			
18			
19			
20			
21			
22			
23			
24			
25			

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ชื่ออาจารย์.....**โรงเรียน**.....

ข้อที่	เปอร์เซ็นต์การตอบถูก		
	รอบที่ 1 (รายบุคคล)	รอบที่ 2 (กลุ่มย่อย)	รอบที่ 3 (กลุ่มใหญ่)
26			
27			
28			
29			
30			
31			
32			
33			
34			
35			
36			
37			
38			
39			
40			
41			
42			
43			
44			
45			
46			
47			
48			
49			
50			

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ชื่ออาจารย์……**โรงเรียน**……

ข้อที่	เปอร์เซ็นต์การตอบถูก		
	รอบที่ 1 (รายบุคคล)	**รอบที่ 2 (กลุ่มย่อย)**	**รอบที่ 3 (กลุ่มใหญ่)**
51			
52			
53			
54			
55			
56			
57			
58			
59			
60			
61			
62			
63			
64			
65			
66			
67			
68			
69			
70			
71			
72			
73			
74			
75			

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ชื่ออาจารย์……**โรงเรียน**……

ข้อที่	เปอร์เซ็นต์การตอบถูก		
	รอบที่ 1 (รายบุคคล)	**รอบที่ 2 (กลุ่มย่อย)**	**รอบที่ 3 (กลุ่มใหญ่)**
76			
77			
78			
79			
80			
81			
82			
83			
84			
85			
86			
87			
88			
89			
90			
91			
92			
93			
94			
95			
96			
97			
98			
99			
100			

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตัวอย่างคำอธิบายลักษณะความสามารถแต่ละระดับของนักเรียน วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ภาคเรียนที่ 1

เนื้อหาในบทเรียนวิชาคณิตศาสตร์ มัธยมศึกษาปีที่ 3 ภาคเรียนที่ 1 ที่นักเรียนควรทำได้

- อธิบายลักษณะและสมบัติของปริซึม พีระมิด ทรงกระบอก กรวย และทรงกลม
- หาพื้นที่ผิวของปริซึมและทรงกระบอก
- หาปริมาตรของปริซึม ทรงกระบอก พีระมิด กรวย และทรงกลม
- เปรียบเทียบหน่วยความจุ หรือหน่วยปริมาตร
- อ่านและแปลความหมายของกราฟของระบบสมการเชิงเส้น
- แก้ระบบสมการเชิงเส้นสองตัวแปร
- นำระบบสมการเชิงเส้นสองตัวแปรไปใช้แก้ปัญหา
- เขียนกราฟของสมการเชิงเส้นสองตัวแปร
- อ่านและแปลกราฟความหมายของกราฟ
- อ่านและแปลความหมายของกราฟของระบบสมการเชิงเส้น
- แก้ระบบสมการเชิงเส้น
- บอกสมบัติของการคล้ายของรูปสามเหลี่ยม และบอกเงื่อนไขที่ทำให้รูปสามเหลี่ยมสองรูปคล้ายกัน
- ใช้สมบัติของรูปสามเหลี่ยมที่คล้ายกันในการให้เหตุผล และแก้ปัญหา

1. **<u>นักเรียนที่มีความสามารถระดับตก</u>** (F) หมายถึง นักเรียนที่มีความสามารถ**ไม่ถึงระดับที่คาดหวังขั้นต่ำ**ว่านักเรียนสามารถปฏิบัติได้ในวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ตามหลักสูตรกระทรวงศึกษาธิการ โดยทำเนื้อหาในบทเรียนวิชานี้**ไม่ได้** คำนวณหรือคิดไม่ได้ นักเรียนระดับนี้ไม่มีแรงจูงใจ ความสนใจ และความรับผิดชอบที่จะทำงานในวิชานี้

2. **<u>นักเรียนที่มีความสามารถระดับอ่อนมาก</u>** (D) หมายถึง นักเรียนที่มีความสามารถ**ถึงระดับที่คาดหวังขั้นต่ำ**สำหรับการปฏิบัติของนักเรียนในวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ตามหลักสูตรกระทรวงศึกษาธิการ นั้นคือ นักเรียนมีความรู้ที่จำเป็นต้องมี และมีทักษะขั้นพื้นฐานในการปฏิบัติงานในระดับคาบเส้น โดยทำเนื้อหาในบทเรียนวิชานี้**ได้น้อย** คำนวณหรือคิดได้บ้างจากโจทย์ที่ไม่ได้พลิกแพลง คิดช้ามาก นักเรียนระดับนี้ไม่ค่อยมีแรงจูงใจ ความสนใจ และความรับผิดชอบที่จะทำงานในวิชานี้

3. **<u>นักเรียนที่มีความสามารถระดับอ่อน</u>** (D+) หมายถึง นักเรียนที่มีความสามารถ**ถึงระดับที่คาดหวังขั้นต่ำขึ้นมาเล็กน้อย** สำหรับการปฏิบัติของนักเรียนในวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ตามหลักสูตรกระทรวงศึกษาธิการ นั้นคือ นักเรียนมีความรู้ที่จำเป็นต้องมี และมีทักษะขั้นพื้นฐานในการปฏิบัติงานในระดับเกินระดับคาบเส้นขึ้นมาไม่มาก โดยทำเนื้อหาในบทเรียนวิชานี้**ได้น้อย** คำนวณหรือคิดได้บ้างจากโจทย์ที่ไม่ได้พลิกแพลง แต่คิดช้า นักเรียนระดับนี้ไม่ค่อยมีแรงจูงใจ ความสนใจ และความรับผิดชอบที่จะทำงานในวิชานี้

4. **<u>นักเรียนที่มีความสามารถระดับพอใช้</u>** (C) หมายถึง นักเรียนที่มีความสามารถ**ถึงระดับที่คาดหวัง**สำหรับการปฏิบัติของนักเรียนในวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ตามหลักสูตรกระทรวงศึกษาธิการ สามารถทำเนื้อหาในบทเรียนวิชานี้ได้บ้างโดยเป็นโจทย์ที่ไม่พลิกแพลงมากนัก คิดค่อนข้างช้า แต่นักเรียนระดับนี้มีแรงจูงใจ ความสนใจ และความรับผิดชอบที่จะทำงานในวิชานี้บ้าง

5. **<u>นักเรียนที่มีความสามารถระดับดีพอใช้</u>** (C+) หมายถึง นักเรียนที่มีความสามารถ**ถึงระดับที่ความคาดหวังไว้**สำหรับการปฏิบัติของนักเรียนในวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ตามหลักสูตรกระทรวงศึกษาธิการ ระดับนี้เป็นเป้าหมายระยะยาวสำหรับการปฏิบัติของนักเรียน นั้นคือ นักเรียนควรมีความรู้ที่แน่น สามารถประยุกต์ความรู้ในสถานการณ์อื่นพอได้ และมีทักษะการวิเคราะห์ในเนื้อหานั้น มีความคิดสร้างสรรค์บ้าง สามารถทำเนื้อหาในบทเรียนวิชานี้พอได้ คิดไม่เร็ว ไม่ช้านัก นักเรียนระดับนี้มีแรงจูงใจ และความรับผิดชอบที่จะทำงานในวิชานี้

6. **<u>นักเรียนที่มีความสามารถระดับดี</u>** (B) หมายถึง นักเรียนที่มีความสามารถถึงระดับที่**เหนือความคาดหวัง**สำหรับการปฏิบัติของนักเรียนในวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ตามหลักสูตรกระทรวงศึกษาธิการ นั้นคือ นักเรียนมีความรู้ที่แน่น สามารถประยุกต์ความรู้ในสถานการณ์อื่นได้ดี และมีทักษะการวิเคราะห์ สังเคราะห์ในเนื้อหานั้น มีความคิดสร้างสรรค์ สามารถทำเนื้อหาในบทเรียนวิชานี้ได้ คิดได้เร็วแต่ยังผิดพลาด ขาดความรอบคอบบ้าง นักเรียนระดับนี้มีแรงจูงใจ ความสนใจ และความรับผิดชอบที่จะทำงานในวิชานี้

7. **<u>นักเรียนที่มีความสามารถระดับดีมาก</u>** (B+) หมายถึง นักเรียนที่มีความสามารถอยู่**เหนือกว่าระดับที่คาดหวังไว้มาก**สำหรับการปฏิบัติของนักเรียนในวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ตามหลักสูตรกระทรวงศึกษาธิการ นั้นคือ นักเรียนมีความรู้ที่แน่น สามารถประยุกต์ความรู้ในสถานการณ์อื่นได้ดีมาก และคล่องแคล่ว มีทักษะการวิเคราะห์ สังเคราะห์ในเนื้อหานั้น มีความคิดสร้างสรรค์ นักเรียนระดับนี้มีความสนใจศึกษาหาความรู้เพิ่มเติมจากข้างนอก มีความรู้ในเนื้อหาวิชานี้แน่น คิดได้เร็ว แต่ขาดความรอบคอบบ้าง นอกจากนี้นักเรียนมีแรงจูงใจ และความรับผิดชอบต่อวิชานี้สูง

8. **<u>นักเรียนที่มีความสามารถระดับดีเยี่ยม</u>** (A) หมายถึง นักเรียนที่มีความสามารถอยู่**เหนือกว่าระดับที่คาดหวังไว้สูงมาก**สำหรับการปฏิบัติของนักเรียนในวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ตามหลักสูตรกระทรวงศึกษาธิการ นั้นคือ นักเรียนมีความรู้ในเนื้อหาที่ลุ่มลึก สามารถประยุกต์ความรู้ในสถานการณ์อื่นได้ดีเยี่ยม คล่องแคล่ว และมีทักษะการวิเคราะห์สังเคราะห์ในเนื้อหานั้นมีความคิดสร้างสรรค์ มีการเชื่อมโยงความรู้ทางคณิตศาสตร์เข้ากับศาสตร์สาขาอื่นเพื่อเกิดวิธีการใหม่ๆขึ้นมา คิดคำนวณอย่างรอบคอบ และคิดได้เร็ว มีความรับผิดชอบสูงมาก สนใจศึกษาหาความรู้เพิ่มเติมจากข้างนอกตลอดเวลา

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ชื่ออาจารย์.....โรงเรียน.....

แบบสร้างนิยามความสามารถของนักเรียน

ผู้ตัดสินเขียนบรรยายระดับการปฏิบัติของนักเรียนในประเด็นความรู้ ทักษะ และความสามารถของนักเรียน ณ ระดับคะแนนจุดตัดของระดับนั้น รวมทั้งสิ้น 8 ระดับ

1. **ระดับดีเยี่ยม** (A)

.....

.....

.....

.....

.....

.....

.....

.....

2. **ระดับดีมาก** (B+)

.....

.....

.....

.....

.....

.....

.....

.....

3. **ระดับดี** (B)

.....

.....

.....

.....

.....

.....

.....

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

4. **ระดับดีพอใช้** (C+)

……

……

……

……

……

……

……

……

……

5. **ระดับพอใช้** (C)

……

……

……

……

……

……

……

……

……

6. **ระดับอ่อน** (D+)

……

……

……

……

……

……

……

……

……

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

7. ระดับอ่อนมาก (D)

.....

.....

.....

.....

.....

.....

.....

.....

8. ระดับตก (F)

.....

.....

.....

.....

.....

.....

.....

.....

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก ง

คู่มือการใช้การกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์ค

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

คู่มือการใช้
การกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์ค

โดย
นางสุรีพร อนุศาสนนันท์
อาจารย์ที่ปรึกษาวิทยานิพนธ์ รศ.ดร.ศิริเดช สุชีวะ
อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม ดร.ชูศักดิ์ ขัมภลิขิต
ภาควิชาวิจัยและจิตวิทยาการศึกษา คณะครุศาสตร์
จุฬาลงกรณ์มหาวิทยาลัย

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

วัตถุประสงค์ในการพัฒนาวิธีการกำหนดมาตรฐานวิธีบุ๊คมาร์ค

วิธีการกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์คได้พัฒนาขึ้นเพื่อให้ผู้ตัดสินหรือครูที่มีประสบการณ์ในวิชานั้นๆ ใช้เป็นแนวทางในการกำหนดคะแนนมาตรฐานสำหรับแบบสอบในวิชาต่างๆ ผู้ตัดสินสามารถนำคู่มือดังกล่าวมาใช้ในการกำหนดมาตรฐานในสถานการณ์ต่างๆ เช่น การสอบที่มีลักษณะส่งผลกระทบต่ออนาคตนักเรียนสูง (high stake) กล่าวคือ การสอบเข้ามหาวิทยาลัยหรือสถาบันการศึกษาต่างๆ การสอบในแต่ละช่วงชั้น การสอบในชั้นเรียน สำหรับวิธีการบุ๊คมาร์คเป็นวิธีการใหม่ที่ยังไม่เคยมาใช้ในประเทศไทย นอกจากนี้ผู้วิจัยได้เพิ่มเติมขั้นตอนบางประการเข้าไปในวิธีการบุ๊คมาร์คเพื่อให้มีประสิทธิภาพยิ่งขึ้น

คำนิยามปฏิบัติการ

1. **มาตรฐาน** (Standard) หมายถึง ระดับที่บ่งบอกคุณภาพของนักเรียนในวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 1 ในการศึกษาครั้งนี้ได้แบ่งระดับของนักเรียนเป็น 8 ระดับ คือ 1) ระดับดีเยี่ยม (A) 2) ระดับดีมาก (B+) 3) ระดับดี (B) 4) ระดับดีพอใช้ (C+) 5) ระดับพอใช้ (C) 6) ระดับอ่อน (D+) 7) ระดับอ่อนมาก (D) และ 8) ระดับตก (F) และในแต่ละระดับจะมีคำอธิบายระดับการปฏิบัติ (performance level descriptor) ที่เป็นข้อความในแต่ละระดับเพื่อแสดงว่า นักเรียนควรมีความรู้ ทักษะ หรือคุณลักษณะในแต่ละระดับอย่างไร

2. **คะแนนจุดตัด** (cut score) หมายถึง จุดคะแนนที่ใช้เป็นจุดแบ่งนักเรียนในวิชานั้นตามความสามารถของนักเรียน ซึ่งในงานวิจัยนี้แบ่งคะแนนจุดตัดออกเป็น 7 จุด คือ 1) จุดคะแนนที่แบ่งความสามารถนักเรียนที่มีความสามารถอ่อนมาก ออกจากนักเรียนระดับตก 2) จุดคะแนนที่แบ่งนักเรียนที่มีความสามารถระดับอ่อนมาก ออกจากนักเรียนระดับอ่อน 3) จุดคะแนนที่แบ่งนักเรียนที่มีความสามารถระดับอ่อน ออกจากนักเรียนระดับพอใช้ 4) จุดคะแนนที่แบ่งนักเรียนที่มีความสามารถระดับพอใช้ ออกจากนักเรียนระดับดีพอใช้ 5) จุดคะแนนที่แบ่งนักเรียนที่มีความสามารถระดับดีพอใช้ ออกจากนักเรียนระดับดี 6) จุดคะแนนที่แบ่งนักเรียนที่มีความสามารถระดับดีออกจากนักเรียนระดับดีมาก 7) จุดคะแนนที่แบ่งนักเรียนที่มีความสามารถระดับดีมาก ออกจากนักเรียนระดับดีเยี่ยม

3. **วิธีการกำหนดมาตรฐาน** (Standard setting) หมายถึง กระบวนการหามาตรฐานเพื่อแบ่งความสามารถของนักเรียนออกเป็น 8 ระดับ คือ 1) ระดับดีเยี่ยม(A) 2) ระดับดีมาก(B+) 3) ระดับดี(B) 4) ระดับดีพอใช้ (C+) 5)ระดับพอใช้ (C) 6) ระดับอ่อน (D+) 7) ระดับอ่อนมาก (D) และ 8) ระดับตก (F)

4. **ผู้ตัดสิน** หมายถึง ครูผู้สอนวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 1 ในโรงเรียนสังกัดสำนักงานการศึกษาขั้นพื้นฐานทุกคนที่มีคุณสมบัติดังนี้ 1) มีความเชี่ยวชาญในเนื้อหาคณิตศาสตร์มัธยมศึกษาตอนต้น คือ จบการศึกษาระดับปริญญาตรีขึ้นไปในสาขาทางคณิตศาสตร์ และ 2) มีความเชี่ยวชาญในการสอน และการวัดผล คือ มีประสบการณ์การสอนมาไม่น้อยกว่า 5 ปี และ เป็นครูชำนาญการ หรือชำนาญการพิเศษ

5. **วิธีการกำหนดมาตรฐานบุ๊คมาร์ค** (Bookmark method) หมายถึง วิธีการหามาตรฐาน โดยจัดให้มีคู่มือการเรียงข้อสอบจากข้อง่ายสุดไปข้อยากสุด ข้อละหนึ่งหน้า มีการวิเคราะห์ข้อสอบเพื่อหาค่าความยาก และความสามารถของผู้สอบซึ่งจะนำทฤษฎีการตอบสนองข้อสอบ (IRT) มาใช้ในการวิเคราะห์ คู่มือที่ได้จะนำมาให้ผู้ตัดสินพิจารณาประกอบการหาจุดตัด ซึ่งผู้ตัดสินจะพิจารณาทีละข้อ จากข้อง่ายไปยังข้อยาก โดยให้ผู้ตัดสินตอบคำถามว่า *ผู้ที่คาบเส้นระดับพอใช้จะมีโอกาสตอบข้อสอบข้อนั้นถูก 67%"* พิจารณาทีละข้อไปเรื่อยๆ จนถึงข้อที่ผู้สอบกลุ่มคาบเส้นระดับพอใช้ไม่มีโอกาสตอบถูก ก็จะใช้ที่คั่นหนังสือคั่นตรงหน้านั้นไว้" ซึ่งเป็นคะแนนจุดตัดระดับพอใช้ (จุดแรก) จากนั้นผู้ตัดสินจะดำเนินการหาคะแนนจุดตัดระดับดี ระดับดีมาก ระดับดีเยี่ยม ระดับอ่อน ระดับอ่อนมาก และระดับตก ตามลำดับ โดยใช้กระบวนการคล้ายกันกับที่กล่าวมา จากนั้นนำผลการตัดสินของแต่ละคนมาอภิปรายร่วมกัน แล้วเปิดโอกาสให้แต่ละคนหาคะแนนจุดตัดอีกครั้ง นำคะแนนจุดตัดเหล่านั้นมาหาค่าเฉลี่ย สำหรับคู่มือจัดเรียงข้อสอบผู้วิจัยได้พัฒนาเพิ่มเติมจากเดิมโดยเพิ่มร้อยละของผู้ตอบในกลุ่มสูง และกลุ่มต่ำที่เลือกตัวเลือกแต่ละตัวเลือก

การเตรียมการก่อนการสัมมนา

1. จัดเตรียมรายชื่อผู้ตัดสินที่มีคุณสมบัติตามที่ต้องการ อาทิเช่น เป็นผู้มีประสบการณ์การสอนและการวัดประเมินในชั้นเรียนในเนื้อหาวิชาที่ทำการสอนมาไม่น้อยกว่า 5 ปี
2. จัดทำหนังสือเรียนเชิญอาจารย์ที่มีคุณสมบัติดังกล่าวมาเป็นผู้ตัดสิน
3. จัดเตรียมเอกสารการจัดสัมมนาดังต่อไปนี้
    3.1 คู่มือการใช้การกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์ค
    3.2 คู่มือการจัดเรียงข้อสอบ (Order item booklet : OLB)
    3.3 แบบฟอร์มเลขที่หน้าที่เป็นคะแนนจุดตัด
    3.4 แบบฟอร์มเขียนคำบรรยายการปฏิบัติแต่ละระดับ
    3.5 คำอธิบายลักษณะความสามารถในวิชานั้นของนักเรียนในแต่ละระดับ เช่น นักเรียนที่มีความสามารถระดับอ่อนมาก หมายถึง นักเรียนที่มีความสามารถถึงระดับที่คาดหวังขั้นต่ำสำหรับการปฏิบัติของนักเรียนในวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 1 ตามหลักสูตร

กระทรวงศึกษาธิการ นั้นคือ นักเรียนมีความรู้ที่จำเป็นต้องมี และมีทักษะขั้นพื้นฐานในการปฏิบัติงานในระดับคาบเส้น

3.6 แบบฟอร์มแสดงค่าเฉลี่ยผลการตัดสินของผู้ตัดสินแต่ละคนในแต่ละระดับ

การสร้างคู่มือการจัดเรียงข้อสอบ (Order item booklet : OLB)

คู่มือการจัดเรียงข้อสอบดังกล่าวจะจัดเรียงข้อสอบตามค่าความยาก โดยจัดเรียงข้อสอบจากข้อสอบง่ายที่สุดไปยังข้อสอบที่ยากที่สุด โดยมีรูปแบบการจัดเรียงข้อสอบ 1 ข้อ ต่อ 1 หน้า นอกจากนี้ผู้เขียนยังได้เพิ่มสารสนเทศสำหรับผู้ตัดสินลงในคู่มือจัดเรียงข้อสอบแต่ละหน้าเข้าไปอีก คือ จำนวนนักเรียนและเปอร์เซ็นต์ที่เลือกตัวเลือกแต่ละตัวเลือก ดังแผนภาพที่ 1 แสดงตัวอย่างรูปแบบหน้าในคู่มือจัดเรียงข้อสอบ และแผนภาพที่ 2 แสดงตัวอย่างการเรียงข้อสอบจากข้องายสุดไปยังข้อยากสุดในคู่มือจัดเรียงข้อสอบ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ข้อที่ 22 35

ระดับความสามารถที่โอกาสความน่าจะเป็นในการตอบถูก 0.67 **คือ** 1.725

(นักเรียนค่อนข้างเก่ง)

การหาคำตอบของระบบสมการเชิงเส้นสองตัวแปร

คำตอบของระบบสมการ $x + y = 2$ และ $2x - 3y = 1$ คือคู่อันดับในข้อใด

ก. (1, 1)

ข. (2, 0)

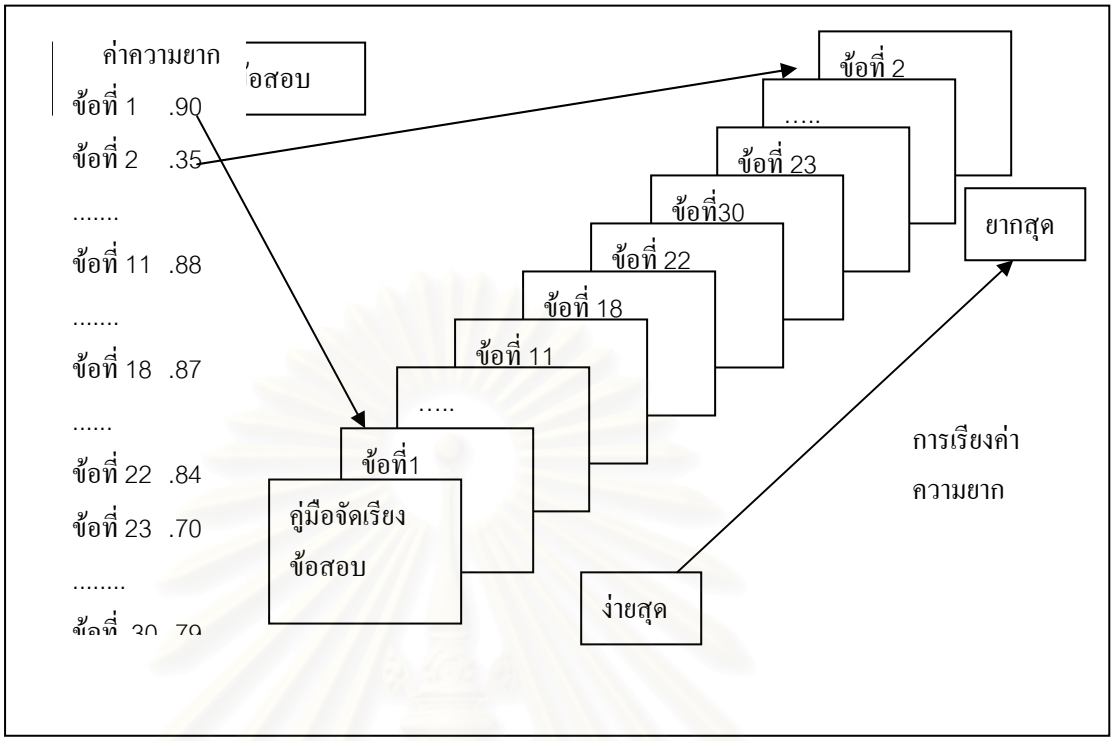
ค. $\left(\frac{5}{3}, \frac{1}{3}\right)$

(ง.) $\left(\frac{7}{5}, \frac{3}{5}\right)$

ตัวเลือก	จำนวนนักเรียนที่เลือก (%)	กลุ่มสูง (%)	กลุ่มต่ำ (%)
ก	4	0	4
ข	8	1	4
ค	12	3	7
(ง)	76	96	85

แผนภาพที่ 1 ตัวอย่างรูปแบบหน้าในคู่มือจัดเรียงข้อสอบ

จากแผนภาพที่ 1 ตัวเข้มบนขวา (35) เป็นตัวเลขที่แสดงเลขที่หน้าของคู่มือจัดเรียงข้อสอบ ตัวเข้มบนซ้าย (ข้อที่ 22) เป็นเลขที่ข้อสอบ ถัดลงมาคือ ระดับความสามารถของผู้สอบ (θ =1.725) จากทฤษฎีการตอบสนองข้อสอบ (item response theory) ที่มีความน่าจะเป็นในการตอบข้อสอบข้อนี้ถูก .67 ข้อมูลถัดลงมาคือ ข้อมูลของข้อสอบที่แสดงว่า ข้อสอบข้อนี้เป็นหนึ่งในเรื่อง การหาคำตอบของระบบสมการเชิงเส้นสองตัวแปร เครื่องหมาย ( ) หน้าตัวเลือก ง. แสดงถึงคำตอบที่ถูก



แผนภาพที่ 2 ตัวอย่างการเรียงข้อสอบจากข้อง่ายสุดไปยังข้อยากสุดในคู่มือจัดเรียงข้อสอบ

สำหรับขั้นตอนการสร้างคู่มือจัดเรียงข้อสอบ มีดังนี้

1. นำผลการตอบจากแบบสอบแบบเลือกหลายคำตอบที่นักเรียนสอบมาวิเคราะห์ข้อสอบเพื่อหาค่าความยาก (b) และอำนาจจำแนก (a) โดยใช้โมเดลการตอบสนองข้อสอบ (IRT) แบบ 2 พารามิเตอร์ สามารถใช้โปรแกรม BILOG หรือ โปรแกรม XCALIBRE วิเคราะห์

2. นำค่าความยาก (b) และอำนาจจำแนก (a) ของข้อสอบแต่ละข้อ มาแทนค่าลงในสูตร (1) เพื่อหาค่าความสามารถ (θ) จากนั้นนำค่าความสามารถดังกล่าวใส่ลงในคู่มือจัดเรียงข้อสอบ ในส่วน “ระดับความสามารถที่โอกาสความน่าจะเป็นในการตอบถูก 0.67 คือ …”

ระดับความสามารถที่โอกาสความน่าจะเป็นในการตอบถูก 0.67 สามารถหาได้จากสูตร

$$\theta = b_j + \frac{.708}{1.7a_j} \quad (1)$$

จากนั้นแปลความหมายค่า θ โดยค่า θ ยิ่งเข้าใกล้ -3 นักเรียนยิ่งอ่อน ส่วนค่า θ ยิ่งเข้าใกล้ +3 นักเรียนยิ่งเก่ง และ ค่า θ เป็น 0 นักเรียนมีความสามารถปานกลาง

3. หาเปอร์เซนต์นักเรียนที่เลือกตัวเลือกแต่ละตัวเลือกในข้อสอบแต่ละข้อ และเปอร์เซนต์นักเรียนที่อยู่ในกลุ่มสูง และกลุ่มต่ำที่เลือกตัวเลือกแต่ละตัวเลือก

สำหรับการหาเปอร์เซนต์นักเรียนที่อยู่ในกลุ่มสูง และกลุ่มต่ำที่เลือกตัวเลือกแต่ละตัวเลือก จะใช้เทคนิค 27% เพื่อแบ่งกลุ่มผู้สอบเป็นกลุ่มสูง และกลุ่มต่ำ โดยนำผู้สอบที่ได้คะแนนสูงสุด 27% แรกของกลุ่มมาจัดเป็นกลุ่มสูง และนำผู้สอบที่ได้คะแนนต่ำสุด 27% สุดท้ายของกลุ่มมาจัดเป็นกลุ่มต่ำ ขั้นตอนการวิเคราะห์ข้อสอบโดยใช้เทคนิค 27% มีดังนี้

(1) ตรวจให้คะแนนข้อสอบเป็นรายข้อ

(2) เรียงลำดับกระดาษคำตอบของผู้ที่ได้คะแนนสูงสุดไปจนถึงผู้ได้คะแนนต่ำสุด

(3) คำนวณว่า 27% ของแต่ละกลุ่ม คิดเป็นจำนวนคนกี่คน

(4) แบ่งกระดาษคำตอบออกเป็นกลุ่มสูงนับจากคะแนนสูงสุดลงมา และกลุ่มต่ำนับจากต่ำสุดขึ้นไป

(5) นำกระดาษคำตอบของนักเรียนกลุ่มสูง และกลุ่มต่ำไปบันทึกความถี่ของการเลือกคำตอบ ว่ามีจำนวนคนในกลุ่มสูง และกลุ่มต่ำเลือกตัวเลือกแต่ละตัวเลือกกี่คน

4. เมื่อคำนวณตัวเลขข้างต้นเรียบร้อยแล้ว ก็จัดรูปแบบแต่ละหน้าดังนี้ นำเลขที่ข้อสอบมาใส่ไว้มุมบนซ้ายมือ และเลขที่หน้าคู่มือจัดเรียงข้อสอบใส่ไว้มุมบนขวามือ เลขที่หน้าดังกล่าวจะจัดเรียงจากข้อที่ง่ายสุดไปยังข้อที่ยากสุดเช่น หน้าที่ 1 คือ ข้อที่ 8 เนื่องจากเป็นข้อสอบที่ง่ายที่สุด ถัดมาคือ หัวข้อเรื่องโจทย์ และตามด้วยโจทย์และตัวเลือก สุดท้ายเป็นตารางแสดงเปอร์เซนต์นักเรียนที่เลือกตัวเลือกแต่ละตัวเลือกในข้อสอบแต่ละข้อ และเปอร์เซนต์นักเรียนที่อยู่ในกลุ่มสูง และกลุ่มต่ำที่เลือกตัวเลือกแต่ละตัวเลือก

การหาค่าความยากจากทฤษฎีการตอบสนองข้อสอบโดยโปรแกรม Xcalibre for Windows95/NT(Version 1.10)

ขั้นที่ 1 การเตรียมข้อมูลลงใน Noteped

หมายเลข 1 มีดังนี้

30 คือ จำนวนข้อสอบ (อยู่ระหว่าง column 1-3) เช่น ถ้าแบบสอบมีข้อสอบ 30 ข้อ ให้พิมพ์ 30 ใน column 2-3 ถ้าแบบสอบมีข้อสอบ 100 ข้อให้พิมพ์ 100 ใน column 1-3

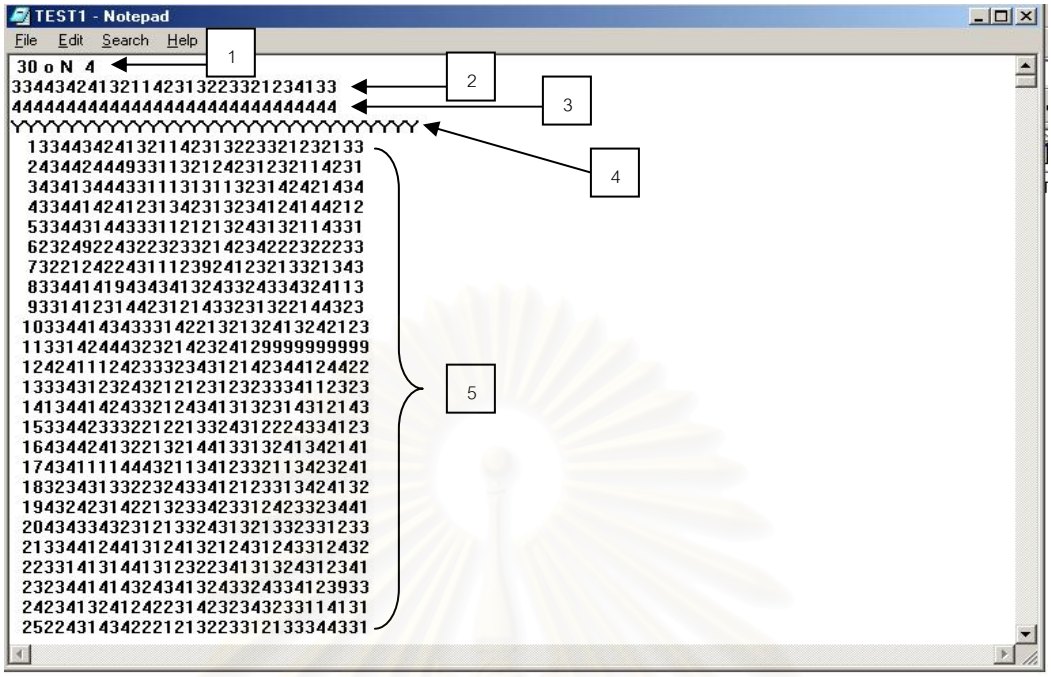
4 คือ จำนวนตัวเลือก

หมายเลข 2 คือ key คำตอบ

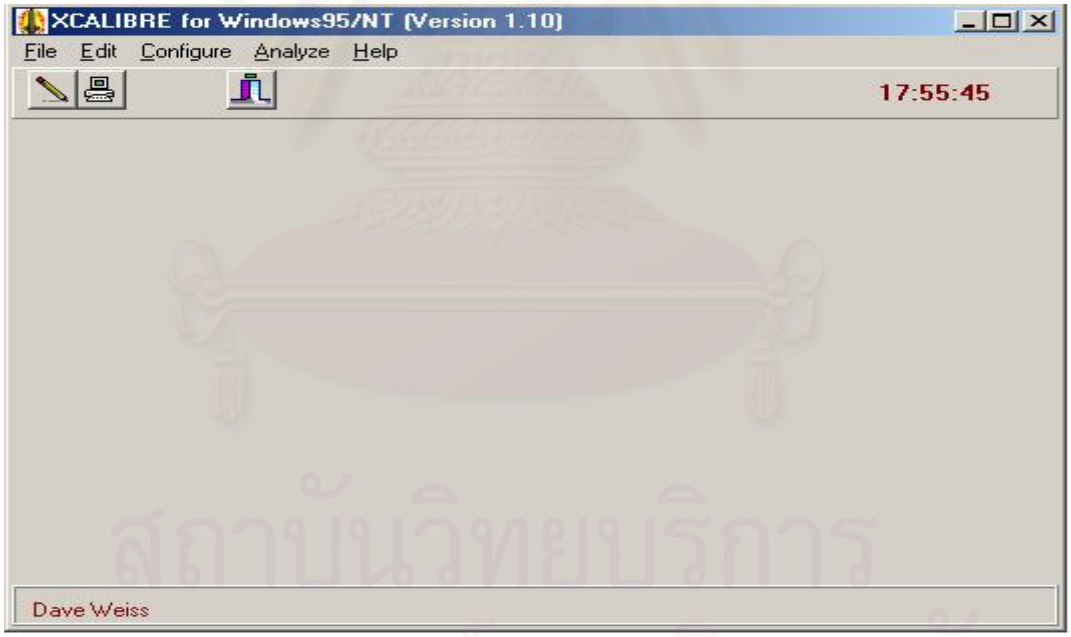
หมายเลข 3 คือ จำนวนตัวเลือก ในที่นี้เป็นข้อสอบแบบเลือกตอบ 4 ตัวเลือก

หมายเลข 4 คือ Y

หมายเลข 5 คือ ข้อมูลการตอบของนักเรียนแต่ละคน

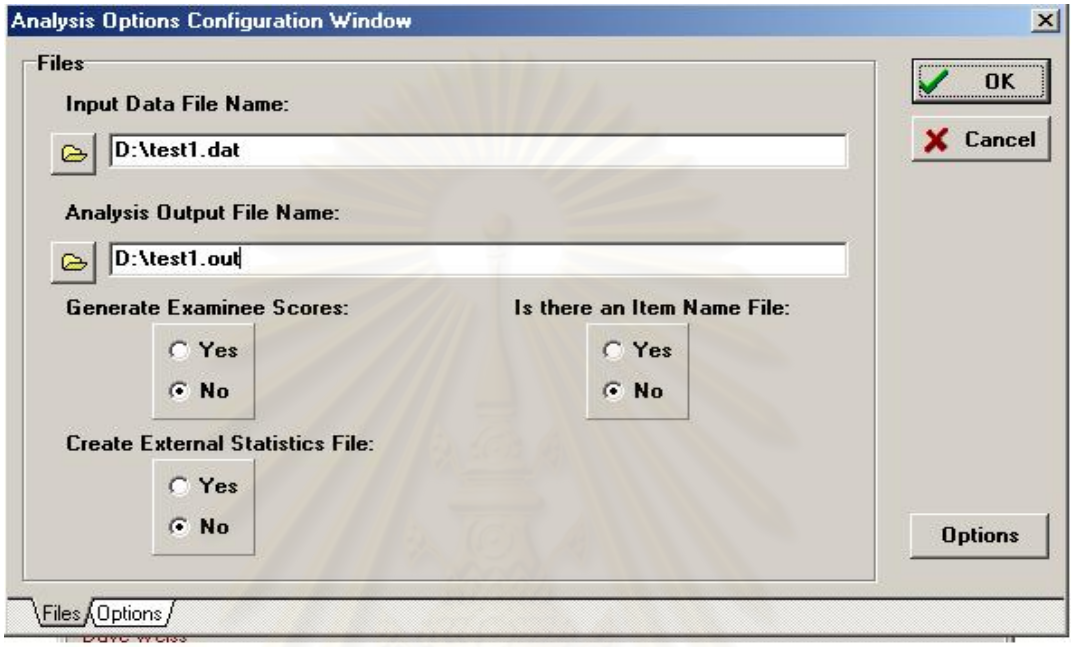


ขั้นที่ 2 เข้าสู่โปรแกรม Xcalibre for Windows95/NT(Version 1.10)



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

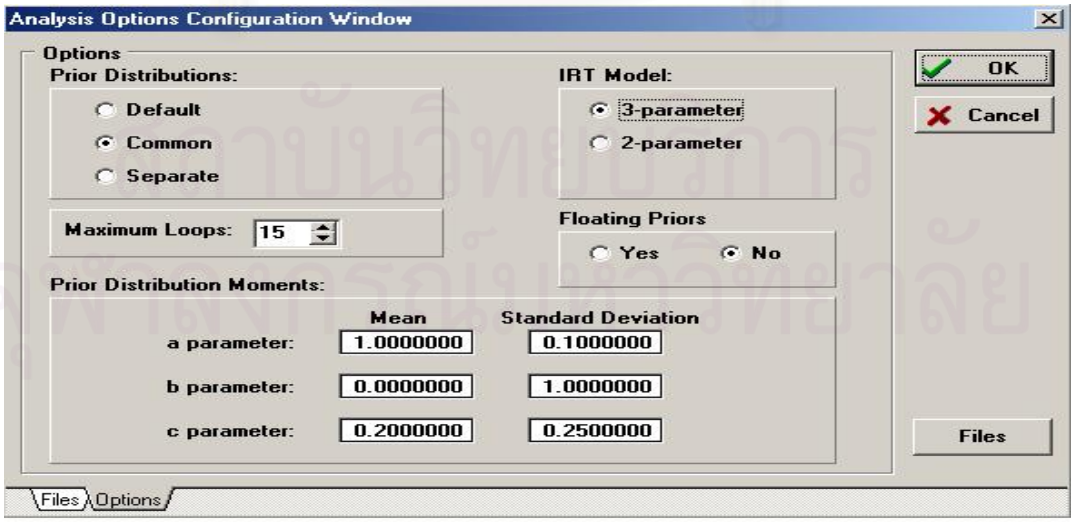
ขั้นที่ 3 พิมพ์ชื่อไฟล์ข้อมูลลงใน Input Data File Name และ
พิมพ์ชื่อไฟล์ผลการวิเคราะห์ลงใน Analysis Output File Name
จากนั้น กดปุ่ม Options



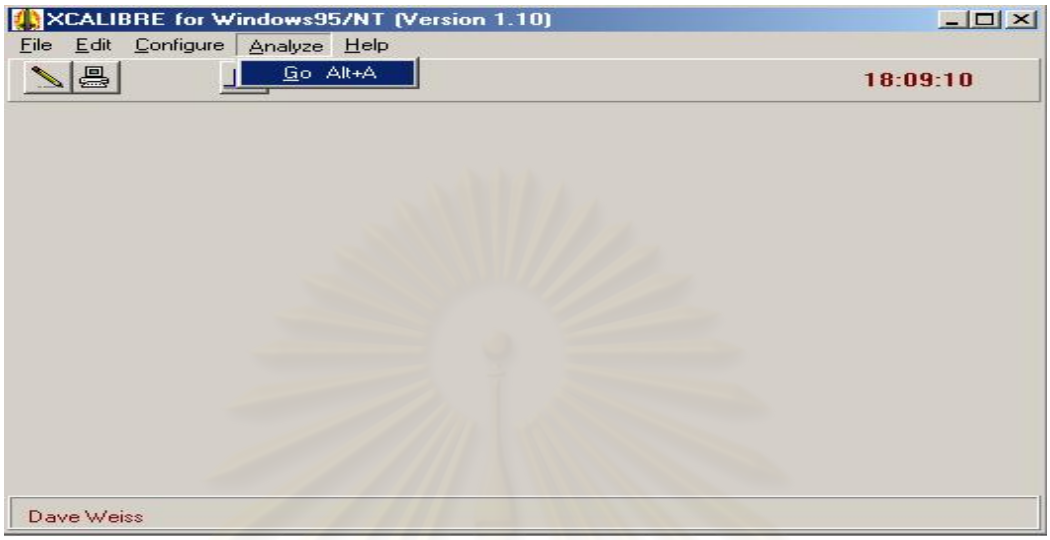
ขั้นที่ 4 เลือกจำนวนพารามิเตอร์ที่ต้องการวิเคราะห์ คือ

2 พารามิเตอร์ ประกอบด้วยค่าความยาก (b) และค่าอำนาจจำแนก (a)

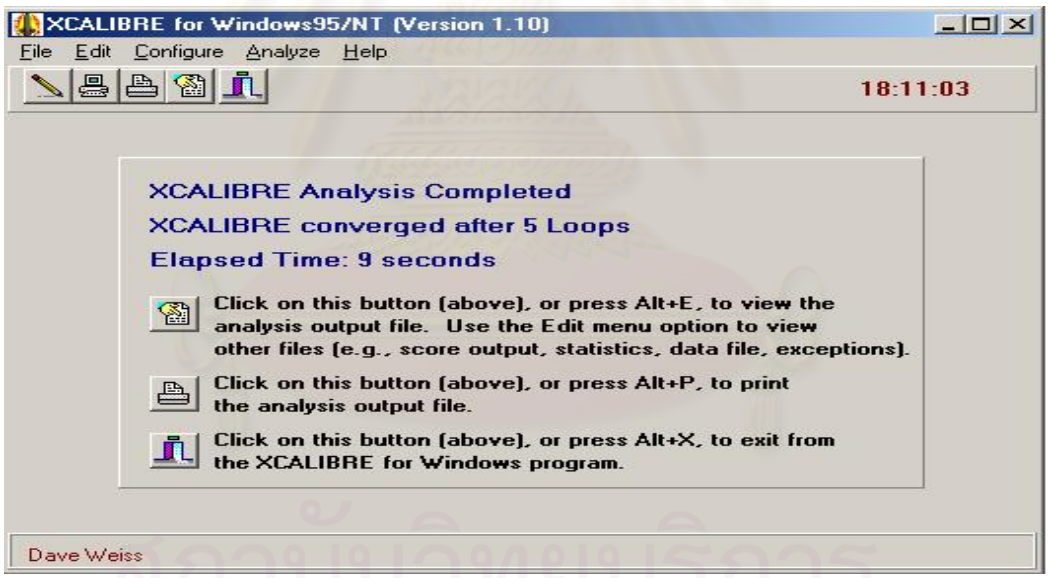
3 พารามิเตอร์ ประกอบด้วยค่าความยาก (b) และค่าอำนาจจำแนก (a) และค่าการเดา (c) จากนั้นกด OK



ขั้นที่ 5 RUN program โดย กดปุ่ม Analyze แล้วไปที่ Go

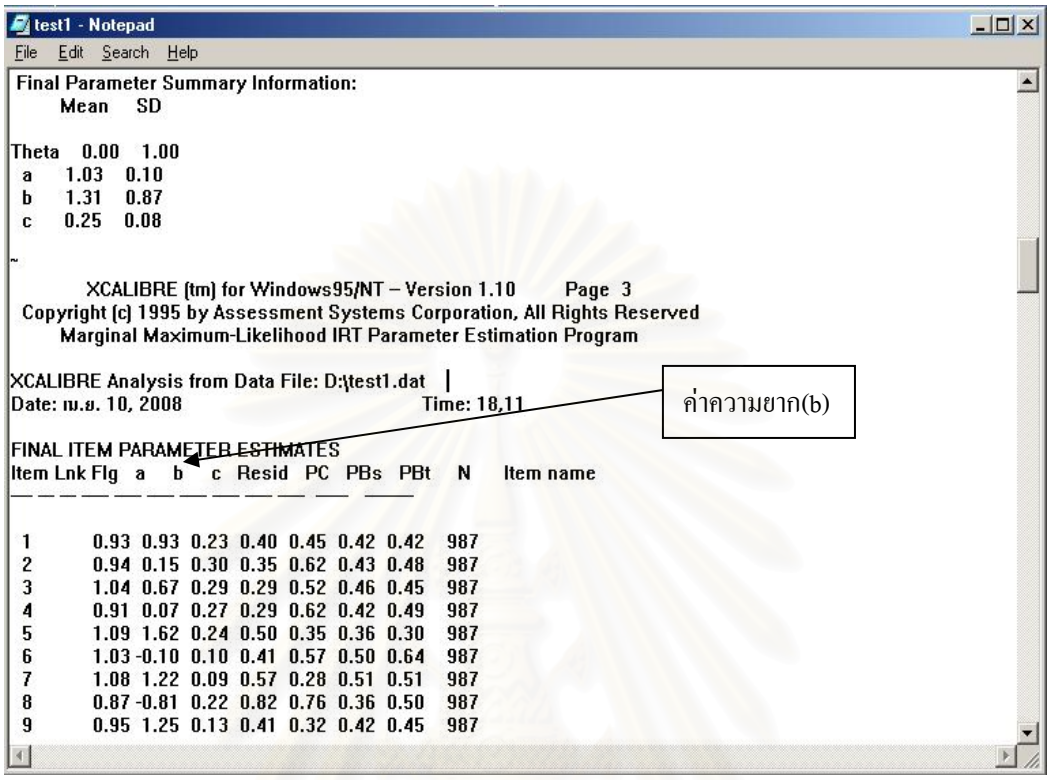


ขั้นที่ 6 เมื่อ Run Program สมบูรณ์จะปรากฏหน้าต่างดังข้างล่าง



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ขั้นที่ 7 ผลการวิเคราะห์ข้อมูลค่าความยาก (b) ปรากฏหน้าต่างข้างล่าง



ตัวอย่างการคำนวณการหาค่าความสามารถของข้อสอบแต่ละข้อโดยใช้โปรแกรม Excel

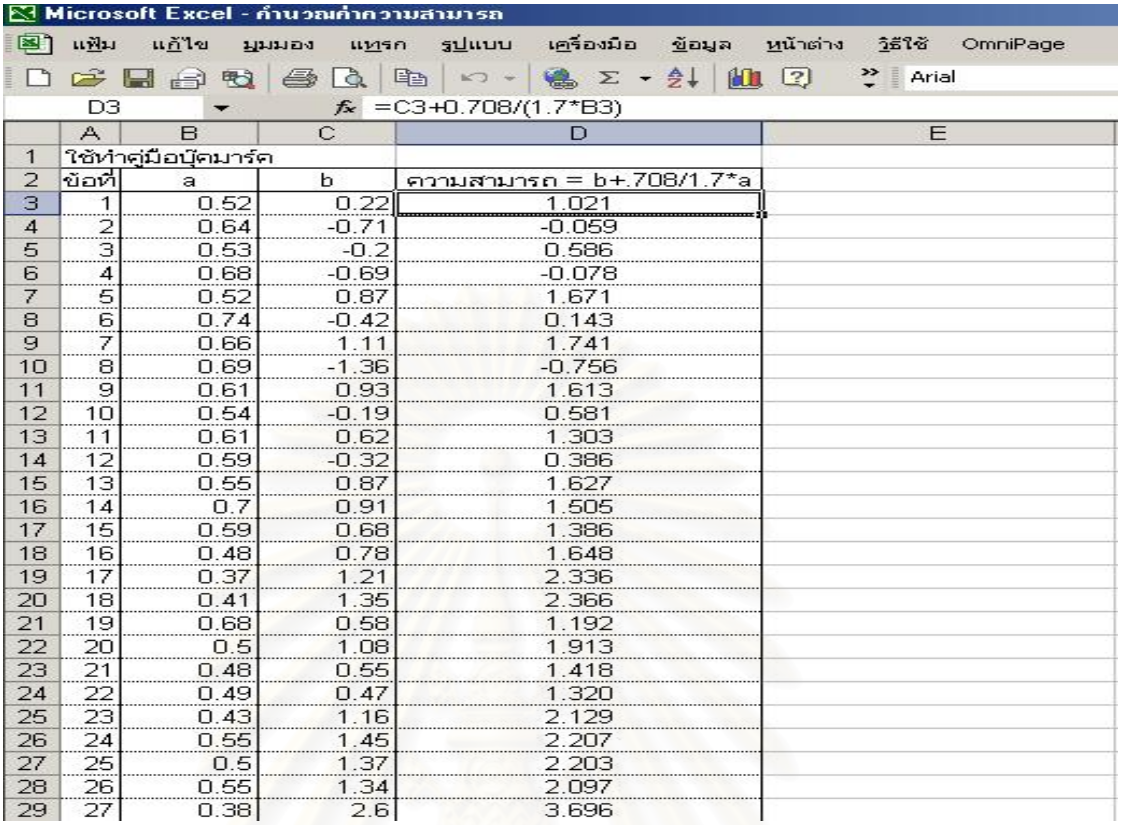
การหาค่าความสามารถของข้อสอบแต่ละข้อจากสูตร

$$\theta \quad = \quad b_j + \frac{.708}{1.7a_j}$$

เมื่อ b_i คือ ค่าความยากของข้อสอบข้อที่ i

a_i คือ ค่าอำนาจจำแนกของข้อสอบข้อที่ i

จากตัวอย่างเป็นการคำนวณหาค่าความสามารถของข้อสอบข้อที่ 1 ใช้คำสั่งในการหาค่าความสามารถคือ =C3+0.708/(1.7*B3) ได้ค่าความสามารถคือ 1.021


Microsoft Excel - คำนวณค่าความสามารถ

แฟ้ม แก้ไข มุมมอง แทรก รูปแบบ เครื่องมือ ข้อมูล หน้าต่าง วิธีใช้ OmniPage

Arial

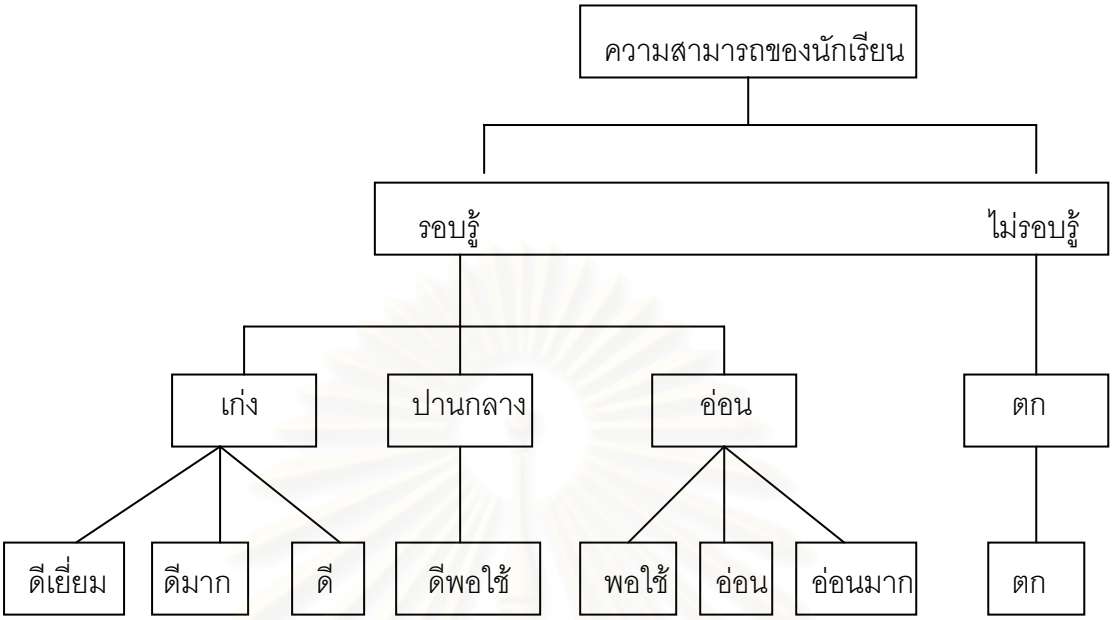
D3 =C3+0.708/(1.7*B3)

	A	B	C	D	E
1	ใช้ทำคู่มือบุ๊คมาร์ค				
2	ข้อที่	a	b	ความสามารถ = b+.708/1.7*a	
3	1	0.52	0.22	1.021	
4	2	0.64	-0.71	-0.059	
5	3	0.53	-0.2	0.586	
6	4	0.68	-0.69	-0.078	
7	5	0.52	0.87	1.671	
8	6	0.74	-0.42	0.143	
9	7	0.66	1.11	1.741	
10	8	0.69	-1.36	-0.756	
11	9	0.61	0.93	1.613	
12	10	0.54	-0.19	0.581	
13	11	0.61	0.62	1.303	
14	12	0.59	-0.32	0.386	
15	13	0.55	0.87	1.627	
16	14	0.7	0.91	1.505	
17	15	0.59	0.68	1.386	
18	16	0.48	0.78	1.648	
19	17	0.37	1.21	2.336	
20	18	0.41	1.35	2.366	
21	19	0.68	0.58	1.192	
22	20	0.5	1.08	1.913	
23	21	0.48	0.55	1.418	
24	22	0.49	0.47	1.320	
25	23	0.43	1.16	2.129	
26	24	0.55	1.45	2.207	
27	25	0.5	1.37	2.203	
28	26	0.55	1.34	2.097	
29	27	0.38	2.6	3.696	

ขั้นตอนการกำหนดมาตรฐานด้วยวิธีบุ๊คมาร์ค

ผู้เขียนขอสรุปขั้นตอนการกำหนดมาตรฐานบุ๊คมาร์คจากการสังเคราะห์เอกสารงานวิจัยที่เกี่ยวข้อง และผู้วิจัยได้พัฒนาปรับปรุงขึ้นในบางขั้นตอนดังนี้ (Buckendahl, 2002; Wang, 2003; Beretvas, 2004; Lewis, Green, Mitzel, Baum & Patz, 1998; Kiplinger, 1997; Lin, 2005; Dawber and Lewis, 2005; Wisconsin Department of Public Instruction, 2004; Cizek, Bunch & Koons, 2004; Ferrara, Johnson & Chen, 2005)

1. ก่อนการกำหนดมาตรฐาน ผู้วิจัยแนะนำลำดับการพิจารณาคะแนนจุดตัดแก่ผู้ตัดสิน มีรายละเอียดคือ ขั้นที่ 1 พิจารณานักเรียน 2 กลุ่มคือ รอบรู้ และไม่รอบรู้ ขั้นที่ 2 กลุ่มรอบรู้แยกเป็น กลุ่มเก่ง กลุ่มปานกลาง และกลุ่มอ่อน ส่วนกลุ่มไม่รอบรู้ แยกเป็นกลุ่มตก และขั้นที่ 3 แยกกลุ่มเก่ง เป็น กลุ่มดีเยี่ยม กลุ่มดีมาก และกลุ่มดี กลุ่มปานกลาง คือ กลุ่มดีพอใช้ กลุ่มอ่อน แยกเป็น กลุ่มพอใช้ กลุ่มอ่อน และ กลุ่มอ่อนมาก ดังแผนภาพที่ 3 ลำดับขั้นการพิจารณาความสามารถของนักเรียน 8 ระดับ



แผนภาพที่ 3 ลำดับขั้นการพิจารณาความสามารถของนักเรียน 8 ระดับ

2. ผู้อำนวยความสะดวกคัดเลือกสิ่งบ่งชี้ที่สำคัญโดยการแจกแจงความถี่เพื่อให้เห็นลำดับความสำคัญ สิ่งบ่งชี้เหล่านี้จะนำมาสร้างความสามารถของนักเรียนทั้ง 8 ระดับ คือ 1) ระดับดีเยี่ยม 2) ระดับดีมาก 3) ระดับดี 4) ระดับดีพอใช้ 5)ระดับพอใช้ 6) ระดับอ่อน 7) ระดับอ่อนมาก และ 8) ระดับตก ว่านักเรียนแต่ละระดับมีความรอบรู้อย่างไร มีลักษณะการตอบในวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ภาคเรียนที่ 1 เป็นการสร้างให้เกิดความเข้าใจตรงกันในการพิจารณาตัดสินคะแนนจุดตัดในแต่ละระดับ

3. ผู้ตัดสินทั้งหมดประมาณ 12 คน มาร่วมสัมมนาเชิงปฏิบัติการเพื่อรับฟังวิธีการกำหนดมาตรฐานด้วยวิธีการบุ๊คมาร์ค พร้อมทั้งฝึกปฏิบัติการเพื่อให้เกิดความเข้าใจในวิธีการกำหนดมาตรฐาน

4. ผู้อำนวยความสะดวกนำสิ่งบ่งชี้ความสามารถของนักเรียนแต่ละระดับมาให้ผู้ตัดสินพิจารณา จากนั้นแบ่งกลุ่มผู้ตัดสินจำนวน 12 คน ออกเป็นกลุ่มย่อย 2 กลุ่ม กลุ่มละ 6 คน ผู้ตัดสินดำเนินงานกำหนดมาตรฐาน

5. ดำเนินการกำหนดคะแนนจุดตัด โดยการคั่นหนังสือลงในคู่มือจัดเรียงข้อสอบ ดำเนินการ 3 รอบดังนี้

รอบที่ 1 เป้าหมายหลักของรอบที่ 1 คือเพื่อให้ผู้ตัดสินมีความคุ้นเคยกับคู่มือการจัดเรียงข้อสอบ ในรอบนี้จะทำงานเป็นกลุ่มย่อย ผู้ตัดสินจะทำงานอิสระจากกัน มีขั้นตอนดังนี้

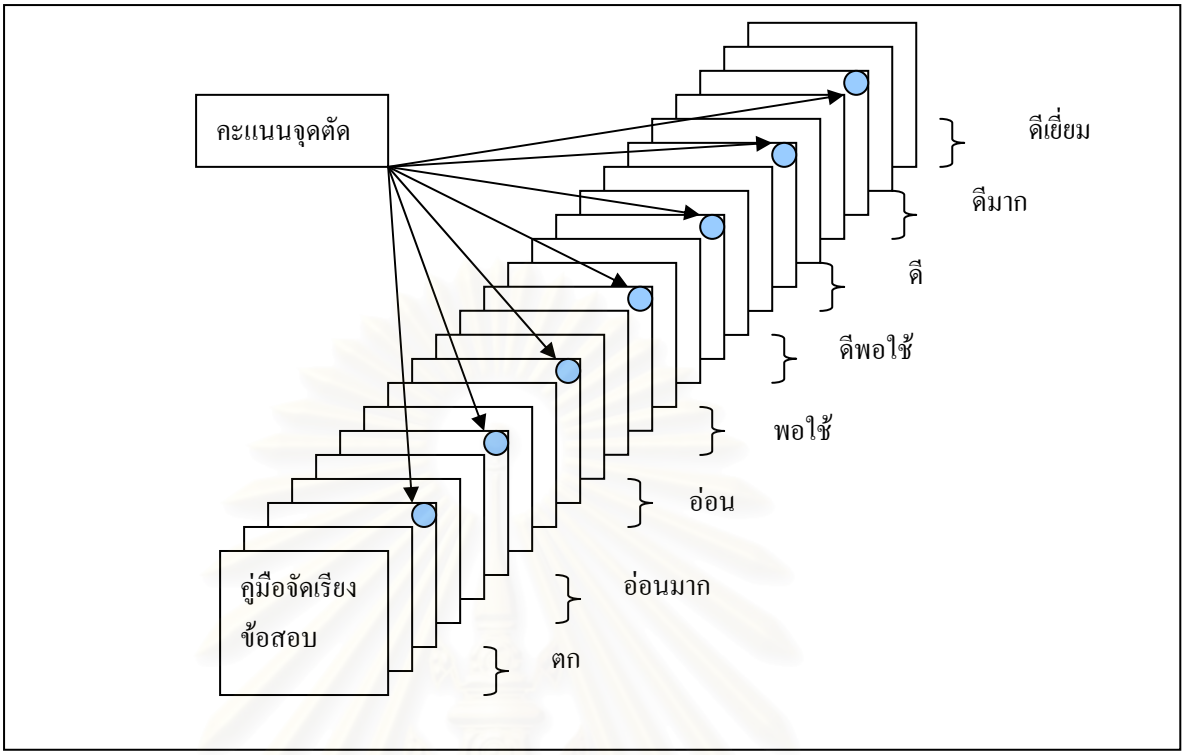
1) ผู้ตัดสินแต่ละคนจะทำการกำหนดคะแนนจุดตัดที่ระดับอ่อนมาก (D) เป็นจุดแรก เพื่อหาคะแนนจุดตัดระหว่างนักเรียนที่ผ่าน และนักเรียนที่ตก โดยให้ผู้ตัดสินแต่ละคนพิจารณาข้อสอบทีละข้อในคู่มือจัดเรียงข้อสอบ โดยพิจารณาตั้งแต่หน้าแรกซึ่งเป็นข้อสอบที่ง่ายที่สุด ไล่ทีละหน้าไปเรื่อยๆ โดยตอบคำถาม 2 คำถามในแต่ละข้อว่า (1) ในระดับชั้นมัธยมศึกษาปีที่ 3 นักเรียนต้องมีความรู้ความสามารถอะไรเพื่อให้ ตอบข้อสอบข้อนี้ถูก ? และ (2) อะไรที่ทำให้ข้อสอบข้อนี้ยากกว่าข้อสอบข้อที่แล้วในคู่มือจัดเรียงข้อสอบ ? การตอบคำถามเหล่านี้เป็นการเตรียมตัวผู้ตัดสินในการทำการตัดสินหาคะแนนจุดตัด

2) ผู้ตัดสินทำการกำหนดบุ๊คมาร์ค(คั่นหนังสือ) บนหน้าที่ผู้ตัดสินคิดว่า "*ผู้ที่คาบเส้นระดับอ่อนมาก (D) จะมีโอกาสตอบข้อสอบข้อนั้นถูก 67%*" โดยผู้ตัดสินจะถูกฝึกให้เข้าใจว่านักเรียนที่คาบเส้นในระดับอ่อนมากควรมีโอกาสตอบถูกอย่างน้อย 67% ในข้อสอบหน้าถัดไป และมีโอกาสตอบถูกสูงกว่า 67% ในหน้าก่อนหน้าที่คั่นหนังสือ

3) หลังจากนั้นก็ดำเนินการกำหนดคะแนนจุดตัดระดับถัดไป คือ หาคะแนนจุดตัดระดับระดับอ่อน (D+) โดยใช้กระบวนการคล้ายกันกับที่กล่าวมา สำหรับระดับนี้ผู้ตัดสินทำการกำหนดบุ๊คมาร์ค(คั่นหนังสือ) บนหน้าที่คิดว่า "*ผู้ที่คาบเส้นระดับอ่อนจะมีโอกาสตอบข้อสอบข้อนั้นถูก 67%*" จากนั้นผู้ตัดสินดำเนินการหาคะแนนจุดตัด ระดับพอใช้ ระดับดีพอใช้ ระดับดี ระดับดีมาก ระดับดีเยี่ยม ตามลำดับเช่นเดียวกับข้างต้น ผลที่ได้จากการหาคะแนนจุดตัดจะได้ระดับคะแนนจุดตัด 8 ระดับสำหรับวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 คือ 1) ระดับดีเยี่ยม 2) ระดับดีมาก 3) ระดับดี 4) ระดับดีพอใช้ 5) ระดับพอใช้ 6) ระดับอ่อน 7) ระดับอ่อนมาก และ 8) ระดับตก

เพื่อให้เห็นภาพที่ชัดเจนขึ้น ผู้วิจัยขอเสนอภาพการกำหนดบุ๊คมาร์ค (การคั่นหนังสือ) ดังแผนภาพที่ 4

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย



แผนภาพที่ 4 ตัวอย่างการกำหนดบุ๊คมาร์ค (การคั่นหนังสือ) 8 ระดับ

รอบที่ 2 ดำเนินการแบ่งผู้ตัดสินเป็นกลุ่มย่อย 2 กลุ่ม มีขั้นตอนดังนี้

1) ผู้ตัดสินแต่ละคนจะได้รับแบบฟอร์มเลขที่หน้าในคู่มือจัดเรียงข้อสอบที่ผู้ตัดสินคนอื่น กำหนดบุ๊คมาร์ค(คั่นหนังสือ) เพื่อใช้เป็นข้อมูลประกอบการพิจารณาคั่นหนังสืออีกครั้ง

2) แต่ละกลุ่มย่อยจัดให้มีการอภิปรายภายในกลุ่ม โดยผู้ตัดสินจะร่วมกันอภิปรายถึงเหตุผล และความเหมาะสมของจุดตัดที่ผู้ตัดสินแต่ละคนตัดสินไว้ โดยเฉพาะให้ผู้ตัดสินคนที่ให้คะแนนจุดตัดสูงสุด และคะแนนจุดตัดต่ำสุดอภิปรายเหตุผลที่ใช้คะแนนจุดตัดตรงนี้ ซึ่งการอภิปรายอาจจะอภิปรายในประเด็นตัวข้อสอบเกี่ยวกับระดับทักษะ และเนื้อหาทางวิชาการที่ควรจะรอบรู้สำหรับนักเรียนในแต่ละระดับคะแนนจุดตัด เหตุผลในการอภิปรายก็เพื่อให้ผู้ตัดสินได้รับฟังเหตุผลของผู้ตัดสินคนอื่น การอภิปรายในกลุ่มย่อยนั้น ผู้ตัดสินจะแสดงความคิดเห็นและประสบการณ์ที่มีความหลากหลาย

3) หลังจากนั้นจะเปิดโอกาสให้ผู้ตัดสินพิจารณาคู่มือจัดเรียงข้อสอบอีกครั้งหนึ่ง เพื่อหาคะแนนจุดตัดในรอบที่ 2 โดยจะพิจารณาทีละระดับจนครบทั้ง 7 จุด เมื่อได้เลขที่หน้าที่ผู้ตัดสินแต่ละคนกำหนดขึ้นมาเพื่อกำหนดคะแนนจุดตัดแล้ว จากนั้นจะนำเลขที่หน้ามาแปลงเป็น

ค่าความสามารถ(θ) (เลขที่หน้าที่คั่นหนังสือไว้นี้ตรงกับค่า θ เท่าไร แล้วนำค่าθ เหล่านี้มาหาค่าเฉลี่ยเพื่อนำมาประมาณความสามารถขั้นต่ำของนักเรียน)

รอบที่ 3 ดำเนินการรวมผู้ตัดสินทั้งสองกลุ่มย่อยเป็นกลุ่มใหญ่ มีขั้นตอนดังนี้

1) นำผู้ตัดสินทั้ง 2 กลุ่มย่อยมารวมกันเป็นกลุ่มใหญ่เพียงกลุ่มเดียว แล้วผู้อำนวยความสะดวกจะนำข้อมูลผลกระทบ (impact data) คือ เปอร์เซ็นต์ของนักเรียนที่อยู่ในแต่ละระดับ ค่าเฉลี่ยของคะแนนจุดตัดจากกลุ่มย่อยทั้งสองกลุ่ม มาแสดงให้ผู้ตัดสินได้พิจารณา

2) จากนั้นให้ผู้ตัดสินอภิปรายร่วมกัน โดยพิจารณาข้อสอบที่อยู่ระหว่างค่าเฉลี่ยของคะแนนจุดตัดทั้งสองกลุ่ม แล้วจึงเปิดโอกาสให้ผู้ตัดสินทำการกำหนดคะแนนจุดตัดอีกครั้งอย่างอิสระจากกัน โดยการกำหนดหน้าที่คั่นหนังสือในคู่มือจัดเรียงข้อสอบ จากนั้นจะนำเลขที่หน้าที่ผู้ตัดสินแต่ละคนจะถูกแปลงเป็นค่าความสามารถ(θ) (เลขที่หน้าที่คั่นหนังสือไว้นี้ตรงกับค่า θ เท่าไร) แล้วนำค่าθ เหล่านี้มาหาค่าเฉลี่ยเพื่อนำมาประมาณความสามารถขั้นต่ำของนักเรียน

3) เมื่อได้คะแนนจุดตัดทั้ง 7 จุดแล้ว ให้ผู้ตัดสินระดมพลังสมองเพื่อเขียนบรรยายระดับการปฏิบัติของนักเรียนในประเด็นความรู้ ทักษะ และความสามารถของนักเรียน ณ ระดับคะแนนจุดตัดของระดับนั้น รวมทั้งสิ้น 8 ระดับ

รายละเอียดต่างๆ ของตัวอย่างแบบฟอร์ม ตลอดจนตัวอย่างสารสนเทศที่ให้ผู้ตัดสินพิจารณา แสดงไว้ในภาคผนวก

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

แบบฟอร์มเลขที่หน้าที่เป็นคะแนนจุดตัดทั้ง 7 ระดับ

ชื่ออาจารย์……..……**โรงเรียน**……

จุดตัดคะแนน	เลขที่หน้าที่คั่น		
	รอบที่ 1 (รายบุคคล)	**รอบที่ 2 (กลุ่มย่อย)**	**รอบที่ 3 (กลุ่มใหญ่)**
ระดับดีเยี่ยม (A)			
ระดับดีมาก (B+)			
ระดับดี (B)			
ระดับดีพอใช้ (C+)			
ระดับพอใช้ (C)			
ระดับอ่อน (D+)			
ระดับอ่อนมาก (D)			

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตัวอย่างคำอธิบายลักษณะความสามารถแต่ละระดับของนักเรียน วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ภาคเรียนที่ 1

เนื้อหาในบทเรียนวิชาคณิตศาสตร์ มัธยมศึกษาปีที่ 3 ภาคเรียนที่ 1 ที่นักเรียนควรทำได้

- อธิบายลักษณะและสมบัติของปริซึม พีระมิด ทรงกระบอก กรวย และทรงกลม
- หาพื้นที่ผิวของปริซึมและทรงกระบอก
- หาปริมาตรของปริซึม ทรงกระบอก พีระมิด กรวย และทรงกลม
- เปรียบเทียบหน่วยความจุ หรือหน่วยปริมาตร
- อ่านและแปลความหมายของกราฟของระบบสมการเชิงเส้น
- แก้ระบบสมการเชิงเส้นสองตัวแปร
- นำระบบสมการเชิงเส้นสองตัวแปรไปใช้แก้ปัญหา
- เขียนกราฟของสมการเชิงเส้นสองตัวแปร
- อ่านและแปลกราฟความหมายของกราฟ
- อ่านและแปลความหมายของกราฟของระบบสมการเชิงเส้น
- แก้ระบบสมการเชิงเส้น
- บอกสมบัติของการคล้ายของรูปสามเหลี่ยม และบอกเงื่อนไขที่ทำให้รูปสามเหลี่ยมสองรูปคล้ายกัน
- ใช้สมบัติของรูปสามเหลี่ยมที่คล้ายกันในการให้เหตุผล และแก้ปัญหา

1. **<u>นักเรียนที่มีความสามารถระดับตก</u>** (F) หมายถึง นักเรียนที่มีความสามารถ**ไม่ถึงระดับที่คาดหวังขั้นต่ำ**ว่านักเรียนสามารถปฏิบัติได้ในวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ตามหลักสูตรกระทรวงศึกษาธิการ โดยทำเนื้อหาในบทเรียนวิชานี้**ไม่ได้** คำนวณหรือคิดไม่ได้ นักเรียนระดับนี้ไม่มีแรงจูงใจ ความสนใจ และความรับผิดชอบที่จะทำงานในวิชานี้

2. **<u>นักเรียนที่มีความสามารถระดับอ่อนมาก</u>** (D) หมายถึง นักเรียนที่มีความสามารถ**ถึงระดับที่คาดหวังขั้นต่ำ**สำหรับการปฏิบัติของนักเรียนในวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ตามหลักสูตรกระทรวงศึกษาธิการ นั้นคือ นักเรียนมีความรู้ที่จำเป็นต้องมี และมีทักษะขั้นพื้นฐานในการปฏิบัติงานในระดับคาบเส้น โดยทำเนื้อหาในบทเรียนวิชานี้**ได้น้อย** คำนวณหรือ

คิดได้บ้างจากโจทย์ที่ไม่ได้พลิกแพลง คิดช้ามาก นักเรียนระดับนี้ไม่ค่อยมีแรงจูงใจ ความสนใจ และความรับผิดชอบที่จะทำงานในวิชานี้

3. **นักเรียนที่มีความสามารถระดับอ่อน** (D+) หมายถึง นักเรียนที่มีความสามารถ**ถึงระดับที่คาดหวังขั้นต่ำขึ้นมาเล็กน้อย** สำหรับการปฏิบัติของนักเรียนในวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ตามหลักสูตรกระทรวงศึกษาธิการ นั้นคือ นักเรียนมีความรู้ที่จำเป็นต้องมี และมีทักษะขั้นพื้นฐานในการปฏิบัติงานในระดับเกินระดับคาบเส้นขึ้นมาไม่มาก โดยทำเนื้อหาในบทเรียนวิชานี้**ได้น้อย** คำนวณหรือคิดได้บ้างจากโจทย์ที่ไม่ได้พลิกแพลง แต่คิดช้า นักเรียนระดับนี้ไม่ค่อยมีแรงจูงใจ ความสนใจ และความรับผิดชอบที่จะทำงานในวิชานี้

4. **นักเรียนที่มีความสามารถระดับพอใช้** (C) หมายถึง นักเรียนที่มีความสามารถ**ถึงระดับที่คาดหวัง**สำหรับการปฏิบัติของนักเรียนในวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ตามหลักสูตรกระทรวงศึกษาธิการ สามารถทำเนื้อหาในบทเรียนวิชานี้ได้บ้างโดยเป็นโจทย์ที่ไม่พลิกแพลงมากนัก คิดค่อนข้างช้า แต่นักเรียนระดับนี้มีแรงจูงใจ ความสนใจ และความรับผิดชอบที่จะทำงานในวิชานี้บ้าง

5. **นักเรียนที่มีความสามารถระดับดีพอใช้** (C+) หมายถึง นักเรียนที่มีความสามารถ**ถึงระดับที่ความคาดหวังไว้**สำหรับการปฏิบัติของนักเรียนในวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ตามหลักสูตรกระทรวงศึกษาธิการ ระดับนี้เป็นเป้าหมายระยะยาวสำหรับการปฏิบัติของนักเรียน นั้นคือ นักเรียนควรมีความรู้ที่แน่น สามารถประยุกต์ความรู้ในสถานการณ์อื่นพอได้ และมีทักษะการวิเคราะห์ในเนื้อหานั้น มีความคิดสร้างสรรค์บ้าง สามารถทำเนื้อหาในบทเรียนวิชานี้พอได้ คิดไม่เร็ว ไม่ช้านัก นักเรียนระดับนี้มีแรงจูงใจ และความรับผิดชอบที่จะทำงานในวิชานี้

6. **นักเรียนที่มีความสามารถระดับดี** (B) หมายถึง นักเรียนที่มีความสามารถถึงระดับที่**เหนือความคาดหวัง**สำหรับการปฏิบัติของนักเรียนในวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ตามหลักสูตรกระทรวงศึกษาธิการ นั้นคือ นักเรียนมีความรู้ที่แน่น สามารถประยุกต์ความรู้ในสถานการณ์อื่นได้ดี และมีทักษะการวิเคราะห์ สังเคราะห์ในเนื้อหานั้น มีความคิดสร้างสรรค์

สามารถทำเนื้อหาในบทเรียนวิชานี้ได้ คิดได้เร็วแต่ยังผิดพลาด ขาดความรอบคอบบ้าง นักเรียนระดับนี้มีแรงจูงใจ ความสนใจ และความรับผิดชอบที่จะทำงานในวิชานี้

7. **<u>นักเรียนที่มีความสามารถระดับดีมาก</u>** (B+) หมายถึง นักเรียนที่มีความสามารถอยู่**เหนือกว่าระดับที่คาดหวังไว้มาก**สำหรับการปฏิบัติของนักเรียนในวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ตามหลักสูตรกระทรวงศึกษาธิการ นั้นคือ นักเรียนมีความรู้ที่แน่น สามารถประยุกต์ความรู้ในสถานการณ์อื่นได้ดีมาก และคล่องแคล่ว มีทักษะการวิเคราะห์ สังเคราะห์ในเนื้อหานั้น มีความคิดสร้างสรรค์ นักเรียนระดับนี้มีความสนใจศึกษาหาความรู้เพิ่มเติมจากข้างนอก มีความรู้ในเนื้อหาวิชานี้แน่น คิดได้เร็ว แต่ขาดความรอบคอบบ้าง นอกจากนี้นักเรียนมีแรงจูงใจ และความรับผิดชอบต่อวิชานี้สูง

8. **<u>นักเรียนที่มีความสามารถระดับดีเยี่ยม</u>** (A) หมายถึง นักเรียนที่มีความสามารถอยู่**เหนือกว่าระดับที่คาดหวังไว้สูงมาก**สำหรับการปฏิบัติของนักเรียนในวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ตามหลักสูตรกระทรวงศึกษาธิการ นั้นคือ นักเรียนมีความรู้ในเนื้อหาที่ลุ่มลึก สามารถประยุกต์ความรู้ในสถานการณ์อื่นได้ดีเยี่ยม คล่องแคล่ว และมีทักษะการวิเคราะห์สังเคราะห์ในเนื้อหานั้นมีความคิดสร้างสรรค์ มีการเชื่อมโยงความรู้ทางคณิตศาสตร์เข้ากับศาสตร์สาขาอื่นเพื่อเกิดวิธีการใหม่ๆขึ้นมา คิดคำนวณอย่างรอบคอบ และคิดได้เร็ว มีความรับผิดชอบสูงมาก สนใจศึกษาหาความรู้เพิ่มเติมจากข้างนอกตลอดเวลา

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ชื่ออาจารย์.....โรงเรียน.....

แบบสร้างนิยามความสามารถของนักเรียน

ผู้ตัดสินเขียนบรรยายระดับการปฏิบัติของนักเรียนในประเด็นความรู้ ทักษะ และความสามารถของนักเรียน ณ ระดับคะแนนจุดตัดของระดับนั้น รวมทั้งสิ้น 8 ระดับ

1. **ระดับดีเยี่ยม** (A)

.....

.....

.....

.....

.....

.....

.....

.....

2. **ระดับดีมาก** (B+)

.....

.....

.....

.....

.....

.....

.....

.....

3. **ระดับดี** (B)

.....

.....

.....

.....

.....

.....

.....

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

4. **ระดับดีพอใช้** (C+)

……

……

……

……

……

……

……

……

……

5. **ระดับพอใช้** (C)

……

……

……

……

……

……

……

……

……

6. **ระดับอ่อน** (D+)

……

……

……

……

……

……

……

……

……

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

7. ระดับอ่อนมาก (D)

.....

.....

.....

.....

.....

.....

.....

.....

8. ระดับตก (F)

.....

.....

.....

.....

.....

.....

.....

.....

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก จ

1. แบบสอบถามความเหมาะสม และความเป็นไปได้ในการนำไปใช้ในกระบวนการการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และ วิธีการบุ๊คมาร์ค
2. แบบสอบถามกระบวนการการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และ วิธีการบุ๊คมาร์ค

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

แบบสอบถามความคิดเห็นเกี่ยวกับวิธีการกำหนดคะแนนจุดตัดด้วย วิธีแองกอฟที่ได้รับการปรับปรุง

<u>คำชี้แจง</u>

1. แบบสอบถามฉบับนี้มีทั้งหมด 3 ตอน ได้แก่

ตอนที่ 1 ข้อมูลทั่วไปของผู้ตอบแบบสอบถาม

ตอนที่ 2 การสำรวจความคิดเห็นเกี่ยวกับการกำหนดมาตรฐานด้วยวิธีแองกอฟที่ได้รับการปรับปรุง

ตอนที่ 3 การสำรวจความคิดเห็นเกี่ยวกับการกำหนดมาตรฐานด้วยวิธีวิธีบุ๊คมาร์ค

การตอบแบบสอบถามฉบับนี้ในตอนที่ 1 ให้ท่านพิจารณาข้อความในแต่ละข้อแล้วตอบคำถามตามความเป็นจริงที่ตรงกับตัวท่าน

2. ในการตอบแบบสอบถามในตอนที่ 2และ3 ให้ท่านพิจารณาข้อความในแต่ละข้อแล้วตอบคำถามแต่ละข้อตามความจริงที่ตรงกับความคิดเห็นของท่านมากที่สุด โดยให้ระดับความคิดเห็นดังนี้

5 หมายถึง มากที่สุด

4 หมายถึง มาก

3 หมายถึง ปานกลาง

2 หมายถึง น้อย

1 หมายถึง น้อยที่สุด

ตอนที่ 1 ข้อมูลทั่วไปเกี่ยวกับผู้ตัดสิน

1. **ตำแหน่งปัจจุบัน**

☐ 1) ครู ☐ 2) ครูชำนาญการ

☐ 3) ครูชำนาญการพิเศษ

2. **ประสบการณ์สอนวิชาคณิตศาสตร์**

☐ 1) 2 – 5 ปี ☐ 2) 6 – 10 ปี

☐ 3) 11 – 15 ปี ☐ 4) มากกว่า 15 ปี

3. **การศึกษา**

☐ 1) ปริญญาตรี ☐ 2) ปริญญาโท

☐ 3) กำลังศึกษาระดับปริญญาเอก

ตอนที่ 2 โปรดกาเครื่องหมาย ✓ ลงในช่องระดับความคิดเห็นในการใช้วิธีการกำหนดคะแนนจุดตัดด้วย**<u>วิธีแองกอฟที่ได้รับการปรับปรุง</u>**

ข้อความ	ระดับความคิดเห็น				
	5	4	3	2	1
1. วิธีการนี้ทำให้ท่านสามารถดำเนินการให้เป็นไปตามวัตถุประสงค์ได้					
2. ผู้บรรยายอธิบายขั้นตอนการกำหนดคะแนนจุดตัดได้อย่างชัดเจน					
3. การฝึกปฏิบัติการกำหนดคะแนนจุดตัดช่วยให้ท่านเข้าใจวิธีการกำหนดคะแนนจุดตัดมากยิ่งขึ้น					
4. วิธีการนี้ง่ายต่อการกำหนดคะแนนจุดตัด					
5. วิธีการนี้เป็นรูปแบบการกำหนดคะแนนที่ยอมรับได้					
6. การบรรยายระดับการปฏิบัติมีความชัดเจน และใช้ประโยชน์ได้					
7. การอภิปรายกลุ่มย่อย และกลุ่มใหญ่ช่วยให้เกิดความเข้าใจในกระบวนการกำหนดคะแนนจุดตัดมากยิ่งขึ้น					
8. เวลาที่ใช้ในการอภิปรายมีความเหมาะสม					
9. มีการเปิดโอกาสให้ผู้ตัดสินภายในกลุ่มได้แสดงความคิดเห็นทุกคน					
10. วิธีการนี้ทำให้ท่านกำหนดคะแนนจุดตัดได้อย่างแคล่วคล่อง					
11. การอภิปรายหลังจากกำหนดคะแนนจุดตัดรอบแรกช่วยให้ท่านกำหนดคะแนนจุดตัดได้แม่นยำขึ้น					
12. การอภิปรายหลังจากกำหนดคะแนนจุดตัดรอบสองช่วยให้ท่านกำหนดคะแนนจุดตัดได้แม่นยำขึ้น					
13. การอภิปรายหลังจากกำหนดคะแนนจุดตัดรอบสามช่วยให้ท่านกำหนดคะแนนจุดตัดได้แม่นยำขึ้น					
14. วิธีการนี้ทำให้ท่านเสียเวลาในการกำหนดคะแนนจุดตัด					
15. ท่านมั่นใจคะแนนจุดตัดที่ได้จากการกำหนดคะแนนจุดตัดด้วยวิธีนี้					
16. วิธีการนี้ทำให้ได้คะแนนจุดตัดที่มีความแม่นยำในการจำแนกความสามารถของนักเรียน					
17. วิธีการนี้เหมาะสมสำหรับกำหนดคะแนนจุดตัดที่ให้น่าเชื่อถือมากน้อยเพียงใด					
18. วิธีการนี้มีความเป็นไปได้ในการนำไปใช้ในสถานการณ์ที่เป็นจริงมากน้อยเพียงใด					
19. คำอธิบายลักษณะความสามารถของนักเรียนแต่ละระดับ ช่วยท่านในการกำหนดคะแนนจุดตัด					

ข้อความ	ระดับความคิดเห็น				
	5	4	3	2	1
20.. ค่าความยากของข้อสอบแต่ละข้อ ช่วยท่านในการกำหนดคะแนนจุดตัด					
21. แบบทดสอบ ช่วยท่านในการกำหนดคะแนนจุดตัด					
22.. กราฟแท่งแจกแจงคะแนนของนักเรียน ช่วยท่านในการกำหนดคะแนนจุดตัด					
23. แบบฟอร์มสำหรับกรอกเปอร์เซ็นต์การตอบถูกของผู้สอบคาบเส้นในแต่ละ					
ระดับ ง่ายต่อการกรอกตัวเลข					
24. คู่มือการใช้อ่านเข้าใจมากน้อยเพียงใด					
25. เมื่ออ่านคู่มือการใช้แล้วท่านสามารถนำไปใช้ปฏิบัติได้					

ตอนที่ 3 โปรดกาเครื่องหมาย ✓ ลงในช่องระดับความคิดเห็นในการใช้วิธีการกำหนดคะแนนจุดตัดด้วย**<u>วิธีบุ๊คมาร์ค</u>**

ข้อความ	ระดับความคิดเห็น				
	5	4	3	2	1
1. วิธีการนี้ทำให้ท่านสามารถดำเนินการให้เป็นไปตามวัตถุประสงค์ได้					
2. ผู้บรรยายอธิบายขั้นตอนการกำหนดคะแนนจุดตัดได้อย่างชัดเจน					
3. การฝึกปฏิบัติการกำหนดคะแนนจุดตัดช่วยให้ท่านเข้าใจวิธีการกำหนดคะแนนจุดตัดมากยิ่งขึ้น					
4. วิธีการนี้ง่ายต่อการกำหนดคะแนนจุดตัด					
5. วิธีการนี้เป็นรูปแบบการกำหนดคะแนนที่ยอมรับได้					
6. การบรรยายระดับการปฏิบัติมีความชัดเจน และใช้ประโยชน์ได้					
7. การอภิปรายกลุ่มย่อย และกลุ่มใหญ่ช่วยให้เกิดความเข้าใจในกระบวนการกำหนดคะแนนจุดตัดมากยิ่งขึ้น					
8. เวลาที่ใช้ในการอภิปรายมีความเหมาะสม					
9. มีการเปิดโอกาสให้ผู้ตัดสินภายในกลุ่มได้แสดงความคิดเห็นทุกคน					
10. วิธีการนี้ทำให้ท่านกำหนดคะแนนจุดตัดได้อย่างแคล่วคล่อง					
11. การอภิปรายหลังจากกำหนดคะแนนจุดตัดรอบแรกช่วยให้ท่านกำหนดคะแนนจุดตัดได้แม่นยำขึ้น					

ข้อความ	ระดับความคิดเห็น				
	5	4	3	2	1
12. การอภิปรายหลังจากกำหนดคะแนนจุดตัดรอบสองช่วยให้ท่านกำหนดคะแนนจุดตัดได้แม่นยำขึ้น					
13. การอภิปรายหลังจากกำหนดคะแนนจุดตัดรอบสามช่วยให้ท่านกำหนดคะแนนจุดตัดได้แม่นยำขึ้น					
14. วิธีนี้ทำให้ท่านเสียเวลาในการกำหนดคะแนนจุดตัด					
15. ท่านมั่นใจคะแนนจุดตัดที่ได้จากการกำหนดคะแนนจุดตัดด้วยวิธีนี้ 16. วิธีการนี้ทำให้ได้คะแนนจุดตัดที่มีความแม่นยำในการจำแนกความสามารถของนักเรียน					
17. วิธีการนี้มีความเป็นไปได้ในการนำไปใช้ในสถานการณ์ที่เป็นจริงมากน้อยเพียงใด					
18. วิธีการนี้เหมาะสมสำหรับกำหนดคะแนนจุดตัดที่ให้น่าเชื่อถือมากน้อยเพียงใด					
19. คำอธิบายลักษณะความสามารถของนักเรียนแต่ละระดับ ช่วยท่านในการกำหนดคะแนนจุดตัด					
20 คู่มือจัดเรียงข้อสอบ ทำให้เกิดความสะดวกในการกำหนดคะแนนจุดตัด					
21. แบบฟอร์มสำหรับใส่เลขที่หน้าง่ายต่อการกรอก					
22. เปอร์เซนต์ของนักเรียนที่อยู่ในแต่ละระดับช่วยท่านในการกำหนดคะแนนจุดตัด					
23. ค่าเฉลี่ยคะแนนจุดตัดของแต่ละกลุ่มย่อย ช่วยท่านในการกำหนดคะแนนจุดตัด					
24. คู่มือการใช้อ่านเข้าใจมากน้อยเพียงใด					
25. เมื่อท่านอ่านคู่มือการใช้แล้วท่านสามารถนำไปใช้ปฏิบัติได้					

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

แบบประเมินความเหมาะสม และความเป็นไปได้ในการใช้การกำหนดคะแนนจุดตัดด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และวิธีบุ๊คมาร์ค

คำชี้แจง จากการที่ท่านได้ปฏิบัติการกำหนดคะแนนจุดตัดด้วยวิธีแองกอฟที่ได้รับการปรับปรุง และวิธีการบุ๊คมาร์คแล้วนั้น ขอให้ท่านประเมินผลความเหมาะสมและความเป็นไปได้ของวิธีการดังกล่าวนี้

แบบประเมิน แบ่งเป็น 3 ตอน

ตอนที่ 1 ข้อมูลทั่วไปของผู้ตัดสิน

ตอนที่ 2 ความเหมาะสมและความเป็นไปได้ในการใช้วิธีการกำหนดคะแนนจุดตัดด้วยวิธีแองกอฟที่ได้รับการปรับปรุง

ตอนที่ 3 ความเหมาะสมและความเป็นไปได้ในการใช้วิธีการกำหนดคะแนนจุดตัดด้วยวิธีบุ๊คมาร์ค

ในการตอบแบบสอบถามในตอนที่ 2และ 3 ให้ท่านพิจารณาข้อความในแต่ละข้อแล้วตอบคำถามแต่ละข้อตามความจริงที่ตรงกับความคิดเห็นของท่านมากที่สุด โดยให้ระดับความคิดเห็นดังนี้

1 หมายถึง วิธีการแองกอฟที่ได้รับการปรับปรุง และวิธีการบุ๊คมาร์คมีความเหมาะสม และความเป็นไปได้ที่จะนำไปใช้น้อยที่สุด

2 หมายถึง วิธีการแองกอฟที่ได้รับการปรับปรุง และวิธีการบุ๊คมาร์คมีความเหมาะสม และความเป็นไปได้ที่จะนำไปใช้น้อย

3 หมายถึง วิธีการแองกอฟที่ได้รับการปรับปรุงและวิธีการบุ๊คมาร์ค มีความเหมาะสม และความเป็นไปได้ที่จะนำไปใช้ปานกลาง

4 หมายถึง วิธีการแองกอฟที่ได้รับการปรับปรุงและวิธีการบุ๊คมาร์ค มีความเหมาะสม และความเป็นไปได้ที่จะนำไปใช้มาก

5 หมายถึง วิธีการแองกอฟที่ได้รับการปรับปรุงและวิธีการบุ๊คมาร์ค มีความเหมาะสม และความเป็นไปได้ที่จะนำไปใช้มากที่สุด

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตอนที่ 1 ข้อมูลทั่วไปเกี่ยวกับผู้ตัดสิน

4. **ตำแหน่งปัจจุบัน**

☐ 1) ครู ☐ 2) ครูชำนาญการ

☐ 3) ครูชำนาญการพิเศษ

5. **ประสบการณ์สอนวิชาคณิตศาสตร์**

☐ 1) 2 – 5 ปี ☐ 2) 6 – 10 ปี

☐ 3) 11 – 15 ปี ☐ 4) มากกว่า 15 ปี

6. **การศึกษา**

1) ปริญญาตรี สาขา.....

สถานศึกษา.....

2) ปริญญาโท สาขา.....

สถานศึกษา.....

3) กำลังศึกษาปริญญาเอก สาขา.....

สถานศึกษา.....

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตอนที่ 2 โปรดกาเครื่องหมาย ✓ ลงในช่องระดับความเหมาะสมและความเป็นไปได้ในการใช้วิธีการ กำหนดคะแนนจุดตัดด้วย**<u>วิธีแองกอฟที่ได้รับการปรับปรุง</u>** และหากท่านเห็นว่าไม่เหมาะสมหรือมีความเป็นไปไม่ได้ โปรดเสนอข้อคิดเห็นลงในช่องข้อเสนอแนะ

รายการประเมิน	ความเหมาะสม					ความเป็นไปได้				
	5	4	3	2	1	5	4	3	2	1
1. วัตถุประสงค์การกำหนดคะแนนจุดตัด										
2. กระบวนการกำหนดคะแนนจุดตัดทำได้ง่าย ไม่ยุ่งยาก										
3. การดำเนินการกำหนดคะแนนจุดตัด 3.1 ความสมเหตุสมผลของกระบวนการ										
3.2 การจัดให้มีการฝึกปฏิบัติหลังฟังบรรยาย										
3.3 การจัดให้มีการอภิปราย และเปิดโอกาสให้เปลี่ยนคะแนนจุดตัด 3 รอบ										
3.4 การจัดให้มีการกำหนดมาตรฐานการปฏิบัติ										
3.5 คุณสมบัติผู้ตัดสิน คือ มีประสบการณ์การสอนคณิตศาสตร์ไม่น้อยกว่า 5 ปี และเป็นครูระดับชำนาญการขึ้นไป										
4. วิธีการนี้มีความน่าเชื่อถือ										
5. เอกสารประกอบการพิจารณาคะแนนจุดตัด 5.1 แบบทดสอบ										
5.2 ค่าความยากของข้อสอบ										
5.3 กราฟแจกแจงความถี่ค่าความยาก										
5.4 คะแนนจุดตัดของผู้ตัดสินแต่ละคน										
5.5 ค่าเฉลี่ยคะแนนจุดตัดของแต่ละกลุ่มย่อย										

ข้อเสนอแนะที่

แก้ไข……

.....

……

.....

ตอนที่ 3 โปรดกาเครื่องหมาย ✓ ลงในช่องระดับความเหมาะสมและความเป็นไปได้ในการใช้วิธีการกำหนดคะแนนจุดตัดด้วย**วิธีบุ๊คมาร์ค** และหากท่านเห็นว่าไม่เหมาะสม หรือมีความเป็นไปไม่ได้ โปรดเสนอข้อคิดเห็นลงในช่องข้อเสนอแนะ

รายการประเมิน	ความเหมาะสม					ความเป็นไปได้				
	5	4	3	2	1	5	4	3	2	1
1. วัตถุประสงค์การกำหนดคะแนนจุดตัด										
2. กระบวนการกำหนดคะแนนจุดตัดทำได้ง่าย ไม่ยุ่งยาก										
3. การดำเนินการกำหนดคะแนนจุดตัด 3.1 ความสมเหตุสมผลของกระบวนการ										
3.2 การจัดให้มีการฝึกปฏิบัติหลังฟังบรรยาย										
3.3 การจัดให้มีการอภิปราย และเปิดโอกาสให้เปลี่ยนคะแนนจุดตัด 3 รอบ										
3.4 การจัดให้มีการกำหนดมาตรฐานการปฏิบัติ										
3.5 คุณสมบัติผู้ตัดสิน คือ มีประสบการณ์การสอนคณิตศาสตร์ไม่น้อยกว่า 5 ปี และเป็นครูระดับชำนาญการขึ้นไป										
4. วิธีการนี้มีความน่าเชื่อถือ										
5. เอกสารประกอบการพิจารณาคะแนนจุดตัด 5.1 คำอธิบายลักษณะความสามารถของนักเรียนแต่ละระดับ										
5.2 คู่มือจัดเรียงข้อสอบ										
5.3 เปอร์เซนต์ของนักเรียนที่อยู่ในแต่ละระดับ										
5.4 ค่าเฉลี่ยของคะแนนจุดตัดจากกลุ่มย่อย										
5.5 แบบฟอร์มกรอกเลขที่หน้าจากในคู่มือจัดเรียงข้อสอบ										

ข้อเสนอแนะที่

แก้ไข.....

.....

ภาคผนวก ฉ

แบบสอบสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์

ชั้นมัธยมศึกษาปีที่ 3 ภาคเรียนที่ 1 จำนวน 4 ฉบับดังนี้

ฉบับที่ 1 พื้นที่ผิวและปริมาตร

ฉบับที่ 2 กราฟ

ฉบับที่ 3 ระบบสมการเชิงเส้น

ฉบับที่ 4 ความคล้าย

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

แบบสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์
ระดับชั้นมัธยมศึกษาปีที่ 3
บทที่ 1 เรื่อง ปริมาตรและพื้นที่ผิว
ใช้เวลา 50 นาที

คำชี้แจง

1. ข้อสอบฉบับนี้มีทั้งหมด 5 หน้า จำนวน 30 ข้อ (ข้อที่ 1 – 30)
2. ก่อนทำข้อสอบ ให้นักเรียนเขียนชื่อ – สกุล ลงในกระดาษคำตอบ และเขียนเลขที่ตรงมุมขวาข้างบนของกระดาษคำตอบ
3. ในการทำข้อสอบ ให้นักเรียนเลือกคำตอบที่ถูกต้องที่สุดเพียงคำตอบเดียวเท่านั้น แล้วให้นักเรียน**ใช้ดินสอขนาด** 2B **ระบายทึบ** ลงในช่องว่างให้เต็มช่อง (ในแต่ละข้อมีคำตอบที่ถูกต้องหรือเหมาะสมที่สุดเพียงคำตอบเดียว) และในกรณีที่ต้องการเปลี่ยนคำตอบที่ระบายไว้แล้ว ให้ลบคำตอบที่ระบายไว้เดิมให้สะอาด แล้วจึงไประบายตัวเลือกที่ต้องการเลือกต่อไป

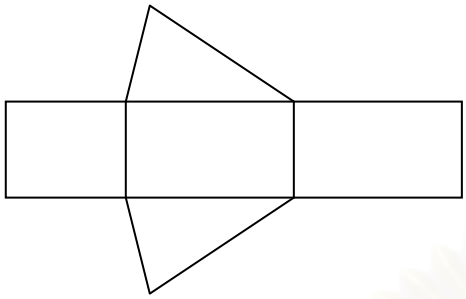
   <u>ตัวอย่าง</u> ถ้าตัวเลือก ก. เป็นคำตอบถูกต้อง ให้นักเรียนทำดังนี้

   ก. ข. ค. ง.
4. ห้ามขีดเขียน หรือทดในกระดาษคำถาม และกระดาษคำตอบ ให้ทดในกระดาษทดที่เตรียมไว้ให้
5. เมื่อหมดเวลาตอบ ให้ส่งทั้งแบบทดสอบ และกระดาษคำตอบ

ขอให้นักเรียนโชคดีในการสอบ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

1.



จากรูปเป็นรูปคลี่ของรูปทรงชนิดใด

ก. พีระมิดฐานสามเหลี่ยม

ข. พีระมิดฐานสี่เหลี่ยม

ค. ปริซึมสามเหลี่ยม

ง. ปริซึมสี่เหลี่ยม

2. ข้อใดเป็นทรงสามมิติที่มีคุณสมบัติ “มีสูงตรงและสูงเอียง”

ก. ทรงกระบอก ข. ปริซึม

ค. กรวย ง. ทรงกลม

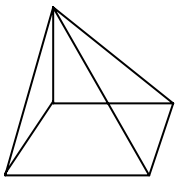
3. ข้อใดเป็นทรงสามมิติที่มีคุณสมบัติ “จุดทุกจุดบนผิวโค้งอยู่ห่างจากจุดคงที่จุดหนึ่งเป็นระยะเท่ากัน”

ก. ทรงกระบอก ข. พีระมิด

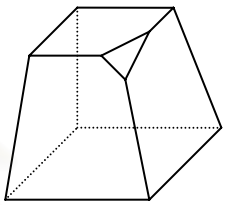
ค. กรวย ง. ทรงกลม

4. ข้อใดเป็นรูปของปริซึม

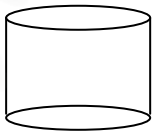
ก.



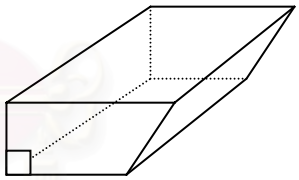
ข.



ค.



ง.



5. ปริซึมสามเหลี่ยมมุมฉาก มีด้านทั้งสามยาว 5 , 12 และ 13 เซนติเมตร สูง 30 เซนติเมตร จงหาว่าปริซึมนี้มีปริมาตรเท่ากับกี่ลิตร

ก. 900 ข. 9

ค. 0.9 ง. 0.09

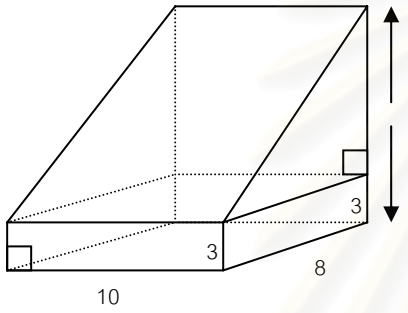
สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

6. ถังทรงสี่เหลี่ยมมุมฉากยาว 20 เมตร กว้าง 12 เมตร ถ้าต้องการเก็บน้ำไว้ในถัง 3,600 ลูกบาศก์เมตร ระดับน้ำจะสูงจากก้นถังกี่เมตร

ก. 12 ข. 13

ค. 14 ง. 15

7.



จากรูปข้างต้น ปริมาตรของทรงสามมิตินี้เท่ากับเท่าไร

ก. 240 ลูกบาศก์หน่วย

ข. 600 ลูกบาศก์หน่วย

ค. 720 ลูกบาศก์หน่วย

ง. 960 ลูกบาศก์หน่วย

8. วัตถุทรงกระบอกมีปริมาตร 1,540 ลูกบาศก์เซนติเมตร มีพื้นที่ฐาน 154 ตารางเซนติเมตร วัตถุทรงกระบอกสูงกี่เซนติเมตร

ก. 2 เซนติเมตร ข. 4 เซนติเมตร

ค. 8 เซนติเมตร ง. 10 เซนติเมตร

9. ข้าวหลามกระบอกหนึ่ง ปล้องสำหรับใส่ข้าวหลามสูง 24 เซนติเมตร มีเส้นผ่านศูนย์กลางภายใน 3.5 เซนติเมตร จุข้าวหลามได้มากที่สุดกี่ลูกบาศก์เซนติเมตร

ก. 231 ลูกบาศก์เซนติเมตร

ข. 234 ลูกบาศก์เซนติเมตร

ค. 343 ลูกบาศก์เซนติเมตร

ง. 512 ลูกบาศก์เซนติเมตร

10. ดินน้ำมันรูปทรงกระบอกมีความยาว 14 เซนติเมตร รัศมี 3 เซนติเมตร ถ้าต้องการปั้นดินน้ำมันรูปทรงกระบอก จำนวน 5 ชิ้น จะต้องใช้ดินน้ำมันทั้งหมดกี่ลูกบาศก์เซนติเมตร

(กำหนด π = 3.143)

ก. 99 ข. 396

ค. 1,980 ง. 2,500

11. พีระมิดตรงฐานสี่เหลี่ยมจัตุรัสยาวด้านละ 2 นิ้ว และพีระมิดสูง 6 นิ้ว วางในแก้วทรงกระบอก ซึ่งมีรัศมีฐานยาว 2 นิ้ว และสูง 7 นิ้ว เมื่อเติมน้ำจนเต็มแก้ว ปริมาตรของน้ำในแก้วเท่ากับกี่ลูกบาศก์นิ้ว (กำหนด $\pi = \frac{22}{7}$)

ก. 78 ลูกบาศก์นิ้ว

ข. 80 ลูกบาศก์นิ้ว

ค. 86 ลูกบาศก์นิ้ว

ง. 92 ลูกบาศก์นิ้ว

12. พีระมิดฐานสี่เหลี่ยมผืนผ้ายาว 10 เซนติเมตร กว้าง 6 เซนติเมตร สูง 15 เซนติเมตร จะมีปริมาตรตรงกับข้อใด

ก. 300 ลูกบาศก์เซนติเมตร

ข. 320 ลูกบาศก์เซนติเมตร

ค. 360 ลูกบาศก์เซนติเมตร

ง. 400 ลูกบาศก์เซนติเมตร

13. พีระมิดตรงฐานสี่เหลี่ยมจัตุรัสยาวด้านละ 24 เซนติเมตร มีสูงเอียงยาว 20 เซนติเมตร ปริมาตรของพีระมิดเป็นเท่าไร

ก. 3,072 ลูกบาศก์เซนติเมตร

ข. 4,608 ลูกบาศก์เซนติเมตร

ค. 7,688 ลูกบาศก์เซนติเมตร

ง. 9,216 ลูกบาศก์เซนติเมตร

14. กรวยอันหนึ่งวัดความยาวรอบฐานได้ 10π เซนติเมตร สูง 12 เซนติเมตร ปริมาตรของกรวยเป็นเท่าไร

ก. 20π ลูกบาศก์เซนติเมตร

ข. 50π ลูกบาศก์เซนติเมตร

ค. 83π ลูกบาศก์เซนติเมตร

ง. 100π ลูกบาศก์เซนติเมตร

15. จงหาความสูงของกรวยอันหนึ่ง ซึ่งมีปริมาตร 72π ลูกบาศก์เซนติเมตร และมีเส้นผ่านศูนย์กลางของฐานยาว 8 เซนติเมตร

ก. 3.4 ข. 13.5

ค. 10.6 ง. 15.5

16. กรวยกระดาษสำหรับใส่น้ำดื่มสูงประมาณ 6 เซนติเมตร มีพื้นที่ฐาน 38 ตารางเซนติเมตร ถ้าถังน้ำกดใบหนึ่งจุน้ำดื่มไว้ 5 ลิตร จะต้องเตรียมกรวยกระดาษไว้ประมาณกี่ใบจึงจะบรรจุน้ำได้ทั้งหมด

ก. 7 ใบ ข. 35 ใบ

ค. 66 ใบ ง. 158 ใบ

17. นำตะกั่วกลม มีรัศมียาว 3 เซนติเมตร มาหลอมทำเป็นทรงกลมเล็กๆ ซึ่งมีรัศมี 1 เซนติเมตร จะได้ทรงกลมเล็กๆ กี่ลูก (π =3.143)

ก. 9 ลูก ข. 12 ลูก

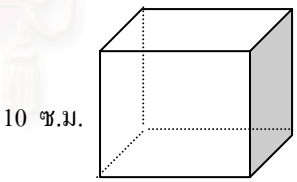
ค. 25 ลูก ง. 27 ลูก

18. ลูกวอลเลย์ลูกหนึ่งวัดความยาวรอบวงได้ 52.8 เซนติเมตร ถ้าลูกวอลเลย์บอลทำด้วยหนังหนา 0.3 เซนติเมตร จะบรรจุลมได้มากที่สุดกี่ลูกบาศก์เซนติเมตร (π =3.143)

ก. 828 ข. 2,478

ค. 2,227 ง. 2,552

19.



พื้นที่ผิวของปริซึมเท่ากับข้อใดต่อไปนี้

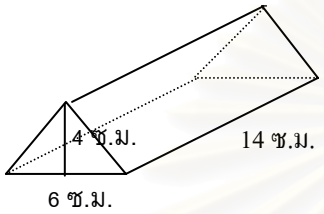
ก. 376 ตารางเซนติเมตร

ข. 504 ตารางเซนติเมตร

ค. 752 ตารางเซนติเมตร

ง. 648 ตารางเซนติเมตร

20. ช็อกโกแลตก้อนหนึ่งมีลักษณะเป็นปริซึมสามเหลี่ยมหน้าจั่วดังรูป จะต้องใช้กระดาษห่อกี่ตารางเซนติเมตร ช็อกโกแลตที่มีพื้นที่อย่างน้อยกี่ตารางเซนติเมตร



ก. 224 ข. 248

ค. 264 ง. 286

21. บ่อเลี้ยงปลาที่บ้านแห่งหนึ่งเป็นทรงสี่เหลี่ยมมุมฉากขนาดกว้าง 3 เมตร ยาว 5 เมตร และลึก 1 เมตร ถ้าเสียค่าทาสีตารางเมตรละ 80 บาทจะเสียค่าทาสีพื้นที่ผิวข้างในบ่อทั้งหมดกี่บาท

ก. 1,420 บาท ข. 2,000 บาท

ค. 2,480 บาท ง. 2,750 บาท

22. จงหาพื้นที่ผิวข้างของรูปทรงกระบอกที่มีเส้นผ่านศูนย์กลางยาว 4 เซนติเมตร และสูง 5 เซนติเมตร ($\pi = 3.14$)

ก. 12.56 ตารางเซนติเมตร

ข. 25.12 ตารางเซนติเมตร

ค. 62.8 ตารางเซนติเมตร

ง. 87.92 ตารางเซนติเมตร

23. ทรงกระบอกอันหนึ่งมีเส้นผ่านศูนย์กลางของฐาน 6 เซนติเมตร และสูง 4 เซนติเมตร พื้นที่ผิวของทรงกระบอกอันนี้เป็นเท่าไร (กำหนด $\pi = 3.143$)

ก. 128 ตารางเซนติเมตร

ข. 132 ตารางเซนติเมตร

ค. 372 ตารางเซนติเมตร

ง. 384 ตารางเซนติเมตร

24. แท่งโลหะทรงกระบอกมีพื้นที่ผิวข้าง 132 ตารางนิ้ว ยาว 15 นิ้ว อยากทราบว่า ถ้าใช้เชือกพันรอบแท่งโลหะตามแนวโค้ง 1 รอบต้องใช้เชือกยาวเท่าไร ($\pi = \frac{22}{7}$)

ก. 8.8 นิ้ว ข. 12.2 นิ้ว

ค. 35.2 นิ้ว ง. 40.8 นิ้ว

25. ข้าวสาร 210 ลิตร คิดเป็นข้าวสารกี่ถัง

ก. 2.1 ถัง ข. 10.5 ถัง

ค. 21 ถัง ง. 24 ถัง

26. รถบรรทุกคันหนึ่งมีส่วนที่เป็นทรงกระบะยาว 4 เมตร กว้าง 2 เมตร และสูง 2 เมตร จะบรรทุกข้าวเปลือกได้ทั้งหมดกี่ลิตร

ก. 1,600 ลิตร ข. 4,800 ลิตร

ค. 16,000 ลิตร ง. 19,000 ลิตร

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

27. ความจุในข้อใดมีค่าน้อยที่สุด

ก. 1,000 ถัง

ข. 2 กิโลลิตร

ค. 20,000 ลิตร

ง. 400,000 ลูกบาศก์เซนติเมตร

28. น้ำยาสระผมชนิดหนึ่งบรรจุขวดมีปริมาณสุทธิ 200 มิลลิลิตร ราคา 120 บาท น้ำยาชนิดเดียวกันบรรจุขวดมีปริมาณสุทธิ 80 มิลลิลิตร ราคา 50 บาท ถ้าต้องการเลือกซื้อควรเลือกซื้ออย่างไร

ก. เลือกซื้อราคาขวดละ 120 บาท เพราะ ได้ปริมาณมากกว่า

ข. เลือกซื้อราคาขวดละ 50 บาท เพราะราคาถูกกว่า

ค. เลือกซื้อราคาขวดละ 120 บาท เพราะราคาถูกกว่า

ง. เลือกซื้ออย่างไรก็ได้ เพราะ ราคาเท่ากัน

29. น้ำอัดลมขวดหนึ่งจุน้ำอัดลม 1,500 ลูกบาศก์เซนติเมตร เทน้ำอัดลมลงในแก้วซึ่งมีความจุ 125 ลูกบาศก์เซนติเมตร จะได้น้ำอัดลมกี่แก้ว

ก. 8 แก้ว ข. 9 แก้ว

ค. 12 แก้ว ง. 15 แก้ว

30. ชาวนามีข้าวเปลือก 10 เกวียน อยากทราบว่าถ้าขายข้าวในราคาถังละ 60 บาท เขาจะได้เงินเท่าไร

ก. 600 บาท ข. 6,000 บาท

ค. 60,000 บาท ง. 600,000 บาท

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

แบบสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์
ระดับชั้นมัธยมศึกษาปีที่ 3
บทที่ 2 เรื่อง กราฟ
ใช้เวลา 50 นาที

คำชี้แจง

1. ข้อสอบฉบับนี้มีทั้งหมด 6 หน้า จำนวน 30 ข้อ (ข้อที่ 31 – 60)
2. ก่อนทำข้อสอบ ให้นักเรียนเขียนชื่อ – สกุล ลงในกระดาษคำตอบ และเขียนเลขที่ตรงมุมขวาข้างบนของกระดาษคำตอบ
3. ในการทำข้อสอบ ให้นักเรียนเลือกคำตอบที่ถูกต้องที่สุดเพียงคำตอบเดียวเท่านั้น แล้วให้นักเรียน**ใช้ดินสอขนาด** 2B **ระบายทึบ** ลงในช่องว่างให้เต็มช่อง (ในแต่ละข้อมีคำตอบที่ถูกต้องหรือเหมาะสมที่สุดเพียงคำตอบเดียว) และในกรณีที่ต้องการเปลี่ยนคำตอบที่ระบายไว้แล้ว ให้ลบคำตอบที่ระบายไว้เดิมให้สะอาด แล้วจึงไประบายตัวเลือกที่ต้องการเลือกต่อไป

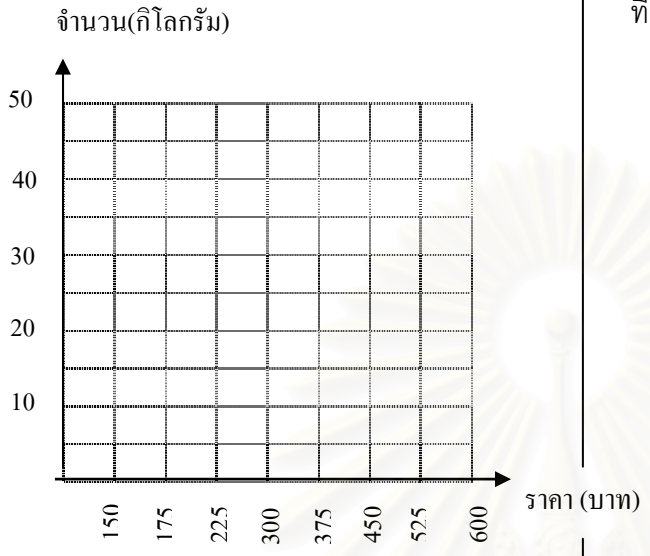
   <u>ตัวอย่าง</u> ถ้าตัวเลือก ก. เป็นคำตอบถูกต้อง ให้นักเรียนทำดังนี้

   ก. ข. ค. ง.
4. ห้ามขีดเขียน หรือทดในกระดาษคำถาม และกระดาษคำตอบ ให้ทดในกระดาษทดที่เตรียมไว้ให้
5. เมื่อหมดเวลาตอบ ให้ส่งทั้งแบบทดสอบ และกระดาษคำตอบ

ขอให้นักเรียนโชคดีในการสอบ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

31. กราฟแสดงราคาและจำนวนสินค้าชนิดหนึ่งเป็นกิโลกรัม



จากกราฟเมื่อนำมาเขียนตารางแสดงความสัมพันธ์ระหว่างราคา และจำนวนสินค้าจะมีลักษณะคือข้อใด

ก.

จำนวน (กิโลกรัม)	10	20	30
ราคา (บาท)	120	225	325

ข.

จำนวน (กิโลกรัม)	0	25	36
ราคา (บาท)	0	300	450

ค.

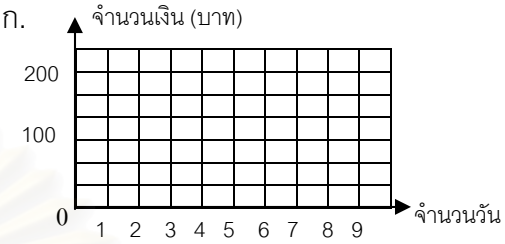
จำนวน (กิโลกรัม)	15	20	36
ราคา (บาท)	180	240	400

ง.

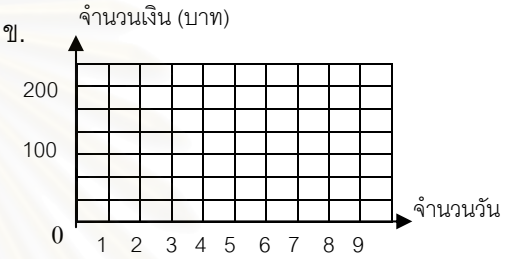
จำนวน (กิโลกรัม)	10	25	50
ราคา (บาท)	120	300	600

32. มีเงิน 200 บาท ใช้ไปวันละ 25 บาท เขียนกราฟเส้นตรงแสดงความสัมพันธ์ระหว่างจำนวนเงินที่เหลือกับจำนวนวันที่ใช้จ่ายได้ดังข้อใดต่อไปนี้

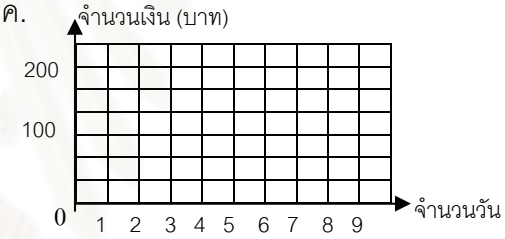
ก.



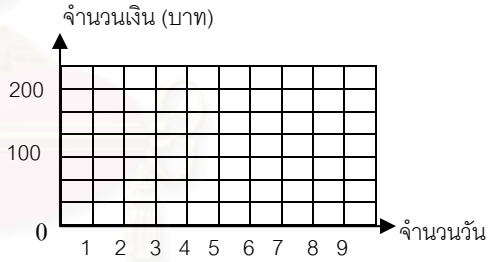
ข.



ค.



ง.



33. จากตาราง จงพิจารณาว่าความสัมพันธ์ที่กำหนดให้ เมื่อนำไปเขียนกราฟจะมีลักษณะดังข้อใด

x	-2	0	2
y	-1	2	5

ก. จุด 3 จุดอยู่แนวเส้นตรงเดียวกัน

ข. จุด 3 จุดไม่อยู่แนวเส้นตรงเดียวกัน

ค. กราฟเส้นโค้ง

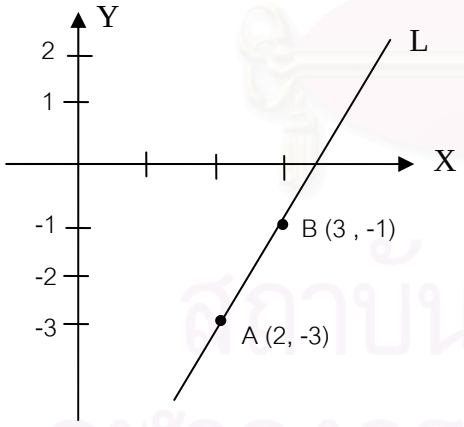
ง. ไม่มีข้อใดถูกต้องเลย

34. จุด (9,12) และ (9,36) อยู่บนเส้นตรงเส้นหนึ่ง จุดใดอยู่บนเส้นตรงเส้นนี้

ก. (200, 12) ข. (0, 0)

ค. (9, 108) ง. (0 , 36)

35.



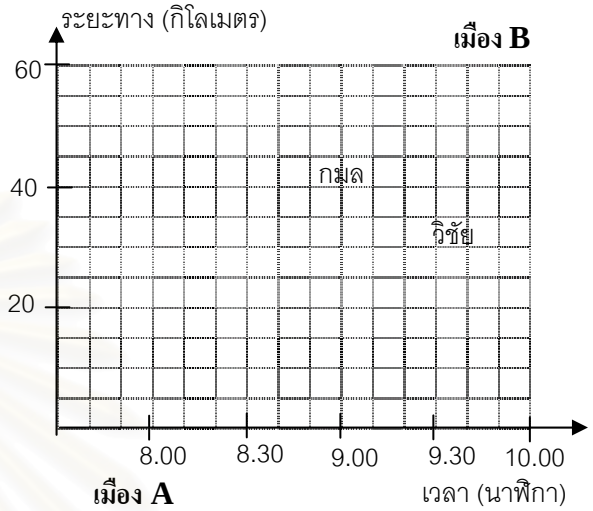
คู่อันดับใดอยู่บนเส้นตรง L

ก. (4 , 0) ข. (4 , 1)

ค. (4 , 1.5) ง. (4 , 2)

ใช้กราฟต่อไปนี้ตอบคำถามข้อ 36 และข้อ 37

กราฟแสดงการเดินทางด้วยรถจักรยานยนต์ของกมล และวิชัยจากเมือง A ไปเมือง B



36. 30 นาทีแรกกมลขี่จักรยานยนต์ด้วยอัตราเร็วกี่กิโลเมตรต่อชั่วโมง

ก. 20 กิโลเมตรต่อชั่วโมง

ข. 30 กิโลเมตรต่อชั่วโมง

ค. 40 กิโลเมตรต่อชั่วโมง

ง. 50 กิโลเมตรต่อชั่วโมง

37. ใครขี่จักรยานยนต์ถึงเมือง B ก่อนกันและถึงก่อนกันกี่นาที

ก. วิชัยถึงก่อนกมล 15 นาที

ข. วิชัยถึงก่อนกมล 10 นาที

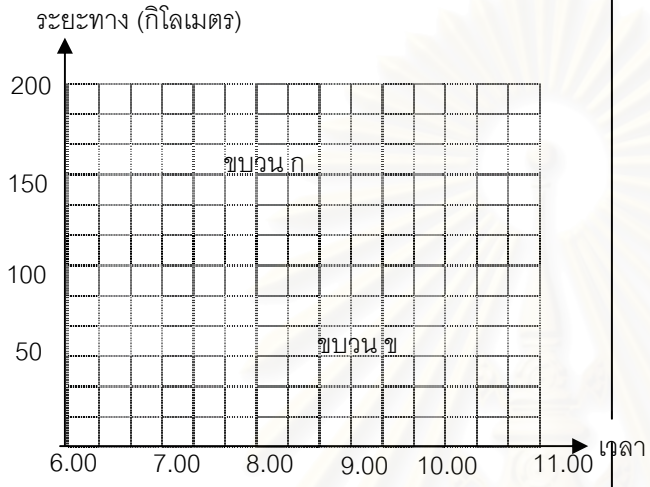
ค. กมลถึงก่อนวิชัย 15 นาที

ง. กมลถึงก่อนวิชัย 10 นาที

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ใช้กราฟต่อไปนี้ตอบคำถามข้อ 38 และ ข้อ 39

กราฟแสดงการเดินทางของรถไฟสองขบวน



38. ถ้ารถไฟสองขบวนนี้แล่นต่อไปเรื่อยๆ จะไปทันกันเมื่อเวลาใด

ก. 11.30 น. ข. 12.00 น.

ค. 12.30 น. ง. 13.00 น.

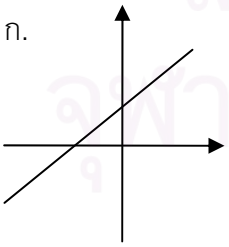
39. เวลาใดที่รถไฟสองขบวนมีระยะห่างกันมากที่สุด

ก. 9.20 น. ข. 9.40 น.

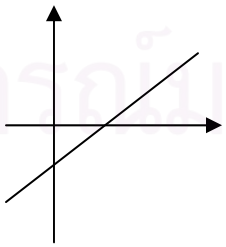
ค. 10.00 น. ง. 10.20 น.

40. กราฟข้อใดมีลักษณะคล้ายกับกราฟของสมการ $2x + 3y - 5 = 0$

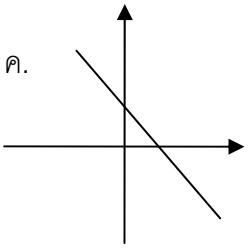
ก.



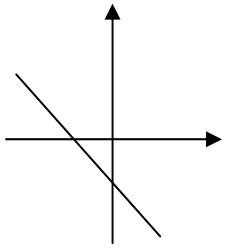
ข.



ค.



ง.



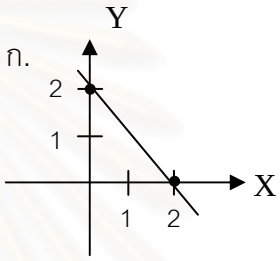
41. สมการในข้อใดมีกราฟผ่านจุด (0,0)

ก. $x = -\frac{1}{2} + y$ ข. $Y = 0.35$

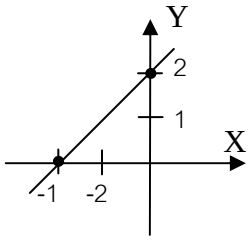
ค. $y = 50X$ ง. $2x + y = 5$

42. ข้อใดเป็นกราฟของสมการ $y = 2 - x$

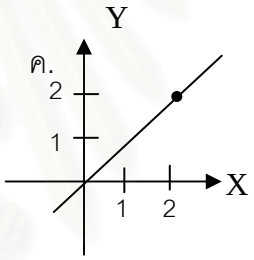
ก.



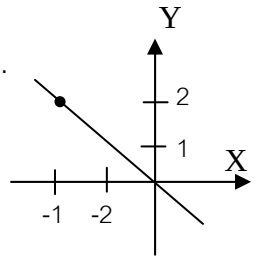
ข.



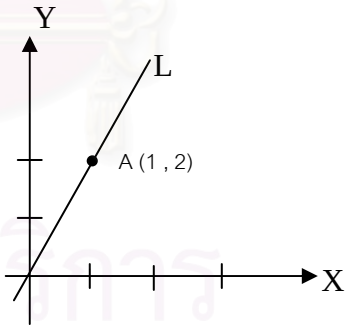
ค.



ง.



43.



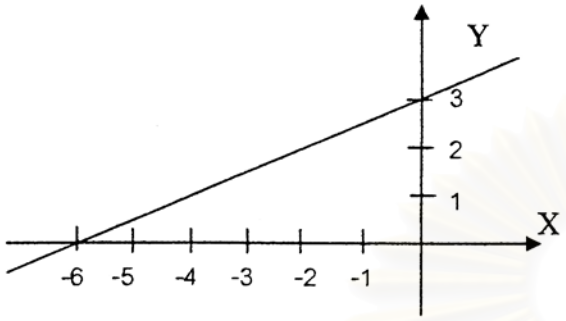
สมการของเส้นตรง L คือข้อใด

ก. $x + y = 2$ ข. $2x + y = 0$

ค. $2x - y = 2$ ง. $2x - y = 0$

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

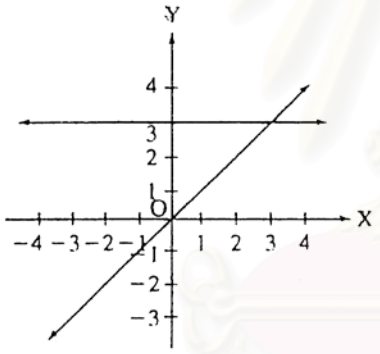
44. จากรูปเป็นกราฟของสมการใด



ก. $2x - 4y = 12$ ข. $2x - 4y = -12$

ค. $2x + 4y = -12$ ง. $2x + 4y = 12$

45. กราฟเส้นตรงที่กำหนดให้เป็นกราฟของสมการใดต่อไปนี้



ก. $y = 3$ และ $y = X$

ข. $X = 3$ และ $y = X$

ค. $y + 3 = 0$ และ $y - \frac{2}{3}X = 0$

ง. $X + 3 = 0$ และ $y + \frac{2}{3}X = 0$

46. ถ้าคู่อันดับ (0 , -8) , (1 , -5) , (2 , -2) และ (3 , Y) อยู่บนเส้นตรงเดียวกัน แล้ว Y มีค่าเท่าใด

ก. –1 ข. 0

ค. 1 ง. 2

47. กราฟของเส้นตรง $4X + kY = 7$ ผ่านจุด (k ,3) จงหาค่า k

ก. 1 ข. 2

ค. 4 ง. 5

48. กราฟเส้นตรงมีสมการ $1.6X + 0.8Y - 1.6 = 0$ ถ้า (c , 0) อยู่บนเส้นตรงนี้ c เท่ากับข้อใด

ก. –1 ข. –0.2

ค. 0.2 ง. 1

49. กราฟของสมการคู่ใดขนานกัน

ก. $2x - y - 1 = 0$ กับ
$4x + 2 + 2y = 0$

ข. $2x - y + 1 = 0$ กับ
$y = 2x + 2$

ค. $y + x = 3$ กับ
$2x + 6y = 2$

ง. $4 + x = y$ กับ
$y - x = 4$

50. เส้นตรง $y = 2x + 7$ กับ $y = ax - 5$ ขนานกัน จงหาค่า a

ก. 1 ข. –1

ค. –2 ง. 2

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

51. สมการเส้นตรงสองตัวแปรคู่ใด มีกราฟเป็นเส้นตรงที่ตัดกัน

ก. $4x = 14 - 2y$ กับ $5x = y + 7$

ข. $5x - 4y - 6 = 0$ กับ $10x - 8y - 6 = 0$

ค. $3x - y = 4$ กับ $2y - 6x - 4 = 0$

ง. $x - 4 = 0$ กับ $2x = 0$

52. กราฟของสมการเส้นตรง $x - 7y - 11 = 0$ กับ $3x + 5y - 7 = 0$ ตัดกันที่จุดใด

ก. (4 , -1) ข. (-1 , 4)

ค. (4 , 1) ง. (-1 , -4)

53. กราฟของสมการ $y = mx + c$ ตัดแกน x ทางขวามือของจุดกำเนิน และตัดแกน y ทางด้านบนของจุดกำเนิด เมื่อ m , c มีลักษณะอย่างไร

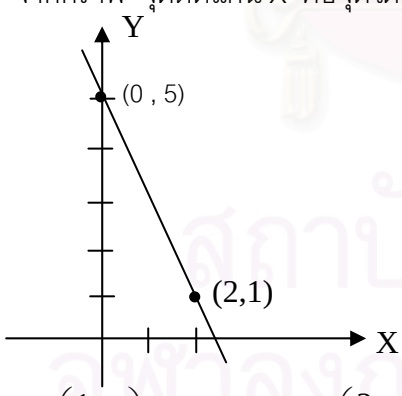
ก. $c < 0$, $m < 0$

ข. $c < 0$, $m > 0$

ค. $c > 0$, $m < 0$

ง. $c > 0$, $m > 0$

54. จากกราฟ จุดตัดแกน X คือจุดใดต่อไปนี้



ก. $\left(\frac{1}{2},0\right)$ ข. $\left(\frac{3}{2},0\right)$

ค. $\left(\frac{5}{2},0\right)$ ง. (3 , 0)

55. ถ้ากราฟของสมการ $y = mx - 5$ ผ่านจุด (2 , -1) แล้วลักษณะของกราฟของสมการนี้จะเป็นอย่างไร

ก. ทำมุมแหลมกับแกน x

ข. ทำมุมป้านกับแกน x

ค. ขนานกับแกน x

ง. ตั้งฉากกับแกน x

56. สมการกราฟในข้อใดให้รูปกราฟทำมุมป้านกับแกน x

ก. $4x - 4y = 3$ ข. $4y = -x - 1$

ค. $2x - y = 12$ ง. $y = 2x + 6$

57. กำหนดสมการ $y = -\frac{1}{3}X - 8$ มีลักษณะของกราฟดังข้อใดต่อไปนี้

ก. ทำมุมแหลมกับแกน X ในทิศทางทวนเข็มนาฬิกา

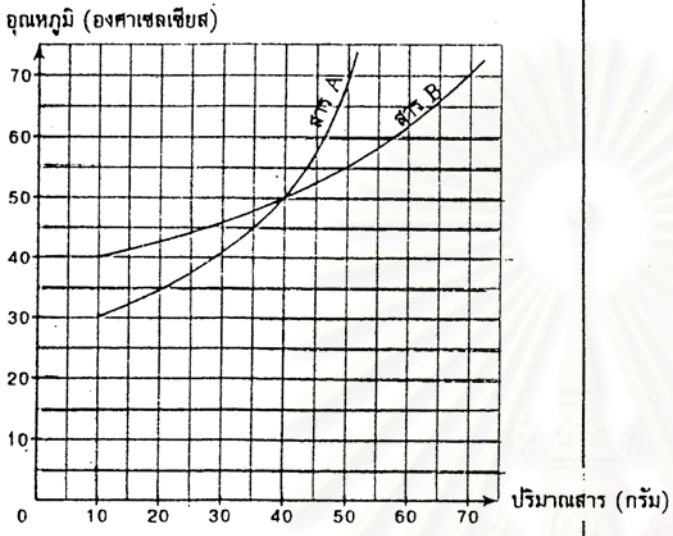
ข. ตัดแกน X และแกน Y ที่จุด (0 , 0)

ค. จุดตัดแกน Y อยู่เหนือแกน X

ง. จุดตัดแกน Y อยู่ใต้แกน X

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

58. สารบางชนิดเมื่อละลายในน้ำจะทำให้อุณหภูมิของน้ำเปลี่ยนไป กราฟแสดงความสัมพันธ์ระหว่างอุณหภูมิของน้ำกับปริมาณสารที่ละลายในน้ำ 100 กรัม เป็นดังนี้

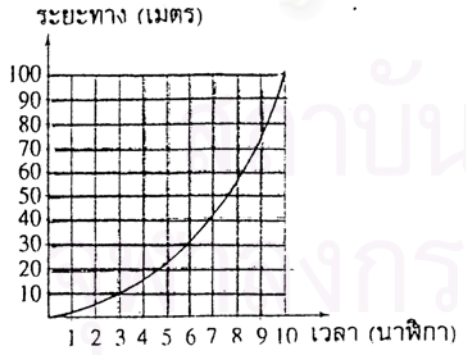


จากกราฟ ปริมาณของสาร A และสาร B ปริมาณเท่าไรที่ละลายในน้ำ 100 กรัม แล้วทำให้อุณหภูมิเท่ากัน

ก. 20 กรัม ข. 30 กรัม

ค. 40 กรัม ง. 50 กรัม

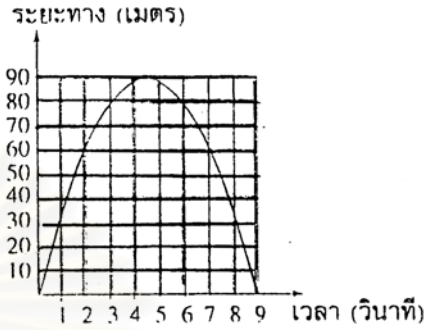
59. จากกราฟ แสดงการเคลื่อนที่ของวัตถุ ถ้าวัตถุเคลื่อนที่ไป 15 วินาที จะได้ทางเท่าไร



ก. 175 เมตร ข. 180 เมตร

ค. 210 เมตร ง. 225 เมตร

60. จากกราฟ แสดงการเคลื่อนที่ของวัตถุ ถ้าวัตถุอยู่สูง 60 เมตร เวลาผ่านไปเท่าไร



ก. 2 และ 7 วินาที

ข. 3 และ 6 วินาที

ค. 3.5 และ 6.5 วินาที

ง. 4 และ 5 วินาที

แบบสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์
ระดับชั้นมัธยมศึกษาปีที่ 3
บทที่ 3 เรื่อง ระบบสมการเชิงเส้น
ใช้เวลา 20 นาที

คำชี้แจง

1. ข้อสอบฉบับนี้มีทั้งหมด 2 หน้า จำนวน 16 ข้อ (ข้อ 61 – 76)
2. ก่อนทำข้อสอบ ให้นักเรียนเขียนชื่อ – สกุล ลงในกระดาษคำตอบ และเขียนเลขที่ตรงมุมขวาข้างบนของกระดาษคำตอบ
3. ในการทำข้อสอบ ให้นักเรียนเลือกคำตอบที่ถูกต้องที่สุดเพียงคำตอบเดียวเท่านั้น แล้วให้นักเรียน**ใช้ดินสอขนาด** 2B **ระบายทึบ** ลงในช่องว่างให้เต็มช่อง (ในแต่ละข้อมีคำตอบที่ถูกต้องหรือเหมาะสมที่สุดเพียงคำตอบเดียว) และในกรณีที่ต้องการเปลี่ยนคำตอบที่ระบายไว้แล้ว ให้ลบคำตอบที่ระบายไว้เดิมให้สะอาด แล้วจึงไประบายตัวเลือกที่ต้องการเลือกต่อไป

   <u>ตัวอย่าง</u> ถ้าตัวเลือก ก. เป็นคำตอบถูกต้อง ให้นักเรียนทำดังนี้

   ก. ข. ค. ง.
4. ห้ามขีดเขียน หรือทดในกระดาษคำถาม และกระดาษคำตอบ ให้ทดในกระดาษทดที่เตรียมไว้ให้
5. เมื่อหมดเวลาตอบ ให้ส่งทั้งแบบทดสอบ และกระดาษคำตอบ

ขอให้นักเรียนโชคดีในการสอบ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

61. ระบบสมการใด ไม่มีคำตอบ

ก. $x + y = 5$; $x - y = 5$

ข. $2x - y = 5$; $4x - 2y = 10$

ค. $-x + y = 15$; $x - y = 5$

ง. $2x - y = 5$; $2x + y = 5$

62. ระบบสมการในข้อใด มีคำตอบมากมาย**ไม่**จำกัดรูป

ก. $x - 2y = 1$; $x + 2y = 1$

ข. $x - 2y = 1$; $2x - 4y = 2$

ค. $x - 2y = 1$; $2x - 2y = 2$

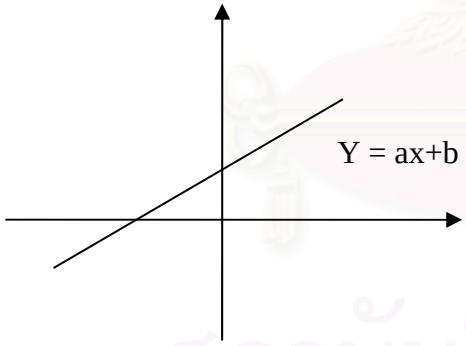
ง. $x - 2y = 1$; $2x - 4y = 4$

63. ถ้า x และ y แทนจำนวนเต็มบวก สมการ $x + y = 10$ จะมีคำตอบทั้งสิ้นกี่คำตอบ

ก. 9 ข. 10

ค. 11 ง. มีมากกมายนับไม่ถ้วน

64. จากกราฟ ข้อใดถูกต้อง



ก. $a > 0$, $b > 0$ ข. $a > 0$, $b < 0$

ค. $a < 0$, $b < 0$ ง. $a < 0$, $b > 0$

65. กราฟของสมการในข้อใดขนานกับกราฟของสมการ $2x - 3y = 6$

ก. $3x - 2y = 6$ ข. $3x + 2y = 0$

ค. $-2x + 3y = 0$ ง. $-2x - 3y = 6$

66. จงพิจารณาว่าข้อใดต่อไปนี้**ผิด**

ก. กราฟของ $y = -3x$ ผ่านจุด (0,0)

ข. กราฟของ $y = -2x + 3$ ผ่านจุด (0, 3)

ค. กราฟของ $y = 4$ ขนานแกน y

ง. กราฟของ $x = 2$ ผ่านจุด (2 , -2)

67. กราฟของเส้นตรงในข้อใดทับกับกราฟของเส้นตรง $2x + y = 5$ เป็นเส้นตรงเดียวกัน

ก. $6x + 3y = 15$ ข. $y - 2x = 5$

ค. $3x + 4y = 5$ ง. $\frac{2}{5}x - \frac{1}{5}y = 1$

68. คำตอบของระบบสมการ

$4(x - 2y) - (5x + 3y) = 30$ และ

$3(3x + 7y) - 2(x + 9y) = 12$ คือข้อใด

ก. (-3 , 3) ข. (3 , -3)

ค. (-3 , -3) ง. (3 , 3)

69. ถ้า (a , b) เป็นคำตอบของสมการ

$x + 3y - 8 = 0$ และ

$x - 2y - 3 = 0$ แล้ว $a - 5b$ มีค่าเท่าไร

ก. -4 ข. -1

ค. 0 ง. 6

70. ระบบสมการใดมีคำตอบเป็น (5,1)

ก. $4x + y = 21$, $5x - y = 26$

ข. $2x - 3y = 8$, $x + 3y = 11$

ค. $4x - 5y = 10$, $2x + 5y = 3$

ง. $2x + y = 11$, $2x - y = 9$

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

71. การแก้ระบบสมการเชิงเส้นด้วยกราฟทำอย่างไร

ก. เขียนกราฟแล้วอ่านจุดที่กราฟตัดแกน x

ข. เขียนกราฟแล้วอ่านจุดที่กราฟตัดแกน y

ค. เขียนกราฟแล้วอ่านจุดที่กราฟตัดกัน

ง. เขียนกราฟแล้วอ่านจุดที่กราฟอยู่ห่างจากแกน y

72. จากสมการ $y = ax + b$ ถ้าสมการ 2 สมการมีค่า a เท่ากัน จะได้กราฟมีลักษณะอย่างไร

ก. ขนานกัน ข. ตัดกันที่จุด (0,0)

ค. ตั้งฉากกัน ง. ต่างก็ขนานกับแกน x

73. เมื่อ 6 ปีก่อน บิดามีอายุเป็น 6 เท่าของบุตร แต่อีก 3 ปี ข้างหน้า บิดาจะมีอายุเป็น 3 เท่าของบุตร อายุของบิดา และบุตรในปัจจุบันตามลำดับตรงกับกี่ปี

ก. 38 และ 8 ปี ข. 40 และ 10 ปี

ค. 42 และ 12 ปี ง. 44 และ 14 ปี

74. แม่ค้าซื้อมะม่วงมาสองชนิด ชนิดแรก 5 ผล ราคา 30 บาท ชนิดที่สอง 3 ผล ราคา 20 บาท ซึ่งมาทั้งสิ้นเป็นเงิน 280 บาท แล้วขายชนิดแรกไป 3 ผล ราคา 20 บาท ชนิดที่สอง 5 ผล ราคา 30 บาท ปรากฏว่าได้กำไร 10 บาท อยากทราบว่าแม่ค้าซื้อมะม่วงทั้งสองชนิดรวมกันกี่ผล

ก. 48 ผล ข. 45 ผล

ค. 42 ผล ง. 30 ผล

75. สมปองมีเงินออม 25,000 บาท เขานำเงินแบ่งเป็น 2 ส่วน และนำไปลงทุนกับกองทุนที่ให้ผลตอบแทน 8 % และ 12 % ตามลำดับ ถึงสิ้นปี เขาได้รับเงินปันผล 2,300 บาท อยากทราบว่าส่วนต่างของเงินลงทุนทั้งสองส่วนเป็นเท่าใด

ก. 7,500 บาท ข. 10,000 บาท

ค. 15,000 บาท ง. 17,500 บาท

76. ส้มมาสองชนิดราคากิโลกรัมละ 35 บาท และ 20 บาท ถ้านำมาปนกันแล้วขายไปกิโลกรัมละ 25 บาท ปรากฏว่าขายได้เงินเท่าทุนพอดี อัตราส่วนในการผสมส้มชนิดแรกต่อชนิดหลังเป็น เท่าไร

ก. 1 : 2 ข. 1 : 3

ค. 2 : 3 ง. 2 : 5

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

แบบสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์
ระดับชั้นมัธยมศึกษาปีที่ 3
บทที่ 4 เรื่อง ความคล้าย
ใช้เวลา 30 นาที

คำชี้แจง

1. ข้อสอบฉบับนี้มีทั้งหมด 7 หน้า จำนวน 24 ข้อ (ข้อ 77 ถึง 100)
2. ก่อนทำข้อสอบ ให้นักเรียนเขียนชื่อ – สกุล ลงในกระดาษคำตอบ และเขียนเลขที่ตรงมุมขวาข้างบนของกระดาษคำตอบ
3. ในการทำข้อสอบ ให้นักเรียนเลือกคำตอบที่ถูกต้องที่สุดเพียงคำตอบเดียวเท่านั้น แล้วให้นักเรียน**ใช้ดินสอขนาด** 2B **ระบายทึบ** ลงในช่องว่างให้เต็มช่อง (ในแต่ละข้อมีคำตอบที่ถูกต้องหรือเหมาะสมที่สุดเพียงคำตอบเดียว) และในกรณีที่ต้องการเปลี่ยนคำตอบที่ระบายไว้แล้ว ให้ลบคำตอบที่ระบายไว้เดิมให้สะอาด แล้วจึงไประบายตัวเลือกที่ต้องการเลือกต่อไป

   <u>ตัวอย่าง</u> ถ้าตัวเลือก ก. เป็นคำตอบถูกต้อง ให้นักเรียนทำดังนี้

   ก. ข. ค. ง.
4. ห้ามขีดเขียน หรือทดในกระดาษคำถาม และกระดาษคำตอบ ให้ทดในกระดาษทดที่เตรียมไว้ให้
5. เมื่อหมดเวลาตอบ ให้ส่งทั้งแบบทดสอบ และกระดาษคำตอบ

ขอให้นักเรียนโชคดีในการสอบ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

77. ข้อความใดบ้างถูกต้อง

(1) รูป n เหลี่ยม จะคล้ายกันก็ต่อเมื่อขนาดของมุมเท่ากันเป็นคู่ๆ ทุกคู่

(2) รูป n เหลี่ยม จะคล้ายกันก็ต่อเมื่ออัตราส่วนของความยาวของด้านคู่ที่สมนัยกันทุกคู่เป็นอัตราส่วนที่เท่ากัน

ก. ข้อ 1 เท่านั้น ข. ข้อ 2 เท่านั้น

ค. ทั้งข้อ 1 และข้อ 2 ง. ไม่ใช่ทั้งข้อ 1 และ ข้อ 2

78. ข้อสรุปใดต่อไปนี้**ไม่**ถูกต้อง

ก. ถ้ารูปห้าเหลี่ยม 2 รูปคล้ายกัน แสดงว่า รูปห้าเหลี่ยม 2 รูปนั้นมีมุมเท่ากัน 5 มุม มุมต่อมุม

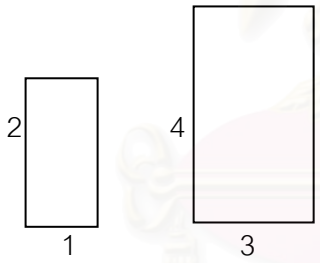
ข. ถ้ารูปห้าเหลี่ยม 2 รูปมีมุมเท่ากัน 5 มุม มุมต่อมุม แสดงว่ารูปห้าเหลี่ยม 2 รูปนั้นคล้ายกัน

ค. ถ้ารูปห้าเหลี่ยม 2 รูปคล้ายกัน แสดงว่ารูปห้าเหลี่ยม 2 รูปนั้นมีอัตราส่วนของความยาวของด้านคู่ที่สมนัยกันจะเท่ากันทุกอัตราส่วน

ง. ถ้ารูปห้าเหลี่ยม 2 รูปนั้นคล้ายกัน แสดงว่ารูปห้าเหลี่ยม 2 รูปมีมุมเท่ากัน 5 มุม มุมต่อมุม และมีความยาวด้านเท่ากัน 5 ด้าน ด้านต่อด้าน

79. รูปเหลี่ยมในข้อใดคล้ายกัน

ก. รูปสี่เหลี่ยมมุมฉาก

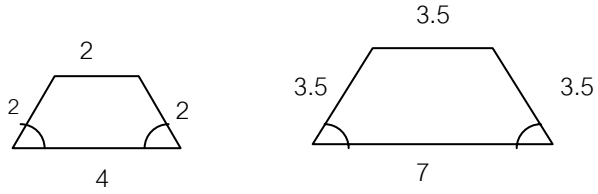


ข. รูปสี่เหลี่ยมด้านขนาน รูปที่ 1 มีมุมยอดเท่ากับ 60 องศา และ 120 องศา และด้านเท่ากับ $\frac{3}{2}$ และ $\frac{4}{5}$ หน่วย รูปสี่เหลี่ยมด้านขนานรูปที่ 2 มีมุมยอดเท่ากับ 60 องศา และ 120 องศา และด้านเท่ากับ 0.4 และ 0.75 หน่วย

ค. รูปสามเหลี่ยมที่มีด้านเท่ากับ 8 , 8 และ 2 หน่วย

รูปสามเหลี่ยมที่มีด้านเท่ากับ 3 , 3 และ 1 หน่วย

ง. รูปสี่เหลี่ยมคางหมูหน้าจั่ว



80. ข้อใดบ้างถูกต้อง

1) รูปสี่เหลี่ยมขนมเปียกปูนสองรูปคล้ายกันเสมอ

2) รูปสี่เหลี่ยมจัตุรัสสองรูปคล้ายกันเสมอ

ก. ข้อ 1) เท่านั้น ข. ข้อ 2) เท่านั้น

ค. ทั้งข้อ 1) และข้อ 2) ง. ไม่ใช่ทั้งข้อ 1) และข้อ 2)

81. ข้อใดต่อไปนี้**ไม่เป็น**รูปที่คล้ายกัน

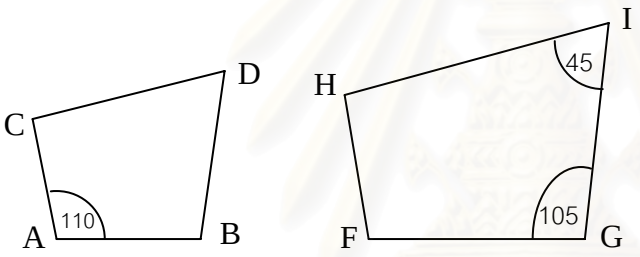
ก. รูปวงกลมสองรูป

ข. รูปสี่เหลี่ยมจัตุรัสสองรูป

ค. รูปสามเหลี่ยมด้านเท่าสองรูป

ง. รูปสามเหลี่ยมหน้าจั่วสองรูป

82. จากรูปกำหนดให้รูป ABCD คล้ายกับ FGHI จงหาขนาดของมุม C และ F

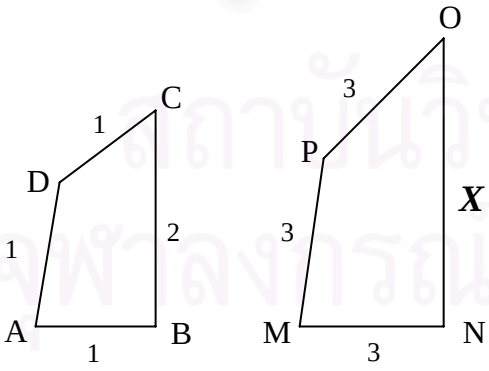


ก. $\hat{C} = 105^\circ$ และ $\hat{F} = 110^\circ$

ข. $\hat{C} = 110^\circ$ และ $\hat{F} = 105^\circ$

ค. $\hat{C} = 45^\circ$ และ $\hat{F} = 105^\circ$

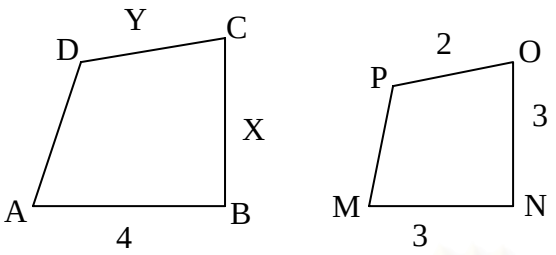
ง. $\hat{C} = 110^\circ$ และ $\hat{F} = 110^\circ$



83. จากรูป ถ้า □ ABCD คล้ายกับ □ MNOP แล้ว X เท่ากับเท่าใด

ก. 2 ข. 3 ค. 6 ง. 9

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

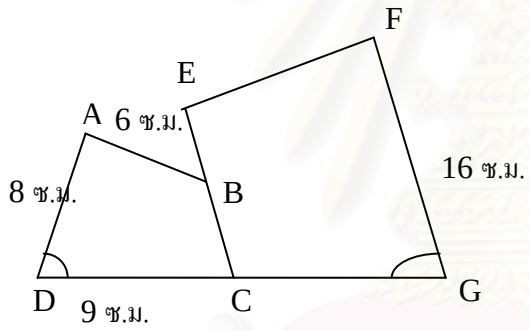


84. จากรูปข้างต้น ถ้า □ ABCD คล้ายกับ □ MNOP แล้ว X+Y เท่ากับเท่าใด

ก. $\frac{15}{3}$ ข. $\frac{21}{3}$ ค. $\frac{22}{3}$ ง. $\frac{25}{3}$

85. จากรูป □ ABCD คล้ายกับ □ CEFG เป็นรูปสี่เหลี่ยมใดๆ

△ ABD ~ △ FEG , $D\hat{A}C = C\hat{F}G$, $B\hat{A}D = G\hat{F}E = 90^\circ$ $B\hat{D}C = E\hat{G}C$ $\overline{AD}$, $\overline{DC}$ และ $\overline{FG}$ ยาว 8 , 9 และ 16 เซนติเมตร ตามลำดับ ดังนั้น $\overline{CG}$ ยาวเท่าไร



ก. 15 เซนติเมตร ข. 18 เซนติเมตร

ค. 20 เซนติเมตร ง. 27 เซนติเมตร

86. ข้อสรุปใดต่อไปนี้**ไม่**ถูกต้อง

ก. ถ้าสามเหลี่ยม 2 รูปคล้ายกัน แสดงว่า สามเหลี่ยม 2 รูปนั้นมีมุมเท่ากัน 3 มุม มุมต่อมุม

ข. ถ้าสามเหลี่ยม 2 รูปมีมุมเท่ากัน 3 มุม มุมต่อมุม แสดงว่าสามเหลี่ยม 2 รูปนั้นคล้ายกัน

ค. ถ้าสามเหลี่ยม 2 รูปคล้ายกัน แสดงว่าสามเหลี่ยม 2 รูปนั้นเท่ากันทุกประการ

ง. ถ้าสามเหลี่ยม 2 รูปนั้นเท่ากันทุกประการ แสดงว่าสามเหลี่ยม 2 รูปนั้นคล้ายกัน

87. ข้อความใดบ้างถูกต้อง

1) รูปสามเหลี่ยมสองรูปจะคล้ายกันก็ต่อเมื่อมีขนาดของมุมเท่ากันทั้งสามมุม

2) รูปสามเหลี่ยมสองรูปจะคล้ายกันก็ต่อเมื่ออัตราส่วนของความยาวของด้านคู่ที่สมนัยกันทุกคู่ของรูปสามเหลี่ยมสองรูปเป็นอัตราส่วนที่เท่ากัน

ก. ข้อ 1) เท่านั้น

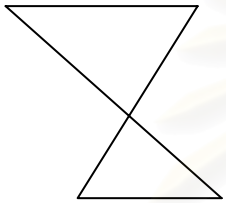
ข. ข้อ 2) เท่านั้น

ค. ทั้งข้อ 1) และข้อ 2)

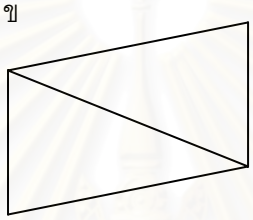
ง. ไม่ใช่ทั้งข้อ 1) และข้อ 2)

88. รูปสามเหลี่ยมคู่ใด**ไม่**คล้ายกัน

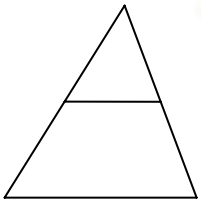
ก.



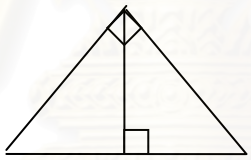
ข



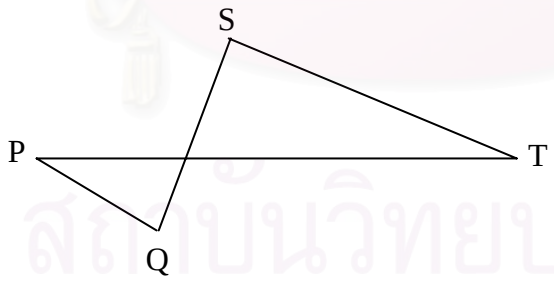
ค



ง



89. จากรูป กำหนดให้ $\overline{PQ} \parallel \overline{ST}$ ข้อใดเป็น**เท็จ**



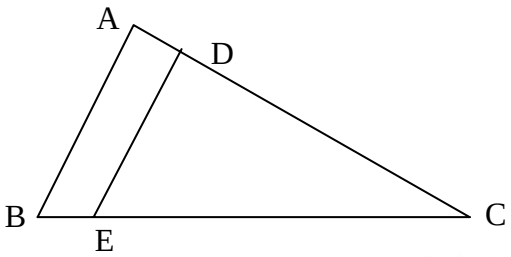
ก. $\frac{RS}{RQ} = \frac{ST}{PQ}$

ข. $R\hat{P}Q = R\hat{T}S$

ค. $P\hat{R}Q = T\hat{R}S$

ง. $R\hat{T}S = P\hat{Q}R$

90.



จากรูปที่กำหนดให้ ข้อใด**ผิด**

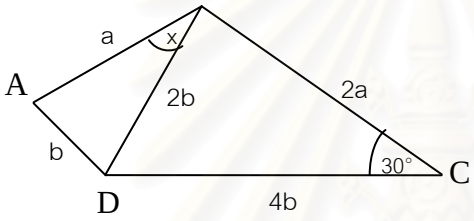
ก. $B\hat{A}C = E\hat{D}C$

ข. $\triangle ABC \sim \triangle DEC$

ค. $\triangle ABC$ ไม่คล้ายกับ $\triangle DEC$

ง. $\frac{AB}{DE} = \frac{BC}{EC}$

91. ให้หาค่าของมุม X จากรูปข้างล่างนี้



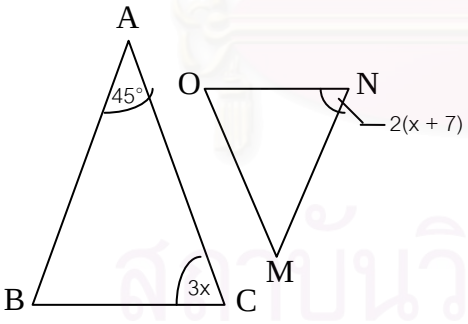
ก. 20 องศา

ข. 30 องศา

ค. 40 องศา

ง. 50 องศา

92. กำหนดให้$\triangle$ABC ~$\triangle$MNO จงหาค่าของ X



ก. 24

ข. 24.2

ค. 24.5

ง. 25

93. จากข้อ 92 จงหาค่าของ $A\hat{B}C$

ก. 60 องศา

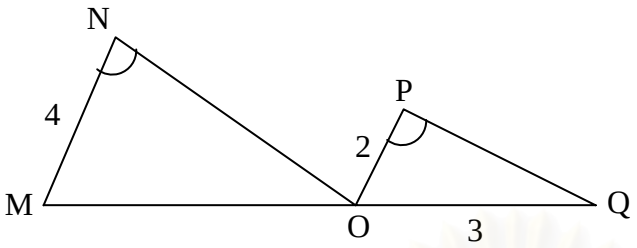
ข. 61.6 องศา

ค. 62.4 องศา

ง. 65 องศา

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

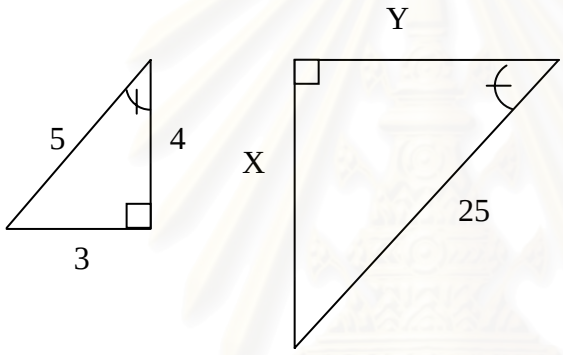
94. จงหาความยาวของ $\overline{MO}$ เท่ากับเท่าใด



ก. 5 ข. 6

ค. 7 ง. 8

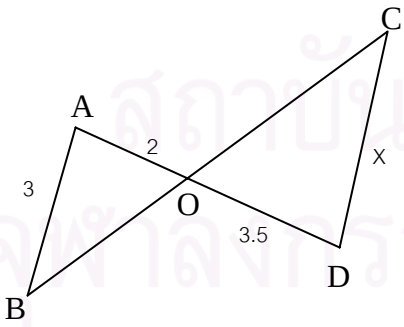
95. จากรูปข้างล่าง ค่าของ X - Y ตรงกับข้อใด



ก. 5 ข. -5

ค. 15 ง. -15

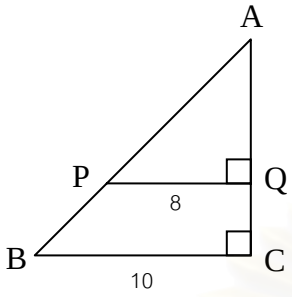
96. จากรูป ข้างล่าง △BOA ~ △DOC จงหาค่าของ X



ก. 5 ข. 5.25

ค. 5.5 ง. 6

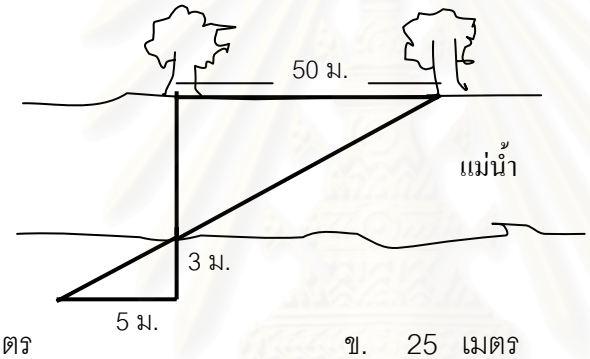
97. จากรูปข้างล่าง ถ้า $A\hat{Q}P = A\hat{C}B = 90°$, $BC = 10$, $PQ = 8$ ถ้า $\triangle APQ$ มีพื้นที่ 32 ตารางหน่วย จงหาระยะ QC



ก. 1 หน่วย ข. 2 หน่วย

ค. 3 หน่วย ง. 4 หน่วย

98. จงหาความกว้างของแม่น้ำ



ก. 20 เมตร ข. 25 เมตร

ค. 30 เมตร ง. 35 เมตร

99. ชายคนหนึ่งสูง 1.6 เมตร ยืนห่างจากเสาไฟฟ้า 5.5 เมตร ถ้าเสาไฟฟ้าสูง 6 เมตร เงาของเขาที่เกิดจากแสงไฟฟ้าบนยอดเสาจะยาวเท่าไร

ก. 1.6 เมตร ข. 1.8 เมตร

ค. 2 เมตร ง. 2.2 เมตร

100. ต้นไม้ต้นหนึ่งสูง 8 เมตร ทอดเงายาว 6 เมตร ถ้านำเสาต้นหนึ่งซึ่งสูง 12 เมตร จะต้องวางไว้ห่างจากต้นไม้กี่เมตร จึงจะทำให้ปลายเงาของเสาและต้นไม้อยู่จุดเดียวกัน

ก. 3 เมตร ข. 4 เมตร

ค. 6 เมตร ง. 7.5 เมตร

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ประวัติผู้เขียนวิทยานิพนธ์

นางสุรีพร (ศิริมาลย์) อนุศาสนนันท์ เกิดเมื่อวันที่ 29 เมษายน 2515 สำเร็จการศึกษาปริญญาครุศาสตรบัณฑิต วิชาเอกวิทยาศาสตร์-คณิตศาสตร์ จากคณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย เมื่อปีการศึกษา 2536 สำเร็จการศึกษาปริญญาครุศาสตรมหาบัณฑิต สาขาวิชาการวัดและประเมินผลการศึกษา จากคณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2539 และเข้าศึกษาต่อหลักสูตรครุศาสตรดุษฎีบัณฑิต คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย สาขาวิชาการวัดและประเมินผลการศึกษา เมื่อปีการศึกษา 2546 ปัจจุบันเป็นอาจารย์ประจำภาควิชาวิจัยและวัดผลการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยบูรพา จังหวัดชลบุรี

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย