การพัฒนาวิธีการกำหนดคะแนนจุดตัดสำหรับการทดสอบทางการศึกษาระดับชาติ ขั้นพื้นฐานวิชาคณิตศาสตร์ และวิทยาศาสตร์ ของนักเรียนมัธยมศึกษาปีที่ 6

นางศิริพันธ์ ติยะวงศ์สุวรรณ

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาครุศาสตรดุษฎีบัณฑิต
สาขาวิชาการวัดและประเมินผลการศึกษา ภาควิชาวิจัยและจิตวิทยาการศึกษา
คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2554

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

บทคัดย่อและแฟ้มข้อมูลฉบับเต็มของวิทยานิพนธ์ตั้งแต่ปีการศึกษา 2554 ที่เก็บบริการในคลังปัญญาจุฬาฯ (CUIR)
เป็นแฟ้มข้อมูลของนิสิตเจ้าของวิทยานิพนธ์ที่ส่งผ่านทางบัณฑิตวิทยาลัย
The abstract and full text of theses from the academic year 2011 in Chulalongkorn University Intellectual Repository (CUIR)
are the thesis authors' files submitted through the Graduate School.

DEVELOPMENT OF CUTSCORES SETTING PROCEDURE ON ORDINARY NATIONAL EDUCATIONAL TEST IN MATHEMATICS AND SCIENCE FOR TWELFTH GRADE STUDENTS

Mrs. Siripan Tiyawongsuwan

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 2011

Copyright of Chulalongkorn University

หัวข้อวิทยานิพนธ์	การพัฒนาวิธีการกำหนดคะแนนจุดตัดสำหรับการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐานวิชาคณิตศาสตร์และวิชาวิทยาศาสตร์ ของนักเรียนมัธยมศึกษาปีที่ 6
โดย	นางศิริพันธ์ ติยะวงศ์สุวรรณ
สาขาวิชา	การวัดและประเมินผลการศึกษา
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	ผู้ช่วยศาสตราจารย์ ดร. ณัฏฐภรณ์ หลาวทอง
อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม	ศาสตราจารย์ ดร. ศิริชัย กาญจนวาสี

คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาดุษฎีบัณฑิต

.....คณบดีคณะครุศาสตร์
(ศาสตราจารย์ ดร. ศิริชัย กาญจนวาสี)

คณะกรรมการสอบวิทยานิพนธ์

.....ประธานกรรมการ
(รองศาสตราจารย์ ดร. ศิริเดช สุชีวะ)

.....อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
(ผู้ช่วยศาสตราจารย์ ดร. ณัฏฐภรณ์ หลาวทอง)

.....อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม
(ศาสตราจารย์ ดร. ศิริชัย กาญจนวาสี)

.....กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร. กมลวรรณ ตังธนกานนท์)

.....กรรมการ
(รองศาสตราจารย์ ดร. โชติกา ภาษีผล)

.....กรรมการภายนอกมหาวิทยาลัย
(ดร. สังวรณ์ งัดกระโทก)

ศิริพันธ์ ติยะวงศ์สุวรรณ: การพัฒนาวิธีการกำหนดคะแนนจุดตัดสำหรับการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐานวิชาคณิตศาสตร์ และวิชาวิทยาศาสตร์ ของนักเรียนมัธยมศึกษาปีที่ 6. (Development of Cutscores Setting Procedure on Ordinary National Educational Test in Mathematics and Science for Twelfth Grade Students) อ.ที่ปรึกษาวิทยานิพนธ์หลัก: ผศ. ดร.ณัฏฐภรณ์ หลาวทอง, อ. ที่ปรึกษาวิทยานิพนธ์ร่วม: ศ. ดร.ศิริชัย กาญจนวาสี, 356 หน้า

การวิจัยครั้งนี้มีวัตถุประสงค์ 1) เพื่อพัฒนาวิธีการกำหนดคะแนนจุดตัดสำหรับการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐานวิชาคณิตศาสตร์ และวิชาวิทยาศาสตร์ ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 2) เพื่อกำหนดคะแนนจุดตัดสำหรับการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐานวิชาคณิตศาสตร์ และวิชาวิทยาศาสตร์ ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ด้วยวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซี 3) เพื่อตรวจสอบคุณภาพของคะแนนจุดตัด และ4) เพื่อวิเคราะห์จำนวนผู้เชี่ยวชาญและจำนวนกลุ่มผู้เชี่ยวชาญที่มีความเหมาะสมเมื่อกำหนดคะแนนจุดตัดโดยวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซี กลุ่มตัวอย่างประกอบด้วยครูผู้สอนวิชาคณิตศาสตร์จำนวน 27 คน และเป็นครูผู้สอนวิชาฟิสิกส์ 10 คน วิชาเคมี 10 คน และวิชาชีววิทยา 10 คน ข้อสอบที่ใช้ในการกำหนดคะแนนจุดตัดเป็นข้อสอบสำหรับการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน ช่วงชั้นที่ 4 (มัธยมศึกษาปีที่ 6) ปีการศึกษา 2553 วิชาคณิตศาสตร์จำนวน 40 ข้อ คะแนนเต็ม 100 คะแนน และวิชาวิทยาศาสตร์ จำนวน 90 ข้อ คะแนนเต็ม 100 คะแนน

ผลการวิจัยพบว่า

1. วิธีกำหนดคะแนนจุดตัดที่พัฒนาขึ้น คือวิธีเอบีซี เป็นวิธีที่ให้ผู้เชี่ยวชาญกำหนดคะแนนจุดตัดจากคู่มือเรียงข้อสอบรายสาระการเรียนรู้ และเสนอสารสนเทศเกี่ยวกับผลการวิเคราะห์กลุ่มประกอบการกำหนดคะแนนจุดตัด เป้าหมายในการพัฒนาวิธีนี้เพื่อลดความซับซ้อนของกระบวนการคิดของผู้เชี่ยวชาญ และเพื่อให้ได้วิธีที่ง่ายต่อการดำเนินการ อันจะนำไปสู่ดำเนินการที่มีประสิทธิภาพ

2. คะแนนจุดตัดวิชาคณิตศาสตร์ระดับ 1/2 ระดับ 2/3 และระดับ 3/4 ที่กำหนดด้วยวิธีแองกอฟแบบใช่/ไม่ใช่ มีค่าเท่ากับ 15.11, 42.63 และ 81.26 ตามลำดับ วิธีบุ๊คมาร์ค มีค่า 39.22, 55.37 และ 70.26 ตามลำดับ และวิธีเอบีซี มีค่า 6.34, 50.74 และ 80.37 ตามลำดับ คะแนนจุดตัดวิชาวิทยาศาสตร์ระดับ 1/2 ระดับ 2/3 และระดับ 3/4 ที่กำหนดด้วยวิธีแองกอฟแบบใช่/ไม่ใช่ มีค่าเท่ากับ 44.17, 72.44 และ88.52 ตามลำดับ วิธีบุ๊คมาร์ค มีค่า 23.52, 53.48 และ 76.93 ตามลำดับ และวิธีเอบีซี มีค่า 21.96, 64.85 และ 85.67 ตามลำดับ

3. ผลวิเคราะห์คุณภาพด้านความตรง พบว่า การกำหนดคะแนนจุดตัดด้วยวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีมีความตรงของปัจจัยนำเข้าโดยภาพรวมอยู่ในระดับมาก สำหรับหลักการที่ผู้เชี่ยวชาญใช้ในการพิจารณาเพื่อกำหนดคะแนนจุดตัด คือ พิจารณาระดับพฤติกรรมด้านพุทธิพิสัยที่ข้อสอบต้องการวัด พิจารณารูปแบบของข้อสอบ และขั้นตอนของการได้มาซึ่งคำตอบที่ถูกต้อง คะแนนจุดตัดที่กำหนดด้วยวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซี มีความสอดคล้องกันน้อย คะแนนจุดตัดที่กำหนดด้วยวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีมีความถูกต้องของการกำหนดกลุ่มสูง มีความคลาดเคลื่อนของการกำหนดกลุ่มแบบบวก และความคลาดเคลื่อนของการกำหนดกลุ่มแบบลบต่ำ สำหรับผลวิเคราะห์คุณภาพด้านความเที่ยง พบว่า วิธีแองกอฟแบบใช่/ไม่ใช่มีความเที่ยงภายในผู้เชี่ยวชาญ 0.36 ถึง 0.97 วิธีบุ๊คมาร์คมีความเที่ยงภายในผู้เชี่ยวชาญ 0.91 ถึง 0.94 วิธีเอบีซีมีความเที่ยงภายในผู้เชี่ยวชาญ 0.33 ถึง 0.90 วิธีแองกอฟแบบใช่/ไม่ใช่มีความเที่ยงระหว่างผู้เชี่ยวชาญ 0.32 ถึง 0.93 วิธีบุ๊คมาร์คมีความเที่ยงระหว่างผู้เชี่ยวชาญ 0.19 ถึง 0.79 วิธีเอบีซีมีความเที่ยงระหว่างผู้เชี่ยวชาญ 0.47 ถึง 0.87

4. จำนวนผู้เชี่ยวชาญที่เหมาะสมสำหรับการกำหนดคะแนนจุดตัดด้วยวิธีแองกอฟแบบใช่ไม่ใช่ควรมีจำนวน 36 คนขึ้นไป โดยแบ่งเป็น 6 กลุ่ม กลุ่มละ 6 คน วิธีบุ๊คมาร์คควรมีจำนวน 49 คนขึ้นไป โดยแบ่งเป็น 7 กลุ่ม กลุ่มละ 7 คน สำหรับวิธีเอบีซีควรมีจำนวน 25 คนขึ้นไป โดยแบ่งเป็น 5 กลุ่ม กลุ่มละ 5 คน

ภาควิชา วิจัยและจิตวิทยาการศึกษา	ลายมือชื่อนิสิต.....
สาขาวิชา การวัดและประเมินผลการศึกษา	ลายมือชื่อ อ. ที่ปรึกษาวิทยานิพนธ์หลัก.....
ปีการศึกษา 2554	ลายมือชื่อ อ. ที่ปรึกษาวิทยานิพนธ์ร่วม.....

5184254427: MAJOR EDUCATIONAL MEASURMENT AND EVALUATION

KEYWORDS: CUTSCORES/ BOOKMARK METHOD/ YES/NO ANGOFF METHOD/ ABC METHOD

SIRIPAN TIYAWONGSUWAN: DEVELOPMENT OF CUTSCORES SETTING PROCEDURE ON ORDINARY NATIONAL EDUCATIONAL TEST IN MATHEMATICS AND SCIENCE FOR TWELFTH GRADE STUDENTS. ADVISOR: ASST. PROF. NUTTAPORN LAWTHONG, Ph.D., CO-ADVISOR: PROF. SIRICHAI KANJANAWASEE, Ph.D., 356 pp.

The objectives of this study were 1) to develop an alternative cutscores setting procedure 2) to set cutscores from Yes/No Angoff method, Bookmark method, and ABC method on Ordinary National Educational Test on twelfth grade in mathematics and sciences 2010 3) to investigate the quality of cutscores and 4) to estimate the appropriate of the number of panelists and the number of table groups for the three methods. The research sample comprised of 27 mathematics teachers, ten physics teachers, ten chemistry teachers and ten biology teachers. The test consisted of 40 items in mathematics and 90 items in sciences. The total score on each test was 100.

The research findings:

1. This study proposed the ABC method, which involved creating separate ordered-item booklets for each strand and providing clustering result. The purpose of this method was to reduce the cognitive complexity of the judgment task and to derive methods that were more intuitive for the panelists, which could be carried out efficiently.

2. For mathematics, the mean cutscores for 1/2, 2/3 and 3/4 levels from the Yes/No Angoff method were 15.11, 42.63, and 81.26, respectively. The mean cutscores for 1/2, 2/3 and 3/4 levels from the Bookmark method were 39.22, 55.37 and 70.26, respectively. The mean cutscores for 1/2, 2/3 and 3/4 levels from the ABC method were 6.34, 50.74, and 80.37, respectively. For sciences, the mean cutscores for 1/2, 2/3 and 3/4 levels from the Yes/No Angoff method were 44.17, 72.44, and 88.52, respectively. The mean cutscores for 1/2, 2/3 and 3/4 levels from the Bookmark method were 23.52, 53.48 and 76.93, respectively. The mean cutscores for 1/2, 2/3 and 3/4 levels from the ABC method were 21.96, 64.85, and 85.67, respectively.

3. The overall validities of standard setting inputs for the three methods were high. The panelists made a decision about cutscores by focusing on the types of cognitive domain, item formats and the step to correct answer. The level of agreement in determining performance level group of cutscores for the three methods was low. The classification accuracy of cutscores for the three methods was high. False positive and false negative error rates were low. In the Intra-rater reliability, intraclass correlation coefficients ranged from 0.36 to 0.97 for Yes/No Angoff method, from 0.91 to 0.94 for Bookmark method and from 0.33 to 0.90 for ABC method. In the Inter-rater reliability, intraclass correlation coefficients ranged from 0.32 to 0.93 for Yes/No Angoff method, from 0.19 to 0.79 for Bookmark method and from 0.47 to 0.87 for ABC method.

4. For Yes/No Angoff method, the recommendations for the numbers of panelists were 36, the numbers of table groups were 6, and the numbers of panelists per table groups were 6. For Bookmark method, the recommendations for the numbers of panelists were 49, the numbers of table groups were 7, and the numbers of panelists per table groups were 7. For ABC method, the recommendations for the numbers of panelists were 25, the numbers of table groups were 5, and the numbers of panelists per table groups were 5.

Department Educational Research and Psychology Student's Signature.....

Field of Study Educational Measurment and Evaluation Advisor's Signature.....

Academic Year 2011 Co-advisor's Signature.....

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ด้วยความกรุณาจากผู้ช่วยศาสตราจารย์ ดร. ณัฏฐภรณ์ หลาวทอง อาจารย์ที่ปรึกษาวิทยานิพนธ์ และศาสตราจารย์ ดร. ศิริชัย กาญจนวาสี อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม ที่ให้คำปรึกษาและข้อเสนอแนะที่เป็นประโยชน์ต่อการศึกษาวิจัยอย่างดีตลอดมา ผู้วิจัยขอกราบขอบพระคุณอย่างสูงมา ณ โอกาสนี้

ขอกราบขอบพระคุณรองศาสตราจารย์ ดร. ศิริเดช สุชีวะประธานกรรมการสอบวิทยานิพนธ์ ดร. สังวรณ์ งัดกระโทก ผู้ทรงคุณวุฒิภายนอก ผู้ช่วยศาสตราจารย์ ดร. กมลวรรณ ตังธนกานนท์ และรองศาสตราจารย์ ดร. โชติกา ภาษีผล ผู้ทรงคุณวุฒิ ที่ช่วยให้ข้อเสนอแนะที่เป็นประโยชน์ต่อการปรับปรุงแก้ไขวิทยานิพนธ์ให้สมบูรณ์ยิ่งขึ้น รวมทั้งคณาจารย์ภาควิชาวิจัยและจิตวิทยาการศึกษาที่ได้ประสิทธ์ประสาทวิชาความรู้ต่าง ๆ แก่ผู้วิจัย และขอขอบคุณบัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัยที่ให้ทุน 90 ปี จุฬาลงกรณ์มหาวิทยาลัย ในการสนับสนุนทุนทำวิทยานิพนธ์ ขอบคุณมหาวิทยาลัยราชภัฏนครราชสีมาที่ให้ทุนสนับสนุนในการศึกษาต่อในระดับดุษฎีบัณฑิต ขอบคุณสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) ที่ให้ข้อมูลสำหรับการศึกษาวิจัยครั้งนี้ และขอขอบคุณ Michael J. Zieky (Educational Testing Service), Barbara S. Plake (University of Nebraska-Lincoln) สำหรับหนังสือที่มอบให้และคำแนะนำต่างๆ

ขอขอบคุณ ดร.วดาภรณ์ พูลผลอำนวย และ คุณกฤตพล ชุมสกุล (ศึกษานิเทศก์สำนักงานเขตพื้นที่การศึกษามัธยมศึกษา เขต 31) คุณกรวิกา ฉินนานนท์ และคุณสุกัญญา ทองนาค ที่ให้ความช่วยเหลือในการประสานงานและเก็บรวบรวมข้อมูล ขอบคุณเพื่อน พี่ และน้อง สาขาการวัดและประเมินผลการศึกษาทุกท่านที่ให้ความช่วยเหลือผู้วิจัยตลอดมา

ท้ายนี้ขอกราบขอบพระคุณคุณพ่อ คุณแม่ และพี่ชายที่ให้การสนับสนุนและเป็นกำลังใจแก่ผู้วิจัยตลอดมา ขอขอบคุณคุณสฤษดิ์ ติยะวงศ์สุวรรณ ที่คอยช่วยเหลือและสนับสนุนขณะที่ผู้วิจัยเรียนและทำวิทยานิพนธ์

สารบัญ

		หน้า
บทคัดย่อภาษาไทย		ง
บทคัดย่อภาษาอังกฤษ		จ
กิตติกรรมประกาศ		ฉ
สารบัญ		ช
สารบัญตาราง		ฌ
สารบัญภาพ		ฎ
บทที่		
1	บทนำ	1
	ความเป็นมาและความสำคัญของปัญหา	1
	คำถามวิจัย	13
	วัตถุประสงค์ของการวิจัย	13
	ขอบเขตของการวิจัย	14
	คำจำกัดความเชิงปฏิบัติการ	15
	ประโยชน์ที่คาดว่าจะได้รับ	20
2	เอกสารและงานวิจัยที่เกี่ยวข้อง	21
	ตอนที่ 1 การกำหนดคะแนนจุดตัด	22
	ความเป็นมาของการกำหนดคะแนนจุดตัด	22
	ความหมายของคะแนนจุดตัด	26
	ความเข้าใจเบื้องต้นเกี่ยวกับคะแนนจุดตัด	27
	ขั้นตอนสำคัญของการกำหนดคะแนนจุดตัด	29
	วิธีกำหนดคะแนนจุดตัด	42
	การตรวจสอบคุณภาพของคะแนนจุดตัด	62
	งานวิจัยที่เกี่ยวข้อง	68
	ตอนที่ 2 ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด	89
	ตอนที่ 3 ทฤษฎีการตอบสนองข้อสอบ	94
	ตอนที่ 4 การทดสอบทางการศึกษาแห่งชาติขั้นพื้นฐาน	100

บทที่		หน้า
3	วิธีดำเนินการวิจัย	110
	ขั้นตอนการดำเนินการวิจัย	110
	ขอบเขตของการวิจัย	112
	เครื่องมือที่ใช้ในการวิจัย	115
	การดำเนินการและการเก็บรวบรวมข้อมูล	120
	วิธีการกำหนดคะแนนจุดตัด	125
	การวิเคราะห์ข้อมูล	131
4	ผลการวิเคราะห์ข้อมูล	137
	ตอนที่ 1 ผลการกำหนดจำนวนของระดับ ชื่อของระดับและคำอธิบายระดับความสามารถจุดตัด	138
	ตอนที่ 2 ผลการพัฒนาวิธีการกำหนดคะแนนจุดตัด	142
	ตอนที่ 3 ผลการกำหนดคะแนนจุดตัด	149
	ตอนที่ 4 ผลการวิเคราะห์คุณภาพของคะแนนจุดตัด	170
5	สรุปผล อภิปรายและข้อเสนอแนะ	236
	สรุปผลการวิจัย	237
	อภิปรายผลการวิจัย	244
	ข้อเสนอแนะ	251
รายการอ้างอิง		255
ภาคผนวก		264
	ภาคผนวก ก	265
	ภาคผนวก ข	270
	ภาคผนวก ค	276
	ภาคผนวก ง	282
	ภาคผนวก จ	311
	ภาคผนวก ฉ	319
	ภาคผนวก ช	333
	ภาคผนวก ซ	340
	ภาคผนวก ฌ	351
ประวัติผู้เขียนวิทยานิพนธ์		356

สารบัญตาราง

ตาราง		หน้า
2.1	ตัวอย่างจำนวนระดับ ชื่อของระดับ และคำอธิบายระดับความสามารถ.....	41
2.2	ตัวอย่างการคำนวณคะแนนจุดตัดที่กำหนดโดยวิธีแองกอฟ.....	43
2.3	สรุปรายละเอียดเกี่ยวกับวิธีการกำหนดคะแนนจุดตัด.....	59
2.4	สรุปการตรวจสอบคุณภาพของคะแนนจุดตัด.....	67
2.5	การคำนวณคะแนนจุดตัดโดยวิธีแองกอฟแบบใช่/ไม่ใช่.....	70
2.6	การปรับปรุงวิธีแองกอฟ ข้อดีและข้อเสีย.....	72
2.7	สรุปการศึกษาวิจัยเกี่ยวกับการกำหนดคะแนนจุดตัดโดยวิธีแองกอฟ.....	76
2.8	สรุปการศึกษาวิจัยเกี่ยวกับการกำหนดคะแนนจุดตัดโดยวิธีบุ๊คมาร์ค.....	84
2.9	สรุปการศึกษาวิจัยเกี่ยวกับการกำหนดคะแนนจุดตัดโดยการวิเคราะห์กลุ่ม.....	88
3.1	แสดงจำนวนกลุ่มตัวอย่างที่เข้าร่วมการกำหนดคะแนนจุดตัด.....	113
3.2	รายละเอียดของข้อสอบสำหรับการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน ช่วงชั้นที่ 4 (มัธยมศึกษาปีที่ 6) ปีการศึกษา 2553 วิชาคณิตศาสตร์.....	114
3.3	รายละเอียดของข้อสอบสำหรับการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน ช่วงชั้นที่ 4 (มัธยมศึกษาปีที่ 6) ปีการศึกษา 2553 วิชาวิทยาศาสตร์.....	114
3.4	สรุปชื่อ วัตถุประสงค์ ลักษณะการวัด ประเด็นหลักในการเก็บข้อมูล.....	115
3.5	การเก็บข้อมูลเพื่อตรวจสอบความตรงของกระบวนการตัดสินใจเพื่อ กำหนดคะแนนจุดตัดของผู้เชี่ยวชาญ.....	124
4.1	จำนวนของระดับ ชื่อของระดับ และคำอธิบายระดับความสามารถ วิชาณิตศาสตร์.....	140
4.2	จำนวนของระดับ ชื่อของระดับ และคำอธิบายระดับความสามารถ วิชาวิทยาศาสตร์.....	141
4.3	แสดงผลการวิเคราะห์ข้อมูลเบื้องต้นของข้อสอบวิชาคณิตศาสตร์และ วิทยาศาสตร์.....	149
4.4	ผลการกำหนดคะแนนจุดตัดวิชาคณิตศาสตร์ด้วยวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซี.....	151
4.5	ผลการกำหนดคะแนนจุดตัดวิชาคณิตศาสตร์ด้วยวิธีแองกอฟแบบใช่/ไม่ใช่.....	154
4.6	ผลการกำหนดคะแนนจุดตัดวิชาคณิตศาสตร์ด้วยวิธีบุ๊คมาร์ค.....	156

ตาราง		หน้า
4.7	ผลการกำหนดคะแนนจุดตัดวิชาคณิตศาสตร์ด้วยวิธีเอบีซี.....	158
4.8	ผลการกำหนดคะแนนจุดตัดวิชาวิทยาศาสตร์ด้วยวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซี.....	160
4.9	ผลการกำหนดคะแนนจุดตัดวิชาคณิตศาสตร์ด้วยวิธีแองกอฟแบบใช่/ไม่ใช่....	163
4.10	ผลการกำหนดคะแนนจุดตัดวิชาคณิตศาสตร์ด้วยวิธีบุ๊คมาร์ค.....	165
4.11	ผลการกำหนดคะแนนจุดตัดวิชาวิทยาศาสตร์ด้วยวิธีเอบีซี.....	167
4.12	ผลการวิเคราะห์ความตรงของปัจจัยนำเข้าของการกำหนดคะแนนจุดตัดวิชาคณิตศาสตร์.....	171
4.13	ผลการวิเคราะห์ข้อมูลเพื่อตรวจสอบความตรงของปัจจัยนำเข้าของการกำหนดคะแนนจุดตัดวิชาวิทยาศาสตร์.....	179
4.14	ผลการวิเคราะห์วิชาและวิธีกำหนดคะแนนจุดตัดที่มีผลต่อความตรงของปัจจัยนำเข้าของการกำหนดคะแนนจุดตัด.....	187
4.15	ผลการวิเคราะห์ความสอดคล้องของคะแนนจุดตัดที่กำหนดโดยวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซี วิชาคณิตศาสตร์.....	200
4.16	ผลการวิเคราะห์ความสอดคล้องของคะแนนจุดตัดที่กำหนดโดยวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซี วิชาวิทยาศาสตร์.....	201
4.17	ผลการวิเคราะห์ความถูกต้อง ความคลาดเคลื่อนแบบบวก และความคลาดเคลื่อนแบบลบของคะแนนจุดตัดในการกำหนดกลุ่มผู้สอบตามระดับความสามารถวิชาคณิตศาสตร์.....	202
4.18	ผลการวิเคราะห์ความถูกต้อง ความคลาดเคลื่อนแบบบวก และความคลาดเคลื่อนแบบลบของคะแนนจุดตัดในการกำหนดกลุ่มผู้สอบตามระดับความสามารถวิชาวิทยาศาสตร์.....	205
4.19	ผลการศึกษาขนาดขององค์ประกอบความแปรปรวนจากแหล่งความคลาดเคลื่อนต่าง ๆ ของการกำหนดคะแนนจุดตัดด้วยวิธีแองกอฟแบบใช่/ไม่ใช่.....	208
4.20	ผลการศึกษาขนาดขององค์ประกอบความแปรปรวนจากแหล่งความคลาดเคลื่อนต่าง ๆ ของการกำหนดคะแนนจุดตัดวิชาคณิตศาสตร์ด้วยวิธีบุ๊คมาร์ค...	210
4.21	ผลการศึกษาขนาดขององค์ประกอบความแปรปรวนจากแหล่งความคลาดเคลื่อน ต่าง ๆ ของการกำหนดคะแนนจุดตัดวิชาคณิตศาสตร์ด้วยวิธีเอบีซี.....	212
	ผลการเปรียบเทียบขนาดขององค์ประกอบความแปรปรวนจากแหล่ง	

ตาราง		หน้า
4.22	ความคลาดเคลื่อนต่าง ๆ ของการกำหนดคะแนนจุดตัดวิชาคณิตศาสตร์ แต่ละวิธี……	214
4.23	ผลวิเคราะห์ค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์ ของการกำหนดคะแนนจุดตัดวิชาคณิตศาสตร์ด้วยวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซี เมื่อจำนวนผู้ตัดสิน และจำนวนกลุ่มผู้ตัดสินต่างกัน….	216
4.24	ผลการศึกษาขนาดขององค์ประกอบความแปรปรวนจากแหล่งความคลาด เคลื่อนต่าง ๆ ของการกำหนดคะแนนจุดตัดวิชาวิทยาศาสตร์ด้วยวิธีแองกอฟ แบบใช่/ไม่ใช่ ……	220
4.25	ผลการศึกษาขนาดขององค์ประกอบความแปรปรวนจากแหล่งความคลาด เคลื่อนต่าง ๆ ของการกำหนดคะแนนจุดตัดวิชาวิทยาศาสตร์ด้วยวิธีบุ๊คมาร์ค…..	222
4.26	ผลการศึกษาขนาดขององค์ประกอบความแปรปรวนจากแหล่งความคลาด เคลื่อน ต่าง ๆ ของการกำหนดคะแนนจุดตัดวิชาวิทยาศาสตร์ด้วยวิธีเอบีซี……..	224
4.27	ผลการเปรียบเทียบขนาดขององค์ประกอบความแปรปรวนจากแหล่ง ความคลาดเคลื่อนต่าง ๆ ของการกำหนดคะแนนจุดตัดวิชาวิทยาศาสตร์ แต่ละวิธี……	226
4.28	ผลวิเคราะห์ค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์ ของการกำหนดคะแนนจุดตัดวิชาวิทยาศาสตร์ด้วยวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซี เมื่อจำนวนผู้ตัดสิน และจำนวนกลุ่มผู้ตัดสินต่างกัน….	228
4.29	ผลการวิเคราะห์ความเที่ยงภายในผู้เชี่ยวชาญ และความเที่ยงระหว่าง ผู้เชี่ยวชาญ ของการกำหนดคะแนนจุดตัดวิชาคณิตศาสตร์ และวิชาวิทยาศาสตร์.	232
5.1	บริบทที่เหมาะสมกับวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซี……	252

สารบัญภาพ

ภาพที่		หน้า
1.1	บทบาทของการทดสอบทางการศึกษาและการตีความคะแนนสอบที่มีต่อการวางแผนและพัฒนาการศึกษาของต่างประเทศ.....	4
1.2	บทบาทของการทดสอบทางการศึกษาและการตีความคะแนนสอบที่มีต่อการวางแผนและพัฒนาการศึกษาของประเทศไทย.....	4
1.3	กระบวนการกำหนดคะแนนจุดตัด.....	10
2.1	บทบาทของการกำหนดคะแนนจุดตัดในกระบวนการวัดและประเมินผลการเรียนรู้.....	23
2.2	จำนวนงานวิจัยเกี่ยวกับการทดสอบแบบอิงเกณฑ์ และการกำหนดคะแนนจุดตัดที่ถูกตีพิมพ์ในวารสารทางวิชาการในแต่ละยุคของวิวัฒนาการของการกำหนดคะแนนจุดตัด.....	26
2.3	ขั้นตอนของการกำหนดคะแนนจุดตัด.....	37
2.4	การแจกแจงคะแนนของกลุ่มสูงกลุ่มต่ำและคะแนนจุดตัด.....	53
2.5	องค์ประกอบความแปรปรวนของ G-study $r \times g \times f$ design เมื่อตัวประกอบที่ศึกษาทั้งหมดเป็นตัวประกอบสุ่ม.....	91
2.6	โค้งลักษณะข้อสอบแบบ 1 พารามิเตอร์ ของข้อสอบ 3 ข้อ.....	96
2.7	โค้งลักษณะข้อสอบแบบ 2 พารามิเตอร์ ของข้อสอบ 3 ข้อ.....	97
2.8	โค้งลักษณะข้อสอบแบบ 3 พารามิเตอร์ ของข้อสอบ 3 ข้อ.....	98
2.9	กรอบแนวคิดในการวิจัย.....	109
3.1	ขั้นตอนการดำเนินการวิจัย.....	110
3.2	ตัวอย่างการคิดคะแนนจุดตัดของวิธีแองกอฟแบบใช่/ไม่ใช่.....	126
3.3	ตัวอย่างคู่มือเรียงข้อสอบ.....	128
3.4	ตัวอย่างการกำหนดคะแนนจุดตัดของวิธีบุ๊คมาร์ค.....	129
3.5	ตัวอย่างการคิดคะแนนจุดตัดของวิธีเอบีซี.....	131
4.1	ตัวอย่างผลคะแนนเฉลี่ยรายสาระการเรียนรู้ของแต่ละกลุ่ม.....	147
4.2	ตัวอย่างช่วงของคะแนนสอบที่คาบเกี่ยวกันของแต่ละกลุ่ม.....	148
4.3	ผลการกำหนดคะแนนจุดตัดวิชาคณิตศาสตร์.....	153

ภาพที่ หน้า

4.4 ผลการกำหนดคะแนนจุดตัดวิชาวิทยาศาสตร์..... 162

4.5 ผลการวิเคราะห์ความตรงของปัจจัยนำเข้าของการกำหนดคะแนนจุดตัดรายด้านวิชาคณิตศาสตร์..... 178

4.6 ผลการวิเคราะห์ความตรงของปัจจัยนำเข้าของการกำหนดคะแนนจุดตัดรายข้อวิชาคณิตศาสตร์..... 178

4.7 ผลการวิเคราะห์ความตรงของปัจจัยนำเข้าของการกำหนดคะแนนจุดตัดรายด้านวิชาวิทยาศาสตร์..... 186

4.8 ผลการวิเคราะห์ความตรงของปัจจัยนำเข้าของการกำหนดคะแนนจุดตัดรายข้อวิชาวิทยาศาสตร์..... 186

4.9 ความถูกต้อง ความคลาดเคลื่อนแบบบวก และความคลาดเคลื่อนแบบลบในการกำหนดกลุ่มผู้สอบตามระดับความสามารถของคะแนนจุดตัดวิชาคณิตศาสตร์ 204

4.10 ความถูกต้อง ความคลาดเคลื่อนแบบบวก และความคลาดเคลื่อนแบบลบ.....ในการกำหนดกลุ่มผู้สอบตามระดับความสามารถของคะแนนจุดตัดวิชาวิทยาศาสตร์..... 207

4.11 ขนาดองค์ประกอบความแปรปรวนของการกำหนดคะแนนจุดตัดวิชาคณิตศาสตร์แต่ละวิธี..... 215

4.12 ค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์ของการกำหนดคะแนนจุดตัดวิชาคณิตศาสตร์แต่ละวิธี..... 219

4.13 ขนาดองค์ประกอบความแปรปรวนของการกำหนดคะแนนจุดตัดวิชาวิทยาศาสตร์แต่ละวิธี..... 227

4.14 ค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์ของการกำหนดคะแนนจุดตัดวิชาวิทยาศาสตร์แต่ละวิธี..... 231

4.15 ความเที่ยงภายในผู้เชี่ยวชาญ และความเที่ยงระหว่างผู้เชี่ยวชาญของคะแนนจุดตัดวิชาคณิตศาสตร์..... 234

4.16 ความเที่ยงภายในผู้เชี่ยวชาญ และความเที่ยงระหว่างผู้เชี่ยวชาญของคะแนนจุดตัดวิชาวิทยาศาสตร์..... 235

บทที่ 1

บทนำ

ความเป็นมาและความสำคัญของปัญหา

ในปัจจุบันหลายประเทศนิยมประเมินผลสัมฤทธิ์ทางการศึกษาแบบอิงเกณฑ์ กล่าวคือ คะแนนจากการสอบจะถูกตีความโดยอิงกับเกณฑ์หรือมาตรฐานที่กำหนดไว้อย่างเหมาะสม (Berk, 1980; Cizek & Bunch, 2007; Zieky, Perie & Livingston, 2008) อาทิ ในประเทศสหรัฐอเมริกา ประเทศอังกฤษ และประเทศออสเตรเลีย เป็นต้น เกณฑ์หรือมาตรฐานที่กำหนดขึ้นเรียกว่า “คะแนนจุดตัด” ซึ่งเป็นคะแนนที่ใช้ในการแบ่งคะแนนเป็น 2 กลุ่มหรือมากกว่าสองกลุ่มที่สะท้อนถึงระดับความสามารถของผู้สอบที่ต่างกัน

ในประเทศสหรัฐอเมริกามีการกำหนดมาตรฐานสำหรับประเมินผลพัฒนาการทางการเรียน ซึ่งแต่ละรัฐจะต้องกำหนดมาตรฐานอย่างน้อย 3 ระดับ ได้แก่ ระดับชำนาญการบางส่วน (partially proficient) ระดับชำนาญการ (proficient) และระดับชำนาญการขั้นสูง (advanced proficient) โดยจำนวนของระดับที่นิยมกำหนดมากที่สุดคือ 4 ระดับ ประกอบด้วย ต่ำกว่าพื้นฐาน (Below Basic), พื้นฐาน (Basic), ชำนาญการ (Proficient) และระดับเชี่ยวชาญ (Advanced) คะแนนจากการสอบจะถูกตีความโดยอิงกับมาตรฐานที่กำหนดขึ้น กล่าวคือ ในรายงานผลการสอบจะนำเสนอเกี่ยวระดับความสามารถของผู้สอบ สำหรับรายละเอียดเกี่ยวกับความรู้และทักษะของผู้สอบที่อยู่ในแต่ละระดับความสามารถจะบรรยายไว้ในอธิบายระดับความสามารถ (Performance Level Descriptor: PLD) ซึ่งทำให้การตีความหมายผลการสอบเป็นไปอย่างมีมาตรฐาน (Perie, 2008) และทำให้ได้สารสนเทศเกี่ยวกับความรู้และทักษะที่ผู้สอบมีด้วย

นอกจากนี้ยังมีการสอบสำคัญจำนวนมากที่นิยมตีความคะแนนสอบโดยการเปรียบเทียบกับมาตรฐานที่กำหนดไว้ ส่งผลให้นักวิชาการหลายท่านพยายามศึกษาเพื่อกำหนดมาตรฐานที่มีความเหมาะสมและน่าเชื่อถือขึ้น เพื่อช่วยให้การตีความคะแนนมีความแม่นยำ อาทิ Gomez (2007) ได้ศึกษาเพื่อกำหนดระดับมาตรฐานความสามารถทางการอ่านภาษาอังกฤษ สำหรับการสอบวัดระดับความรู้ ความสามารถทางการอ่านภาษาอังกฤษผ่านทางอินเทอร์เน็ต (TOEFL iBT reading test) 3 ระดับ ได้แก่ ระดับต่ำ (Low), ระดับปานกลาง (Intermediate) และระดับสูง (High)

นอกจากนี้ Watermann และ Klieme (2002 cited in Gomez, 2007) ศึกษาเพื่อกำหนดมาตรฐาน 4 ระดับ สำหรับการทดสอบเพื่อการประเมินผลการศึกษาวิทยาศาสตร์และคณิตศาสตร์ (Third International Mathematics Science Study: TIMSS) รวมถึงพัฒนาคำอธิบายระดับความสามารถที่เป็นที่ยอมรับ ซึ่งในรายงานผลสอบจะไม่นำเสนอเพียงคะแนนสอบเท่านั้นแต่จะอธิบายถึงระดับความรู้และทักษะของผู้สอบเมื่อเปรียบเทียบกับเกณฑ์ด้วย

มาตรฐานที่กำหนดขึ้นส่วนใหญ่มิได้เป็นเพียงเกณฑ์สำหรับการตีความคะแนนสอบเท่านั้น แต่ยังเป็นเสมือนเป้าหมายสำหรับท้าทายให้ไปให้ถึงอีกด้วย ดังปรากฏในเป้าหมายของประเทศสหรัฐอเมริกาที่ตั้งเป้าหมายเกี่ยวกับคุณภาพผู้เรียนไว้ว่า ในปี ค.ศ. 2014 ผู้เรียนทุกคนจะต้องมีความสามารถในระดับชำนาญการ (Proficient) ทุกรายวิชา โดยอาศัยสารสนเทศจากการทดสอบเพื่อวัดพัฒนาการทางการเรียนประกอบการวางแผนเพื่อดำเนินการไปสู่เป้าหมายที่คาดหวัง

นอกจากที่กล่าวมาข้างต้นแล้วมาตรฐานยังสามารถนำไปใช้ในหลายๆ บทบาท อาทิ ใช้เพื่อการรับรอง (Certifiers) เช่น ใช้เพื่อการรับรองการจบการศึกษาและรับรองความสามารถในการทำงาน ใช้เพื่อทำนาย (Predictors) เช่น ใช้เพื่อคัดเลือกบุคคลเข้าศึกษาต่อในมหาวิทยาลัย ใช้เพื่อบรรยาย (Descriptors) เช่น ใช้แบ่งกลุ่มผู้สอบตามระดับความสามารถ และใช้เพื่อกระตุ้น (Motivators) เช่น ใช้เป็นคะแนนเป้าหมายของการจัดการศึกษา หรือการเรียนการสอน เป็นต้น (Green, 1995)

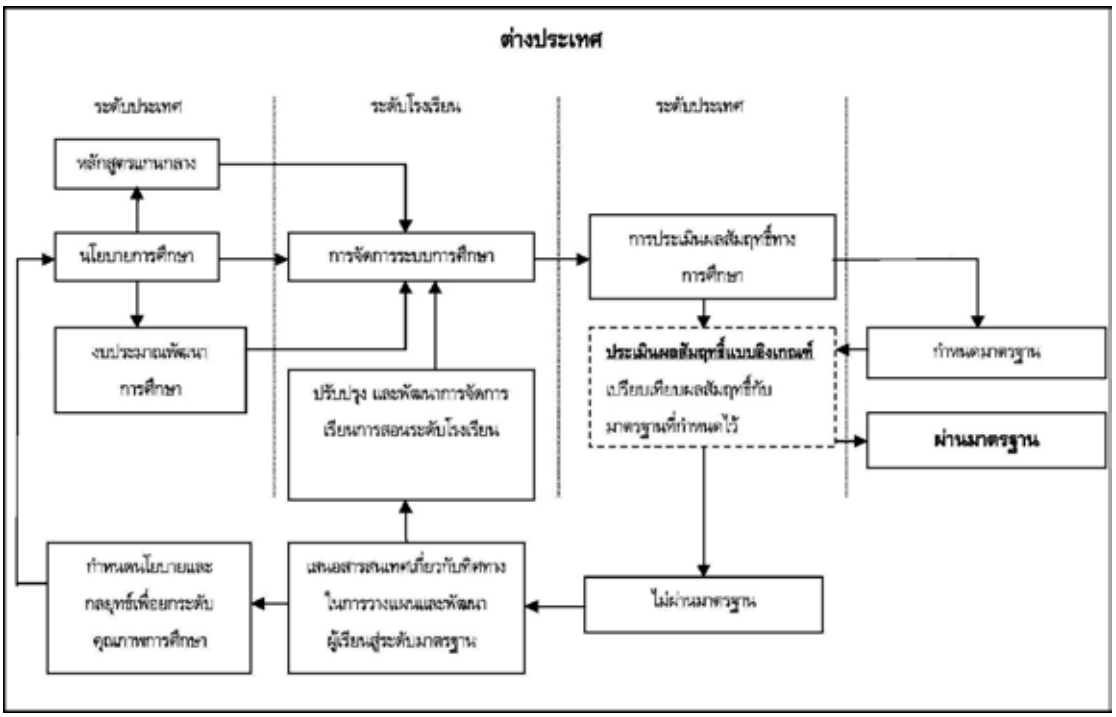
สำหรับประเทศไทยการกำหนดมาตรฐานเพื่อใช้ในการตีความหมายคะแนนสอบยังไม่เป็นที่แพร่หลายนัก แม้แต่การสอบที่มีความสำคัญ อาทิ การทดสอบทางการศึกษาแห่งชาติขั้นพื้นฐาน ซึ่งเป็นการทดสอบเพื่อประเมินคุณภาพผู้เรียนในระดับชาติตามมาตรฐานการเรียนรู้ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน ผลคะแนนจากการทดสอบทางการศึกษาแห่งชาติขั้นพื้นฐานถูกตีความแบบอิงกลุ่ม กล่าวคือ เป็นการตีความคะแนนสอบโดยการเปรียบเทียบคะแนนของผู้สอบด้วยกัน ซึ่งสารสนเทศที่ได้ประกอบด้วยข้อมูลเกี่ยวกับ ค่าเฉลี่ย ค่าสูงสุด ค่าต่ำสุด ค่าเปอร์เซ็นต์ไทล์ ค่าการกระจาย เป็นต้น (สถาบันทดสอบทางการศึกษาแห่งชาติ, 2550)

ทั้งนี้เนื่องจากการทดสอบทางการศึกษาแห่งชาติขั้นพื้นฐานเป็นการสอบเพื่อประเมินผลการจัดการศึกษา และผลจากการสอบจะเป็นข้อมูลสำคัญสำหรับการวางแผนเพื่อดำเนินการปรับปรุงหลักสูตร วิธีการสอน อันจะนำไปสู่การพัฒนาคุณภาพการศึกษาของประเทศ ซึ่งถือว่าผลจากการ

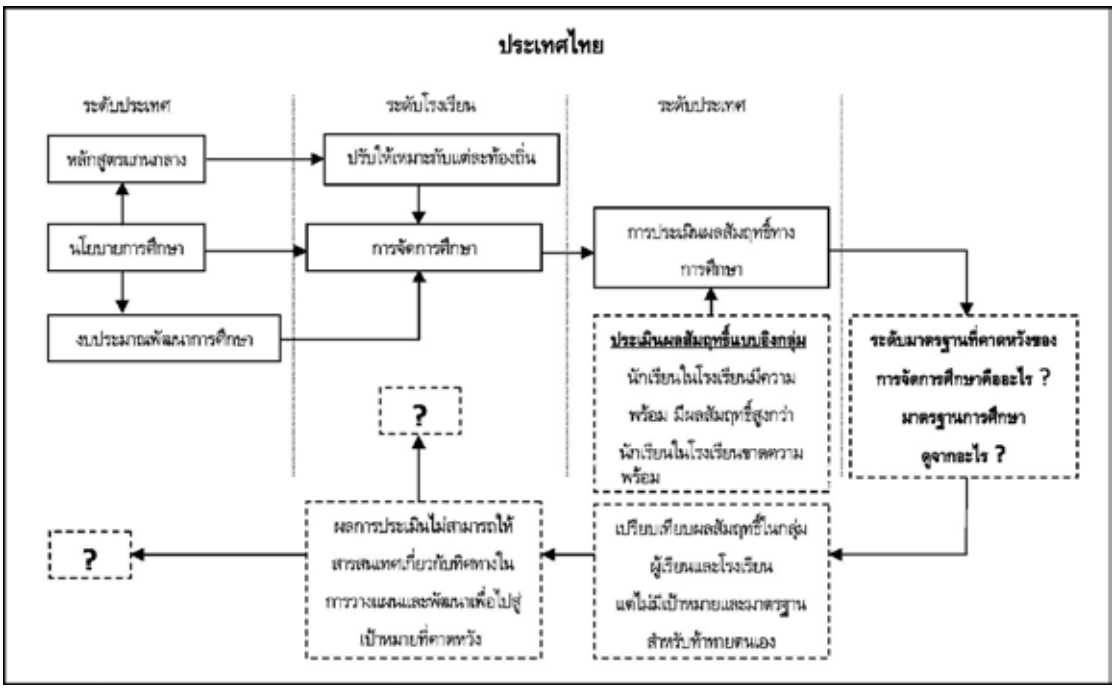
สอบเป็นสารสนเทศที่มีความสำคัญอย่างยิ่ง แต่สารสนเทศที่ได้จากการสอบจะเป็นประโยชน์หรือไม่ก็ขึ้นอยู่กับวิธีการตีความคะแนนสอบ

เมื่อพิจารณารายงานผลการทดสอบทางการศึกษาแห่งชาติขั้นพื้นฐาน พบว่า สารสนเทศที่ได้ทำให้ทราบเพียงว่านักเรียนที่เรียนในโรงเรียนขนาดใหญ่และมีความพร้อมในทุก ๆ ด้านมีคะแนนสูงกว่านักเรียนที่เรียนในโรงเรียนขนาดเล็กซึ่งขาดแคลนวัสดุ อุปกรณ์ รวมถึงโอกาสการในเข้าถึงทรัพยากรต่าง ๆที่เอื้อต่อการเรียนรู้ ซึ่ง Green (1995) กล่าวว่า ผลจากการประเมินแบบอิงกลุ่ม เป็นสารสนเทศที่มีรายละเอียดไม่เพียงพอต่อการนำใช้ กล่าวคือ ยังขาดความชัดเจนเกี่ยวกับระดับความรู้และทักษะที่ผู้สอบแต่ละคนมีเมื่อเปรียบเทียบกับเป้าหมายหรือมาตรฐานที่กำหนดไว้ จึงทำให้ไม่รู้ว่าการจัดการเรียนการสอนวิธีเดิม หลักสูตรแบบเดิม และแผนด้านการศึกษาแบบเดิมส่งผลให้คุณภาพการศึกษาเข้าใกล้เป้าหมายที่กำหนดไว้มากขึ้นหรือน้อยลง กล่าวคือ สารสนเทศจากการสอบไม่สามารถกระตุ้นให้เกิดการพัฒนาคุณภาพการศึกษา ซึ่งถือว่าไม่ได้ใช้ประโยชน์จากการประเมินผลสัมฤทธิ์ทางการศึกษาอย่างคุ้มค่า จึงไม่น่าสงสัยว่าเหตุใดคะแนนเฉลี่ยจากการทดสอบทางการศึกษาแห่งชาติขั้นพื้นฐานของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ต่ำอย่างต่อเนื่อง กล่าวคือ คะแนนเฉลี่ยต่ำกว่าร้อยละ 50 ใน 5 วิชาหลัก คือ ภาษาไทย สังคมศึกษา คณิตศาสตร์ วิทยาศาสตร์ และภาษาอังกฤษ ตั้งแต่ปี 2548-2551 (สำนักงานเลขาธิการสภาการศึกษา, 2551ข)

จากที่กล่าวมาเห็นได้ว่าการทดสอบทางการศึกษาจะสามารถมีบทบาทสำคัญต่อการพัฒนาคุณภาพการศึกษาหรือไม่นั้น ส่วนหนึ่งขึ้นอยู่กับวิธีการตีความคะแนนจากการสอบ เพื่อแสดงให้เห็นความแตกต่างของบทบาทของการทดสอบทางการศึกษาและการตีความคะแนนสอบที่มีต่อการวางแผนและพัฒนาการศึกษาระหว่างต่างประเทศกับประเทศไทย ผู้วิจัยขอนำเสนอเป็นภาพดังต่อไปนี้



ภาพที่ 1.1 บทบาทของการทดสอบทางการศึกษาและการตีความคะแนนสอบที่มีต่อการวางแผนและพัฒนาการศึกษาของต่างประเทศ



ภาพที่ 1.2 บทบาทของการทดสอบทางการศึกษาและการตีความคะแนนสอบที่มีต่อการวางแผนและพัฒนาการศึกษาของประเทศไทย

จากที่กล่าวมาการทดสอบทางการศึกษาแห่งชาติขั้นพื้นฐานควรมีการกำหนดมาตรฐานขึ้น เพื่อใช้เป็นเกณฑ์ในการประเมินผลจากการจัดการศึกษา และเพื่อให้ได้สารสนเทศที่เป็นประโยชน์ต่อการนำไปใช้เพื่อพัฒนาคุณภาพการศึกษามากขึ้น โดยควรเลือกวิธีการที่เหมาะสมในการกำหนดมาตรฐานเนื่องจากเป็นการกำหนดมาตรฐานสำหรับการสอบที่มีความสำคัญ

วิธีการกำหนดมาตรฐานมีหลายวิธี โดยวิธีที่ได้รับการยอมรับและถูกใช้สำหรับการกำหนดมาตรฐานที่มีความสำคัญมากที่สุด คือ วิธีแองกอฟและวิธีบุ๊คมาร์ค (Zieky et al., 2008)

วิธีแองกอฟถูกคิดขึ้นโดย William H. Angoff ในปีค.ศ.1970 เพื่อกำหนดคะแนนจุดตัดสำหรับข้อสอบแบบเลือกตอบ (Multiple choice tests) วิธีนี้ได้รับความนิยมอย่างกว้างขวางโดยเฉพาะการทดสอบที่มีความสำคัญ (high stakes Testing) (Angoff, 1971, cited in Kane, 1994; Zieky et al., 2008) ทั้งนี้เนื่องจากว่าวิธีนี้มีขั้นตอนการดำเนินการที่ง่าย เน้นการพิจารณาข้อสอบเป็นหลัก และสามารถดำเนินการได้ทั้งก่อนและหลังการสอบ โดยวิธีนี้จะให้ผู้เชี่ยวชาญเป็นผู้พิจารณากำหนดคะแนนจุดตัด ซึ่งผู้เชี่ยวชาญจะต้องพิจารณาความน่าจะเป็นที่ผู้มีความสามารถคาบเส้น (Borderline Candidate) จะสามารถตอบข้อสอบได้ถูกต้อง โดยคะแนนจุดตัดจะได้มาจากผลรวมของค่าความน่าจะเป็นของข้อสอบทุกข้อ

ต่อมาวิธีแองกอฟได้ถูกพัฒนาให้ง่ายต่อการปฏิบัติมากขึ้นโดย Impara และ Plake (1997) ซึ่งใช้ชื่อว่า วิธีแองกอฟแบบใช่/ไม่ใช่ เพื่อให้สามารถกำหนดคะแนนจุดตัดได้ในกรณีที่ข้อสอบแต่ละข้อให้คะแนนไม่เท่ากัน การพิจารณากำหนดคะแนนจุดตัดจะต่างจากวิธีแองกอฟตรงที่วิธีแองกอฟแบบใช่/ไม่ใช่จะให้ผู้เชี่ยวชาญพิจารณาว่าผู้มีความสามารถคาบเส้นจะสามารถตอบข้อสอบได้ถูกต้องใช่หรือไม่ ซึ่ง Impara และ Plake (1997) ได้ทำการทดลองเพื่อเปรียบคะแนนจุดตัดที่กำหนดโดยใช้วิธีแองกอฟและวิธีแองกอฟแบบใช่/ไม่ใช่ พบว่า ได้คะแนนจุดตัดเท่ากันและจากการศึกษาของ Chinn และ Hertz (2002) พบว่า ผู้เชี่ยวชาญมีความคิดเห็นว่าวิธีแองกอฟแบบใช่/ไม่ใช่ ง่ายต่อการปฏิบัติมากกว่าวิธีแองกอฟ

วิธีบุ๊คมาร์คถูกเสนอขึ้นครั้งแรกในปี 1996 โดย Lewis และ Green (1996 cited in Zieky et al., 2008) และได้รับความนิยมอย่างรวดเร็ว โดยเฉพาะในปี ค.ศ. 2000 วิธีบุ๊คมาร์คถูกนำไปใช้กันอย่างแพร่หลายในสหรัฐอเมริกา โดยถูกนำไปใช้ใน 18 รัฐ ซึ่งถือว่าถูกใช้มากที่สุด (Nellhaus, 2000) ต่อมาในปี ค.ศ. 2001 วิธีนี้ก็ได้ถูกนำมาใช้เพื่อกำหนดคะแนนจุดตัดมากกว่า

28 รัฐ (Egan, 2001 cited in Lin, 2008) และเมื่อปี 2005 มี 31 รัฐ ที่ใช้วิธีนี้เพื่อกำหนดคะแนนจุดตัดและใช้ในการกำหนดคะแนนจุดตัดสำหรับการสอบที่สำคัญ (Perie, 2005 cited in Lee & Lewis, 2008) วิธีบุ๊คมาร์คเป็นวิธีที่สามารถใช้ได้ทั้งกับข้อสอบที่เป็นแบบเลือกตอบที่ให้คะแนนแบบ 2 ค่า และข้อสอบแบบให้คะแนนมากกว่า 2 ค่า นอกจากนี้วิธีบุ๊คมาร์คยังมีจุดเด่นอยู่ที่คู่มือเรียงข้อสอบ (Ordered Item Booklet: OIB) ซึ่งจัดทำโดยการเรียงข้อสอบจากข้อง่ายไปข้อยาก โดยใช้ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory) ในการวิเคราะห์ค่าความยากของข้อสอบ เป็นการทำให้ความสามารถของนักเรียนอยู่ในระดับเดียวกัน (same continuum) ซึ่งช่วยให้ผู้เชี่ยวชาญกำหนดจุดตัดได้ง่ายขึ้น

วิธีแองกอฟแบบใช่/ไม่ใช่ และวิธีบุ๊คมาร์คเป็นวิธีที่ถูกพัฒนาขึ้นมาเพื่อลดจุดอ่อนของวิธีแองกอฟ กล่าวคือ มีเป้าหมายเพื่อลดความซับซ้อนของกระบวนการคิดของผู้เชี่ยวชาญ (cognitive complexity) (Skaggs & Hein, 2011, p. 586) โดยวิธีแองกอฟแบบใช่/ไม่ใช่จะให้ผู้เชี่ยวชาญตัดสินใจว่าผู้มีความสามารถคาบเส้นของแต่ละระดับความสามารถจะสามารถตอบข้อสอบได้ถูกต้องใช่หรือไม่ แทนการพิจารณาความน่าจะเป็นที่ผู้มีความสามารถคาบเส้นของแต่ละระดับความสามารถจะสามารถตอบข้อสอบได้ถูกต้อง ส่วนวิธีบุ๊คมาร์คจะให้ผู้เชี่ยวชาญตัดสินใจเพื่อกำหนดคะแนนจุดตัดโดยพิจารณาข้อสอบโดยภาพรวมทั้งฉบับในคู่มือเรียงข้อสอบแล้วให้ผู้เชี่ยวชาญพิจารณากำหนดคะแนนจุดตัดโดยนำที่คั่นหนังสือมาคั่นที่ข้อสอบที่ผู้มีความสามารถคาบเส้นของระดับความสามารถที่กำลังพิจารณาอยู่ มีความน่าจะเป็น 0.67 ที่จะตอบได้ถูกต้องแทนการพิจารณาเป็นรายข้อเหมือนวิธีแองกอฟ

วิธีแองกอฟแบบใช่/ไม่ใช่ และวิธีบุ๊คมาร์คต่างก็มีจุดเด่นที่ต่างกัน โดยวิธีแองกอฟแบบใช่/ไม่ใช่มีจุดเด่นคือการให้ความสำคัญกับการกำหนดคะแนนจุดตัดของผู้เชี่ยวชาญโดยปราศจากกรอบเรื่องค่าความยากของข้อสอบ กล่าวคือ เป็นการกำหนดคะแนนจุดตัดที่ให้ความสำคัญกับการตัดสินใจของผู้เชี่ยวชาญที่พิจารณาจากเนื้อหาที่สอบวัดเป็นหลัก สำหรับวิธีบุ๊คมาร์คมีจุดเด่นอยู่ที่คู่มือเรียงข้อสอบ ที่ช่วยให้ผู้เชี่ยวชาญกำหนดคะแนนจุดตัดได้ง่ายและเร็ว

สำหรับข้อด้อยของวิธีแองกอฟแบบใช่/ไม่ใช่นั้น ผู้เชี่ยวชาญจะต้องใช้กระบวนการคิดอย่างมากในการพิจารณาว่าผู้มีความสามารถคาบเส้นจะสามารถตอบข้อสอบได้ถูกต้องใช่หรือไม่ โดยเฉพาะในกรณีที่มีจำนวนข้อสอบที่ต้องพิจารณามาก ๆ และต้องกำหนดคะแนนจุดตัดหลายรอบ

อาจทำให้ผู้เชี่ยวชาญเกิดความอ่อนล้าในขณะที่กำหนดคะแนนจุดตัดซึ่งอาจส่งผลให้คะแนนจุดตัดที่กำหนดเกิดความคลาดเคลื่อนได้ (Clauser, et al., 2009) ส่วนวิธีบุ๊คมาร์คมีข้อด้อย คืออาจเกิดกรณีที่ผู้เชี่ยวชาญรู้สึกว่าข้อสอบที่ถูกเรียงไว้ตามลำดับความยากนั้นบางข้อมีตำแหน่งที่ไม่เหมาะสม ซึ่งกรณีเช่นนี้อาจส่งผลต่อคุณภาพของคะแนนจุดตัดที่กำหนดได้

ผู้วิจัยจึงสนใจที่จะพัฒนาวิธีการกำหนดคะแนนจุดตัดขึ้นมาโดยผสมผสานระหว่างวิธีแองกอฟแบบใช่/ไม่ใช่ และวิธีบุ๊คมาร์คเพื่อดึงเอาจุดเด่นและลดข้อด้อยของแต่ละวิธีลง เรียกว่าวิธีเอบีซี โดยมีเป้าหมายเพื่อลดความซับซ้อนของกระบวนการคิดของผู้เชี่ยวชาญ และเพื่อให้ง่ายต่อการดำเนินการ ซึ่งจะทำให้ผู้เชี่ยวชาญสามารถกำหนดคะแนนจุดตัดได้ง่ายและเร็วขึ้น

วิธีเอบีซีจะให้ผู้เชี่ยวชาญกำหนดคะแนนจุดตัดจากคู่มือเรียงข้อสอบซึ่งถือเป็นจุดเด่นของวิธีบุ๊คมาร์ค แต่วิธีเอบีซีจะแบ่งคู่มือเรียงข้อสอบเป็นหลาย ๆ ฉบับตามสาระการเรียนรู้ เรียกว่า คู่มือเรียงข้อสอบรายสาระการเรียนรู้ (Domain–Ordered Item Booklet: D-OIB) ซึ่งเป็นคู่มือที่จัดทำขึ้นโดยแบ่งข้อสอบเป็นกลุ่มตามสาระการเรียนรู้ กล่าวคือ คู่มือเรียงข้อสอบรายสาระการเรียนรู้จะมีหลายฉบับขึ้นอยู่กับจำนวนสาระการเรียนรู้ที่สอบวัด ข้อสอบในแต่ละสาระการเรียนรู้จะถูกจัดเรียงจากข้องายไปข้อยาก ข้อละ 1 หน้า โดยใช้ทฤษฎีการตอบสนองข้อสอบ โมเดล 3 พารามิเตอร์ (Three parameter model) ในการวิเคราะห์ค่าความยากของข้อสอบ (b) ทั้งนี้ก็เพื่อให้การพิจารณากำหนดคะแนนจุดตัดทำได้ง่ายขึ้น เนื่องจากคู่มือเรียงข้อสอบรายสาระการเรียนรู้แต่ละฉบับจะประกอบด้วยข้อสอบที่วัดสาระการเรียนรู้เดียวกัน ส่วนหลักการพิจารณาเพื่อกำหนดคะแนนจุดตัดจะดำเนินคล้ายกับวิธีแองกอฟแบบใช่/ไม่ใช่ กล่าวคือ ให้ผู้เชี่ยวชาญพิจารณาข้อสอบตามลำดับความยาก และตัดสินว่าผู้ที่สามารถตอบข้อสอบแต่ละข้อได้ถูกต้องควรมีความสามารถอย่างน้อยระดับใดแทนที่จะให้ผู้เชี่ยวชาญนำที่คั่นหนังสือมาคั่นตรงข้อที่ผู้เชี่ยวชาญคิดว่าเป็นจุดตัด ซึ่งการพิจารณาแบบนี้น่าจะช่วยลดปัญหาที่เกิดจากการที่ผู้เชี่ยวชาญคิดว่าข้อสอบที่ถูกจัดเรียงไว้ตามความยากมีลำดับที่ไม่เหมาะสม เพราะวิธีเอบีซีใช้คู่มือเรียงข้อสอบเป็นเพียงสารสนเทศประกอบการกำหนดคะแนนจุดตัดที่เสนอเกี่ยวกับค่าความยากของข้อสอบที่ได้จากการวิเคราะห์จากผลการตอบของผู้สอบเท่านั้น กล่าวได้ว่าวิธีเอบีซียังคงให้ความสำคัญกับความคิดเห็นของผู้เชี่ยวชาญที่พิจารณาจากเนื้อหาที่สอบเป็นหลักแทนที่จะต้องพิจารณาคะแนนจุดตัดที่ถูกจำกัดไว้กับลำดับของความยากของข้อสอบจนเกินไป

นอกจากนี้วิธีเอบีซีจะเสนอสารสนเทศที่มากกว่าค่าความยากของข้อสอบ โดยจะเสนอสารสนเทศเกี่ยวกับผลการวิเคราะห์กลุ่ม (Clustering – Result: CR) ซึ่งเป็นสารสนเทศที่ได้จากการวิเคราะห์กลุ่ม (Cluster Analysis) เป็นการวิเคราะห์เพื่อจัดให้กลุ่มผู้สอบที่มีลักษณะที่สนใจเหมือนหรือคล้ายกันอยู่กลุ่มเดียวกัน และแต่ละกลุ่มจะต้องมีลักษณะที่สนใจต่างกันให้มากที่สุด ซึ่งกล่าวได้ว่าเป็นการวิเคราะห์เพื่อการกำหนดกลุ่มให้ผู้สอบจากลักษณะที่สนใจโดยปราศจากความลำเอียงใดๆ เนื่องจากเป็นผลการจัดกลุ่มที่มิใช่เป็นการกำหนดโดยอาศัยความคิดเห็นของผู้เชี่ยวชาญแต่เป็นการวิเคราะห์ข้อมูลทางสถิติที่ใช้ข้อมูลจากผลการสอบของผู้สอบมาทำการวิเคราะห์ (Berk, 1996; Plake, 2008; Lee & Lewis, 2008; Näsström & Nyström, 2008) ตัวแปรที่ใช้ในการแบ่งกลุ่ม ได้แก่ คะแนนรวมรายสาระการเรียนรู้ สารสนเทศนี้จะนำเสนอเป็นผลคะแนนเฉลี่ยรายสาระการเรียนรู้ของแต่ละกลุ่ม ซึ่งข้อมูลนี้จะทำให้ทราบว่าผู้สอบแต่ละกลุ่มสามารถทำคะแนนในแต่ละสาระการเรียนรู้เฉลี่ยเป็นเท่าใด สาระการเรียนรู้ใดได้มากและสาระการเรียนรู้ใดได้น้อย และเสนอช่วงของคะแนนสอบที่คาบเกี่ยวกันของแต่ละกลุ่ม ซึ่งได้จากคะแนนที่ซ้อนทับกันระหว่างแต่ละกลุ่ม เป็นคะแนนระหว่างคะแนนสูงที่สุดของกลุ่มที่ต่ำกว่าไปถึงคะแนนที่ต่ำที่สุดของกลุ่มสูงกว่า (รายละเอียดดังแสดงในภาคผนวก ฉ) ผู้วิจัยคิดว่าคะแนนเฉลี่ยรายสาระการเรียนรู้ของแต่ละกลุ่มน่าจะเป็นข้อมูลที่เป็นประโยชน์ต่อการตัดสินใจกำหนดข้อสอบไว้ในแต่ละกลุ่ม ส่วนข้อมูลเกี่ยวกับช่วงของคะแนนสอบที่คาบเกี่ยวกันของแต่ละกลุ่มน่าจะเป็นประโยชน์ต่อผู้เชี่ยวชาญในแง่การเปรียบเทียบผลการกำหนดคะแนนจุดตัดของผู้เชี่ยวชาญกับช่วงคะแนนที่คาบเกี่ยวกันของแต่ละกลุ่มที่ถูกกำหนดโดยการวิเคราะห์กลุ่ม

ดังนั้นในการศึกษาครั้งนี้ผู้วิจัยได้ทำการกำหนดคะแนนจุดตัดสำหรับการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐานวิชาคณิตศาสตร์ และวิชาวิทยาศาสตร์ ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ด้วยวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซี รวมทั้งตรวจสอบคุณภาพของคะแนนจุดตัดด้วย เนื่องจากเป็นเกณฑ์ที่ใช้สำหรับการตัดสินที่มีความสำคัญและมีผลกระทบสูง จึงควรนำเสนอสารสนเทศเกี่ยวกับคุณภาพของคะแนนจุดตัดเพื่อสร้างความเชื่อมั่นและการยอมรับต่อคะแนนจุดตัดสำหรับผู้ที่เกี่ยวข้อง

คุณภาพของคะแนนจุดตัดพิจารณาได้จากคุณภาพด้านความเที่ยง และคุณภาพด้านความตรง ซึ่งงานวิจัยส่วนใหญ่จะตรวจสอบคุณภาพด้านความเที่ยงในประเด็นต่อไปนี้คือ ความเที่ยงภายใน

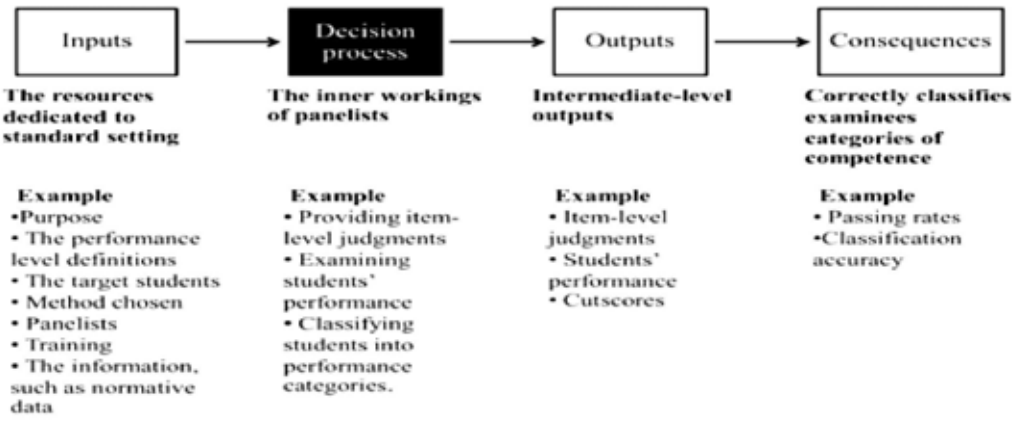
ผู้เชี่ยวชาญ (Intra-rater reliability) เป็นการตรวจสอบจากผลการกำหนดคะแนนจุดตัดของผู้เชี่ยวชาญในแต่ละรอบว่าเท่าเดิมหรือใกล้เคียงกันทุกรอบหรือไม่ และความเที่ยงระหว่างผู้เชี่ยวชาญ (Inter-rater reliability) เป็นการตรวจสอบว่าผู้เชี่ยวชาญมีความเห็นในทิศทางเดียวกันหรือไม่ (Kane, 1994; Näsström & Nyström, 2008; Papageorgiou, 2010)

นอกจากนี้เนื่องจากการกำหนดคะแนนจุดตัดส่วนใหญ่เป็นการดำเนินการกำหนดโดยอาศัยผู้เชี่ยวชาญเป็นหลัก และมีการดำเนินการกำหนดคะแนนจุดตัดหลายรอบ ซึ่งกล่าวได้ว่าในสถานการณ์หรือเงื่อนไขต่าง ๆ ของการกำหนดคะแนนจุดตัดย่อมเกิดความแปรปรวนของคะแนนจุดตัดอันเนื่องมาจากแหล่งความคลาดเคลื่อนต่าง ๆ เช่น ความคลาดเคลื่อนจากจำนวนผู้เชี่ยวชาญ ความคลาดเคลื่อนจากจำนวนกลุ่มย่อย และความคลาดเคลื่อนจากจำนวนรอบของการพิจารณา เป็นต้น ดังนั้นจึงควรมีการวิเคราะห์ข้อมูลเพื่อให้ทราบถึงแหล่งความแปรปรวนที่ส่งผลต่อการกำหนดคะแนนจุดตัด โดยอาศัยทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือ (Generalizability Theory) ซึ่งเป็นทฤษฎีที่เหมาะกับการประเมินการปฏิบัติของผู้เชี่ยวชาญ (Brennan, 2001, p.117) และสามารถนำมาศึกษาในกรณีที่มีความคลาดเคลื่อนหลายแหล่งได้ (ศิริชัย กาญจนวาสี, 2550, p.11)

ผู้วิจัยจึงสนใจที่จะศึกษาขนาดความแปรปรวนจากแหล่งความคลาดเคลื่อนต่าง ๆที่เกิดขึ้นจากการกำหนดคะแนนจุดตัด โดยฟาเซทของการวัดมี 3 ฟาเซท คือ ผู้เชี่ยวชาญ (p) กลุ่มผู้เชี่ยวชาญ (g) และจำนวนรอบของการพิจารณา (r) สิ่งที่ต้องการวัดคือผลการกำหนดคะแนนจุดตัด รูปแบบความสัมพันธ์ระหว่างฟาเซทเป็น (p:g)xr Design กล่าวคือ แต่ละกลุ่มประกอบด้วยผู้เชี่ยวชาญที่ไม่ซ้ำกันและ ทุกกลุ่มกำหนดคะแนนจุดตัดทุกรอบ ซึ่งการศึกษาครั้งนี้จะศึกษาขนาดองค์ประกอบความแปรปรวนและวิเคราะห์ค่าสัมประสิทธิ์การสรุปอ้างอิง เพื่อให้ทราบถึงขนาดของแหล่งความแปรปรวนที่ส่งผลต่อคะแนนจุดตัด อันจะเป็นข้อมูลสำคัญสำหรับการวางแผนเพื่อให้สามารถควบคุมแหล่งความคลาดเคลื่อนที่จะเกิดขึ้นจากการกำหนดคะแนนจุดตัดด้วยวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซีได้ตรงประเด็น และได้สารสนเทศสำหรับการตัดสินใจเกี่ยวกับจำนวนผู้เชี่ยวชาญ และจำนวนกลุ่มย่อยของผู้เชี่ยวชาญที่เหมาะสม เพื่อให้สามารถกำหนดคะแนนจุดตัดที่มีความเที่ยงสูงถึงระดับที่ต้องการต่อไป

นอกจากนี้การศึกษาครั้งนี้ได้ทำการตรวจสอบความตรงของคะแนนจุดตัดด้วย ซึ่งถือเป็นคุณสมบัติของคะแนนจุดตัดที่สำคัญที่สุด แต่ก็เป็นคุณสมบัติที่ตรวจสอบได้ยากที่สุดด้วย

McGinty (2005) เสนอว่า การตรวจสอบความตรง ควรตรวจสอบถึงการได้มาซึ่งคะแนนจุดตัดว่ามีความตรงหรือไม่ โดยเสนอให้ตรวจสอบกระบวนการกำหนดคะแนนจุดตัด ซึ่งประกอบด้วย ปัจจัยนำเข้า (input) กระบวนการตัดสินใจ (decision processes) ผลผลิต (output) และ ผลที่ตามมา (consequence) ดังแสดงในภาพต่อไปนี้



ภาพที่ 1.3 กระบวนการกำหนดคะแนนจุดตัด

ที่มา: ปรับปรุงจาก McGinty (2005)

ปัจจัยนำเข้าประกอบด้วย เป้าหมายของการกำหนดคะแนนจุดตัด (Purpose) คำอธิบายระดับความสามารถ (Performance level descriptor) วิธีการกำหนดคะแนนจุดตัด (Method) ผู้เชี่ยวชาญ (Panelists) การฝึกอบรม (Training) และสารสนเทศที่เสนอต่อผู้เชี่ยวชาญ (Information) ส่วนกระบวนการตัดสินใจเป็นกระบวนการคิดและตัดสินใจของผู้เชี่ยวชาญในการกำหนดคะแนนจุดตัด ผลผลิตคือคะแนนจุดตัดที่กำหนดขึ้น และผลที่ตามมา (Consequence) คือผลจากการนำคะแนนจุดตัดไปใช้ในการกำหนดกลุ่มผู้สอบตามระดับความสามารถ ได้แก่ ความถูกต้องของการกำหนดกลุ่ม (Classification accuracy) และความคลาดเคลื่อนของการกำหนดกลุ่ม (Classification errors) อัตราการผ่านหรือตก (passing rates) เป็นต้น McGinty (2005) กล่าวว่าการตรวจสอบคุณภาพด้านความตรงของคะแนนจุดตัดควรตรวจสอบความตรงขององค์ประกอบทั้ง 4 ในกระบวนการกำหนดคะแนนจุดตัดที่กล่าวมา

การตรวจสอบความตรงของปัจจัยนำเข้านั้น Kane (1994) เสนอว่าควรให้ผู้เชี่ยวชาญที่เข้าร่วมการกำหนดคะแนนจุดตัดเป็นผู้ตรวจสอบ ในประเด็นต่อไปนี้ ได้แก่ ความถูกต้อง ความเป็นไปได้ ความเหมาะสม ความเข้าใจและความมั่นใจของผู้เชี่ยวชาญ และความเป็นประโยชน์ของเอกสาร

ความถูกต้อง หมายถึง การดำเนินการมีลำดับขั้นตอนที่ถูกต้อง และสารสนเทศประกอบการดำเนินการมีเนื้อหาที่ถูกต้อง สื่อความหมายชัดเจน สำหรับความเป็นไปได้ หมายถึง วิธีการดำเนินการสามารถปฏิบัติได้จริงและง่ายต่อการดำเนินการ ความเหมาะสม หมายถึง เป้าหมาย การดำเนินการ และเอกสารประกอบการดำเนินการ มีความสอดคล้องและเชื่อมโยงกัน ความเข้าใจและความมั่นใจของผู้เชี่ยวชาญ หมายถึง ผู้เชี่ยวชาญมีความเข้าใจเกี่ยวกับความสำคัญ เป้าหมาย คำอธิบายระดับความสามารถ การดำเนินการ และสารสนเทศประกอบการดำเนินการ รวมถึงมีความมั่นใจต่อคะแนนจุดตัดที่กำหนดขึ้น และความเป็นประโยชน์ของเอกสารประกอบการกำหนดคะแนนจุดตัด หมายถึง สารสนเทศประกอบการดำเนินการเป็นประโยชน์ต่อการกำหนดคะแนนจุดตัดของผู้เชี่ยวชาญ

กระบวนการตัดสินใจของผู้เชี่ยวชาญเปรียบเสมือนเป็นกล่องดำ (Black box) เนื่องจากเป็นกระบวนการที่เกิดขึ้นภายในจิตใจของผู้เชี่ยวชาญที่ไม่สามารถล่วงรู้ได้ McGinty (2005) กล่าวว่าการตรวจสอบความตรงของกระบวนการตัดสินใจมีความสำคัญอย่างยิ่ง เนื่องจากทำให้ทราบเกี่ยวกับเหตุผลและหลักการที่ผู้เชี่ยวชาญยึดถือในการคะแนนจุดตัด และทราบเกี่ยวกับความคิดความกังวลขณะตัดสินใจ รวมถึงปัญหาและอุปสรรคที่เกิดขึ้น มีนักวิชาการหลายท่านที่คิดว่าการเก็บรวบรวมข้อมูลเกี่ยวกับกระบวนการคิดและการตัดสินใจของผู้เชี่ยวชาญในระหว่างที่ผู้เชี่ยวชาญกำลังใช้ความคิดหรือกำลังตัดสินใจ อาจจะเป็นการรบกวนผู้เชี่ยวชาญและอาจส่งผลต่อคุณภาพของการตัดสินใจได้ จึงได้ทำการเก็บรวบรวมข้อมูลเกี่ยวกับกระบวนการคิดและการตัดสินใจของผู้เชี่ยวชาญในช่วงที่ผู้เชี่ยวชาญกำหนดคะแนนจุดตัดเสร็จสิ้นแล้ว โดยใช้วิธีการในการเก็บรวบรวมข้อมูลที่แตกต่างกันออกไป อาทิ Skorupski และ Hambleton (2005) ใช้แบบสอบถามแบบปลายเปิด Ferdous และ Plake (2005) ใช้วิธีการสนทนากลุ่ม (Focus group) และ Hein และ Skaggs (2009) ใช้วิธีสนทนากลุ่มเชิงลึก (In-depth focus group) ส่วน McGinty (2005) ใช้วิธีการสังเกตพฤติกรรม (Observing behavior) และวิธีการอภิปรายกลุ่ม (Group discussions) สำหรับ Papageorgiou (2010) ใช้วิธีการอภิปรายกลุ่ม

van Someren, Barnard และ Sandberg (1994) กล่าวว่าวิธีที่ดีที่สุดที่จะทำให้ล่วงรู้ถึงความคิดของบุคคล คือการให้คนๆ นั้น พูดในสิ่งที่กำลังคิดออกมาดัง ๆ และบันทึกสิ่งที่บุคคลนั้นพูดเอาไว้ ซึ่งเรียกว่าเป็นการเก็บข้อมูลโดยใช้เทคนิคการคิดออกเสียง (Think-aloud method) เทคนิคนี้ได้รับการยอมรับและถูกใช้เพื่อเก็บข้อมูลที่เกี่ยวกับความคิดหรือกระบวนการตัดสินใจของบุคคลอย่างแพร่หลาย กล่าวคือ มีนักวิชาการหลากหลายสาขานำเทคนิคนี้มาใช้ (อาทิ Branch, 2000; Whitting, López, Schley & Fisher, 2000; Katalin, 2000; Charters, 2003; Johnstone, Bottsford-Miller, & Thompson, 2006 เป็นต้น) ซึ่งเทคนิคนี้เหมาะที่จะใช้ในกรณีที่กลุ่มตัวอย่างมีขนาดไม่ใหญ่นัก (Bowen, 1994; Ericsson & Simon, 1984 cited in Branch, 2000)

ในการศึกษาครั้งนี้ผู้วิจัยได้เก็บรวบรวมข้อมูลเกี่ยวกับเหตุผลและหลักการที่ผู้เชี่ยวชาญยึดถือในการคะแนนจุดตัด รวมถึงความคิด ความกังวลขณะตัดสินใจ ปัญหาและอุปสรรคที่เกิดขึ้น โดยใช้หลาย ๆ วิธี ซึ่งวิธีการที่ใช้ประกอบด้วย การสังเกตพฤติกรรม การใช้เทคนิคการคิดออกเสียง การบันทึกข้อสนทนาในระหว่างที่ผู้เชี่ยวชาญร่วมกันอภิปรายกลุ่มและระหว่างที่พักรับประทานอาหารร่วมกัน และการสนทนากลุ่ม

การตรวจสอบความตรงของผลผลิตนั้น McGinty (2005) กล่าวว่า งานวิจัยส่วนใหญ่ตรวจสอบโดยการหาความสัมพันธ์ระหว่างคะแนนจุดตัดที่กำหนดขึ้นจากวิธีที่ต่างกันโดยวิเคราะห์ความสอดคล้องของการกำหนดกลุ่มของคะแนนจุดตัด

ความตรงของผลที่ตามมาสามารถพิจารณาได้จากความถูกต้องของการกำหนดกลุ่ม (Classification accuracy) และความคลาดเคลื่อนของการกำหนดกลุ่ม (Classification errors) ได้แก่ ความคลาดเคลื่อนแบบบวก (False positive) และความคลาดเคลื่อนแบบลบ (False negative)

จากที่กล่าวมาสรุปได้ว่าการศึกษาครั้งนี้ได้ตรวจสอบคุณภาพของคะแนนจุดตัดที่กำหนดด้วยวิธีบุ๊คมาร์ค วิธีแองกอฟแบบใช่/ไม่ใช่ และวิธีเอบีซี ด้านความตรงของคะแนนจุดตัด ได้แก่ ความตรงของปัจจัยนำเข้า ความตรงของกระบวนการตัดสินใจเพื่อกำหนดคะแนนจุดตัด ความตรงของผลลัพธ์ และความตรงของผลที่ตามมา และ คุณภาพด้านความเที่ยงของคะแนนจุดตัด ได้แก่ ค่าสัมประสิทธิ์การสรุปอ้างอิง ความเที่ยงภายในผู้เชี่ยวชาญ ความเที่ยงระหว่างผู้เชี่ยวชาญรวมถึงวิเคราะห์จำนวนผู้เชี่ยวชาญ และจำนวนกลุ่มผู้เชี่ยวชาญที่ความเหมาะสม

คำถามวิจัย

1. วิธีกำหนดคะแนนจุดตัดสำหรับการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐานควรมีลักษณะอย่างไร

2. คะแนนจุดตัดของแต่ละระดับความสามารถสำหรับการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐานวิชาคณิตศาสตร์ และวิชาวิทยาศาสตร์ ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ที่กำหนดด้วยวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซีเป็นเท่าใด

3. คะแนนจุดตัดที่กำหนดด้วยวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซี มีคุณภาพเป็นอย่างไร

4. ในการกำหนดคะแนนจุดตัดด้วยวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซี จำนวนผู้เชี่ยวชาญ และจำนวนกลุ่มผู้เชี่ยวชาญที่มีความเหมาะสมควรเป็นเท่าใด

วัตถุประสงค์ของการวิจัย

1. เพื่อพัฒนาวิธีการกำหนดคะแนนจุดตัดสำหรับการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐานวิชาคณิตศาสตร์ และวิชาวิทยาศาสตร์ ของนักเรียนชั้นมัธยมศึกษาปีที่ 6

2. เพื่อกำหนดคะแนนจุดตัดสำหรับการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐานวิชาคณิตศาสตร์ และวิชาวิทยาศาสตร์ ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ด้วยวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซี ซึ่งเป็นวิธีที่ผู้วิจัยพัฒนาขึ้น

3. เพื่อตรวจสอบคุณภาพของคะแนนจุดตัดที่กำหนดด้วยวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซี

4. เพื่อวิเคราะห์จำนวนผู้เชี่ยวชาญและจำนวนกลุ่มผู้เชี่ยวชาญ ที่มีความเหมาะสมเมื่อกำหนดคะแนนจุดตัดโดยวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซี

ขอบเขตของการวิจัย

1. ประชากรและกลุ่มตัวอย่าง

ประชากรผู้เชี่ยวชาญเพื่อกำหนดคะแนนจุดตัด ได้แก่ ครูผู้สอนวิชาคณิตศาสตร์ วิชาฟิสิกส์ วิชาเคมี และวิชาชีววิทยาในระดับมัธยมศึกษาตอนปลาย

การศึกษาครั้งนี้ผู้วิจัยใช้วิธีการสุ่มอย่างง่ายเพื่อให้ได้กลุ่มตัวอย่างผู้เชี่ยวชาญ ซึ่งเป็นครูผู้สอนวิชาคณิตศาสตร์จำนวน 27 คน และเป็นครูผู้สอนวิชาฟิสิกส์ จำนวน10 คน วิชาเคมี จำนวน10 คน และวิชาชีววิทยา จำนวน10 คน รวมทั้งสิ้น 57 คน จากนั้นจึงใช้วิธีการสุ่มอย่างง่าย เพื่อแบ่งครูผู้สอนวิชาคณิตศาสตร์เป็น 3 กลุ่ม และแบ่งครูผู้สอนวิชาฟิสิกส์ วิชาเคมี และวิชาชีววิทยา เป็น 3 กลุ่ม แล้วจับสลากเพื่อกำหนดวิธีการกำหนดคะแนนจุดตัดให้แต่ละกลุ่ม

2. ข้อสอบ

การศึกษาครั้งนี้ใช้ข้อสอบสำหรับการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน ช่วงชั้นที่ 4 (มัธยมศึกษาปีที่ 6) ปีการศึกษา 2553 วิชาคณิตศาสตร์และวิชาวิทยาศาสตร์ โดยข้อสอบวิชาคณิตศาสตร์มีจำนวน 40 ข้อ แบ่งเป็นข้อสอบแบบเลือกตอบ 4 ตัวเลือก จำนวน 20 ข้อ และข้อสอบแบบเติมคำตอบ จำนวน 20 ข้อ ประกอบด้วย 5 สาระการเรียนรู้ คะแนนเต็ม 100 คะแนน ส่วนวิชาวิทยาศาสตร์ มีจำนวน 90 ข้อ แบ่งเป็นข้อสอบแบบเลือกตอบที่มีคำตอบถูกเพียงข้อเดียว จำนวน 80 ข้อ และข้อสอบแบบเลือกตอบที่มีคำตอบถูก 2 ข้อ จำนวน 10 ข้อ ประกอบด้วย 8 สาระการเรียนรู้ คะแนนเต็ม 100 คะแนน

3. วิธีที่ใช้ในการกำหนดคะแนนจุดตัด

วิธีที่ใช้ในการกำหนดคะแนนจุดตัดประกอบด้วยวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซี ซึ่งวิธีเป็นวิธีการกำหนดคะแนนจุดตัดโดยผู้เชี่ยวชาญและเป็นวิธีแบบสอบเป็นศูนย์กลาง (test-center methods) กล่าวคือ ผู้เชี่ยวชาญกำหนดคะแนนจุดตัดโดยพิจารณาข้อสอบเป็นหลัก

4. ตัวแปรที่ศึกษา

4.1 ตัวแปรต้น คือ

4.1.1 วิธีการกำหนดคะแนนจุดตัด 3 วิธี ได้แก่ วิธีแองกอฟแบบใช่/ไม่ใช่วิธีบุ๊คมาร์ค และวิธีเอบีซี

4.1.2 จำนวนผู้เชี่ยวชาญ

4.1.3 จำนวนกลุ่มผู้เชี่ยวชาญ

4.2 ตัวแปรตาม คือ คุณภาพของคะแนนจุดตัด ประกอบด้วย

4.2.1 ความตรง ได้แก่

4.2.1.1 ความตรงของปัจจัยนำเข้า

4.2.1.2 ความตรงของกระบวนการตัดสินใจ

4.2.1.3 ความตรงของผลลัพธ์

4.2.1.4 ความตรงของผลที่ตามมา

4.2.2 ความเที่ยง ได้แก่

4.2.2.1 ค่าสัมประสิทธิ์การสรุปอ้างอิง

4.2.2.2 ความเที่ยงภายในผู้เชี่ยวชาญ

4.2.2.3 ความเที่ยงระหว่างผู้เชี่ยวชาญ

คำจำกัดความเชิงปฏิบัติการ

คะแนนจุดตัด หมายถึง คะแนน 3 ค่า ที่ใช้เป็นเกณฑ์ในการแบ่งกลุ่มนักเรียนตามระดับความสามารถ 4 ระดับ ได้แก่ ระดับที่ 1 ระดับที่ 2 ระดับที่ 3 และระดับที่ 4 ซึ่งได้จากการกำหนดด้วยวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซี

วิธีแองกอฟแบบใช่/ไม่ใช่ หมายถึง วิธีการกำหนดคะแนนจุดตัดที่มีขั้นตอนการดำเนินการดังนี้ รอบที่ 1 ผู้เชี่ยวชาญพิจารณาข้อสอบทีละข้อว่าเป็นข้อสอบที่ผู้มีความสามารถคาบเส้นของระดับความสามารถที่กำลังพิจารณาอยู่สามารถตอบได้ถูกต้องใช่หรือไม่ รอบที่ 2 ผู้เชี่ยวชาญร่วมกันอภิปรายกลุ่มย่อยเกี่ยวกับคะแนนจุดตัดที่กำหนดในรอบที่ 1 จากนั้นกำหนดคะแนนจุดตัดอีกครั้ง โดยอาศัยหลักการเดียวกับรอบที่ 1 สำหรับรอบที่ 3 ผู้เชี่ยวชาญร่วมกันอภิปรายกลุ่มใหญ่

เกี่ยวกับคะแนนจุดตัดที่กำหนดในรอบที่ 2 และกำหนดคะแนนจุดตัดอีกครั้ง โดยอาศัยหลักการเดียวกับรอบที่ 1

วิธีบุ๊คมาร์ค หมายถึง วิธีการกำหนดคะแนนจุดตัดที่มีขั้นตอนการดำเนินการดังนี้ รอบที่ 1 ผู้เชี่ยวชาญพิจารณากำหนดคะแนนจุดตัดโดยนำที่คั่นหนังสือมาคั่นที่ข้อสอบในคู่มือเรียงข้อสอบ โดยข้อที่อยู่ก่อนหน้าที่คั่นหนังสือจะเป็นข้อสอบที่ผู้มีความสามารถคาบเส้นของระดับความสามารถที่กำลังพิจารณาอยู่ มีความน่าจะเป็น 0.67 ที่จะตอบได้ถูกต้อง รอบที่ 2 ผู้เชี่ยวชาญร่วมกันอภิปรายกลุ่มย่อยเกี่ยวกับคะแนนจุดตัดที่กำหนดในรอบที่ 1 จากนั้นกำหนดคะแนนจุดตัดอีกครั้ง โดยอาศัยหลักการเดียวกับรอบที่ 1 สำหรับรอบที่ 3 ผู้เชี่ยวชาญร่วมกันอภิปรายกลุ่มใหญ่เกี่ยวกับคะแนนจุดตัดที่กำหนดในรอบที่ 2 และกำหนดคะแนนจุดตัดอีกครั้ง โดยอาศัยหลักการเดียวกับรอบที่ 1

คู่มือเรียงข้อสอบ หมายถึง คู่มือที่จัดทำขึ้นเพื่อใช้ในการกำหนดคะแนนจุดตัดโดยวิธีบุ๊คมาร์ค คู่มือเรียงข้อสอบจะนำเสนอข้อสอบตามลำดับความยากโดยเรียงจากข้อสอบที่ง่ายที่สุดไปหาข้อสอบที่ยากที่สุด โดยใช้ทฤษฎีการตอบสนองข้อสอบช่วยในการวิเคราะห์ค่าความยากของข้อสอบ

วิธีเอบีซี หมายถึง วิธีการกำหนดคะแนนจุดตัดที่ผู้วิจัยพัฒนาขึ้น โดยมีขั้นตอนการดำเนินการดังนี้ รอบที่ 1 ผู้เชี่ยวชาญพิจารณาคู่มือเรียงข้อสอบรายสาระการเรียนรู้ และผลการวิเคราะห์กลุ่ม จากนั้นก็ให้ผู้เชี่ยวชาญพิจารณาว่าผู้ที่สามารถตอบข้อสอบแต่ละข้อได้ถูกต้องควรมีความสามารถอย่างน้อยระดับใด โดยมีหลักการพิจารณาเพิ่มเติมคือหากข้อใดที่ผู้เชี่ยวชาญไม่มั่นใจก็ให้ระบุไว้ในข้อสอบ ดังนั้นข้อสอบที่ผู้เชี่ยวชาญกำหนดไว้ในแต่ละระดับความสามารถก็จะมีสองกลุ่ม คือกลุ่มที่ผู้เชี่ยวชาญมั่นใจและกลุ่มที่ผู้เชี่ยวชาญไม่มั่นใจ รอบที่ 2 ผู้เชี่ยวชาญร่วมกันอภิปรายกลุ่มย่อยเกี่ยวกับข้อสอบที่ผู้เชี่ยวชาญยังไม่มั่นใจในรอบที่ 1 และกำหนดคะแนนจุดตัดอีกครั้ง โดยอาศัยหลักการเดียวกับรอบที่ 1 และรอบที่ 3 ผู้เชี่ยวชาญร่วมกันอภิปรายกลุ่มใหญ่เกี่ยวกับข้อสอบที่ผู้เชี่ยวชาญยังไม่มั่นใจในรอบที่ 2 และกำหนดคะแนนจุดตัดอีกครั้ง โดยอาศัยหลักการเดียวกับรอบที่ 1

วิธีการวิเคราะห์กลุ่ม หมายถึง วิธีการกำหนดคะแนนจุดตัดโดยใช้เทคนิคทางสถิติ ในการจัดกลุ่มนักเรียนโดยอาศัยผลจากการตอบข้อสอบคณิตศาสตร์และวิชาวิทยาศาสตร์สำหรับการทดสอบ

ทางการศึกษาระดับชาติขั้นพื้นฐาน โดยนักเรียนที่มีผลการตอบข้อสอบเหมือนกันหรือคล้ายกัน จะถูกจัดให้อยู่ในกลุ่มเดียวกัน ส่วนนักเรียนที่มีผลการตอบข้อสอบแตกต่างกัน จะถูกจัดให้อยู่ต่างกลุ่มกัน โดยตัวแปรในการแบ่งกลุ่ม ได้แก่ คะแนนรายมาตรฐานการเรียนรู้

ผู้มีความสามารถคาบเส้น หมายถึง ผู้ที่มีความรู้และมีทักษะขั้นต่ำของแต่ละระดับความสามารถ

ผู้เชี่ยวชาญ หมายถึง ครูผู้สอนระดับมัธยมศึกษาตอนปลาย ที่สอนวิชาคณิตศาสตร์ วิชาฟิสิกส์ วิชาเคมี และวิชาชีววิทยา โดยมีประสบการณ์การสอนไม่น้อยกว่า 5 ปี

กลุ่มผู้เชี่ยวชาญ หมายถึง กลุ่มของครูผู้สอนระดับมัธยมศึกษาตอนปลายที่ถูกแบ่งเป็นกลุ่มย่อย เพื่อเข้าร่วมการกำหนดคะแนนจุดตัดด้วยวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซี

รอบของการกำหนดคะแนนจุดตัด หมายถึง จำนวนรอบในการพิจารณากำหนดคะแนนจุดตัด

ความสามารถทางคณิตศาสตร์ หมายถึง ความสามารถของนักเรียนที่ได้จากการตอบแบบสอบวิชาคณิตศาสตร์สำหรับการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน ของนักเรียนชั้นมัธยมศึกษาปีที่ 6

ความสามารถทางวิทยาศาสตร์ หมายถึง ความสามารถของนักเรียนที่ได้จากการตอบแบบสอบวิชาวิทยาศาสตร์สำหรับการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน ของนักเรียนชั้นมัธยมศึกษา ปีที่ 6

คุณภาพของคะแนนจุดตัด หมายถึง ความเหมาะสมของคะแนนจุดตัดที่กำหนดโดยวิธีบุ๊คมาร์ค วิธีแองกอฟแบบใช่/ไม่ใช่ และวิธีเอบีซี โดยพิจารณาจากความตรงและความเที่ยง

ความตรงของคะแนนจุดตัด หมายถึง คุณสมบัติของคะแนนจุดตัดที่สะท้อนถึงความรู้และทักษะที่นักเรียนในแต่ละระดับความสามารถจำเป็นต้องมีได้อย่างเหมาะสม โดยพิจารณาจากความตรง 5 ด้าน ดังนี้

1. ความตรงของปัจจัยนำเข้า หมายถึง คุณภาพของเป้าหมาย คำอธิบายระดับความสามารถ ผู้เชี่ยวชาญ วิธีการกำหนดคะแนนจุดตัด และสารสนเทศที่ใช้ประกอบการกำหนดคะแนนจุดตัด ซึ่งจะเป็นการประเมินโดยผู้เชี่ยวชาญ โดยตรวจสอบในประเด็นที่เสนอโดย Kane

(1994) 5 ประเด็นได้แก่ ความถูกต้อง ความเป็นไปได้ ความเหมาะสม ความเข้าใจและความมั่นใจของผู้เชี่ยวชาญ และความเป็นประโยชน์ของเอกสาร

1.1 ความถูกต้อง หมายถึง การดำเนินการมีลำดับขั้นตอนที่ถูกต้อง และเอกสารประกอบการดำเนินการ มีเนื้อหาที่ถูกต้อง สื่อความหมาย

1.2 ความเป็นไปได้ หมายถึง วิธีการดำเนินการสามารถปฏิบัติได้จริงและง่ายต่อการดำเนินการ

1.3 ความเหมาะสม หมายถึง เป้าหมาย การดำเนินการ และเอกสารประกอบการดำเนินการ มีความสอดคล้องและเชื่อมโยงกัน

1.4 ความเข้าใจและความมั่นใจของผู้เชี่ยวชาญ หมายถึง ผู้เชี่ยวชาญมีความเข้าใจเกี่ยวกับความสำคัญ เป้าหมาย คำนิยาม การดำเนินการ และเอกสารประกอบการดำเนินการ และ มีความมั่นใจต่อคะแนนจุดตัดที่กำหนด

1.5 ความเป็นประโยชน์ของเอกสาร หมายถึง เอกสารประกอบการดำเนินการเป็นประโยชน์ต่อการกำหนดคะแนนจุดตัดของผู้เชี่ยวชาญ

2. ความตรงของกระบวนการตัดสินใจเพื่อกำหนดคะแนนจุดตัด หมายถึง คุณภาพของการตัดสินใจเพื่อกำหนดคะแนนจุดตัดของผู้เชี่ยวชาญที่สะท้อนว่าคะแนนจุดตัดที่กำหนดขึ้นมาจากกระบวนการคิด และการตัดสินใจที่มีหลักการ และเหตุผลสนับสนุนที่สมเหตุสมผล ปราศจากความกังวลและปัญหาที่อาจส่งผลต่อคุณภาพของคะแนนจุดตัด

3. ความตรงของผลลัพธ์ หมายถึง ความสอดคล้องกันของคะแนนจุดตัดที่กำหนดขึ้นด้วยวิธีที่ต่างกัน

4. ความตรงของผลที่ตามมา หมายถึง ความถูกต้องในการกำหนดกลุ่มผู้สอบตามระดับความสามารถของคะแนนจุดตัด

ความเที่ยงของคะแนนจุดตัด หมายถึง คุณสมบัติในด้านความคงเส้นคงวาของการกำหนดคะแนนจุดตัด ประกอบด้วย

1. ค่าสัมประสิทธิ์การสรุปอ้างอิง หมายถึง การวิเคราะห์ความคงเส้นคงวาของคะแนนจุดตัดโดยการใช้ทฤษฎีการสรุปอ้างอิง ซึ่งคำนวณโดยการวิเคราะห์ความแปรปรวนโดยใช้โปรแกรมสำเร็จรูป GENOVA โดยมีการกำหนดรายละเอียดดังนี้

1.1 เอกภพ หมายถึง เงื่อนไขทั้งหมดในการวัด ในการศึกษาครั้งนี้ ได้แก่ ผู้เชี่ยวชาญทุกคน และกลุ่มผู้เชี่ยวชาญทุกกลุ่ม

1.2 ฟาเซท หมายถึง ชุดของเงื่อนไขการวัดที่คล้ายคลึงกัน ในการศึกษาครั้งนี้ได้แก่ ผู้เชี่ยวชาญ กลุ่มผู้เชี่ยวชาญ และรอบในการตัดสิน

1.3 เงื่อนไขการวัด หมายถึง สถานการณ์ของการตัดสินใจ ในการศึกษาครั้งนี้ได้แก่ ผู้เชี่ยวชาญแต่ละคน กลุ่มผู้เชี่ยวชาญแต่ละกลุ่ม และการกำหนดคะแนนจุดตัดแต่ละรอบ

2. ความเที่ยงภายในผู้เชี่ยวชาญ หมายถึง ความสอดคล้องของคะแนนจุดตัดที่กำหนดโดยผู้เชี่ยวชาญในแต่ละรอบ พิจารณาจากค่าสัมประสิทธิ์สหสัมพันธ์ภายในกลุ่ม วิเคราะห์โดยใช้โปรแกรมคอมพิวเตอร์สำเร็จรูป SPSS for Windows

3. ความเที่ยงระหว่างผู้เชี่ยวชาญ หมายถึง ความสอดคล้องของคะแนนจุดตัดระหว่างผู้เชี่ยวชาญ พิจารณาจากค่าสัมประสิทธิ์สหสัมพันธ์ภายในกลุ่ม วิเคราะห์โดยใช้โปรแกรมคอมพิวเตอร์สำเร็จรูป SPSS for Windows

ความคลาดเคลื่อนของการกำหนดกลุ่ม หมายถึง ความผิดพลาดที่เกิดจากการนำคะแนนจุดตัดไปใช้เป็นเกณฑ์ในการกำหนดระดับความสามารถของผู้สอบ โดยความคลาดเคลื่อนมี 2 แบบ ได้แก่ ความคลาดเคลื่อนแบบลบ และความคลาดเคลื่อนของแบบบวก

ความคลาดเคลื่อนแบบลบ หมายถึง ความผิดพลาดที่เกิดจากการนำคะแนนจุดตัดไปใช้เป็นเกณฑ์ในการการตัดสินเกี่ยวกับระดับความสามารถของผู้สอบ โดยผลการกำหนดกลุ่มผู้สอบมีความผิดพลาดในลักษณะที่ในสภาพความเป็นจริงแล้วผู้สอบเป็นผู้ที่มีความรู้พอที่จะอยู่ในระดับความสามารถนั้น ๆ แต่เมื่อใช้คะแนนจุดตัดเป็นเกณฑ์ในการกำหนดกลุ่มผลจากการกำหนดกลุ่มปรากฏว่าเป็นผู้ที่มีความรู้ไม่เพียงพอที่จะอยู่ในระดับความสามารถนั้น ๆ

ความคลาดเคลื่อนของแบบบวก หมายถึง ความผิดพลาดเกิดจากการนำคะแนนจุดตัดไปใช้เป็นเกณฑ์ในการการตัดสินเกี่ยวกับระดับความสามารถของผู้สอบโดยผลการกำหนดกลุ่มผู้สอบมีความผิดพลาดในลักษณะที่ในสภาพความเป็นจริงแล้วเป็นผู้ที่ไม่มีความรู้พอที่จะอยู่ในระดับความสามารถนั้น ๆ แต่เมื่อใช้คะแนนจุดตัดเป็นเกณฑ์ในการกำหนดกลุ่มผลจากการกำหนดกลุ่มปรากฏว่าเป็นผู้ที่มีความรู้เพียงพอที่จะอยู่ในระดับความสามารถนั้น ๆ ได้

ประโยชน์ที่คาดว่าจะได้รับ

ประโยชน์ในเชิงวิชาการ

1. ทำให้ทราบถึงคุณภาพของคะแนนจุดตัดที่กำหนดโดยวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซี

2. ทำให้ทราบจำนวนผู้เชี่ยวชาญ จำนวนกลุ่มผู้เชี่ยวชาญที่เหมาะสม สำหรับการกำหนดคะแนนจุดตัดโดยวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซี

ประโยชน์ในด้านการนำไปใช้

1. สถาบันทดสอบทางการศึกษาแห่งชาติสามารถพิจารณาเลือกใช้วิธีการกำหนดคะแนนจุดตัด จำนวนผู้เชี่ยวชาญ จำนวนกลุ่มผู้เชี่ยวชาญที่เหมาะสมไปประยุกต์ใช้ในการกำหนดคะแนนจุดตัดสำหรับวิชาอื่นๆ ของทุกระดับต่อไป

2. สถาบันทดสอบทางการศึกษาแห่งชาติได้สารสนเทศเกี่ยวกับคะแนนจุดตัดสำหรับแต่ละดับความสามารถทางคณิตศาสตร์และวิทยาศาสตร์ เพื่อใช้ในการพิจารณาตัดสินเกี่ยวกับระดับความสามารถของนักเรียน

3. นักเรียน ครู นักวิชาการด้านการศึกษารวมถึงผู้ที่มีส่วนเกี่ยวข้องในการจัดการศึกษาได้สารสนเทศเพื่อการตีความคะแนนจากการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐานวิชาคณิตศาสตร์และวิชาวิทยาศาสตร์ ทำให้สามารถนำสารสนเทศที่ได้ไปใช้ประโยชน์ได้มากขึ้น

บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

ผู้วิจัยได้แบ่งการนำเสนอเนื้อหาที่ได้จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องเป็น 4 ตอน ดังนี้

ตอนที่ 1 การกำหนดคะแนนจุดตัด ประกอบด้วย

1.1 ความเป็นมาของการกำหนดคะแนนจุดตัด

1.2 ความหมายของคะแนนจุดตัด

1.3 ความเข้าใจเบื้องต้นเกี่ยวกับคะแนนจุดตัด

1.4 ขั้นตอนสำคัญของการกำหนดคะแนนจุดตัด

1.5 วิธีการกำหนดคะแนนจุดตัด

1.5.1 วิธีแบบสอบเป็นศูนย์กลาง (test-center methods)

1.5.1.1 วิธีแองกอฟ (Angoff method)

1.5.1.2 วิธีบุ๊คมาร์ค (Bookmark Method)

1.5.1.3 วิธีของอีเบล (Ebel's method)

1.5.1.4 วิธีของนีเดลสกี (Nedelsky method)

1.5.1.5 วิธีเอบีซี (ABC method)

1.5.2 วิธีผู้สอบเป็นศูนย์กลาง (Examinee-centered methods)

1.5.2.1 วิธีกลุ่มคาบเส้น (Borderline-group method)

1.5.2.2 วิธีกลุ่มตรงข้าม (Contrasting-group method)

1.5.2.3 วิธีเลือกงานผู้สอบ (Examinee Paper Selections Method)

1.5.2.4 วิธีองค์รวม (Holistic Method)

1.5.2.5 วิธีการวิเคราะห์กลุ่ม (Cluster analysis)

1.6 การเลือกวิธีการกำหนดคะแนนจุดตัด

1.7 การตรวจสอบคุณภาพของคะแนนจุดตัด

1.8 งานวิจัยที่เกี่ยวข้อง

ตอนที่ 2 ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด

ตอนที่ 3 ทฤษฎีการตอบสนองข้อสอบ

ตอนที่ 4 การทดสอบทางการศึกษาแห่งชาติขั้นพื้นฐาน

รายละเอียดดังแสดงต่อไปนี้

ตอนที่ 1 การกำหนดคะแนนจุดตัด

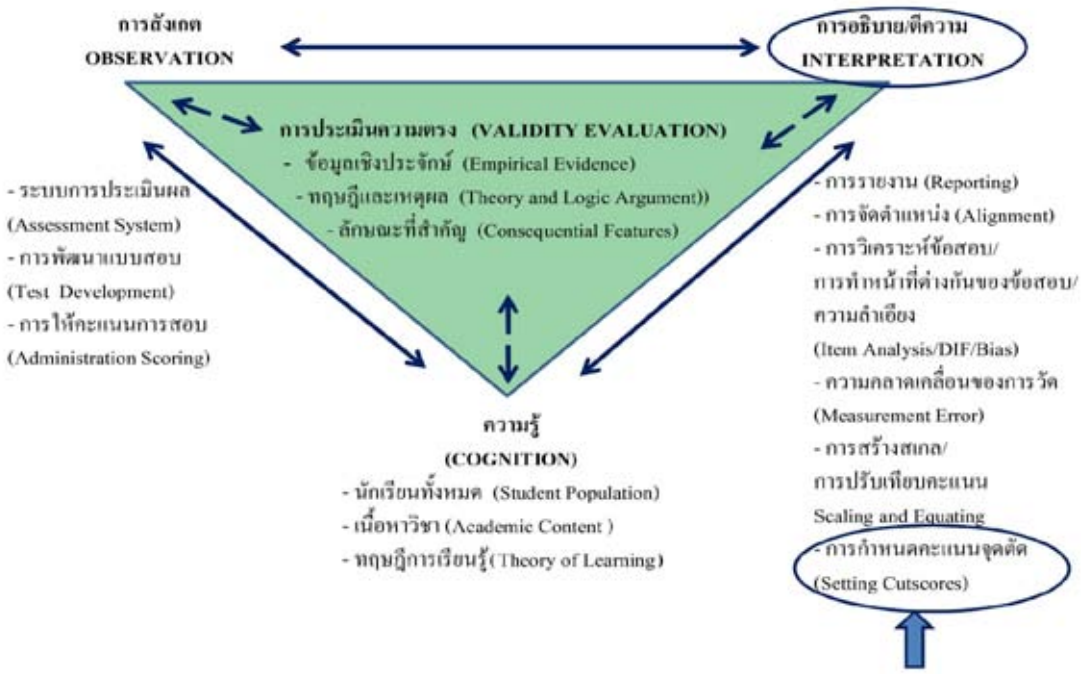
1.1 ความเป็นมาของการกำหนดคะแนนจุดตัด

การกำหนดคะแนนจุดตัด (Setting Cutscores) เป็นการกำหนดคะแนนเพื่อใช้เป็นขีดจำกัด (Threshold) ในการแบ่งคะแนนสอบตามระดับความสามารถของผู้สอบ (Towles-Reeves, 2008) ผู้เชี่ยวชาญบางท่านใช้คำว่า มาตรฐาน (Standard) ในความหมายเดียวกับคำว่า คะแนนจุดตัด (Cutscores) (Zieky et al., 2008: 1) จึงมีการเรียกกระบวนการกำหนดคะแนนจุดตัดว่าการกำหนดมาตรฐาน (Standard Setting) ซึ่งการกำหนดคะแนนจุดตัดมีวิวัฒนาการมากว่า 60 ปี โดยคะแนนจุดตัดถูกนำมาใช้เพื่อการตีความหมายและอธิบายคะแนนสอบโดยเฉพาะการทดสอบแบบอิงเกณฑ์ (Criterion referenced testing) ตั้งแต่ ปี ค.ศ.1970 ซึ่งหากจะพิจารณาถึงวิวัฒนาการของการกำหนดคะแนนจุดตัด Zieky (1994) กล่าวว่าสามารถจำแนกวิวัฒนาการของการกำหนดคะแนนจุดตัด ได้เป็น 4 ยุค หลัก ๆ ดังนี้

ยุคที่ 1 ยุคแห่งความไร้เดียงสา (Age of innocence) เริ่มตั้งแต่มีการบันทึกจนถึง ค.ศ. 1960 ยุคนี้เป็นยุคที่ไม่ให้ความสนใจกับวิธีการได้มาซึ่งมาตรฐานหรือคะแนนจุดตัด ดังนั้นในยุคนี้จึงนิยมกำหนดคะแนนจุดตัดด้วย 3 วิธีการต่อไปนี้คือ วิธีแรกเป็นวิธีการแบบอิงประเพณีนิยม (based on tradition) คือเป็นการกำหนดคะแนนจุดตัดตามค่าในอดีตที่นิยมใช้ เช่น กำหนดคะแนนจุดตัด เป็น 70 % ของคะแนนที่ตอบถูก วิธีที่สองเป็นวิธีแบบอิงผู้มีอำนาจ หรือผู้ที่น่าเชื่อถือในวงการนั้น (based on authority) และวิธีสุดท้ายเป็นวิธีแบบอิงความพอดี (based on goldilocks) เช่น กำหนดคะแนนจุดตัด เป็น 80 % ของคะแนนที่ตอบถูก เพราะเห็นว่า 70 % เป็นค่าที่น้อยเกินไป และ 90 % ก็เป็นค่าที่มากเกินไป เป็นต้น

ยุคที่ 2 ยุคแห่งการตื่นรู้ (Age of awakening) เริ่มตั้งแต่ ค.ศ. 1961 เป็นต้นมา ยุคนี้เป็นยุคที่การประเมินผลสัมฤทธิ์แบบอิงเกณฑ์เริ่มได้รับความสนใจในแวดวงการศึกษา จึงทำให้คนเริ่ม

มองเห็นความสำคัญของเกณฑ์หรือคะแนนจุดตัดมากขึ้นและพร้อมที่จะพัฒนาวิธีการกำหนดคะแนนจุดตัดที่มีความเหมาะสมขึ้นมา เป็นยุคที่มองว่าวิธีการได้มาซึ่งเกณฑ์นั้นมีความสำคัญที่จะต้องมีการศึกษาอย่างถ่องแท้ เพื่อให้สามารถกำหนดเกณฑ์ได้อย่างเหมาะสมและน่าเชื่อถือ ซึ่งเป็นยุคที่มองว่าการกำหนดคะแนนจุดตัดเป็นส่วนหนึ่งของกระบวนการวัดและประเมินผลการเรียนรู้ ดังแสดงในภาพต่อไปนี้



ภาพที่ 2.1 บทบาทของการกำหนดคะแนนจุดตัดในกระบวนการวัดและประเมินผลการเรียนรู้
ที่มา : Marion และ Pellegrino (2006)

จากภาพที่ 2.1 การกำหนดคะแนนจุดตัดมีบทบาทเพื่อการตีความหมายของคะแนนสอบในกระบวนการวัดและประเมินผลการเรียนรู้ โดยใช้เป็นเกณฑ์ในการแบ่งคะแนนสอบเป็นกลุ่มๆ ซึ่งคะแนนแต่ละกลุ่มก็จะสามารถอธิบายความหมายตามเป้าหมายของการกำหนดคะแนนจุดตัด

ยุคที่ 3 ยุคแห่งความท้อแท้ (Age of disillusionment) ยุคนี้เป็นยุคที่มีความคาบเกี่ยวกันกับยุคแห่งการตื่นรู้ ซึ่งยุคนี้ได้รับอิทธิพลโดยตรงจากบทวิจารณ์เกี่ยวกับวิธีการกำหนดคะแนนจุดตัดของ Glass (1978) ซึ่งกล่าวว่า ธรรมชาติของการกำหนดคะแนนจุดตัดถือเป็นอัตวิสัย (Subjective) กล่าวคือ เป็นกระบวนการที่อิงอยู่กับมุมมองหรือความคิดเห็นของบุคคล ซึ่งเป็นแนวคิดที่มองว่าคะแนนจุดตัดที่ได้ขึ้นอยู่กับผู้รู้แต่ละคน สิ่งที่ใช่และยอมรับได้สำหรับคนหนึ่ง ๆ อาจจะไม่ใช่และยอมรับได้สำหรับคนอื่นๆ จากบทความนี้ทำให้นักวิชาการหลายท่านเริ่มรู้สึกว่าวิธีการกำหนดคะแนนจุดตัดไม่ได้เป็นวิธีการที่ดีอย่างที่คิด การที่บทวิจารณ์ฉบับนี้มีอิทธิพลอย่างมากต่อนักวิชาการทั้งหลาย มีสาเหตุมาจากหลาย ๆ ปัจจัย ดังต่อไปนี้

ปัจจัยแรก คือ มนุษย์มีความคิดที่หยั่งรากลึกเกี่ยวกับความเชื่อในความจริงสูงสุด (Absolute truth) มนุษย์จึงเชื่อว่าคะแนนจุดตัดควรจะเป็นสากลและเป็นมาตรฐานเดียวกัน มีความเที่ยงแท้แน่นอน

ปัจจัยที่สอง คือ การให้ค่าความสำคัญกับคะแนนจุดตัดที่กำหนดขึ้นว่ามีผลกระทบต่อผู้ที่มีส่วนเกี่ยวข้องอย่างมาก ดังนั้นผู้ที่กำหนดคะแนนจุดตัดต้องสามารถอธิบายการได้มาซึ่งคะแนนจุดตัดที่มีเหตุผลที่ยอมรับได้

ยุคที่ 4 ยุคแห่งการยอมรับความจริง (Age of realistic acceptance) ยุคนี้เริ่มต้นเมื่อ ค.ศ. 1983 ถึงปัจจุบัน เป็นยุคที่คนเริ่มเปลี่ยนความคิดและความเชื่อในความจริงสูงสุดไปยอมรับแนวคิดและความเชื่อที่ว่าความจริงสูงสุดที่ไม่เปลี่ยนแปลงไม่มี ความจริงคือ การตีความ (Truth is an interpretation) คือการให้ค่าของสิ่งต่างๆ ความจริงทุกความจริงสัมพันธ์กับมนุษย์ มนุษย์เป็นผู้สร้าง “ความจริง” นั้นๆ ขึ้นมา ดังนั้นจึงได้ข้อสรุปที่ว่าคะแนนจุดตัดที่เป็นสากล แน่นอนตายตัวสำหรับทุกคนไม่มี ยุคนี้จึงเป็นยุคที่มีการพัฒนาวิธีการกำหนดคะแนนจุดตัดขึ้นมาจำนวนมาก นักวิชาการจึงพยายามพัฒนาวิธีการกำหนดคะแนนจุดตัดที่สามารถกำหนดคะแนนจุดตัดที่มีความแม่นยำในการจำแนกผู้สอบตามระดับความสามารถและมีความน่าเชื่อถือ ในปี ค.ศ. 1986 มีวิธีที่ใช้ในการกำหนดคะแนนจุดตัดทั้งสิ้น 38 วิธี (Berk, 1986) ต่อมาในปี ค.ศ.1996 พบว่ามีวิธีการกำหนดคะแนนจุดตัด 50 วิธี (Berk, 1996)

ในปี ค.ศ. 2001 ประเทศสหรัฐอเมริกามีการกำหนดคะแนนจุดตัดอย่างกว้างขวางมาก เนื่องจากรัฐบาลกลางของสหรัฐอเมริกาออกกฎหมายปฏิรูปการศึกษาที่ชื่อว่า No Child Left Behind

ที่มีผลบังคับทั้งโรงเรียนประถมศึกษาและมัธยมศึกษาทั่วประเทศ โดยมีสาระสำคัญว่านักเรียนเกรด 3–8 จะต้องมีพัฒนาการทางการเรียนด้านการอ่านและการคำนวณ จึงมีการทดสอบเพื่อวัดผลของพัฒนาการทางการเรียนเป็นรายปี เรียกว่า Adequate Yearly Progress ดังนั้นจึงมีกำหนดคะแนนจุดตัดเพื่อใช้เป็นเกณฑ์ในตีความหมายของคะแนนสอบ (Zieky & Perie, 2006)

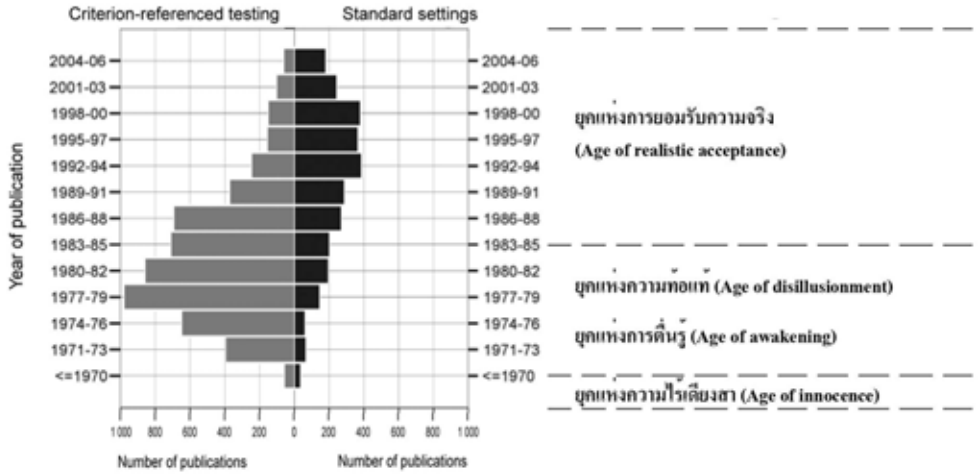
ทิศทางการพัฒนาวิธีการกำหนดคะแนนจุดตัดได้รับอิทธิพลมาจากการเปลี่ยนแปลงของสิ่งต่างๆ ดังต่อไปนี้

1) อิทธิพลจากการเปลี่ยนแปลงความนิยมในรูปแบบข้อสอบ (Item format) กล่าวคือ ยุคนี้เป็นช่วงที่คนเริ่มเปลี่ยนความนิยมจากข้อสอบแบบเลือกตอบไปเป็นข้อสอบแบบตอบสั้น ดังนั้นแนวทางการพัฒนาวิธีการกำหนดคะแนนจุดตัดจึงเป็นไปเพื่อใช้สำหรับการกำหนดคะแนนจุดตัดสำหรับข้อสอบแบบตอบสั้น

2) อิทธิพลจากการเปลี่ยนแปลงวิธีการตีความผลจากการสอบ กล่าวคือ แต่เดิมผลจากการสอบจะถูกนำมาตีความเพื่อแบ่งผู้สอบเป็น 2 กลุ่ม (Dichotomous) เช่น คนที่สอบได้กับคนที่สอบตก (pass/fail) เป็นต้น สำหรับยุคนี้เริ่มนิยมนำผลจากการสอบมาตีความเพื่อแบ่งผู้สอบเป็น หลาย ๆ กลุ่ม (Polytomous) มากขึ้น เพื่อให้สามารถอธิบายผลจากการสอบได้ละเอียดยิ่งขึ้น ดังนั้นแนวทางการพัฒนาวิธีการกำหนดคะแนนจุดตัดจึงเป็นไปเพื่อใช้สำหรับการกำหนดคะแนนจุดตัดหลาย ๆ ค่า

3) อิทธิพลจากการเปลี่ยนแปลงแนวคิดเกี่ยวกับการวิเคราะห์ข้อสอบ กล่าวคือ ยุคนี้เป็นยุคที่ทฤษฎีการทดสอบแนวใหม่ เช่น ทฤษฎีการตอบสนองข้อสอบ (Item response theory) เริ่มถูกนำมาใช้แทนทฤษฎีการทดสอบแบบดั้งเดิม (Classical test theory) ดังนั้นแนวทางการพัฒนาวิธีการกำหนดคะแนนจุดตัดจึงเป็นการบูรณาการความรู้ของทฤษฎีการทดสอบแนวใหม่มาใช้มากขึ้น

Kaftandjieva (2010) ได้แสดงจำนวนงานวิจัยที่ศึกษาเกี่ยวกับการทดสอบแบบอิงเกณฑ์ และการกำหนดคะแนนจุดตัด ในแต่ละยุคของวิวัฒนาการของการกำหนดคะแนนจุดตัด



ภาพที่ 2.2 จำนวนงานวิจัยเกี่ยวกับการทดสอบแบบอิงเกณฑ์ และการกำหนดคะแนนจุดตัดที่ถูกตีพิมพ์ในวารสารทางวิชาการในแต่ละยุคของวิวัฒนาการของการกำหนดคะแนนจุดตัด
ที่มา: ปรับปรุงมาจาก Kaftandjieva (2010)

จากภาพที่ 2.2 การกำหนดคะแนนจุดตัดเกิดขึ้นมาพร้อมกับความนิยมในการใช้การทดสอบแบบอิงเกณฑ์ กล่าวคือ การกำหนดคะแนนจุดตัดมีขึ้นในยุคแห่งการตื่นรู้ แต่ปริมาณงานวิจัยยังคงมีไม่มากนักเนื่องจากยุคแห่งการตื่นรู้กับยุคแห่งความท้อแท้ยังคงมีความคาบเกี่ยวกัน ทำให้การกำหนดคะแนนจุดตัดไม่ได้รับความสนใจจากนักวิชาการมากนัก แต่ช่วงของยุคนี้เป็นช่วงเวลาที่งานวิจัยเกี่ยวกับการทดสอบแบบอิงเกณฑ์มีปริมาณการตีพิมพ์สูงที่สุด หลังจากนั้นก็ลดลงมาเรื่อย ๆ จนมาถึงยุคแห่งการยอมรับความจริง นักวิชาการหันมาให้ความสนใจกับการกำหนดคะแนนจุดตัดมากขึ้น ซึ่งเห็นได้จากจำนวนงานวิจัยเกี่ยวกับการกำหนดคะแนนจุดตัดที่เพิ่มมากขึ้นในยุคนี้

1.2 ความหมายของคะแนนจุดตัด

จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องกับการกำหนดคะแนนจุดตัดมีนักวิชาการได้ให้ความหมายเกี่ยวกับคะแนนจุดตัดไว้ดังต่อไปนี้

Berk (1986) ได้ให้ความหมายของคะแนนจุดตัดไว้ว่าเป็นจุดแบ่งผู้สอบออกเป็น 2 กลุ่มหรือมากกว่า 2 กลุ่มเช่น กลุ่มผ่านกับกลุ่มไม่ผ่าน กลุ่มที่มีความรู้ในระดับพื้นฐาน กลุ่มที่มีความรู้ในระดับเชี่ยวชาญและกลุ่มที่มีความเชี่ยวชาญพิเศษ เป็นต้น

Morgan และ Michaelides (2005) ให้ความหมายของคะแนนจุดตัดว่าเป็นจุดบนการแจกแจงของคะแนนทั้งหมดที่แบ่งคะแนนเป็นสองกลุ่ม หรือมากกว่าสองกลุ่ม

Kurpuis (2006) กล่าวว่า คะแนนจุดตัดคือ คะแนนการผ่านที่ต่ำที่สุด

Bejar (2008) กล่าวว่า คะแนนจุดตัด คือ คะแนนหรือจุดที่ใช้แบ่งผู้สอบออกเป็นสองกลุ่มหรือมากกว่าสองกลุ่มตามคะแนนที่ได้

Zieky และคณะ (2008) ให้ความหมายของคะแนนจุดตัดไว้ว่าเป็นคะแนนหนึ่งบนสเกลของคะแนนสอบทั้งหมดที่กำหนดเพื่อเป็นเกณฑ์สำหรับแบ่งผู้สอบเป็นกลุ่ม

สรุปได้ว่าคะแนนจุดตัด หมายถึง คะแนนที่ใช้เป็นเกณฑ์ในการแบ่งนักเรียนออกเป็นกลุ่มๆ ตามคะแนนของนักเรียน

1.3 ความเข้าใจเบื้องต้นเกี่ยวกับคะแนนจุดตัด

ผู้ที่จะนำคะแนนจุดตัดมาใช้ควรทำความเข้าใจเกี่ยวกับคะแนนจุดตัดก่อนที่จะนำไปใช้ในการตัดสินใจ เพื่อให้สามารถใช้ได้อย่างถูกต้องและเหมาะสม ผู้วิจัยได้รวบรวมไว้ดังนี้ (Livingston & Zieky, 1982; Halpin, 1987; Kane, 1994; Boursicot et al., 2006; Norcini, 2003; Zieky et al., 2008)

1.3.1 คะแนนจุดตัดมีบทบาทเพื่อการตัดสินใจ

ในการตัดสินใจโดยใช้คะแนนจากการสอบซึ่งเป็นข้อมูลเพียงบางส่วนของผู้สอบ จะต้องหาวิธีการที่ดีที่สุดในการนำคะแนนจากการสอบมาใช้เพื่อการตัดสินใจ เพื่อให้การตัดสินใจเป็นไปอย่างถูกต้องที่สุด หนึ่งในวิธีการนั้นคือการกำหนดคะแนนจุดตัดซึ่งจะมีบทบาทเมื่อมีการตัดสินใจเชิงสัมบูรณ์ (Absolute Decision) คือสินใจโดยอิงกับเกณฑ์ โดยคะแนนจุดตัดจะมีบทบาทเป็นเกณฑ์ประกอบการตัดสินใจ เช่น หากต้องการตัดสินใจว่านักเรียนมีความรู้ตามเกณฑ์ที่ต้องการหรือไม่ก็สามารถทำได้โดยนำผลการสอบมาเปรียบเทียบกับเกณฑ์ที่กำหนดไว้ซึ่งก็คือคะแนนจุดตัดนั้นเอง แต่หากเป็นการตัดสินใจเชิงสัมพัทธ์ (Relative Decision) ก็ไม่จำเป็นต้องใช้คะแนนจุดตัด สามารถตัดสินใจโดยการเปรียบเทียบคะแนนของผู้สอบภายในกลุ่มได้เลย

การใช้คะแนนจุดตัดเพื่อการตัดสินใจ จะทำให้เกิดลักษณะของการตัดสินใจ 4 ลักษณะดังนี้คือ

1) การตัดสินว่าเป็นผู้ไม่ผ่านเกณฑ์ แต่ในสภาพความเป็นจริงแล้วเป็นผู้ที่มีความรู้พอที่จะผ่านเกณฑ์ได้ ซึ่งถือว่าเป็นการตัดสินใจคลาดเคลื่อนแบบไม่ยอมรับ (Error of rejection) หรือเรียกว่า ความคลาดเคลื่อนแบบลบ (False negative)

2) การตัดสินว่าเป็นผู้ผ่านเกณฑ์ แต่ในสภาพความเป็นจริงแล้วเป็นผู้ที่ไม่มีความรู้พอที่จะผ่านเกณฑ์ ซึ่งถือว่าเป็นการตัดสินใจคลาดเคลื่อนแบบยอมรับ (Error of acceptance) หรือเรียกว่า ความคลาดเคลื่อนแบบบวก (False positive)

3) การตัดสินว่าผ่านเกณฑ์และในสภาพความเป็นจริงก็เป็นผู้ที่มีความรู้พอที่จะผ่านเกณฑ์ ถือว่าเป็นการตัดสินที่ถูกต้องไม่มีความคลาดเคลื่อนเกิดขึ้น

4) การตัดสินว่าไม่ผ่านเกณฑ์และในสภาพความเป็นจริงก็เป็นผู้ที่ไม่มีความรู้พอที่จะผ่านเกณฑ์ ถือว่าเป็นการตัดสินที่ถูกต้องไม่มีความคลาเคลื่อนเกิดขึ้น

นอกจากที่กล่าวมาข้างต้นแล้ว คะแนนจุดตัดยังสามารถนำไปใช้ในหลายๆ บทบาท ดังนี้ (Green, 1995)

1) เพื่อการรับรอง (certifiers) เช่น ใช้เพื่อการรับรองการจบการศึกษาและรับรองความสามารถในการทำงาน

2) เพื่อทำนาย (predictors) เช่น ใช้เพื่อคัดเลือกบุคคลเข้าศึกษาต่อในมหาวิทยาลัย

3) เพื่อบรรยาย (descriptors) เช่น ใช้แบ่งกลุ่มผู้สอบตามระดับความสามารถ

4) เพื่อกระตุ้น (motivators) เช่น ใช้เป็นคะแนนที่เป็นเป้าหมายของการจัดการศึกษา หรือการเรียนการสอน เป็นต้น

1.3.2 คะแนนจุดตัดเป็นคะแนนซึ่งขึ้นอยู่กับการตัดสิน

คะแนนจุดตัดเป็นคะแนนที่ได้จากการตัดสินของผู้เชี่ยวชาญซึ่งจะตัดสินโดยอาศัยความรู้และประสบการณ์ในขณะนั้นเท่านั้น ดังนั้นการเลือกผู้เชี่ยวชาญจึงเป็นกระบวนการที่สำคัญมาก โดยต้องเลือกผู้ที่มีความเชี่ยวชาญในเรื่องนั้น ๆ และต้องมีเอกสารประกอบการตัดสินใจที่เพียงพอเพื่อให้คะแนนจุดตัดที่ได้มีความน่าเชื่อถือ

1.3.3 ไม่มีคะแนนจุดตัดจริง

คะแนนจุดตัดที่กำหนดขึ้นเป็นเพียงคะแนนที่ได้จากการตัดสินของผู้เชี่ยวชาญกลุ่มนั้นๆ เท่านั้นและใช้กับการตัดสินใจหนึ่งๆ เฉพาะเท่านั้น

1.3.4 อาจจะมีการกำหนดระดับความสามารถผิดพลาดได้

การใช้คะแนนจุดตัดเพื่อตัดสินใจกำหนดกลุ่มนั้น อาจมีโอกาสในการกำหนดกลุ่มผิดได้ เนื่องจากการที่จะเปลี่ยนจากกลุ่มหนึ่งไปเป็นอีกกลุ่มหนึ่งนั้นคะแนนต่างกันเพียงแค่หนึ่งค่าเท่านั้น ทำให้มีโอกาสกำหนดกลุ่มผิดได้

1.4 ขั้นตอนสำคัญของการกำหนดคะแนนจุดตัด

การกำหนดคะแนนจุดตัดประกอบด้วยขั้นตอนที่ต้องดำเนินการหลายขั้นตอน ซึ่งมีทั้งขั้นตอนที่ต้องดำเนินการก่อนที่จะทำการกำหนดคะแนนจุดตัด ขั้นตอนระหว่างที่ดำเนินการกำหนดคะแนนจุดตัด และขั้นตอนหลังจากที่กำหนดคะแนนจุดตัดเสร็จสิ้นแล้ว ซึ่งมีนักวิชาการได้เสนอขั้นที่สำคัญต่อไปนี้

Cizek และคณะ (2004) ได้เสนอขั้นตอนสำคัญที่ต้องดำเนินการก่อนที่จะกำหนดคะแนนจุดตัด ดังนี้

1) กำหนดจุดมุ่งหมายของการกำหนดคะแนนจุดตัด

2) เลือกผู้เชี่ยวชาญที่เป็นตัวแทนที่ดี และฝึกอบรมผู้เชี่ยวชาญเพื่อให้มีความรู้และความเข้าใจเกี่ยวกับการกำหนดคะแนนจุดตัด

3) กำหนดระดับความสามารถตามจุดมุ่งหมายที่ต้องการนำไปใช้

4) เขียนคำอธิบายระดับความสามารถในแต่ละระดับให้ชัดเจน

5) กำหนดกลุ่มเป้าหมายที่จะนำคะแนนจุดตัดไปใช้เพื่อการจัดกลุ่มตามระดับความสามารถ

Perie (2006) เสนอขั้นตอนที่ต้องให้ความสำคัญเป็นพิเศษในการกำหนดคะแนนจุดตัด ดังนี้

1) การเขียนคำอธิบายระดับความสามารถต้องเขียนให้มีความชัดเจนและถูกต้อง การศึกษาของ Ferdous และ Plake (2005) ที่ศึกษาปัจจัยที่ส่งผลต่อการกำหนดคะแนนจุดตัด พบว่า ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดมีความคิดเห็นสอดคล้องกันว่าปัจจัยที่ส่งผลต่อการกำหนดคะแนนจุดตัด คือ ความชัดเจนของคำอธิบายระดับความสามารถซึ่งหากเขียนไม่ชัดเจนย่อมส่งผลให้การกำหนดคะแนนจุดตัดผิดพลาดได้

2) เลือกวิธีการกำหนดคะแนนจุดตัดที่เหมาะสมกับข้อสอบที่ใช้สอบ

3) พิจารณาความถูกต้องของวิธีการกำหนดคะแนนจุดตัดที่กำหนดไว้

4) เตรียมเอกสารหรือข้อมูลสารสนเทศที่ใช้เพื่อประกอบการกำหนดคะแนนจุดตัด ให้มีความถูกต้องและชัดเจน

Zieky และคณะ (2008) กล่าวว่า การกำหนดคะแนนจุดตัดมีขั้นตอนที่จำเป็นหลายขั้นตอน แต่ก็มีขั้นตอนที่ต้องดำเนินการให้แล้วเสร็จก่อนที่จะมีการกำหนดคะแนนจุดตัด ดังต่อไปนี้

1) ตัดสินใจว่าจะใช้คะแนนจุดตัดเพื่อการตัดสินใจหรือไม่

การตัดสินใจว่าจะใช้หรือไม่ใช้คะแนนจุดตัดถือเป็นขั้นตอนแรกที่สำคัญ ควรตั้งคำถามเพื่อถามตนเองก่อนที่จะเลือกใช้คะแนนจุดตัด โดยมีคำถามดังนี้

1.1) จะใช้คะแนนจุดตัดเพื่อการตัดสินใจในเรื่องใด ?

1.2) มีวิธีการอื่นที่ใช้ในการตัดสินใจนอกจากการใช้คะแนนจุดตัดหรือไม่ ?

1.3) มีความเสียหายที่อาจจะเกิดขึ้นจากการใช้คะแนนจุดตัดหรือไม่ ? อะไรบ้าง ?

1.4) มีความเสียหายที่อาจเกิดจากการใช้วิธีการแบบอื่นเพื่อการตัดสินใจหรือไม่ ? อะไรบ้าง ?

1.5) ข้อดีของการใช้คะแนนจุดตัดเพื่อการตัดสินใจมีอะไรบ้าง ?

1.6) ข้อดีของการใช้วิธีการอื่นเพื่อการตัดสินใจมีอะไรบ้าง ?

1.7) ข้อดีของการใช้คะแนนจุดตัดมีมากกว่าข้อเสียหรือไม่ ?

1.8) วัตถุประสงค์ของการใช้คะแนนจุดตัดคืออะไร ?

สรุปได้ว่าการตัดสินใจเลือกใช้คะแนนจุดตัดเพื่อการตัดสินใจจะต้องมีเหตุผลที่ดีเพียงพอ โดยต้องพิจารณาทั้งผลดี ผลเสียและเปรียบเทียบกับเกณฑ์อื่นๆ ก่อนตัดสินใจ

2) วางแผนล่วงหน้า (Plan ahead)

การกำหนดคะแนนจุดตัดมีกระบวนการที่ค่อนข้างซับซ้อน ซึ่งต้องใช้เวลาและค่าใช้จ่ายมาก เนื่องจากว่าในการกำหนดคะแนนจุดตัดแต่ละครั้งต้องอาศัยผู้เชี่ยวชาญจำนวนมาก โดยผู้เชี่ยวชาญแต่ละท่านจะต้องมีความเชี่ยวชาญในเรื่องที่จะกำหนดคะแนนจุดตัดและก่อนที่จะมีการกำหนดคะแนนจุดตัดจะต้องมีการให้ความรู้เกี่ยวกับขั้นตอน วิธีการกำหนดคะแนนจุดตัดแก่ผู้เชี่ยวชาญ ดังนั้นเพื่อให้การดำเนินการทุกอย่างเป็นไปอย่างถูกต้องและเหมาะสม ควรมีการวางแผนล่างหน้าอย่างรัดกุมเพื่อป้องกันและหลีกเลี่ยงปัญหาที่อาจจะเกิดขึ้น โดยการวางแผนต้องเริ่มจากการระบุกิจกรรมที่ต้องดำเนินการทั้งหมด ระบุจำนวนและคุณสมบัติของผู้เชี่ยวชาญที่ต้องการ รวมทั้งระบุ

อุปกรณ์และแหล่งของอุปกรณ์ที่ต้องใช้ในการดำเนินการทั้งหมด นอกจากนี้ก่อนการดำเนินการกำหนดคะแนนจุดตัดควรมีการอธิบายหลักการและเหตุผลของการกำหนดคะแนนจุดตัดต่อผู้ที่เกี่ยวข้องทั้งหมด เพื่อให้รับทราบและอาจมีข้อเสนอแนะที่เป็นประโยชน์ต่อการดำเนินการ

3) การพัฒนาคำอธิบายระดับความสามารถ

คำอธิบายระดับความสามารถเป็นการเขียนบรรยายถึงความรู้ และทักษะของผู้ซึ่งอยู่ในแต่ละระดับความสามารถต้องมี หากสามารถเขียนคำอธิบายระดับความสามารถได้ดีย่อมเป็นฐานที่ดีในการสร้างข้อสอบ การกำหนดคะแนนจุดตัด และการรายงานผล

4) ตัดสินใจว่าจะใช้คะแนนจุดตัดที่ได้จากคะแนนรวมทั้งหมดหรือคะแนนย่อย

ในกรณีที่ข้อสอบแบ่งออกเป็นข้อสอบย่อยหลาย ๆ ส่วน การให้คะแนนก็สามารถให้แยกกันเป็นส่วนๆ ดังนั้นต้องตัดสินใจว่าจะกำหนดคะแนนจุดตัดจากคะแนนรวมของข้อสอบทั้งฉบับหรือจะกำหนดคะแนนจุดตัดเป็นคะแนนย่อย

5) กำหนดขั้นตอนที่จำเป็นสำหรับการกำหนดคะแนนจุดตัด

ขั้นตอนที่จำเป็นสำหรับการกำหนดคะแนนจุดตัดประกอบด้วย การอภิปรายร่วมกันระหว่างผู้เชี่ยวชาญ การกำหนดจำนวนผู้เชี่ยวชาญ ข้อมูลที่ต้องเตรียมให้ผู้เชี่ยวชาญเพื่อประกอบการกำหนดคะแนนจุดตัด วิธีการรวบรวมและคำนวณคะแนนจุดตัด รายละเอียดดังแสดงต่อไปนี้

5.1) การอภิปรายร่วมกันระหว่างผู้เชี่ยวชาญ

ในการอภิปรายร่วมกันของผู้เชี่ยวชาญนั้น ต้องกำหนดประเด็นที่จะอภิปรายให้ผู้เชี่ยวชาญอภิปรายและกำหนดระยะเวลาสำหรับการอภิปราย ยกตัวอย่างเช่นบางวิธีของการกำหนดคะแนนจุดตัดนั้นจะต้องให้ผู้เชี่ยวชาญร่วมกันอภิปรายเกี่ยวกับความสามารถต่ำสุดที่ผู้สอบต้องมีในแต่ละระดับความสามารถ หรือบางวิธีต้องอภิปรายเกี่ยวกับเนื้อหาที่จะใช้สอบ จำนวนข้อสอบ รวมทั้งอภิปรายเกี่ยวกับความรู้และความสามารถที่นักเรียนต้องมีเพื่อตอบข้อสอบให้ถูกต้องเป็นต้น นอกจากนี้ต้องวางแผนเกี่ยวกับการอภิปรายผลการกำหนดคะแนนจุดตัดเพื่อให้ผู้เชี่ยวชาญได้แสดงความคิดเห็นเกี่ยวกับผลที่ได้และเพื่อเปิดโอกาสให้ผู้เชี่ยวชาญได้ตัดสินใจในการกำหนดคะแนนจุดตัดอีกครั้ง โดยต้องตัดสินใจเกี่ยวกับจำนวนครั้งของการอภิปราย รวมทั้งห้องที่ใช้อภิปราย ซึ่งอาจให้อภิปรายร่วมกันทั้งหมดในห้องใหญ่หรือแบ่งกลุ่มเพื่ออภิปรายเป็นกลุ่มย่อยๆ ซึ่งข้อดีของการร่วมกันอภิปรายทั้งหมดคือทำให้ผู้ตัดสินทุกท่านได้แลกเปลี่ยนความคิดเห็น

ซึ่งกันและกันทั้งหมด ส่วนข้อดีของการอภิปรายกลุ่มย่อยคือง่ายต่อการตัดสินใจหาข้อสรุป เนื่องจากคนน้อย (Livingston & Zieky, 1982; Cizek, & Bunch, 2007)

5.2) จำนวนรอบในการพิจารณาคะแนนจุดตัด

การกำหนดคะแนนจุดตัดแต่ละวิธีนั้น จำนวนรอบในการพิจารณาตัดสินจะแตกต่างกันไป บางวิธีมีการกำหนดคะแนนจุดตัดเพียงรอบเดียว บางวิธีพิจารณาหลายรอบ ซึ่งอาจมีการพิจารณามากถึง 5 รอบ จากการศึกษาส่วนใหญ่จะพบว่าผลการตัดสินเพื่อกำหนดคะแนนจุดตัดจะมีการปรับคะแนนจุดตัดน้อยลงหลังจากที่พิจารณาผ่านไปแล้ว 3 รอบ

5.3) การสรุปคะแนนจุดตัดจากผู้เชี่ยวชาญแต่ละท่าน

ในการรวบรวมคะแนนเพื่อกำหนดเป็นคะแนนจุดตัดนั้นมีวิธีการหลายวิธี ได้แก่ การใช้คะแนนเฉลี่ย ซึ่งเป็นการเฉลี่ยคะแนนจากผู้เชี่ยวชาญทุกท่าน ข้อเสียของการใช้วิธีนี้คือ หากในกรณีที่มีคะแนนที่สูงและต่ำมากๆ ย่อมส่งผลทำให้คะแนนเฉลี่ยที่ได้ไม่เป็นตัวแทนที่ดีนัก ดังนั้นการคำนวณคะแนนจุดตัดโดยการใช้ค่าเฉลี่ยอาจจะมีความคลาดเคลื่อนได้ถ้าคะแนนจุดตัดของผู้เชี่ยวชาญบางท่านมีค่าที่ผิดปกติไปจากกลุ่มหรือที่เรียกว่า Extreme Value ส่วนวิธีที่สองคือการใช้คะแนนมัธยฐาน (Median) เป็นคะแนนจุดตัด ซึ่งทำได้โดยการเรียงคะแนนจุดตัดของผู้เชี่ยวชาญแต่ละท่านจากคะแนนมากไปหาคะแนนน้อยตามลำดับ โดยคะแนนที่อยู่ตรงกลางคือคะแนนมัธยฐาน ข้อดีของการใช้คะแนนมัธยฐาน คือ เป็นการลดผลของคะแนนที่สูงและต่ำเกินไป ส่วนข้อเสีย คือ การละเลยต่อข้อมูลที่ซ้ำๆ กันโดยสนใจเฉพาะคะแนนที่อยู่ตรงกลางเท่านั้น หากข้อมูลมีค่าผิดปกติไปจากกลุ่มมาก คะแนนจุดตัดควรคำนวณโดยใช้ค่า Trimmed Means ซึ่งสามารถคำนวณได้โดยการเรียงลำดับข้อมูลจากน้อยไปมาก แล้วลบข้อมูลที่น้อยที่สุดและมากที่สุดออก จากนั้นคำนวณหาค่าเฉลี่ยของข้อมูลที่เหลือ จำนวนของข้อมูลที่จะถูกลบออกนั้นจะมีประมาณ 5% จนถึง 25% ของข้อมูลทั้งหมด

6) การเตรียมคณะทำงานและอุปกรณ์เพื่อใช้ในการดำเนินการกำหนดคะแนนจุดตัด

6.1) คณะทำงาน

คณะทำงานที่จะช่วยในการดำเนินการกำหนดคะแนนจุดตัด ประกอบด้วยผู้อำนวยความสะดวก (Facilitator) ผู้เชี่ยวชาญด้านเนื้อหา (Subject Matter Expert) ผู้เชี่ยวชาญด้านการทดสอบ (Test Expert) นักสถิติ (Statistician) และ ผู้ที่จัดการเกี่ยวกับเอกสาร (Document

Management) คนหนึ่งคนสามารถทำให้หลายบทบาท ซึ่ง Zieky และคณะ (2008) เสนอว่าในการดำเนินการกำหนดคะแนนจุดตัดแต่ละครั้งควรมีคณะทำงานอย่างน้อย 3 คน คือผู้อำนวยความสะดวก นักสถิติ และ ผู้ที่จัดการเกี่ยวกับเอกสาร โดยผู้อำนวยความสะดวกจะทำหน้าที่ในการดำเนินการประชุมเพื่อกำหนดคะแนนจุดตัด รวมทั้งฝึกอบรมเพื่อให้ความรู้เกี่ยวกับการกำหนดคะแนนจุดตัดแก่ผู้เชี่ยวชาญและคณะทำงานคนอื่นๆ รวมทั้งประสานงานกับสมาชิกที่ร่วมทำงาน ซึ่งผู้อำนวยความสะดวกต้องเป็นผู้ที่มีความเข้าใจในกระบวนการกำหนดคะแนนจุดตัดเป็นอย่างดี และควรเป็นผู้ที่เคยมีประสบการณ์ในการดำเนินการประชุมและการฝึกอบรมเป็นอย่างดี โดยต้องสามารถดำเนินการแต่ละขั้นตอนได้อย่างถูกต้องและเหมาะสมแก่เวลารวมทั้งต้องเป็นผู้ที่สามารถแก้ไขปัญหาเฉพาะหน้าต่างๆ ได้ดี ส่วนบทบาทของนักสถิติคือการคำนวณค่าสถิติต่าง ๆ ที่ต้องใช้เพื่อประกอบการดำเนินการกำหนดคะแนนจุดตัดโดยขึ้นอยู่กับวิธีการกำหนดคะแนนจุดตัดแต่ละวิธีว่าจะต้องใช้ค่าสถิติใดบ้าง เช่น วิธีการกำหนดคะแนนจุดตัดบางวิธีต้องใช้ทฤษฎีการตอบสนองข้อสอบหรืออาจใช้การวิเคราะห์การถดถอยโลจิสติก เป็นต้น ดังนั้นจึงจำเป็นที่จะต้องมีนักสถิติที่มีความรู้เกี่ยวกับการวิเคราะห์ในเรื่องนั้นๆ สำหรับบทบาทของผู้จัดเกี่ยวกับเอกสารนั้นคือเป็นผู้จัดเตรียมและจัดระบบเอกสารที่จำเป็นต้องใช้เพื่อประกอบการดำเนินการกำหนดคะแนนจุดตัดและต้องตรวจสอบความถูกต้องของเอกสารที่จะต้องใช้ พร้อมทั้งเก็บรักษาเอกสารต่างๆ เป็นอย่างดี (Livingston & Zieky, 1982; Cizek & Bunch, 2007)

6.2) สิ่งอำนวยความสะดวก (Facilities)

การดำเนินการกำหนดคะแนนจุดตัดจำเป็นต้องจัดประชุมเพื่อพิจารณาคะแนนจุดตัดโดยผู้เชี่ยวชาญหลายท่าน ดังนั้นต้องมีการเตรียมห้องประชุมที่มีความพร้อมและขนาดห้องที่พอเหมาะกับจำนวนผู้เชี่ยวชาญ ในการกำหนดจุดตัดบางวิธีจำเป็นต้องให้ผู้เชี่ยวชาญประชุมกลุ่มย่อย ดังนั้นต้องเตรียมห้องประชุมขนาดเล็กหลายๆ ห้อง รวมทั้งต้องเตรียมสิ่งอำนวยความสะดวกต่างๆ เช่น กระดาน เครื่องเล่นวิดีโอ อาหารว่าง เป็นต้น นอกจากนี้หากต้องมีการพิจารณากำหนดคะแนนจุดตัดหลายวันต้องมีการเตรียมห้องพักสำหรับผู้เชี่ยวชาญโดยห้องพักและห้องประชุมควรอยู่ใกล้กันเพื่อให้สามารถเดินทางได้สะดวก สำหรับบางการประชุมอาจใช้การประชุมทางไกล (Web conferencing) ซึ่งช่วยประหยัดค่าใช้จ่ายในเรื่องค่าเดินทาง ค่าอาหารและค่าห้องประชุม ซึ่งขึ้นอยู่กับการบริหารจัดการของผู้ดำเนินการกำหนดคะแนนจุดตัด

7) การเลือกผู้ตัดสิน

การคัดเลือกผู้เชี่ยวชาญที่จะมาทำการกำหนดคะแนนจุดตัดนั้น ควรเป็นผู้ที่มีความเชี่ยวชาญในเรื่องที่ต้องกำหนดจุดตัดและควรเป็นตัวแทนที่ดีของผู้เชี่ยวชาญทั้งหมด ซึ่งผู้ดำเนินการกำหนดคะแนนจุดตัดจะต้องระบุคุณลักษณะของผู้เชี่ยวชาญที่ต้องการให้ชัดเจนแล้วจึงดำเนินการเลือกผู้เชี่ยวชาญมากำหนดคะแนนจุดตัด

7.1) ผู้ที่จะเป็นผู้เชี่ยวชาญ

เนื่องจากว่าคะแนนจุดตัดเป็นคะแนนที่ขึ้นอยู่กับการตัดสินของผู้เชี่ยวชาญ ดังนั้นการเลือกผู้เชี่ยวชาญจึงเป็นเรื่องที่สำคัญอย่างยิ่ง เกณฑ์การเลือกผู้เชี่ยวชาญเพื่อกำหนดคะแนนจุดตัดเป็นประเด็นได้รับการกล่าวถึงอย่างแพร่หลาย นักวิชาการได้เสนอหลักเกณฑ์ในการเลือกผู้เชี่ยวชาญ ซึ่งสามารถสรุปได้ 3 แนวคิด ดังนี้ แนวคิดแรก เป็นแนวคิดที่ว่าการกำหนดคะแนนจุดตัดควรสะท้อนการมีส่วนร่วมของผู้มีส่วนได้ส่วนเสียทุกกลุ่ม “แม้ว่าผู้มีส่วนได้ส่วนเสียบางคนอาจไม่มีความเชี่ยวชาญในเรื่องที่จะกำหนดคะแนนจุดตัดก็ตาม” (Kane, 1994: 441) แนวคิดที่สองไม่ยอมรับการเลือกผู้เชี่ยวชาญที่ไม่มีความเชี่ยวชาญในเรื่องที่จะกำหนดคะแนนจุดตัด (Impara, 1997) โดยแนวคิดนี้เน้นว่าผู้เชี่ยวชาญควรเป็นผู้ที่มีความเชี่ยวชาญในเนื้อหาและควรมีความเข้าใจเกี่ยวกับตัวผู้สอบเป็นอย่างดี ส่วนแนวคิดสุดท้าย เป็นแนวคิดที่ว่าผู้เชี่ยวชาญไม่จำเป็นต้องเป็นผู้ที่มีความเชี่ยวชาญในเนื้อหาที่จะทำการกำหนดคะแนนจุดตัด แต่ควรมีความรู้ทั่วๆ ไปเกี่ยวกับเนื้อหาพอสมควร(Norcini, Shea, & Kanya, 1988; Brandon, 2004; Zieky et al., 2008)

7.2) จำนวนผู้เชี่ยวชาญ

จำนวนผู้เชี่ยวชาญที่เหมาะสมสำหรับการกำหนดคะแนนจุดตัดนั้น โดยทั่วๆ ไปกำหนดไว้ที่ 12–18 ท่านและไม่ควรต่ำกว่า 8 ท่านสำหรับการกำหนดคะแนนจุดตัดที่เป็นการตัดสินใจที่สำคัญ (Zieky et al., 2008) ทั้งนี้หากผู้เชี่ยวชาญในเรื่องนั้นๆ มีจำนวนน้อยก็สามารถกำหนดให้ทุกท่านเป็นผู้เชี่ยวชาญตัดสินได้ทั้งหมด และหากผู้เชี่ยวชาญในเรื่องนั้นๆ มีจำนวนมากก็ควรติดต่อเพื่อให้เป็นผู้เชี่ยวชาญมากกว่าจำนวนที่กำหนดไว้เนื่องจากว่าบางท่านอาจไม่สามารถมาประชุมได้หรือบางท่านอาจไม่สามารถร่วมดำเนินการกำหนดคะแนนจุดตัดได้จนสิ้นสุดกระบวนการ

7.3) การฝึกผู้เชี่ยวชาญ

การฝึกเกี่ยวกับวิธีการกำหนดคะแนนจุดตัดที่เลือกใช้ ถือเป็นขั้นตอนที่สำคัญ เนื่องจากผู้เชี่ยวชาญบางท่านอาจจะยังไม่เคยใช้วิธีการกำหนดคะแนนจุดตัดที่เลือก จึงอาจทำให้เกิดการสับสนในการปฏิบัติได้ ดังนั้นจึงควรมีการฝึกอบรมเกี่ยวกับวิธีการกำหนดคะแนนจุดตัดเพื่อให้ผลการตัดสินถูกต้อง ดังนั้นก่อนที่จะมีการดำเนินการกำหนดคะแนนจุดตัดควรมีการอบรมเพื่อทำความเข้าใจกับผู้เชี่ยวชาญเกี่ยวกับกระบวนการและรายละเอียดเกี่ยวกับขั้นตอนวิธีของการกำหนดคะแนนจุดตัดที่เลือกใช้ รวมทั้งรายละเอียดเกี่ยวกับเป้าหมายของการกำหนดคะแนนจุดตัดและอธิบายเกี่ยวกับผลกระทบของการตัดสินที่มีต่อผู้สอบที่ถูกจัดให้อยู่ในแต่ละระดับความสามารถ ดังนั้นก่อนที่จะมีการกำหนดคะแนนจุดตัดควรมีการทดลองปฏิบัติเพื่อซักซ้อมความเข้าใจก่อนที่จะปฏิบัติจริง โดยระยะเวลาที่ใช้เพื่อการอบรมไม่ควรน้อยกว่าครึ่งวัน (Brennan, 2006)

Plake (2008) กล่าวว่า การดำเนินการเพื่อกำหนดคะแนนจุดตัด จะมีการดำเนินการตามลำดับดังนี้

1) กำหนดเป้าหมายของการกำหนดคะแนนจุดตัด เช่น เพื่อรับรองคุณสมบัติ เพื่อจัดกลุ่มนักเรียนตามระดับความสามารถ หรือเพื่อตัดสินผลการสอบว่าได้หรือตก

2) กำหนดผู้เชี่ยวชาญที่จะมากำหนดคะแนนจุดตัด

3) กำหนดกระบวนการประเมินความสามารถของผู้สอบ

4) ฝึกอบรมผู้เชี่ยวชาญเกี่ยวกับการกำหนดคะแนนจุดตัด

5) กำหนดขั้นตอนและวิธีการสร้างข้อคำถาม

6) ประเมินความถูกต้องของคะแนนจุดตัด

Brennan (2006) ได้เสนอขั้นตอนที่สำคัญในการกำหนดคะแนนจุดตัดดังนี้

1) เลือกวิธีที่จะใช้ในการกำหนดคะแนนจุดตัด

หลักเกณฑ์ในการเลือกใช้วิธีที่เหมาะสมจะต้องพิจารณาจากสิ่งต่อไปนี้

1.1) ชนิดของเครื่องมือที่ใช้ในการทดสอบ

1.2) ระยะเวลาและแหล่งข้อมูลที่ใช้สำหรับการกำหนดคะแนนจุดตัด

1.3) ประสบการณ์เกี่ยวกับการกำหนดคะแนนจุดตัดของผู้ที่ทำหน้าที่ในการกำหนดคะแนนจุดตัด

1.4) ระดับความน่าเชื่อถือของผลการกำหนดคะแนนจุดตัด

2) เลือกผู้เชี่ยวชาญและออกแบบกระบวนการกำหนดคะแนนจุดตัด

3) เตรียมคำบรรยายเกี่ยวกับความสามารถแต่ละระดับ

4) ฝึกอบรมผู้เชี่ยวชาญเกี่ยวกับการกำหนดคะแนนจุดตัด

การฝึกอบรมผู้เชี่ยวชาญเกี่ยวกับการกำหนดคะแนนจุดตัดเพื่อช่วยให้ผู้เชี่ยวชาญทุกท่านมีความเข้าใจอย่างถูกต้องและตรงกันเกี่ยวกับการกำหนดคะแนนจุดตัด ซึ่งการอบรมนั้นจะต้องมีรายละเอียดที่ต้องแจ้งให้ผู้เชี่ยวชาญทราบดังนี้

4.1) อธิบายเป้าหมายของการกำหนดคะแนนจุดตัดเพื่อนำไปใช้ในการตัดสินใจในเรื่องใด

4.2) อธิบายเกี่ยวกับขั้นตอนในการกำหนดคะแนนจุดตัด

4.3) นำข้อสอบและการแปลผลการสอบเสนอต่อผู้เชี่ยวชาญ

4.4) อธิบายขั้นตอนและกระบวนการสร้างแบบสอบ

4.5) ให้เวลาแก่ผู้เชี่ยวชาญในการทำข้อสอบเหมือนกับการสอบจริง

4.6) อธิบายเกี่ยวกับข้อมูลที่จะนำเสนอต่อผู้เชี่ยวชาญเพื่อใช้ประกอบการตัดสินกำหนดคะแนนจุดตัด

4.7) เสนอรายละเอียดเกี่ยวกับความสามารถแต่ละระดับ

5) ดำเนินการกำหนดคะแนนจุดตัด

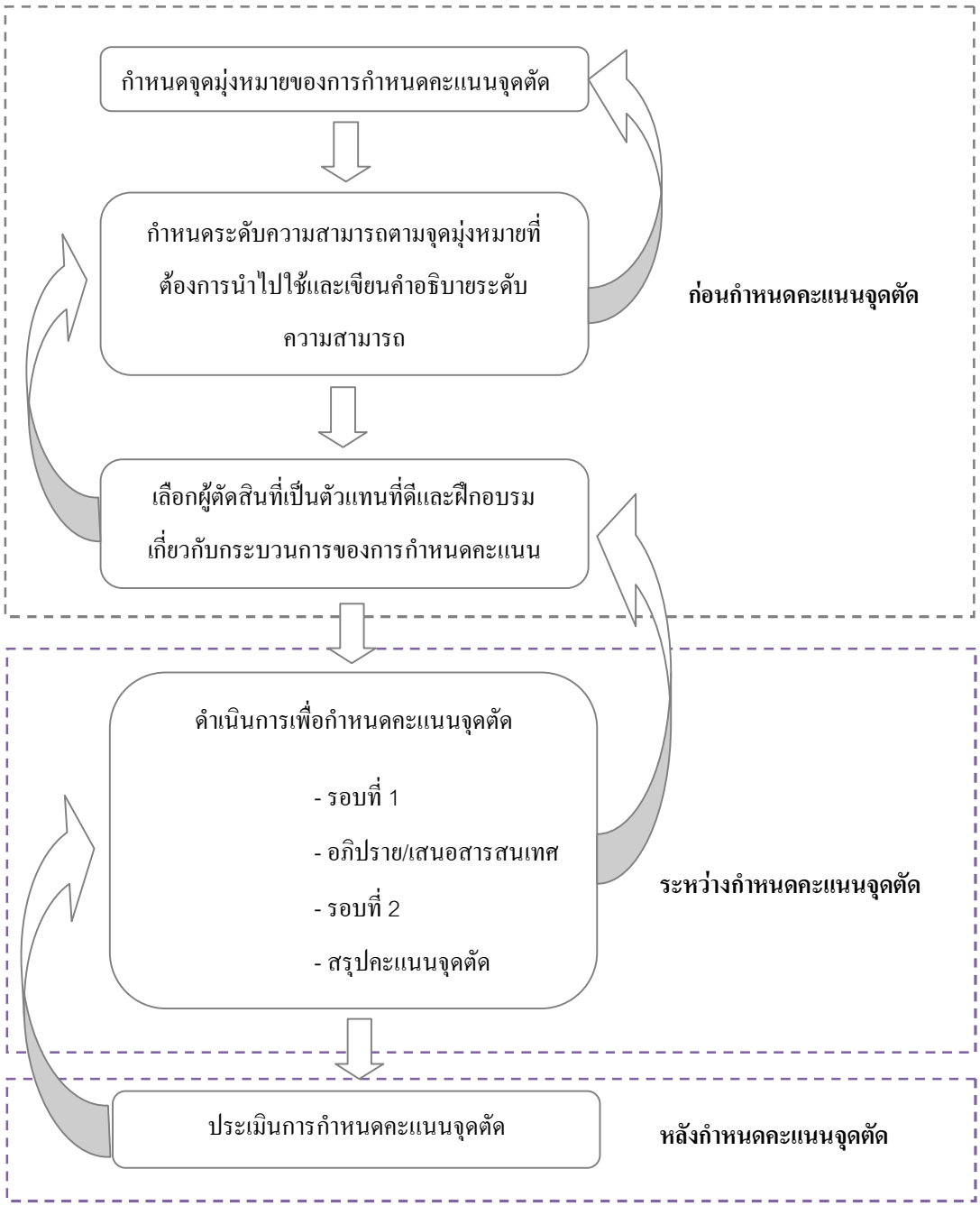
6) เสนอผลการกำหนดคะแนนจุดตัดเพื่อให้ผู้เชี่ยวชาญได้ร่วมกันอภิปรายผล

7) รวบรวมผลการกำหนดคะแนนจุดตัดเพื่อกำหนดเป็นคะแนนจุดตัดขั้นสุดท้าย

หลังจากที่กำหนดคะแนนจุดตัดแล้วต้องประเมินกระบวนการกำหนดคะแนนจุดตัดตั้งแต่เริ่มแรกจนถึงขั้นตอนสุดท้ายเพื่อดูความถูกต้องเหมาะสม (Livingston & Zieky, 1982; Plake, 2008; Zieky et al., 2008)

จากการศึกษาขั้นตอนของการกำหนดคะแนนจุดตัด สรุปได้ว่าการกำหนดคะแนนจุดตัด มีขั้นตอนที่ต้องดำเนินการหลายขั้นตอน ซึ่งต้องดำเนินการอย่างครบถ้วนและถูกต้องเพื่อให้ผลการกำหนดคะแนนจุดตัดมีความแม่นยำและน่าเชื่อถือ โดยขั้นตอนที่สำคัญประกอบด้วยการกำหนดจุดมุ่งหมายของการกำหนดคะแนนจุดตัด กำหนดระดับความสามารถตามจุดมุ่งหมายที่ต้องการ

นำไปใช้และเขียนคำอธิบายแต่ละระดับความสามารถ เลือกผู้เชี่ยวชาญที่เป็นตัวแทนที่ดีและฝึกอบรมเกี่ยวกับกระบวนการของการกำหนดคะแนนจุดตัด ดำเนินการเพื่อกำหนดคะแนนจุดตัด และขั้นตอนสุดท้ายที่สำคัญคือ การประเมินคุณภาพของคะแนนจุดตัด ผู้วิจัยได้สรุปขั้นตอนที่สำคัญในการกำหนดคะแนนจุดตัด ดังแสดงในภาพต่อไปนี้



ภาพที่ 2.3 ขั้นตอนของการกำหนดคะแนนจุดตัด

จากที่กล่าวมาข้างต้น ขั้นตอนที่มีความสำคัญและต้องดำเนินการก่อนที่จะกำหนดจุดตัดคือการพัฒนาคำอธิบายระดับความสามารถ (Performance – Level Descriptors: PLDs) ซึ่งจะเริ่มจากการกำหนดสิ่งต่อไปนี้ ได้แก่

1) กำหนดจำนวนของระดับและชื่อของระดับความสามารถ

2) กำหนดคำอธิบายระดับความสามารถ

3) กำหนดคะแนนจุดตัด

ชื่อและระดับความสามารถที่กำหนดจะขึ้นอยู่กับเป้าหมายของการนำผลไปใช้ ส่วนคำอธิบายระดับความสามารถนั้น เป็นการเขียนบรรยายถึงความรู้ และทักษะของผู้ซึ่งอยู่ในแต่ละระดับความสามารถต้องมี โดยต้องเขียนรายละเอียดให้เห็นความแตกต่างระหว่างแต่ละระดับความสามารถ หากสามารถเขียนคำอธิบายระดับความสามารถได้ดี ย่อมเป็นฐานที่ดีในการประเมินผล นับตั้งแต่การสร้างข้อสอบ การกำหนดคะแนนจุดตัดและการรายงานผล (Perie, 2008)

Fehrmann และคณะ (1991) กล่าวว่า การเขียนคำอธิบายระดับความสามารถเป็นขั้นตอนที่ต้องให้เวลาในการทำมากกว่าขั้นตอนอื่นๆ เพื่อให้การกำหนดความรู้และทักษะสำหรับแต่ละระดับความสามารถมีความถูกต้องมากที่สุด Giraud และคณะ (2005) กล่าวว่า คำอธิบายระดับความสามารถมีผลต่อความแม่นยำของคะแนนจุดตัดที่จะนำไปใช้เพื่อการตัดสินใจกำหนดนักเรียนเข้ากลุ่มต่างๆ มาก หากเขียนรายละเอียดได้ชัดเจนและถูกต้องย่อมส่งผลให้ผู้เชี่ยวชาญสามารถกำหนดคะแนนจุดตัดได้อย่างความถูกต้องมากขึ้น

Perie (2008) ได้สรุปขั้นตอนการพัฒนาคำอธิบายระดับความสามารถเป็น 2 ขั้นตอนหลักๆ ดังนี้

ขั้นที่ 1 การกำหนดจำนวนของระดับและชื่อของแต่ละระดับความสามารถ

การกำหนดจำนวนของระดับความสามารถถือว่าเป็นสิ่งแรกที่ต้องตัดสินใจ โดยพิจารณาจากเป้าหมายของการนำผลไปใช้ ส่วนชื่อของแต่ละระดับความสามารถก็ควรตั้งให้สื่อความหมายต่อการนำไปใช้ เช่น หากต้องการนำผลไปใช้เพื่อการรับรอง (Certification) หรือ อนุมัติ (Licensure) ก็อาจกำหนดระดับความสามารถเพียง 2 ระดับ คือ ผ่านและตก (Pass / Fail) ก็เพียงพอ แต่หากต้องการจำแนกผู้สอบตามระดับความสามารถ (Classified) ก็อาจกำหนดเป็น 4 ระดับ เช่น ต่ำกว่าพื้นฐาน/ พื้นฐาน/ เชี่ยวชาญ/ เชี่ยวชาญมาก (below basic/ basic/ proficient/ advanced) ไม่ได้

มาตรฐาน/ ได้มาตรฐานบางส่วน/ ได้มาตรฐาน/ เหนือกว่ามาตรฐาน (does not meet standards/ partially meets standards/ meets standards/ exceeds standards) เป็นต้น

Beck (2003 cited in Perie, 2008) กล่าวว่า การกำหนดชื่อสำหรับแต่ละระดับความสามารถนั้นถือเป็นขั้นตอนแรกของการนิยามเกี่ยวกับความสามารถและได้เสนอสิ่งที่ควรหลีกเลี่ยงในการกำหนดชื่อความสามารถแต่ละระดับดังนี้

1) ชื่อที่คลุมเครือ (Nebulous) ไม่ชัดเจน (Unclear) ไม่เหมาะสม (unreasonable) หรือใช้ถ้อยคำหรือภาษาที่ขัดกัน (Oxymoron) เช่น ต้องการพัฒนา (Needs Improvement) รอบรู้พอสมควร (Reasonable Mastery) เป็นต้น

2) ชื่อที่เป็นแบบแผน (Normative Terms) เช่น เฉลี่ย

3) ชื่อที่แสดงถึงการเคลื่อนไหว (Moving Terms) เช่น กำลังปรากฏ (Emerging), กำลังก้าวหน้า (Progressing) เป็นต้น ซึ่งชื่อเหล่านี้ทำให้ยากต่อการนิยามผู้มีความสามารถคาบเส้น (Borderline Performance)

4) ชื่อที่ไม่เกี่ยวข้องกับการศึกษา (Noneducational Terms) เช่น ปกติ(normal) ไม่เพียงพอ (inadequate) เป็นต้น

5) ชื่อที่ไม่เป็นประเภทเดียวกันหรือไม่ขนานกัน (Nonparallel Terms) เช่น ยังไม่สำเร็จ, ผ่าน และเตือน (Outstanding, Pass, Warning) เป็นต้น

เนื่องจากแนวทางการกำหนดชื่อแต่ละระดับมาตรฐานจะยังไม่ชัดเจน Cizek และ Bunch (2007) จึงเสนอแนะเพิ่มเติมว่าการกำหนดชื่อควรกำหนดให้สอดคล้องกับวัตถุประสงค์ของการวัด โครงสร้างของการวัด ควรเป็นชื่อที่สนับสนุนการสรุปผลและการกำหนดกลุ่ม ส่วน Zieky และคณะ (2008) แนะนำว่า ควรใช้คำกลาง ๆ ในการตั้งชื่อระดับ เช่น ความสามารถระดับที่ 1, ระดับที่ 2 และระดับที่ 3 เป็นต้น เพื่อหลีกเลี่ยงการแปลความหมายตามคำที่ใช้จนเกินที่ต้องการสื่อจริง

จากการศึกษาเกี่ยวกับการกำหนดระดับความสามารถในประเทศสหรัฐอเมริกา Perie (2008) สรุปได้ว่าจำนวนระดับ และชื่อของระดับของแต่ละรัฐที่ดำเนินการกำหนดคะแนนจุดตัด นั้น พบว่ามีการแบ่งจำนวนของระดับความสามารถเป็น 3, 4 และ 5 ระดับ โดยจำนวนของระดับที่นิยมใช้มากที่สุดคือ 4 ระดับ รองลงมาคือ 5 ระดับและ 3 ระดับตามลำดับ ส่วนชื่อของระดับที่นิยมใช้มากที่สุด

คือ ต่ำกว่าพื้นฐาน/ พื้นฐาน/ เชี่ยวชาญ/ เชี่ยวชาญพิเศษ (Below Basic/ Basic/ Proficient/ Advanced)

ขั้นที่ 2 การเขียนคำอธิบายระดับความสามารถ

การพัฒนาเป็นคำอธิบายระดับความสามารถเป็นการเขียนบรรยายเกี่ยวกับความรู้ และทักษะที่กำหนดให้สำหรับแต่ละระดับความสามารถ โดยการเขียนคำอธิบายแต่ละระดับความสามารถต้องดำเนินการให้เสร็จก่อนที่จะทำการกำหนดคะแนนจุดตัด เนื่องจากต้องใช้เพื่ออธิบายให้ผู้เชี่ยวชาญเข้าใจให้ตรงกันเกี่ยวกับคุณสมบัติของผู้ที่มีความสามารถแต่ละระดับจำเป็นต้องมี นอกจากนี้ยังเป็นข้อมูลสำหรับรายงานเกี่ยวกับระดับความสามารถของนักเรียนต่อผู้ที่มีส่วนเกี่ยวข้อง เช่น พ่อแม่ผู้ปกครอง ครู หรือแม้แต่ตัวนักเรียนเอง เพื่อให้ทราบถึงคุณสมบัติที่ต้องมีในแต่ละระดับความสามารถและหากต้องการขยับไปสู่ระดับที่สูงขึ้น จะต้องพัฒนาในเรื่องใด (Perie, 2008) ผู้วิจัยขอแสดงตัวอย่างจำนวนระดับ ชื่อของระดับ และคำอธิบายระดับความสามารถ ดังแสดงในตารางต่อไปนี้

ตารางที่ 2.1 ตัวอย่างจำนวนระดับ ชื่อของระดับ และคำอธิบายระดับความสามารถ

รัฐ/หน่วยงาน	ระดับ	ชื่อระดับความสามารถ	คำอธิบายระดับความสามารถ
The National Assessment Educational Progress (Zieky, et al., 2008)	3	ความสามารถระดับเชี่ยวชาญพิเศษ (Advance)	มีความสามารถยอดเยี่ยม
		ความสามารถระดับเชี่ยวชาญ (Proficient)	มีความรู้ความสามารถในเชิงวิชาการที่ควรรู้ทั้งหมดในระดับชั้นนั้น ๆ และแสดงความสามารถในสถานการณ์การทดสอบที่ท้าทายโดยสามารถนำความรู้และความสามารถที่มีมาใช้ในการวิเคราะห์สถานการณ์และแก้ปัญหาได้อย่างเหมาะสม
		ความสามารถระดับพื้นฐาน (Basic)	มีความรู้บางส่วนจากความรู้ทั้งหมดที่ควรรู้ และมีทักษะพื้นฐานสำหรับระดับชั้นที่เรียนอยู่
รัฐเพนซิวาเนีย (Pennsylvania) (Perie, 2008)	4	ความสามารถระดับเชี่ยวชาญพิเศษ (Advance)	มีความรู้ ความสามารถในเชิงวิชาการยอดเยี่ยม คือมีความเข้าใจในเนื้อหาอย่างลึกซึ้ง และมีทักษะยอดเยี่ยม
		ความสามารถระดับเชี่ยวชาญ (Proficient)	มีความรู้ ความสามารถในเชิงวิชาการเพียงพอ คือเข้าในเนื้อหาทั้งหมดและมีทักษะที่เหมาะสม
		ความสามารถระดับพื้นฐาน (Basic)	มีความรู้ ความสามารถในเชิงวิชาการเล็กน้อย คือเข้าใจเนื้อหาเพียงบางส่วน และมีทักษะที่จำกัด ไม่สามารถปฏิบัติงานที่ได้รับมอบหมายได้สำเร็จ
		ความสามารถต่ำกว่าระดับพื้นฐาน (Below Basic)	มีความรู้ ความสามารถในเชิงวิชาการไม่ดีพอ คือไม่เข้าใจในเนื้อหาและมีทักษะเพียงเล็กน้อย
รัฐ อัลบามา (Alabama) (Perie, 2008)	4	ระดับที่ 4 (Level IV)	มีความรู้ ความสามารถในเชิงวิชาการเหนือกว่ามาตรฐาน
		ระดับที่ 3 (Level III)	มีความรู้ ความสามารถในเชิงวิชาการตามมาตรฐาน
		ระดับที่ 2 (Level II)	มีความรู้ ความสามารถทางวิชาการถึงระดับมาตรฐานเพียงบางส่วน
		ระดับที่ 1 (Level I)	มีความรู้ ความสามารถทางวิชาการไม่ถึงระดับมาตรฐาน
รัฐอริโซนา (Arizona) (Perie, 2008)	4	เหนือกว่ามาตรฐาน (Exceeds the Standard)	มีความสามารถทางวิชาการระดับยอดเยี่ยม
		ถึงมาตรฐาน (Meets the Standard)	มีความรู้ความสามารถในเชิงวิชาการทั้งหมด สามารถประยุกต์ใช้ ความรู้ในสถานการณ์จริงได้
		เข้าใกล้มาตรฐาน (Approaches the Standard)	มีความเข้าใจในเนื้อหา และสามารถประยุกต์ ทักษะพื้นฐานต่างๆ ได้
		ต่ำกว่ามาตรฐานมาก (Falls Far Below the Standard)	มีความรู้และทักษะเบื้องต้นที่ต้องมี

Zieky และคณะ (2008) เสนอแนะว่า หากเป็นไปได้ควรเขียนอธิบายระดับความสามารถให้เสร็จก่อนที่จะดำเนินการสร้างข้อสอบ ทั้งนี้เนื่องจากว่าหากผู้ที่สร้างข้อสอบทราบว่าระดับความสามารถแต่ละระดับจะต้องสามารถทำอะไรหรือรู้อะไรบ้าง ก็จะสามารถสร้างข้อคำถามเพื่อ

วัดความสามารถในระดับต่างๆ ได้อย่างถูกต้องและครอบคลุมทุกระดับความสามารถที่ผู้กำหนดมาตรฐานต้องการ

นอกจากคำอธิบายระดับความสามารถจะถูกใช้เพื่อการกำหนดคะแนนจุดตัดแล้วยังใช้เพื่อการรายงานผล เพื่อเป็นข้อมูลย้อนกลับเกี่ยวกับความรู้และทักษะสำหรับผู้ที่ได้คะแนนสอบต่างๆ กัน ซึ่งเป็นสารสนเทศที่ถูกกำหนดขึ้นก่อนที่จะมีการสอบโดยการตัดสินของผู้ตัดสิน จากการศึกษาพบว่า มีการแปลความหมายเกี่ยวกับคะแนนสอบที่แตกต่างกันของผู้สอบแต่ละคน เพื่อสะท้อนความรู้และทักษะที่ต่างกันของผู้สอบแต่ละคน โดยใช้วิธี Anchoring Scaling เป็นวิธีการแปลความหมายคะแนนสอบโดยอาศัยข้อมูลเชิงประจักษ์ ซึ่งดำเนินการกำหนดขึ้นหลังจากที่การสอบเสร็จสิ้นแล้ว

1.5 วิธีกำหนดคะแนนจุดตัด

วิธีการกำหนดคะแนนจุดตัด นิยมแบ่งออกเป็น 2 กลุ่ม คือ วิธีแบบสอบเป็นศูนย์กลาง (test-center methods) และวิธีผู้สอบเป็นศูนย์กลาง (Examinee-centered methods) (Berk, 1996; Plake, 2008; Lee & Lewis, 2008; Näsström & Nyström, 2008)

รายละเอียดเกี่ยวกับวิธีการกำหนดคะแนนจุดตัด ดังแสดงต่อไปนี้

1.5.1 วิธีแบบสอบเป็นศูนย์กลาง (Test-Center Methods)

ผู้เชี่ยวชาญจะกำหนดคะแนนจุดตัดโดยพิจารณาจากข้อสอบ วิธีแบบสอบเป็นศูนย์กลางมีรายละเอียดดังแสดงต่อไปนี้

1.5.1.1 วิธีแองกอฟ (Angoff method)

ความเป็นมา

วิธีแองกอฟ (Angoff method) ถูกคิดขึ้นโดย William H. Angoff ในปีค.ศ. 1970 เพื่อใช้ในการกำหนดคะแนนจุดตัดสำหรับข้อสอบแบบเลือกตอบ ซึ่งวิธีแองกอฟถูกนำไปใช้ในการกำหนดคะแนนจุดตัดอย่างกว้างขวางสำหรับการทดสอบที่มีความสำคัญ (high stakes Testing) (Angoff, 1971, cited in Kane, 1994) และยังได้รับความนิยมอย่างต่อเนื่องจนถึงปัจจุบัน (Zieky et al., 2008) ทั้งนี้เนื่องจากว่าวิธีแองกอฟเป็นวิธีที่ง่ายต่อการปฏิบัติและเป็นการกำหนดคะแนนจุดตัดโดยการพิจารณาจากข้อสอบจึงสามารถดำเนินการได้ทั้งก่อนและหลังจากที่มีการจัดสอบแล้ว โดยให้ผู้เชี่ยวชาญพิจารณาความน่าจะเป็นที่ผู้มีความสามารถคาบเส้นจะสามารถตอบข้อสอบแต่ละข้อ

ได้ถูกต้อง ซึ่งผู้มีความสามารถคาบเส้น หมายถึง คนที่มีระดับความรู้ขั้นต่ำในระดับความสามารถนั้น ๆ เมื่อพิจารณาครบทุกข้อแล้วก็นำค่าความน่าจะเป็นที่ผู้มีความสามารถคาบเส้นจะสามารถตอบข้อสอบได้ถูกต้อง ซึ่งกำหนดโดยผู้เชี่ยวชาญแต่ละท่านมาหาค่าเฉลี่ยก็จะได้ความน่าจะเป็นที่ผู้มีความสามารถคาบเส้นจะสามารถตอบข้อสอบได้ถูกต้องเป็นรายข้อ จากนั้นก็รวมค่าความน่าจะเป็นที่ผู้มีความสามารถคาบเส้นจะสามารถตอบข้อสอบได้ถูกต้องจากทุกข้อในแบบสอบเพื่อกำหนดเป็นคะแนนจุดตัด

ขั้นตอน

การกำหนดคะแนนจุดตัดโดยวิธีแองกอฟมีขั้นตอนการดำเนินการดังต่อไปนี้

1) ผู้เชี่ยวชาญกำหนดระดับการผ่านขั้นต่ำ ซึ่งหมายถึงความรู้ขั้นต่ำในเรื่องที่จำเป็นต้องมี เพื่อนิยามผู้มีความสามารถคาบเส้น

2) ผู้เชี่ยวชาญพิจารณาข้อสอบทีละข้อเพื่อพิจารณาและกำหนดความน่าจะเป็นที่ผู้มีความสามารถคาบเส้นจะสามารตอบข้อสอบข้อนั้นได้ถูกต้อง

ตัวอย่างการคำนวณคะแนนจุดตัดที่กำหนดโดยวิธีแองกอฟ ดังแสดงในตารางต่อไปนี้

ตารางที่ 2.2 ตัวอย่างการคำนวณคะแนนจุดตัดที่กำหนดโดยวิธีแองกอฟ

ข้อสอบ	ผู้ตัดสินคนที่ 1	ผู้ตัดสินคนที่ 2
1	1.00	0.85
2	0.65	0.50
3	0.80	0.75
4	0.45	0.50
5	0.30	0.40
รวม	3.20	3.00

ที่มา: Zieky และคณะ (2008)

คะแนนจุดตัด = (3.2+3.0)/2 = 3.1 นั่นคือผู้ที่จะผ่านได้ ต้องทำข้อสอบให้ถูก 3 ข้อ จาก 5 ข้อ

ต่อมาวิธีแองกอฟได้ถูกพัฒนาให้ง่ายต่อการปฏิบัติมากขึ้นโดย Impara และ Plake (1997) ซึ่งใช้ชื่อว่า วิธีแองกอฟแบบใช่/ไม่ใช่ เพื่อให้สามารถกำหนดคะแนนจุดตัดได้ในกรณีที่ข้อสอบแต่ละข้อให้คะแนนไม่เท่ากัน การพิจารณากำหนดคะแนนจุดตัดจะต่างจากวิธีแองกอฟตรงที่วิธีแองกอฟแบบใช่/ไม่ใช่จะให้ผู้เชี่ยวชาญพิจารณาว่าผู้มีความสามารถคาบเส้นจะสามารถตอบข้อสอบได้ถูกต้องใช่หรือไม่ ซึ่ง Impara และ Plake (1997) ได้ทำการทดลองเพื่อเปรียบเทียบ

คะแนนจุดตัดที่กำหนดโดยใช้วิธีแองกอฟและวิธีแองกอฟแบบใช่/ไม่ใช่ พบว่า ได้คะแนนจุดตัดเท่ากัน

การกำหนดคะแนนจุดตัดโดยวิธีแองกอฟแบบใช่/ไม่ใช่ จะดำเนินการกำหนดคะแนนจุดตัด 3 รอบ เนื่องจากเป็นจำนวนรอบที่นักวิชาการส่วนใหญ่นิยม (Busch & Jaeger, 1990; Fehrmann, et al., 1991; Hambleton & Plake, 1995; Hess, et al., 2007) แต่ละรอบผู้เชี่ยวชาญสามารถเปลี่ยนแปลงคะแนนจุดตัดได้ โดยดำเนินการดังต่อไปนี้

รอบที่ 1 ผู้เชี่ยวชาญแต่ละท่านพิจารณาข้อสอบทีละข้อและตัดสินว่าผู้มีความสามารถคาบเส้นของแต่ละระดับความสามารถจะสามารถตอบข้อสอบแต่ละข้อได้ถูกต้องใช่หรือไม่ หากผู้เชี่ยวชาญพิจารณาว่าผู้มีความสามารถคาบเส้นของระดับความสามารถที่กำลังพิจารณาอยู่สามารถตอบข้อสอบข้อนั้นได้ถูกต้องก็ให้ใส่เลข 1 ลงในช่องผลการพิจารณาของข้อนั้น แต่หากผู้เชี่ยวชาญพิจารณาว่าผู้มีความสามารถคาบเส้นของระดับความสามารถที่กำลังพิจารณาอยู่ไม่สามารถตอบข้อสอบข้อนั้นได้ถูกต้องก็ให้ใส่เลข 0 ลงในช่องผลการพิจารณาของข้อนั้น โดยผู้เชี่ยวชาญจะต้องพิจารณาทีละระดับความสามารถจนครบทุกระดับ

รอบที่ 2 ผู้เชี่ยวชาญร่วมกันอภิปรายกลุ่มย่อยเกี่ยวกับคะแนนจุดตัดที่กำหนดในรอบที่ 1 และกำหนดคะแนนจุดตัดอีกครั้งในรอบนี้ โดยใช้หลักการเดียวกันกับรอบที่ 1

รอบที่ 3 ผู้เชี่ยวชาญร่วมกันอภิปรายกลุ่มใหญ่เกี่ยวกับคะแนนจุดตัดที่กำหนดในรอบที่ 2 และกำหนดคะแนนจุดตัดครั้งสุดท้าย โดยใช้หลักการเดียวกันกับรอบที่ 1

ข้อดีและข้อด้อย

<u>ข้อดี</u>

วิธีแองกอฟ และวิธีแองกอฟแบบใช่/ไม่ใช่มีข้อดีที่เด่นที่สุดคือเป็นวิธีที่ง่ายต่อการนำไปปฏิบัติ ซึ่งทำให้ผู้เชี่ยวชาญหรือผู้ที่เกี่ยวข้องสามารถปฏิบัติได้โดยไม่ต้องใช้เวลาในการเรียนรู้มากนอกจากนี้วิธีแองกอฟใช้สถิติในการวิเคราะห์ไม่ยาก

<u>ข้อด้อย</u>

วิธีแองกอฟ และวิธีแองกอฟแบบใช่/ไม่ใช่มีข้อบกพร่องที่สำคัญคือการพิจารณากำหนดคะแนนจุดตัดนั้นต้องพิจารณาข้อสอบทุกข้อ ดังนั้นหากข้อสอบมีจำนวนมากก็ต้องใช้เวลาในการพิจารณานานจึงไม่เหมาะกับการกำหนดคะแนนจุดตัดสำหรับข้อสอบที่มีจำนวนข้อมากๆ

แต่หากสามารถปรับปรุงวิธีการแองกอฟเพื่อลดข้อบกพร่องนี้ เช่นการสุ่มข้อสอบมาพิจารณา โดยที่ยังคงกำหนดคะแนนจุดตัดได้เท่ากับการกำหนดจากแบบสอบฉบับเต็ม ซึ่งถือว่ามีประโยชน์มาก

1.5.1.2 วิธีบุ๊คมาร์ค (Bookmark Method)

ความเป็นมา

นับตั้งแต่ Lewis และ Green (1996 cited in Zieky et al., 2008) ได้เสนอวิธีบุ๊คมาร์คครั้งแรกในปี 1996 ก็ได้รับความนิยมอย่างรวดเร็ว โดยเฉพาะในปี ค.ศ. 2000 วิธีบุ๊คมาร์คถูกนำไปใช้กันอย่างแพร่หลายในสหรัฐอเมริกา โดยจากการสำรวจพบว่าวิธีบุ๊คมาร์คถูกนำไปใช้ใน 18 รัฐ ซึ่งถือว่าถูกใช้มากที่สุดและวิธีแองกอฟถูกนำไปใช้ 6 รัฐ (Nellhaus, 2000) ต่อมาในปี ค.ศ. 2001 วิธีบุ๊คมาร์คก็ได้ถูกนำมาใช้เพื่อกำหนดคะแนนจุดตัดมากกว่า 28 รัฐ (Egan, 2001 cited in Lin, 2008) และเมื่อปี 2005 มี 31 รัฐ ที่ใช้วิธีบุ๊คมาร์คในการกำหนดคะแนนจุดตัดสำหรับการสอบที่สำคัญ (Perie, 2005 cited in Lee & Lewis, 2008) นอกจากนี้ Karantonis และ Sireci (2006) กล่าวว่าในแวดวงของการวัดผลการศึกษานั้น วิธีบุ๊คมาร์คเป็นวิธีที่นิยมใช้มากที่สุด สำหรับการศึกษาในประเทศไทยโดย สุรีพร อนุศาสนนันท์ (2550) ได้ศึกษาเพื่อเปรียบเทียบคุณภาพของการกำหนดคะแนนจุดตัดระหว่างวิธีแองกอฟที่ได้รับการปรับปรุงกับ วิธีบุ๊คมาร์ค ผลการศึกษาพบว่าวิธีบุ๊คมาร์คมีความเหมาะสมและมีความเป็นไปได้ในการนำมาใช้มากกว่าวิธีแองกอฟ

วิธีบุ๊คมาร์คเป็นวิธีที่สามารถช่วยในการกำหนดคะแนนจุดตัดได้หลายระดับโดยการใช้แบบสอบเพียงฉบับเดียว และสามารถใช้ได้ทั้งข้อสอบที่เป็นแบบเลือกตอบที่ให้คะแนน 2 ค่า และข้อสอบแบบให้คะแนนมากกว่า 2 ค่า นอกจากนี้วิธีบุ๊คมาร์คยังมีจุดเด่นอยู่ที่การจัดทำคู่มือเรียงข้อสอบ (Ordered Item Booklet: OIB) ซึ่งจัดทำโดยการเรียงข้อสอบจากข้อง่ายไปข้อยาก โดยใช้ทฤษฎีการตอบสนองข้อสอบ เพื่อช่วยในการวิเคราะห์หาค่าความยากของข้อสอบ และเป็นการทำให้ความสามารถของนักเรียนอยู่ในระดับเดียวกัน ซึ่งทำให้ผู้เชี่ยวชาญสามารถตัดสินกำหนดจุดตัดได้ง่ายขึ้น Lewis (1998 cited in Lee & Lewis, 2008) เสนอว่า ในการวิเคราะห์ข้อสอบแบบให้คะแนน 2 ค่า ควรใช้โมเดลการตอบสนองข้อสอบแบบ 3 พารามิเตอร์ (Three parameter model) ส่วนการวิเคราะห์ข้อสอบแบบให้คะแนนคำตอบถูกบางส่วนนั้นควรวิเคราะห์ด้วยโมเดลการตอบสนองข้อสอบแบบ Partial credit model 2 พารามิเตอร์ Cizek และคณะ (2004) เสนอว่า การวิเคราะห์ข้อสอบด้วยโมเดลการตอบสนองข้อสอบแบบ 1 พารามิเตอร์ สามารถใช้ในการวิเคราะห์

ข้อสอบทั้งที่เป็นแบบให้คะแนน 2 ค่า หรือให้คะแนนมากกว่า 2 ค่าได้ดี เมื่อสร้างคู่มือเรียงข้อสอบเรียบร้อยแล้วก็เป็นขั้นตอนของการกำหนดคะแนนจุดตัด โดยผู้เชี่ยวชาญจะต้องพิจารณาข้อสอบทีละข้อตามลำดับความยาก ผู้เชี่ยวชาญจะต้องพิจารณาว่าผู้มีความสามารถคาบเส้นมีโอกาสตอบข้อสอบข้อนั้นได้ถูก (Response Probability: RP) หรือไม่

ขั้นตอน

วิธีบุ๊คมาร์คจะดำเนินการกำหนดคะแนนจุดตัด 3 รอบ ดังนี้

รอบที่ 1 ผู้เชี่ยวชาญแต่ละท่านพิจารณาคู่มือเรียงข้อสอบในแต่ละหน้า และพิจารณาในประเด็นต่อไปนี้ 1) ผู้สอบต้องมีความรู้ความสามารถในเรื่องใดจึงจะสามารถตอบข้อสอบแต่ละข้อได้ถูกต้อง และ 2) เหตุใดข้อสอบข้อนี้จึงยากกว่าข้อก่อนๆ จากนั้นผู้เชี่ยวชาญกำหนดคะแนนจุดตัดอย่างอิสระโดยนำที่คั่นหนังสือมาคั่นที่ข้อสอบซึ่งผู้เชี่ยวชาญคิดว่ามีความเหมาะสมที่จะเป็นจุดตัด โดยข้อที่อยู่ก่อนหน้าที่คั่นหนังสือจะเป็นข้อที่ผู้มีความสามารถคาบเส้นของระดับความสามารถที่กำลังพิจารณาอยู่ มีความน่าจะเป็น 0.67 ที่จะตอบข้อสอบได้ถูกต้อง โดยผู้เชี่ยวชาญจะพิจารณาจากระดับความสามารถต่ำที่สุดไปจนถึงระดับความสามารถที่สูงที่สุด

รอบที่ 2 ผู้เชี่ยวชาญร่วมกันอภิปรายกลุ่มย่อยเกี่ยวกับคะแนนจุดตัดที่กำหนดในรอบที่ 1 และกำหนดคะแนนจุดตัดอีกครั้ง โดยอาศัยหลักการเดียวกับรอบที่ 1

รอบที่ 3 ผู้เชี่ยวชาญร่วมกันอภิปรายกลุ่มใหญ่เกี่ยวกับคะแนนจุดตัดที่กำหนดในรอบที่ 2 และกำหนดคะแนนจุดตัดอีกครั้ง โดยอาศัยหลักการเดียวกับรอบที่ 1

ข้อดีและข้อด้อย

<u>ข้อดี</u>

1. วิธีนี้เป็นการตัดสินคะแนนจุดตัดโดยพิจารณาข้อสอบโดยภาพรวมและสามารถกำหนดคะแนนจุดตัดหลายค่าโดยใช้ข้อสอบเพียงฉบับเดียว

2. วิธีนี้ใช้ได้ทั้งกับข้อสอบที่ให้คะแนน 2 ค่าและแบบให้คะแนนมากกว่า 2 ค่า เนื่องจากคะแนนอยู่บนเสกลเดียวกัน

3. วิธีนี้ได้คำนวณความยากเพื่อประกอบการพิจารณาทำให้ตัดสินได้ง่ายขึ้น และทำให้คะแนนจุดตัดมีความน่าเชื่อถือมากขึ้น

<u>ข้อด้อย</u>

1. วิธีการนี้ต้องใช้ข้อสอบจำนวนมากมิฉะนั้นอาจเกิด floor and ceiling effect ได้

2. ผลการตัดสินขึ้นอยู่ที่คณะผู้เชี่ยวชาญที่ใช้ตัดสิน

1.5.1.3 วิธีของอีเบล (Ebel's method)

เนื่องจากการที่ข้อสอบมีความยากมาก หรือง่ายเกินไป หรือมีอำนาจจำแนกต่ำมากเกินกว่าที่ผู้ออกข้อสอบคาดหวังไว้ ทำให้การที่ผู้สอบจะสอบผ่านหรือไม่นั้นขึ้นอยู่กับคำถามในแบบสอบมากกว่าการพิจารณาจากระดับความสามารถของผู้สอบ อีเบลจึงเสนอวิธีการแก้ไขจุดอ่อนดังกล่าวโดยพิจารณาความเกี่ยวข้อง (Relevance) หรือความสำคัญ (Important) ของแต่ละข้อคำถามในแบบสอบ สำหรับในเรื่องของความเกี่ยวข้องอีเบลได้แบ่งออกเป็น 4 ประเภท คือ มีความจำเป็น (Essential) มีความสำคัญ (Important) เป็นที่ยอมรับได้ (Acceptable) และน่าสงสัย (Questionable) ส่วนในเรื่องระดับความยาก อีเบลได้แบ่งออกเป็น 3 ระดับ คือ ง่าย ปานกลาง และยาก โดยให้ผู้เชี่ยวชาญเป็นผู้พิจารณาและกำหนดว่าข้อใดถือเป็นข้อยาก ปานกลาง หรือข้อง่าย จากนั้นก็กำหนดร้อยละที่คาดหวังที่ผู้มีความสามารถคาบเส้นจะสามารถผ่านได้ถือว่าเป็นปริมาณความคาดหวังของผู้มีความสามารถคาบเส้นที่จะสอบผ่าน (Zieky et al., 2008)

สรุปขั้นตอนการกำหนดคะแนนจุดตัดโดยวิธีอีเบล สามารถสรุปได้ดังนี้

1) ให้ผู้เชี่ยวชาญแต่ละคนพิจารณาข้อคำถาม แล้วจัดแยกเป็น 12 ประเภท ตามระดับความเกี่ยวข้อง 4 ลักษณะ และลักษณะความยาก 3 ระดับ อภิปรายแต่ละข้อคำถามดังนี้

2) ถามผู้เชี่ยวชาญว่ามีกี่คนที่จัดคำถามอยู่ในระดับง่าย ปานกลางและยาก ถ้าผู้เชี่ยวชาญจัดไม่เหมือนกัน ก็ควรจะให้อธิบายเหตุผล

3) ถามผู้เชี่ยวชาญว่ามีกี่คนที่แยกประเภทของข้อคำถามออกเป็น 1) มีความจำเป็น 2) มีความสำคัญ 3) เป็นที่ยอมรับ และ 4) น่าสงสัย ถ้าผู้เชี่ยวชาญแยกประเภทของข้อคำถามไม่เหมือนกัน ก็ควรให้อธิบายเหตุผลเช่นเดียวกัน

4) ให้โอกาสแก่ผู้เชี่ยวชาญจัดประเภทของข้อคำถามใหม่ถ้าต้องการเปลี่ยนแปลง

5) ให้ผู้เชี่ยวชาญกำหนดร้อยละของผู้สอบที่คาดหวังว่าจะตอบในแต่ละข้อคำถามได้ถูกต้องทั้ง 12 ประเภท

6) รวบรวมผลการตัดสิน และสรุปคะแนนจุดตัด

วิธีการคำนวณคะแนนจุดตัดของอีเบล คือ จำนวนข้อสอบในแต่ลักษณะคูณกับค่าร้อยละที่คาดหวังแล้วนำผลคูณที่ได้มาหาค่าเฉลี่ย เพื่อกำหนดเป็นคะแนนจุดตัดหรือคะแนนที่ผู้มีความสามารถคาบเส้นจะสอบผ่าน

ข้อดีและข้อด้อย

ข้อดี

วิธีของอีเบลนั้นกำหนดให้ผู้เชี่ยวชาญพิจารณาเพื่อกำหนดคะแนนจุดตัดเพียง 12 ครั้งเท่านั้นคือพิจารณาตามลักษณะที่แบ่งไว้ทั้ง 12 ลักษณะซึ่งถือว่าเป็นการตัดสินที่ทำได้ง่ายและใช้เวลาไม่มาก

ข้อด้อย

การคำนวณคะแนนจุดตัดต้องนำจำนวนข้อสอบในแต่ละลักษณะไปคูณกับค่าร้อยละที่คาดหวังนั้นเป็นเรื่องที่ค่อนข้างยุ่งยากและเสียเวลามาก

1.5.1.4 วิธีของนีเดลสกี (Nedelsky method)

นีเดลสกีได้เสนอเทคนิคนี้ในปี ค.ศ. 1954 (Linn, 1989) ซึ่งเป็นเทคนิคที่เหมาะกับข้อสอบแบบเลือกตอบซึ่งพิจารณาโดยให้ผู้เชี่ยวชาญพิจารณาข้อคำถามและตัวเลือกทุกข้อเพื่อตัดสินว่าตัวเลือกที่ผิดใดบ้างที่ผู้มีความสามารถคาบเส้นจะไม่เลือก โดยจำนวนตัวเลือกที่ยังคงเหลืออยู่จะนำมาคำนวณเป็นคะแนนจุดตัดรายข้อจากนั้นก็นำผลการคำนวณทุกข้อรวมกันเพื่อกำหนดคะแนนจุดตัด (Zieky et al., 2008)

สรุปวิธีกำหนดคะแนนจุดตัดด้วยวิธีนีเดลสกี มีรายละเอียดดังนี้

1) ให้ผู้เชี่ยวชาญแต่ละคนพิจารณาคำถามและตัวเลือกทั้งหมด จากนั้นใช้ดินสอขีดลงไปที่ตัวเลือกที่ผิดที่เห็นว่าผู้สอบจะไม่เลือก

2) อภิปรายคำถามแต่ละข้อ ตามแนวทางดังนี้

- ให้ผู้เชี่ยวชาญพิจารณาคำตอบที่ผิดก่อน และดูว่ามีผู้เชี่ยวชาญกี่ท่านที่คิดว่าผู้สอบจะไม่เลือก และมีกี่ท่านไม่เห็นด้วย
- หากผู้เชี่ยวชาญมีความเห็นไม่ตรงกัน จะต้องให้ผู้เชี่ยวชาญแต่ละท่านอธิบายเหตุผลของตนเอง โดยยังไม่ต้องหาข้อสรุปใด ๆ จนกว่าผู้เชี่ยวชาญจะแน่ใจว่าได้เข้าใจความหมายของคำว่า “ผู้มีความสามารถคาบเส้น” ได้ตรงกัน

- พิจารณาตัวเลือกที่เป็นตัวผิดตัวต่อไป

3) หลังจากที่ได้พิจารณาตัวเลือกจนครบแล้ว ขั้นตอนต่อไป คือ การให้ผู้เชี่ยวชาญทบทวนผลการตัดสินโดยผู้เชี่ยวชาญจะต้องทำเครื่องหมายกากบาทลงบนตัวเลือกที่แน่ใจว่าผู้สอบที่มีความสามารถต่ำจะไม่เลือก

4) รวบรวมผลการตัดสิน แล้วดำเนินการเช่นเดียวกันในคำถามต่อไป

วิธีนีเดลสกีนี้ สามารถให้ผู้เชี่ยวชาญแต่ละคนตัดสินเองโดยปราศจากการอภิปรายร่วมกันก็ได้ แต่อาจเกิดปัญหาในกรณีที่ผู้เชี่ยวชาญจะยังไม่คุ้นเคยกับวิธีการตัดสินเพราะฉะนั้นถ้าเป็นไปได้ก็ควรจะให้โอกาสแก่ผู้เชี่ยวชาญในการฝึกตัดสินใจ จากตัวอย่างข้อสอบและฝึกอภิปรายร่วมกันก่อนที่จะมีการพิจารณาคำถามที่เหลือด้วยตนเองต่อไป

ข้อดีและข้อด้อย

<u>ข้อดี</u>

การกำหนดคะแนนจุดตัดโดยวิธีนีเดลสกีนี้ เป็นการพิจารณาที่ตัวเลือกที่ผิดดังนั้นจึงมีประโยชน์มากในศาสตร์ที่ตัวเลือกที่ผิดมีความสำคัญ เพราะตัวเลือกที่ผิดถือเป็นตัวลวง ดังนั้นหากผู้สอบเข้าใจแนวคิดเกี่ยวกับเรื่องที่สอบผิด (Miss Concept) ก็จะทำให้เลือกตอบตัวเลือกซึ่งเป็นตัวลวง ซึ่งการเลือกตอบตัวลวงแต่ละตัวก็จะแสดงให้เห็นว่าผู้สอบมีแนวคิดเกี่ยวกับเรื่องที่สอบผิด ซึ่งจะเป็นประโยชน์มากสำหรับผู้สอนที่จะสามารถอธิบายเพิ่มเติมในเรื่องที่ผู้สอบเข้าใจผิด

<u>ข้อด้อย</u>

การกำหนดคะแนนจุดตัดโดยวิธีนีเดลสกีใช้เวลาในการกำหนดคะแนนจุดตัดมาก

1.5.1.5 วิธีเอบีซี (ABC Method)

ความเป็นมา

Downing และคณะ (2006) กล่าวว่า ไม่มีวิธีการกำหนดคะแนนจุดตัดใดที่สมบูรณ์แบบ และได้รับการยอมรับว่าเป็นวิธีที่ควรนำไปใช้มากที่สุด ดังนั้นจึงมีนักวิชาการได้พยายามเสนอวิธีการกำหนดคะแนนจุดตัดวิธีใหม่ ๆ ขึ้นมามากมายที่มีความน่าเชื่อถือ เพื่อนำเสนอต่อผู้ที่ต้องการกำหนดคะแนนจุดตัด ผู้วิจัยจึงคิดวิธีกำหนดคะแนนจุดตัดขึ้นมาเรียกว่า วิธีเอบีซี เป็นวิธีที่ผสมผสานระหว่างวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีวิเคราะห์กลุ่ม ซึ่งวิธีแองกอฟแบบใช่/ไม่ใช่ และวิธีบุ๊คมาร์คเป็นวิธีแบบสอบเป็นศูนย์กลาง (test-center methods) (Zieky et al., 2008)

กล่าวคือเป็นการกำหนดคะแนนจุดตัดโดยเน้นการพิจารณาจากข้อสอบเป็นหลัก ส่วนวิธีการวิเคราะห์กลุ่มเป็นวิธีผู้สอบเป็นศูนย์กลาง (Examinee-centered methods) ซึ่งเป็นการกำหนดคะแนนจุดตัดที่เน้นการพิจารณาจากผลการตอบของผู้สอบเป็นหลัก (Berk, 1996; Plake, 2008; Lee & Lewis, 2008; Näsström & Nyström, 2008) วิธีเอบีซีจะดึงเอาจุดเด่นและขณะเดียวกันก็ลดจุดด้อยของแต่ละวิธีลง ซึ่งวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีการวิเคราะห์กลุ่มต่างก็มีจุดเด่นที่ต่างกัน โดยวิธีแองกอฟแบบใช่/ไม่ใช่มีจุดเด่นคือการให้ความสำคัญกับการกำหนดคะแนนจุดตัดของผู้ตัดสินโดยปราศจากกรอบเรื่องค่าความยากง่ายของข้อสอบ กล่าวคือเป็นการกำหนดคะแนนจุดตัดที่ให้ความสำคัญกับการตัดสินใจของผู้เชี่ยวชาญที่พิจารณาจากเนื้อหาที่สอบวัดเป็นหลัก สำหรับวิธีบุ๊คมาร์คมีจุดเด่นอยู่ที่คู่มือเรียงข้อสอบที่ช่วยให้ผู้เชี่ยวชาญกำหนดจุดตัดได้ง่ายและเร็ว ส่วนวิธีการวิเคราะห์กลุ่มเป็นการวิเคราะห์ทางสถิติเพื่อจัดให้ผู้สอบที่มีลักษณะที่สนใจเหมือนหรือคล้ายกันอยู่กลุ่มเดียวกัน และแต่ละกลุ่มจะต้องมีลักษณะที่สนใจต่างกันให้มากที่สุด ซึ่งกล่าวได้ว่าเป็นวิเคราะห์เพื่อการกำหนดกลุ่มให้ผู้สอบจากลักษณะที่สนใจโดยปราศจากความลำเอียงใดๆ

สำหรับข้อด้อยของทั้ง 3 วิธีข้างต้น คือวิธีแองกอฟแบบใช่/ไม่ใช่นั้น ผู้เชี่ยวชาญจะต้องใช้กระบวนการคิดอย่างมาก ในการพิจารณาว่าผู้มีความสามารถคาบเส้นจะสามารถตอบข้อสอบได้ถูกต้องใช่หรือไม่ โดยเฉพาะในกรณีที่มีจำนวนข้อสอบที่ต้องพิจารณามาก ๆ และต้องกำหนดคะแนนจุดตัดหลายรอบ อาจทำให้ผู้เชี่ยวชาญเกิดความอ่อนล้าในขณะที่กำหนดคะแนนจุดตัดซึ่งอาจส่งผลให้คะแนนจุดตัดที่กำหนดเกิดความคลาดเคลื่อนได้ (Clauser, et al., 2009) ส่วนวิธีบุ๊คมาร์คมีจุดด้อย คือหากเกิดกรณีที่ผู้เชี่ยวชาญรู้สึกว่าข้อสอบที่ถูกเรียงไว้ตามลำดับความยากง่ายนั้นบางข้อมีตำแหน่งที่ไม่เหมาะสม ซึ่งกรณีเช่นนี้อาจส่งผลต่อคุณภาพของคะแนนจุดตัดที่กำหนดได้ ส่วนวิธีการวิเคราะห์กลุ่มก็มีจุดด้อยคือไม่สามารถระบุได้ว่าแต่ละกลุ่มที่แบ่งมีความสามารถระดับใดและไม่สามารถกำหนดคะแนนจุดตัดที่สูงกว่าหรือต่ำกว่าผลการตอบข้อสอบของผู้สอบกลุ่มที่เลือกมาวิเคราะห์ได้ (Violato, Marini & Lee, 2003; Hess, Subhiyah & Giordano, 2007)

จากที่กล่าวมาข้างต้น วิธีเอบีซีจึงเป็นวิธีที่ถูกพัฒนาขึ้นโดยมีเป้าหมายที่จะทำให้ผู้เชี่ยวชาญสามารถกำหนดคะแนนจุดตัดได้ง่ายและเร็ว โดยให้ผู้เชี่ยวชาญกำหนดคะแนนจุดตัดจากคู่มือเรียงข้อสอบซึ่งถือเป็นจุดเด่นของบุ๊คมาร์ค แต่วิธีเอบีซีจะแบ่งคู่มือเรียงข้อสอบเป็นหลาย ๆ ฉบับตามสาระการเรียนรู้เรียกว่า คู่มือเรียงข้อสอบรายสาระการเรียนรู้ (Domain–Ordered Item Booklet: D-OIB) ซึ่ง

เป็นคู่มือที่จัดทำขึ้นโดยแบ่งข้อสอบเป็นกลุ่มตามสาระการเรียนรู้ กล่าวคือคู่มือเรียงข้อสอบรายสาระการเรียนรู้จะมีหลายฉบับขึ้นอยู่กับจำนวนสาระการเรียนรู้ที่สอบวัด ข้อสอบในแต่ละสาระการเรียนรู้จะถูกจัดเรียงจากข้อง่ายไปข้อยาก ข้อละ 1 หน้า โดยใช้ทฤษฎีการตอบสนองข้อสอบโมเดล 3 พารามิเตอร์ (Three parameter model) ในการวิเคราะห์ค่าความยาก ทั้งนี้ก็เพื่อให้การพิจารณากำหนดคะแนนจุดตัดทำได้ง่ายขึ้น เนื่องจากคู่มือเรียงข้อสอบรายสาระการเรียนรู้แต่ละฉบับจะประกอบด้วยข้อสอบที่วัดสาระการเรียนรู้เดียวกัน ส่วนหลักการพิจารณาเพื่อกำหนดคะแนนจุดตัดจะดำเนินการคล้ายกับวิธีแองกอฟแบบใช่/ไม่ใช่ กล่าวคือให้ผู้เชี่ยวชาญพิจารณาข้อสอบตามลำดับความยากง่าย และตัดสินว่าผู้ที่สามารถตอบข้อสอบแต่ละข้อได้ถูกต้องควรมีความสามารถอย่างน้อยระดับใดแทนที่จะให้ผู้เชี่ยวชาญนำที่คั่นหนังสือมาขั้นตรงข้อที่ผู้เชี่ยวชาญคิดว่าเป็นจุดตัด ซึ่งการพิจารณาแบบนี้น่าจะช่วยลดปัญหาที่เกิดจากการที่ผู้เชี่ยวชาญคิดว่าข้อสอบที่ถูกจัดเรียงไว้ตามความยากง่ายมีลำดับที่ไม่เหมาะสม เพราะวิธีเอบีซีใช้คู่มือการจัดเรียงข้อสอบเป็นเพียงสารสนเทศประกอบการกำหนดคะแนนจุดตัดที่เสนอเกี่ยวกับค่าความยากง่ายของข้อสอบที่ได้จากการวิเคราะห์จากผลการตอบของผู้สอบเท่านั้น กล่าวได้ว่าวิธีเอบีซียังคงให้ความสำคัญกับความคิดเห็นของผู้เชี่ยวชาญที่พิจารณาจากเนื้อหาที่สอบเป็นหลักแทนที่จะต้องพิจารณาคะแนนจุดตัดที่ถูกจำกัดไว้กับลำดับของความยากง่ายของข้อสอบจนเกินไป นอกจากนี้วิธีเอบีซีจะเสนอสารสนเทศที่มากกว่าค่าความยากง่ายของข้อสอบ โดยจะเสนอสารสนเทศเกี่ยวกับผลการวิเคราะห์กลุ่ม (Clustering – Result: CR) ซึ่งเป็นผลการจัดกลุ่มที่มิใช่เป็นการกำหนดโดยอาศัยความคิดเห็นของผู้เชี่ยวชาญแต่เป็นการวิเคราะห์ข้อมูลทางสถิติที่ใช้ข้อมูลจากผลการสอบของผู้สอบมาทำการวิเคราะห์ โดยตัวแปรที่ใช้ในการแบ่งกลุ่ม ได้แก่ คะแนนรวมของข้อสอบของแต่ละรายสาระการเรียนรู้ สารสนเทศนี้จะนำเสนอเป็นผลคะแนนเฉลี่ยรายสาระการเรียนรู้ของแต่ละกลุ่ม และช่วงของคะแนนสอบที่คาบเกี่ยวกันของแต่ละระดับความสามารถ

วิธีเอบีซีจะดำเนินการกำหนดคะแนนจุดตัด 3 รอบ ดังนี้

รอบที่ 1 ผู้เชี่ยวชาญพิจารณาคู่มือเรียงข้อสอบรายสาระการเรียนรู้ และผลการวิเคราะห์กลุ่ม จากนั้นก็ให้ผู้เชี่ยวชาญพิจารณาว่าผู้ที่สามารถตอบข้อสอบแต่ละข้อได้ถูกต้องควรมีความสามารถอย่างน้อยระดับใด โดยมีหลักการพิจารณาเพิ่มเติมคือหากข้อใดที่ผู้เชี่ยวชาญไม่มั่นใจก็ให้ระบุไว้ในข้อสอบ ดังนั้นข้อสอบที่ผู้เชี่ยวชาญกำหนดไว้ในแต่ละระดับความสามารถก็จะมีสอง

กลุ่ม คือกลุ่มที่ผู้เชี่ยวชาญมั่นใจและกลุ่มที่ผู้เชี่ยวชาญไม่มั่นใจ ผู้วิจัยคิดว่าข้อมูลเกี่ยวกับผลคะแนนเฉลี่ยรายสาระการเรียนรู้ของแต่ละกลุ่มน่าจะเป็นข้อมูลที่เป็นประโยชน์ต่อการตัดสินใจกำหนดข้อสอบไว้ในแต่ละกลุ่ม ส่วนข้อมูลเกี่ยวกับช่วงของคะแนนสอบที่คาบเกี่ยวกันของแต่ละกลุ่มน่าจะเป็นประโยชน์ต่อผู้เชี่ยวชาญในแง่การเปรียบเทียบผลการกำหนดคะแนนจุดตัดของผู้เชี่ยวชาญเองกับช่วงคะแนนที่คาบเกี่ยวกันของแต่ละกลุ่มที่ถูกกำหนดโดยการวิเคราะห์กลุ่ม

รอบที่ 2 ผู้เชี่ยวชาญร่วมกันอภิปรายกลุ่มย่อยเกี่ยวกับข้อสอบที่ผู้เชี่ยวชาญยังไม่มั่นใจในรอบที่ 1 เพื่อให้การอภิปรายกระชับและตรงประเด็นมากขึ้น และกำหนดคะแนนจุดตัดอีกครั้ง โดยอาศัยหลักการเดียวกับรอบที่ 1

รอบที่ 3 ผู้ตัดสินร่วมกันอภิปรายกลุ่มใหญ่เกี่ยวกับข้อสอบที่ผู้ตัดสินยังไม่มั่นใจในรอบที่ 2 และกำหนดคะแนนจุดตัดอีกครั้ง โดยอาศัยหลักการเดียวกับรอบที่ 1

ข้อดีและข้อด้อย

ข้อดี

วิธีเอบีซีเสนอคู่มือการจัดกลุ่มผู้สอบด้วยวิธีการวิเคราะห์กลุ่ม ซึ่งถือว่าเป็นสารสนเทศที่ช่วยให้ผู้ตัดสินมีความมั่นใจในการกำหนดคะแนนจุดตัดมากขึ้น

ข้อด้อย

วิธีเอบีซี เป็นวิธีที่มีการเสนอสารสนเทศจำนวนมาก ดังนั้นจึงอาจต้องใช้เวลาและค่าใช้จ่ายในการดำเนินการเพื่อเตรียมสาระสนเทศมาก

1.5.2. วิธีผู้สอบเป็นศูนย์กลาง (Examinee-centered methods)

1.5.2.1 วิธีกลุ่มคาบเส้น (Borderline-group method)

การกำหนดคะแนนจุดตัดโดยวิธีกลุ่มคาบเส้นนั้น ผู้เชี่ยวชาญจะต้องพิจารณาเพื่อระบุผู้สอบที่มีความสามารถระดับคาบเส้น (Borderline-group) โดยผู้เชี่ยวชาญควรเป็นครู หรือผู้ควบคุมดูแล (supervisors) ซึ่งเป็นผู้ที่ใกล้ชิดกับผู้สอบแต่ละคน โดยผู้เชี่ยวชาญจะระบุผู้สอบที่มีความสามารถระดับคาบเส้น จากนั้นก็จะคำนวณคะแนนมัธยฐานของคะแนนสอบของผู้สอบที่มีความสามารถระดับคาบเส้น ก็จะได้คะแนนจุดตัด (Näsström & Nyström, 2008; Zieky et al., 2008)

สรุปขั้นตอนวิธีกลุ่มคาบเส้นดังนี้

1) นิยามความหมายของคำว่าระดับความรู้และทักษะของผู้สอบที่อยู่ในระดับคาบเส้น

2) ระบุกลุ่มผู้สอบที่เป็นกลุ่มคาบเส้น

3) เก็บรวบรวมคะแนนที่ได้จากการสอบของผู้สอบในกลุ่มคาบเส้น

4) หาคะแนนมัธยฐานของผู้สอบในกลุ่มคาบเส้นให้เป็นคะแนนจุดตัด

ข้อดีและข้อด้อย

ข้อดี

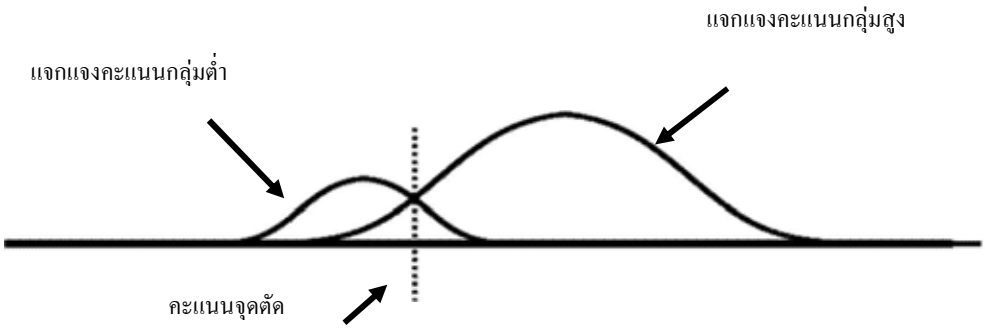
การกำหนดคะแนนจุดตัดโดยวิธีกลุ่มคาบเส้น คือ ง่ายต่อการทำความเข้าใจและนำไปใช้

ข้อด้อย

การกำหนดคะแนนจุดตัดโดยวิธีกลุ่มคาบเส้น คือ การจะตัดสินว่าผู้สอบคนใดมีความสามารถในระดับคาบเส้นนั้นทำได้ยากหากผู้ตัดสินไม่ได้คลุกคลี และสังเกตผู้สอบอย่างจริงจังอาจทำให้การระบุผู้สอบที่มีความสามารถคาบเส้นผิดพลาดได้

1.5.2.2 วิธีกลุ่มตรงข้าม (Contrasting-group method) (Linn, 1989; Zieky, et al., 2008)

การกำหนดคะแนนจุดตัดด้วยวิธีกลุ่มตรงข้ามนั้น ผู้เชี่ยวชาญควรจะเป็นครูหรือผู้ที่มีประสบการณ์ในการให้คะแนนและจัดอันดับการปฏิบัติงานโดยเริ่มจากการแบ่งกลุ่มผู้สอบออกเป็น 2 กลุ่ม คือ กลุ่มที่มีระดับการปฏิบัติงานสูงกว่ามาตรฐาน และกลุ่มที่มีระดับการปฏิบัติงานต่ำกว่ามาตรฐาน หลังจากนั้นก็แจกแจงคะแนนสำหรับสองกลุ่มเพื่อใช้ตัดสินและกำหนดคะแนนจุดตัดเพื่อจำแนกความเป็นไปได้ระหว่างกลุ่มสูงและกลุ่มต่ำโดยจุดที่คะแนนการแจกแจงของผู้สอบทั้งสองกลุ่มตัดกันจะเป็นคะแนนจุดตัด



ภาพที่ 2.4 การแจกแจงคะแนนของกลุ่มสูงกลุ่มต่ำและคะแนนจุดตัด

ข้อดีและข้อด้อย

ข้อดี

การกำหนดคะแนนจุดตัดโดยวิธีกลุ่มตรงข้ามเป็นวิธีที่ตัดสินจากการปฏิบัติงานของผู้สอบและเป็นวิธีที่ง่ายต่อการอธิบายให้ผู้เชี่ยวชาญเข้าใจวิธีการกำหนดคะแนนจุดตัด

ข้อด้อย

การกำหนดคะแนนจุดตัดโดยวิธีกลุ่มตรงข้ามมีข้อบกพร่อง คือหากกำหนดคะแนนจุดตัดหลายค่า ผู้เชี่ยวชาญก็ต้องกำหนดกลุ่มเป็นหลาย ๆ กลุ่มตามระดับความสามารถดังนั้นต้องใช้เวลาในการกำหนดคะแนนจุดตัดนานและการกำหนดผู้สอบเป็นกลุ่มระดับความสามารถต่างๆ ทำได้ค่อนข้างยากหากผู้เชี่ยวชาญไม่มีความรู้ในเรื่องที่ตัดสินดีพอ

1.5.2.3 วิธีเลือกงานผู้สอบ (Examinee Paper Selections Method)

วิธีเลือกงานผู้สอบเหมาะสำหรับการกำหนดคะแนนจุดตัดสำหรับมาตรฐานการปฏิบัติงานที่ให้คะแนนมากกว่า 2 ค่า ซึ่งวิธีนี้คล้ายกับวิธีกลุ่มคาบเส้น (borderline group method) โดยผู้เชี่ยวชาญจะเลือกพิจารณาชิ้นงานที่เป็นตัวแทนของงานทั้งหมดของผู้สอบคาบเส้น คะแนนเฉลี่ยจากชิ้นงานที่ถูกเลือกจะถูกใช้เป็นค่าการผ่านขั้นต่ำ (minimum passing value) และจะถูกรวบรวมเพื่อใช้กำหนดเป็นคะแนนจุดตัด

ข้อดีและข้อด้อย

ข้อดี

เป็นการกำหนดคะแนนจุดตัดโดยการใช้ผลงานประกอบการพิจารณา ทำให้ผลการกำหนดคะแนนจุดตัดมีความน่าเชื่อถือ เนื่องจากเป็นการตัดสินจากผลการปฏิบัติจริงและมีผลงานเป็นหลักฐานประกอบการตัดสิน

ข้อด้อย

วิธีนี้เหมาะสำหรับการตัดสินกับกลุ่มเล็ก ๆ เท่านั้น และการเลือกชิ้นงานเพื่อเป็นตัวแทนในการพิจารณานั้นทำได้ยาก (Hambleton, 2000)

1.5.2.4 วิธีองค์รวม (Holistic Method)

วิธีองค์รวมเป็นวิธีที่คล้ายกับวิธีเลือกงานผู้สอบ (Examinee Paper Selections Method) โดยผู้เชี่ยวชาญจะกำหนดคะแนนจุดตัดโดยตัดสินจากผลงานที่สมบูรณ์ และตัดสินว่าชิ้นงานชิ้นใดเป็นตัวแทนของผลงานของกลุ่มผู้สอบคาบเส้น คะแนนเฉลี่ยจากชิ้นงานที่ถูกเลือกจะถูกใช้เป็นค่าการผ่านขั้นต่ำ (minimum passing value) และถูกรวบรวมเพื่อใช้กำหนดเป็นคะแนนจุดตัด

ข้อดีและข้อด้อย

<u>ข้อดี</u>

ผู้เชี่ยวชาญกำหนดคะแนนจุดตัดจากชิ้นงานของผู้สอบจริงๆ

<u>ข้อด้อย</u>

วิธีนี้เหมาะสำหรับการตัดสินกับกลุ่มเล็กๆ เท่านั้น และการเลือกชิ้นงานเพื่อเป็นตัวแทนในการพิจารณานั้นทำได้ยาก (Hambleton, 2000)

1.5.2.5 วิธีการวิเคราะห์กลุ่ม (Cluster analysis)

ความเป็นมา

การวิเคราะห์กลุ่มเป็นเทคนิคทางสถิติ ใช้ในการแบ่งหน่วยข้อมูลออกเป็นกลุ่มย่อย โดยให้หน่วยที่อยู่ในกลุ่มเดียวกันมีลักษณะที่สนใจเหมือนหรือคล้ายกัน แต่หน่วยที่อยู่ต่างกลุ่มกันจะมีลักษณะที่สนใจต่างกัน ในปี 1995 Sireci (1995) ได้นำวิธีการวิเคราะห์กลุ่ม มาใช้เพื่อการกำหนดคะแนนจุดตัด ในสถานการณ์ของการกำหนดคะแนนจุดตัดโดยส่วนใหญ่จะกำหนดคะแนนจุดตัดโดยยึดคำอธิบายระดับความสามารถ จากนั้นจึงกำหนดคะแนนที่เป็นจุดตัดของคะแนนผู้มีความสามารถคาบเส้นในแต่ละระดับความสามารถ แต่การวิเคราะห์กลุ่มจะไม่ยึดคำอธิบายระดับความสามารถและกำหนดโดยปราศจากความลำเอียงใดๆ โดยจะกำหนดกลุ่มจากตัวแปรหรือลักษณะที่สนใจ โดยพยายามให้กลุ่มเดียวกันมีลักษณะที่สนใจเหมือนหรือคล้ายกันมากที่สุดและหน่วยที่อยู่ต่างกลุ่มกันจะต้องมีลักษณะที่สนใจต่างกันให้มากที่สุด สำหรับการกำหนดคะแนนจุดตัดนั้น ลักษณะที่สนใจคือผลการตอบข้อสอบ และผลจากการวิเคราะห์กลุ่มคือผู้สอบจะถูกแบ่งออกเป็นกลุ่มๆ โดยแต่ละกลุ่มจะแทนผู้มีความสามารถแต่ละระดับ (Kane, 2001 cited in Hess, 2007)

ประเภทของการวิเคราะห์กลุ่ม

การวิเคราะห์กลุ่มจะมีเทคนิคการรวมกลุ่มที่นิยมใช้มากที่สุด 2 เทคนิค คือ การวิเคราะห์กลุ่มแบบลดหลั่น (Hierarchical Cluster Analysis: HCA) และการวิเคราะห์กลุ่มแบบไม่ลดหลั่น (Nonhierarchical Cluster Analysis หรือ K-Means Clustering) โดยมีรายละเอียดดังนี้

1) การวิเคราะห์กลุ่มแบบลดหลั่น (Hierarchical Cluster Analysis: HCA)

การวิเคราะห์กลุ่มแบบลดหลั่นเป็นการจัดกลุ่มที่เกี่ยวกับโครงสร้างของการลดหลั่น แบ่งได้เป็น 2 วิธี คือ

1.1) การสร้างกลุ่มแบบรวมเข้า (Agglomerative) เป็นการรวมกลุ่มที่แต่ละหน่วยที่ใกล้กันจะรวมกันเป็นกลุ่มด้วยตนเอง ต่อมาแต่ละกลุ่มที่ใกล้กันก็จะรวมกันเป็นกลุ่มที่ใหญ่มากขึ้นเรื่อยๆ

ขั้นตอนของเทคนิคการสร้างกลุ่มแบบรวมเข้า (Agglomerative)

ขั้นที่ 1 เลือกตัวแปรที่คาดว่าจะทำให้หน่วยต่างกัน

ขั้นที่ 2 เลือกว่าจะวัดความคล้ายหรือความต่าง

ขั้นที่ 3 คำนวณค่าความคล้ายหรือความต่างที่เลือกแต่ละคู่ โดยคู่ที่มีค่าความคล้ายมากที่สุด หรือมีค่าความต่างต่ำที่สุดจะรวมเป็นกลุ่มเดียวกัน

ขั้นที่ 4 คำนวณค่าความคล้ายหรือความต่างของแต่ละคู่ของกลุ่มที่เหลือ แล้วเลือกวิธีการรวมกลุ่มเพื่อลดจำนวนกลุ่ม

1.2) การสร้างกลุ่มแบบแบ่งออก (Divisive) เป็นการรวมกลุ่มที่ทุกหน่วยจะรวมเป็นกลุ่มเดียวกัน จากนั้นกลุ่มที่ไม่เหมือนก็จะถูกแยกออกไปเป็นกลุ่มย่อย ๆ จนกระทั่งแต่ละกลุ่มที่ได้เป็นกลุ่มที่แตกต่างกัน

การแบ่งกลุ่มแบบเป็นขั้นตอน มีข้อจำกัดคือเมื่อนำหน่วยใดไว้ในกลุ่มใดแล้ว จะไม่มีการย้ายหน่วยนั้นไว้กลุ่มอื่นอีก และไม่เหมาะสำหรับการวิเคราะห์กับข้อมูลจำนวนมาก (Sireci, 2001 cited in Hess, 2007) และเหมาะสำหรับการวิเคราะห์ในกรณีที่ข้อมูลมีจำนวนไม่มาก ($n < 200$) (กัลยา วานิชบัญชา, 2550)

2) การวิเคราะห์กลุ่มแบบไม่ลดหลั่น (Nonhierarchical Cluster Analysis หรือ K-Means Clustering)

การวิเคราะห์กลุ่มแบบไม่เป็นขั้นตอน เป็นเทคนิคการจัดกลุ่มโดยการกำหนดจำนวนกลุ่มที่ต้องการแบ่งไว้ก่อน โดย Hartigan (1975 อ้างถึงใน กัลยา วานิชบัญชา, 2550) สรุปขั้นตอนการวิเคราะห์ K-Means Clustering ได้ดังนี้

ขั้นที่ 1 แบ่งข้อมูลออกเป็น k กลุ่ม ด้วยวิธีใดวิธีหนึ่งดังนี้

- แบ่งอย่างสุ่ม
- แบ่งด้วยผู้ศึกษาเอง

ขั้นที่ 2 คำนวณหาจุดกึ่งกลางกลุ่มของแต่ละกลุ่ม

ขั้นที่ 3 คำนวณหาระยะห่าง

แบบที่ 1 คำนวณระยะห่างจากแต่ละหน่วยไปยังจุดกลางกลุ่มของทุกกลุ่มและจะพิจารณาย้ายหน่วยไปยังกลุ่มที่มีระยะห่างต่ำที่สุด

แบบที่ 2 คำนวณระยะห่างกำลังสองของแต่ละหน่วยไปยังจุดกลางกลุ่มที่หน่วยนั้นอยู่ โดยให้ ESS (Error Sum Square) เท่ากับระยะห่างกำลังสองของแต่ละหน่วยไปยังจุดกลางกลุ่ม

$$\text{ESS} = \sum_{i=1}^{n} \left(x_i - \bar{x}_{c_{(i)}}\right)' \left(x_i - \bar{x}_{c_{(i)}}\right)$$

เมื่อ $C_{(i)}$ หมายถึง กลุ่มของหน่วยที่ i

ESS หมายถึง ผลบวกของระยะห่างจากแต่ละหน่วยในกลุ่มไปยังจุดกลางกลุ่มรวมทุกกลุ่ม

กลุ่มใดมีค่า ESS ต่ำ แสดงว่าหน่วยที่อยู่ในกลุ่มนั้นมีความคล้ายคลึงกัน

ขั้นที่ 4 พิจารณาย้ายกลุ่ม

แบบที่ 1 ย้ายหน่วยที่ i ไปยังกลุ่มที่ทำให้ระยะห่างจากหน่วยที่ i ไปยังจุดกลางกลุ่มมีค่าต่ำสุด

แบบที่ 2 ย้ายหน่วยที่ i ไปยังกลุ่มที่ทำให้ค่า ESS มีค่าต่ำที่สุด

ถ้าในขั้นตอนที่ 4 ไม่มีการย้ายกลุ่มอีกแล้ว แสดงว่ากลุ่มที่แบ่งได้นั้นเหมาะสมแล้ว แต่ถ้าในขั้นตอนที่ 4 มีการย้ายกลุ่ม กลุ่มที่มีการย้ายเข้าหรือออกจะต้องทำการคำนวณหาจุดกลางกลุ่มใหม่ และกลับไปทำตามขั้นตอนที่ 2 อีกครั้ง

<u>การวัดความคล้าย</u>

การจัดข้อมูลเข้ากลุ่มจะพิจารณาจากระยะห่างของข้อมูล ซึ่งการวัดระยะทางของข้อมูลที่สำคัญมี 6 วิธี ดังนี้

1. ระยะทางยูคลิด (Euclidean Distance) เป็นรากที่สองของผลรวมของผลต่างกำลังสองของข้อมูลแต่ละตัว

ระยะทาง (X,Y) = $\sqrt{\sum_i (X_i - Y_i)^2}$

2. ระยะทางยูคลิดกำลังสอง (Square Euclidean Distance) เป็นผลรวมของผลต่างกำลังสองของข้อมูลแต่ละตัว ใช้กับการจัดกลุ่มแบบศูนย์กลาง (Center) มัธยฐาน (Median) และ วิธีของ Ward

ระยะทาง (X, Y) = $\sum_i (X_i - Y_i)^2$

3. โคซายน์ (Cosine, Cosine of Vectors of Variables)

ความคล้ายกัน (X, Y) = $\dfrac{\sum_i (X_i - Y_i)}{\sqrt{\sum_i (X_i)^2 (Y_i)^2}}$

4. บล็อก (Block, City-Block หรือ Manhattan Distance) เป็นผลรวมของค่าสัมบูรณ์ของผลต่างของข้อมูลแต่ละตัว

ระยะทาง (X, Y) = $\sum_i |X_i - Y_i|$

5. เชบิ้เชฟ (Chebechev, Chebechev Distance Metric) เป็นค่าสูงสุดของค่าสัมบูรณ์ของผลต่างของข้อมูลแต่ละตัว

ระยะทาง (X, Y) = $Max|X_i - Y|$

6. กำลัง (Power (p.r), Distance in an Absolute Power Metric) เป็นความแตกต่างระหว่าง 2 กรณีในรากที่ r (r^{th} root) ของความแตกต่างสัมบูรณ์ในกำลังที่ p (p^{th} Power) ในค่าของตัวแปรแต่ละตัว

กำลัง (X, Y) = $\left(\sum_{i\ i} |X_i - Y_i|^p\right)^{1/r}$

ข้อดีและข้อด้อย

ข้อดี

เป็นการกำหนดคะแนนจุดตัดโดยปราศจากความลำเอียงใดๆ

ข้อด้อย

เนื่องจากเป็นการวิเคราะห์เพื่อแบ่งกลุ่มจากตัวแปรที่สนใจ อันได้แก่ ผลการตอบข้อสอบ ดังนั้นจึงไม่สามารถกำหนดคะแนนจุดตัดที่สูงกว่าหรือต่ำกว่าผลการตอบข้อสอบของนักเรียนกลุ่มที่เลือกมาวิเคราะห์

จากการศึกษาสรุปได้ว่า วิธีแบบสอบเป็นศูนย์กลางเป็นวิธีที่ใช้ง่าย เหมาะกับแบบสอบปรนัยหรือแบบตอบสั้น แต่วิธีแบบสอบเป็นศูนย์กลางไม่เหมาะกับแบบสอบการปฏิบัติ ส่วนวิธีผู้สอบเป็นศูนย์กลาง เหมาะสำหรับการปฏิบัติตามสภาพจริง

ตารางที่ 2.3 สรุปรายละเอียดเกี่ยวกับวิธีการกำหนดคะแนนจุดตัด

วิธี	การพิจารณา	ข้อดี	ข้อด้อย
1. วิธีแบบสอบเป็นศูนย์กลาง			
1.1 วิธีแองกอฟ	ผู้ตัดสินพิจารณาความน่าจะเป็นที่ผู้มีความสามารถคาบเส้นจะสามารถตอบข้อสอบแต่ละข้อได้ถูกต้องอย่างอิสระ	1. วิธีเป็นวิธีที่ง่ายต่อการนำไปปฏิบัติ 2.ใช้สถิติในการวิเคราะห์ไม่ยาก	การพิจารณากำหนดคะแนนจุดตัดต้องต้องใช้เวลาในการพิจารณามาก
1.2 วิธีบุ๊คมาร์ค	ผู้ตัดสินพิจารณากำหนดคะแนนจุดตัดโดยพิจารณาจากคู่มือเรียงข้อสอบ ซึ่งจัดทำโดยการเรียงข้อสอบจากข้อง่ายไปข้อยาก	1. กำหนดคะแนนจุดตัดหลายค่าโดยใช้ข้อสอบเพียงฉบับเดียว 2. ใช้ได้ทั้งกับข้อสอบที่ให้คะแนน 2 ค่า และแบบให้คะแนนมากกว่า 2 ค่า 3. วิธีนี้ได้คำนวณความยากเพื่อประกอบการพิจารณาทำให้ตัดสินได้ง่ายขึ้น และทำให้คะแนนจุดตัดมีความน่าเชื่อถือมากขึ้น	1.วิธีการนี้ต้องใช้ข้อสอบจำนวนมาก มิฉะนั้นอาจเกิด floor and ceiling effect ได้ 2. ผลการตัดสินขึ้นอยู่ที่คณะผู้เชี่ยวชาญที่ใช้ตัดสิน

ตารางที่ 2.3 สรุปรายละเอียดเกี่ยวกับวิธีการกำหนดคะแนนจุดตัด

วิธี	การพิจารณา	ข้อดี	ข้อด้อย
1.3 วิธีของอีเบล	ผู้ตัดสินกำหนดร้อยละที่คาดหวังที่ผู้มีความสามารถคาบเส้นจะสามารถผ่านได้ถือว่าเป็นปริมาณความคาดหวังของผู้มีความสามารถคาบเส้นที่จะสอบผ่าน	ผู้ตัดสินพิจารณาเพื่อกำหนดคะแนนจุดตัดเพียง 12 ลักษณะเท่านั้นคือ พิจารณาตามลักษณะที่แบ่งไว้ทั้ง 12 ลักษณะซึ่งถือว่าเป็นการตัดสินที่ทำได้ง่ายและใช้เวลาไม่มาก	การคำนวณคะแนนจุดตัดที่ต้องนำจะจำนวนข้อสอบในแต่ละลักษณะไปคูณกับค่าร้อยละที่คาดหวังนั้นเป็นเรื่องที่ค่อนข้างยุ่งยากและเสียเวลามาก
1.4 วิธีของนีเดลสกี	ผู้ตัดสินพิจารณาตัดตัวที่ผู้มีความสามารถคาบเส้นน่าจะตอบผิดระดับการผ่านขั้นต่ำสำหรับข้อสอบถูกคำนวณจากจำนวนตัวลวงที่ยังคงเหลืออยู่	มีประโยชน์มากในศาสตร์ที่ตัวเลือกที่ผิดมีความสำคัญเพราะทำให้เข้าใจว่าผู้สอบมีแนวคิดเกี่ยวกับเรื่องที่สอบผิด ในเรื่องใด	ใช้เวลาในการกำหนดคะแนนจุดตัดมาก
2. วิธีผู้สอบเป็นศูนย์กลาง			
2.1 วิธีกลุ่มคาบเส้น	ผู้ตัดสินจะต้องพิจารณาเพื่อระบุผู้สอบที่มีความสามารถระดับคาบเส้น ซึ่งผู้สอบที่มีความสามารถคาบเส้นคือผู้ที่มีความสามารถในระดับที่ผ่านมาตรฐานพอดี	ง่ายในการนำไปใช้และง่ายในการทำความเข้าใจ	การจะตัดสินว่าผู้สอบคนใดมีความสามารถในระดับคาบเส้นนั้นทำได้ยากหากผู้ตัดสินไม่ได้คลุกคลี และสังเกตผู้สอบอย่างจริงจังอาจทำให้การระบุผู้สอบที่มีความสามารถคาบเส้นทำได้ยาก
2.2 วิธีกลุ่มตรงข้าม	ผู้ตัดสินควรจะเป็นครู หรือผู้ที่มีประสบการณ์ในการให้คะแนนและจัดอันดับการปฏิบัติงาน เพื่อพิจารณาแบ่งผู้สอบออกเป็น 2 กลุ่ม คือ กลุ่มที่มีระดับการปฏิบัติงานสูงกว่ามาตรฐาน และกลุ่มที่มีระดับการปฏิบัติงานต่ำกว่ามาตรฐาน หลังจากนั้นก็แจกแจงคะแนนสำหรับสองกลุ่มเพื่อใช้ตัดสินและกำหนดคะแนนจุดตัด	เป็นวิธีที่ง่ายต่อการอธิบายให้ผู้ตัดสินเข้าใจวิธีการกำหนดคะแนนจุดตัด	หากกำหนดคะแนนจุดตัดหลายค่า ผู้ตัดสินก็ต้องกำหนดกลุ่มเป็นหลาย ๆ กลุ่มระดับความสามารถซึ่งต้องใช้เวลาในการพิจารณามาก

ตารางที่ 2.3 สรุปรายละเอียดเกี่ยวกับวิธีการกำหนดคะแนนจุดตัด

วิธี	การพิจารณา	ข้อดี	ข้อด้อย
2.3 วิธีเลือกงานผู้สอบ	ผู้ตัดสินจะเลือกพิจารณาชิ้นงานที่เป็นตัวแทนของงานทั้งหมด คะแนนเฉลี่ยจากชิ้นงานที่ถูกเลือกในแต่ละระดับจะถูกใช้เป็นค่าการผ่านขั้นต่ำ และจะถูกรวบรวมเพื่อใช้ตัดสินมาตรฐานการปฏิบัติแต่ละมาตรฐาน	เป็นการกำหนดคะแนนจุดตัดโดยการใช้ผลงานประกอบการพิจารณา	วิธีนี้เหมาะสำหรับการตัดสินกับกลุ่มเล็ก ๆ เท่านั้น และการเลือกชิ้นงานเพื่อเป็นตัวแทนในการพิจารณานั้นทำได้ยาก
2.4 วิธีองค์รวม	ผู้ตัดสินจะกำหนดคะแนนจุดตัดโดยตัดสินจากผลงานที่สมบูรณ์ และตัดสินว่าชิ้นงานชิ้นใดเป็นตัวแทนของผลงานของกลุ่มผู้สอบคาบเส้น	ผู้ตัดสินกำหนดคะแนนจุดตัดจากชิ้นงานของผู้สอบจริงๆ	เหมาะสำหรับการตัดสินกับกลุ่มเล็ก ๆ และการเลือกชิ้นงานเพื่อเป็นตัวแทนในการพิจารณานั้นทำได้ยาก
2.5 การวิเคราะห์กลุ่ม	วิเคราะห์เพื่อกำหนดกลุ่มจากตัวแปรหรือลักษณะที่สนใจ โดยพยายามให้กลุ่มเดียวกันมีลักษณะที่สนใจเหมือนหรือคล้ายกันมากที่สุดและหน่วยที่อยู่ต่างกลุ่มกันจะต้องมีลักษณะที่สนใจต่างกันให้มากที่สุด	เป็นการกำหนดมาตรฐานความสามารถโดยปราศจากความลำเอียงใดๆ	เนื่องจากเป็นการวิเคราะห์เพื่อแบ่งกลุ่มจากตัวแปรที่สนใจ อันได้แก่ ผลการตอบข้อสอบ ดังนั้นจึงไม่สามารถกำหนดคะแนนจุดตัดที่สูงกว่าหรือต่ำกว่าผลการตอบข้อสอบของนักเรียนกลุ่มที่เลือกมาวิเคราะห์

1.6 การเลือกวิธีกำหนดคะแนนจุดตัด (Reckase, 2006; Cizek, 2005; Plake, 2008; Zieky et al., 2008)

จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้อง พบว่า มีผู้ให้ความคิดเห็นเกี่ยวกับการเลือกใช้วิธีการกำหนดจุดตัดซึ่งสามารถแบ่งเป็น 2 กลุ่มดังนี้

กลุ่มแรก กล่าวว่า การเลือกวิธีกำหนดจุดตัดควรเลือกหลายวิธี หรืออย่างน้อย 2 วิธีในการกำหนดจุดตัดเรื่องใดเรื่องหนึ่ง เพื่อเป็นการยืนยันผลที่ได้จากหลายๆ วิธีเพื่อให้ได้คะแนนจุดตัดที่มีความถูกต้องและเหมาะสมที่สุด

กลุ่มที่สอง กล่าวว่า การเลือกวิธีการกำหนดจุดตัดควรเลือกใช้เพียงวิธีเดียวและดำเนินการตามวิธีนั้นอย่างเคร่งครัด และทำให้ถูกต้องที่สุดเท่าที่จะทำได้เพื่อให้ผลการกำหนดคะแนนจุดตัดมีความถูกต้องและน่าเชื่อถือ

1.7 การตรวจสอบคุณภาพของคะแนนจุดตัด

การกำหนดคะแนนจุดตัดมิได้สิ้นสุดเพียงแค่การได้มาซึ่งคะแนนจุดตัดที่ต้องการ แต่ยังมีกระบวนการที่มีความสำคัญอย่างยิ่งอีกกระบวนการหนึ่งที่ต้องดำเนินการหลังจากการกำหนดคะแนนจุดตัดแล้วคือ การตรวจสอบคุณภาพของคะแนนจุดตัด (Plake, 2008; Zieky et al., 2008) มีนักวิชาการได้เสนอวิธีการตรวจสอบคุณภาพของคะแนนจุดตัด มีรายละเอียดดังนี้

1.7.1 ความเที่ยง

งานวิจัยส่วนใหญ่จะตรวจสอบความเที่ยงโดยการศึกษาความสอดคล้อง (Consistency) ที่กำหนดโดยผู้เชี่ยวชาญต่างกลุ่ม ต่างโอกาส (occasion) เป็นต้น โดยงานวิจัยส่วนใหญ่มีการตรวจสอบความเที่ยงใน 2 ประเด็นดังนี้ 1) ความเที่ยงภายในผู้เชี่ยวชาญ (Intra-rater reliability) และ 2) ความเที่ยงระหว่างผู้เชี่ยวชาญ (Inter-rater reliability) (Kane, 1994; Näsström & Nyström, 2008; Papageorgiou, 2010) โดยมีรายละเอียดดังนี้

1.7.1.1 ความเที่ยงภายในผู้เชี่ยวชาญ เป็นการตรวจสอบจากผลการกำหนดคะแนนจุดตัดของผู้เชี่ยวชาญในแต่ละรอบว่าเท่าเดิมหรือใกล้เคียงกันทุกรอบหรือไม่ (Näsström & Nyström, 2008) โดยการวิเคราะห์ค่าสัมประสิทธิ์แอลฟ่า (Cronbach's alpha coefficient : α) เป็นการพิจารณาความสอดคล้องภายใน (Internal Consistency) ด้วยการประเมินความใกล้เคียงของผลการกำหนดคะแนนจุดตัดของผู้เชี่ยวชาญ

1.7.1.2 ความเที่ยงระหว่างผู้เชี่ยวชาญ เป็นการตรวจสอบความความสอดคล้องของผลการกำหนดคะแนนจุดตัดระหว่างผู้เชี่ยวชาญ (Consistent across Panelists) เป็นการตรวจสอบว่าผู้เชี่ยวชาญมีความเห็นในทิศทางเดียวกันหรือไม่ในการจัดกลุ่มตัวอย่างลงในกลุ่มต่างๆ ซึ่งสามารถทำได้โดยการวัดความสอดคล้องกัน (Measure of Agreement) โดยการวิเคราะห์ค่าสัมประสิทธิ์สหสัมพันธ์ภายในกลุ่ม (Intraclass Correlation: ICC) และการวิเคราะห์ Kendall's W ซึ่งเป็นค่าที่บอกความสอดคล้องกันของคะแนนจุดตัดระหว่างผู้เชี่ยวชาญ (Generalitat de Catalunya, 2006 cited in Papageorgiou, 2010)

นอกจากนี้เนื่องจากการกำหนดคะแนนจุดตัดส่วนใหญ่เป็นการดำเนินการกำหนดโดยอาศัยกลุ่มผู้เชี่ยวชาญเป็นหลัก และมีการดำเนินการกำหนดคะแนนจุดตัดหลายรอบ ซึ่งกล่าวได้ว่าในสถานการณ์หรือเงื่อนไขต่าง ๆ ของการกำหนดคะแนนจุดตัดย่อมเกิดความความแปรปรวนของคะแนนจุดตัดอันเนื่องมาจากแหล่งความคลาดเคลื่อนต่าง ๆ เช่น ความคลาดเคลื่อนจากจำนวนผู้เชี่ยวชาญ ความคลาดเคลื่อนจากจำนวนกลุ่มย่อย และความคลาดเคลื่อนจากจำนวนรอบเป็นต้น ดังนั้นจึงควรมีการวิเคราะห์ข้อมูลเพื่อให้ทราบถึงแหล่งความแปรปรวนที่ส่งผลต่อการกำหนดคะแนนจุดตัด โดยอาศัยทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือ (Generalizability Theory) ซึ่งเป็นทฤษฎีที่เหมาะกับการประเมินการปฏิบัติของผู้เชี่ยวชาญ (Brennan, 2001, p.117) และสามารถนำมาศึกษาในกรณีที่มีความคลาดเคลื่อนหลายแหล่งได้ (ศิริชัย กาญจนวาสี, 2550, p.11) ดังนั้นจึงควรศึกษาขนาดองค์ประกอบความแปรปรวน และวิเคราะห์ค่าสัมประสิทธิ์การสรุปอ้างอิงเพื่อให้ทราบถึงขนาดของแหล่งความแปรปรวนที่ส่งผลต่อคะแนนจุดตัด อันจะเป็นข้อมูลสำคัญสำหรับการวางแผนเพื่อให้สามารถควบคุมแหล่งความคลาดเคลื่อนที่จะเกิดขึ้นจากการกำหนดคะแนนจุดตัดได้ตรงประเด็น และได้สารสนเทศสำหรับการตัดสินใจเกี่ยวกับจำนวนผู้เชี่ยวชาญ จำนวนกลุ่มย่อยของผู้เชี่ยวชาญ และจำนวนรอบที่เหมาะสมเพื่อให้สามารถกำหนดคะแนนจุดตัดที่มีความเที่ยงสูงถึงระดับที่ต้องการต่อไป

1.7.2 ความตรง

ความตรงของคะแนนจุดตัด คือคุณสมบัติของคะแนนจุดตัดที่สะท้อนถึงความรู้และทักษะที่ผู้มีความสามารถแต่ละระดับมี (McGinty, 2005) โดยคุณสมบัติด้านความตรงถือว่าเป็นคุณสมบัติที่สำคัญที่สุด แต่ก็เป็นคุณสมบัติที่ตรวจสอบได้ยากที่สุด

McGinty (2005) เสนอว่า การตรวจสอบความตรงควรตรวจสอบถึงการได้มาซึ่งคะแนนจุดตัดว่ามีความตรงหรือไม่ โดยเสนอให้ตรวจสอบความตรงของกระบวนการกำหนดคะแนนจุดตัดอันประกอบด้วย ปัจจัยนำเข้า (input) กระบวนการ (processes) ผลผลิต (output) และ ผลที่ตามมา (consequence) โดยปัจจัยนำเข้า ประกอบด้วย เป้าหมายของการกำหนดคะแนนจุดตัด คำอธิบายระดับความสามารถ วิธีการกำหนดคะแนนจุดตัดที่เลือก การเลือกผู้ตัดสิน การฝึกอบรม และสารสนเทศที่เสนอต่อผู้ตัดสิน ส่วนกระบวนการเป็นกระบวนการคิดตัดสินใจของผู้ตัดในการกำหนดคะแนนจุดตัดสำหรับแต่ละระดับความสามารถ ผลผลิตคือคะแนนจุดตัดที่กำหนด ผลที่

ตามมา คือผลจากการนำคะแนนจุดตัดไปใช้ตามวัตถุประสงค์ ได้แก่ อัตราการผ่านหรือตก ความคลาดเคลื่อนของการจัดกลุ่ม เป็นต้น

McGinty (2005) กล่าวว่า การตรวจสอบคุณภาพด้านความตรงของคะแนนจุดตัดควรมีการตรวจสอบความตรงขององค์ประกอบทั้ง 4 ในกระบวนการกำหนดคะแนนจุดตัดที่กล่าวมา

1.7.2.1 การตรวจสอบความตรงของปัจจัยนำเข้า Kane (1994) เสนอว่าควรให้ผู้เชี่ยวชาญเป็นผู้ตรวจสอบในประเด็นต่อไปนี้

1.7.2.1.1. ความถูกต้อง

การตรวจสอบระดับความถูกต้องของปัจจัยนำเข้าซึ่งประกอบด้วย 1) เป้าหมายในการกำหนดคะแนนจุดตัด และ 2) วิธีการและขั้นตอนในการกำหนดคะแนนจุดตัด นอกจากนี้ Van Der Linden (1995 cited in Brennan, 2006) เสนอให้ตรวจสอบความถูกต้องของวิธีการในการดำเนินการเช่นกัน

1.7.2.1.2 ความเป็นไปได้

การตรวจสอบในเรื่องของความเป็นไปได้นั้น (Kane, 1994) เสนอให้พิจารณา 2 เรื่อง โดยเรื่องแรกคือ ความเป็นไปได้ในการปฏิบัติตามขั้นตอนที่ได้กำหนดไว้ และความเป็นไปได้ในการวิเคราะห์ข้อมูลทางสถิติ แม้คนทั่วไปที่ไม่เคยกำหนดคะแนนจุดตัดมาก่อนก็ต้องปฏิบัติได้ นอกจากนี้ได้มีนักวิชาการที่พิจารณาในเรื่องความเป็นไปได้ของการปฏิบัติ อันได้แก่ Chinn และ Hertz (2002) และFerdous และ Plake (2008) โดยให้ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดโดยใช้วิธีแองกอฟแบบใช่/ไม่ใช่ ประเมินในเรื่องของความเป็นไปได้ของวิธีการที่ได้ดำเนินการ ซึ่งผู้เชี่ยวชาญประเมินว่าวิธีแองกอฟแบบใช่/ไม่ใช่ ความเป็นไปได้กว่าการกำหนดคะแนนจุดตัดโดยวิธีแองกอฟส่วน Ferdous และ Plake (2005, 2007) ได้ศึกษาเพื่อหาวิธีลดจำนวนข้อสอบที่ใช้ในการกำหนดคะแนนจุดตัดด้วยวิธีแองกอฟลง และหลังจากที่กำหนดคะแนนจุดตัดแล้วก็ได้ประเมินความง่ายและความสะดวกในการดำเนินการกำหนดคะแนนจุดตัด

1.7.2.1.3 ความเหมาะสม

การตรวจสอบในเรื่องความเหมาะสม นั้น Kane (1994) เสนอว่าควรตรวจสอบความเหมาะสมและคุณภาพของการปฏิบัติในกระบวนการต่อไปนี้คือ 1) กระบวนการฝึกอบรมผู้เชี่ยวชาญ 2) กระบวนการกำหนดคะแนนจุดตัด

1.7.2.1.4 ความเข้าใจและความมั่นใจของผู้ตัดสิน

เป็นการตรวจสอบความตรงโดยอาศัยข้อมูลย้อนกลับของผู้เชี่ยวชาญ มีความสำคัญต่อความน่าเชื่อถือของคะแนนจุดตัดอย่างยิ่งโดย Kane (1994) เสนอให้ตรวจสอบในเรื่องต่อไปนี้ 1) ความเข้าใจเกี่ยวกับนิยามระดับความสามารถ 2) ความเข้าใจเกี่ยวกับกระบวนการกำหนดคะแนนจุดตัด ซึ่งสอดคล้องกับ Skorupski และ Hambleton (2003 cited in Ferdous & Plake, 2005) และ 3) ความมั่นใจต่อคะแนนจุดตัดที่กำหนด ซึ่ง Kane (1994) กล่าวว่า หากผู้เชี่ยวชาญไม่มีความมั่นใจในคะแนนจุดตัดที่กำหนดขึ้น ก็ยากที่ผู้อื่นจะเชื่อถือในคะแนนจุดตัดนั้น นอกจากนี้ Clauser และคณะ (2009) และ Norcini และคณะ (1988) ได้เสนอให้สอบถาม ผู้เชี่ยวชาญในประเด็นเกี่ยวกับผลของการอภิปรายกลุ่มและการเสนอสารสนเทศต่อการตัดสินใจ

1.7.2.1.5 ความเป็นประโยชน์ของเอกสาร

เป็นการตรวจสอบเอกสารที่ใช้ประกอบการกำหนดคะแนนจุดตัด Kane (1994) และ Cizek (1996 cited in Brennan, 2006) กล่าวว่า ต้องมีรายละเอียดที่ถูกต้องและชัดเจน โดยการที่จัดทำเอกสารได้ดีและมีความถูกต้องย่อมทำให้การปฏิบัติมีระบบและเป็นไปอย่างถูกต้องในทุกๆ ขั้นตอน

1.7.2.2 การตรวจสอบความตรงของกระบวนการ หมายถึง การประเมินกระบวนการคิดตัดสินใจของผู้เชี่ยวชาญในการกำหนดคะแนนจุดตัดสำหรับแต่ละระดับความสามารถ ซึ่งกระบวนการนี้เปรียบเสมือนเป็นกล่องดำ (black box) เนื่องจากเป็นกระบวนการที่เกิดขึ้นภายในจิตใจของผู้เชี่ยวชาญที่ไม่สามารถล่วงรู้ได้ McGinty (2005) กล่าวว่า การยืนยันถึงความตรงของคะแนนจุดตัดควรมีการตรวจสอบความตรงของกระบวนการตัดสินใจด้วย เพื่อให้ทราบเกี่ยวกับเหตุผลและหลักการที่ผู้เชี่ยวชาญยึดถือในการคะแนนจุดตัด และทราบเกี่ยวกับความคิด ความกังวลขณะตัดสินใจ และปัญหาที่เกิดขึ้น

นอกจากนี้เพื่อวิเคราะห์ถึงปัจจัยที่มีอิทธิพลต่อการตัดสินใจของผู้ตัดสิน Dawber และคณะ (2002 อ้างถึงใน McGinty, 2005) สำรวจข้อมูลเกี่ยวกับกระบวนการคิดของผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดโดยวิธีบุ๊คมาร์ค โดยใช้แบบสอบถามปลายเปิดและปลายปิด และหลังจากเสร็จสิ้นการตัดสินในแต่ละรอบ ผู้นำแต่ละโต๊ะ (table leader) จะสรุปความคิดของผู้เชี่ยวชาญแต่ละคนการอัดเสียงการสรุปของตน ส่วน Skorupski และ Hambleton (2003 อ้างถึงใน McGinty, 2005)

ศึกษากระบวนการคิดของผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดโดยใช้แบบสอบถาม ส่วน McGinty, 2005 ศึกษากระบวนการคิดของผู้เชี่ยวชาญโดยการร่วมสังเกตการณ์การกำหนดคะแนนจุดตัด

จากรายงานวิจัยที่กล่าวมาข้างต้น สรุปข้อค้นพบที่ได้คือ ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดส่วนใหญ่คิดว่าการกำหนดคะแนนจุดตัดเป็นการตัดสินใจที่ยาก โดยปัจจัยที่มีอิทธิพลต่อการตัดสินใจได้แก่ ความรู้ความสามารถ อารมณ์ สารสนเทศที่เสนอ ความวิตกกังวลต่าง ๆ เช่น ผู้เชี่ยวชาญบางท่านต้องการสร้างมาตรฐานที่สูงแต่ก็กังวลเกี่ยวกับผลของการนำคะแนนจุดตัดไปใช้ บางท่านเกิดกังวลเกรงว่าผลการกำหนดคะแนนจุดตัดจะไม่ถูกนำไปใช้ รวมทั้งปัจจัยทางร่างกาย เช่น ความเมื่อยล้า เป็นต้น

1.7.2.3 การตรวจสอบความตรงของผลผลิต McGinty (2005) กล่าวว่า งานวิจัยส่วนใหญ่ตรวจสอบโดยการหาความสัมพันธ์ระหว่างคะแนนจุดตัดที่กำหนดขึ้นในเรื่องเดียวกัน แต่ใช้ต่างวิธีกัน โดยการนำผลที่ได้จากการกำหนดคะแนนจุดตัดแต่ละวิธีไปใช้แบ่งกลุ่มผู้สอบตามระดับความสามารถที่กำหนดไว้ และวิเคราะห์เพื่อพิจารณาความสอดคล้องของการกำหนดกลุ่มโดยการวิเคราะห์ค่าสัมประสิทธิ์แคปป้า หรืออาจหาความสัมพันธ์ระหว่างผลการกำหนดกลุ่มโดยใช้คะแนนจุดตัดกับการกำหนดกลุ่มโดยใช้ข้อมูลอื่น ๆ ที่สัมพันธ์หรือเกี่ยวข้องกับเรื่องที่กำหนดคะแนนจุดตัด เช่น เกรดหรือผลการปฏิบัติงานในเรื่องเกี่ยวข้องกับเรื่องที่กำหนดคะแนนจุดตัดเป็นต้น มี 2 งานวิจัยที่ตรวจสอบความตรงคือ งานวิจัยของ (Giraud, 2000 cited in Brennan, 2006) และงานวิจัยของสุรีพร อนุศาสนนันท์ (2550) ซึ่งตรวจสอบความตรงโดยการเปรียบเทียบคะแนนจุดตัดกับเกรดที่ได้จากการเรียนในห้องเรียนในวิชาที่เกี่ยวข้องกับเรื่องที่กำหนดคะแนนจุดตัด

1.7.2.4 การตรวจสอบความตรงของผลที่ตามมา สามารถตรวจสอบโดยพิจารณาจากความถูกต้อง และค่าความคลาดเคลื่อนของการแบ่งกลุ่ม ซึ่งประกอบด้วยความคลาดเคลื่อน 2 แบบ ได้แก่ ความคลาดเคลื่อนแบบไม่ยอมรับ หรือเรียกว่า ความคลาดเคลื่อนแบบลบ เป็นการตัดสินว่าเป็น ผู้ไม่ผ่านเกณฑ์ แต่ในสภาพความเป็นจริงแล้วเป็นผู้ที่มีความรู้พอที่จะผ่านเกณฑ์ได้ และความคลาดเคลื่อนแบบยอมรับหรือเรียกว่า ความคลาดเคลื่อนแบบบวกเป็นการตัดสินว่าเป็นผู้ผ่านเกณฑ์ แต่ในสภาพความเป็นจริงแล้วเป็นผู้ที่ไม่มีความรู้พอที่จะผ่านเกณฑ์ โดยคะแนนจุดตัดที่กำหนดควรมีความคลาดเคลื่อนในการแบ่งกลุ่มต่ำ ๆ (McGinty, 2005)

จากการศึกษาสรุปได้ว่า ขั้นตอนของการตรวจสอบคุณภาพของคะแนนจุดตัดถือว่ามีความสำคัญมากเนื่องจากเป็นตรวจสอบเพื่อยืนยันคุณภาพของคะแนนจุดตัดโดยมีวิธีในการตรวจสอบหลายวิธี ดังที่สรุปไว้ในตารางต่อไปนี้

ตารางที่ 2.4 สรุปการตรวจสอบคุณภาพของคะแนนจุดตัด

การตรวจสอบคุณภาพ	รายละเอียด
1. ความเที่ยง	
1.1 ความเที่ยงภายในผู้เชี่ยวชาญ	ตรวจสอบจากผลการกำหนดคะแนนจุดตัดของผู้เชี่ยวชาญในแต่ละรอบว่าเท่าเดิมหรือใกล้เคียงกันทุกรอบหรือไม่
1.2 ความเที่ยงระหว่างผู้เชี่ยวชาญ	ตรวจสอบความความสอดคล้องของผลการกำหนดคะแนนจุดตัดระหว่างผู้เชี่ยวชาญเป็นการตรวจสอบว่าผู้เชี่ยวชาญมีความเห็นในทิศทางเดียวกันหรือไม่ในการจัดกลุ่มตัวอย่างลงในกลุ่มต่างๆ
1.3 ค่าสัมประสิทธิ์การสรุปอ้างอิง	เป็นการศึกษาขนาดองค์ประกอบความแปรปรวน และวิเคราะห์ค่าสัมประสิทธิ์การสรุปอ้างอิงเพื่อให้ทราบถึงขนาดของแหล่งความแปรปรวนที่ส่งผลต่อคะแนนจุดตัด และได้สารสนเทศสำหรับการตัดสินใจเกี่ยวกับจำนวนผู้เชี่ยวชาญ จำนวนกลุ่มย่อยของผู้เชี่ยวชาญ และจำนวนรอบที่เหมาะสมเพื่อให้สามารถกำหนดคะแนนจุดตัดที่มีความเที่ยงสูงถึงระดับที่ต้องการ
2. ความตรง	
2.1 ความตรงของปัจจัยนำเข้า	1) ความถูกต้อง 2) ความเป็นไปได้ 3) ความเหมาะสม 4) ความเข้าใจและความมั่นใจของผู้ตัดสิน 5) ความเป็นประโยชน์ของเอกสาร
2.2 ความตรงของกระบวนการตัดสินใจ	เป็นการตรวจสอบความตรงของกระบวนการตัดสินใจเพื่อให้ทราบเกี่ยวกับเหตุผลและหลักการที่ผู้ตัดสินยึดถือในการคะแนนจุดตัด และทราบเกี่ยวกับความคิด ความกังวลขณะตัดสินใจ รวมถึงปัญหาที่เกิดขึ้น
2.3 ความตรงของผลผลิต	เป็นการหาความสอดคล้องของการกำหนดกลุ่มระหว่างคะแนนจุดตัดที่กำหนดขึ้นในเรื่องเดียวกันแต่ใช้ต่างวิธีกัน
2.4 ความตรงของผลที่ตามมา	เป็นการตรวจสอบความถูกต้อง และความคลาดเคลื่อนของการแบ่งกลุ่ม ซึ่งประกอบด้วยความคลาดเคลื่อน 2 แบบ ได้แก่ ความคลาดเคลื่อนแบบลบและความคลาดเคลื่อนแบบบวก

1.8 งานวิจัยที่เกี่ยวข้อง

1.8.1 วิจัยเกี่ยวกับวิธีแองกอฟและแองกอฟแบบใช่/ไม่ใช่

การศึกษาวิจัยเกี่ยวกับวิธีแองกอฟส่วนใหญ่เป็นการศึกษาเพื่อพยายามปรับปรุงวิธีการนี้ โดยมีเป้าหมายของเพื่อปรับปรุงวิธีวิธีแองกอฟให้คะแนนจุดตัดที่กำหนดขึ้นมีความเที่ยงภายในผู้เชี่ยวชาญสูง และมีความเที่ยงระหว่างผู้เชี่ยวชาญมากขึ้น นอกจากนี้ก็ปรับปรุงเพื่อช่วยลดความคลาดเคลื่อนมาตรฐานให้น้อยลง วิธีแองกอฟถูกปรับปรุงในหลายลักษณะที่ต่างกันภายใต้ชื่อดังนี้ คือ Modified Angoff หรือ Extended Angoff นอกจากนี้ก็มีชื่อว่า Yes/No Angoff

การปรับปรุงวิธีแองกอฟวิธีแรกคือ การเพิ่มจำนวนรอบของการตัดสินให้มากขึ้น เนื่องจากการกำหนดคะแนนจุดตัดมีการตัดสินซ้ำหลายๆ ครั้ง เพื่อให้ได้คะแนนจุดตัดที่ผู้เชี่ยวชาญมีความมั่นใจที่สุด (Linn, 1989) ส่วนช่วงเวลาระหว่างรอบของการตัดสินก็ให้ผู้เชี่ยวชาญได้ร่วมกันอภิปรายเกี่ยวกับคะแนนจุดตัดที่แต่ละคนกำหนด เพื่อแลกเปลี่ยนความคิดเห็นซึ่งกันและกัน ซึ่งจะช่วยเพิ่มความเที่ยงระหว่างผู้เชี่ยวชาญ Busch และ Jaeger (1990) กล่าวว่า จำนวนรอบของการตัดสินจะกำหนดโดยพิจารณาจากความคงที่ของผลการกำหนดคะแนนจุดตัด นั่นคือหากผู้เชี่ยวชาญไม่เปลี่ยนแปลงผลการตัดสินหรือเปลี่ยนแปลงเพียงเล็กน้อยก็ให้ยุติการกำหนดคะแนนจุดตัด โดยนักวิจัยส่วนใหญ่กำหนดจำนวนรอบในการพิจารณา 2 รอบ (Giraud et al., 2005; Hurtz et al., 2008; Näsström, & Nyström, 2008; MacCann, 2008; Clauser, et al., 2009) และมีบางงานวิจัยกำหนดจำนวนรอบในการพิจารณา 3 รอบ (Busch & Jaeger, 1990; Fehrmann, et al., 1991; Hambleton & Plake, 1995 ; Hess, et al., 2007) ซึ่งข้อดีของวิธีการนี้ คือ เป็นการเปิดโอกาสให้ผู้เชี่ยวชาญได้ทบทวนผลการกำหนดคะแนนจุดตัด และระหว่างรอบของการตัดสินนั้นผู้เชี่ยวชาญจะได้อภิปรายและแลกเปลี่ยนความคิดเห็นกันเกี่ยวกับคะแนนจุดตัดที่กำหนดในรอบก่อนก็จะทำให้การกำหนดคะแนนจุดตัดในรอบต่อไปมีความสอดคล้องกันมากขึ้นระหว่างผู้เชี่ยวชาญแต่ละท่าน ซึ่งจะทำให้การกำหนดคะแนนจุดตัดมีความน่าเชื่อถือเนื่องจากเป็นคะแนนจุดตัดที่ผู้เชี่ยวชาญได้พิจารณาอย่างถี่ถ้วนแล้ว ส่วนข้อบกพร่อง คือ ใช้เวลาในการกำหนดคะแนนจุดตัดมาก

วิธีที่สองคือการนำเสนอข้อมูลสารสนเทศเพื่อใช้เป็นข้อมูลประกอบการกำหนดคะแนนจุดตัด Busch และ Jaeger (1990) กล่าวว่า การนำเสนอสารสนเทศโดยนำเสนอข้อมูลที่สำคัญเพื่อประกอบการตัดสินใจรอบสุดท้าย เช่น ร้อยละของนักเรียนที่ตอบข้อสอบแต่ละข้อได้ถูกต้อง ข้อมูล

เกี่ยวกับสัดส่วนของนักเรียนที่ถูกตัดสินให้อยู่ในระดับความสามารถต่าง ๆ และข้อมูลเกี่ยวกับคะแนนจุดตัดที่ผู้เชี่ยวชาญแต่และท่านกำหนดเป็นต้น จะช่วยเพิ่มความเที่ยงภายในผู้เชี่ยวชาญ และมีความเที่ยงระหว่างผู้เชี่ยวชาญได้ และงานวิจัยของ Clauser และคณะ (2009) ศึกษาผลกระทบของการอภิปรายกลุ่มและการเสนอสารสนเทศต่อการพิจารณากำหนดคะแนนจุดตัด ซึ่งผลจากการศึกษา พบว่า การอภิปรายกลุ่มและการเสนอสารสนเทศเพื่อประกอบการอภิปรายช่วยให้ผลการตัดสินของผู้เชี่ยวชาญแต่ละท่านใกล้เคียงกันมากขึ้น ส่วน Norcini และคณะ (1988) กล่าวว่า การที่ให้ผู้เชี่ยวชาญพิจารณาข้อมูลสารสนเทศเพียงร้อยละ 25 ของเวลาในการตัดสินทั้งหมดมีผลทำให้ผู้เชี่ยวชาญเปลี่ยนแปลงคะแนนจุดเพียงเล็กน้อย ดังนั้นควรให้เวลาในการพิจารณาข้อมูลสารสนเทศมากขึ้น ซึ่งการเสนอข้อมูลสารสนเทศประกอบการพิจารณากำหนดคะแนนจุดตัดนั้นมีข้อดีคือช่วยให้ผู้เชี่ยวชาญมีข้อมูลสนับสนุนคะแนนจุดตัดที่กำหนดและช่วยให้ผู้เชี่ยวชาญกำหนดคะแนนจุดตัดได้ง่ายขึ้น แต่หากให้ข้อมูลสารสนเทศมากเกินไปอาจส่งผลในด้านลบต่อการกำหนดคะแนนจุดตัดเนื่องจากเป็นการชี้นำผลการตัดสินของผู้เชี่ยวชาญ Brandon (2004) กล่าวว่า การปรับปรุงวิธีแองกอฟนั้น วิธีได้รับความนิยมมากที่สุดคือการเพิ่มจำนวนรอบในการพิจารณาตัดสินให้มากขึ้นและการนำเสนอสารสนเทศที่สำคัญต่อผู้เชี่ยวชาญในระหว่างรอบของการตัดสิน ซึ่งโดยส่วนใหญ่จะกำหนดรอบของการตัดสินเป็น 2 ส่วนข้อมูลที่มักนำมาเสนอต่อผู้เชี่ยวชาญคือ ข้อมูลเกี่ยวกับสัดส่วนของนักเรียนที่ถูกตัดสินให้อยู่ในระดับความสามารถต่างๆ และข้อมูลเกี่ยวกับคะแนนจุดตัดที่ผู้เชี่ยวชาญแต่และท่านกำหนด ส่วนการนำทฤษฎีการตอบข้อสอบมาประยุกต์ใช้นั้นก็กำลังได้รับความสนใจจากนักวิจัยหลายท่านในการนำมาประยุกต์ใช้เพื่อให้คะแนนจุดตัดที่ได้มีความถูกต้องและน่าเชื่อถือยิ่งขึ้น

วิธีที่สามคือการเปลี่ยนแปลงวิธีการพิจารณาให้ง่ายต่อการปฏิบัติ ผู้เสนอวิธีนี้คือ Impara และ Plake (1997) โดยผู้เชี่ยวชาญจะต้องพิจารณาเพื่อตัดสินว่า ผู้มีความสามารถคาบเส้นจะสามารถตอบข้อสอบได้ถูกต้องใช่หรือไม่ ซึ่งผู้เชี่ยวชาญจะตอบเพียงว่าใช่ หรือไม่ใช่ ซึ่งเป็นที่มาของชื่อวิธีแองกอฟแบบใช่/ไม่ใช่ (Yes/No Angoff Method) เมื่อผู้เชี่ยวชาญพิจารณาครบทุกข้อแล้วจากนั้นก็ทำเหมือนวิธีแองกอฟเดิมคือเฉลี่ยผลการพิจารณาจากผู้เชี่ยวชาญเป็นระดับการผ่านขั้นต่ำรายข้อ ระดับการผ่านขั้นต่ำรายข้อจะถูกรวมจากทุกข้อในแบบสอบเพื่อกำหนดเป็นจุดตัด ถือว่าเป็นคะแนนการผ่านขั้นต่ำ ซึ่ง Impara และ Plake (1997)ได้ทำการทดลองเพื่อเปรียบคะแนนจุดตัดที่

กำหนดโดยใช้วิธีแองกอฟและวิธีแองกอฟแบบใช่/ไม่ใช่ พบว่า ได้คะแนนจุดตัดเท่ากันแต่ผลการทดลองก็ยังไม่เป็นที่น่าเชื่อถือนัก เนื่องจากใช้กลุ่มผู้เชี่ยวชาญเพียงกลุ่มเดียวและให้พิจารณาทั้งสองวิธีซึ่งก็ยังเป็นที่น่าสังเกตว่าผลการกำหนดคะแนนจุดตัดโดยวิธีหนึ่งอาจได้รับอิทธิพลจากวิธีที่ใช้กำหนดก่อนหน้านี้ได้ ต่อมา Chinn และ Hertz (2002) ได้ศึกษาเพื่อเปรียบเทียบผลการกำหนดคะแนนจุดตัดโดยใช้วิธีแองกอฟและวิธีแองกอฟแบบใช่/ไม่ใช่ ซึ่งทำการทดลอง 2 กลุ่ม ผลการศึกษาพบว่า ผู้เชี่ยวชาญมีความคิดเห็นว่าวิธีแองกอฟแบบใช่/ไม่ใช่ ง่ายต่อการปฏิบัติมากกว่าวิธีแองกอฟ แต่วิธีแองกอฟแบบใช่/ไม่ใช่นั้น คะแนนจุดตัดที่กำหนดจะได้รับอิทธิพลจากสารสนเทศที่ให้ได้มากกว่าคะแนนจุดตัดที่กำหนดโดยวิธีแองกอฟ นอกจากนี้ Ferdous และ Plake (2008) ได้ศึกษาวิธีการนำทฤษฎีการตอบสนองข้อสอบแบบ 3 พารามิเตอร์ มาประยุกต์ใช้ร่วมกับการกำหนดคะแนนจุดตัดด้วยวิธีแองกอฟแบบใช่/ไม่ใช่ โดยนำทฤษฎีการตอบสนองข้อสอบมาประยุกต์ใช้เพื่อกำหนดคะแนนการผ่านขั้นต่ำ ซึ่งข้อดีของการพิจารณาแบบวิธีแองกอฟแบบใช่/ไม่ใช่ คือ ทำให้การพิจารณาเพื่อกำหนดคะแนนจุดตัดง่ายขึ้น ส่วนข้อด้อยคือเป็นการพิจารณาตัดสินที่หยาบเกินไปนั่นคือการตัดสินเพียงใช่หรือไม่ใช่นั้นส่งผลให้คะแนนจุดตัดที่ได้ต่างกันมาก

ตารางที่ 2.5 แสดงการคำนวณคะแนนจุดตัดโดยวิธีแองกอฟแบบใช่/ไม่ใช่

ข้อสอบ	ผู้ตัดสินคนที่ 1	ผู้ตัดสินคนที่ 2
1	1	1
2	0	1
3	1	0
4	1	1
5	0	0
รวม	3	3

ที่มา : Zieky และคณะ (2008)

คะแนนจุดตัด = (3+3)/2 = 3 นั่นคือผู้ที่จะผ่านมาตรฐานต้องทำข้อสอบให้ถูก 3 ข้อ จาก 5 ข้อ

Hambleton และ Plake (1995) ได้เสนอวิธีการปรับปรุงวิธีแองกอฟให้สามารถกำหนดคะแนนจุดตัดสำหรับข้อสอบที่วัดหลายมิติ โดยการให้ค่าน้ำหนัก (Weight) แก่ข้อสอบแต่ละมิติซึ่งน้ำหนักที่ให้จะแสดงถึงความสำคัญของมิตินั้น ๆ จากนั้นก็ให้ผู้เชี่ยวชาญพิจารณาเพื่อกำหนดคะแนนจุดตัดเป็นรายมิติแล้วนำคะแนนจุดตัดที่ได้มาคูณกับค่าน้ำหนักก่อนจึงรวมกันเป็นคะแนนจุดตัดขั้นสุดท้ายซึ่งข้อดีคือใช้ในการตัดสินสำหรับข้อสอบที่วัดหลายมิติได้ ส่วนข้อด้อยคือใช้เวลาค่อนข้างมาก

นอกจากวิธีการปรับปรุงที่กล่าวมาแล้วได้มีนักวิชาการนำทฤษฎีการตอบสนองข้อสอบ (Item Response Theory) มาประยุกต์ใช้ เนื่องจากว่าวิธีแองกอฟไม่ให้ความสนใจต่อลักษณะของข้อสอบเช่น ค่าพารามิเตอร์ความยาก (b) อำนาจจำแนก (a) ที่คำนวณโดยใช้ทฤษฎีการตอบสนองข้อสอบซึ่งเป็นทฤษฎีการวัดที่อธิบายความสัมพันธ์ระหว่างความสามารถที่มีอยู่ในตัวบุคคล (Latent trait) กับผลการตอบข้อสอบ โดยใช้โค้งลักษณะข้อสอบ (Item Characteristic Curve: ICC) ซึ่งมีการกำหนดลักษณะของข้อสอบด้วยพารามิเตอร์ความยาก อำนาจจำแนกและโอกาสการเดาข้อสอบถูก Van Der Linden (1982) กล่าวว่า การใช้ข้อมูลจากการวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบมีประโยชน์ต่อการพิจารณากำหนดคะแนนจุดตัด นอกจากนี้ MacCann และ Stanley (2006) เสนอให้นำทฤษฎีการตอบสนองแบบ 1 พารามิเตอร์ (Rash Model) มาใช้วิเคราะห์ข้อสอบเพื่อให้ได้ข้อมูลเกี่ยวกับความยากของข้อสอบ และโอกาสการตอบข้อสอบได้ถูก โดยจะนำมาใช้เป็นข้อมูลสารสนเทศเสนอต่อผู้เชี่ยวชาญเพื่อใช้ประกอบการพิจารณากำหนดคะแนนจุดตัด

ต่อมา Ferdous และ Plake (2008) ได้ศึกษาวิธีการนำทฤษฎีการตอบสนองข้อสอบแบบ 3 พารามิเตอร์ มาประยุกต์ใช้ร่วมกับการกำหนดคะแนนจุดตัดด้วยวิธีแองกอฟ และ วิธีแองกอฟ แบบใช่/ไม่ใช่ โดยนำทฤษฎีการตอบสนองข้อสอบมาประยุกต์ใช้เพื่อกำหนดคะแนนการผ่านขั้นต่ำซึ่งศึกษา 2 วิธี วิธีแรกคือ Judge Theta Approach ใช้กับการกำหนดคะแนนจุดตัดโดยวิธีแองกอฟซึ่งเป็นวิธีการเปลี่ยนค่าความน่าจะเป็นในการตอบถูกของผู้มีความสามารถคาบเส้นซึ่งเป็นผลจากการพิจารณาโดยผู้เชี่ยวชาญไปเป็นค่าความสามารถ (Ability: θ) โดยใช้ทฤษฎีการตอบสนองข้อสอบแบบ 3 พารามิเตอร์ ซึ่งแสดงความสัมพันธ์ระหว่างความสามารถของบุคคลกับโอกาสในการทำข้อสอบได้ถูก และวิธีที่สอง คือวิธี Item Response Vector Approach เป็นวิธีที่ใช้กับการกำหนดมาตรฐานวิธีแองกอฟแบบใช่/ไม่ใช่ โดยผู้เชี่ยวชาญจะตัดสินว่าผู้มีความสามารถคาบเส้นจะมี

โอกาสอย่างน้อย 0.67 ที่จะตอบข้อสอบได้ถูกใช่หรือไม่ ผลจากการศึกษาพบว่าคะแนนจุดตัดที่กำหนดโดยวิธีแองกอฟและวิธีแองกอฟแบบใช่/ไม่ใช่ มีค่าใกล้เคียงกับจุดตัดที่มีการนำทฤษฎีการตอบสนองข้อสอบมาประยุกต์ใช้เพื่อกำหนดคะแนนการผ่านขั้นต่ำ แสดงว่าวิธีแองกอฟและวิธีแองกอฟแบบใช่/ไม่ใช่สามารถนำมาใช้ได้กับข้อสอบที่มีการวิเคราะห์ข้อสอบโดยใช้ทฤษฎีการตอบสนองข้อสอบซึ่งถือว่าเป็นทฤษฎีในการวิเคราะห์ข้อสอบที่ได้รับความนิยมอยู่ในปัจจุบัน ซึ่งข้อดีของการนำทฤษฎีการตอบสนองข้อสอบมาประยุกต์ใช้คือทำให้ผลการตัดสินน่าเชื่อถือมากขึ้นและช่วยให้ตัดสินใจได้ง่ายขึ้น ส่วนข้อด้อยคือต้องมีความรู้ในเรื่องสถิติ ใช้เวลาในการวิเคราะห์และจัดเตรียมข้อมูลมาก

จากการศึกษางานวิจัยที่เกี่ยวกับการพัฒนาปรับปรุงวิธีแองกอฟ สรุปได้ว่า นักวิจัยพยายามที่จะทำให้การกำหนดคะแนนจุดตัดสามารถนำไปใช้สำหรับข้อสอบหลายๆ แบบ ปฏิบัติได้ง่ายและมีความคงที่ของผลการกำหนดคะแนนจุดตัด โดยแต่ละวิธีก็มีจุดเด่นและข้อบกพร่องแตกต่างกันไป ดังสรุปในตารางต่อไปนี้

ตารางที่ 2.6 การปรับปรุงวิธีแองกอฟ ข้อดีและข้อเสีย

	การปรับปรุงวิธีแองกอฟ	ข้อดี	ข้อเสีย
1.	การเพิ่มจำนวนรอบของการพิจารณาตัดสิน	- เป็นการเปิดโอกาสให้ผู้ตัดสินได้ทบทวนผลการกำหนดคะแนนจุดตัด - ผู้ตัดสินได้อภิปรายและแลกเปลี่ยนความคิดเห็นกันเกี่ยวกับคะแนนจุดตัด ทำให้การกำหนดคะแนนมีความสอดคล้องกันมากขึ้น	ใช้เวลาในการกำหนดคะแนนจุดตัดมาก
2.	การนำเสนอข้อมูลสารสนเทศที่สำคัญในการตัดสิน	- ช่วยให้ผู้ตัดสินมีข้อมูลสนับสนุนคะแนนจุดตัดที่กำหนด - ช่วยให้ผู้ตัดสินกำหนดคะแนนจุดตัดได้ง่ายขึ้น	หากให้ข้อมูลสารสนเทศมากเกินไปอาจส่งผลในด้านลบต่อการกำหนดคะแนนจุดตัดเนื่องจากเป็นการชี้นำผลการตัดสินของผู้ตัดสิน
3.	การเปลี่ยนวิธีการพิจารณาให้ง่ายขึ้น	- ง่ายต่อการพิจารณาตัดสิน - ทำให้กำหนดคะแนนจุดตัดได้ในกรณีที่ข้อสอบคะแนนไม่เท่ากัน	เป็นการพิจารณาตัดสินที่หยาบเกินไป
4.	การให้ค่าน้ำหนักแก่ข้อสอบที่วัดแต่ละมิติ	ใช้ในการตัดสินสำหรับข้อสอบที่วัดหลายมิติได้	ใช้เวลาค่อนข้างมาก
5.	การนำทฤษฎีการตอบสนองข้อสอบมาประยุกต์ใช้	- ทำให้ผลการตัดสินน่าเชื่อถือมากขึ้น - ช่วยให้ตัดสินใจได้ง่ายขึ้น	- ต้องมีความรู้ในเรื่องสถิติที่ต้องใช้ - ต้องใช้เวลาในการวิเคราะห์และจัดเตรียมข้อมูลมาก

การศึกษาวิจัยเกี่ยวกับการกำหนดคะแนนจุดตัดโดยวิธีแองกอฟ มี 4 งานวิจัยที่ศึกษาวิจัยเพื่อเปรียบเทียบผลการกำหนดจุดตัดโดยวิธีแองกอฟกับวิธีอื่นๆ ประกอบด้วยงานวิจัยของ Buckendahl และคณะ (2002), Nijlen และ Janssen (2008), Näsström และ Nyström (2008) และงานวิจัย Hess และคณะ (2007)

Buckendahl และคณะ (2002) ได้ทำการศึกษาเพื่อเปรียบเทียบผลการกำหนดจุดตัดโดยวิธีบุ๊คมาร์คและแองกอฟ พบว่า คะแนนจุดตัดที่กำหนดโดยทั้งสองวิธีมีค่าเท่ากันทั้งนี้อาจเนื่องจากการกำหนดคะแนนจุดตัดด้วยวิธีบุ๊คมาร์คและวิธีแองกอฟ ใช้สารสนเทศในการพิจารณาเหมือนกัน คือพิจารณาจากข้อสอบ ดังนั้นจึงทำให้ผลการกำหนดคะแนนจุดตัดมีค่าเท่ากัน

Niljen และ Janssen (2008) เปรียบเทียบผลการกำหนดคะแนนจุดตัดโดยวิธีแองกอฟและวิธีกลุ่มตรงข้าม กษาพบวซึ่งจากการศึาคะแนนจุดตัดที่กำหนดโดยวิธีกลุ่มตรงข้ามต่ำกว่าคะแนนจุดตัดที่กำหนดโดยวิธีแองกอฟ ทั้งนี้ผู้วิจัยกล่าวว่าการที่คะแนนจุดตัดที่กำหนดโดยวิธีกลุ่มตรงข้ามต่ำกว่าคะแนนจุดตัดที่กำหนดโดยวิธีแองกอฟนั้นเนื่องจากว่าครูที่ทำหน้าที่เป็นผู้เชี่ยวชาญเป็นครูผู้สอนจึงอาจเกิดความลำเอียงในการกำหนดคะแนนจุดตัดและกำหนดคะแนนจุดตัดต่ำกว่าคะแนนจุดตัดที่กำหนดโดยวิธีแองกอฟ

Näsström และ Nyström (2008) ได้ทำการศึกษาเกี่ยวกับความถูกต้องของคะแนนจุดตัดที่ได้จากการกำหนดโดยใช้วิธีแองกอฟและวิธีกลุ่มคาบเส้น ซึ่งมีเป้าหมายของการศึกษาเพื่อตรวจสอบความถูกต้องของคะแนนจุดตัด และเปรียบเทียบผลการกำหนดคะแนนจุดตัดจากทั้งสองวิธี โดยทำการศึกษากับการทดสอบที่มีความสำคัญของประเทศสวีเดน วิชาคณิตศาสตร์ ข้อสอบเป็นแบบเขียนตอบจำนวน 22 ข้อ เพื่อกำหนดคะแนนจุดตัดสำหรับระดับผ่านและระดับยอดเยี่ยม ทำการตรวจสอบความถูกต้องของคะแนนจุดตัด โดยตรวจสอบในประเด็นต่อไปนี้ 1) ความถูกต้องของกระบวนการ (Procedural Validation) 2) ตรวจสอบความถูกต้องภายใน (Internal Validation) 3) ตรวจสอบความถูกต้องภายนอก (External Validation) พบว่า การกำหนดคะแนนจุดตัดครั้งนี้ได้ดำเนินการตามขั้นตอนที่สำคัญทั้งหมดยกเว้นขั้นตอนของการฝึกอบรมผู้เชี่ยวชาญซึ่งถือว่าเป็นขั้นตอนที่สำคัญมากของกระบวนการกำหนดคะแนนจุดตัด เนื่องจากเป็นขั้นตอนที่ใช้เพื่ออธิบายเกี่ยวกับวิธีการกำหนดคะแนนจุดตัดและฝึกทักษะที่จำเป็นสำหรับการพิจารณาเพื่อกำหนดคะแนนจุดตัด (Hambleton & Pitoniak, 2006 cited in Näsström & Nyström, 2008) แต่เมื่อพิจารณาจากภูมิ

หลังของผู้เชี่ยวชาญพบว่า ผู้เชี่ยวชาญที่ได้รับการคัดเลือกล้วนเป็นผู้ที่มีประสบการณ์ในการกำหนดคะแนนจุดตัดและมีประสบการณ์เกี่ยวกับการพิจารณาผลการสอบที่มีความสำคัญ ดังนั้นการที่ไม่มีการฝึกอบรมผู้เชี่ยวชาญน่าจะไม่ส่งผลกระทบต่อการกำหนดคะแนนจุดตัด

นอกจากนี้ผลการพิจารณากำหนดคะแนนจุดตัดครั้งแรกกับครั้งที่สองมีการเปลี่ยนแปลงเพียงเล็กน้อยซึ่งถือได้ว่าผลการกำหนดคะแนนจุดตัดมีความคงเส้นคงวา เมื่อพิจารณาผู้เชี่ยวชาญของวิธีกลุ่มคาบเส้น พบว่า ผู้เชี่ยวชาญเป็นครูที่มีประสบการณ์ในการสอนอย่างน้อย 2 ปี ดังนั้นย่อมมีประสบการณ์ในการกำหนดเกรดให้กับนักเรียน จึงสันนิษฐานได้ว่าครูน่าจะมีทักษะในการกำหนดกลุ่มนักเรียนตามความสามารถได้เป็นอย่างดี ดังนั้นการที่ไม่มีการอบรมผู้เชี่ยวชาญก่อนการกำหนดคะแนนจุดตัดน่าจะไม่ส่งผลกระทบต่อการกำหนดคะแนนจุดตัด จึงสรุปได้ว่าทั้งสองวิธีดำเนินการกำหนดคะแนนจุดตัดได้อย่างถูกต้อง ตรวจสอบความถูกต้องภายในจะพิจารณาค่าความเที่ยงระหว่างผู้เชี่ยวชาญและความเที่ยงระหว่างวิธี ผลการพิจารณาพบว่า คะแนนจุดตัดที่กำหนดโดยวิธีแองกอฟและวิธีกลุ่มคาบเส้นมีความคลาดเคลื่อนมาตรฐานต่ำ แสดงว่าคะแนนจุดตัดที่กำหนดมีความถูกต้อง นอกจากนี้การตรวจสอบความถูกต้องภายนอก (External Validation) จะพิจารณาโดยการเปรียบเทียบผลการกำหนดคะแนนจุดตัดที่กำหนดด้วยวิธีต่างกันซึ่งการศึกษาครั้งนี้ทำการเปรียบเทียบคะแนนจุดตัดที่ได้จากการกำหนดโดยวิธีแองกอฟและวิธีกลุ่มคาบเส้น ซึ่งพบว่าคะแนนจุดตัดมีค่าใกล้เคียงกัน นอกจากนี้ก็ยังพิจารณาจากการหาความสัมพันธ์ระหว่างคะแนนที่ได้จากการสอบกับเกรดที่ได้จากการเรียนในห้องเรียนโดยถือว่าเกรดที่ได้จากห้องเรียนเป็นการประเมินจากผลการสอบ ที่มีจุดมุ่งหมายและเนื้อหาที่เหมือนกับการสอบในครั้งนี้จึงสามารถนำมาเป็นเกณฑ์ภายนอกได้ผลการศึกษาพบว่า ผลการกำหนดระดับความสามารถโดยใช้คะแนนจุดตัดที่กำหนดขึ้นกับเกรดมีความสัมพันธ์ทางบวกโดยมีค่าสหสัมพันธ์ระหว่าง 0.81 – 0.83

Hess และคณะ (2007) เพื่อเปรียบเทียบผลการกำหนดคะแนนจุดตัดโดยวิธีแองกอฟกับการวิเคราะห์กลุ่ม ผลจากการศึกษาพบว่า การจัดกลุ่มผู้สอบโดยใช้คะแนนจุดตัดที่ได้จากการกำหนดโดยวิธีแองกอฟให้ผลเหมือนกับการวิเคราะห์กลุ่ม

เนื่องจากการกำหนดคะแนนจุดตัดโดยวิธีแองกอฟนั้น ผู้เชี่ยวชาญต้องพิจารณาข้อสอบทุกข้อเพื่อกำหนดคะแนนจุดตัดซึ่งถือว่าเป็นงานที่หนัก ต้องใช้เวลาและค่าใช้จ่ายในการดำเนินการมาก (Green, 1995) ดังนั้นเพื่อแก้ไขข้อบกพร่องดังกล่าวจึงมีนักวิจัยได้ทำการศึกษาเพื่อหาวิธีในการ

ลดจำนวนข้อสอบที่จะนำมาใช้ในการกำหนดคะแนนจุดตัด แต่ยังคงให้คะแนนจุดตัดเท่ากับการพิจารณาจากแบบสอบทั้งฉบับ โดยมีงานวิจัยที่ทำการศึกษาเรื่องนี้จำนวน 2 เรื่องคืองานวิจัยของ Ferdous และ Plake ในปี ค.ศ. 2005 และต่อมาได้ศึกษาเกี่ยวกับเรื่องนี้อีก ในปี ค.ศ. 2007

Ferdous และ Plake (2005) ได้ศึกษาเพื่อหาวิธีในการลดจำนวนข้อสอบลงโดยศึกษาวิธีการคัดเลือกข้อสอบจากแบบสอบทั้งฉบับมาเป็นข้อสอบฉบับย่อย ซึ่งในการศึกษาครั้งนี้ใช้วิธีการสุ่มข้อสอบโดยสุ่มแบบแบ่งชั้น (Stratified Random Sampling) โดยกำหนดชั้นของข้อสอบตามกลุ่มเนื้อหา(Content Categories) จากนั้นก็แบ่งแต่ละกลุ่มเนื้อหาตามระดับความยาก (Difficulty Categories) และทำการสุ่มข้อสอบตามสัดส่วนเพื่อใช้เป็นข้อสอบย่อย โดยจะทำการสุ่มมา 10%, 20%, 30%, 40%, และ 50% ของแบบสอบฉบับเต็ม จากนั้นก็นำแบบสอบฉบับเต็มและแบบสอบย่อยทั้ง 5 ฉบับ มาให้ผู้เชี่ยวชาญพิจารณาเพื่อกำหนดคะแนนจุดตัด แล้วนำคะแนนจุดตัดที่ได้ไปใช้ตัดสินผู้สอบเพื่อแบ่งเป็นกลุ่มตามระดับความสามารถ 3 ระดับ ผลการจัดกลุ่มพบว่า คะแนนจุดตัดที่พิจารณาจากแบบสอบย่อยที่สุ่มมา 50% ของฉบับเต็ม สามารถจัดกลุ่มผู้สอบในแต่ละระดับความสามารถได้เหมือนกับการจัดกลุ่มโดยใช้คะแนนจุดตัดที่พิจารณาจากแบบสอบฉบับเต็ม นอกจากนี้การกำหนดคะแนนจุดตัดโดยการพิจารณาจากแบบสอบย่อยใช้เวลาในการพิจารณาน้อยกว่าการพิจารณาแบบสอบฉบับเต็มมาก

Ferdous และ Plake (2007) ได้ทำการศึกษาวิธีการคัดเลือกข้อสอบจากแบบสอบทั้งฉบับมาเป็นข้อสอบฉบับย่อย โดยในการศึกษาครั้งนี้ใช้วิธีการสุ่มข้อสอบโดยสุ่มแบบแบ่งชั้น และกำหนดชั้นข้อสอบตามกลุ่มเนื้อหา จากนั้นก็แบ่งแต่ละกลุ่มเนื้อหาตามระดับความยาก เหมือนงานวิจัยเดิมและเพิ่มการแบ่งแต่ละระดับความยากตามค่าอำนาจจำแนก (Discrimination) จากนั้นก็ทำการสุ่มข้อสอบตามสัดส่วนเพื่อใช้เป็นข้อสอบย่อย โดยจะทำการสุ่มมา 10%, 20%, 30%, 40% และ 50% ของแบบสอบฉบับเต็ม แล้วนำแบบสอบฉบับเต็มและแบบสอบย่อยทั้ง 5 ฉบับ มาให้ผู้เชี่ยวชาญพิจารณาเพื่อกำหนดคะแนนจุดตัด จากนั้นก็นำคะแนนจุดตัดที่ได้ไปใช้ตัดสินผู้สอบเพื่อแบ่งเป็นกลุ่มตามระดับความสามารถ 3 ระดับ ผลการจัดกลุ่ม พบว่า คะแนนจุดตัดที่พิจารณาจากแบบสอบย่อยที่สุ่มมา 40% และ 50% ของฉบับเต็มสามารถจัดกลุ่มผู้สอบในแต่ละระดับความสามารถได้เหมือนคะแนนจุดตัดที่พิจารณาจากฉบับเต็ม ซึ่งถือว่าช่วยลดจำนวนข้อสอบได้น้อยลงกว่างานวิจัยที่แล้ว

ตารางที่ 2.7 สรุปการศึกษาวิจัยเกี่ยวกับการกำหนดคะแนนจุดตัดโดยวิธีแองกอฟ

นักวิชาการ	ชื่อ	กลุ่มที่กำหนดคะแนนจุดตัด	ระดับความสามารถ	ข้อสอบ	จำนวนรอบการตัดสิน	การปรับปรุงวิธีแองกอฟ			
						ข้อมูลสารสนเทศ	อภิปรายกลุ่ม	วิธีการพิจารณา	นำ IRT มาประยุกต์ใช้
Ferdous และ Plake (2005)	modified method	เกรด 3-8 และ 11	Basic/ proficient/ advanced	คำถามปลายเปิดวิชาการอ่านและคณิตศาสตร์	2 รอบ	ความยากของข้อสอบ	✓	ประมาณค่าโอกาสการตอบข้อสอบถูก	-
Giraud และคณะ (2005)	Angoff	เกรด 12	รอบรู้/ไม่รอบรู้	ข้อสอบแบบเลือกตอบ	2 รอบ	-	✓	ประมาณค่าโอกาสการตอบข้อสอบถูก	-
Hess และคณะ (2007)	modified Angoff	การสอบที่มีความสำคัญสูง	ผ่าน/ตก	ข้อสอบแบบเลือกตอบ	3 รอบ	ความยากของข้อสอบ	✓	ประมาณค่าโอกาสการตอบข้อสอบถูก	-
Hurtz และคณะ (2008)	Angoff	การสอบที่มีความสำคัญสูง	ผ่าน/ตก	ข้อสอบแบบเลือกตอบ	2 รอบ	-	-	ประมาณค่าโอกาสการตอบข้อสอบถูก	three-parameter logistic
Yudkowsky และคณะ (2008)	Yes/No Angoff	นักศึกษาแพทย์ปี 4	2 ระดับ	สำรวจรายการ	1 รอบ	-	-	ตอบใช่หรือไม่ใช่	-
Nijlen และJanssen (2008)	Angoff	เกรด 6	2 ระดับ	คำถามปลายเปิดและข้อสอบแบบเลือกตอบ	1 รอบ	-	-	ประมาณค่าโอกาสการตอบข้อสอบถูก	-
Näsström และ Nyström (2008)	extended Angoff-	การสอบที่มีความสำคัญสูง	4 ระดับ	ข้อสอบเติมคำวิชาคณิตศาสตร์	2 รอบ	ความยากของข้อสอบ	✓	ประมาณค่าโอกาสการตอบข้อสอบถูก	-

ตารางที่ 2.7 สรุปการศึกษาวิจัยเกี่ยวกับการกำหนดคะแนนจุดตัดโดยวิธีแองกอฟ

นักวิชาการ	ชื่อ	กลุ่มที่กำหนดคะแนนจุดตัด	ระดับความสามารถ	ข้อสอบ	จำนวนรอบการตัดสิน	การปรับปรุงวิธีแองกอฟ			
						ข้อมูลสารสนเทศ	อภิปรายกลุ่ม	วิธีการพิจารณา	นำ IRT มาประยุกต์ใช้
Fowell และคณะ (2008)	modified Angoff	นักศึกษาแพทย์	2 ระดับ	เขียนตอบสั้นๆ	2 รอบ	สัดส่วนผู้ที่ตกกับผู้ที่ผ่าน	✓	ประมาณค่าโอกาสการตอบข้อสอบถูก	-
Ferdous และ Plake (2008)	Yes/No Angoff	การรับรองนักวิเคราะห์ทางการเงิน	2 ระดับ	ข้อสอบแบบเลือกตอบ	รอบเดียว	-	-	ตอบใช่หรือไม่ใช่	three-parameter logistic
MacCann (2008)	modified Angoff	เกรด 12	2 ระดับ	ข้อสอบแบบเลือกตอบ	2 รอบ	-	✓	-	-
Clauser และคณะ (2009)	Extended Angoff	การสอบที่มีความสำคัญสูง	2 ระดับ	ข้อสอบแบบเลือกตอบ	2 รอบ	- ความน่าจะเป็นที่จะตอบข้อสอบถูก - ผลการใช้คะแนนจุดตัด - ความยากของข้อสอบ	✓	ประมาณค่าโอกาสการตอบข้อสอบถูก	Rash model

1.8.2 งานวิจัยเกี่ยวกับวิธีบุ๊คมาร์ค

จากการศึกษางานวิจัยที่เกี่ยวกับการกำหนดคะแนนจุดตัดโดยใช้วิธีบุ๊คมาร์ค พบว่า มีงานวิจัยที่ทำการศึกษาเปรียบเทียบผลของการกำหนดคะแนนจุดตัดโดยวิธีบุ๊คมาร์คกับอื่นๆ 3 เรื่อง ได้แก่ งานวิจัยของ Green และคณะ (2003) งานวิจัยของ Yin และ Schulz (2005 cited in Karantonis & Sireci 2006) และ งานวิจัยของ Buckendahl และคณะ (2002)

งานวิจัยของ Green และคณะ (2003) เป็นงานวิจัยเดียวที่ใช้วิธีบุ๊คมาร์คแบบดั้งเดิมโดยไม่ได้มีการปรับปรุงเปลี่ยนแปลงใดๆ เลย โดยศึกษาเพื่อเปรียบเทียบผลของการกำหนดคะแนนจุดตัดจาก 3 วิธีคือ วิธีบุ๊คมาร์ค วิธีกลุ่มตรงข้าม และวิธีจีเกอร์ มิลล์ ทั้งนี้เนื่องจากว่าผู้มีอำนาจในการตัดสินใจ เชื่อว่าการกำหนดคะแนนจุดตัดโดยใช้เพียงวิธีการเดียวไม่สามารถให้ข้อมูลที่เพียงพอที่จะทำให้เข้าใจเกี่ยวกับระดับสามารถของนักเรียนที่แท้จริงได้ ดังนั้นการศึกษาเพื่อกำหนดคะแนนจุดตัดครั้งนี้จึงใช้ถึง 3 วิธีดังที่กล่าวมา การวิจัยนี้ศึกษาในรัฐเคนตักกี้ เพื่อกำหนดระดับความสามารถของนักเรียน 4 ระดับ คือ เริ่มต้น (Novice) กำลังฝึก (Apprentice) เชี่ยวชาญ (Proficient) และ ยอดเยี่ยม (Distinguish) โดยศึกษาในหลากหลายพื้นที่และศึกษากับนักเรียน 18 ห้องซึ่งเป็นระดับชั้นที่ต่างกันและผู้เชี่ยวชาญเป็นครู การกำหนดคะแนนจุดตัดโดยวิธีบุ๊คมาร์คใช้ผู้เชี่ยวชาญจำนวน 278 คน วิธีกลุ่มตรงข้ามใช้ผู้เชี่ยวชาญ 299 คน และวิธีจีเกอร์ มิลล์ใช้ผู้เชี่ยวชาญ 907 คน โดยจะมีผู้เชี่ยวชาญกลาง จำนวน 133 คน ที่ทำหน้าที่ในการพิจารณาคะแนนจุดตัดครั้งสุดท้ายโดยพิจารณาจากผลการกำหนดคะแนนจุดตัดจากทั้ง 3 วิธี ผลจากการศึกษาพบว่า คะแนนจุดตัดที่กำหนดโดยใช้วิธีจีเกอร์มิลล์ มีค่าสูงที่สุดทุกค่า ส่วนคะแนนจุดตัดที่กำหนดโดยวิธีบุ๊คมาร์คมีค่าต่ำที่สุด ที่จุดระหว่างระดับเริ่มต้นกับกำลังฝึก และระหว่างระดับกำลังฝึกกับเชี่ยวชาญ คะแนนจุดตัดที่กำหนดโดยวิธีกลุ่มตรงข้ามมีค่าต่ำที่สุด ที่จุดระหว่างระดับเชี่ยวชาญกับยอดเยี่ยม เมื่อพิจารณาที่คะแนนจุดตัดขั้นสุดท้ายที่พิจารณาโดยผู้เชี่ยวชาญกลาง พบว่า คะแนนจุดตัดมีค่าใกล้เคียงกับคะแนนจุดตัดที่กำหนดโดยวิธีบุ๊คมาร์ค ซึ่งจากการศึกษาครั้งนี้แสดงให้เห็นว่าการกำหนดคะแนนจุดตัดโดยใช้วิธีที่ต่างกันให้คะแนนจุดตัดต่างกัน ทั้งนี้ผู้วิจัยกล่าวว่า การที่คะแนนจุดตัดที่ต่างกันนั้นอาจเนื่องมากว่าการกำหนดคะแนนจุดตัดแต่ละวิธีพิจารณากำหนดคะแนนจุดตัดโดยอาศัยสารสนเทศ (Information) ที่ต่างกัน โดยวิธีบุ๊คมาร์คจะพิจารณาโดยดูจากเนื้อหาและระดับความยากง่ายของข้อสอบเป็นหลัก ส่วนวิธีกลุ่มตรงข้ามจะตัดสินโดยอาศัยประสบการณ์ของครูที่

รับรู้ถึงความสามารถของนักเรียนแต่ละคน และวิธีจีเกอร์มิลล์เป็นการพิจารณาแบบองค์รวม ดังนั้นจึงทำให้ผลการกำหนดคะแนนจุดตัดต่างกัน

Yin และ Schulz (2005 cited in Yin & Sconing, 2008) ได้ทำการศึกษาเพื่อเปรียบผลของการกำหนดคะแนนจุดตัดระหว่างวิธีบุ๊คมาร์คและวิธีแองกอฟ ซึ่งการศึกษาครั้งนี้ได้ปรับปรุงวิธีบุ๊คมาร์คและใช้ชื่อใหม่ว่าแมปมาร์ค ซึ่งวิธีแมปมาร์คจะมีการพิจารณา 3 รอบ และการพิจารณาในรอบแรกจะเหมือนกับวิธีบุ๊คมาร์ค คือคณะผู้เชี่ยวชาญจะถูกแบ่งเป็นกลุ่มย่อย โดยคณะผู้เชี่ยวชาญแต่ละกลุ่มจะต้องร่วมกันพิจารณาคู่มือเรียงข้อสอบ และอภิปรายในประเด็นต่อไปนี้ 1) ผู้สอบต้องมีความรู้ความสามารถในเรื่องใดจึงจะสามารถตอบข้อสอบแต่ละข้อได้ถูกต้อง และ 2) เหตุใดข้อสอบข้อนี้จึงยากกว่าข้อก่อนๆ เมื่ออภิปรายครบทุกข้อแล้ว ผู้เชี่ยวชาญจะต้องกำหนดจุดตัดโดยที่ผู้เชี่ยวชาญแต่ละท่านจะมีอิสระในการกำหนดจุดตัดโดยนำที่คั่นหนังสือมาขั้นที่ข้อสอบที่เป็นตัวแทนของคะแนนจุดตัด ผลการศึกษาพบว่า คะแนนจุดตัดที่กำหนดโดยวิธีแมปมาร์คมีค่าต่ำกว่าคะแนนจุดตัดที่กำหนดโดยวิธีแองกอฟ และยังพบว่าคะแนนจุดตัดที่กำหนดโดยวิธีแมปมาร์คมีการกระจายแบบไม่เป็นโค้งปกติ และมีช่วงคะแนนระหว่างคะแนนสูงที่สุดกับคะแนนต่ำที่สุดต่างกันมากๆ ดังนั้นผู้วิจัยจึงเสนอว่าในการพิจารณาเพื่อกำหนดคะแนนจุดตัดโดยวิธีแมปมาร์ค ควรใช้ค่ามัธยฐานแทนการใช้ค่าเฉลี่ย และเป็นที่น่าสังเกตว่าการกำหนดคะแนนจุดตัดด้วยวิธีแมปมาร์คนั้น ผลคะแนนจุดตัดที่ได้มีค่าน้อยกว่าคะแนนจุดตัดที่ใช้วิธีแองกอฟ ซึ่งคล้ายกับงานวิจัยของ Green และคณะ (2003) ซึ่งนำเสนอไปในตอนต้น

Buckendahl และคณะ (2002) ศึกษาเปรียบเทียบผลการกำหนดจุดตัด โดยวิธีบุ๊คมาร์คและแองกอฟ ซึ่งศึกษากับนักเรียนเกรด 7 ในวิชาคณิตศาสตร์โดยข้อสอบที่ใช้เป็นข้อสอบ เป็นแบบให้คะแนน 2 ค่า การศึกษาครั้งนี้มีการปรับปรุงวิธีบุ๊คมาร์คโดยในการเรียงลำดับข้อสอบตามความยากง่ายนั้นจะกำหนดโดยใช้ค่าความยากที่วิเคราะห์โดยวิธีการตามทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory) แทนการวิเคราะห์โดยใช้ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory) ซึ่งการกำหนดคะแนนจุดตัดทั้งสองวิธีจะให้ผู้เชี่ยวชาญกำหนดคะแนนจุดตัดโดยไม่มีการอภิปรายเกี่ยวกับคะแนนจุดตัดเลย มีเพียงข้อมูลเกี่ยวกับค่าเฉลี่ยของคะแนนจุดตัดที่กำหนดในรอบแรกและสัดส่วนของนักเรียนที่ตอบข้อสอบ แต่ละข้อได้ถูกต้องเท่านั้น โดยให้ผู้เชี่ยวชาญกำหนดคะแนนจุดตัดทั้งหมด 3 รอบ ผลการศึกษาพบว่า คะแนนจุดตัดที่กำหนดโดยวิธีแองกอฟในรอบ

แรกมีค่าสูงกว่าคะแนนจุดตัดที่กำหนดในรอบที่สอง และค่าความเบี่ยงเบนมาตรฐานมีแนวโน้มเพิ่มขึ้นโดยรอบแรกมีค่าส่วนเบี่ยงเบนมาตรฐานเป็น 7.79 ส่วนในรอบที่สองค่าส่วนเบี่ยงเบนมาตรฐานเพิ่มขึ้นเป็น 10.96 ซึ่งแสดงให้เห็นว่าผลการกำหนดคะแนนจุดตัดในรอบที่สอง มีความแตกต่างกันมากระหว่างผู้เชี่ยวชาญแต่ละท่าน สำหรับคะแนนจุดตัดที่กำหนดโดย วิธีบุ๊คมาร์คในรอบแรกมีค่าน้อยกว่าคะแนนจุดตัดในรอบที่สอง และค่าส่วนความเบี่ยงเบนมาตรฐานมีแนวโน้มลดลงโดยรอบแรกมีค่าส่วนเบี่ยงเบนมาตรฐานเป็น 11.03 ส่วนในรอบที่สองมีค่าส่วนเบี่ยงเบนมาตรฐานลดลงเป็น 8.66 ซึ่งแสดงให้เห็นว่าผลการกำหนดคะแนนจุดตัดในรอบที่สองมีความใกล้เคียงกันมากขึ้นนั่นเอง และเมื่อกำหนดคะแนนจุดตัดครั้งสุดท้ายพบว่าทั้งสองวิธีให้คะแนนจุดตัดเท่ากัน ซึ่งเป็นการยืนยันเกี่ยวกับความตรงของผลการตัดสินเมื่อกำหนดด้วยวิธีที่ต่างกัน ทั้งนี้อาจเนื่องจากว่าการกำหนดคะแนนจุดตัดด้วยวิธีบุ๊คมาร์คและวิธีแองกอฟใช้สารสนเทศในการพิจารณาเหมือนกันคือพิจารณาจากข้อสอบดังนั้นจึงทำให้ผลการกำหนดคะแนนจุดตัดมีค่าเท่ากัน

จากการศึกษางานวิจัยที่เปรียบเทียบผลการกำหนดคะแนนจุดตัดระหว่างวิธีบุ๊คมาร์คกับวิธีอื่นๆ สรุปได้ว่าการที่ผลของการกำหนดคะแนนจุดตัดต่างกัน ส่วนหนึ่งมีผลมาจากสารสนเทศที่ใช้ในการพิจารณาเพื่อกำหนดคะแนนจุดตัด ซึ่งถ้าใช้สารสนเทศต่างกัน เช่น ใช้ข้อสอบหรือใช้ผลการปฏิบัติงานของผู้สอบในการพิจารณาย่อมให้คะแนนจุดตัดที่ต่างกัน ดังนั้นหากจะกำหนดคะแนนจุดตัดควรจะให้หลายๆ วิธี และใช้หลายๆ สารสนเทศในการกำหนด เพื่อให้ได้ข้อมูลที่เพียงพอในการกำหนดคะแนนจุดตัดที่จะใช้ในการจัดนักเรียนเข้าในแต่ละระดับความสามารถได้อย่างถูกต้องและเหมาะสมที่สุด

งานวิจัยที่ศึกษาในประเด็นเกี่ยวกับคุณภาพของคะแนนจุดตัดประกอบด้วยงานวิจัยของ Yin และ Sconing (2008) ทำการศึกษาเพื่อประมาณค่าความคลาดเคลื่อนมาตรฐานของคะแนนจุดตัดโดยใช้ทฤษฎีสรุปอ้างอิง ซึ่งกำหนดคะแนนจุดตัดโดยวิธี Item rating ปรับปรุงมาจากวิธีแองกอฟ และวิธีแมปมาร์คปรับจากวิธีบุ๊คมาร์ค ผลจากการศึกษาพบว่าแหล่งของความคลาดเคลื่อนส่วนใหญ่เกิดจากผู้เชี่ยวชาญ และหากต้องการให้ผลการกำหนดคะแนนจุดตัดมีความคลาดเคลื่อนน้อยลงควรเพิ่มจำนวนผู้เชี่ยวชาญให้มากขึ้น ซึ่งสอดคล้องกับผลการศึกษาของสุรีพร อนุศาสนนันท์ (2550) ผลจากการศึกษาพบว่าความเที่ยงของการกำหนดคะแนนจุดตัดด้วยวิธีบุ๊คมาร์คจะเพิ่มขึ้นเมื่อเพิ่มจำนวนผู้เชี่ยวชาญให้มากขึ้น

Lee และ Lewis (2008) ได้ทำการศึกษาเกี่ยวความคลาดเคลื่อนมาตรฐานของการกำหนดคะแนนจุดตัดด้วยวิธีบุ๊คมาร์คโดยใช้ทฤษฎีสรุปอ้างอิง พบว่า แหล่งความแปรปรวนหลักมาจากกลุ่มผู้เชี่ยวชาญและเสนอว่าในการกำหนดคะแนนจุดตัดโดยวิธีบุ๊คมาร์คนั้น ต้องให้ความสำคัญกับกลุ่มผู้เชี่ยวชาญเป็นพิเศษ โดยแนะนำว่าการลดความคลาดเคลื่อนของผลการกำหนดคะแนนจุดตัดให้น้อยลงนั้น ควรเพิ่มจำนวนกลุ่มย่อยของผู้เชี่ยวชาญให้มีจำนวนมากขึ้นแทนที่จะเพิ่มจำนวนผู้เชี่ยวชาญในแต่ละกลุ่มย่อย

จากงานวิจัยที่ศึกษาเกี่ยวกับความถูกต้องคะแนนจุดตัดโดยใช้ทฤษฎีสรุปอ้างอิง สรุปได้ว่า แหล่งความคลาดเคลื่อนส่วนใหญ่ของการกำหนดคะแนนจุดตัดมาจากผู้ตัดสิน และกลุ่มผู้ตัดสินเป็นส่วนใหญ่ ดังนั้นในการวางแผนเพื่อกำหนดคะแนนจุดตัดโดยวิธีบุ๊คมาร์คควรพิจารณาเกี่ยวกับจำนวนผู้ตัดสิน และจำนวนกลุ่มย่อยให้เหมาะสม นอกจากนี้ต้องดำเนินการกำหนดคะแนนจุดตัดให้มีระบบ เพื่อให้ผลการกำหนดคะแนนจุดตัดมีความถูกต้อง ซึ่งจะเป็นสารสนเทศสำหรับการตัดสินใจที่มีประสิทธิภาพ

นักวิจัยจำนวนไม่น้อยสนใจทำการศึกษาเกี่ยวกับความน่าจะเป็นของการตอบข้อสอบได้ถูกต้อง (Response Probability: RP) เนื่องจากการกำหนดคะแนนจุดตัดโดยวิธีบุ๊คมาร์คนั้น ผู้เชี่ยวชาญจะต้องพิจารณาความน่าจะเป็นของการตอบข้อสอบได้ถูกต้องซึ่งเป็นค่าความน่าจะเป็นต่ำที่สุด (Minimum probability) ที่ผู้มีความสามารถคาบเส้นจะตอบข้อได้ถูกต้อง และในการจัดทำคู่มือเรียงข้อสอบ สำหรับใช้ประกอบการพิจารณากำหนดคะแนนจุดตัดด้วยวิธีบุ๊คมาร์คนั้น การเรียงข้อสอบจะเรียงตามลำดับความยากง่ายของข้อสอบ ณ ระดับความน่าจะเป็นของการตอบข้อสอบได้ถูกที่กำหนดไว้

การกำหนดคะแนนจุดตัดด้วยวิธีบุ๊คมาร์คส่วนใหญ่กำหนดความน่าจะเป็นของการตอบถูกไว้ที่ 0.67 หรือ 2 ใน 3 ของผู้มีความสามารถคาบเส้นสามารถตอบถูก โดยมีเหตุผลของการกำหนดความน่าจะเป็นค่านี้ด้วย 2 เหตุผล คือ 1) โมเดลการตอบสนองข้อสอบ แบบ 3 พารามิเตอร์ ฟังก์ชั่นสารสนเทศของข้อสอบจะมีค่าสูงสุดที่ θ สำหรับ $P\theta(=(c+2)/3$ เมื่อไม่พิจารณาโอกาสการเดาถูก และ 2) การกำหนดความน่าจะเป็นของการตอบถูกไว้ที่ 0.67 นั้นทำความเข้าใจและปฏิบัติได้ง่าย (Williams & Schulz, 2005 cited in Yin & Sconing, 2008) ส่วน Wang (2003 cited in Beretvas, 2004) เสนอว่า หากนำโมเดลการวิเคราะห์ข้อสอบแบบ 1 พารามิเตอร์ (Rash model) มา

ใช้ในการวิเคราะห์ข้อสอบ ฟังก์ชั่นสารสนเทศของข้อสอบจะมีค่าสูงสุดเมื่อโอกาสการตอบถูกมีค่าเท่ากับ 0.50

Beretvas (2004) ศึกษาความน่าจะเป็นของการตอบข้อสอบถูก 3 ค่า ดังนี้ 0.50, 0.67 และ 0.80 และศึกษาการวิเคราะห์ข้อสอบด้วยโมเดลการตอบสนองข้อสอบ 4 โมเดล คือ โมเดลแบบ 1 พารามิเตอร์ 2 โมเดล คือ Rash model และ Partial credit โมเดลแบบ 3 พารามิเตอร์ 2 โมเดล คือ Generalize Partial Credit (GPC) และ Graded Response (GR) จากการศึกษา พบว่า การจัดเรียงข้อสอบตามค่าความยากที่ได้จากการวิเคราะห์ข้อสอบโดยใช้โมเดล 1 พารามิเตอร์ คือทั้ง Rash model และ Partial credit นั้นลำดับของข้อสอบที่จัดเรียงจะต่างจากข้อสอบที่เรียงโดยการใช้ค่าความยากที่วิเคราะห์โดยใช้โมเดล Generalize Partial Credit (GPC) 3 พารามิเตอร์ และโมเดล Graded Response (GR) 3 พารามิเตอร์ แต่การจัดเรียงข้อสอบตามระดับความยากง่ายให้ผลเหมือนกันเมื่อใช้ค่าความยากที่วิเคราะห์โดยใช้โมเดล Generalize Partial Credit (GPC) 3 พารามิเตอร์และโมเดล Graded Response (GR) 3 พารามิเตอร์ เมื่อพิจารณาค่าความน่าจะเป็น พบว่า การจัดเรียงข้อสอบแบบให้คะแนน 2 ค่า ตามค่าความยากที่ได้จากการวิเคราะห์ข้อสอบโดยใช้โมเดล 1 พารามิเตอร์ (Rash model) เมื่อกำหนดความน่าจะเป็นของการตอบถูกต่างๆ กัน ผลการจัดเรียงข้อสอบก็ยังคงเป็นลำดับเดิม ส่วนการจัดเรียงข้อสอบแบบให้คะแนนมากกว่า 2 ค่า ตามระดับค่าความยากที่ได้จากการวิเคราะห์ข้อสอบโดยใช้โมเดล 1 พารามิเตอร์ (Partial credit) เมื่อกำหนดความน่าจะเป็นของการตอบถูกไว้ที่ระดับต่างกันผลการจัดเรียงข้อสอบจะแตกต่างกัน และการจัดเรียงข้อสอบแบบให้คะแนน 2 ค่า และให้คะแนนมากกว่า 2 ค่า ตามระดับค่าความยากที่ได้จากการวิเคราะห์ข้อสอบโดยใช้โมเดล 3 พารามิเตอร์แบบ Generalize Partial Credit (GPC) และโมเดล Graded Response (GR) เมื่อกำหนดความน่าจะเป็นของการตอบถูกต่างๆ กัน ผลการจัดเรียงข้อสอบจะต่างกัน

Williams และ Schulz (2005 cited in Yin & Sconing, 2008) ได้ศึกษาความน่าจะเป็นของการตอบข้อสอบถูกแตกต่างกัน คือ 0.50 และ 0.67 ผลจากการศึกษาพบว่า ผู้เชี่ยวชาญมีความคิดเห็นว่าการกำหนดความน่าจะเป็นของการตอบข้อสอบถูกไว้ที่ 0.67 นั้นผู้เชี่ยวชาญมีความเข้าใจและนำไปปฏิบัติได้ง่ายกว่ากำหนดความน่าจะเป็นไว้ที่ 0.50

จากการศึกษาสรุปได้ว่าความน่าจะเป็นของการตอบถูกมีผลต่อการจัดเรียงข้อสอบตามระดับความยากโดยเฉพาะข้อสอบที่วิเคราะห์ด้วยโมเดลการตอบสนองข้อสอบแบบ 3 พารามิเตอร์ และผู้เชี่ยวชาญส่วนใหญ่สนับสนุนการกำหนดความน่าจะเป็นของการตอบถูกที่ 0.67 เนื่องจากให้สารสนเทศของข้อสอบได้สูงที่สุดเมื่อวิเคราะห์ข้อสอบด้วยโมเดลแบบ 3 พารามิเตอร์ แต่หากวิเคราะห์ข้อสอบด้วยโมเดล 1 พารามิเตอร์ ความน่าจะเป็นของการตอบถูกควรเป็น 0.50 จึงจะทำให้ได้สารสนเทศของข้อสอบสูงที่สุด

นอกจากงานวิจัยที่กล่าวมาข้างต้น ยังมีการศึกษาเกี่ยวกับการปรับคะแนนจุดตัด ในกรณีที่ข้อสอบที่ใช้มีความเที่ยงต่ำ 1 เรื่อง คืองานวิจัยของ MacCann (2008) เป็นการศึกษาวิธีการปรับคะแนนจุดตัด (Adjusted Cutscore) ที่ได้จากการกำหนดโดยวิธีบุ๊คมาร์คและวิธีแองกอฟ โดยสูตรในการคำนวณคะแนนจุดตัดปรับนี้จะนำค่าความเที่ยงของข้อสอบมาคำนวณด้วย เป็นการช่วยลดปัญหาเกี่ยวกับการกำหนดคะแนนจุดตัดในกรณีที่ข้อสอบมีความเที่ยงต่ำได้ โดยผลจากการนำคะแนนจุดตัดปรับไปใช้ พบว่า จำนวนนักเรียนที่สอบตกมีจำนวนน้อยกว่าการนำคะแนนจุดตัดที่ยังไม่ได้มีการปรับไปใช้

งานวิจัยที่ศึกษาเกี่ยวกับผู้เชี่ยวชาญ มี 2 งานวิจัย ได้แก่งานวิจัยของ Skorupski และ Hambleton (2003 cited in Ferdous & Plake, 2005) ได้ทำการศึกษาเกี่ยวกับการรับรู้และความเข้าใจเกี่ยวกับการกำหนดคะแนนจุดตัดของผู้เชี่ยวชาญโดยวิธีบุ๊คมาร์ค โดยการสัมภาษณ์ผู้เชี่ยวชาญ ซึ่งช่วงเวลาที่สัมภาษณ์คือ 1) ตอนที่ผู้เชี่ยวชาญมาเข้ารับการอบรม 2) หลังจากอบรมเสร็จแล้ว 3) หลังจากการกำหนดคะแนนจุดตัดรอบที่ 1 4) หลังการอภิปรายกลุ่ม 5) หลังจากกำหนดคะแนนจุดตัดรอบที่สอง และ6) หลังจากกำหนดคะแนนจุดตัดรอบสุดท้าย ซึ่งผลจากการสัมภาษณ์ พบว่า ผู้เชี่ยวชาญที่มาเข้าร่วมการกำหนดคะแนนจุดตัดมีแนวคิด และมุมมองเกี่ยวกับการกำหนดคะแนนจุดตัดแตกต่างกันไป และผู้เชี่ยวชาญล้วนมีความมั่นใจว่าตนเองเข้าใจเกี่ยวกับระดับความสามารถที่จะต้องพิจารณากำหนดคะแนนจุดตัดเป็นอย่างดี นอกจากนี้ผู้เชี่ยวชาญยังให้ความคิดเห็นว่าการให้ข้อมูลสารสนเทศช่วยให้กำหนดคะแนนจุดตัดง่าย

Egan และ Green (2003 cited in Ferdous & Plake, 2005) ได้ทำการศึกษาเกี่ยวกับปัจจัยที่ส่งผลต่อการตัดสินใจกำหนดคะแนนจุดตัดด้วยวิธีบุ๊คมาร์คของผู้เชี่ยวชาญ ซึ่งจากการศึกษา พบว่า

ปัจจัยด้านอายุ ประสบการณ์ในการกำหนดคะแนนจุดตัด และความชัดเจนของคำนิยามความสามารถ ล้วนส่งผลต่อผลการกำหนดคะแนนจุดตัด

ตารางที่ 2.8 สรุปการศึกษาวิจัยเกี่ยวกับการกำหนดคะแนนจุดตัดโดยวิธีบุ๊คมาร์ค

นักวิชาการ	วิชา	ระดับความสามารถ	ข้อสอบ	Response Probability : RP	การใช้ทฤษฎีตอบสนองข้อสอบ (IRT)
Buckendahl และคณะ (2002)	คณิตศาสตร์	2 ระดับ	ข้อสอบแบบเลือกตอบ	0.67	three-parameter logistic (3PL)
Beretvas (2004)	คณิตศาสตร์	4 ระดับ	- ข้อสอบแบบเลือกตอบ - คำถามปลายเปิด	0.50 และ 0.80	- Rash model - Partial credit - Generalize Partial Credit - Graded Response
Lewis (2006)	คณิตศาสตร์และภาษาอังกฤษ	2 ระดับ	ข้อสอบแบบเลือกตอบ	0.67	three-parameter logistic (3PL)
Lee และ Lewis (2008)	คณิตศาสตร์	4 ระดับ	ข้อสอบแบบเลือกตอบ	0.67	three-parameter logistic (3PL)
Yin และ Sconing (2008)	คณิตศาสตร์	3 ระดับ	ข้อสอบแบบเลือกตอบ	0.67	three-parameter logistic (3PL)
MacCann (2008)	คณิตศาสตร์และภาษาอังกฤษ	2 ระดับ	ข้อสอบแบบเลือกตอบ	0.67	three-parameter logistic (3PL)
Lissitz และ Kroopnick (2008)	คณิตศาสตร์และภาษาอังกฤษ	2 ระดับ	ข้อสอบแบบเลือกตอบ	0.50	Rash model

1.8.3 งานวิจัยเกี่ยวกับวิธีการวิเคราะห์กลุ่ม

Sireci (1995) ได้ทำการศึกษาเพื่อประเมินถึงประสิทธิภาพของการวิเคราะห์กลุ่มในกรณีที่นำมาใช้เพื่อกำหนดมาตรฐานความสามารถด้านการอ่านของนักเรียน ซึ่งเป็นเงื่อนไขของการสำเร็จการศึกษา โดยกำหนดมาตรฐานความสามารถเป็นสองระดับ คือ ผ่านกับไม่ผ่าน ศึกษากับข้อมูลจากการสอบ The General Educational Development test (GED) ซึ่งเป็นการสอบเพื่อวัดทักษะการเขียนของนักเรียน โดยสอบกับนักเรียนระดับมัธยมศึกษาตอนปลายเพื่อวัดระดับความสามารถในการเขียนภาษาอังกฤษ ข้อสอบประกอบด้วย 4 ประเด็น ประกอบด้วย การเขียนเรียงความ โครงสร้างประโยค การใช้คำ และการเขียนบรรยายรายละเอียด ซึ่งข้อสอบมี 2 รูปแบบ คือ แบบให้เขียนตอบ และแบบเลือกตอบ ผู้วิจัยได้ทำการวิเคราะห์กลุ่มโดยใช้ผลการสอบในแต่ละสาระเป็นตัวแปรในการแบ่งกลุ่ม ซึ่งสามารถแบ่งกลุ่มได้ 5 กลุ่ม ซึ่งผลจากการวิเคราะห์กลุ่มทำให้ทราบว่าผู้สอบแต่ละกลุ่มสามารถทำคะแนนได้มากในสาระใดซึ่งถือว่าเป็นจุดเข็งของผู้สอบ และผู้สอบทำข้อสอบได้น้อยในสาระใดซึ่งถือว่าเป็นจุดอ่อนของผู้สอบ ซึ่งผู้วิจัยกำหนดให้กลุ่มที่ 3 เป็นกลุ่มคาบเส้นเนื่องจากเป็นกลุ่มที่ได้คะแนนต่ำกว่าคะแนนเฉลี่ยเล็กน้อย โดยใช้คะแนนมัธยฐานของกลุ่มที่ 3 เป็นคะแนนจุดตัด นอกจากนี้ผู้วิจัยได้ทำการศึกษาความตรงของการกำหนดคะแนนจุดตัดด้วยวิธีวิเคราะห์กลุ่ม โดยการหาความสัมพันธ์ระหว่างกลุ่มกับเกรดที่ผู้สอบได้จากการเรียนวิชาภาษาอังกฤษ ซึ่งพบว่าสัมพันธ์กันทางบวก และหาความสัมพันธ์กับคะแนนจุดตัดที่ได้จากการกำหนดก่อนหน้านี้โดยวิธีแองกอฟ พบว่า มีคะแนนจุดตัดใกล้เคียงกัน จากการศึกษาครั้งนี้ทำให้ได้ความรู้เกี่ยวกับการนำการวิเคราะห์กลุ่มเพื่อใช้ในการกำหนดคะแนนจุดตัด ซึ่งการวิเคราะห์กลุ่มสามารถแสดงให้เห็นถึงจุดแข็งและจุดอ่อนในแต่สาระที่สอบวัดของผู้สอบ และแสดงให้เป็นว่าการกำหนดคะแนนจุดตัดโดยวิธีวิเคราะห์กลุ่มมีความตรงเป็นที่น่าพอใจ

นอกจากนี้ Sireci และคณะ (1995) และ Sireci (2001 cited in Hess, 2007) ได้เสนอวิธีในการกำหนดคะแนนจุดตัดจากการวิเคราะห์กลุ่ม ซึ่งสามารถสรุปได้ 3 วิธี ดังนี้คือ วิธีแรก กำหนดคะแนนจุดตัดเป็นช่วงซึ่งได้จากคะแนนที่ซ้อนทับกันระหว่างแต่ละระดับความสามารถ มีค่าระหว่างคะแนนสูงที่สุดของระดับความสามารถที่ต่ำกว่าไปถึงคะแนนที่ต่ำที่สุดของระดับความสามารถที่สูงกว่า วิธีที่สองกำหนดคะแนนจุดตัดเป็นค่าคะแนนเดียวคือ คะแนนสอบที่ตัดกันระหว่างแต่ละระดับความสามารถ นั้นคือต้องทำการแจกแจงคะแนนของแต่ละกลุ่มที่เป็นตัวแทน

แต่ละระดับความสามารถ และจุดที่คะแนนตัดกันจะถือว่าเป็นคะแนนจุดตัด ส่วนวิธีสุดท้าย คือ การใช้การวิเคราะห์ลอจิสติกส์ (Logistic Regression Analysis) ซึ่งเป็นการวิเคราะห์เพื่อกำหนดคะแนนจุดตัดเพียงค่าเดียวที่สามารถจำแนกผู้สอบตามความแตกต่างได้ดีที่สุด

ต่อมา Sireci และคณะ (1999) ได้ใช้วิธีการวิเคราะห์กลุ่มกำหนดคะแนนจุดตัดสำหรับการสอบ The Connecticut Mastery Testing (CMT) วิชาคณิตศาสตร์ ของนักเรียนเกรด 7 ข้อสอบจำนวน 134 ข้อ ประกอบด้วย ข้อสอบแบบเลือกตอบและแบบเติมคำตอบ สอบวัดทั้งหมด 4 เนื้อหา ได้แก่ แนวคิด การคำนวณและการประมาณ การแก้ปัญหาและการประยุกต์ใช้ การวัดและเรขาคณิต โดยกำหนดจุดตัดเพื่อแบ่งผู้สอบตามระดับความสามารถ 3 ระดับ คือ intervention, proficient และ excellence ทำการวิเคราะห์กับกลุ่มผู้สอบจำนวน 2 กลุ่มๆ ละ 405 คน เป็นตัวแปรในการแบ่งกลุ่ม มีทั้งหมด 3 แปร ประกอบด้วย 1) คะแนนสอบทั้งหมด 2) คะแนนสอบแต่ละองค์ประกอบ โดยวิเคราะห์ปัจจัยโดยใช้การวิเคราะห์องค์ประกอบและ 3) คะแนนสอบแต่ละเนื้อหา ใช้การวิเคราะห์กลุ่มแบบลดหลั่น (Hierarchical Cluster Analysis: HCA) เพื่อพิจารณาเกี่ยวกับจำนวนกลุ่มที่เหมาะกับข้อมูล จากนั้นจึงวิเคราะห์โดยใช้ K-Means Clustering เพื่อแบ่งกลุ่มผู้สอบตามตัวแปรที่กำหนด จากนั้นก็กำหนดคะแนนจุดตัดโดยใช้การวิเคราะห์ลอจิสติกส์ และใช้ค่ามัธยฐาน

Violato และคณะ (2003) ศึกษาเพื่อเปรียบเทียบความเหมือนและความแตกต่างระหว่างคะแนนจุดตัดที่กำหนดโดยผู้เชี่ยวชาญ ได้แก่ วิธีของอีเบล และวิธีของนีเดลสกี กับการวิเคราะห์กลุ่ม เพื่อให้ได้ข้อมูลสนับสนุนเกี่ยวกับความตรงของคะแนนจุดตัดที่กำหนดขึ้น เนื่องจากเป็นการเปรียบเทียบคะแนนจุดตัดที่กำหนดโดยใช้ 3 วิธี ผู้วิจัยวิเคราะห์ร้อยละของผู้ที่อยู่ในกลุ่มผ่าน/ไม่ผ่าน ที่ถูกกำหนดโดยใช้จุดตัดที่ได้จากการกำหนดทั้ง 3 วิธี ในการวิจัยครั้งนี้ศึกษาข้อมูลจากผู้สอบจำนวน 243 คน ที่เข้าสอบ Canadian Standard Assessment in Optometry (CSAO) ใช้การวิเคราะห์กลุ่มแบบ K–Means Clustering โดยตัวแปรที่ใช้คือ เนื้อหาที่สอบวัดซึ่งมีทั้งหมด 6 ฉบับ ผลจากการศึกษา พบว่า คะแนนจุดตัดที่กำหนดโดยการวิเคราะห์กลุ่มสอดคล้องกับคะแนนจุดตัดที่ได้จากการกำหนดโดยวิธีของอีเบล และวิธีของนีเดลสกี ซึ่งสามารถใช้เป็นหลักฐานด้านความตรง ของคะแนนจุดตัดที่กำหนดโดยวิธีให้ผู้เชี่ยวชาญเป็นผู้กำหนด ทั้งสองวิธี

Hess และคณะ (2007) ศึกษาความสอดคล้องระหว่างคะแนนจุดตัดที่กำหนดโดยใช้สองวิธีคือ ผู้สอบเป็นศูนย์กลางได้แก่การวิเคราะห์กลุ่ม และวิธีข้อสอบเป็นศูนย์กลางได้แก่ วิธีแองกอฟ

สำหรับการสอบเพื่อการรับรอง โดยข้อสอบเป็นแบบเลือกตอบ จำนวน 240 ข้อ ซึ่งประกอบด้วยสองมิติ คือ 1 มิติด้านเนื้อหา (Content dimension) ซึ่งสอบวัดทั้งหมด 13 เนื้อหา เช่น เนื้อหาเกี่ยวกับหลอดเลือดหัวใจและมิติด้านกระบวนการ สอบวัดทั้งหมด 7 กระบวนการ เช่น กระบวนการคิดวิเคราะห์ เป็นต้น ผู้เชี่ยวชาญจำนวน 15 คน ร่วมกันกำหนดคะแนนจุดตัดโดยวิธีแองกอฟ ส่วนการกำหนดคะแนนจุดตัดโดยวิธีการวิเคราะห์กลุ่มใช้การวิเคราะห์กลุ่มแบบไม่เป็นขั้นตอน และใช้วิธีการวิเคราะห์ลอจิสติกส์ เพื่อกำหนดคะแนนจุดตัดเพียงค่าเดียวที่สามารถจำแนกผู้สอบตามความแตกต่างได้ดีที่สุด พิจารณาความสอดคล้องระหว่างคะแนนจุดตัดที่กำหนดโดยวิธีวิเคราะห์กลุ่มและวิธีแองกอฟ โดยพิจารณาจากร้อยละของผู้สอบที่ถูกจัดไว้แต่ละกลุ่ม และค่าสัมประสิทธิ์แคปป้า ผลการศึกษาพบว่า ผลการกำหนดกลุ่มระหว่างคะแนนจุดตัดที่ได้จากวิธีวิเคราะห์กลุ่ม และวิธีแองกอฟที่ได้จากการกำหนดในรอบแรกมีความสอดคล้องกัน แต่คะแนนจุดตัดที่กำหนดโดยวิธีวิเคราะห์กลุ่มมีค่าสูงกว่าคะแนนจุดตัดที่ได้จากการกำหนดโดยวิธีแองกอฟที่ได้จากการกำหนดในรอบที่สาม ซึ่งผู้วิจัยกล่าวว่า สาเหตุที่ผลการศึกษาเป็นเช่นนั้นเนื่องจากว่า การกำหนดคะแนนจุดตัดโดยวิธีแองกอฟในรอบแรกนั้น ผู้เชี่ยวชาญตัดสินจากข้อสอบเพียงอย่างเดียว โดยไม่มีสารสนเทศใดๆ ประกอบการกำหนดคะแนนจุดตัด ส่วนในรอบที่ 3 นั้น ผู้เชี่ยวชาญได้รับสารสนเทศเกี่ยวกับร้อยละของนักเรียนที่ตอบข้อสอบแต่ละข้อจึงอาจส่งผลต่อการตัดสินใจกำหนดจุดตัด จึงทำให้ผลการกำหนดคะแนนจุดตัดต่ำกว่าคะแนนจุดตัดที่กำหนดโดยวิธีวิเคราะห์กลุ่ม

ตารางที่ 2.9 สรุปการศึกษาวิจัยเกี่ยวกับการกำหนดคะแนนจุดตัดโดยการวิเคราะห์กลุ่ม

นักวิชาการ	วัตถุประสงค์	ข้อสอบ	ระดับ	ตัวแปรในการแบ่งกลุ่ม	วิธีการรวมกลุ่ม	กำหนดคะแนนจุดตัด	การพิจารณาความสอดคล้อง
Sireci (1995)	กำหนดคะแนนจุดตัดสำหรับการสอบการอ่าน	เขียนตอบและแบบเลือกตอบ	2	ผลการสอบในแต่ละสาระการเรียนรู้	K-Means Clustering	ใช้ค่ามัธยฐาน	1. หาความสัมพันธ์กับคะแนนจุดตัดที่ได้จากการกำหนดโดยวิธีแองกอฟ 2. หาความสัมพันธ์กับเกรดที่ผู้สอบได้จากการเรียนวิชาภาษาอังกฤษ
Sireci และคณะ (1999)	กำหนดคะแนนจุดตัดสำหรับการสอบ The Connecticut Mastery Testing (CMT) วิชาคณิตศาสตร์	1) ข้อสอบแบบเลือกตอบ 2) คำถามปลายเปิด	3	1) คะแนนสอบทั้งหมด 2) คะแนนสอบแต่ละองค์ประกอบ โดยวิเคราะห์ปัจจัยโดยใช้การวิเคราะห์องค์ประกอบและ 3) คะแนนสอบแต่ละเนื้อหา	การวิเคราะห์กลุ่มแบบขั้นตอนและวิเคราะห์แบบ K–Means Clustering	ใช้การวิเคราะห์ลอจิสติกส์ และใช้ค่ามัธยฐาน	-
Violato และคณะ 2003	เพื่อเปรียบเทียบความเหมือนและความแตกต่างระหว่างคะแนนจุดตัดที่กำหนดโดยวิธีให้ผู้เชี่ยวชาญเป็นผู้กำหนด ได้แก่ วิธีของอีเบล และวิธีของนีเดลสกี กับการวิเคราะห์กลุ่ม	ข้อสอบแบบเลือกตอบ	2	เนื้อหาที่สอบวัด ซึ่งมีทั้งหมด 8 เรื่อง	K–Means Clustering	ค่ามัธยฐาน	1. ร้อยละของผู้ที่อยู่ในกลุ่มผ่าน/ไม่ผ่าน 2. ค่าสัมประสิทธิ์แคปป้า
Hess และคณะ (2007)	ศึกษาความสอดคล้อง ระหว่างคะแนนจุดตัดที่กำหนดโดยใช้สองวิธีคือ ผู้สอบเป็นศูนย์กลาง ได้แก่ การวิเคราะห์กลุ่ม และวิธีข้อสอบเป็นศูนย์กลาง ได้แก่ วิธีแองกอฟ	แบบเลือกตอบ	2	1. มิติด้านเนื้อหา 2. มิติด้านกระบวนการ	K–Means Clustering	กำหนดคะแนนจุดตัดค่าเดียวด้วยวิธีการวิเคราะห์ลอจิสติกส์	1. ร้อยละของผู้สอบที่ถูกจัดไว้แต่ละกลุ่ม 2. ค่าสัมประสิทธิ์แคปป้า

ตอนที่ 2 ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (Generalizbility Theory)

2.1 แนวคิดพื้นฐานของทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด

ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด ได้เสนอโมเดลวิเคราะห์ความน่าเชื่อถือของผลการวัดหรือความเที่ยงในสถานการณ์หรือเงื่อนไขต่าง ๆ ของการของการวัด ทำให้ทราบและสามารถควบคุมแหล่งความแปรปรวนได้ตรงประเด็น เพื่อให้ผลการวัดมีความน่าเชื่อถือหรือมีความเที่ยงสูงถึงระดับที่ต้องการ (ศิริชัย กาญจนวาสี, 2550)

ครอนบาค และคณะ (1972, อ้างถึงใน ศิริชัย กาญจนวาสี, 2550) ได้เสนอทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (G-Theory) ซึ่งเป็นวิธีการวิเคราะห์ค่าสัมพัทธ์ของความแปรปรวนของคะแนนความคลาดเคลื่อนอย่างเป็นระบบจากแหล่งต่างๆ อันเป็นสถานการณ์หรือเงื่อนไขของการวัด เช่น ความยาวของแบบสอบ ชุดของแบบสอบ จำนวนครั้งของการทดสอบ จำนวนผู้ตรวจ เป็นต้น รวมทั้งปฏิสัมพันธ์ระหว่างผู้สอบกับเงื่อนไขของการวัด เช่น ปฏิสัมพันธ์ระหว่างผู้สอบกับชุดของแบบสอบ ผู้สอบกับครั้งของการทดสอบ เป็นต้น ซึ่งจะทำให้ทราบและสามารถควบคุมแหล่งความคลาดคลาดเคลื่อนได้ตรงประเด็น ทำให้สัมประสิทธิ์ความน่าเชื่อถือหรือคล้ายกับสัมประสิทธิ์ความเที่ยง (จากทฤษฎีการทดสอบแบบดังเดิม) ของการวัดสูงขึ้น ดังนั้น ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัดเป็นวิธีการประเมินความน่าเชื่อถือหรือความเที่ยงของเครื่องมือวัดและกลยุทธ์การออกแบบการวัดให้ผลการวัดมีความน่าเชื่อถือ หรือมีความเที่ยงถึงระดับที่ต้องการ เพื่อนำผลไปใช้เป็นสารสนเทศสำหรับการตัดสินใจได้อย่างมีประสิทธิภาพ

2.2 คำศัพท์ที่สำคัญ

ในการทำความเข้าใจเกี่ยวกับทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด มีความจำเป็นต้องทำความเข้าใจเกี่ยวกับคำศัพท์สำคัญ ดังต่อไปนี้

2.2.1 ประชากร (Population) หมายถึง สิ่งที่เป็นเป้าหมายของการวัดหรือสิ่งที่มุ่งวัดทั้งหมด (Object of Measurement) สำหรับสถานการณ์ของการทดสอบทั่วไป สิ่งที่มุ่งวัดได้แก่ ผู้สอบ

2.2.2 เอกภพ (Universe) หมายถึง เงื่อนไขการวัดที่สนใจทั้งหมดของแต่ละองค์ประกอบ ประชากร (Population) หมายถึง สิ่งที่มุ่งวัดทั้งหมด ในสถานการณ์ของการทดสอบทั่วไป สิ่งที่มุ่งวัดมักได้แก่ ผู้ทำการทดสอบ หรือนักเรียน

2.2.3 องค์ประกอบ (Facet) หมายถึง กลุ่มเงื่อนไขของการวัด ซึ่งเป็นองค์ประกอบที่ คาดว่ามีผลต่อความคลาดเคลื่อนของการวัด เช่น ความยาวของแบบสอบ รูปแบบของข้อสอบ จำนวนครั้งของการทดสอบ จำนวนผู้ตรวจให้คะแนนเป็นต้น

2.2.4 คะแนนเอกภพ หมายถึง คะแนนของเงื่อนไขการวัดทั้งหมดแต่ละองค์ประกอบ ซึ่งคล้ายกับคะแนนจริง (True score) ในทฤษฎีการทดสอบแบบดั้งเดิม

2.2.5 รูปแบบความสัมพันธ์ระหว่างองค์ประกอบ หมายถึง องค์ประกอบที่เป็นสิ่งที่ชี้ถึงแหล่งความแปรปรวนของรูปแบบการวัด ในแบบจำลองการวิเคราะห์ความแปรปรวน ซึ่งสามารถแบ่งรูปแบบความสัมพันธ์ได้ 3 รูปแบบ ได้แก่

2.2.5.1 ความสัมพันธ์แบบไขว้ (Crossed) หมายถึง ความสัมพันธ์ที่แต่ละระดับของสิ่งที่ถูกวัด ถูกวัดภายใต้เงื่อนไขเดียวกันทั้งหมด โดยใช้สัญลักษณ์ “x” อ่านว่า “crossed with”

2.2.5.2 ความสัมพันธ์แบบแฝง (Nested) หมายถึง ความสัมพันธ์ที่แต่ละระดับของสิ่งที่ถูกวัด ถูกวัดภายใต้เงื่อนไขที่ต่างกัน ใช้สัญลักษณ์ “:” อ่านว่า “nested within”

2.2.5.3 ความสัมพันธ์แบบผสม (Confounded) หมายถึง รูปแบบความสัมพันธ์ที่มีทั้งความสัมพันธ์แบบไขว้และความสัมพันธ์แบบแฝงอยู่ด้วยกัน

2.2.6 เอกภพของการสังเกตที่ยอมรับได้ (Universe of Admissible Observation) หมายถึง กลุ่มเงื่อนไขของการวัดที่สามารถวัดหรือสังเกตได้แต่ละองค์ประกอบ

2.2.7 เอกภพของการสรุปอ้างอิง (Universe of Generalizability) หมายถึง เงื่อนไขของการวัดทั้งหมดขององค์ประกอบที่ต้องการสรุปอ้างอิงผลการวัด จากกลุ่มตัวอย่างของเงื่อนไขการวัดไปยังกลุ่มเงื่อนไขการวัดทั้งหมดขององค์ประกอบนั้น ๆ

2.3 การศึกษา G (G-Study) และการศึกษา D (D-Study)

ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด ประกอบด้วย 2 ขั้นตอนสำคัญ ได้แก่ การศึกษาเชิงสรุปอ้างอิง หรือการศึกษา G (Generalizability Study: G-Study) กับการศึกษาเชิงตัดสินใจ (Decision Study: D-Study) (Cronbach & other, 1972; Brennen, 1983 อ้างถึงใน ศิริชัย กาญจนวาสี, 2550) ซึ่งมีรายละเอียดของแต่ละขั้นตอนดังต่อไปนี้

2.3.1 การศึกษาเชิงสรุปอ้างอิง (G-Study) เป็นการสรุปอ้างอิงผลจากการศึกษาตัวอย่างการวัดตามเงื่อนไขที่สนใจ ประมาณความแปรปรวนจากแหล่งต่างๆ เพื่ออ้างอิงไปยังเอกภพของการวัด ประกอบด้วยขั้นตอนต่อไปนี้

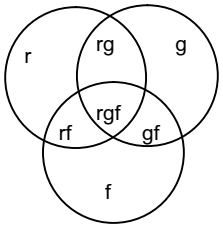
2.3.1.1 กำหนดสิ่งที่จะวัด เช่น ความสามารถทางคณิตศาสตร์ และความสามารถทางวิทยาศาสตร์ เป็นต้น

2.3.1.2 กำหนดองค์ประกอบ (Facet) ของการวัด เช่น กำหนดให้มี 3 องค์ประกอบ คือ จำนวนผู้ตัดสิน จำนวนกลุ่มผู้ตัดสิน และจำนวนชุดข้อสอบ เป็นต้น

2.3.1.3 กำหนดเอกภพของการสังเกตที่ยอมรับได้ ในที่นี้ประกอบด้วย องค์ประกอบของผู้ตัดสิน องค์ประกอบของกลุ่มผู้ตัดสินและองค์ประกอบของชุดข้อสอบ

2.3.1.4 กำหนดความสัมพันธ์ระหว่างองค์ประกอบ ซึ่งจะมีทั้งความสัมพันธ์แบบไขว้ (Crossed) หรือแบบแฝง (Nested) หากกำหนดความสัมพันธ์แบบไขว้ (Crossed) จะทำให้ได้จำนวนแหล่งความแปรปรวนมากกว่าแบบแฝง (Nested)

2.3.1.5 เก็บรวบรวมข้อมูล โดยการสุ่มตัวอย่างมาศึกษา เช่น ในที่นี้สุ่มตัวอย่างผู้ตัดสิน n_r คน ตัวอย่างกลุ่มผู้ตัดสิน n_g คน และตัวอย่างชุดข้อสอบ n_f ชุด เป็นการศึกษา G-study $r \times g \times f$ design ซึ่งมีแหล่งความแปรปรวนทั้งสิ้น 7 แหล่ง ประกอบด้วย ผลหลัก (main effect) 3 แหล่ง คือ ผลจากผู้ตัดสิน ผลจากกลุ่มผู้ตัดสินและผลจากชุดข้อสอบ ผลของปฏิสัมพันธ์ 4 แหล่ง คือ ปฏิสัมพันธ์ระหว่างผู้ตัดสินกับกลุ่มผู้ตัดสิน ผู้ตัดสินกับชุดข้อสอบ กลุ่มผู้ตัดสินกับชุดข้อสอบ และผู้ตัดสินกับกลุ่มผู้ตัดสินและชุดข้อสอบ ดังแสดงในภาพต่อไปนี้



ภาพที่ 2.5 องค์ประกอบความแปรปรวนของ G-study $r \times g \times f$ design เมื่อตัวประกอบที่ศึกษาทั้งหมดเป็นตัวประกอบสุ่ม

จากภาพ เป็นการแสดงองค์ประกอบของความแปรปรวน ของ G-study $r \times g \times f$ design โดยมีแหล่งความแปรปรวนทั้งสิ้น 7 แหล่ง ประกอบด้วย ผลหลัก (main effect) 3 แหล่ง คือ ผลจากผู้ตัดสิน (*r*) ผลจากกลุ่มผู้ตัดสิน (*g*) และผลจากชุดข้อสอบ (*f*) ผลของปฏิสัมพันธ์ 4 แหล่ง คือ ปฏิสัมพันธ์ระหว่างผู้ตัดสินกับกลุ่มผู้ตัดสิน (*rg*) ผู้ตัดสินกับชุดข้อสอบ (*rf*) กลุ่มผู้ตัดสินกับชุดข้อสอบ (*gf*) และผู้ตัดสินกับกลุ่มผู้ตัดสินและชุดข้อสอบ (*rgf*)

2.3.1.6 คำนวณความแปรปรวนของแหล่งต่างๆ โดยการวิเคราะห์ความแปรปรวนแบบ Factorial Design $r \times g \times f$ เพื่อประมาณค่าความแปรปรวนจากค่าเฉลี่ยกำลังสอง โดยใช้สูตรเฉพาะที่สอดคล้องกับรูปแบบ

2.3.1.6.1 การคำนวณค่าเฉลี่ยกำลังสองของความเบี่ยงเบน (Mean Square of Deviation) จากสูตรต่อไปนี้

$$\mathrm{MS}(\alpha) = \mathrm{SS}(\alpha)/\mathrm{df}(\alpha)$$

เมื่อ $\mathrm{MS}(\alpha)$ แทน ค่าเฉลี่ยกำลังสองของความเบี่ยงเบน

$\mathrm{SS}(\alpha)$ แทน ผลรวมกำลังสองของความเบี่ยงเบน

$\mathrm{df}(\alpha)$ แทน ชั้นแห่งความเป็นอิสระที่สอดคล้องกับผลขององค์ประกอบ

2.3.1.6.2 วิเคราะห์ข้อมูล โดยใช้กระบวนการวิเคราะห์ความแปรปรวน (Analysis of Variance Procedure)

2.3.2 การศึกษาเชิงตัดสินใจ (Decision Study: D-Study) เป็นขั้นตอนการนำผลที่ได้จากการประมาณความแปรปรวนในการศึกษาเชิงสรุปอ้างอิง(G-Study) มีขั้นตอนดังนี้

2.3.2.1 กำหนดเอกภพของการสรุปอ้างอิงที่ผู้ตัดสินใจต้องการสรุปอ้างอิงในการศึกษาเชิงตัดสินใจ (D-Study) ซึ่งเอกภพของการสรุปอ้างอิงอาจประกอบด้วยเงื่อนไขทั้งหมดหรือเพียงเงื่อนไขย่อยในเอกภพของการสังเกตที่ยอมรับได้

2.3.2.2 เลือกแบบแผนการศึกษาเพื่อการตัดสินใจหรือความสัมพันธ์ขององค์ประกอบที่ต้องการศึกษา โดยแบบแผนการศึกษาเพื่อการตัดสินใจ เช่น ต้องการตัดสินใจว่าใน

การที่ผู้ตัดสินแต่ละคนอยู่ต่างกลุ่มกัน พิจารณาข้อสอบทุกฉบับ แบบแผนดังกล่าวเป็น $(r:g)\times f$ เป็นต้น

2.3.2.3 กำหนดขนาดตัวอย่างของตัวอย่างศึกษาเพื่อการตัดสินใจ โดยจำนวนเงื่อนไขขององค์ประกอบในการศึกษาเพื่อตัดสินใจ สามารถกำหนดให้แตกต่างจากจำนวนเงื่อนไขในการศึกษาเพื่อการสรุปอ้างอิง

2.3.2.4 ประมาณค่าความแปรปรวน เป็นการประมาณค่าความแปรปรวนที่สอดคล้องกับแบบแผนและขนาดของตัวอย่างที่ต้องการตัดสินใจ

2.3.2.5 ประมาณค่าความแปรปรวนของความคลาดเคลื่อน ซึ่งแบ่งได้เป็น 2 ชนิด คือ ความแปรปรวนของความคลาดเคลื่อนแบบสัมบูรณ์ (Absolute Error Variance) เป็นค่าที่คำนวณจากผลบวกของค่าองค์ประกอบความแปรปรวนอื่น ๆ ทั้งหมด ยกเว้นความแปรปรวนของคะแนนเอกภพ และความแปรปรวนของความคลาดเคลื่อนแบบสัมพัทธ์ (Relative Error Variance) คำนวณจากผลบวกค่าขององค์ประกอบความแปรปรวนของผลร่วมระหว่างสิ่งที่ถูกวัดกับองค์ประกอบอื่นๆ

2.3.2.6 ประมาณค่าสัมประสิทธิ์การสรุปอ้างอิง (Generalizability Coefficient) เป็นดัชนีชี้บ่งความน่าเชื่อถือของผลการวัด คำนวณจากอัตราส่วนระหว่างความแปรปรวนของคะแนนเอกภพและความแปรปรวนของคะแนนสังเกตที่คาดหวัง ซึ่งสัมประสิทธิ์การสรุปอ้างอิงมี 2 ชนิด

สัมประสิทธิ์การสรุปอ้างอิงเพื่อการตัดสินใจเชิงสัมพัทธ์ ซึ่งสามารถคำนวณได้ดังสูตรต่อไปนี้

$$\rho^2_{Rel} = \frac{\sigma^2_p}{\sigma^2_p + \sigma^2_{Rel}}$$

สัมประสิทธิ์การสรุปอ้างอิงเพื่อการตัดสินใจเชิงสัมบูรณ์ ซึ่งสามารถคำนวณได้ดังสูตรต่อไปนี้

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

ตอนที่ 3 ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT)

3.1 แนวคิดพื้นฐานของทฤษฎีการตอบสนองข้อสอบ

ทฤษฎีการตอบสนองข้อสอบเป็นทฤษฎีการวัดที่อธิบายความสัมพันธ์ระหว่างความสามารถที่อยู่ภายในบุคคล (Latent taint or ability) กับพฤติกรรมการตอบสนองข้อสอบของบุคคลว่ามีโอกาสตอบถูกต้องมากเพียงไร ทฤษฎีนี้มีพื้นฐานความเชื่อว่า พฤติกรรมการตอบสนองต่อข้อสอบซึ่งเป็นสิ่งที่สังเกตได้โดยตรงว่าถูกหรือผิด จะถูกกำหนดโดยคุณลักษณะภายในหรือความสามารถที่อยู่ภายในตัวบุคคล ซึ่งเป็นสิ่งที่ไม่สามารถสังเกตได้โดยตรง ความสัมพันธ์ระหว่างความน่าจะเป็นในการตอบข้อสอบแต่ละข้อได้ถูกต้อง [$P_i(\theta)$] กับระดับความสามารถของผู้สอบ (θ) เมื่อนำมาเขียนเป็นกราฟจะได้โค้งลักษณะข้อสอบ (Item Characteristic Curve: ICC) ซึ่งโดยทั่วไปมีลักษณะเป็นโค้งรูปตัวเอส (S-Shape) มีฟังก์ชั่นทางคณิตศาสตร์ 2 ประเภท ได้แก่ ฟังก์ชั่นโลจิสติก (Logistic Function) กับฟังก์ชันปกติสะสม (Normal Ogive Function) โค้งลักษณะข้อสอบที่คำนวณจากฟังก์ชั่นโลจิสติกมีวิธีการคำนวณที่ง่าย สะดวก และทนทานต่อความคลาดเคลื่อนที่เกิดขึ้นกับผู้สอบที่มีความสามารถสูงจะตอบได้ดีกว่าการคำนวณจากฟังก์ชันปกติสะสม จึงทำให้โมเดลโลจิสติกเป็นที่นิยมนำไปใช้จริง (ศิริชัย กาญจนวาสี, 2550)

พารามิเตอร์ของโมเดลการตอบสนองข้อสอบประกอบด้วยพารามิเตอร์และค่าคงที่ดังนี้

พารามิเตอร์ของผู้สอบ (Examinee's Parameter) (θ_p) คือ ระดับความสามารถของผู้สอบคนที่ P ซึ่งประมาณได้จากโมเดลตามทฤษฎีการตอบสนองข้อสอบ และปรับให้เป็นคะแนนมาตรฐานที่มีค่าเฉลี่ยเป็น 0 และส่วนเบี่ยงเบนมาตรฐานเป็น 1 ค่า θ มีพิสัยอยู่ระหว่าง $-\infty$ ถึง $+\infty$ แต่ในทางปฏิบัตินิยมใช้ในช่วง -3 ถึง +3

พารามิเตอร์ของข้อสอบ (Item Parameters) ประกอบด้วย

b_i คือ ค่าความยากของข้อสอบข้อที่ i (Item Difficulty) ซึ่งเป็นการวัดตำแหน่งของโค้งคุณลักษณะของข้อสบ (ICC) ตามแกนนอนบนสเกลของ θ ณ จุดที่โค้งมีความชันมากที่สุด (จุดเปลี่ยนโค้ง) หรือที่ตำแหน่งต่อไปนี้

$b_i = \theta$ ที่ $P_i(\theta) = 0.50$ (สำหรับ 1 – และ 2 – Parameter Model)

$b_i = \theta$ ที่ $P_i(\theta) = \frac{1+c}{2}$ (สำหรับ 3 - Parameter Model)

ค่า b มีพิสัยอยู่ระหว่าง - ∞ ถึง + ∞ แต่ในทางปฏิบัตินิยมใช้ข้อสอบที่ค่า b_i อยู่ระหว่าง -2.5 ถึง +2.5 ค่า b_i ที่อยู่ใกล้ -2.5 แสดงว่าเป็นข้อสอบที่ง่าย ส่วน b_i ที่อยู่ใกล้ +2.5 แสดงว่าเป็นข้อสอบที่ยาก

a_i คือ ค่าอำนาจจำแนกของข้อสอบที่ i (Item Discrimination) ซึ่งเป็นความชันของโค้ง ICC ณ จุดเปลี่ยนโค้ง หรือที่จุด $\theta = b_i$ ค่า a มีพิสัยอยู่ระหว่าง - ∞ ถึง + ∞ ในทางปฏิบัติจึงนิยมใช้ข้อสอบที่ค่า a_i อยู่ระหว่าง 0.5 ถึง 2.5 ค่า a_i ที่สูง แสดงว่าข้อสอบข้อนั้นมี Slope ที่ชันจึงจำแนกผู้สอบที่มีความสามารถแตกต่างกันได้ดี

c_i คือ ความน่าจะเป็นของการเดาถูก (Guessing Parameter หรือ pseudo-chance score level) ซึ่งเป็นความน่าจะเป็นที่ผู้สอบที่มีความสามารถต่ำมาก ๆ จะทำข้อสอบข้อที่ i ได้ถูก c_i เป็นค่ากำกับต่ำที่สุด (lower asymptote ของ ICC) ค่า c_i มีพิสัยอยู่ระหว่าง 0 ถึง 1 โดยทั่วไปนิยมใช้ข้อสอบมีค่า c_i อยู่ระหว่าง 0 ถึง 0.3

ค่าคงที่ (Constant)

e = ค่าคงที่ ซึ่งมีค่าประมาณ 2.71828

D = ค่าการปรับสเกล (Scaling Factior) ซึ่งมีค่าเท่ากับ 1.70

โมเดลโลจิสติกที่สำคัญมี 3 โมเดล ได้แก่ โมเดลโลจิสติกแบบหนึ่งพารามิเตอร์ (One–Parameter Logistic Model), โมเดลโลจิสติกแบบสองพารามิเตอร์ (Two–Parameter Logistic Model) และ โมเดลโลจิสติกแบบสามพารามิเตอร์ (Three–Parameter Logistic Model), (ศิริชัย กาญจนวาสี, 2550) มีรายละเอียดดังต่อไปนี้ (ศิริชัย กาญจนวาสี, 2550)

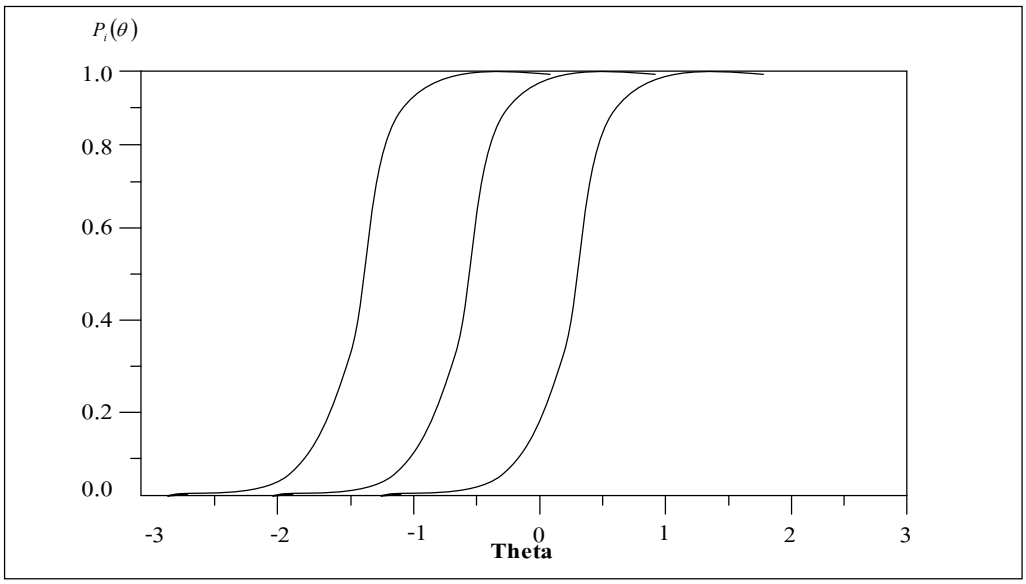
1) โมเดลโลจิสติกแบบหนึ่งพารามิเตอร์ (One–Parameter Logistic Model) เรียกอีกชื่อหนึ่งว่า โมเดลราชซ์ (Simple Rash Model) เป็นโมเดลที่มีความซับซ้อนน้อยที่สุด โดยมีพารามิเตอร์เพียงค่าเดียว คือ ค่าความยากของข้อสอบ (b) โดยโมเดลนี้เกิดจากความเชื่อว่าอำนาจจำแนกของทุกข้อเท่ากันและค่าการเดาของข้อสอบต่ำมากหรือมีค่าเป็น 0 สูตรทางคณิตศาสตร์ของโมเดลนี้คือ

$$P_i(\theta) = \frac{1}{1 + e^{-(\theta - b_i)}}$$

เมื่อ $P_i(\theta)$ แทน ความน่าจะเป็นซึ่งผู้ตอบที่มีระดับความสามารถ θ ตอบข้อ i ได้ถูกต้อง

b_i แทน ค่าพารามิเตอร์ความยากของข้อสอบข้อที่ i

e แทน ค่าคงที่ซึ่งมีค่าเท่ากับ 2.718



ภาพที่ 2.6 โค้งลักษณะข้อสอบแบบ 1 พารามิเตอร์ ของข้อสอบ 3 ข้อ

2) โมเดลโลจิสติกแบบสองพารามิเตอร์ (Two–Parameter Logistic Model) มีพารามิเตอร์สองค่า คือ ค่าความยากของข้อสอบ (b) และ ค่าอำนาจจำแนกของข้อสอบ (a) สูตรทางคณิตศาสตร์ของโมเดลนี้คือ

$$P_i(\theta) = \frac{1}{1 + e^{-D_{ai}(\theta - b_i)}}$$

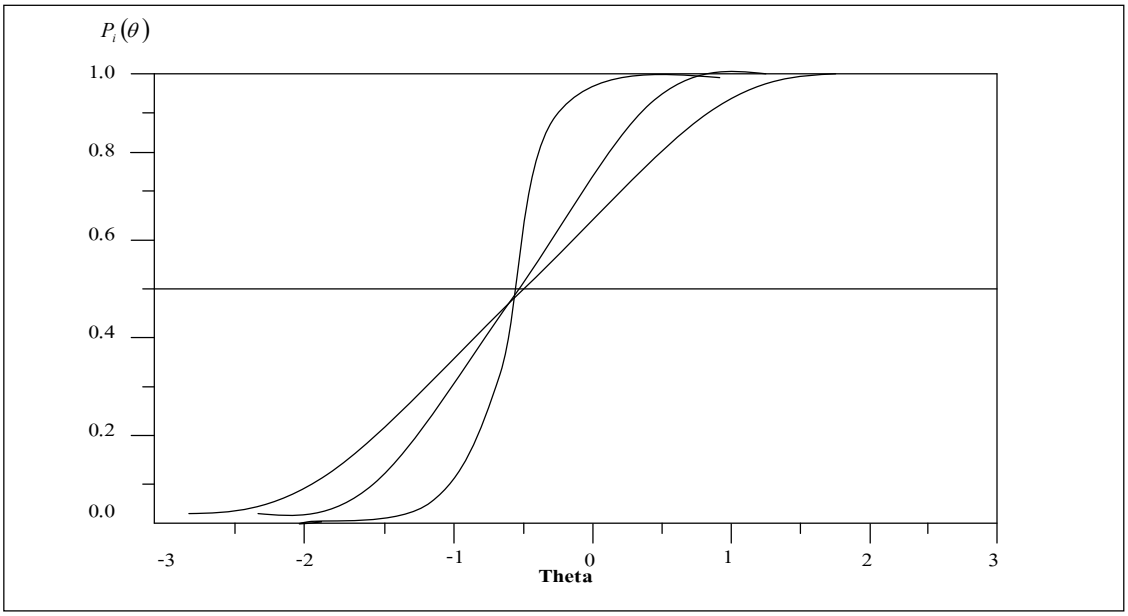
เมื่อ $P_i(\theta)$ แทน ความน่าจะเป็นซึ่งผู้ตอบที่มีระดับความสามารถ θ ตอบข้อ i ได้ถูกต้อง

b_i แทน ค่าพารามิเตอร์ความยากของข้อสอบข้อที่ I เป็นค่าที่แสดงตำแหน่งของโค้งลักษณะข้อสอบ ณ จุด θ ที่มีโอกาสตอบข้อสอบถูก 0.5

a_i แทน ค่าพารามิเตอร์อำนาจจำแนกของข้อสอบข้อที่ i เป็นค่าความชันของโค้งลักษณะข้อสอบ ณ ตำแหน่ง b_i

e แทน ค่าคงที่ซึ่งมีค่าเท่ากับ 2.718

D แทน ค่าคงที่ซึ่งมีค่าเท่ากับ 1.70



ภาพที่ 2.7 โค้งลักษณะข้อสอบแบบ 2 พารามิเตอร์ ของข้อสอบ 3 ข้อ

3) โมเดลโลจิสติกแบบสามพารามิเตอร์ (Three–Parameter Logistic Model) มีพารามิเตอร์สามค่า คือ ค่าความยากของข้อสอบ (b) ค่าอำนาจจำแนกของข้อสอบ (a) และค่าโอกาสในการเดาข้อสอบ (c) สูตรทางคณิตศาสตร์ของโมเดลนี้คือ

$$P_i(\theta) = c_i + \frac{(1 - c_i)}{1 + e^{-D_{ai}(\theta - b_i)}}$$

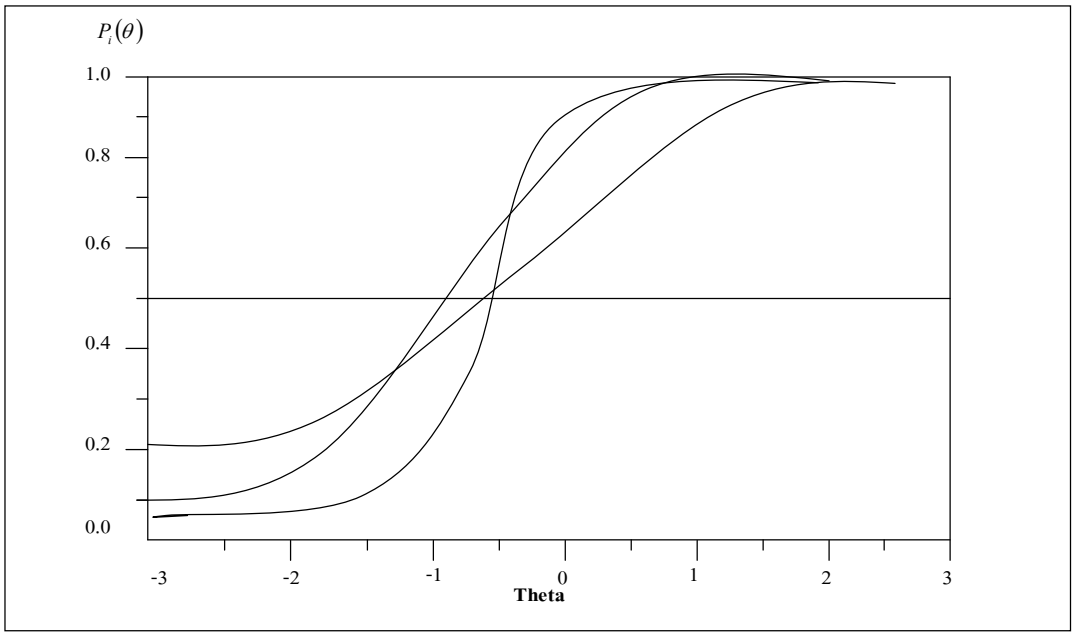
เมื่อ $P_i(\theta)$ แทน ความน่าจะเป็นซึ่งผู้ตอบที่มีระดับความสามารถ θ ตอบข้อ i ได้ถูกต้อง

b_i แทน ค่าพารามิเตอร์ความยากของข้อสอบข้อที่ I เป็นค่าที่แสดงตำแหน่งของโค้งลักษณะข้อสอบ ณ จุด θ ที่มีโอกาสตอบข้อสอบถูก $\frac{1+c_i}{2}$

a_i แทน ค่าพารามิเตอร์อำนาจจำแนกของข้อสอบข้อที่ i เป็นค่าความชันของโค้งลักษณะข้อสอบ ณ ตำแหน่ง b_i

e แทน ค่าคงที่ซึ่งมีค่าเท่ากับ 2.718

D แทน ค่าคงที่ซึ่งมีค่าเท่ากับ 1.70



ภาพที่ 2.8 โค้งลักษณะข้อสอบแบบ 3 พารามิเตอร์ ของข้อสอบ 3 ข้อ

3.2 การประมาณค่าพารามิเตอร์ของข้อสอบและค่าความสามารถ

ในการกำหนดคะแนนจุดตัดด้วยวิธีบุ๊คมาร์คที่ได้รับการปรับปรุง มีจุดเด่นอยู่ที่การจัดทำคู่มือเรียงข้อสอบ (Ordered Item Booklet: OIB) ซึ่งจัดทำโดยการเรียงข้อสอบจากข้อง่ายไปข้อยาก โดยใช้ทฤษฎีการตอบสนองข้อสอบ ช่วยในการประมาณค่าความยากของข้อสอบ ซึ่งการประมาณค่าพารามิเตอร์มักอยู่ในสถานการณ์ที่ไม่ทราบทั้งค่าความสามารถของผู้สอบ (θ) และค่าพารามิเตอร์ของข้อสอบ (β) ซึ่งในกรณีนี้จะต้องพิจารณาทั้งแบบแผนการตอบข้อสอบทุกข้อของผู้สอบทุกคน พร้อมกันไปในคราวเดียวกัน วิธีการประมาณค่าพารามิเตอร์ มีรายละเอียดดังนี้

3.2.1 วิธีการประมาณค่าที่เป็นไปได้สูงสุดแบบประมาณค่าร่วมกัน (Joint Maximum Likelihood Estimation Procedure)

การคำนวณเพื่อประมาณค่าด้วยวิธีการประมาณค่าที่เป็นไปได้สูงสุดแบบประมาณค่าร่วมกัน เริ่มต้นด้วยการกำหนดค่าเริ่มต้นของค่าความสามารถของผู้สอบ (θ) โดยใช้ค่า log ของ

อัตราส่วนจำนวนข้อที่ตอบถูก ต่อจำนวนข้อที่ตอบผิด สำหรับผู้ตอบแต่ละคน แปลงเป็นคะแนนมาตรฐานเพื่อใช้เป็นค่าเริ่มต้นของค่าความสามารถของผู้สอบ (θ) จากนั้นทำเสมือนทราบค่าความสามารถของผู้สอบ (θ) เพื่อประมาณค่าพารามิเตอร์ของข้อสอบ (β) และนำค่าค่าพารามิเตอร์ของข้อสอบ (β) ที่ประมาณได้มาประมาณค่าความสามารถของผู้สอบ (θ) กระทำซ้ำจนกระทั่งได้ค่าประมาณสองครั้งหลังที่ไม่เปลี่ยนแปลง โดยมีสูตรทางคณิตศาสตร์ในการคำนวณดังนี้

$$\text{Maximize } L(\theta;a,b,c)=\prod_{a=1}^{N}\prod_{i=1}^{n}[P_i(\theta_a)]^{u_{ia}}[Q_i(\theta_a)]^{1-u_{ia}}$$

$$\text{หรือ } \log L(\theta;a,b,c)=\sum_{a=1}^{N}\sum_{i=1}^{n}[u_{ai}\log P_i(\theta_a)+(1+u_{ai})\log Q_i(\theta_a)]$$

เมื่อ N แทน จำนวนผู้สอบ

$P_i(\theta_a)$ แทน ความน่าจะเป็นในการตอบข้อสอบได้ถูก ข้อที่ i ณ ระดับความสามารถ θ และ $Q_i(\theta_a)$= 1- $P_i(\theta_a)$

u_{ai} แทน คำตอบข้อที่ i (0,1)

ข้อจำกัดของวิธีการประมาณค่าที่เป็นไปได้สูงสุดแบบประมาณค่าร่วมกัน คือ 1) ไม่สามารถประมาณค่าความสามารถของผู้สอบ (θ) สำหรับผู้สอบที่ได้คะแนนเต็มหรือศูนย์ 2) ไม่สามารถประมาณค่าพารามิเตอร์ของข้อสอบที่ผู้สอบตอบถูกทุกคนหรือตอบผิดหมดทุกคน 3) ค่าพารามิเตอร์ที่ประมาณได้สำหรับโมเดล 2 และ 3 พารามิเตอร์ จะมีความคงเส้นคงวาก็ต่อเมื่อมีจำนวนข้อสอบมากและกลุ่มผู้สอบมีขนาดใหญ่ และการประมาณค่าพารามิเตอร์สำหรับโมเดล 3 พารามิเตอร์ อาจมีปัญหา ถ้าไม่จำกัดค่าความสามารถของผู้สอบ (θ) และค่าค่าพารามิเตอร์ของข้อสอบ (β) (Hembelton, et. al. 1991 อ้างถึงใน ศิริชัย กาจนวาสี, 2550)

3.2.2 วิธีการประมาณค่าเป็นไปได้สูงสุดแบบประมาณค่าปลายทาง (Marginal Maximum Likelihood Estimation Procedure)

วิธีการประมาณค่าเป็นไปได้สูงสุดแบบประมาณค่าปลายทาง จะประมาณค่าความสามารถของผู้สอบ (θ) และค่าพารามิเตอร์ของข้อสอบ (β) ไปพร้อมกัน ซึ่งค่าพารามิเตอร์ของข้อสอบจะไม่อิงกับค่าความสามารถของผู้สอบ และมีการประมาณการแจกแจงความสามารถของผู้สอบ เพื่อให้ได้ฟังก์ชันความน่าจะเป็นปลายทางของพารามิเตอร์ของข้อสอบ เมื่อประมาณค่าพารามิเตอร์ของข้อสอบได้แล้วก็ใช้ประมาณค่าความสามารถของผู้สอบต่อไป เพื่อให้การ

ประมาณค่าพารามิเตอร์ของข้อสอบมีค่าคงที่ ต้องมีกลุ่มผู้สอบขนาดใหญ่และจำนวนข้อสอบมาก เพื่อให้ประมาณค่าความสามารถของผู้สอบได้อย่างถูกต้อง เมื่อใช้กับโมเดล 3 พารามิเตอร์ ควรมีผู้สอบที่มีความสามารถต่ำเพียงพอ เพื่อให้การประมาณค่าโอกาสในการเดาข้อสอบ (c) น่าเชื่อถือ

ตอนที่ 4 การทดสอบทางการศึกษาแห่งชาติขั้นพื้นฐาน

4.1 ความเป็นมา

สถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) เป็นหน่วยงานทดสอบระดับชาติ โดยมีวัตถุประสงค์ เพื่อบริหารจัดการและดำเนินการเกี่ยวกับการศึกษา วิจัย พัฒนา และให้บริการทางการประเมินผลทางการศึกษาและทดสอบทางการศึกษาอย่างต่อเนื่อง รวมทั้งเป็นศูนย์กลางความร่วมมือด้านการทดสอบทางการศึกษาในระดับชาติ และระดับนานาชาติ การจัดสอบความรู้รวบยอดระดับชาติจัดเป็นภารกิจสำคัญที่สถาบันทดสอบทางการศึกษาแห่งชาติ เรียกว่า การทดสอบทางการศึกษาแห่งชาติขั้นพื้นฐาน หรือ O-NET เป็นการสอบความรู้รวบยอดปลายช่วงชั้น ได้แก่ ช่วงชั้นที่ 2 ช่วงชั้นที่ 3 และช่วงชั้นที่ 4 โดยดำเนินการจัดสอบให้แก่ผู้สอบใน 9 สังกัด โดยไม่คิดค่าใช้จ่าย ได้แก่ สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน กระทรวงศึกษาธิการ สำนักบริหารคณะกรรมการส่งเสริมการศึกษาเอกชน กระทรวงศึกษาธิการ สำนักงานคณะกรรมการการอุดมศึกษา (โรงเรียนสาธิตแห่งมหาวิทยาลัย) กระทรวง ศึกษาธิการ กรมส่งเสริมการปกครองส่วนท้องถิ่น กระทรวงมหาดไทย กรมพัฒนาสังคมและสวัสดิการ กระทรวงการพัฒนาสังคมและความมั่นคงของมนุษย์ สำนักการศึกษากรุงเทพมหานคร สำนักการศึกษาเมืองพัทยา สถาบันการพลศึกษา กระทรวงการท่องเที่ยวและกีฬา กองบัญชาการตำรวจตระเวนชายแดน สำนักงานตำรวจแห่งชาติ (สถาบันทดสอบทางการศึกษาแห่งชาติ, 2550)

4.2 วัตถุประสงค์

4.2.1 เพื่อประเมินคุณภาพการศึกษาระดับชาติตามหลักสูตรการศึกษาขั้นพื้นฐาน พ.ศ. 2544

4.2.2 เพื่อนำผลการสอบไปใช้ในการปรับปรุงคุณภาพการเรียนการสอนของโรงเรียน

4.2.3 เพื่อนำผลการทดสอบไปใช้ในวัตถุประสงค์อื่น

4.3 สาระและมาตรฐานการเรียนรู้การประเมิน

การทดสอบทางการศึกษาแห่งชาติขั้นพื้นฐาน (O-NET) จะประเมินผลสัมฤทธิ์ทางการเรียนของผู้เรียนตามจุดมุ่งหมายของหลักสูตรการศึกษาขั้นพื้นฐาน พุทธศักราช 2544 โดยมีสาระการประเมิน ดังนี้

ช่วงชั้นที่ 2 (ประถมศึกษาปีที่ 6) สอบสาระการเรียนรู้ภาษาไทย คณิตศาสตร์ และวิทยาศาสตร์

ช่วงชั้นที่ 3 (มัธยมศึกษาปีที่ 3) สอบสาระการเรียนรู้ภาษาไทย คณิตศาสตร์ วิทยาศาสตร์ ภาษาอังกฤษ และสังคมศึกษา ศาสนาและวัฒนธรรม

ช่วงชั้นที่ 4 (มัธยมศึกษาปีที่ 6) สอบสาระการเรียนรู้ภาษาไทย คณิตศาสตร์ วิทยาศาสตร์ ภาษาอังกฤษ สังคมศึกษาศาสนาและวัฒนธรรม สุขศึกษาและพลศึกษา ศิลปะ และการงานอาชีพ และเทคโนโลยี

4.4 สาระและมาตรฐานการเรียนรู้ ช่วงชั้นที่ 4 (มัธยมศึกษาปีที่ 6) ตามหลักสูตรการศึกษาขั้นพื้นฐาน พ.ศ. 2551 ดังนี้

วิชาคณิตศาสตร์

สาระที่ 1 จำนวนและการดำเนินการ

มาตรฐาน ค 1.1 เข้าใจถึงความหลากหลายของการแสดงจำนวนและการใช้จำนวนในชีวิตจริง

มาตรฐาน ค 1.2 เข้าใจถึงผลที่เกิดขึ้นจากการดำเนินการของจำนวนและความสัมพันธ์ระหว่างการดำเนินการต่าง ๆ และใช้การดำเนินการในการแก้ปัญหา

มาตรฐาน ค 1.3 ใช้การประมาณค่าในการคำนวณและแก้ปัญหา

มาตรฐาน ค 1.4 เข้าใจระบบจำนวนและนำสมบัติเกี่ยวกับจำนวนไปใช้

สาระที่ 2 การวัด

มาตรฐาน ค 2.1 เข้าใจพื้นฐานเกี่ยวกับการวัด วัดและคาดคะเนขนาดของสิ่งที่ต้องการวัด

มาตรฐาน ค 2.2 แก้ปัญหาเกี่ยวกับการวัด

สาระที่ 3 เรขาคณิต

มาตรฐาน ค 3.1 อธิบายและวิเคราะห์รูปเรขาคณิตสองมิติและสามมิติ

มาตรฐาน ค 3.2 ใช้การนึกภาพ (visualization) ใช้เหตุผลเกี่ยวกับปริภูมิ (spatial reasoning) และใช้แบบจำลองทางเรขาคณิต (geometric model) ในการแก้ปัญหา

สาระที่ 4 พีชคณิต

มาตรฐาน ค 4.1 เข้าใจและวิเคราะห์แบบรูป (pattern) ความสัมพันธ์ และฟังก์ชัน

มาตรฐาน ค 4.2 ใช้นิพจน์ สมการ อสมการ กราฟ และตัวแบบเชิงคณิตศาสตร์ (mathematical model) อื่น ๆ แทนสถานการณ์ต่าง ๆ ตลอดจนแปลความหมาย และนำไปใช้แก้ปัญหา

สาระที่ 5 การวิเคราะห์ข้อมูลและความน่าจะเป็น

มาตรฐาน ค 5.1 เข้าใจและใช้วิธีการทางสถิติในการวิเคราะห์ข้อมูล

มาตรฐาน ค 5.2 ใช้วิธีการทางสถิติและความรู้เกี่ยวกับความน่าจะเป็นในการคาดการณ์ได้อย่างสมเหตุสมผล

มาตรฐาน ค 5.3 ใช้ความรู้เกี่ยวกับสถิติและความน่าจะเป็นช่วยในการตัดสินใจและแก้ปัญหา

สาระที่ 6 ทักษะและกระบวนการทางคณิตศาสตร์

มาตรฐาน ค 6.1 มีความสามารถในการแก้ปัญหา การให้เหตุผล การสื่อสาร การสื่อความหมายทางคณิตศาสตร์ และการนำเสนอ การเชื่อมโยงความรู้ต่าง ๆ ทางคณิตศาสตร์และเชื่อมโยงคณิตศาสตร์กับศาสตร์อื่น ๆ และมีความคิดริเริ่มสร้างสรรค์

วิทยาศาสตร์

สาระที่ 1 สิ่งมีชีวิตกับกระบวนการดำรงชีวิต

มาตรฐาน ว 1.1 เข้าใจหน่วยพื้นฐานของสิ่งมีชีวิต ความสัมพันธ์ของโครงสร้าง และหน้าที่ของระบบต่างๆ ของสิ่งมีชีวิตที่ทำงานสัมพันธ์กัน มีกระบวนการสืบเสาะหาความรู้ สื่อสารสิ่งที่เรียนรู้และนำความรู้ไปใช้ในการดำรงชีวิตของตนเองและดูแลสิ่งมีชีวิต

มาตรฐาน ว 1.2 เข้าใจกระบวนการและความสำคัญของการถ่ายทอดลักษณะทางพันธุกรรมวิวัฒนาการของสิ่งมีชีวิต ความหลากหลายทางชีวภาพ การใช้เทคโนโลยีชีวภาพที่มีผลกระทบต่อมนุษย์และสิ่งแวดล้อม มีกระบวนการสืบเสาะหาความรู้และจิตวิทยาศาสตร์ สื่อสารสิ่งที่เรียนรู้ และนำความรู้ไปใช้ประโยชน์

สาระที่ 2 ชีวิตกับสิ่งแวดล้อม

มาตรฐาน ว 2.1 เข้าใจสิ่งแวดล้อมในท้องถิ่น ความสัมพันธ์ระหว่างสิ่งแวดล้อมกับสิ่งมีชีวิต ความสัมพันธ์ระหว่างสิ่งมีชีวิตต่าง ๆ ในระบบนิเวศ มีกระบวนการสืบเสาะหาความรู้และจิตวิทยาศาสตร์สื่อสารสิ่งที่เรียนรู้และนำความรู้ไปใช้ประโยชน์

มาตรฐาน ว 2.1 เข้าใจความสำคัญของทรัพยากรธรรมชาติ การใช้ทรัพยากรธรรมชาติในระดับท้องถิ่น ประเทศ และโลกนำความรู้ไปใช้ในในการจัดการทรัพยากรธรรมชาติและสิ่งแวดล้อมในท้องถิ่นอย่างยั่งยืน

สาระที่ 3 สารและสมบัติของสาร

มาตรฐาน ว 3.1 เข้าใจสมบัติของสาร ความสัมพันธ์ระหว่างสมบัติของสารกับโครงสร้างและแรงยึดเหนี่ยวระหว่างอนุภาค มีกระบวนการสืบเสาะหาความรู้และจิตวิทยาศาสตร์สื่อสารสิ่งที่เรียนรู้ นำความรู้ไปใช้ประโยชน์

มาตรฐาน ว 3.2 เข้าใจหลักการและธรรมชาติของการเปลี่ยนแปลงสถานะของสาร การเกิดสารละลาย การเกิดปฏิกิริยา มีกระบวนการสืบเสาะหาความรู้และจิตวิทยาศาสตร์ สื่อสารสิ่งที่เรียนรู้และนำความรู้ไปใช้ประโยชน์

สาระที่ 4 แรงและการเคลื่อนที่

มาตรฐาน ว 4.1 เข้าใจธรรมชาติของแรงแม่เหล็กไฟฟ้า แรงโน้มถ่วง และแรงนิวเคลียร์ มีกระบวนการสืบเสาะหาความรู้ สื่อสารสิ่งที่เรียนรู้และนำความรู้ไปใช้ประโยชน์อย่างถูกต้องและมีคุณธรรม

มาตรฐาน ว 4.2 เข้าใจลักษณะการเคลื่อนที่แบบต่างๆ ของวัตถุในธรรมชาติมีกระบวนการสืบเสาะหาความรู้และจิตวิทยาศาสตร์ สื่อสารสิ่งที่เรียนรู้และนำความรู้ไปใช้ประโยชน์

สาระที่ 5 พลังงาน

มาตรฐาน ว 5.2 เข้าใจความสัมพันธ์ระหว่างพลังงานกับการดำรงชีวิต การเปลี่ยนรูปพลังงาน ปฏิสัมพันธ์ระหว่างสารและพลังงาน ผลของการใช้พลังงานต่อชีวิตและสิ่งแวดล้อม มีกระบวนการสืบเสาะหาความรู้ สื่อสารสิ่งที่เรียนรู้และนำความรู้ไปใช้ประโยชน์

สาระที่ 6 กระบวนการเปลี่ยนแปลงของโลก

มาตรฐาน ว 6.1 เข้าใจกระบวนการต่าง ๆ ที่เกิดขึ้นบนผิวโลกและภายในโลก ความสัมพันธ์ของกระบวนการต่าง ๆ ที่มีผลต่อการเปลี่ยนแปลงภูมิอากาศ ภูมิประเทศ และสัณฐานของโลก มีกระบวนการสืบเสาะหาความรู้และจิตวิทยาศาสตร์ สื่อสารสิ่งที่เรียนรู้และนำความรู้ไปใช้ประโยชน์

สาระที่ 7 ดาราศาสตร์และอวกาศ

มาตรฐาน ว 7.1 เข้าใจวิวัฒนาการของระบบสุริยะ กาแล็กซีและเอกภพการปฏิสัมพันธ์ภายในระบบสุริยะและผลต่อสิ่งมีชีวิตบนโลก มีกระบวนการสืบเสาะหาความรู้และจิตวิทยาศาสตร์ การสื่อสารสิ่งที่เรียนรู้และนำความรู้ไปใช้ประโยชน์

มาตรฐาน ว 7.2 เข้าใจความสำคัญของเทคโนโลยีอวกาศที่นำมาใช้ในการสำรวจอวกาศและทรัพยากรธรรมชาติ ด้านการเกษตรและการสื่อสาร มีกระบวนการสืบเสาะ หาความรู้และจิตวิทยาศาสตร์ สื่อสารสิ่งที่เรียนรู้และนำความรู้ไปใช้ประโยชน์อย่างมีคุณธรรมต่อชีวิตและสิ่งแวดล้อม

สาระที่ 8 ธรรมชาติของวิทยาศาสตร์และเทคโนโลยี

มาตรฐาน ว 8.1 ใช้กระบวนการทางวิทยาศาสตร์และจิตวิทยาศาสตร์ในการสืบเสาะหาความรู้ การแก้ปัญหา รู้ว่าปรากฏการณ์ทางธรรมชาติที่เกิดขึ้นส่วนใหญ่มีรูปแบบที่แน่นอน สามารถอธิบายและตรวจสอบได้ ภายใต้ข้อมูลและเครื่องมือที่มีอยู่ในช่วงเวลานั้นๆ เข้าใจว่า วิทยาศาสตร์ เทคโนโลยี สังคม และสิ่งแวดล้อมมีความเกี่ยวข้องสัมพันธ์กัน

4.4 กลุ่มเป้าหมาย

กลุ่มเป้าหมาย คือ นักเรียนในช่วงชั้นที่ 2 (ประถมศึกษาปีที่ 6) ช่วงชั้นที่ 3 (มัธยมศึกษาปีที่ 3) และช่วงชั้นที่ 4 (มัธยมศึกษาปีที่ 6) ปีการศึกษา 2551 ทุกโรงเรียน ทุกคน ในแต่ละเขตพื้นที่การศึกษา

4.5 ระบบการพัฒนาข้อสอบ

4.5.1 คุณสมบัติของผู้ออกและกลั่นกรองข้อสอบ

4.5.1.1 มีความรู้ในเนื้อหาที่จะออกข้อสอบอย่างดี (Content specialist)

4.5.1.2 มีความสามารถในการออกข้อสอบ

4.5.1.3 เก็บความลับได้ ไม่กวดวิชา

4.5.1.4 มีเวลามาทำงานจนแล้วเสร็จ

4.5.2 ขั้นตอนการพัฒนาข้อสอบ

ขั้นตอนการพัฒนาข้อสอบ ประกอบด้วยการออกข้อสอบ การกลั่นกรองข้อสอบและการพัฒนาผู้ออกข้อสอบ

4.5.2.1 การออกข้อสอบ

4.5.2.1.1 การทำผังการออกข้อสอบ (Test Blueprint) โดยอิงมาตรฐานการเรียนรู้ตามหลักสูตรการศึกษาขั้นพื้นฐาน พ.ศ. 2544 (ปรับปรุง พ.ศ. 2551)

4.5.2.1.2 การสังเคราะห์เนื้อหาร่วมของแบบเรียนต่างๆ ที่กระทรวงศึกษาธิการให้ความเห็นชอบ

4.5.2.1.3 วางระดับความยาก ง่าย ดังนี้

ระดับความยาก ง่ายของข้อสอบ O-NET

ยาก	20%
ปานกลาง	60%
ง่าย	20%

4.5.2.1.4 กำหนดรูปแบบของข้อสอบมี 3 รูปแบบ คือ

- ปรนัย 4 ตัวเลือก
- มีบทความให้อ่าน และให้เลือกคำตอบ แบบ 4 ตัวเลือกหรือมากกว่า 4 ตัวเลือก
- ให้ระบุค่า/คะแนนที่ตัวเลือก (กรณีวิชาวิทยาศาสตร์ และคณิตศาสตร์)

4.5.2.1.5 ทำรายละเอียดของผังการออกข้อสอบเป็น 3 ระดับ คือ

ประเด็นหลัก

⬇

ประเด็นรอง

⬇

ประเด็นย่อย

4.5.2.1.6 ทำผังการออกข้อสอบ 2 ที่ละเอียดกว่าผังการออกข้อสอบ 1

4.5.2.1.7 เชิญผู้ออกข้อสอบที่มีคุณสมบัติตามข้อ 1 มาออกข้อสอบ

O-NET ป.6, ม.3 ใช้ครูโรงเรียนออกข้อสอบ วิชาละ 5–10 คน ถ้า O-NET ม.6 ใช้อาจารย์มหาวิทยาลัยเป็นผู้ออกข้อสอบ

4.5.2.1.8 จัดทำฉบับคู่ขนานตามจำนวนที่ต้องการ

4.5.2.1.9 ทำผังการออกข้อสอบ 3 ที่ระบุเลขที่ข้อตามระดับความยาก-ง่าย หรือความซับซ้อน

4.5.2.1 การกลั่นกรองข้อสอบ 2 ระยะ ดังนี้

4.5.2.1.1 ระยะตรวจความถูกต้องของภาษา ตัวสะกด ระยะห่างของข้อสอบ สถาบันทดสอบทางการศึกษาแห่งชาติมีผู้เชี่ยวชาญด้านภาษาไทยมาตรวจสอบให้

4.5.2.1.2 ระยะกลั่นกรองความถูกต้องของเนื้อหา สถาบันทดสอบทางการศึกษาแห่งชาติเชิญผู้เชี่ยวชาญภายนอกวิชาละ 3 -5 คน มาตรวจสอบความถูกต้องของเนื้อหา หลังตรวจสอบได้เชิญประธานของกลุ่มออกข้อสอบมาพิจารณาแก้ไข พร้อมเฉลยคำตอบด้วยการระบายในกระดาษคำตอบ และกำหนดเกณฑ์การให้คะแนนหลังจากกลั่นกรองทั้ง 2 ระยะแล้ว สถาบันทดสอบทางการศึกษาแห่งชาติจัดพิมพ์ต้นฉบับลงไข ตรวจสอบอีกครั้งก่อนส่งโรงพิมพ์

4.5.2.2 การพัฒนาผู้ออกข้อสอบ

เนื่องจากสถาบันทดสอบทางการศึกษาแห่งชาติเปิดเผยข้อสอบให้สาธารณชนทราบหลังสอบ ข้อสอบที่ใช้สอบแล้วจึงไม่สามารถนำมาเข้าคลังข้อสอบได้ สถาบันทดสอบทางการศึกษาแห่งชาติจึงได้จัดให้มีการพัฒนาผู้ออกข้อสอบแทนในการพัฒนาผู้ออกข้อสอบ สถาบันทดสอบทางการศึกษาแห่งชาติได้วิเคราะห์คะแนนและข้อสอบเพื่อให้เป็นข้อมูลแก่ผู้ออกข้อสอบดังนี้

4.5.2.2.1 ค่าเฉลี่ย มัธยฐาน ค่าสูงสุด ค่าต่ำสุด และค่าคลาดเคลื่อนมาตรฐาน ระดับโรงเรียน สังกัด และประเทศ จำแนกตามกลุ่มสาระการเรียนรู้ สาระย่อย และมาตรฐานการเรียนรู้

4.5.2.2.2 จำนวนผู้เข้าสอบจำแนกตามช่วงคะแนน

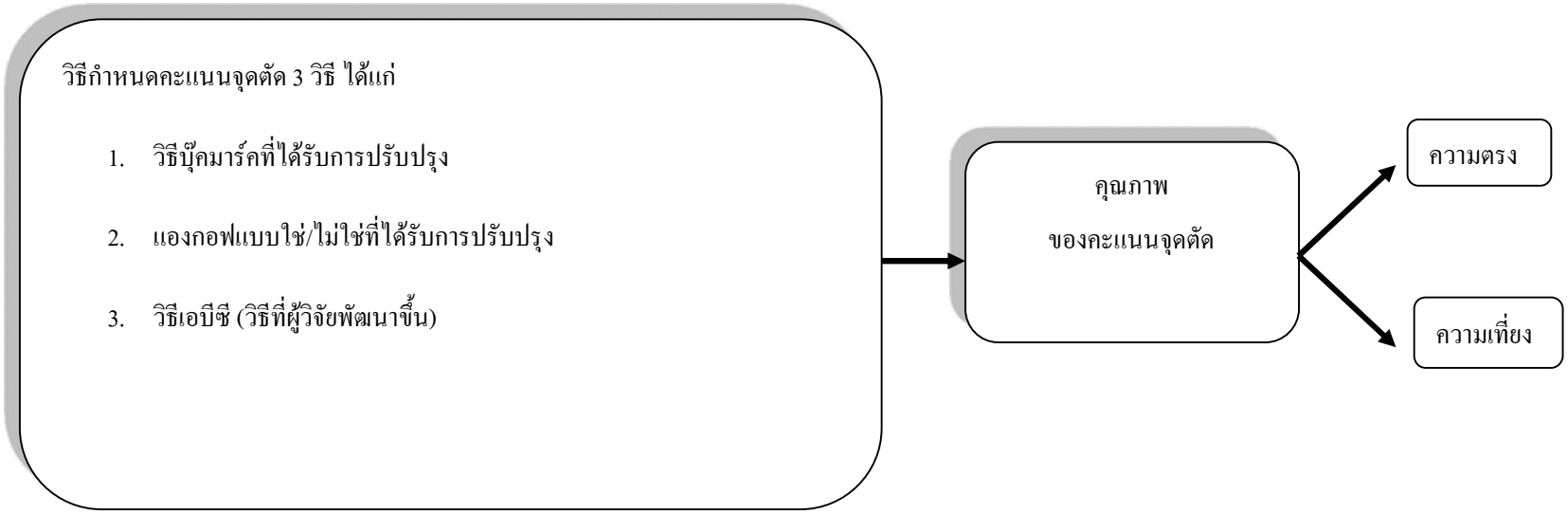
4.5.2.2.3 จำนวนร้อยละผู้ตอบถูกรายข้อ

4.5.2.2.4 ค่าความยาก อำนาจจำแนก และค่าการเดา จากการวิเคราะห์ข้อสอบ (Item Analysis) ด้วยทฤษฎี CTT และ IRT

4.5.2.2.5 ค่าความเที่ยง (Reliability) และความคลาดเคลื่อนในการวัด (Standard Error of Measurement: SEM) จากการวิเคราะห์แบบสอบ (Test Analysis) ในการพัฒนาผู้ออกข้อสอบใช้การสัมมนาโดยอธิบายผลวิเคราะห์ควบคู่กับข้อสอบเพื่อให้เห็นจุดอ่อน/ จุดแข็งรายข้อ เพื่อให้ผู้ออกข้อสอบนำไปปรับปรุง การออกข้อสอบในครั้งต่อไป

จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องกับวิธีการกำหนดคะแนนจุดตัด พบว่า วิธีการกำหนดคะแนนจุดตัดมีหลายวิธี โดยวิธีที่ได้รับการยอมรับและถูกใช้สำหรับการกำหนดคะแนนจุดตัดที่มีความสำคัญมากที่สุด คือ วิธีแองกอฟและวิธีบุ๊คมาร์ค ต่อมาวิธีแองกอฟได้ถูกพัฒนาให้ง่ายต่อการปฏิบัติมากขึ้นโดย Impara และ Plake (1997) ซึ่งใช้ชื่อว่า วิธีแองกอฟแบบใช่/ไม่ใช่ ดังนั้นในการกำหนดคะแนนจุดตัดสำหรับการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐานซึ่งถือว่ามีความสำคัญจึงควรกำหนดคะแนนจุดตัดด้วยวิธีแองกอฟแบบใช่/ไม่ใช่ และวิธีบุ๊คมาร์ค ทั้งนี้วิธีแองกอฟแบบใช่/ไม่ใช่ และวิธีบุ๊คมาร์คต่างก็มีข้อด้อยบางประการ โดยวิธีแองกอฟแบบใช่/ไม่ใช่นั้น ผู้เชี่ยวชาญจะต้องใช้กระบวนการคิดอย่างมากในการพิจารณาว่าผู้มีความสามารถคาบเส้นจะสามารถตอบข้อสอบได้ถูกต้องใช่หรือไม่ โดยเฉพาะในกรณีที่มีจำนวนข้อสอบที่ต้องพิจารณามาก ๆ และต้องกำหนดคะแนนจุดตัดหลายรอบ อาจทำให้ผู้เชี่ยวชาญเกิดความอ่อนล้าในขณะที่กำหนดคะแนนจุดตัดซึ่งอาจส่งผลให้คะแนนจุดตัดที่กำหนดเกิดความคลาดเคลื่อนได้ (Clauser, et al., 2009) ส่วนวิธีบุ๊คมาร์คมีข้อด้อย คืออาจเกิดกรณีที่ผู้เชี่ยวชาญรู้สึกว่าข้อสอบที่ถูกเรียงไว้ตามลำดับความยากนั้นบางข้อมีตำแหน่งที่ไม่เหมาะสม ซึ่งกรณีเช่นนี้อาจส่งผลต่อคุณภาพของคะแนนจุดตัดที่กำหนดได้ ดังนั้นผู้วิจัยจึงพัฒนาวิธีการกำหนดคะแนนจุดตัดขึ้นมาเรียกว่า วิธีเอบีซี ซึ่งเป็นวิธีที่ลดความซับซ้อนของกระบวนการคิดของผู้เชี่ยวชาญ และมีขั้นตอนในการดำเนินการที่ง่าย ดังนั้นในการกำหนดคะแนนจุดตัดสำหรับการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐานวิชาคณิตศาสตร์ และวิชาวิทยาศาสตร์ ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 จะกำหนดด้วยวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซี รวมทั้งตรวจสอบคุณภาพของคะแนนจุดตัดด้วย เนื่องจากเป็นเกณฑ์ที่ใช้สำหรับการตัดสินที่มีความสำคัญและมีผลกระทบสูง จึงควรนำเสนอสารสนเทศเกี่ยวกับคุณภาพของคะแนนจุดตัดเพื่อสร้างความเชื่อมั่นและการยอมรับต่อคะแนนจุดตัดสำหรับผู้ที่เกี่ยวข้อง

จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องทำให้ได้กรอบแนวคิดในการวิจัยดังนี้



ภาพที่ 2.9 กรอบแนวคิดในการวิจัย

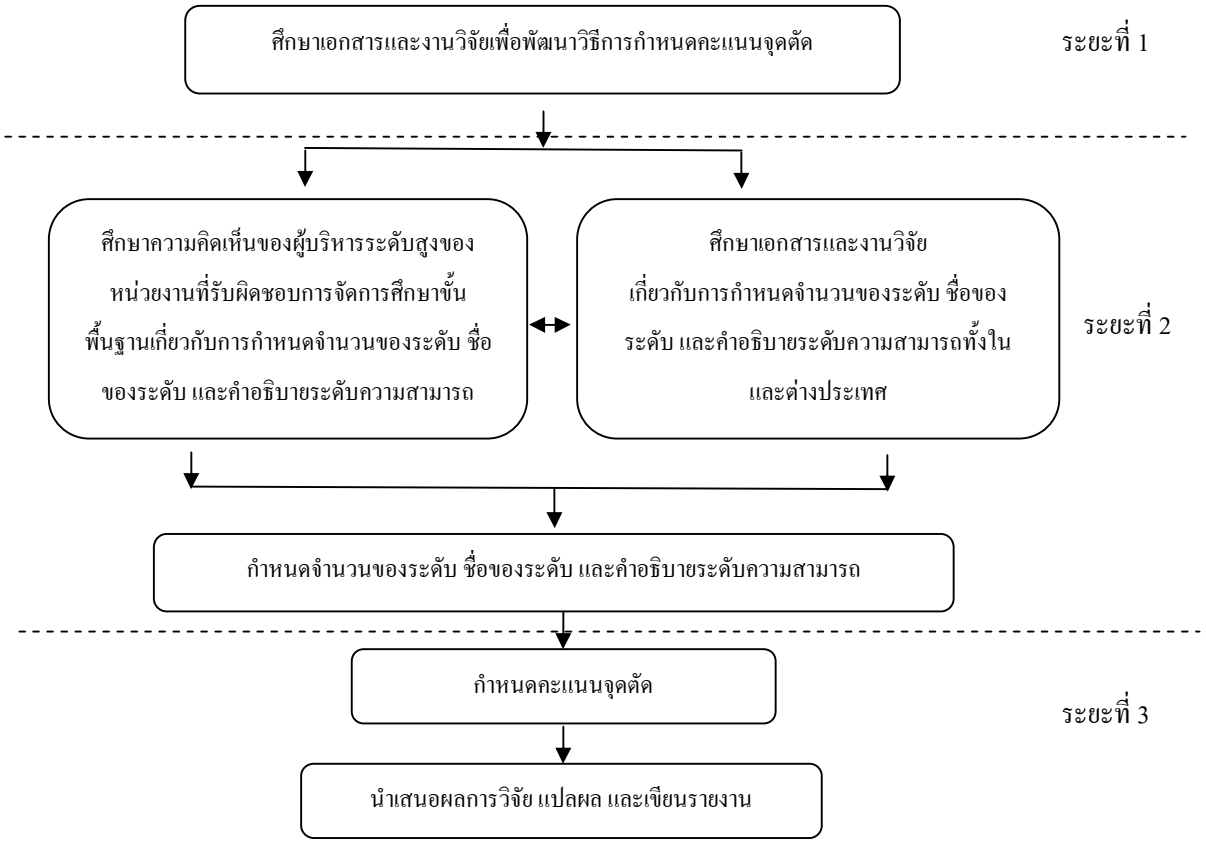
บทที่ 3

วิธีดำเนินการวิจัย

การวิจัยครั้งนี้ผู้วิจัยขอเสนอรายละเอียดการดำเนินการดังนี้

ขั้นตอนการดำเนินการวิจัย

ในการศึกษาครั้งนี้ ผู้วิจัยกำหนดขั้นตอนการดำเนินการวิจัยเป็น 3 ระยะ ได้แก่ ระยะที่ 1 การพัฒนาวิธีการกำหนดคะแนนจุดตัด ระยะที่ 2 การกำหนดจำนวนของระดับ ชื่อของระดับ และคำอธิบายระดับความสามารถ และระยะที่ 3 การกำหนดคะแนนจุดตัด ผู้วิจัยขอสรุปขั้นตอนการดำเนินการวิจัยดังภาพต่อไปนี้



ภาพที่ 3.1 ขั้นตอนการดำเนินการวิจัย

รายละเอียดในการดำเนินการแต่ละขั้นตอนเป็นดังนี้

ระยะที่ 1 การพัฒนาวิธีการกำหนดคะแนนจุดตัด

ในขั้นตอนการพัฒนาวิธีการกำหนดคะแนนจุดตัด ผู้วิจัยได้แบ่งการดำเนินการเป็น 2 ส่วน โดยส่วนแรก เป็นการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องกับการกำหนดคะแนนจุดตัดเพื่อศึกษาแนวคิด และวิธีการในการกำหนดคะแนนจุดตัด รวมทั้งพิจารณาถึงจุดดีและจุดด้อยของวิธีการต่างๆ เพื่อนำไปสู่การดำเนินการในส่วนที่สอง ซึ่งเป็นส่วนของการพัฒนาวิธีการกำหนดคะแนนจุดตัด โดยผู้วิจัยคาดหวังไว้ว่าจะพัฒนาวิธีการกำหนดคะแนนจุดตัดให้มีประสิทธิภาพ โดยมีแนวคิดตั้งต้นที่จะบูรณาการวิธีการกำหนดคะแนนจุดตัดหลาย ๆ วิธี โดยดึงเอาจุดเด่นและลดจุดด้อยของวิธีกำหนดคะแนนจุดตัดที่มีอยู่เดิม เพื่อให้ได้วิธีการกำหนดคะแนนจุดตัดที่มีประสิทธิภาพขึ้น

ระยะที่ 2 การกำหนดจำนวนของระดับ ชื่อของระดับ และคำอธิบายระดับความสามารถ

ผู้วิจัยได้แบ่งการดำเนินการเพื่อพัฒนาสารสนเทศเกี่ยวกับจำนวนของระดับ ชื่อของระดับ และคำอธิบายระดับความสามารถ เป็น 3 ส่วน โดยส่วนแรก คือการศึกษาเอกสารและงานวิจัยทั้งในและต่างประเทศเพื่อศึกษาแนวคิดและหลักการในการกำหนดจำนวนของระดับ ชื่อของระดับ และคำอธิบายระดับความสามารถ ส่วนที่สอง คือการสำรวจความคิดเห็น แนวคิด และข้อเสนอแนะในการกำหนดจำนวนของระดับ ชื่อของระดับ และคำอธิบายระดับความสามารถ จากผู้บริหารระดับสูงของหน่วยงานที่รับผิดชอบการจัดการศึกษาขั้นพื้นฐาน จำนวน 5 คน (รายชื่อดังแสดงในภาคผนวก ข) จากนั้นนำข้อมูลจากทั้งสองส่วนมาประกอบกันเพื่อดำเนินการในส่วนที่สาม คือการร่างจำนวนของระดับ ชื่อของระดับ และคำอธิบายระดับความสามารถสำหรับการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐานวิชาวิชาคณิตศาสตร์และวิทยาศาสตร์ ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 จากนั้นนำร่างจำนวนของระดับ ชื่อของระดับ และคำอธิบายระดับความสามารถไปให้ผู้เชี่ยวชาญพิจารณา ประกอบด้วย ครูผู้สอนวิชาคณิตศาสตร์และวิชาวิทยาศาสตร์ โรงเรียนอุตรดิตถ์ดรุณี จำนวน 7 ท่าน และอาจารย์ประจำสาขาคณิตศาสตร์คณะครุศาสตร์ มหาวิทยาลัยราชภัฏอุตรดิตถ์ จำนวน 2 ท่าน (รายนามดังแสดงในภาคผนวก ข) รวมถึงผู้เชี่ยวชาญที่เข้าร่วมการกำหนดคะแนนจุดตัด จำนวน 57 คน เพื่อขอข้อเสนอแนะในการปรับปรุงร่างจำนวนของระดับ ชื่อ

ของระดับ และคำอธิบายระดับความสามารถ ให้มีความเหมาะสมและสมบูรณ์ยิ่งขึ้น สำหรับนำไปกำหนดคะแนนจุดตัดในระยะที่สามต่อไป

ระยะที่ 3 การกำหนดคะแนนจุดตัด

ผู้วิจัยได้แบ่งการดำเนินการเป็น 2 ส่วน โดยส่วนแรกเป็นการศึกษานำร่อง (Pilot Study) ซึ่งเป็นการกำหนดคะแนนจุดตัดโดยนักศึกษาชั้นปีที่ 1, 2 และ 3 คณะศึกษาศาสตร์ มหาวิทยาลัยราชภัฏนครราชสีมา วิชาเอกการสอนคณิตศาสตร์ จำนวน 27 คน และวิชาเอกการสอนวิทยาศาสตร์ จำนวน 27 คนโดยการรับสมัครผู้ที่สนใจเข้าร่วมการกำหนดคะแนนจุดตัด และส่วนที่สองเป็นการกำหนดคะแนนจุดตัดโดยครูผู้สอนระดับมัธยมศึกษาตอนปลายวิชาคณิตศาสตร์ จำนวน 27 คน และวิชาฟิสิกส์ จำนวน 10 คน วิชาเคมี จำนวน 10 คน และวิชาชีววิทยา จำนวน 10 คน (รายชื่อดังแสดงในภาคผนวก ข) การศึกษาทั้งสองส่วนใช้วิธีกำหนดคะแนนจุดตัด 3 วิธี ได้แก่ วิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซี

ขอบเขตของการวิจัย

1. ประชากรและกลุ่มตัวอย่าง

ประชากรผู้เชี่ยวชาญเพื่อกำหนดคะแนนจุดตัด ได้แก่ ครูผู้สอนวิชาคณิตศาสตร์และวิชาฟิสิกส์ วิชาเคมี และวิชาชีววิทยา ในระดับมัธยมศึกษาตอนปลาย ซึ่งผู้วิจัยยึดหลักเกณฑ์ในการเลือกผู้เชี่ยวชาญคือต้องเป็นผู้ที่มีความเชี่ยวชาญในเนื้อหาที่จะทำการกำหนดคะแนนจุดตัดมาก (Impara, 1997) ดังนั้นผู้วิจัยจึงกำหนดให้ผู้เชี่ยวชาญต้องเป็นครูผู้สอนที่มีประสบการณ์การสอนไม่ต่ำกว่า 5 ปี

การวิจัยครั้งนี้ผู้วิจัยใช้วิธีการสุ่มอย่างง่ายเพื่อให้ได้กลุ่มตัวอย่างผู้เชี่ยวชาญ ซึ่งเป็นครูผู้สอนวิชาคณิตศาสตร์จำนวน 27 คน และเป็นครูผู้สอนวิชาฟิสิกส์ จำนวน 10 คน วิชาเคมี จำนวน 10 คน และวิชาชีววิทยา จำนวน 10 คน รวมทั้งสิ้นจำนวน 57 คน โดยผู้เชี่ยวชาญมีประสบการณ์การสอนโดยเฉลี่ย 18.71 ปี ผู้เชี่ยวชาญที่มีประสบการณ์การสอนสูงที่สุดมีประสบการณ์การสอน 37 ปี ส่วนผู้เชี่ยวชาญที่มีประสบการณ์การสอนน้อยที่สุดมีประสบการณ์การสอน 2 ปี โดยผู้เชี่ยวชาญที่มีประสบการณ์การสอนน้อยกว่า 5 ปี ซึ่งไม่เป็นไปตามเกณฑ์ที่ผู้วิจัยกำหนดไว้ มีจำนวน 10 คน คิดเป็นร้อยละ 17.5 ทั้งนี้เนื่องจากว่าครูที่มีประสบการณ์การสอน

มากกว่า 5 ปีที่ผู้วิจัยได้ติดต่อไว้ ติดภารกิจด่วนในวันที่จัดการกำหนดคะแนนจุดตัดจึงส่งตัวแทนของโรงเรียนมาร่วมกำหนดคะแนนจุดตัดแทน

จากนั้นจึงใช้วิธีการสุ่มอย่างง่ายเพื่อกำหนดกลุ่มให้ผู้เชี่ยวชาญ โดยครูผู้สอนวิชาคณิตศาสตร์แบ่งเป็น 3 กลุ่ม และแบ่งครูผู้สอนวิชาฟิสิกส์ วิชาเคมี และวิชาชีววิทยา เป็น 3 กลุ่ม จากนั้นจับสลากเพื่อกำหนดวิธีการกำหนดคะแนนจุดตัดให้แต่ละกลุ่ม ดังรายละเอียดในตารางต่อไปนี้

ตารางที่ 3.1 จำนวนกลุ่มตัวอย่างที่เข้าร่วมการกำหนดคะแนนจุดตัด

วิชา	ผู้เชี่ยวชาญ (จำนวนคน)
วิชาคณิตศาสตร์	
กลุ่มที่ 1 วิธีแองกอฟแบบใช่/ไม่ใช่	9
กลุ่มที่ 2 วิธีบุ๊คมาร์ค	9
กลุ่มที่ 3 วิธีเอบีซี	9
รวม	27
วิชาวิทยาศาสตร์	
กลุ่มที่ 1 วิธีแองกอฟแบบใช่/ไม่ใช่	
ฟิสิกส์	3
เคมี	3
ชีววิทยา	3
รวม	9
กลุ่มที่ 2 วิธีบุ๊คมาร์ค	
ฟิสิกส์	3
เคมี	3
ชีววิทยา	3
รวม	9
กลุ่มที่ 3 วิธีเอบีซี	
ฟิสิกส์	4
เคมี	4
ชีววิทยา	4
รวม	12
รวมทั้งสิ้น	57

2. ข้อสอบ

การวิจัยครั้งนี้ใช้ข้อสอบสำหรับการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน ช่วงชั้นที่ 4 (มัธยมศึกษาปีที่ 6) ปีการศึกษา 2553 วิชาคณิตศาสตร์และวิชาวิทยาศาสตร์ โดยข้อสอบวิชาคณิตศาสตร์มีจำนวน 40 ข้อ แบ่งเป็นข้อสอบแบบเลือกตอบ 4 ตัวเลือก จำนวน 20 ข้อ และข้อสอบแบบเติมคำตอบ จำนวน 20 ข้อ ประกอบด้วย 5 สาระการเรียนรู้ ส่วนวิชาวิทยาศาสตร์ มีจำนวน 90 ข้อ แบ่งเป็นข้อสอบแบบเลือกตอบที่มีคำตอบถูกเพียงข้อเดียว จำนวน 80 ข้อ และข้อสอบแบบเลือกตอบที่มีคำตอบถูก 2 ข้อ จำนวน 10 ข้อ ประกอบด้วย 8 สาระการเรียนรู้ รายละเอียดดังแสดงในตารางต่อไปนี้

ตารางที่ 3.2 รายละเอียดของข้อสอบสำหรับการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน ช่วงชั้นที่ 4 (มัธยมศึกษาปีที่ 6) ปีการศึกษา 2553 วิชาคณิตศาสตร์

สาระการเรียนรู้ที่สอบวัด	ข้อสอบ (จำนวน)	ค่าความยาก (b)	คะแนนเต็ม
สาระการเรียนรู้ที่ 1 จำนวนและการดำเนินการ	7	0.60 – 1.86	18
สาระการเรียนรู้ที่ 2 การวัด	4	0.74 – 1.44	11
สาระการเรียนรู้ที่ 3 เรขาคณิต	2	1.40 – 1.72	4
สาระการเรียนรู้ที่ 4 พีชคณิต	16	0.17 – 1.57	37
สาระการเรียนรู้ที่ 5 การวิเคราะห์ข้อมูลและความน่าจะเป็น	11	0.43 – 1.71	30
ทั้งฉบับ	**40**	0.17 – 1.86	100

ตารางที่ 3.3 รายละเอียดของข้อสอบสำหรับการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน ช่วงชั้นที่ 4 (มัธยมศึกษาปีที่ 6) ปีการศึกษา 2553 วิชาวิทยาศาสตร์

สาระการเรียนรู้ที่สอบวัด	ข้อสอบ (จำนวน)	ค่าความยาก (b)	คะแนนเต็ม
สาระการเรียนรู้ที่ 1 สิ่งมีชีวิตกับกระบวนการดำรงชีวิต	10	-0.71 – 5.5	11
สาระการเรียนรู้ที่ 2 ชีวิตกับสิ่งแวดล้อม	10	-1.98 – 0.84	11
สาระการเรียนรู้ที่ 3 สารและสมบัติของสาร	18	-0.57 – 1.66	18
สาระการเรียนรู้ที่ 4 แรงและการเคลื่อนที่	10	-1.69 – 1.15	11
สาระการเรียนรู้ที่ 5 พลังงาน	10	0.28 – 1.12	11
สาระการเรียนรู้ที่ 6 กระบวนการเปลี่ยนแปลงของโลก	11	-1.39 – 18.45	13
สาระการเรียนรู้ที่ 7 ดาราศาสตร์และอวกาศ	11	-1.94 – 1.41	13
สาระการเรียนรู้ที่ 8 ธรรมชาติของวิทยาศาสตร์และเทคโนโลยี	10	-3.91 – 2.88	12
ทั้งฉบับ	**90**	-3.91 – 18.45	100

3. วิธีที่ใช้ในการกำหนดคะแนนจุดตัด

วิธีที่ใช้ในการกำหนดคะแนนจุดตัดประกอบด้วยวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซี ซึ่งวิธีทั้งหมดเป็นวิธีการกำหนดคะแนนจุดตัดโดยผู้เชี่ยวชาญและเป็นวิธีแบบสอบเป็นศูนย์กลาง (test-center methods) กล่าวคือ ให้ผู้เชี่ยวชาญกำหนดคะแนนจุดตัดโดยพิจารณาข้อสอบเป็นหลัก

เครื่องมือที่ใช้ในการวิจัย

ในการศึกษาครั้งนี้มีเครื่องมือในการวิจัยจำนวน 5 ฉบับ ซึ่งผู้วิจัยได้สรุปเกี่ยวกับชื่อ วัตถุประสงค์ ลักษณะ และประเด็นหลักในการเก็บข้อมูล ดังแสดงในตารางต่อไปนี้

ตารางที่ 3.4 สรุปชื่อ วัตถุประสงค์ ลักษณะการวัด ประเด็นหลักในการเก็บข้อมูล

ชื่อเครื่องมือ	วัตถุประสงค์	ลักษณะ	ประเด็นหลัก	ระยะการเก็บข้อมูล
1. แบบสอบถามความคิดเห็นเกี่ยวกับจำนวนของระดับ ชื่อของระดับ และคำอธิบายระดับความสามารถ	เพื่อศึกษาความคิดเห็นเกี่ยวกับจำนวนของระดับ ชื่อของระดับ และคำอธิบายระดับความสามารถ	แบบสอบถามปลายเปิด	จำนวนของระดับความสามารถทางคณิตศาสตร์และวิทยาศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ที่เหมาะสมและเป็นไปได้ควรมีกี่ระดับ แต่ละระดับควรชื่อว่าอะไรและแต่ละระดับควรมีคำอธิบายแต่ละระดับความสามารถว่าอย่างไร	ระยะที่ 2
2. แบบสอบถามความคิดเห็นเกี่ยวกับร่างจำนวนของระดับ ชื่อของระดับ และคำอธิบายระดับความสามารถ	เพื่อขอข้อเสนอแนะในการปรับปรุงเกี่ยวกับจำนวนของระดับ ชื่อของระดับ และคำอธิบายระดับความสามารถ	แบบสอบถามปลายเปิด	ข้อเสนอแนะในการปรับปรุงเกี่ยวกับจำนวนของระดับ ชื่อของระดับ และคำอธิบายระดับความสามารถ	ระยะที่ 2
3. แบบสังเกตพฤติกรรม	เพื่อบันทึกผลการสังเกตพฤติกรรมของผู้เข้าร่วมกำหนดคะแนนจุดตัด	แบบบันทึกข้อมูลตามประเด็นที่กำหนด	บันทึกพฤติกรรม ในประเด็นหลักต่อไปนี้ได้แก่ ความสนใจ ความวิตกกังวล ปฏิสัมพันธ์ระหว่างผู้เชี่ยวชาญ ปัญหา อุปสรรคในการกำหนดคะแนนจุดตัด และอื่น ๆ	ระยะที่ 3
4. แบบบันทึกการกำหนดคะแนนจุดตัด	เพื่อรวบรวมข้อมูลผลการกำหนดคะแนนจุดตัด	แบบบันทึกคะแนนจุดตัด	บันทึกผลการกำหนดคะแนนจุดตัด	ระยะที่ 3

ตารางที่ 3.4 สรุปชื่อ วัตถุประสงค์ ลักษณะการวัด ประเด็นหลักในการเก็บข้อมูล

ชื่อเครื่องมือ	วัตถุประสงค์	ลักษณะ	ประเด็นหลัก	ระยะการเก็บข้อมูล
5. แบบสอบถามความคิดเห็นเกี่ยวกับความตรงของปัจจัยนำเข้า	เพื่อสอบถามความคิดเห็นของผู้เชี่ยวชาญเกี่ยวกับความตรงของปัจจัยนำเข้า	แบบสอบถามมาตรประเมินค่า 5 ระดับ	สอบถามเกี่ยวกับความตรงของคะแนนจุดตัดใน 5 ด้าน ได้แก่ ด้านความถูกต้อง ด้านความเป็นไปได้ ด้านความเหมาะสม ด้านความเข้าใจ และความมั่นใจของผู้เชี่ยวชาญ และด้านความเป็นประโยชน์ของสารสนเทศ	ระยะที่ 3

การสร้างและการตรวจสอบคุณภาพเครื่องมือ

ผู้วิจัยขอเสนอรายละเอียดในการสร้างและตรวจสอบคุณภาพของเครื่องมือที่ใช้ในการวิจัยดังนี้

1. แบบสอบถามความคิดเห็นเกี่ยวกับจำนวนของระดับ ชื่อของระดับ และคำอธิบายระดับความสามารถ และแบบสอบถามความคิดเห็นเกี่ยวกับร่างจำนวนของระดับ ชื่อของระดับ และคำอธิบายระดับความสามารถ

แบบสอบถามความคิดเห็นเกี่ยวกับจำนวนของระดับ ชื่อของระดับ และคำอธิบายระดับความสามารถ และแบบสอบถามความคิดเห็นเกี่ยวกับร่างจำนวนของระดับ ชื่อของระดับ และคำอธิบายระดับความสามารถ ผู้วิจัยได้แนวคิดจาก Perie (2008) ในการสร้างประเด็นคำถามเพื่อสอบถามความคิดเห็นของผู้เชี่ยวชาญเกี่ยวกับจำนวนของระดับ ชื่อของระดับ และคำอธิบายระดับความสามารถที่มีความเหมาะสมและมีความเป็นไปได้ โดยนิยามของความเหมาะสม หมายถึง ความสอดคล้องของจำนวนของระดับ ชื่อของระดับ และคำอธิบายระดับความสามารถทางคณิตศาสตร์และวิทยาศาสตร์กับวัตถุประสงค์ของการกำหนดคะแนนจุดตัด ซึ่งการกำหนดคะแนนจุดตัดครั้งนี้มีวัตถุประสงค์เพื่อใช้ในการรายงานผลการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน วิชาคณิตศาสตร์ และวิชาวิทยาศาสตร์ ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 และนิยามของความเป็นไปได้ หมายถึง การเป็นที่ยอมรับได้ของผู้ที่เกี่ยวข้อง ของจำนวนของระดับ ชื่อของระดับ และคำอธิบายระดับความสามารถทางคณิตศาสตร์และวิทยาศาสตร์ โดยประเด็นคำถามใน

แบบสอบถามความคิดเห็นเกี่ยวกับจำนวนของระดับ ชื่อของระดับ และคำอธิบายระดับความสามารถ ได้แก่ จำนวนของระดับ ชื่อของระดับ และคำอธิบายระดับความสามารถทางคณิตศาสตร์และวิทยาศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ที่เหมาะสมและเป็นไปได้ควรมีกี่ระดับ แต่ละระดับควรชื่อว่าอะไรและแต่ละระดับควรจะมีคำอธิบายระดับความสามารถว่าอย่างไร และประเด็นคำถามในแบบสอบถามความคิดเห็นเกี่ยวกับร่างจำนวนของระดับ ชื่อของระดับ และคำอธิบายระดับความสามารถ ได้แก่ จำนวนของระดับ ชื่อของระดับ และคำอธิบายระดับความสามารถทางคณิตศาสตร์และวิทยาศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ที่กำหนดขึ้นมีความเหมาะสมและเป็นไปได้หรือไม่ อย่างไร

2. แบบสังเกตพฤติกรรม

ผู้วิจัยศึกษาแนวคิดเกี่ยวกับการตรวจสอบความตรงของกระบวนการตัดสินใจกำหนดคะแนนจุดตัดของผู้เชี่ยวชาญของ McGinty (2005) ซึ่งใช้วิธีการสังเกตพฤติกรรมของผู้เชี่ยวชาญเพื่อศึกษาปัญหาหรือปัจจัยที่ส่งผลต่อคุณภาพในการตัดสินใจเพื่อกำหนดคะแนนจุดตัด โดยกำหนดประเด็นหลักในการสังเกต ได้แก่ ความสนใจ ความวิตกกังวล ปฏิสัมพันธ์ระหว่างผู้เชี่ยวชาญ ปัญหา อุปสรรคในการกำหนดคะแนนจุดตัด และอื่น ๆ

3. แบบบันทึกการกำหนดคะแนนจุดตัด

ผู้วิจัยศึกษาแนวคิดเกี่ยวกับการสร้างแบบบันทึกการกำหนดคะแนนจุดตัดของแต่ละวิธีจากแนวคิดของ Cizek และ Bunch (2007) และ Zieky และคณะ (2008)

4. แบบสอบถามความคิดเห็นเกี่ยวกับความตรงของปัจจัยนำเข้า

ผู้วิจัยได้สร้างและตรวจสอบคุณภาพเครื่องมือดังนี้

ขั้นตอนที่ 1 การกำหนดนิยามเชิงปฏิบัติการ

ผู้วิจัยศึกษาเอกสารและงานวิจัยที่เกี่ยวข้อง และกำหนดนิยามเชิงปฏิบัติของประเด็นที่จะประเมินปัจจัยนำเข้า โดยมีรายละเอียดของนิยามเชิงปฏิบัติการแต่และประเด็นดังนี้

ปัจจัยนำเข้า หมายถึง องค์ประกอบของการดำเนินการกำหนดคะแนนจุดตัด ประกอบด้วย เป้าหมายของการกำหนดคะแนนจุดตัด คำอธิบายระดับความสามารถ ผู้เชี่ยวชาญ วิธีการกำหนดคะแนนจุดตัด และสารสนเทศที่ใช้ประกอบการกำหนดคะแนนจุดตัด

ความตรงของปัจจัยนำเข้า หมายถึง คุณภาพของเป้าหมาย คำอธิบายระดับความสามารถ ผู้เชี่ยวชาญ วิธีการกำหนดคะแนนจุดตัด และสารสนเทศที่ใช้ประกอบการกำหนดคะแนนจุดตัด โดยพิจารณา 5 ด้าน ได้แก่ ความถูกต้อง ความเป็นไปได้ ความเหมาะสม ความเข้าใจและความมั่นใจของผู้เชี่ยวชาญ และความเป็นประโยชน์ของเอกสาร

ความถูกต้อง หมายถึง การดำเนินการมีลำดับขั้นตอนที่ถูกต้อง และสารสนเทศประกอบการดำเนินการมีเนื้อหาที่ถูกต้อง สื่อความหมายชัดเจน

ความเป็นไปได้ หมายถึง วิธีการดำเนินการสามารถปฏิบัติได้จริงและง่ายต่อการดำเนินการ

ความเหมาะสม หมายถึง เป้าหมาย การดำเนินการ และเอกสารประกอบการดำเนินการมีความสอดคล้องและเชื่อมโยงกัน

ความเข้าใจและความมั่นใจของผู้เชี่ยวชาญ หมายถึง ผู้เชี่ยวชาญมีความเข้าใจเกี่ยวกับความสำคัญ เป้าหมาย คำอธิบายระดับความสามารถ การดำเนินการ และสารสนเทศประกอบการดำเนินการ รวมถึงมีความมั่นใจต่อคะแนนจุดตัดที่กำหนด

ความเป็นประโยชน์ของเอกสารประกอบการกำหนดคะแนนจุดตัด หมายถึง เอกสารประกอบการดำเนินการเป็นประโยชน์ต่อการกำหนดคะแนนจุดตัดของผู้เชี่ยวชาญ

ขั้นตอนที่ 2 การสร้างแบบสอบถามความคิดเห็นเกี่ยวกับความตรงของปัจจัยนำเข้า

ผู้วิจัยแปลแบบสอบถามของ Skaggs และคณะ (2007) Cizek และคณะ (2004) และ Skorupski และ Hambleton (2005) จากนั้นจัดข้อคำถามไว้ในความตรงแต่ละด้าน นอกจากนี้ผู้วิจัยได้สร้างข้อคำถามขึ้นเพิ่มเติมในแต่ละด้านด้วย จากนั้นเสนอให้อาจารย์ที่ปรึกษาวิทยานิพนธ์พิจารณาความเหมาะสม และนำไปให้ผู้ทรงคุณวุฒิด้านการวัดและประเมินผลการศึกษา จำนวน 3 ท่าน (รายนามดังแสดงในภาคผนวก ข) ประเมินความตรงตามเนื้อหาของเครื่องมือ และวิเคราะห์ค่า IOC ซึ่งผลจากการวิเคราะห์ พบว่า ทุกข้อมีค่า IOC เป็น 1

ตัวอย่างแบบประเมินความตรงตามเนื้อหาของแบบสอบถามความคิดเห็นเกี่ยวกับความตรงของปัจจัยนำเข้า

<u>คำชี้แจง</u> แบบสอบถามความคิดเห็นเกี่ยวกับความตรงของปัจจัยนำเข้า เป็นแบบสอบถามเพื่อรวบรวมความคิดเห็นของผู้เชี่ยวชาญที่เข้าร่วมการกำหนดคะแนนจุดตัด ใน 5 ด้าน ได้แก่ ความถูกต้อง ความเป็นไปได้ ความเหมาะสม ความเข้าใจและความมั่นใจของผู้เชี่ยวชาญ และความเป็นประโยชน์ของเอกสารประกอบการกำหนดคะแนนจุดตัด

ขอความอนุเคราะห์ท่านพิจารณาข้อความแต่ละข้อ ว่าท่านมีความเห็นอยู่ในระดับใด แล้วทำเครื่องหมาย ✓ในช่องที่ตรงกับความคิดเห็นของท่าน โดยมีเกณฑ์ในการพิจารณาดังนี้

1 หมายถึง ผู้เชี่ยวชาญเห็นด้วยว่าเนื้อหาของข้อคำถามสอดคล้องกับนิยามเชิงปฏิบัติการ

0 หมายถึง ผู้เชี่ยวชาญไม่แน่ใจว่าเนื้อหาของข้อคำถามสอดคล้องกับนิยามเชิงปฏิบัติการ

-1 หมายถึง ผู้เชี่ยวชาญไม่เห็นด้วยว่าเนื้อหาของข้อคำถามสอดคล้องกับนิยามเชิงปฏิบัติการ

ในกรณีให้คะแนน -1 (ไม่เห็นด้วย) ขอความอนุเคราะห์ให้ข้อคิดเห็นเพิ่มเติมในข้อนั้น

นิยาม	รายการ	ผลการพิจารณา			ข้อเสนอแนะ
		1	0	-1	
1. ด้านความถูกต้อง หมายถึง การดำเนินการมีลำดับขั้นตอนที่ถูกต้อง และสารสนเทศประกอบการดำเนินการมีเนื้อหาที่ถูกต้องสื่อความหมายชัดเจน	1. การกำหนดคะแนนจุดครั้งนี้มีวิธีดำเนินการที่ถูกต้อง				
	2. คำอธิบายระดับความสามารถมีเนื้อหาที่ถูกต้องสื่อความหมาย (The performance level descriptions were clear and useful)				
2. ด้านความเป็นไปได้ หมายถึง วิธีการดำเนินการสามารถปฏิบัติได้จริงและง่ายต่อการดำเนินการ	3. ขั้นตอนการกำหนดคะแนนจุดตัดสามารถปฏิบัติได้จริง				
	4. ระยะเวลาที่ให้สำหรับการกำหนดคะแนนจุดตัดเพียงพอ (The amount of time I had to complete the tasks was sufficient)				

ขั้นตอนที่ 3 การจัดทำแบบสอบถามความคิดเห็นเกี่ยวกับความตรงของปัจจัยนำเข้า

ผู้วิจัยได้จัดทำแบบสอบถามมาตราส่วนประมาณค่า 5 ระดับ โดยคัดเลือกข้อคำถามที่มีค่า IOC ตั้งแต่ 0.50 ขึ้นไป และปรับปรุงข้อคำถามตามข้อเสนอแนะของผู้เชี่ยวชาญ

ขั้นตอนที่ 4 การนำเครื่องมือไปทดลองใช้กับการศึกษานำร่อง

ผู้วิจัยได้นำแบบสอบถามไปทดลองใช้กับการศึกษานำร่องซึ่งเป็นการกำหนดคะแนนจุดตัดโดยนักศึกษาชั้นปีที่ 1, 2 และ 3 คณะศึกษาศาสตร์ มหาวิทยาลัยราชภัฏนครราชสีมา วิชาเอกการสอนคณิตศาสตร์ จำนวน 27 คน วันที่ 6 พฤศจิกายน 2554 และวิชาเอกการสอนวิทยาศาสตร์ จำนวน 27 คน วันที่ 7 พฤศจิกายน 2554 เพื่อทดสอบความเป็นปรนัยและความชัดเจนของข้อคำถาม รวมถึงวิเคราะห์ความเที่ยงของแบบสอบถามมีค่า 0.84

ขั้นตอนที่ 5 จัดทำแบบสอบถามฉบับสมบูรณ์

ผู้วิจัยได้ปรับแก้ข้อคำถามที่ยังไม่ชัดเจนและจัดทำแบบสอบถามฉบับสมบูรณ์

การดำเนินการและการเก็บรวบรวมข้อมูล

ในการศึกษาครั้งนี้ ผู้วิจัยกำหนดขั้นตอนการดำเนินการและเก็บรวบรวมข้อมูลไว้ 3 ระยะ ดังนี้

ระยะที่ 1 การพัฒนาวิธีการกำหนดคะแนนจุดตัด ผู้วิจัยได้ดำเนินการและเก็บรวบรวมข้อมูลดังนี้

ผู้วิจัยศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องกับการกำหนดคะแนนจุดตัดเพื่อศึกษาแนวคิด และวิธีการในการกำหนดคะแนนจุดตัด รวมทั้งพิจารณาถึงจุดเด่นและจุดด้อยของวิธีการต่างๆ โดยผู้วิจัยใช้วิธีการสืบค้นข้อมูลจากอินเตอร์เน็ตเป็นหลัก นอกจากนี้ได้สอบถามความรู้จากผู้เชี่ยวชาญในต่างประเทศผ่านทาง E-mail เมื่อผู้วิจัยได้ศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องเพื่อทำความเข้าใจเกี่ยวกับเนื้อหาที่เกี่ยวกับแนวคิด และวิธีการในการกำหนดคะแนนจุดตัด รวมถึงจุดเด่นจุดด้อยของแต่ละวิธีแล้ว ผู้วิจัยจึงได้พัฒนาวิธีการกำหนดคะแนนจุดตัดขึ้นเรียกว่า “วิธีเอบีซี”

ระยะที่ 2 การกำหนดจำนวนของระดับ ชื่อของระดับ และคำอธิบายระดับความสามารถ ผู้วิจัยได้ดำเนินการและเก็บรวบรวมข้อมูลดังนี้

การกำหนดจำนวนของระดับ ชื่อของระดับ และคำอธิบายระดับความสามารถ ผู้วิจัยได้ดำเนินการดังนี้

1) วิเคราะห์เนื้อหาจากเอกสารและงานวิจัยเพื่อศึกษาแนวคิดและหลักการในการกำหนดจำนวนของระดับ ชื่อของระดับ และคำอธิบายระดับความสามารถ

2) ศึกษาความคิดเห็นของผู้บริหารระดับสูงในหน่วยงานที่รับผิดชอบการจัดการศึกษาขั้นพื้นฐานจำนวน 5 ท่าน (รายนามดังแสดงในภาคผนวก ข) เกี่ยวกับจำนวนของระดับ ชื่อของระดับ และคำอธิบายระดับความสามารถ เพื่อใช้ในการกำหนดคะแนนจุดตัดสำหรับการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐานวิชาคณิตศาสตร์ และวิชาวิทยาศาสตร์ ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 โดยแต่เดิมได้กำหนดวิธีการเก็บรวบรวมข้อมูลโดยใช้วิธีการสัมภาษณ์แบบมีโครงสร้าง แต่เนื่องด้วยข้อจำกัดในเรื่องเวลาของผู้บริหารระดับสูงจึงเปลี่ยนวิธีการเก็บข้อมูลเป็นการใช้แบบสอบถามความคิดเห็นซึ่งเป็นแบบสอบถามแบบปลายเปิด โดยมีประเด็นคำถามเกี่ยวกับจำนวนของระดับ ชื่อของระดับ และคำอธิบายระดับความสามารถทางคณิตศาสตร์และวิทยาศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ที่เหมาะสมและเป็นไปได้ควรมีกี่ระดับ แต่ละระดับควรชื่อว่าอะไรและแต่ละระดับควรจะมีคำอธิบายระดับความสามารถว่าอย่างไร นอกจากนี้ผู้วิจัยได้สรุปตัวอย่างจำนวนของระดับ ชื่อของระดับ และคำอธิบายระดับความสามารถที่กำหนดในต่างประเทศ เพื่อใช้เป็นสารสนเทศประกอบการตอบแบบสอบถามของผู้บริหารระดับสูง ซึ่งรายละเอียดคำชี้แจง คำจำกัดความ และประเด็นคำถามของแบบสอบถามแบบปลายเปิด รวมถึงสารสนเทศเกี่ยวกับตัวอย่างจำนวนของระดับ ชื่อของระดับ และคำอธิบายระดับความสามารถของต่างประเทศ ดังแสดงในภาคผนวก ค ในการเก็บรวบรวมข้อมูล ผู้วิจัยนำแบบแบบสอบถามไปให้ผู้ทรงคุณวุฒิด้วยตนเอง และนัดหมายวันเวลาในการขอรับแบบสอบถามกลับคืน การรับเอกสารคืนมีทั้งผู้วิจัยไปรับด้วยตัวเองและผู้เชี่ยวชาญส่งข้อมูลให้ผู้วิจัยทาง E-mail

3) ผู้วิจัยสรุปความคิดเห็นจากผู้บริหารระดับสูงในหน่วยงานที่รับผิดชอบการจัดการศึกษาขั้นพื้นฐานทุกท่านเกี่ยวกับจำนวนของระดับ ชื่อของระดับ และคำอธิบายระดับความสามารถ ที่เหมาะสมและเป็นไปได้ จากนั้นจึงร่างจำนวนของระดับ ชื่อของระดับ และคำอธิบายระดับความสามารถ

4) ผู้วิจัยนำร่างจำนวนของระดับ ชื่อของระดับ และคำอธิบายระดับความสามารถวิชาคณิตศาสตร์ และวิชาวิทยาศาสตร์ไปให้ผู้เชี่ยวชาญพิจารณา โดยผู้เชี่ยวชาญเป็นครูผู้สอนที่มีประสบการณ์การสอนไม่น้อยกว่า 5 ปี และมีตำแหน่งวิทยฐานะไม่ต่ำกว่าครูชำนาญการได้แก่ ครูผู้สอนโรงเรียนอุตรดิตถ์ดรุณี จำนวน 7 คน และอาจารย์ประจำสาขาคณิตศาสตร์ คณะครุศาสตร์ มหาวิทยาลัยราชภัฏอุตรดิตถ์ จำนวน 2 คน (รายนามดังแสดงในภาคผนวก ข)

รวมถึงผู้เชี่ยวชาญที่เข้าร่วมการกำหนดคะแนนจุดตัด จำนวน 57 คน เพื่อขอข้อเสนอแนะในการปรับปรุงร่างจำนวนของระดับ ชื่อของระดับ และคำอธิบายระดับความสามารถให้มีความสมบูรณ์ยิ่งขึ้น

ระยะที่ 3 การกำหนดคะแนนจุดตัด ผู้วิจัยได้ดำเนินการดังนี้

1) ผู้วิจัยติดต่อนัดหมายวัน เวลา และสถานที่สำหรับการเข้าร่วมเพื่อกำหนดคะแนนจุดตัดกับผู้เชี่ยวชาญ และทำจดหมายถึงสำนักงานเขตพื้นที่การศึกษามัธยมศึกษา เขต 31 จังหวัดนครราชสีมา เพื่อขออนุญาตให้ครูเข้าร่วมการกำหนดคะแนนจุดตัดครั้งนี้ จากนั้นผู้วิจัยได้ส่งคู่มือการกำหนดคะแนนจุดตัดไปให้ครูที่เข้าร่วมการกำหนดคะแนนจุดตัดทางไปรษณีย์ และทาง E-mail

2) ผู้วิจัยจัดการกำหนดคะแนนจุดตัดตามวัน เวลา และสถานที่ที่ได้นัดหมายไว้ โดยดำเนินการตามกำหนดการที่กำหนดไว้ (ดังแสดงในภาคผนวก ง) ซึ่งมีรายละเอียดหลัก ๆ ดังนี้

ขั้นที่ 1 ผู้วิจัยอธิบายเกี่ยวกับเป้าหมายและวิธีการกำหนดคะแนนจุดตัด

ขั้นที่ 2 ผู้เชี่ยวชาญสุ่มเลือกวิธีการกำหนดคะแนนจุดตัดโดยการจับฉลาก

ขั้นที่ 3 ผู้เชี่ยวชาญฝึกปฏิบัติตามวิธีที่จับฉลากได้ ในระหว่างที่ผู้เชี่ยวชาญฝึกปฏิบัติ ผู้วิจัยรวมถึงนักศึกษาช่วยงานจะทำการสังเกตพฤติกรรมของผู้เชี่ยวชาญ รวมถึงบันทึกข้อมูลเกี่ยวกับประเด็นคำถามที่ถูกถามบ่อย ๆ เพื่อเป็นข้อมูลประกอบการวางแผนกำหนดคะแนนจุดตัด และเป็นข้อมูลสำหรับการตรวจสอบความตรงของกระบวนการตัดสินใจเพื่อกำหนดคะแนนจุดตัดของผู้เชี่ยวชาญ

ขั้นที่ 4 ฝึกอบรมเกี่ยวกับการคิดออกเสียงแก่ผู้เชี่ยวชาญ เนื่องจากในการวิจัยครั้งนี้ ผู้วิจัยต้องการตรวจสอบความตรงของกระบวนการตัดสินใจเพื่อกำหนดคะแนนจุดตัดของผู้เชี่ยวชาญ ซึ่งเป็นการศึกษาเกี่ยวกับกระบวนการคิดที่เกิดขึ้นภายในจิตใจของผู้เชี่ยวชาญขณะกำหนดคะแนนจุดตัดเกี่ยวกับเหตุผล หลักการที่ผู้เชี่ยวชาญยึดถือในการกำหนดคะแนนจุดตัด รวมถึงความคิด ความกังวลขณะตัดสินใจ และปัญหาที่เกิดขึ้น ซึ่งผู้วิจัยได้เก็บข้อมูลโดยใช้เทคนิคการคิดออกเสียง (Think-aloud method) ดังนั้นในระหว่างที่ผู้เชี่ยวชาญกำหนดคะแนนจุดตัด ผู้เชี่ยวชาญจะต้องพูดในสิ่งที่กำลังคิดอยู่ ซึ่งโดยหลักการแล้วเทคนิคการคิดออกเสียงนั้นจะต้องให้ผู้

ที่จะคิดออกเสียงพูดในสิ่งที่กำลังคิดดัง ๆ คนเดียว แต่จากการศึกษานำร่องกับนักศึกษาชั้นปีที่ 1, 2 และ 3 คณะศึกษาศาสตร์ มหาวิทยาลัยราชภัฏนครราชสีมา เมื่อวันที่ 6 พฤศจิกายน 2554 และวันที่ 7 พฤศจิกายน 2554 พบว่า นักศึกษาที่กำหนดคะแนนจุดตัดจะไม่ค่อยพูดในสิ่งที่กำลังคิด เนื่องจากเขินอายที่จะพูดกับเครื่องบันทึกเสียงคนเดียว อาจเป็นเพราะรู้สึกเหมือนพูดคนเดียวและบางครั้งก็ไม่มีประเด็นที่จะพูด ดังนั้นในกำหนดคะแนนจุดตัดหากผู้เชี่ยวชาญนิ่งเงียบไปหรือไม่ค่อยพูด นักศึกษาช่วยงานก็จะคอยซักถามเป็นระยะ ๆ ในประเด็นเกี่ยวกับเหตุผล หลักการ ความคิด ความกังวลขณะตัดสินใจ และปัญหาที่เกิดขึ้น

ขั้นที่ 5 ผู้เชี่ยวชาญกำหนดคะแนนจุดตัด ในระหว่างที่ผู้เชี่ยวชาญกำหนดคะแนนจุดตัดนั้น ผู้เชี่ยวชาญก็จะพูดความคิดของตนเอง หากผู้เชี่ยวชาญนิ่งเงียบไปหรือไม่ค่อยพูดนักศึกษาช่วยงานก็ถามผู้เชี่ยวชาญในประเด็นเหตุผล หลักการ ความคิด ความกังวลขณะตัดสินใจ และปัญหาที่เกิดขึ้น และทำการบันทึกเสียงไว้ นอกจากนี้ผู้วิจัยก็ได้ให้นักศึกษาช่วยงานจดบันทึกคำพูดของผู้เชี่ยวชาญในประเด็นที่กำหนดไว้ ทั้งนี้เนื่องจากว่าในการศึกษานำร่องกับนักศึกษาชั้นปีที่ 1, 2 และ 3 คณะศึกษาศาสตร์ มหาวิทยาลัยราชภัฏนครราชสีมา พบปัญหาเกี่ยวกับการถอดเทปความคิดของนักศึกษาขณะกำหนดคะแนนจุดตัด ปัญหาที่พบคือ ในกรณีที่ข้อสอบมีจำนวนมาก ๆ เช่น ข้อสอบวิชาวิทยาศาสตร์ คำพูดที่บันทึกไว้มีความยาวมาก บางครั้งเป็นคำพูดเกี่ยวกับการวิเคราะห์ข้อสอบที่เป็นแนวทางการวิเคราะห์ที่ซ้ำ ๆ ซึ่งทำให้การถอดเทปใช้เวลานานมากและยากต่อการจับประเด็นสำคัญ ๆ ผู้วิจัยจึงให้นักศึกษาช่วยงานจดประเด็นหลัก ๆไว้ด้วย เพื่อช่วยในการจับประเด็นในการดำเนินการถอดเทป จากนี้ผู้วิจัยได้ให้นักศึกษาช่วยงานสังเกตและบันทึกพฤติกรรมของผู้เชี่ยวชาญในขณะที่กำหนดคะแนนจุดตัดด้วย หลังจากที่ผู้เชี่ยวชาญกำหนดคะแนนจุดตัดรอบที่ 1 เรียบร้อยแล้ว จากนั้นผู้เชี่ยวชาญจะพักรับประทานอาหารกลางวันร่วมกัน ในระหว่างนี้นักศึกษาช่วยงานก็จะจดบันทึกประเด็นสนทนาที่เกี่ยวกับการกำหนดคะแนนจุดตัดของผู้เชี่ยวชาญ หลังจากรับประทานอาหารกลางวันเรียบร้อยแล้ว ผู้เชี่ยวชาญก็ร่วมกันอภิปรายกลุ่มย่อยเกี่ยวกับคะแนนจุดตัดที่กำหนดในรอบที่ 1 และกำหนดคะแนนจุดตัดรอบที่ 2 จากนั้นผู้เชี่ยวชาญร่วมกันอภิปรายกลุ่มใหญ่เกี่ยวกับคะแนนจุดตัดที่กำหนดในรอบที่ 2 และกำหนดคะแนนจุดตัดอีกครั้ง ซึ่งนักศึกษาช่วยงานจะจดบันทึกประเด็นที่อภิปรายและพฤติกรรมของผู้เชี่ยวชาญทุกรอบไว้ หลังจากเสร็จสิ้นการกำหนดคะแนนจุดตัดแล้ว ผู้เชี่ยวชาญทำแบบสอบถามความคิดเห็นเกี่ยวกับความตรงของปัจจัยนำเข้า

จากนั้นผู้เชี่ยวชาญทุกท่านร่วมกันสนทนากลุ่ม โดยมีประเด็นในการสนทนาคือสารสนเทศที่มีผลต่อการตัดสินใจ รวมถึงสารสนเทศที่ต้องการเพิ่มเติมเพื่อให้การกำหนดคะแนนจุดตัดทำได้ง่ายขึ้น และข้อเสนอแนะเพื่อการปรับปรุงกระบวนการกำหนดคะแนนจุดตัด

ในการกำหนดคะแนนจุดตัดผู้วิจัยได้เก็บข้อมูลเพื่อตรวจสอบความตรงของกระบวนการตัดสินใจด้วยวิธีการที่หลากหลาย โดยมีวัตถุประสงค์เพื่อให้ได้ข้อมูลมากที่สุด ผู้วิจัยได้สรุปรายละเอียดของช่วงเวลาที่เก็บข้อมูล วิธีการ และข้อมูลที่ต้องการดังแสดงในตารางต่อไปนี้

ตารางที่ 3.5 การเก็บข้อมูลเพื่อตรวจสอบความตรงของกระบวนการตัดสินใจของผู้เชี่ยวชาญ

ช่วงเวลา	วิธีการ	ข้อมูลที่ต้องการ
ฝึกอบรม	การสังเกตพฤติกรรม	พฤติกรรมและประเด็นคำถามที่ถูกถามบ่อย ๆ
กำหนดคะแนนจุดตัด	การสังเกตพฤติกรรมและเทคนิคการคิดออกเสียง	พฤติกรรมและกระบวนการตัดสินใจ
อภิปรายกลุ่มย่อยและกลุ่มใหญ่	การสังเกตพฤติกรรมและการบันทึกข้อสนทนา	พฤติกรรมและประเด็นสนทนา
พัก	การบันทึกข้อสนทนา	ประเด็นที่สนทนาเกี่ยวกับการกำหนดคะแนนจุดตัด
หลังจากเสร็จสิ้นการกำหนดคะแนนจุดตัด	การสนทนากลุ่ม	ข้อมูลเกี่ยวกับสารสนเทศที่มีผลต่อการตัดสินใจ รวมถึงสารสนเทศที่ต้องการเพิ่มเติมเพื่อให้การกำหนดคะแนนจุดตัดทำได้ง่ายขึ้น และข้อเสนอแนะเพื่อการปรับปรุงกระบวนการกำหนดคะแนนจุดตัด

วิธีการกำหนดคะแนนจุดตัด

ในการศึกษาครั้งนี้ได้กำหนดคะแนนจุดตัดโดยใช้ข้อสอบวิชาคณิตศาสตร์ และวิชาวิทยาศาสตร์ด้วยวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซี ขั้นตอนในการกำหนดคะแนนจุดตัดแต่ละวิธีมีรายละเอียดดังนี้

1. วิธีแองกอฟแบบใช่/ไม่ใช่

การกำหนดคะแนนจุดตัดโดยวิธีแองกอฟแบบใช่/ไม่ใช่ จะดำเนินการกำหนดคะแนนจุดตัด 3 รอบ เนื่องจากเป็นจำนวนรอบที่นักวิชาการส่วนใหญ่นิยม (Busch & Jaeger, 1990; Fehrmann, et al., 1991; Hambleton & Plake, 1995 ; Hess, et al., 2007) แต่ละรอบผู้เชี่ยวชาญสามารถเปลี่ยนแปลงคะแนนจุดตัดได้ โดยดำเนินการดังต่อไปนี้

รอบที่ 1 ผู้เชี่ยวชาญแต่ละท่านพิจารณาข้อสอบทีละข้อและตัดสินว่าผู้มีความสามารถคาบเส้นของแต่ละระดับความสามารถจะตอบข้อสอบแต่ละข้อได้ถูกต้องหรือไม่ หากผู้เชี่ยวชาญพิจารณาว่าผู้มีความสามารถคาบเส้นของระดับความสามารถที่กำลังพิจารณาอยู่สามารถตอบข้อสอบข้อนั้นได้ถูกต้องก็ให้ใส่เลข 1 ลงในช่องผลการพิจารณาของข้อนั้น แต่หากผู้เชี่ยวชาญพิจารณาว่าผู้มีความสามารถคาบเส้นของระดับความสามารถที่กำลังพิจารณาอยู่ไม่สามารถตอบข้อสอบข้อนั้นได้ถูกต้องก็ให้ใส่เลข 0 ลงในช่องผลการพิจารณาของข้อนั้น โดยผู้เชี่ยวชาญจะต้องพิจารณาทีละระดับความสามารถจนครบทุกระดับ

รอบที่ 2 ผู้เชี่ยวชาญร่วมกันอภิปรายกลุ่มย่อยเกี่ยวกับคะแนนจุดตัดที่กำหนดในรอบที่ 1 และกำหนดคะแนนจุดตัดอีกครั้งในรอบนี้ โดยใช้หลักการเดียวกันกับรอบที่ 1

รอบที่ 3 ผู้เชี่ยวชาญร่วมกันอภิปรายกลุ่มใหญ่เกี่ยวกับคะแนนจุดตัดที่กำหนดในรอบที่ 2 และกำหนดคะแนนจุดตัดครั้งสุดท้าย โดยใช้หลักการเดียวกันกับรอบที่ 1

วิธีการคิดคะแนนจุดตัด

การคิดคะแนนจุดตัดที่กำหนดโดยผู้เชี่ยวชาญแต่ละท่าน สามารถดำเนินการได้ดังนี้

การศึกษาครั้งนี้เป็นการกำหนดคะแนนจุดตัดสำหรับใช้เป็นเกณฑ์ในการแบ่งผู้สอบออกเป็น 4 ระดับความสามารถ ได้แก่ ระดับที่ 1 ระดับที่ 2 ระดับที่ 3 และระดับที่ 4 ดังนั้นจะต้องกำหนดคะแนนจุดตัดทั้งสิ้น 3 ค่า โดยค่าแรกเป็นคะแนนจุดตัดระหว่างความสามารถระดับที่ 1 กับระดับที่ 2 ค่าที่สองเป็นคะแนนจุดตัดระหว่างความสามารถระดับที่ 2 กับระดับที่ 3 และค่าที่สาม

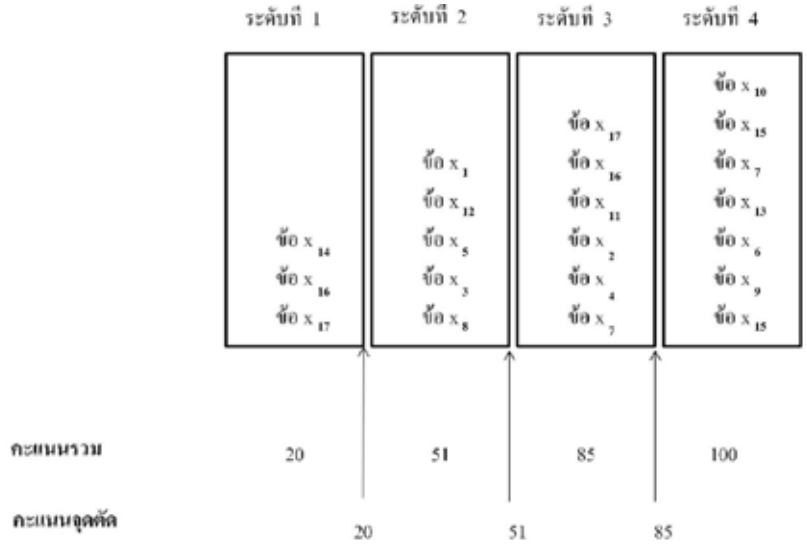
เป็นคะแนนจุดตัดระหว่างความสามารถระดับที่ 3 กับระดับที่ 4 ตามลำดับ หลักในการคิดคะแนนจุดตัดแต่และค่าเป็นดังนี้

คะแนนจุดตัดระหว่างความสามารถระดับที่ 1 กับระดับที่ 2 คิดจากผลคะแนนรวมของข้อสอบข้อที่เชี่ยวชาญคิดว่าผู้มีความสามารถคาบเส้นของระดับที่ 1 สามารถตอบได้ถูกต้อง

คะแนนจุดตัดระหว่างความสามารถระดับที่ 2 กับระดับที่ 3 คิดจากผลคะแนนรวมของข้อสอบข้อที่ผู้เชี่ยวชาญคิดว่าผู้มีความสามารถคาบเส้นของระดับที่ 2 สามารถตอบได้ถูกต้อง

คะแนนจุดตัดระหว่างความสามารถระดับที่ 3 กับระดับที่ 4 คิดจากผลคะแนนรวมของข้อสอบข้อที่ผู้เชี่ยวชาญคิดว่าผู้มีความสามารถคาบเส้นของระดับที่ 3 สามารถตอบได้ถูกต้อง

ผู้วิจัยขอเสนอภาพประกอบคำอธิบายวิธีการคิดคะแนนจุดตัดของวิธีแองกอฟแบบใช่/ไม่ใช่ โดย “ข้อ x” แสดงถึงข้อสอบข้อที่ผู้เชี่ยวชาญคิดว่าผู้มีความสามารถคาบเส้นของแต่ละระดับจะสามารถตอบได้ถูกต้อง รายละเอียดต่าง ๆ ดังแสดงในภาพต่อไปนี้



ภาพที่ 3.2 ตัวอย่างการคิดคะแนนจุดตัดของวิธีแองกอฟแบบใช่/ไม่ใช่

2. วิธีบุ๊คมาร์ค

วิธีบุ๊คมาร์คจะดำเนินการกำหนดคะแนนจุดตัด 3 รอบ ในแต่ละรอบผู้เชี่ยวชาญสามารถเปลี่ยนแปลงคะแนนจุดตัดได้ ขั้นตอนการดำเนินการแต่ละรอบมีรายละเอียดดังนี้

รอบที่ 1 ผู้เชี่ยวชาญแต่ละท่านพิจารณาคู่มือเรียงข้อสอบในแต่ละหน้า และพิจารณาในประเด็นต่อไปนี้ 1) ผู้สอบต้องมีความรู้ความสามารถในเรื่องใดจึงจะสามารถตอบข้อสอบแต่ละข้อได้ถูกต้อง และ 2) เหตุใดข้อสอบข้อนี้จึงยากกว่าข้อก่อนๆ จากนั้นผู้เชี่ยวชาญกำหนดคะแนนจุดตัดอย่างอิสระโดยนำที่คั่นหนังสือมาคั่นที่ข้อสอบที่ผู้เชี่ยวชาญคิดว่ามีความเหมาะสมที่จะเป็นจุดตัด โดยข้อที่อยู่ก่อนหน้าที่คั่นหนังสือจะเป็นข้อที่ผู้มีความสามารถคาบเส้นของระดับความสามารถที่กำลังพิจารณาอยู่ มีความน่าจะเป็น 0.67 ที่จะตอบข้อสอบได้ถูกต้อง โดยผู้เชี่ยวชาญจะพิจารณาจากระดับความสามารถต่ำที่สุดไปจนถึงระดับความสามารถที่สูงที่สุด

รอบที่ 2 ผู้เชี่ยวชาญร่วมกันอภิปรายกลุ่มย่อยเกี่ยวกับคะแนนจุดตัดที่กำหนดในรอบที่ 1 และกำหนดคะแนนจุดตัดอีกครั้ง โดยอาศัยหลักการเดียวกับรอบที่ 1

รอบที่ 3 ผู้เชี่ยวชาญร่วมกันอภิปรายกลุ่มใหญ่เกี่ยวกับคะแนนจุดตัดที่กำหนดในรอบที่ 2 และกำหนดคะแนนจุดตัดอีกครั้ง โดยอาศัยหลักการเดียวกับรอบที่ 1

การจัดทำคู่มือเรียงข้อสอบ

คู่มือเรียงข้อสอบจัดทำโดยแสดงรายละเอียดของข้อสอบไว้หน้าละ 1 ข้อ แต่ละหน้าจะถูกจัดเรียงตามค่าความยากของข้อสอบโดยเรียงลำดับหน้าจากข้อสอบง่ายที่ง่ายที่สุดไปหาข้อที่ยากที่สุด โดยค่าความยากของข้อสอบได้มาจากการวิเคราะห์โดยใช้ทฤษฎีการตอบสนองข้อสอบโมเดล 3 พารามิเตอร์ ตัวอย่างรายละเอียดคู่มือเรียงข้อสอบในแต่ละหน้าดังแสดงในภาพต่อไปนี้

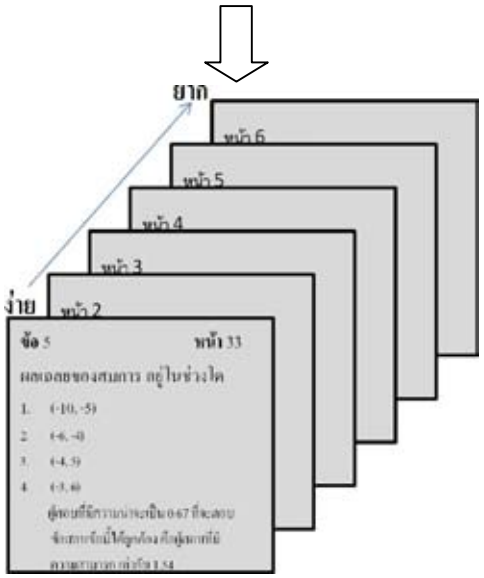
หน้า 33

ข้อ 5

ผลเฉลยของสมการ $2|5-x|=1$ อยู่ในช่วงใด

1. (-10, -5)
2. (-6, -4)
3. (-4, 5)
4. (-3, 6)

ผู้สอบที่มีความน่าจะเป็น 0.67 ที่จะตอบข้อสอบข้อนี้ได้ถูกต้อง คือผู้สอบที่มีความสามารถ (θ) เท่ากับ 1.54



ภาพที่ 3.3 ตัวอย่างคู่มือเรียงข้อสอบ

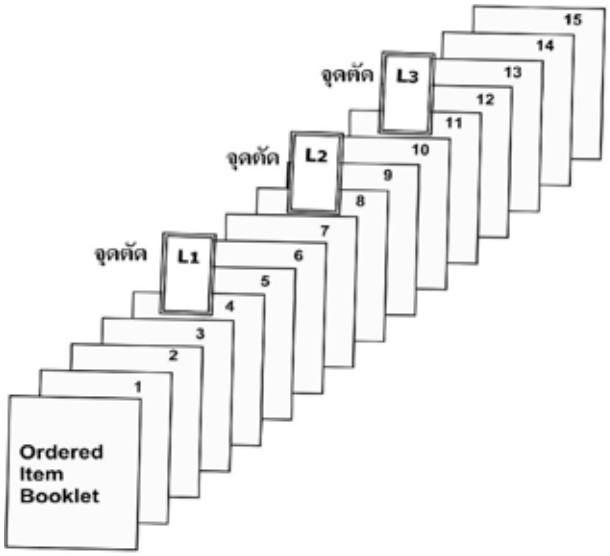
จากตัวอย่างคู่มือการจัดเรียงข้อสอบแต่ละหน้า มีรายละเอียดที่แสดงดังนี้

1) แสดงลำดับข้อสอบในแบบสอบไว้ที่มุมซ้ายมือ

2) แสดงลำดับหน้าในคู่มือเรียงข้อสอบ ไว้ที่มุมขวามือ

3) แสดงข้อคำถาม และตัวเลือก ไว้ในย่อหน้าแรก

4) แสดงค่าความสามารถของผู้สอบ (θ) ที่มีความน่าจะเป็น 0.67 ที่จะตอบข้อสอบข้อนั้นได้ถูกต้อง ไว้ในบรรทัดสุดท้าย



ภาพที่ 3.4 ตัวอย่างการกำหนดคะแนนจุดตัดของวิธีบุ๊คมาร์ค

วิธีการคิดคะแนนจุดตัด

การคิดคะแนนจุดตัดที่กำหนดโดยผู้เชี่ยวชาญแต่ละท่าน สามารถดำเนินการได้ดังนี้

คะแนนจุดตัดแต่ละค่าหาได้จากการเปลี่ยนค่าความสามารถของผู้สอบ (θ) ที่มีความน่าจะเป็น 0.67 ที่จะตอบข้อสอบข้อที่อยู่ก่อนหน้าที่คั่นหนังสือได้ถูกต้อง ไปเป็นคะแนนสอบ (Expected score) โดยสามารถดูคะแนนสอบที่ระดับความสามารถต่าง ๆ ได้จากสารสนเทศเกี่ยวกับการแปลงค่าความสามารถเป็นคะแนนจุดตัด ดังแสดงในภาคผนวก ฉ

3. วิธีเอบีซี

การกำหนดคะแนนจุดตัดโดยวิธีเอบีซีจะดำเนินการ 3 รอบ ซึ่งแต่ละรอบผู้เชี่ยวชาญสามารถเปลี่ยนแปลงคะแนนจุดตัดได้ โดยแต่ละรอบของการตัดสินจะดำเนินการดังต่อไปนี้

รอบที่ 1 ผู้เชี่ยวชาญพิจารณาคู่มือเรียงข้อสอบรายสาระการเรียนรู้ และผลการวิเคราะห์กลุ่ม จากนั้นก็ให้ผู้เชี่ยวชาญพิจารณาว่าผู้ที่สามารถตอบข้อสอบแต่ละข้อได้ถูกต้องควรมีความสามารถอย่างน้อยระดับใด โดยมีหลักการพิจารณาเพิ่มเติมคือหากข้อใดที่ผู้เชี่ยวชาญไม่มั่นใจก็ให้ระบุไว้ในข้อสอบ ดังนั้นข้อสอบที่ผู้เชี่ยวชาญกำหนดไว้ในแต่ละระดับความสามารถก็จะมีสองกลุ่ม คือกลุ่มที่ผู้เชี่ยวชาญมั่นใจและกลุ่มที่ผู้เชี่ยวชาญไม่มั่นใจ

รอบที่ 2 ผู้เชี่ยวชาญร่วมกันอภิปรายกลุ่มย่อยเกี่ยวกับข้อสอบที่ผู้เชี่ยวชาญยังไม่มั่นใจในรอบที่ 1 เพื่อให้การอภิปรายกระชับและตรงประเด็นมากขึ้น และกำหนดคะแนนจุดตัดอีกครั้ง โดยอาศัยหลักการเดียวกับรอบที่ 1

รอบที่ 3 ผู้เชี่ยวชาญร่วมกันอภิปรายกลุ่มใหญ่เกี่ยวกับข้อสอบที่ผู้เชี่ยวชาญยังไม่มั่นใจในรอบที่ 2 และกำหนดคะแนนจุดตัดอีกครั้ง โดยอาศัยหลักการเดียวกับรอบที่ 1

วิธีการคิดคะแนนจุดตัด

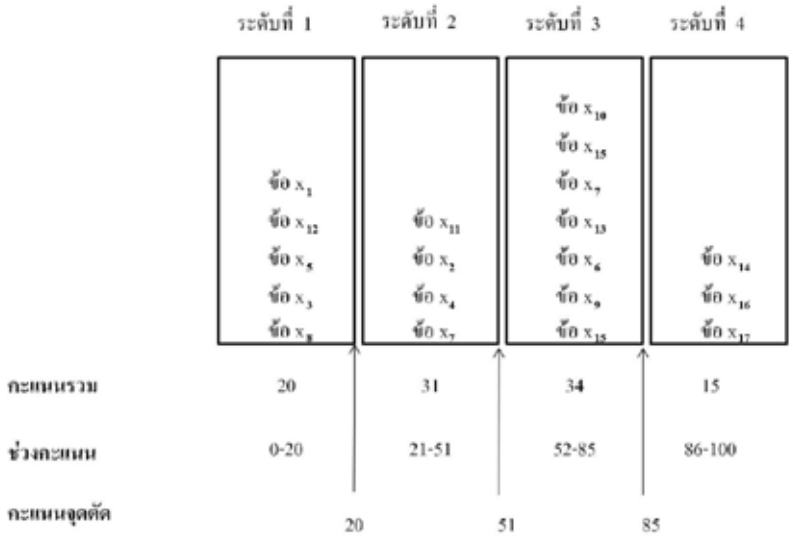
การคิดคะแนนจุดตัดที่กำหนดโดยผู้เชี่ยวชาญแต่ละท่าน สามารถดำเนินการได้ดังนี้

คะแนนจุดตัดระหว่างความสามารถระดับที่ 1 กับระดับที่ 2 คิดจากคะแนนรวมของข้อสอบที่ผู้เชี่ยวชาญคิดว่าผู้ที่สามารถตอบได้ถูกต้องควรมีความสามารถอย่างน้อยระดับที่ 1

คะแนนจุดตัดระหว่างความสามารถระดับที่ 2 กับระดับที่ 3 คิดจากคะแนนรวมของข้อสอบที่ผู้เชี่ยวชาญคิดว่าผู้ที่สามารถตอบได้ถูกต้องควรมีความสามารถอย่างน้อยระดับที่ 1 และข้อสอบที่ผู้เชี่ยวชาญคิดว่าผู้ที่สามารถตอบได้ถูกต้องควรมีความสามารถอย่างน้อยระดับที่ 2

คะแนนจุดตัดระหว่างความสามารถระดับที่ 3 กับระดับที่ 4 คิดจากคะแนนรวมของข้อสอบที่ผู้เชี่ยวชาญคิดว่าผู้ที่สามารถตอบได้ถูกต้องควรมีความสามารถอย่างน้อยระดับที่ 1 ข้อสอบที่ผู้เชี่ยวชาญคิดว่าผู้ที่สามารถตอบได้ถูกต้องควรมีความสามารถอย่างน้อยระดับที่ 2 และข้อสอบที่ผู้เชี่ยวชาญคิดว่าผู้ที่สามารถตอบได้ถูกต้องควรมีความสามารถอย่างน้อยระดับที่ 3

ผู้วิจัยขอเสนอภาพประกอบคำอธิบายวิธีการคิดคะแนนจุดตัดของวิธีเอบีซีโดย “ข้อ x” แสดงถึงข้อสอบที่ผู้เชี่ยวชาญคิดว่าผู้ที่สามารถตอบได้ถูกต้องควรมีความสามารถอย่างน้อยระดับนั้น ๆ รายละเอียดต่าง ๆ ดังแสดงในภาพต่อไปนี้



ภาพที่ 3.5 ตัวอย่างการคิดคะแนนจุดตัดของวิธีเอบีซี

การวิเคราะห์ข้อมูล

ผู้วิจัยนำข้อมูลที่ได้มาวิเคราะห์หาค่าต่างๆ ดังนี้

1. การวิเคราะห์ความตรงตามเนื้อหาของแบบสอบถาม โดยใช้สูตรในการคำนวณดังนี้ (ศิริชัย กาญจนวาสี, 2548)

$$IOC = \frac{\sum R}{N}$$

เมื่อ IOC แทน ค่าดัชนีความสอดคล้องระหว่างข้อคำถามกับจุดประสงค์ (Item Objective Congruence)

R แทน คะแนนความคิดเห็นของผู้เชี่ยวชาญ

+ 1 หมายถึง แน่ใจว่าข้อคำถามวัดได้ตรงกับจุดประสงค์

0 หมายถึง ไม่แน่ใจว่าข้อคำถามวัดได้ตรงกับจุดประสงค์

-1 หมายถึง แน่ใจว่าข้อคำถามวัดได้ไม่ตรงกับจุดประสงค์

N แทน จำนวนผู้เชี่ยวชาญ

2. การวิเคราะห์ข้อมูลในขั้นตอนการกำหนดจำนวนของระดับความสามารถ ชื่อแต่ละระดับและคำอธิบายระดับความสามารถ

การวิเคราะห์ข้อมูลที่ได้จากแบบสอบถามความคิดเห็นเกี่ยวกับจำนวนของระดับ ชื่อของระดับ และคำอธิบายระดับความสามารถ แบบสอบถามความคิดเห็นเกี่ยวกับร่างจำนวนของระดับ ชื่อของระดับ และคำอธิบายระดับความสามารถ ผู้วิจัยวิเคราะห์โดยการจำแนกประเภทข้อมูล (Typological Analysis) การเปรียบเทียบข้อมูล (Comparison) และการสร้างข้อสรุปแบบอุปนัย (Analytic Induction)

3. การวิเคราะห์ข้อมูลเพื่อกำหนดคะแนนจุดตัด

3.1 การวิเคราะห์ค่าพารามิเตอร์ของข้อสอบ ได้แก่ ค่าความยากของข้อสอบ (b) และค่าพารามิเตอร์ของผู้สอบ (θ) โดยใช้โปรแกรม MULTILOG

3.1.1 สำหรับข้อสอบที่ให้คะแนนแบบ 2 ค่า วิเคราะห์ค่าความยากของข้อสอบ (b) และค่าความสามารถของผู้สอบ (θ) ด้วยโมเดลการตอบสนองข้อสอบแบบ 3 พารามิเตอร์ โดยกำหนดให้โอกาสของการเดาข้อสอบ (c) เป็นศูนย์

$$P(X=1|\theta)=\left(\frac{1}{1+e^{\left[-D\hat{a}\left(\theta-\hat{b}\right)\right]}}\right)$$

$$\theta_{3PL}=\left(\frac{1}{D\hat{a}}\right)\ln\left(\frac{x}{1-x}\right)+\hat{b}$$

เมื่อ $P(X=1|\theta)$ = ความน่าจะเป็นที่ผู้ตอบซึ่งมีระดับความสามารถ θ จะตอบข้อสอบได้ถูกต้อง

$\hat{b}$ = ค่าความยากของข้อสอบ

$\hat{a}$ = อำนาจจำแนกของข้อสอบ

θ= ความสามารถของผู้สอบ

e = ค่าคงที่ ซึ่งมีค่าประมาณ 2.71828

D = มีค่าเท่ากับ 1.70

x = ความน่าจะเป็นของการตอบข้อสอบถูก

3.1.2 สำหรับข้อสอบที่ให้คะแนนแบบมากกว่า 2 ค่า วิเคราะห์ค่าความยากของข้อสอบ (b) และค่าความสามารถของผู้สอบ (θ) ด้วยโมเดลการตอบสนองข้อสอบแบบ graded-response (GR) models

$$P(X \geq k|\theta) = \left(\frac{1}{1+e^{\left[-D\hat{a}\left(\theta-\hat{b}_k\right)\right]}} \right)$$

$$\theta_{GR} = \left(\frac{1}{D\hat{a}}\right)\ln\left(\frac{x}{1-x}\right) + \hat{b}_k$$

เมื่อ $P(X \geq k|\theta)$= ความน่าจะเป็นที่ผู้ตอบซึ่งมีระดับความสามารถ θ จะตอบข้อสอบได้ถูกต้องในแต่ละรายการคำตอบ

$\hat{b}$ = ค่าความยากของข้อสอบข้อ

$\hat{a}$ = อำนาจจำแนกของข้อสอบ

θ = ความสามารถของผู้สอบ

e = ค่าคงที่ ซึ่งมีค่าประมาณ 2.71828

D = มีค่าเท่ากับ 1.70

x = ความน่าจะเป็นของการตอบข้อสอบถูก

3) การวิเคราะห์เพื่อเปลี่ยนค่าความสามารถของผู้สอบ (θ) เป็นคะแนนจุดตัดโดยอาศัยโค้งลักษณะของแบบสอบ (Test Characteristic Curve: TCC) ซึ่งเป็นผลรวมของโค้งลักษณะข้อสอบ (Item Characteristic Curve: ICC) ของข้อสอบทุกข้อ

$$\mathrm{TCC}(\theta_j) = \sum_{j=1}^{n} \mathrm{P}(u_{ij} = 1|\theta_i)$$

เมื่อ $P(u = 1|\theta)$ = โอกาสของการตอบข้อสอบถูกในแต่ละข้อที่ระดับความสามารถต่าง ๆ

3.2 การวิเคราะห์กลุ่ม

ผู้วิจัยวิเคราะห์กลุ่มแบบไม่เป็นขั้นตอน (Nonhierarchical Cluster Analysis) หรือ K-Means Clustering โดยตัวแปรในการแบ่งกลุ่ม ได้แก่ คะแนนรายสาระการเรียนรู้ โดยใช้โปรแกรมคอมพิวเตอร์สำเร็จรูป SPSS for Windows

4. การวิเคราะห์ข้อมูลเพื่อวิเคราะห์คุณภาพของวิธีการกำหนดคะแนนจุดตัด

4.1 การวิเคราะห์ข้อมูลเพื่อศึกษาขนาดองค์ประกอบความแปรปรวน และค่าสัมประสิทธิ์ความน่าเชื่อถือสำหรับการตัดสินใจเชิงสัมบูรณ์ของการกำหนดคะแนนจุดตัดด้วยวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซี โดยใช้โปรแกรม GENOVA (GENeralized analysis Of Variance System) ฟาเซทของการวัดมี 3 ฟาเซท คือ ผู้เชี่ยวชาญ (p) กลุ่มผู้เชี่ยวชาญ (g) และจำนวนรอบ (r) โดยสิ่งที่ต้องการวัด คือ ผลการกำหนดคะแนนจุดตัด รูปแบบความสัมพันธ์ระหว่างฟาเซทเป็น $(p:g)xr$ Design กล่าวคือ แต่ละกลุ่มประกอบด้วยผู้เชี่ยวชาญที่ไม่ซ้ำกันและทุกกลุ่มกำหนดคะแนนจุดตัดทุกรอบ ซึ่งสามารถเขียนเป็นโมเดลเชิงเส้นตรงสำหรับคะแนนจุดตัดที่สังเกตได้ของผู้เชี่ยวชาญแต่ละคน (X_{pgr}) ได้ดังนี้

$$X_{pgr} = \mu + \upsilon_{p:g} + \upsilon_{g} + \upsilon_{r} + \upsilon_{gr} + \upsilon_{pr:g}$$

เมื่อ X_{pgr} = คะแนนจุดตัดที่สังเกตได้ของผู้เชี่ยวชาญแต่ละคน

μ = ค่าเฉลี่ยของคะแนนจุดตัดทั้งหมดจากประชากรในเอกภพ (grand mean)

υ = ผล (effect) จากองค์ประกอบความแปรปรวนจากแหล่งต่าง ๆ ที่ส่งผลให้คะแนนจุดตัดของผู้เชี่ยวชาญแต่ละคนเบี่ยงเบนไปจาก μ

$\upsilon_{p:g}$ = ผลจากองค์ประกอบความแปรปรวนของผู้เชี่ยวชาญในแต่ละกลุ่ม

υ_{g} = ผลจากองค์ประกอบความแปรปรวนของกลุ่มผู้เชี่ยวชาญ

υ_{r} = ผลจากองค์ประกอบความแปรปรวนของรอบในการตัดสิน

υ_{gr} = ผลจากองค์ประกอบความแปรปรวนของปฏิสัมพันธ์ระหว่างกลุ่มผู้เชี่ยวชาญและรอบในการตัดสิน

$\upsilon_{pr:g}$ = ผลจากองค์ประกอบความแปรปรวนส่วนที่เหลือ

4.2 การศึกษาเชิงสรุปอ้างอิง (G-Study) เพื่อประมาณความแปรปรวนของความคลาดเคลื่อนสัมบูรณ์ (Absolute error variance: σ^2_{Abs}) จากแหล่งต่างๆ เมื่อฟาเซทของผู้เชี่ยวชาญ (p) และกลุ่มผู้เชี่ยวชาญ (g) เป็นองค์ประกอบสุ่ม (random) และฟาเซทจำนวนรอบ (r) เป็นองค์ประกอบเจาะจง (fix)

$$\sigma^2_{Abs} = \frac{\hat{\sigma}^2(g)}{n'_g} + \frac{\hat{\sigma}^2(gr)}{n'_r n'_g} + \frac{\hat{\sigma}^2(p:g)}{n'_p n'_g} + \frac{\hat{\sigma}^2(pr:g)}{n'_p n'_r n'_g}$$

เมื่อ $\hat{\sigma}^2(g)$ = ความแปรปรวนของกลุ่มผู้เชี่ยวชาญ

$\hat{\sigma}^2(gr)$ = ความแปรปรวนของปฏิสัมพันธ์ระหว่างกลุ่มผู้เชี่ยวชาญและรอบในการตัดสิน

$\hat{\sigma}^2(p:g)$ = ความแปรปรวนของผู้เชี่ยวชาญในแต่ละกลุ่ม

$\hat{\sigma}^2(pr:g)$ = ความแปรปรวนส่วนที่เหลือ

n'_p = จำนวนผู้เชี่ยวชาญในแต่ละกลุ่ม

n'_r = จำนวนรอบ

n'_g = จำนวนกลุ่มผู้เชี่ยวชาญ

4.3 คำนวณค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์

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

5. การวิเคราะห์ความเที่ยงภายในผู้เชี่ยวชาญ (Intra-rater reliability) และวิเคราะห์ความเที่ยงระหว่างผู้เชี่ยวชาญ (Inter-rater reliability) โดยการวิเคราะห์สัมประสิทธิ์สหสัมพันธ์ภายในชั้น (Intra-Class Correlation: ICC) โดยใช้โปรแกรมคอมพิวเตอร์สำเร็จรูป SPSS for Windows

6. การวิเคราะห์ความตรงของปัจจัยนำเข้า

วิเคราะห์ข้อมูลเพื่อตรวจสอบความตรงของปัจจัยนำเข้า โดยการวิเคราะห์ข้อมูลจากแบบสอบถามความคิดเห็นเกี่ยวกับความตรงของปัจจัยนำเข้า โดยวิเคราะห์ค่าเฉลี่ย (Mean) ของระดับความคิดเห็นและส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation) โดยใช้โปรแกรมคอมพิวเตอร์สำเร็จรูป SPSS for Windows โดยมีรายละเอียดการแปลความหมายคะแนนดังนี้

ค่าเฉลี่ย (Mean) ใช้การแปลความหมายคะแนนเป็น 5 ระดับคือ

ระดับความคิดเห็น	1.00 – 1.49	หมายถึง	เห็นด้วยน้อยที่สุด
	1.50 – 2.49	หมายถึง	เห็นด้วยน้อย
	2.50 – 3.49	หมายถึง	เห็นด้วยปานกลาง

3.50 – 4.49	หมายถึง	เห็นด้วยมาก
4.50 – 5.00	หมายถึง	เห็นด้วยมากที่สุด

7. การวิเคราะห์ความตรงของกระบวนการตัดสินใจ

วิเคราะห์ข้อมูลที่เก็บรวบรวมจากการสังเกตพฤติกรรม การใช้เทคนิคการคิดออกเสียง การบันทึกข้อสนทนา และการสนทนากลุ่มโดยการสร้างข้อสรุปแบบอุปนัย การเปรียบเทียบข้อมูล และการสร้างข้อสรุปแบบอุปนัย

8. การวิเคราะห์ความตรงของผลลัพธ์

วิเคราะห์ข้อมูลเพื่อตรวจสอบความตรงของผลลัพธ์ โดยการวิเคราะห์ความสอดคล้องของคะแนนจุดตัดที่กำหนดโดยวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซี โดยการวิเคราะห์ค่าสัมประสิทธิ์แคปป้า (Kappa coefficient: K) ด้วยโปรแกรมคอมพิวเตอร์สำเร็จรูป SPSS for Windows

$$K = \frac{P(A) - P(E)}{1 - P(E)}$$

เมื่อ P(A) = ค่าสัดส่วนที่คะแนนจุดตัดจากแต่ละวิธีกำหนดกลุ่มผู้สอบได้สอดคล้องกัน

P(E) = ค่าสัดส่วนที่คะแนนจุดตัดจากแต่ละวิธีกำหนดกลุ่มผู้สอบได้สอดคล้องกันโดยบังเอิญ(by chance)

การแปลความหมายความสอดคล้อง (Landis & Koch, 1977) เป็นดังนี้

ค่าความสอดคล้อง ≤ 0.2	หมายถึง	สอดคล้องกันน้อยมาก
0.3 - 0.4	หมายถึง	สอดคล้องกันน้อย
0.5 - 0.6	หมายถึง	สอดคล้องกันปานกลาง
0.7 - 0.8	หมายถึง	สอดคล้องกันมาก
>0.8	หมายถึง	สอดคล้องกันอย่างยิ่ง

9. การวิเคราะห์ความตรงของผลที่ตามมา

วิเคราะห์ข้อมูลเพื่อตรวจสอบความตรงของผลที่ตามมา โดยการวิเคราะห์ความถูกต้องของการกำหนดกลุ่ม (Classification accuracy) และความคลาดเคลื่อนของการกำหนดกลุ่ม (Classification errors) ได้แก่ ความคลาดเคลื่อนแบบบวก และความคลาดเคลื่อนแบบลบของคะแนนจุดตัดในการกำหนดกลุ่มผู้สอบตามระดับความสามารถโดยใช้โปรแกรม MULT-CLASS (Lee, 2008) ในการวิเคราะห์

บทที่ 4

ผลการวิเคราะห์ข้อมูล

การศึกษาครั้งนี้มีวัตถุประสงค์เพื่อกำหนดคะแนนจุดตัดและตรวจสอบคุณภาพของคะแนนจุดตัดสำหรับการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐานวิชาคณิตศาสตร์ และวิชาวิทยาศาสตร์ ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 รวมทั้งวิเคราะห์จำนวนผู้เชี่ยวชาญ และจำนวนกลุ่มผู้เชี่ยวชาญที่มีความเหมาะสมเมื่อกำหนดคะแนนจุดตัดด้วยวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซี ในการวิเคราะห์ข้อมูล ผู้วิจัยเสนอผลการวิเคราะห์ให้สอดคล้องกับวัตถุประสงค์และขั้นตอนในการดำเนินการวิจัย โดยแบ่งการนำเสนอเป็น 4 ตอน ได้แก่

ตอนที่ 1 ผลการกำหนดจำนวนของระดับ ชื่อของระดับ และคำอธิบายระดับความสามารถ

ตอนที่ 2 ผลการพัฒนาวิธีการกำหนดคะแนนจุดตัด

ตอนที่ 3 ผลการกำหนดคะแนนจุดตัด

ตอนที่ 4 ผลการวิเคราะห์คุณภาพของคะแนนจุดตัด

ผู้วิจัยขอเสนอรายละเอียดผลการวิเคราะห์ข้อมูลดังต่อไปนี้

ตอนที่ 1 ผลการกำหนดจำนวนของระดับ ชื่อของระดับ และคำอธิบายระดับความสามารถ

ผลการวิเคราะห์ข้อมูลเพื่อกำหนดจำนวนของระดับ ชื่อของระดับ และคำอธิบายระดับความสามารถ ผู้วิจัยขอนำเสนอเป็น 2 ส่วน ประกอบด้วย ส่วนแรกเป็นผลการวิเคราะห์เนื้อหาจากการศึกษาเอกสารและงานวิจัยเพื่อศึกษาแนวคิดและหลักการในการกำหนดจำนวนของระดับ ชื่อของระดับ และคำอธิบายระดับความสามารถ ส่วนที่สอง เป็นการวิเคราะห์ความคิดเห็น แนวคิดและข้อเสนอแนะในการกำหนดจำนวนของระดับ ชื่อของระดับ และคำอธิบายระดับความสามารถของผู้บริหารระดับสูงของหน่วยงานที่รับผิดชอบการจัดการศึกษาขั้นพื้นฐาน อาจารย์มหาวิทยาลัยและครูผู้สอนในวิชาคณิตศาสตร์ วิชาฟิสิกส์ วิชาเคมีและวิชาชีววิทยา

ส่วนที่ 1 ผลการวิเคราะห์เนื้อหาจากเอกสารและงานวิจัยที่เกี่ยวข้องเพื่อศึกษาแนวคิดและหลักการในการกำหนดจำนวนของระดับ ชื่อของระดับ และคำอธิบายระดับความสามารถ ผู้วิจัยขอเสนอเป็นหัวข้อย่อยได้แก่ แนวทางการกำหนดจำนวนระดับความสามารถ แนวทางการกำหนดชื่อของระดับ และแนวทางการพัฒนาคำอธิบายระดับความสามารถ

แนวทางการกำหนดจำนวนระดับความสามารถ

จากการทบทวนเอกสารและงานวิจัยเกี่ยวกับการกำหนดจำนวนของระดับความสามารถ พบว่า การกำหนดจำนวนของระดับความสามารถจะกำหนดโดยพิจารณาจากเป้าหมาย (Goals) ของการนำผลไปใช้ โดยเป้าหมายของการกำหนดระดับความสามารถมีหลายประการ ได้แก่ เป้าหมายเพื่อการรับรอง (Certifiers) เช่น ใช้เพื่อการรับรองการจบการศึกษาและรับรองความสามารถในการทำงาน เป้าหมายเพื่อทำนาย (Predictors) เช่น ใช้เพื่อคัดเลือกบุคคลเข้าศึกษาต่อในมหาวิทยาลัย เป้าหมายเพื่อบรรยาย (Descriptors) เช่น ใช้แบ่งกลุ่มผู้สอบตามระดับความสามารถ และเป้าหมายเพื่อกระตุ้น (Motivators) เช่น ใช้เป็นคะแนนที่เป็นเป้าหมายของการจัดการศึกษา หรือการเรียนการสอน เป็นต้น (Green, 1995) หากมีเป้าหมายเพื่อต้องการนำผลไปใช้ในการรับรอง หรือ อนุมัติ (Licensure) ควรกำหนดระดับความสามารถเพียง 2 ระดับ คือ ผ่านและตก (Pass / Fail) ก็เพียงพอ แต่หากมีเป้าหมายเพื่อทำนาย บรรยาย หรือเพื่อกระตุ้น ก็ควรกำหนดจำนวนระดับความสามารถเป็นหลายระดับเพื่อใช้จำแนก(Classified) ผู้สอบตามระดับความสามารถ โดยควรกำหนดระดับความสามารถมากกว่า 2 ระดับ เช่น กำหนดระดับความสามารถเป็น 3, 4 หรือ 5 ระดับ เป็นต้น (Fehrmann et al., 1991, Giraud et al., 2005, Perie, 2008)

ในต่างประเทศ เช่นประเทศสหรัฐอเมริกา มีการกำหนดให้แต่ละรัฐดำเนินการกำหนดระดับความสามารถของผู้เรียนเพื่อใช้ในการตัดสินเกี่ยวกับระดับความสามารถของผู้เรียนที่ทำการทดสอบเพื่อวัดผลของพัฒนาการทางการเรียนเป็นรายปี (Adequate Yearly Progress) จากการศึกษาจำนวนของระดับความสามารถที่แต่ละรัฐกำหนด พบว่าแต่ละรัฐมีการกำหนดระดับความสามารถแตกต่างกัน โดยจำนวนของระดับที่นิยมกำหนดมากที่สุดคือ 4 ระดับ รองลงมาคือ 5 ระดับและ 3 ระดับตามลำดับ (Perie, 2008)

จากที่กล่าวมาข้างต้นสรุปได้ว่า การทดสอบทางการศึกษาระดับชาติขั้นพื้นฐานมีเป้าหมายสำคัญเพื่อตรวจสอบผลการจัดการศึกษา ดังนั้นผลการสอบควรจะสามารถใช้เพื่ออธิบายระดับความสามารถของผู้สอบอันเป็นการสะท้อนถึงระดับคุณภาพของการจัดการศึกษา ดังนั้นการจะกำหนดระดับความสามารถควรกำหนดหลายระดับเพื่อใช้จำแนกผู้สอบตามระดับความสามารถ

แนวทางการกำหนดชื่อของระดับ

จากการศึกษาเอกสารและงานวิจัยเกี่ยวกับการกำหนดชื่อของระดับ พบว่า ชื่อของระดับต้องมีความชัดเจน (clear) เหมาะสม (reasonable) ใช้ภาษาที่ไม่ขัดแย้งกันเอง และไม่เป็นชื่อที่คลุมเครือ (Nebulous) โดยควรกำหนดให้สอดคล้องกับวัตถุประสงค์ของการวัด (Purpose) โครงสร้างของการวัด (Construct) ควรเป็นชื่อที่สนับสนุนการสรุปผล (Inferences) และการกำหนดกลุ่ม (Classifications) (Cizek and Bunch, 2007 cited in Perie, 2008; Perie, 2008) นอกจากนี้ควรใช้คำกลาง ๆ ในการตั้งชื่อระดับ เช่น ความสามารถระดับที่ 1, ระดับที่ 2 และระดับที่ 3 เป็นต้น เพื่อหลีกเลี่ยงการแปลความหมายตามคำที่ใช้จนเกินที่ต้องการสื่อจริง(Zieky et al., 2008) ในประเทศสหรัฐอเมริกา ชื่อของระดับที่นิยมใช้มากที่สุดคือ ต่ำกว่าพื้นฐาน/ พื้นฐาน/ เชี่ยวชาญ/ เชี่ยวชาญพิเศษ (Below Basic/ Basic/ Proficient/ Advanced) รองลงมาเป็นระดับที่ 1, ระดับที่ 2 และระดับที่ 3 เป็นต้น (Perie, 2008)

แนวทางการกำหนดคำอธิบายระดับความสามารถ

จากการศึกษาเอกสารและงานวิจัยเกี่ยวกับการกำหนดคำอธิบายระดับความสามารถ พบว่า คำอธิบายระดับความสามารถเป็นการเขียนบรรยายเกี่ยวกับความรู้ และทักษะที่ผู้สอบในแต่ละระดับความสามารถควรจะมี หากเป็นไปได้ควรเขียนอธิบายระดับความสามารถให้เสร็จก่อนที่จะดำเนินการสร้างข้อสอบ ทั้งนี้เนื่องจากว่าหากผู้ที่สร้างข้อสอบทราบว่าระดับความสามารถแต่ละ

ระดับจะต้องสามารถทำอะไรหรือรู้อะไรบ้าง ก็จะสามารถสร้างข้อคำถามเพื่อวัดความสามารถในระดับต่างๆ ได้อย่างถูกต้องและครอบคลุมทุกระดับความสามารถที่กำหนด (Perie, 2008; Zieky et al., 2008) การเขียนคำอธิบายระดับความสามารถต้องเขียนให้มีความชัดเจน เนื่องจากคำอธิบายระดับความสามารถถือเป็นปัจจัยสำคัญปัจจัยหนึ่งที่ส่งผลต่อคุณภาพของคะแนนจุดตัด กล่าวคือ หากเขียนไม่ชัดเจนย่อมส่งผลให้การกำหนดคะแนนจุดตัดผิดพลาดได้ (Ferdous and Plake, 2005; Perie, 2006)

ส่วนที่ 2 ผลการกำหนดจำนวนของระดับ ชื่อของระดับ และคำอธิบายระดับความสามารถ

จากการศึกษาความคิดเห็น และข้อเสนอแนะในการกำหนดจำนวนของระดับ ชื่อของระดับ และคำอธิบายระดับความสามารถ ของผู้บริหารระดับสูงของหน่วยงานที่รับผิดชอบการจัดการศึกษาขั้นพื้นฐาน อาจารย์มหาวิทยาลัย และครู ผลการกำหนดจำนวนของระดับ ชื่อของระดับ และคำอธิบายระดับความสามารถดังแสดงในตารางต่อไปนี้

ตารางที่ 4.1 จำนวนของระดับ ชื่อของระดับ และคำอธิบายระดับความสามารถ วิชาคณิตศาสตร์

ชื่อระดับความสามารถ	คำอธิบายระดับความสามารถ
ระดับที่ 1	นักเรียนมีความรู้ ความสามารถทางวิชาการ และมีทักษะพื้นฐานเล็กน้อย คือ นักเรียนเข้าใจแนวคิดพื้นฐานทางคณิตศาสตร์ สามารถนำความรู้ และทักษะพื้นฐานทางคณิตศาสตร์ไปใช้ในการคำนวณและแก้ปัญหา ง่าย ๆ ได้
ระดับที่ 2	นักเรียนมีความรู้ ความสามารถทางวิชาการ และมีทักษะพื้นฐานอย่างดี คือ นักเรียนเข้าใจแนวคิดพื้นฐานทางคณิตศาสตร์ สามารถนำความรู้ และทักษะพื้นฐานทางคณิตศาสตร์ไปใช้ในการคำนวณและแก้ปัญหาที่มีความซับซ้อนได้
ระดับที่ 3	นักเรียนมีความรู้ ความสามารถทางวิชาการ และมีทักษะดีเยี่ยม คือ นักเรียนสามารถเชื่อมโยงความรู้ และทักษะทางคณิตศาสตร์ในการคำนวณและแก้ปัญหาที่มีความซับซ้อนได้อย่างสมเหตุสมผล
ระดับที่ 4	นักเรียนมีความรู้ ความสามารถทางวิชาการ และมีทักษะดีเยี่ยม สามารถนำความรู้ที่มีไปประยุกต์ใช้ได้อย่างดี คือ นักเรียนสามารถเชื่อมโยงความรู้ และทักษะทางคณิตศาสตร์ในการคำนวณและแก้ปัญหาที่มีความซับซ้อนได้อย่างสมเหตุสมผล ตลอดจนสามารถใช้ความรู้และทักษะทางคณิตศาสตร์ช่วยในการแปลความหมาย ตัดสินใจและแก้ปัญหาในสถานการณ์ต่าง ๆ ได้

ตารางที่ 4.2 จำนวนของระดับ ชื่อของระดับ และคำอธิบายระดับความสามารถ วิชาวิทยาศาสตร์

ชื่อระดับความสามารถ	คำอธิบายระดับความสามารถ
ระดับที่ 1	นักเรียนมีความรู้ ความสามารถทางวิชาการ และมีทักษะพื้นฐานเล็กน้อย คือ นักเรียนเข้าใจแนวคิดพื้นฐานทางวิทยาศาสตร์ สามารถนำความรู้ และทักษะพื้นฐานทางวิทยาศาสตร์ไปใช้ในการแก้ปัญหา ง่าย ๆ ได้
ระดับที่ 2	นักเรียนมีความรู้ ความสามารถทางวิชาการ และมีทักษะพื้นฐานอย่างดี คือ นักเรียนเข้าใจแนวคิดพื้นฐานทางวิทยาศาสตร์ สามารถนำความรู้ และทักษะพื้นฐานทางวิทยาศาสตร์ไปใช้ในการแก้ปัญหาที่มีความซับซ้อนได้
ระดับที่ 3	นักเรียนมีความรู้ ความสามารถทางวิชาการ และมีทักษะดีเยี่ยม คือ นักเรียนสามารถเชื่อมโยงความรู้ และทักษะทางวิทยาศาสตร์ในการแก้ปัญหาที่มีความซับซ้อนได้อย่างสมเหตุสมผล
ระดับที่ 4	นักเรียนมีความรู้ ความสามารถทางวิชาการ และมีทักษะดีเยี่ยม สามารถนำความรู้ที่มีไปประยุกต์ใช้ได้อย่างดี คือ นักเรียนสามารถเชื่อมโยงความรู้ และทักษะทางวิทยาศาสตร์ในการแก้ปัญหาที่มีความซับซ้อนได้อย่างสมเหตุสมผล ตลอดจนสามารถใช้ความรู้และทักษะทางวิทยาศาสตร์ช่วยในการแปลความหมาย ตัดสินใจและแก้ปัญหาในสถานการณ์ต่าง ๆ ได้

ตอนที่ 2 ผลการพัฒนาวิธีการกำหนดคะแนนจุดตัด

วิธีกำหนดคะแนนจุดตัดที่พัฒนาขึ้น คือวิธีเอบีซี เป็นวิธีที่ให้ผู้เชี่ยวชาญกำหนดคะแนนจุดตัดจากคู่มือเรียงข้อสอบรายสาระการเรียนรู้ และผู้เชี่ยวชาญจะได้สารสนเทศเกี่ยวกับผลการวิเคราะห์กลุ่มประกอบการกำหนดคะแนนจุดตัดด้วย เป้าหมายในการพัฒนาวิธีนี้เพื่อลดความซับซ้อนของกระบวนการคิดของผู้เชี่ยวชาญ และเพื่อให้ได้วิธีที่ง่ายต่อการดำเนินการ โดยการผสมผสานระหว่างวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีวิเคราะห์กลุ่ม ดังนั้นผู้วิจัยขอนำเสนอผลการวิเคราะห์เนื้อหาจากเอกสารและงานวิจัยเกี่ยวกับแนวคิดและหลักการของการกำหนดคะแนนจุดตัด รวมทั้งจุดดีและจุดด้อยของวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีการวิเคราะห์กลุ่ม และขอเสนอแนวคิดเกี่ยวกับการพัฒนาวิธีเอบีซี รวมถึงขั้นตอนการดำเนินการของวิธีเอบีซีด้วย ซึ่งมีรายละเอียดดังต่อไปนี้

2.1 วิธีแองกอฟแบบใช่/ไม่ใช่

2.1.1 ความเป็นมา

วิธีแองกอฟแบบใช่/ไม่ใช่ ถูกพัฒนาขึ้นมาโดย Impara และ Plake (1997) โดยมีเป้าหมายของการพัฒนาวิธีนี้เพื่อลดจุดอ่อนของวิธีแองกอฟ กล่าวคือมีเป้าหมายเพื่อลดความซับซ้อนของกระบวนการคิดของผู้เชี่ยวชาญ (Skaggs & Hein, 2011, p. 586) โดยวิธีแองกอฟแบบใช่/ไม่ใช่จะให้ผู้เชี่ยวชาญตัดสินใจว่าผู้มีความสามารถคาบเส้นของแต่ละระดับความสามารถจะตอบข้อสอบได้ถูกต้องหรือไม่ แทนการให้ผู้เชี่ยวชาญพิจารณาความน่าจะเป็นที่ผู้มีความสามารถคาบเส้นจะสามารถตอบข้อสอบแต่ละข้อได้ถูกต้องตามแบบวิธีแองกอฟ

2.1.2 ขั้นตอนการดำเนินการ

วิธีแองกอฟแบบใช่/ไม่ใช่มีขั้นตอนการดำเนินการกำหนดคะแนนจุดตัด 3 รอบ ดังนี้

รอบที่ 1 ผู้เชี่ยวชาญแต่ละท่านพิจารณาข้อสอบทีละข้อและตัดสินว่าผู้มีความสามารถคาบเส้นของแต่ละระดับความสามารถจะตอบข้อสอบแต่ละข้อได้ถูกต้องหรือไม่ หากผู้เชี่ยวชาญพิจารณาว่าผู้มีความสามารถคาบเส้นของระดับความสามารถที่กำลังพิจารณาอยู่สามารถตอบข้อสอบข้อนั้นได้ถูกต้องก็ให้ใส่เลข 1 ลงในช่องผลการพิจารณาของข้อนั้น แต่หากผู้เชี่ยวชาญพิจารณาว่าผู้มีความสามารถคาบเส้นของระดับความสามารถที่กำลังพิจารณาอยู่ไม่สามารถตอบ

ข้อสอบข้อนั้นได้ถูกต้องก็ให้ใส่เลข 0 ลงในช่องผลการพิจารณาของข้อนั้น โดยผู้เชี่ยวชาญจะต้องพิจารณาทีละระดับความสามารถจนครบทุกระดับ

รอบที่ 2 ผู้เชี่ยวชาญร่วมกันอภิปรายกลุ่มย่อยเกี่ยวกับคะแนนจุดตัดที่กำหนดในรอบที่ 1 และกำหนดคะแนนจุดตัดอีกครั้งในรอบนี้ โดยใช้หลักการเดียวกันกับรอบที่ 1

รอบที่ 3 ผู้เชี่ยวชาญร่วมกันอภิปรายกลุ่มใหญ่เกี่ยวกับคะแนนจุดตัดที่กำหนดในรอบที่ 2 และกำหนดคะแนนจุดตัดครั้งสุดท้าย โดยใช้หลักการเดียวกันกับรอบที่ 1

2.1.3 ข้อดีและข้อเสีย

วิธีแองกอฟแบบใช่/ไม่ใช่มีข้อดีคือเป็นวิธีที่ง่ายต่อการนำไปปฏิบัติ กล่าวคือสามารถปฏิบัติได้โดยไม่ต้องใช้เวลาในการเรียนรู้มาก และใช้สถิติในการวิเคราะห์ไม่ยาก สำหรับข้อด้อยที่สำคัญคือการพิจารณากำหนดคะแนนจุดตัดนั้นต้องพิจารณาข้อสอบทุกข้อในทุก ๆ รอบของการพิจารณา ดังนั้นหากข้อสอบมีจำนวนมากก็ต้องใช้เวลาในการพิจารณานานจึงไม่เหมาะกับการกำหนดคะแนนจุดตัดสำหรับข้อสอบที่มีจำนวนข้อมากๆ

2.2 วิธีบุ๊คมาร์ค

2.2.1 ความเป็นมา

วิธีบุ๊คมาร์คถูกพัฒนาขึ้นโดย Lewis และ Green (1996 cited in Zieky et al., 2008) ในปี 1996 โดยมีเป้าหมายของการพัฒนาวิธีนี้เพื่อลดความซับซ้อนของกระบวนการคิดของผู้เชี่ยวชาญ (Skaggs & Hein, 2011, p. 586) เนื่องจากวิธีบุ๊คมาร์คให้ผู้เชี่ยวชาญกำหนดคะแนนจุดตัดโดยพิจารณาจากคู่มือเรียงข้อสอบ ซึ่งเป็นคู่มือที่ข้อสอบถูกเรียงไว้ตามลำดับความยาก ทำให้ผู้เชี่ยวชาญสามารถกำหนดจุดตัดได้ง่ายขึ้น วิธีบุ๊คมาร์คเป็นวิธีที่สามารถใช้ได้ทั้งกับข้อสอบที่เป็นแบบเลือกตอบที่ให้คะแนน 2 ค่า และข้อสอบแบบให้คะแนนมากกว่า 2 ค่า และเป็นวิธีที่เหมาะกับการกำหนดคะแนนจุดตัดสำหรับข้อสอบที่มีจำนวนมาก

2.2.2 ขั้นตอนการดำเนินการ

วิธีบุ๊คมาร์คมีขั้นตอนการดำเนินการกำหนดคะแนนจุดตัด 3 รอบ ดังนี้

รอบที่ 1 ผู้เชี่ยวชาญแต่ละท่านพิจารณาคู่มือเรียงข้อสอบในแต่ละหน้า และพิจารณาในประเด็นต่อไปนี้ 1) ผู้สอบต้องมีความรู้ความสามารถในเรื่องใดจึงจะสามารถตอบข้อสอบแต่ละข้อได้ถูกต้อง และ 2) เหตุใดข้อสอบข้อนี้จึงยากกว่าข้อก่อนๆ จากนั้นผู้เชี่ยวชาญกำหนดคะแนน

จุดตัดอย่างอิสระโดยนำที่คั่นหนังสือมาคั่นที่ข้อสอบซึ่งผู้เชี่ยวชาญคิดว่ามีความเหมาะสมที่จะเป็นจุดตัด โดยข้อที่อยู่ก่อนหน้าที่คั่นหนังสือจะเป็นข้อที่ผู้มีความสามารถคาบเส้นของระดับความสามารถที่กำลังพิจารณาอยู่ มีความน่าจะเป็น 0.67 ที่จะตอบข้อสอบได้ถูกต้อง โดยผู้เชี่ยวชาญจะพิจารณาจากระดับความสามารถต่ำที่สุดไปจนถึงระดับความสามารถที่สูงที่สุด

รอบที่ 2 ผู้เชี่ยวชาญร่วมกันอภิปรายกลุ่มย่อยเกี่ยวกับคะแนนจุดตัดที่กำหนดในรอบที่ 1 และกำหนดคะแนนจุดตัดอีกครั้ง โดยอาศัยหลักการเดียวกับรอบที่ 1

รอบที่ 3 ผู้เชี่ยวชาญร่วมกันอภิปรายกลุ่มใหญ่เกี่ยวกับคะแนนจุดตัดที่กำหนดในรอบที่ 2 และกำหนดคะแนนจุดตัดอีกครั้ง โดยอาศัยหลักการเดียวกับรอบที่ 1

2.2.3 ข้อดีและข้อเสีย

วิธีนี้มีข้อดีคือเป็นวิธีที่ใช้ได้ทั้งกับข้อสอบที่ให้คะแนนแบบ 2 ค่าและแบบให้คะแนนมากกว่า 2 ค่า และมีจุดเด่นอยู่ที่คู่มือเรียงข้อสอบ ที่ช่วยให้ผู้เชี่ยวชาญกำหนดคะแนนจุดตัดได้ง่ายและเร็ว ส่วนข้อด้อยคือ เป็นวิธีที่ต้องใช้กับข้อสอบจำนวนมากมิฉะนั้นอาจเกิด floor and ceiling effect ได้

2.3 วิธีวิเคราะห์กลุ่ม

2.3.1 ความเป็นมา

การวิเคราะห์กลุ่มเป็นการกำหนดคะแนนจุดตัดที่ไม่ต้องกำหนดคำอธิบายระดับความสามารถ และไม่อาศัยการตัดสินใจของผู้เชี่ยวชาญ แต่จะกำหนดกลุ่มโดยการวิเคราะห์ทางสถิติ (Kane, 2001 cited in Hess, 2007) กล่าวคือพยายามให้หน่วยที่อยู่ในกลุ่มเดียวกันมีลักษณะที่สนใจเหมือนหรือคล้ายกันมากที่สุดและหน่วยที่อยู่ต่างกลุ่มกันมีลักษณะที่สนใจต่างกันให้มากที่สุด การวิเคราะห์กลุ่มจะมีเทคนิคการรวมกลุ่มที่นิยมใช้มากที่สุด 2 เทคนิค คือ การวิเคราะห์กลุ่มแบบลดหลั่น (Hierarchical Cluster Analysis: HCA) และการวิเคราะห์กลุ่มแบบไม่ลดหลั่น (Nonhierarchical Cluster Analysis หรือ K-Means Clustering) โดยการวิเคราะห์กลุ่มแบบขั้นตอนมีข้อจำกัดคือเมื่อนำหน่วยใดไว้ในกลุ่มใดแล้ว จะไม่มีการย้ายหน่วยนั้นไว้กลุ่มอื่นอีก และเหมาะสำหรับการวิเคราะห์ในกรณีที่ข้อมูลมีจำนวนไม่มาก ($n < 200$) (Sireci, 2001 cited in Hess, 2007; กัลยา วานิชบัญชา, 2550) ส่วนการวิเคราะห์กลุ่มแบบไม่ลดหลั่น เป็นเทคนิคการจัดกลุ่มโดยการกำหนดจำนวนกลุ่มที่ต้องการแบ่งไว้ก่อน Kane (2001 cited in Hess, 2007) กล่าวว่าการนำวิธีการ

วิเคราะห์กลุ่มมาใช้ในการกำหนดคะแนนจุดตัดถือเป็นการกำหนดคะแนนจุดตัดที่ปราศจากความลำเอียงใดๆ เนื่องจากไม่ได้อาศัยการตัดสินใจโดยผู้เชี่ยวชาญ นอกจากนี้ผลจากการวิเคราะห์กลุ่มยังทำให้ได้ข้อมูลเกี่ยวกับคะแนนสอบที่แต่ละกลุ่มมีความคาบเกี่ยวกันรวมถึงรายละเอียดเกี่ยวกับคะแนนเฉลี่ยของแต่ละตัวแปรที่ใช้ในการกำหนดกลุ่มด้วย ซึ่งข้อมูลเหล่านี้น่าจะเป็นประโยชน์ต่อการกำหนดคะแนนจุดตัดวิธีอื่น ๆ ได้

2.3.2 ข้อดีและข้อด้อย

ข้อดีของวิธีนี้คือเป็นการกำหนดคะแนนจุดตัดที่ปราศจากความลำเอียงใดๆ เนื่องจากเป็นการกำหนดคะแนนจุดตัดที่ไม่ได้อาศัยการตัดสินของผู้เชี่ยวชาญ แต่เป็นการวิเคราะห์ข้อมูลจากผลการตอบข้อสอบ และสารสนเทศที่ได้จากการวิเคราะห์กลุ่ม อาทิ คะแนนสอบที่แต่ละกลุ่มคาบเกี่ยวกัน รวมถึงรายละเอียดเกี่ยวกับคะแนนเฉลี่ยของแต่ละตัวแปรที่ใช้ในการกำหนดกลุ่มสามารถใช้เป็นข้อมูลประกอบการกำหนดคะแนนจุดตัดวิธีอื่น ๆ ที่อาศัยผู้เชี่ยวชาญเป็นผู้ตัดสินได้ สำหรับข้อด้อย คือ เนื่องจากเป็นการวิเคราะห์ที่อาศัยข้อมูลจากผลการตอบข้อสอบ ดังนั้นจึงไม่สามารถกำหนดคะแนนจุดตัดที่สูงกว่าหรือต่ำกว่าผลการตอบข้อสอบของนักเรียนกลุ่มที่เลือกมาวิเคราะห์ได้

2.4 วิธีเอบีซี

2.4.1 ความเป็นมา

วิธีเอบีซีเป็นวิธีที่ผสมผสานระหว่างวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีวิเคราะห์กลุ่ม โดยวิธีแองกอฟแบบใช่/ไม่ใช่ และวิธีบุ๊คมาร์คเป็นวิธีแบบสอบเป็นศูนย์กลาง (test-center methods) (Zieky et al., 2008) ที่ได้รับความนิยมมากที่สุด ซึ่งเป็นการกำหนดคะแนนจุดตัดโดยเน้นการพิจารณาจากข้อสอบเป็นหลัก ส่วนวิธีการวิเคราะห์กลุ่มเป็นวิธีผู้สอบเป็นศูนย์กลาง (Examinee-centered methods) ซึ่งเป็นการกำหนดคะแนนจุดตัดที่เน้นการพิจารณาจากผลการตอบของผู้สอบเป็นหลัก (Berk, 1996; Plake, 2008; Lee & Lewis, 2008; Näsström & Nyström, 2008)

วิธีแองกอฟแบบใช่/ไม่ใช่มีจุดเด่นคือการให้ความสำคัญกับการกำหนดคะแนนจุดตัดของผู้เชี่ยวชาญโดยปราศจากกรอบเรื่องค่าความยากของข้อสอบ กล่าวคือเป็นการกำหนดคะแนนจุดตัดที่ให้ความสำคัญกับการตัดสินใจของผู้เชี่ยวชาญที่พิจารณาจากเนื้อหาที่สอบวัดเป็นหลัก

สำหรับวิธีบุ๊คมาร์คมีจุดเด่นอยู่ที่คู่มือเรียงข้อสอบ ซึ่งเป็นคู่มือที่เรียงข้อสอบจากข้อง่ายไปข้อยาก ซึ่งคู่มือนี้จะช่วยให้ผู้เชี่ยวชาญกำหนดคะแนนจุดตัดได้ง่ายและเร็ว ส่วนวิธีการวิเคราะห์กลุ่มเป็นการวิเคราะห์ทางสถิติเพื่อจัดให้กลุ่มผู้สอบที่มีลักษณะที่สนใจเหมือนหรือคล้ายกันอยู่กลุ่มเดียวกัน และแต่ละกลุ่มจะต้องมีลักษณะที่สนใจต่างกันให้มากที่สุด ซึ่งกล่าวได้ว่าเป็นวิเคราะห์เพื่อกำหนดกลุ่มให้ผู้สอบจากลักษณะที่สนใจโดยปราศจากความลำเอียงใดๆ

ข้อด้อยของวิธีแองกอฟแบบใช่/ไม่ใช่นั้น ผู้เชี่ยวชาญจะต้องใช้กระบวนการคิดอย่างมากในการพิจารณาว่าผู้มีความสามารถคาบเส้นจะสามารถตอบข้อสอบได้ถูกต้องใช่หรือไม่ โดยเฉพาะในกรณีที่มีจำนวนข้อสอบที่ต้องพิจารณามาก ๆ และต้องกำหนดคะแนนจุดตัดหลายรอบ อาจทำให้ผู้เชี่ยวชาญเกิดความอ่อนล้าในขณะที่กำหนดคะแนนจุดตัดซึ่งอาจส่งผลต่อคุณภาพของคะแนนจุดตัดได้ (Clauser, et al., 2009) ส่วนวิธีบุ๊คมาร์คมีจุดด้อยในทางปฏิบัติ คืออาจเกิดกรณีที่ผู้เชี่ยวชาญคิดว่าข้อสอบที่ถูกเรียงไว้ตามลำดับความยากมีตำแหน่งที่ไม่เหมาะสม ซึ่งกรณีเช่นนี้อาจส่งผลต่อคุณภาพของคะแนนจุดตัดที่กำหนดได้ ส่วนวิธีการวิเคราะห์กลุ่มมีจุดด้อยคือไม่สามารถระบุได้ว่าแต่ละกลุ่มที่แบ่งมีความสามารถระดับใดและไม่สามารถกำหนดคะแนนจุดตัดที่สูงกว่าหรือต่ำกว่าผลการตอบข้อสอบของผู้สอบกลุ่มที่เลือกมาวิเคราะห์ได้ (Violato, Marini & Lee, 2003; Hess, Subhiyah & Giordano, 2007)

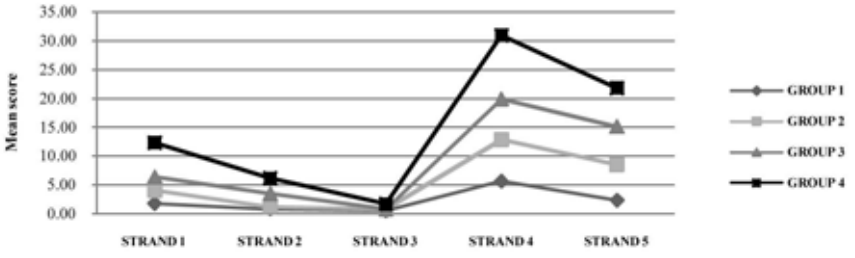
2.4.2 ขั้นตอนการดำเนินการ

วิธีเอบีซีถูกพัฒนาขึ้นจากการผสมผสานระหว่างวิธีแองกอฟแบบใช่/ไม่ใช่ และวิธีบุ๊คมาร์คเพื่อดึงเอาจุดเด่นและลดข้อด้อยของแต่ละวิธีลง โดยวิธีเอบีซีมีขั้นตอนการดำเนินการ 3 รอบ ตามลำดับดังนี้

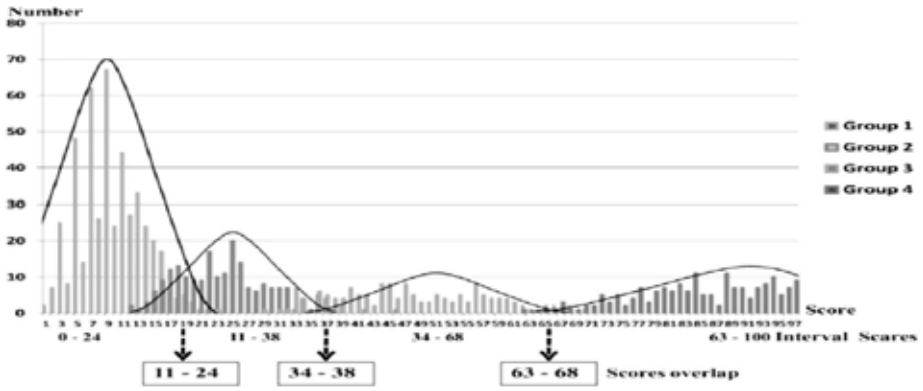
รอบที่ 1 ผู้เชี่ยวชาญกำหนดคะแนนจุดตัดจากคู่มือเรียงข้อสอบซึ่งถือเป็นจุดเด่นของบุ๊คมาร์ค แต่วิธีเอบีซีจะแบ่งคู่มือเรียงข้อสอบเป็นหลาย ๆ ฉบับตามสาระการเรียนรู้ เรียกว่าคู่มือเรียงข้อสอบรายสาระการเรียนรู้ (Domain–Ordered Item Booklet: D-OIB) ซึ่งเป็นคู่มือที่จัดทำขึ้นโดยแบ่งข้อสอบเป็นกลุ่มตามสาระการเรียนรู้ กล่าวคือคู่มือเรียงข้อสอบรายสาระการเรียนรู้จะมีหลายฉบับขึ้นอยู่กับจำนวนสาระการเรียนรู้ที่สอบวัด ข้อสอบในแต่ละสาระการเรียนรู้จะถูกจัดเรียงจากข้อง่ายไปข้อยาก ข้อละ 1 หน้า โดยใช้ทฤษฎีการตอบสนองข้อสอบ โมเดล 3 พารามิเตอร์ ในการวิเคราะห์ค่าความยากของข้อสอบ (b) ทั้งนี้เพื่อให้การพิจารณากำหนดคะแนนจุดตัดทำได้ง่าย

ขึ้น เนื่องจากคู่มือเรียงข้อสอบรายสาระการเรียนรู้แต่ละฉบับจะประกอบด้วยข้อสอบที่วัดสาระการเรียนรู้เดียวกัน ส่วนหลักการพิจารณาเพื่อกำหนดคะแนนจุดตัดจะดำเนินการคล้ายกับวิธีแองกอฟแบบใช่/ไม่ใช่ กล่าวคือให้ผู้เชี่ยวชาญพิจารณาข้อสอบตามลำดับความยาก และตัดสินว่าผู้ที่สามารถตอบข้อสอบแต่ละข้อได้ถูกต้องควรมีความสามารถอย่างน้อยระดับใดแทนที่จะให้ผู้เชี่ยวชาญนำที่คั่นหนังสือมาคั่นตรงข้อที่ผู้เชี่ยวชาญคิดว่าเป็นจุดตัด ซึ่งการพิจารณาแบบนี้น่าจะช่วยลดปัญหาที่เกิดจากการที่ผู้เชี่ยวชาญคิดว่าข้อสอบที่ถูกจัดเรียงไว้ตามความยากมีลำดับที่ไม่เหมาะสม เพราะวิธีเอบีซีใช้คู่มือเรียงข้อสอบเป็นเพียงสารสนเทศประกอบการกำหนดคะแนนจุดตัดที่เสนอเกี่ยวกับค่าความยากของข้อสอบที่ได้จากการวิเคราะห์จากผลการตอบของผู้สอบเท่านั้น กล่าวได้ว่าวิธีเอบีซียังคงให้ความสำคัญกับความคิดเห็นของผู้เชี่ยวชาญที่พิจารณาจากเนื้อหาที่สอบเป็นหลักแทนที่จะต้องพิจารณากำหนดคะแนนจุดตัดที่ถูกจำกัดไว้กับลำดับของความยากของข้อสอบจนเกินไป

นอกจากนี้วิธีเอบีซีจะเสนอสารสนเทศที่มากกว่าค่าความยากของข้อสอบ โดยจะเสนอสารสนเทศเกี่ยวกับผลการวิเคราะห์กลุ่ม (Clustering – Result: CR) ซึ่งเป็นผลการจัดกลุ่มที่มิใช่เป็นการกำหนดโดยอาศัยความคิดเห็นของผู้เชี่ยวชาญแต่เป็นการวิเคราะห์ข้อมูลทางสถิติที่ใช้ข้อมูลจากผลการสอบของผู้สอบมาทำการวิเคราะห์ โดยตัวแปรที่ใช้ในการแบ่งกลุ่ม ได้แก่ คะแนนรวมแต่ละสาระการเรียนรู้ สารสนเทศนี้จะนำเสนอเป็นผลคะแนนเฉลี่ยรายสาระการเรียนรู้ของแต่ละกลุ่ม และช่วงของคะแนนสอบที่คาบเกี่ยวกันของแต่ละกลุ่ม (รายละเอียดดังแสดงในภาคผนวก ฉ) ผู้วิจัยขอเสนอตัวอย่างดังแสดงในภาพต่อไปนี้



ภาพที่ 4.1 ตัวอย่างผลคะแนนเฉลี่ยรายสาระการเรียนรู้ของแต่ละกลุ่ม



ภาพที่ 4.2 ตัวอย่างช่วงของคะแนนสอบที่คาบเกี่ยวกันของแต่ละกลุ่ม

ผู้วิจัยคิดว่าข้อมูลเกี่ยวกับผลคะแนนเฉลี่ยรายสาระการเรียนรู้ของแต่ละกลุ่มน่าจะเป็นข้อมูลที่เป็นประโยชน์ต่อการตัดสินใจกำหนดข้อสอบไว้ในแต่ละกลุ่ม ส่วนข้อมูลเกี่ยวกับช่วงของคะแนนสอบที่คาบเกี่ยวกันของแต่ละกลุ่มน่าจะเป็นประโยชน์ต่อผู้เชี่ยวชาญในแง่การเปรียบเทียบผลการกำหนดคะแนนจุดตัดของผู้เชี่ยวชาญเองกับช่วงคะแนนที่คาบเกี่ยวกันของแต่ละกลุ่มที่ถูกกำหนดโดยการวิเคราะห์กลุ่ม

ทั้งนี้วิธีเอบีซีมีหลักการพิจารณาเพิ่มเติมคือหากข้อใดที่ผู้เชี่ยวชาญไม่มั่นใจก็ให้ระบุไว้ในข้อสอบ ดังนั้นข้อสอบที่ผู้เชี่ยวชาญกำหนดไว้ในแต่ละระดับความสามารถก็จะมีสองกลุ่ม คือ กลุ่มที่ผู้เชี่ยวชาญมั่นใจและกลุ่มที่ผู้เชี่ยวชาญไม่มั่นใจ ซึ่งสำหรับข้อสอบที่ผู้เชี่ยวชาญยังไม่มั่นใจนั้น ผู้เชี่ยวชาญสามารถนำประเด็นที่ไม่มั่นใจมาอภิปรายร่วมกันกับผู้เชี่ยวชาญท่านอื่น ๆ ในรอบที่ 2 และ 3 ได้

รอบที่ 2 ผู้เชี่ยวชาญร่วมกันอภิปรายกลุ่มย่อยเกี่ยวกับข้อสอบที่ผู้เชี่ยวชาญยังไม่มั่นใจในรอบที่ 1 เพื่อให้การอภิปรายกระชับและตรงประเด็นมากขึ้น และกำหนดคะแนนจุดตัดอีกครั้ง โดยอาศัยหลักการเดียวกับรอบที่ 1

รอบที่ 3 ผู้ตัดสินร่วมกันอภิปรายกลุ่มใหญ่เกี่ยวกับข้อสอบที่ผู้ตัดสินยังไม่มั่นใจในรอบที่ 2 และกำหนดคะแนนจุดตัดอีกครั้ง โดยอาศัยหลักการเดียวกับรอบที่ 1

ตอนที่ 3 ผลการกำหนดคะแนนจุดตัด

ผู้วิจัยแบ่งการนำเสนอผลการกำหนดคะแนนจุดตัดเป็น 2 ส่วน ได้แก่ ส่วนที่ 1 เป็นผลการวิเคราะห์ข้อมูลเบื้องต้นของข้อสอบ และส่วนที่ 2 เป็นผลการกำหนดคะแนนจุดตัด รายละเอียดดังแสดงต่อไปนี้

ส่วนที่ 1 ผลการวิเคราะห์ข้อมูลเบื้องต้นของข้อสอบ รายละเอียดดังแสดงในตารางต่อไปนี้

ตารางที่ 4.3 ผลการวิเคราะห์ข้อมูลเบื้องต้นของข้อสอบวิชาคณิตศาสตร์และวิชาวิทยาศาสตร์

ข้อสอบ	ผู้สอบ (จำนวนคน)	ข้อสอบ (จำนวนข้อ)	คะแนน				
			ต่ำสุด	สูงสุด	เต็ม	*M*	*SD*
วิชาคณิตศาสตร์							
สาระการเรียนรู้ที่ 1 จำนวนและการดำเนินการ	356,591	7	0	10.00	18	2.21	2.89
สาระการเรียนรู้ที่ 2 การวัด	356,591	4	0	11.00	11	1.10	1.90
สาระการเรียนรู้ที่ 3 เรขาคณิต	356,591	2	0	4.00	4	0.81	1.14
สาระการเรียนรู้ที่ 4 พีชคณิต	356,591	16	0	37.00	37	7.35	5.88
สาระการเรียนรู้ที่ 5 การวิเคราะห์ข้อมูลและความน่าจะเป็น	356,591	11	0	30.00	30	3.51	4.89
ทั้งฉบับ	**356,591**	**40**	**0**	**100.00**	**100**	**14.99**	**13.50**
วิชาวิทยาศาสตร์							
สาระการเรียนรู้ที่ 1 สิ่งมีชีวิตกับกระบวนการดำรงชีวิต	349,210	10	0	11.00	11	3.55	1.70
สาระการเรียนรู้ที่ 2 ชีวิตกับสิ่งแวดล้อม	349,210	10	0	11.00	11	4.24	1.99
สาระการเรียนรู้ที่ 3 สารและสมบัติของสาร	349,210	18	0	18.00	18	4.49	2.24
สาระการเรียนรู้ที่ 4 แรงและการเคลื่อนที่	349,210	10	0	11.00	11	3.42	1.76
สาระการเรียนรู้ที่ 5 พลังงาน	349,210	10	0	11.00	11	3.62	1.85
สาระการเรียนรู้ที่ 6 กระบวนการเปลี่ยนแปลงของโลก	349,210	11	0	13.00	13	3.70	1.69
สาระการเรียนรู้ที่ 7 ดาราศาสตร์และอวกาศ	349,210	11	0	13.00	13	4.01	1.78
สาระการเรียนรู้ที่ 8 ธรรมชาติของวิทยาศาสตร์และเทคโนโลยี	349,210	10	0	12.00	12	3.87	1.65
ทั้งฉบับ	**349,210**	**90**	**0**	**92.00**	**100**	**30.90**	**9.30**

จากตาราง ผลการวิเคราะห์ผลจากการสอบ พบว่า วิชาคณิตศาสตร์มีคะแนนเฉลี่ยทั้งฉบับเท่ากับ 14.99 คะแนน และเมื่อพิจารณารายสาระการเรียนรู้ พบว่า สาระการเรียนรู้ที่ 4 พีชคณิต มีคะแนนเฉลี่ยสูงที่สุด เท่ากับ 7.35 คะแนน รองลงมา ได้แก่ สาระการเรียนรู้ที่ 5 การวิเคราะห์ข้อมูล

และความน่าจะเป็น เท่ากับ 3.51 คะแนน สาระการเรียนรู้ที่ 1 จำนวนและการดำเนินการ เฉลี่ยเท่ากับ 2.21 คะแนน สาระการเรียนรู้ที่ 2 การวัด เฉลี่ยเท่ากับ 1.10 คะแนน และสาระการเรียนรู้ที่ 3 เรขาคณิต เฉลี่ยเท่ากับ 0.81 คะแนนตามลำดับ

ผลการสอบวิชาวิทยาศาสตร์เฉลี่ยทั้งฉบับ เท่ากับ 30.90 คะแนน และเมื่อพิจารณารายสาระการเรียนรู้ พบว่า สาระการเรียนรู้ที่ 3 สารและสมบัติของสาร มีคะแนนเฉลี่ยสูงที่สุด 4.24 คะแนน รองลงมา ได้แก่ สาระการเรียนรู้ที่ 2 ชีวิตกับสิ่งแวดล้อม เฉลี่ยเท่ากับ 4.24 คะแนน สาระการเรียนรู้ที่ 7 ดาราศาสตร์และอวกาศ เฉลี่ยเท่ากับ 4.01 คะแนน สาระการเรียนรู้ที่ 8 ธรรมชาติของวิทยาศาสตร์และเทคโนโลยี เฉลี่ยเท่ากับ 3.87 คะแนน สาระการเรียนรู้ที่ 6 กระบวนการเปลี่ยนแปลงของโลก เฉลี่ยเท่ากับ 3.70 คะแนน สาระการเรียนรู้ที่ 5 พลังงาน เฉลี่ยเท่ากับ 3.62 คะแนน สาระการเรียนรู้ที่ 1 สิ่งมีชีวิตกับกระบวนการดำรงชีวิต เฉลี่ยเท่ากับ 3.55 คะแนน และสาระการเรียนรู้ที่ 4 แรงและการเคลื่อนที่ เฉลี่ยเท่ากับ 3.42 คะแนน ตามลำดับ

ส่วนที่ 2 ผลการกำหนดคะแนนจุดตัด

2.1 ผลการกำหนดคะแนนจุดตัดวิชาคณิตศาสตร์

2.1.1 ผลการกำหนดคะแนนจุดตัดวิชาคณิตศาสตร์ด้วยวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซี ดังแสดงในต่อไปนี้

ตารางที่ 4.4 ผลการกำหนดคะแนนจุดตัดวิชาคณิตศาสตร์ด้วยวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซี

จุดตัด	วิธีกำหนดจุดตัด	คะแนนจุดตัด							
		รอบที่ 1		รอบที่ 2		รอบที่ 3		เฉลี่ย	
		Score	Theta	Score	Theta	Score	Theta	Score	Theta
ระดับ 1/2	วิธีแองกอฟแบบใช่/ไม่ใช่	20.22	0.14	16.33	0.03	8.78	-0.29	**15.11**	**-0.04**
	วิธีบุ๊คมาร์ค	41.89	0.70	37.89	0.62	37.89	0.62	**39.22**	**0.65**
	วิธีเอบีซี	8.78	-0.25	5.56	-0.48	4.67	-0.57	**6.34**	**-0.43**
	เฉลี่ย	**23.63**	**0.20**	**19.93**	**0.06**	**17.11**	**-0.08**	**20.22**	**0.06**
ระดับ 2/3	วิธีแองกอฟแบบใช่/ไม่ใช่	56.56	0.98	47.33	0.78	24.00	0.31	**42.63**	**0.69**
	วิธีบุ๊คมาร์ค	56.78	0.98	54.67	0.94	54.67	0.94	**55.37**	**0.95**
	วิธีเอบีซี	54.11	0.92	49.33	0.82	48.78	0.81	**50.74**	**0.85**
	เฉลี่ย	**55.82**	**0.96**	**50.44**	**0.85**	**42.48**	**0.69**	**49.58**	**0.83**
ระดับ 3/4	วิธีแองกอฟแบบใช่/ไม่ใช่	82.67	1.59	79.33	1.50	81.78	1.58	**81.26**	**1.56**
	วิธีบุ๊คมาร์ค	72.11	1.30	69.33	1.24	69.33	1.24	**70.26**	**1.26**
	วิธีเอบีซี	84.89	1.86	77.22	1.43	79.00	1.48	**80.37**	**1.59**
	เฉลี่ย	**79.89**	**1.58**	**75.29**	**1.39**	**76.70**	**1.43**	**77.30**	**1.47**

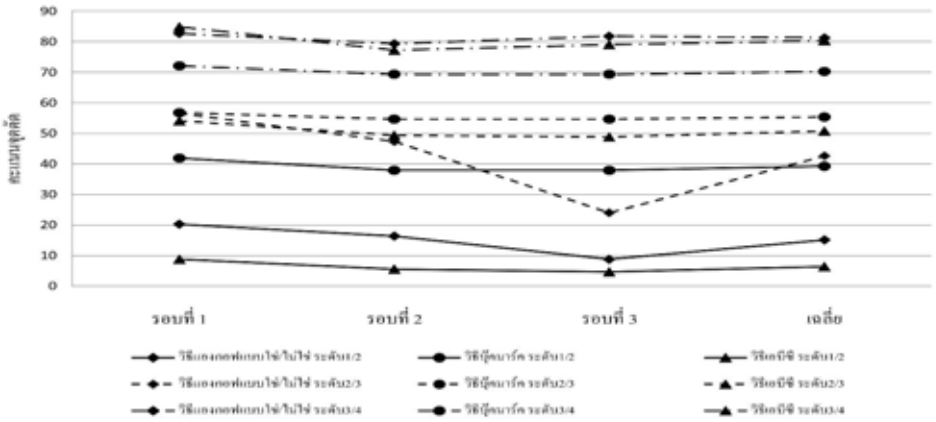
จากตาราง พบว่า คะแนนจุดตัดระดับ 1/2 เฉลี่ยทุกวิธีเท่ากับ 20.22 คะแนน เมื่อพิจารณาแต่ละรอบ พบว่า คะแนนจุดตัดระดับ 1/2 ในรอบที่ 1 เฉลี่ยเท่ากับ 23.63 คะแนน รอบที่ 2 เฉลี่ยเท่ากับ 19.93 คะแนน รอบที่ 3 เฉลี่ยเท่ากับ 17.11 คะแนน สำหรับคะแนนจุดตัดระดับ 2/3 เฉลี่ยทุกวิธีเท่ากับ 49.58 คะแนน เมื่อพิจารณาแต่ละรอบ พบว่า คะแนนจุดตัดระดับ 2/3 ในรอบที่ 1 เฉลี่ย

เท่ากับ 55.82 คะแนน รอบที่ 2 เฉลี่ยเท่ากับ 50.44 คะแนน รอบที่ 3 เฉลี่ยเท่ากับ 42.48 คะแนน ส่วนคะแนนจุดตัดระดับ 3/4 เฉลี่ยทุกวิธีเท่ากับ 77.30 คะแนน เมื่อพิจารณาแต่ละรอบ พบว่า คะแนนจุดตัดระดับ 2/3 ในรอบที่ 1 เฉลี่ยเท่ากับ 79.89 คะแนน รอบที่ 2 เฉลี่ยเท่ากับ 75.29 คะแนน รอบที่ 3 เฉลี่ยเท่ากับ 76.70 คะแนน

คะแนนจุดตัดวิชาคณิตศาสตร์ที่กำหนดด้วยวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซี ระดับ 1/2 เฉลี่ยมีค่าเท่ากับ 15.11, 39.22 และ 6.34 ระดับ 2/3 มีค่า 42.63, 55.37 และ50.74 ระดับ 3/4 มีค่า 81.26, 70.26 และ80.37

เมื่อเปรียบเทียบคะแนนจุดตัดที่กำหนดโดยวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซี พบว่า คะแนนจุดตัดระดับ 1/2 ในแต่ละรอบที่สูงที่สุดได้แก่ คะแนนจุดตัดที่กำหนดโดยวิธีบุ๊คมาร์ค มีค่า 41.89, 37.89 และ 37.89 ตามลำดับ รองลงมาคือคะแนนจุดตัดที่กำหนดโดยวิธีแองกอฟแบบใช่/ไม่ใช่ มีค่า 20.22, 16.33 และ 8.78 ตามลำดับ และคะแนนจุดตัดที่ต่ำที่สุดเป็นคะแนนจุดตัดที่กำหนดโดยวิธีเอบีซี มีค่า 8.78, 5.56 และ 4.67 ตามลำดับ สำหรับคะแนนจุดตัดที่ระดับ 2/3 ในแต่ละรอบที่สูงที่สุด ได้แก่ คะแนนจุดตัดที่กำหนดโดยวิธีบุ๊คมาร์ค มีค่า 56.78, 54.67 และ 54.67 ตามลำดับ รองลงมาคือคะแนนจุดตัดที่กำหนดโดยวิธีเอบีซี มีค่า 54.11, 49.33 และ 48.78 ตามลำดับ และคะแนนจุดตัดที่ต่ำที่สุดเป็นคะแนนจุดตัดที่กำหนดโดยวิธีแองกอฟแบบใช่/ไม่ใช่ มีค่า 54.11, 49.33 และ 48.78 ตามลำดับ ส่วนคะแนนจุดตัดที่ระดับ 3/4 ในแต่ละรอบที่สูงที่สุด ได้แก่ คะแนนจุดตัดที่กำหนดโดยวิธีแองกอฟแบบใช่/ไม่ใช่ มีค่า 82.67, 79.33 และ 81.78 ตามลำดับ รองลงมาคือคะแนนจุดตัดที่กำหนดโดยเอบีซี มีค่า 84.89, 77.22 และ 79.00 ตามลำดับ และคะแนนจุดตัดที่ต่ำที่สุดเป็นคะแนนจุดตัดที่กำหนดโดยวิธีบุ๊คมาร์ค มีค่า 72.11, 69.33 และ 69.33 ตามลำดับ

เพื่อแสดงให้เห็นความแตกต่างของคะแนนจุดตัดแต่ละระดับ ผู้วิจัยขอนำเสนอผลเป็นกราฟเส้น ดังแสดงในภาพต่อไปนี้



ภาพที่ 4.3 ผลการกำหนดคะแนนจุดตัดวิชาคณิตศาสตร์

2.1.2 ผลการกำหนดคะแนนจุดตัดวิชาคณิตศาสตร์ด้วยวิธีแองกอฟแบบใช่/ไม่ใช่ ดังแสดงในตารางต่อไปนี้

ตารางที่ 4.5 ผลการกำหนดคะแนนจุดตัดวิชาคณิตศาสตร์ด้วยวิธีแองกอฟแบบใช่/ไม่ใช่

จุดตัด	ผู้เชี่ยวชาญ	ผลการพิจารณา					
		รอบที่ 1		รอบที่ 2		รอบที่ 3	
		Score	Theta	Score	Theta	Score	Theta
ระดับ1/2	1	7	-0.34	12	-0.06	8	-0.27
	2	44	0.73	38	0.62	8	-0.27
	3	18	0.16	24	0.32	11	-0.01
	4	21	0.24	18	0.16	14	0.02
	5	7	-0.34	7	-0.34	4	-0.63
	6	14	0.02	17	0.12	13	-0.02
	7	16	0.09	16	0.09	11	-0.1
	8	17	0.12	7	-0.34	8	-0.27
	9	38	0.62	8	-0.27	2	-1.02
	เฉลี่ย	**20.22**	**0.14**	**16.33**	**0.03**	**8.78**	**-0.29**
	SD	**12.76**	**0.37**	**9.96**	**0.32**	**3.96**	**0.34**
ระดับ2/3	1	70	1.24	68	1.19	26	0.37
	2	53	0.89	49	0.82	22	0.27
	3	34	0.54	46	0.76	21	0.24
	4	46	0.76	49	0.82	26	0.37
	5	65	1.12	70	1.24	23	0.29
	6	48	0.80	48	0.80	27	0.39
	7	43	0.71	49	0.82	25	0.34
	8	72	1.28	25	0.34	32	0.50
	9	78	1.44	22	0.27	14	0.02
	เฉลี่ย	**56.56**	**0.98**	**47.33**	**0.78**	**24.00**	**0.31**
	SD	**15.17**	**0.30**	**16.17**	**0.32**	**4.95**	**0.13**
ระดับ3/4	1	85	1.67	79	1.47	79	1.47
	2	82	1.57	79	1.47	79	1.47
	3	76	1.39	76	1.39	90	1.86
	4	85	1.67	85	1.67	85	1.67
	5	83	1.60	85	1.67	83	1.60
	6	85	1.67	86	1.71	88	1.78
	7	82	1.57	67	1.17	67	1.17
	8	82	1.57	76	1.39	80	1.50
	9	84	1.63	81	1.53	85	1.67
	เฉลี่ย	**82.67**	**1.59**	**79.33**	**1.50**	**81.78**	**1.58**
	SD	**2.83**	**0.09**	**5.98**	**0.17**	**6.76**	**0.20**

จากตาราง พบว่า คะแนนจุดตัดที่กำหนดด้วยวิธีแองกอฟแบบใช่/ไม่ใช่ ระดับ1/2 รอบที่ 1 เฉลี่ยเท่ากับ 20.22 คะแนน เมื่อพิจารณาเป็นรายบุคคล พบว่า ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดสูงที่สุด มีค่าเท่ากับ 38 คะแนน ส่วนคะแนนจุดตัดต่ำที่สุด มีค่าเท่ากับ 7 คะแนน รอบที่ 2 เฉลี่ยเท่ากับ 16.33คะแนน เมื่อพิจารณาเป็นรายบุคคล พบว่า ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดสูงที่สุด มีค่าเท่ากับ 38 คะแนน ส่วนคะแนนจุดตัดต่ำที่สุด มีค่าเท่ากับ 7 คะแนน และรอบที่ 3 เฉลี่ยเท่ากับ 8.78 คะแนน เมื่อพิจารณาเป็นรายบุคคล พบว่า ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดสูงที่สุด มีค่าเท่ากับ 14 คะแนน ส่วนคะแนนจุดตัดต่ำที่สุด มีค่าเท่ากับ 2 คะแนน

ระดับ2/3 รอบที่ 1 เฉลี่ยเท่ากับ 56.56 คะแนน เมื่อพิจารณาเป็นรายบุคคล พบว่า ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดสูงที่สุด มีค่าเท่ากับ 78 คะแนน ส่วนคะแนนจุดตัดต่ำที่สุด มีค่าเท่ากับ 34 คะแนน รอบที่ 2 เฉลี่ยเท่ากับ 47.33 คะแนน เมื่อพิจารณาเป็นรายบุคคล พบว่า ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดสูงที่สุด มีค่าเท่ากับ 70 คะแนน ส่วนคะแนนจุดตัดต่ำที่สุด มีค่าเท่ากับ 22 คะแนน และรอบที่ 3 เฉลี่ยเท่ากับ 24.00 คะแนน เมื่อพิจารณาเป็นรายบุคคล พบว่า ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดสูงที่สุด มีค่าเท่ากับ 32 คะแนน ส่วนคะแนนจุดตัดต่ำที่สุด มีค่าเท่ากับ 14 คะแนน

ระดับที่ 3/4 รอบที่ 1 เฉลี่ยเท่ากับ 82.67 เมื่อพิจารณาเป็นรายบุคคล พบว่า ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดสูงที่สุด มีค่าเท่ากับ 85 คะแนน ส่วนคะแนนจุดตัดต่ำที่สุด มีค่าเท่ากับ 76 คะแนน รอบที่ 2 เฉลี่ยเท่ากับ 79.33 เมื่อพิจารณาเป็นรายบุคคล พบว่า ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดสูงที่สุด มีค่าเท่ากับ 86คะแนน ส่วนคะแนนจุดตัดต่ำที่สุด มีค่าเท่ากับ 67 คะแนน และรอบที่ 3 เฉลี่ยเท่ากับ 81.78 เมื่อพิจารณาเป็นรายบุคคล พบว่า ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดสูงที่สุด มีค่าเท่ากับ 90 คะแนน ส่วนคะแนนจุดตัดต่ำที่สุด มีค่าเท่ากับ 67 คะแนน

2.1.3 ผลการกำหนดคะแนนจุดตัดวิชาคณิตศาสตร์ด้วยวิธีบุ๊คมาร์ค ดังแสดงในตารางต่อไปนี้

ตารางที่ 4.6 ผลการกำหนดคะแนนจุดตัดวิชาคณิตศาสตร์ด้วยวิธีบุ๊คมาร์ค

จุดตัด	ผู้เชี่ยวชาญ	ผลการพิจารณา					
		รอบที่ 1		รอบที่ 2		รอบที่ 3	
		Score	Theta	Score	Theta	Score	Theta
ระดับ1/2	1	42	0.71	42	0.71	42	0.71
	2	47	0.79	42	0.71	42	0.71
	3	42	0.71	42	0.71	42	0.71
	4	47	0.79	42	0.71	42	0.71
	5	42	0.71	42	0.71	42	0.71
	6	42	0.71	28	0.42	28	0.42
	7	59	1.02	47	0.79	47	0.79
	8	28	0.42	28	0.42	28	0.42
	9	28	0.42	28	0.42	28	0.42
	เฉลี่ย	41.89	0.70	37.89	0.62	37.89	0.62
	SD	9.56	0.19	7.59	0.15	7.59	0.15
ระดับ2/3	1	52	0.89	52	0.89	52	0.89
	2	59	1.02	52	0.89	52	0.89
	3	57	0.98	57	0.98	57	0.98
	4	57	0.98	58	1	58	1
	5	52	0.89	52	0.89	52	0.89
	6	65	1.14	58	1	58	1
	7	65	1.14	59	1.02	59	1.02
	8	52	0.89	52	0.89	52	0.89
	9	52	0.89	52	0.89	52	0.89
	เฉลี่ย	56.78	0.98	54.67	0.94	54.67	0.94
	SD	5.38	0.10	3.20	0.06	3.20	0.06
ระดับ3/4	1	72	1.29	72	1.29	72	1.29
	2	67	1.18	58	1	58	1
	3	72	1.29	72	1.29	72	1.29
	4	67	1.18	60	1.03	60	1.03
	5	72	1.29	72	1.29	72	1.29
	6	81	1.54	72	1.29	72	1.29
	7	74	1.36	74	1.36	74	1.36
	8	72	1.29	72	1.29	72	1.29
	9	72	1.29	72	1.29	72	1.29
	เฉลี่ย	72.11	1.30	69.33	1.24	69.33	1.24
	SD	4.11	0.11	5.92	0.13	5.92	0.13

จากตาราง พบว่า คะแนนจุดตัดที่กำหนดด้วยวิธีบุ๊คมาร์ค ระดับ1/2 รอบที่ 1 เฉลี่ยเท่ากับ 41.89 คะแนน เมื่อพิจารณาเป็นรายบุคคล พบว่า ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดสูงที่สุด มีค่าเท่ากับ 59 คะแนน ส่วนคะแนนจุดตัดต่ำที่สุด มีค่าเท่ากับ 28 คะแนน รอบที่ 2 เฉลี่ยเท่ากับ 37.89 คะแนน เมื่อพิจารณาเป็นรายบุคคล พบว่า ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดสูงที่สุด มีค่าเท่ากับ 47 คะแนน ส่วนคะแนนจุดตัดต่ำที่สุด มีค่าเท่ากับ 28 คะแนน รอบที่ 3 เฉลี่ยเท่ากับ 37.89 คะแนน เมื่อพิจารณาเป็นรายบุคคล พบว่า ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดสูงที่สุด มีค่าเท่ากับ 47 คะแนน ส่วนคะแนนจุดตัดต่ำที่สุด มีค่าเท่ากับ 28 คะแนน

ระดับ 2/3 รอบที่ 1 เฉลี่ยเท่ากับ 56.78 คะแนน เมื่อพิจารณาเป็นรายบุคคล พบว่า ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดสูงที่สุด มีค่าเท่ากับ 65 คะแนน ส่วนคะแนนจุดตัดต่ำที่สุด มีค่าเท่ากับ 52 คะแนน รอบที่ 2 เฉลี่ยเท่ากับ 54.67 คะแนน เมื่อพิจารณาเป็นรายบุคคล พบว่า ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดสูงที่สุด มีค่าเท่ากับ 59 คะแนน ส่วนคะแนนจุดตัดต่ำที่สุด มีค่าเท่ากับ 52 คะแนน รอบที่ 3 เฉลี่ยเท่ากับ 54.67 คะแนน เมื่อพิจารณาเป็นรายบุคคล พบว่า ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดสูงที่สุด มีค่าเท่ากับ 59 คะแนน ส่วนคะแนนจุดตัดต่ำที่สุด มีค่าเท่ากับ 52 คะแนน

ระดับ 3/4 รอบที่ 1 เฉลี่ยเท่ากับ 72.11 คะแนน เมื่อพิจารณาเป็นรายบุคคล พบว่า ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดสูงที่สุด มีค่าเท่ากับ 81 คะแนน ส่วนคะแนนจุดตัดต่ำที่สุด มีค่าเท่ากับ 67 คะแนน รอบที่ 2 เฉลี่ยเท่ากับ 69.33 คะแนน เมื่อพิจารณาเป็นรายบุคคล พบว่า ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดสูงที่สุด มีค่าเท่ากับ 74 คะแนน ส่วนคะแนนจุดตัดต่ำที่สุด มีค่าเท่ากับ 60 คะแนน รอบที่ 3 เฉลี่ยเท่ากับ 69.33 คะแนน เมื่อพิจารณาเป็นรายบุคคล พบว่า ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดสูงที่สุด มีค่าเท่ากับ 74 คะแนน ส่วนคะแนนจุดตัดต่ำที่สุด มีค่าเท่ากับ 60 คะแนน

2.1.4 ผลการกำหนดคะแนนจุดตัดวิชาคณิตศาสตร์ด้วยวิธีเอบีซี ดังแสดงในตารางต่อไปนี้

ตารางที่ 4.7 ผลการกำหนดคะแนนจุดตัดวิชาคณิตศาสตร์ด้วยวิธีเอบีซี

จุดตัด	ผู้เชี่ยวชาญ	ผลการพิจารณา					
		รอบที่ 1		รอบที่ 2		รอบที่ 3	
		Score	Theta	Score	Theta	Score	Theta
ระดับ1/2	1	6	-0.42	6	-0.42	4	-0.63
	2	10	-0.15	4	-0.63	4	-0.63
	3	8	-0.27	7	-0.34	4	-0.63
	4	6	-0.42	4	-0.63	4	-0.63
	5	8	-0.27	4	-0.63	4	-0.63
	6	9	-0.21	7	-0.34	7	-0.34
	7	8	-0.27	7	-0.34	4	-0.63
	8	7	-0.34	7	-0.34	7	-0.34
	9	17	0.12	4	-0.63	4	-0.63
	เฉลี่ย	**8.78**	**-0.25**	**5.56**	**-0.48**	**4.67**	**-0.57**
	SD	**3.35**	**0.16**	**1.51**	**0.15**	**1.32**	**0.13**
ระดับ2/3	1	63	1.08	50	0.83	52	0.87
	2	44	0.73	56	0.95	56	0.95
	3	66	1.15	50	0.83	50	0.83
	4	47	0.78	44	0.73	42	0.69
	5	69	1.21	50	0.83	50	0.83
	6	37	0.6	44	0.73	44	0.73
	7	54	0.91	50	0.83	45	0.74
	8	50	0.83	50	0.83	50	0.83
	9	57	0.96	50	0.83	50	0.83
	เฉลี่ย	**54.11**	**0.92**	**49.33**	**0.82**	**48.78**	**0.81**
	SD	**10.68**	**0.20**	**3.61**	**0.06**	**4.35**	**0.08**
ระดับ3/4	1	94	2.08	78	1.44	82	1.57
	2	54	0.91	76	1.39	87	1.75
	3	96	2.28	69	1.21	80	1.5
	4	83	2.01	82	1.57	77	1.41
	5	95	2.17	81	1.53	81	1.53
	6	79	1.47	73	1.31	75	1.36
	7	85	1.67	80	1.5	73	1.31
	8	80	1.5	73	1.31	73	1.31
	9	98	2.64	83	1.6	83	1.6
	เฉลี่ย	**84.89**	**1.86**	**77.22**	**1.43**	**79.00**	**1.48**
	SD	**13.68**	**0.52**	**4.79**	**0.13**	**4.82**	**0.15**

จากตาราง พบว่า คะแนนจุดตัดที่กำหนดด้วยวิธีเอบีซี ระดับ1/2 รอบที่ 1 เฉลี่ยเท่ากับ 8.78 คะแนน เมื่อพิจารณาเป็นรายบุคคล พบว่า ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดสูงที่สุด มีค่าเท่ากับ 17 คะแนน ส่วนคะแนนจุดตัดต่ำที่สุด มีค่าเท่ากับ 6 คะแนน รอบที่ 2 เฉลี่ยเท่ากับ 5.56 คะแนน เมื่อพิจารณาเป็นรายบุคคล พบว่า ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดสูงที่สุด มีค่าเท่ากับ 7 คะแนน ส่วนคะแนนจุดตัดต่ำที่สุด มีค่าเท่ากับ 4 คะแนน รอบที่ 3 เฉลี่ยเท่ากับ 4.67 คะแนน เมื่อพิจารณาเป็นรายบุคคล พบว่า ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดสูงที่สุด มีค่าเท่ากับ 7 คะแนน ส่วนคะแนนจุดตัดต่ำที่สุด มีค่าเท่ากับ 4 คะแนน

ระดับ 2/3 รอบที่ 1 เฉลี่ยเท่ากับ 54.11 คะแนน เมื่อพิจารณาเป็นรายบุคคล พบว่า ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดสูงที่สุด มีค่าเท่ากับ 69 คะแนน ส่วนคะแนนจุดตัดต่ำที่สุด มีค่าเท่ากับ 37 คะแนน รอบที่ 2 เฉลี่ยเท่ากับ 49.33 คะแนน เมื่อพิจารณาเป็นรายบุคคล พบว่า ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดสูงที่สุด มีค่าเท่ากับ 56 คะแนน ส่วนคะแนนจุดตัดต่ำที่สุด มีค่าเท่ากับ 44 คะแนน รอบที่ 3 เฉลี่ยเท่ากับ 48.78 คะแนน เมื่อพิจารณาเป็นรายบุคคล พบว่า ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดสูงที่สุด มีค่าเท่ากับ 56 คะแนน ส่วนคะแนนจุดตัดต่ำที่สุด มีค่าเท่ากับ 42 คะแนน

ระดับ 3/4 รอบที่ 1 เฉลี่ยเท่ากับ 84.89 คะแนน เมื่อพิจารณาเป็นรายบุคคล พบว่า ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดสูงที่สุด มีค่าเท่ากับ 98 คะแนน ส่วนคะแนนจุดตัดต่ำที่สุด มีค่าเท่ากับ 54 คะแนน รอบที่ 2 เฉลี่ยเท่ากับ 77.22 คะแนน เมื่อพิจารณาเป็นรายบุคคล พบว่า ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดสูงที่สุด มีค่าเท่ากับ 83 คะแนน ส่วนคะแนนจุดตัดต่ำที่สุด มีค่าเท่ากับ 69 คะแนน รอบที่ 3 เฉลี่ยเท่ากับ 79.00 คะแนน เมื่อพิจารณาเป็นรายบุคคล พบว่า ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดสูงที่สุด มีค่าเท่ากับ 87 คะแนน ส่วนคะแนนจุดตัดต่ำที่สุด มีค่าเท่ากับ 73 คะแนน

2.2 ผลการกำหนดคะแนนจุดตัดวิชาวิทยาศาสตร์

การกำหนดคะแนนจุดตัดวิชาวิทยาศาสตร์ ใช้วิธีการกำหนดคะแนนจุดตัด 3 วิธี ประกอบด้วยวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซี โดยมีรายละเอียดดังต่อไปนี้

2.2.1 ผลการกำหนดคะแนนจุดตัดวิชาวิทยาศาสตร์ด้วยวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซี ดังแสดงในตารางต่อไปนี้

ตารางที่ 4.8 ผลการกำหนดคะแนนจุดตัดวิชาวิทยาศาสตร์ด้วยวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซี

จุดตัด	วิธีกำหนดจุดตัด	คะแนนจุดตัด							
		รอบที่ 1		รอบที่ 2		รอบที่ 3		เฉลี่ย	
		Score	Theta	Score	Theta	Score	Theta	Score	Theta
ระดับ 1/2	วิธีแองกอฟแบบใช่/ไม่ใช่	36.28	0.26	47.83	0.59	48.39	0.61	**44.17**	**0.49**
	วิธีบุ๊คมาร์ค	20.56	-0.49	25.00	-0.21	25.00	-0.21	**23.52**	**-0.30**
	วิธีเอบีซี	24.88	-0.18	21.08	-0.31	19.92	-0.36	**21.96**	**-0.28**
	เฉลี่ย	**27.24**	**-0.14**	**31.30**	**0.02**	**31.10**	**0.01**	**29.88**	**-0.03**
ระดับ 2/3	วิธีแองกอฟแบบใช่/ไม่ใช่	67.44	1.09	75.06	1.35	74.83	1.34	**72.44**	**1.26**
	วิธีบุ๊คมาร์ค	54.67	0.77	52.89	0.73	52.89	0.73	**53.48**	**0.74**
	วิธีเอบีซี	68.29	1.09	63.71	0.98	62.54	0.95	**64.85**	**1.01**
	เฉลี่ย	**63.47**	**0.98**	**63.89**	**1.02**	**63.42**	**1.01**	**63.59**	**1.00**
ระดับ 3/4	วิธีแองกอฟแบบใช่/ไม่ใช่	86.94	2.05	87.17	2.03	91.44	2.47	**88.52**	**2.18**
	วิธีบุ๊คมาร์ค	78.56	1.4	76.11	1.33	76.11	1.33	**76.93**	**1.35**
	วิธีเอบีซี	89.25	2.36	84.79	1.93	82.96	1.8	**85.67**	**2.03**
	เฉลี่ย	**84.92**	**1.94**	**82.69**	**1.76**	**83.50**	**1.87**	**83.70**	**1.86**

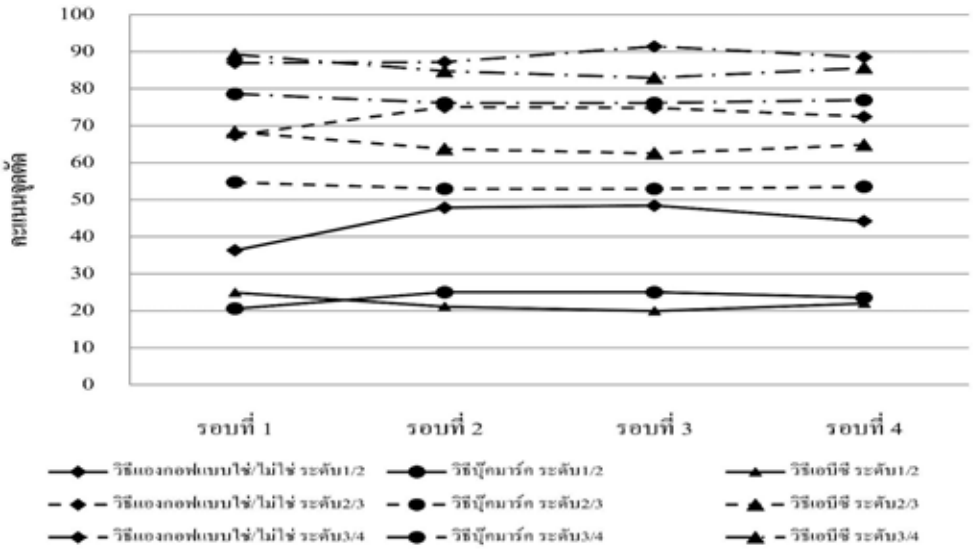
จากตาราง พบว่า คะแนนจุดตัดระดับ 1/2 เฉลี่ยทุกวิธีเท่ากับ 29.88 เมื่อพิจารณาแต่ละรอบ พบว่า คะแนนจุดตัดระดับ 1/2 ในรอบที่ 1 เฉลี่ยเท่ากับ 27.24 รอบที่ 2 เฉลี่ยเท่ากับ 31.30 รอบที่ 3 เฉลี่ยเท่ากับ 31.10 สำหรับคะแนนจุดตัดระดับ 2/3 เฉลี่ยทุกวิธีเท่ากับ 63.59 เมื่อพิจารณาแต่ละรอบ พบว่า คะแนนจุดตัดระดับ 2/3 ในรอบที่ 1 เฉลี่ยเท่ากับ 63.47 รอบที่ 2 เฉลี่ยเท่ากับ 63.89 รอบที่ 3 เฉลี่ยเท่ากับ 63.59 ส่วนคะแนนจุดตัดระดับ 3/4 เฉลี่ยทุกวิธีเท่ากับ 83.70 เมื่อพิจารณาแต่ละรอบ

พบว่า คะแนนจุดตัดระดับ 3/4 ในรอบที่ 1 เฉลี่ยเท่ากับ 79.95 รอบที่ 2 เฉลี่ยเท่ากับ 84.92 รอบที่ 3 เฉลี่ยเท่ากับ 83.50

คะแนนจุดตัดวิชาวิทยาศาสตร์ที่กำหนดด้วยวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซี ระดับ 1/2 เฉลี่ยมีค่า 44.17, 23.52 และ 21.96 ระดับ 2/3 มีค่าเท่ากับ 72.44, 53.48 และ 64.85 ระดับ 3/4 มีค่า 88.52, 76.93 และ 85.67

เมื่อเปรียบเทียบคะแนนจุดตัดที่กำหนดโดยวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซี พบว่า คะแนนจุดตัดระดับ 1/2 ในแต่ละรอบที่สูงที่สุดได้แก่ คะแนนจุดตัดที่กำหนดโดยวิธีแองกอฟแบบใช่/ไม่ใช่ มีค่า 36.28, 47.83 และ 48.39 ตามลำดับ รองลงมาคือคะแนนจุดตัดที่กำหนดโดยวิธีบุ๊คมาร์ค มีค่า 20.56, 25.00 และ 25.00 ตามลำดับ และคะแนนจุดตัดที่ต่ำที่สุดเป็นคะแนนจุดตัดที่กำหนดโดยวิธีเอบีซี มีค่า 24.88, 21.08 และ 19.92 ตามลำดับ สำหรับคะแนนจุดตัดที่ระดับ 2/3 ในแต่ละรอบที่สูงที่สุด ได้แก่ คะแนนจุดตัดที่กำหนดโดยวิธีแองกอฟแบบใช่/ไม่ใช่ มีค่า 67.44, 75.06 และ 74.83 ตามลำดับ รองลงมาคือคะแนนจุดตัดที่กำหนดโดยวิธีเอบีซี มีค่า 68.29, 63.71 และ 62.54 ตามลำดับ และคะแนนจุดตัดที่ต่ำที่สุดเป็นคะแนนจุดตัดที่กำหนดโดยวิธีบุ๊คมาร์ค มีค่า 54.67, 52.89และ 52.89 ตามลำดับ ส่วนคะแนนจุดตัดที่ระดับ 3/4 ในแต่ละรอบที่สูงที่สุด ได้แก่ คะแนนจุดตัดที่กำหนดโดยวิธีแองกอฟแบบใช่/ไม่ใช่ มีค่า 86.94, 87.17 และ 91.44 ตามลำดับ รองลงมาคือคะแนนจุดตัดที่กำหนดโดยวิธีเอบีซี มีค่า 89.25 , 84.79 และ 82.96 ตามลำดับ และคะแนนจุดตัดที่ต่ำที่สุดเป็นคะแนนจุดตัดที่กำหนดโดยวิธีบุ๊คมาร์ค มีค่า 78.56, 76.11 และ 76.11 ตามลำดับ

เพื่อแสดงให้เห็นความแตกต่างของคะแนนจุดตัดแต่ละระดับที่กำหนดโดยวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซี ผู้วิจัยขอนำเสนอผลเป็นกราฟเส้น ดังแสดงในภาพต่อไปนี้



ภาพที่ 4.4 ผลการกำหนดคะแนนจุดตัดวิชาวิทยาศาสตร์

2.1.5 ผลการกำหนดคะแนนจุดตัดวิชาวิทยาศาสตร์ด้วยวิธีแองกอฟแบบใช่/ไม่ใช่ ดังแสดงในตารางต่อไปนี้

ตารางที่ 4.9 ผลการกำหนดคะแนนจุดตัดวิชาวิทยาศาสตร์ด้วยวิธีแองกอฟแบบใช่/ไม่ใช่

จุดตัด	ผู้เชี่ยวชาญ	ผลการพิจารณา					
		รอบที่ 1		รอบที่ 2		รอบที่ 3	
		Score	Theta	Score	Theta	Score	Theta
ระดับ1/2	1	42	0.46	53	0.72	53	0.72
	2	33	0.22	49.5	0.64	49.5	0.64
	3	31	0.16	56	0.79	56	0.79
	4	27.5	0.04	46	0.56	46	0.56
	5	36.5	0.32	52.5	0.71	52.5	0.71
	6	54.5	0.76	54.5	0.76	54.5	0.76
	7	16.5	-0.56	33	0.22	34.5	0.27
	8	32	0.19	32	0.19	35.5	0.29
	9	53.5	0.74	54	0.75	54	0.75
	เฉลี่ย	**36.28**	**0.26**	**47.83**	**0.59**	**48.39**	**0.61**
	SD	**12.18**	**0.40**	**9.18**	**0.23**	**8.15**	**0.20**
ระดับ2/3	1	46	0.56	57	0.82	57	0.82
	2	72.5	1.19	76.5	1.29	76.5	1.31
	3	59.5	0.88	79	1.37	79	1.37
	4	83	1.55	89.5	2.04	87.5	1.88
	5	74	1.22	80	1.41	80	1.41
	6	82.5	1.52	88	1.93	88	1.93
	7	52.5	0.71	66	1.03	66	1.03
	8	60.5	0.90	60.5	0.90	60.5	0.90
	9	76.5	1.29	79	1.37	79	1.37
	เฉลี่ย	**67.44**	**1.09**	**75.06**	**1.35**	**74.83**	**1.34**
	SD	**13.29**	**0.35**	**11.48**	**0.42**	**11.18**	**0.39**
ระดับ3/4	1	80.5	1.43	87.5	1.88	92	2.54
	2	76.5	1.29	75.5	1.27	92	2.54
	3	88	1.93	88	1.93	92	2.54
	4	93.5	2.95	93.5	2.95	90.5	2.26
	5	92.5	2.66	90.5	2.26	89.5	2.11
	6	92	2.54	92	2.54	94	3.14
	7	80.5	1.43	80.5	1.43	92	2.54
	8	90.5	2.26	89.5	2.11	89	2.04
	9	88.5	1.99	87.5	1.88	92	2.54
	เฉลี่ย	**86.94**	**2.05**	**87.17**	**2.03**	**91.44**	**2.47**
	SD	**6.20**	**0.59**	**5.72**	**0.52**	**1.53**	**0.32**

จากตาราง พบว่า คะแนนจุดตัดที่กำหนดด้วยวิธีแองกอฟแบบใช่/ไม่ใช่ ระดับ1/2 รอบที่ 1 เฉลี่ยเท่ากับ 36.28 คะแนน เมื่อพิจารณาเป็นรายบุคคล พบว่า ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดสูงที่สุด มีค่าเท่ากับ 54.5 คะแนน ส่วนคะแนนจุดตัดต่ำที่สุด มีค่าเท่ากับ 27.5 คะแนน รอบที่ 2 เฉลี่ยเท่ากับ 47.83 คะแนน เมื่อพิจารณาเป็นรายบุคคล พบว่า ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดสูงที่สุด มีค่าเท่ากับ 56 คะแนน ส่วนคะแนนจุดตัดต่ำที่สุด มีค่าเท่ากับ 32 คะแนน รอบที่ 3 เฉลี่ยเท่ากับ 48.39 คะแนน เมื่อพิจารณาเป็นรายบุคคล พบว่า ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดสูงที่สุด มีค่าเท่ากับ 56 คะแนน ส่วนคะแนนจุดตัดต่ำที่สุด มีค่าเท่ากับ 34.5 คะแนน

ระดับ 2/3 รอบที่ 1 เฉลี่ยเท่ากับ 67.44 คะแนน เมื่อพิจารณาเป็นรายบุคคล พบว่า ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดสูงที่สุด มีค่าเท่ากับ 83 คะแนน ส่วนคะแนนจุดตัดต่ำที่สุด มีค่าเท่ากับ 46 คะแนน รอบที่ 2 เฉลี่ยเท่ากับ 75.06 คะแนน เมื่อพิจารณาเป็นรายบุคคล พบว่า ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดสูงที่สุด มีค่าเท่ากับ 89.5 คะแนน ส่วนคะแนนจุดตัดต่ำที่สุด มีค่าเท่ากับ 57 คะแนน รอบที่ 3 เฉลี่ยเท่ากับ 74.83 คะแนน เมื่อพิจารณาเป็นรายบุคคล พบว่า ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดสูงที่สุด มีค่าเท่ากับ 87.5 คะแนน ส่วนคะแนนจุดตัดต่ำที่สุด มีค่าเท่ากับ 57 คะแนน

ระดับ 3/4 รอบที่ 1 เฉลี่ยเท่ากับ 86.94 คะแนน เมื่อพิจารณาเป็นรายบุคคล พบว่า ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดสูงที่สุด มีค่าเท่ากับ 93.5 คะแนน ส่วนคะแนนจุดตัดต่ำที่สุด มีค่าเท่ากับ 80.5 คะแนน รอบที่ 2 เฉลี่ยเท่ากับ 87.17 คะแนน เมื่อพิจารณาเป็นรายบุคคล พบว่า ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดสูงที่สุด มีค่าเท่ากับ 93.5 คะแนน ส่วนคะแนนจุดตัดต่ำที่สุด มีค่าเท่ากับ 75.5 คะแนน รอบที่ 3 เฉลี่ยเท่ากับ 91.44 คะแนน เมื่อพิจารณาเป็นรายบุคคล พบว่า ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดสูงที่สุด มีค่าเท่ากับ 94 คะแนน ส่วนคะแนนจุดตัดต่ำที่สุด มีค่าเท่ากับ 89 คะแนน

2.1.6 ผลการกำหนดคะแนนจุดตัดวิชาวิทยาศาสตร์ด้วยวิธีบุ๊คมาร์ค ดังแสดงในตารางต่อไปนี้

ตารางที่ 4.10 ผลการกำหนดคะแนนจุดตัดวิชาวิทยาศาสตร์ด้วยวิธีบุ๊คมาร์ค

จุดตัด	ผู้เชี่ยวชาญ	ผลการพิจารณา					
		รอบที่ 1		รอบที่ 2		รอบที่ 3	
		Score	Theta	Score	Theta	Score	Theta
ระดับ1/2	1	9	-1.33	9	-1.33	9	-1.33
	2	9	-1.31	9	-1.31	9	-1.31
	3	35	0.29	35	0.29	35	0.29
	4	9	-1.31	29	0.08	29	0.08
	5	35	0.29	35	0.29	35	0.29
	6	25	-0.05	29	0.08	29	0.08
	7	25	-0.05	25	-0.05	25	-0.05
	8	13	-0.85	25	-0.05	25	-0.05
	9	25	-0.05	29	0.08	29	0.08
	เฉลี่ย	**20.56**	**-0.49**	**25.00**	**-0.21**	**25.00**	**-0.21**
	SD	**10.81**	**0.71**	**9.75**	**0.64**	**9.75**	**0.64**
ระดับ2/3	1	64	1	64	1	64	1
	2	58	0.84	54	0.76	54	0.76
	3	47	0.59	47	0.59	47	0.59
	4	50	0.67	52	0.71	52	0.71
	5	51	0.68	51	0.68	51	0.68
	6	62	0.95	54	0.76	54	0.76
	7	58	0.84	52	0.71	52	0.71
	8	49	0.64	49	0.64	49	0.64
	9	53	0.73	53	0.73	53	0.73
	เฉลี่ย	**54.67**	**0.77**	**52.89**	**0.73**	**52.89**	**0.73**
	SD	**6.04**	**0.14**	**4.76**	**0.11**	**4.76**	**0.11**
ระดับ3/4	1	83	1.55	83	1.55	83	1.55
	2	74	1.24	54	0.76	54	0.76
	3	83	1.55	83	1.55	83	1.55
	4	84	1.62	83	1.55	83	1.55
	5	83	1.55	83	1.55	83	1.55
	6	75	1.27	74	1.24	74	1.24
	7	75	1.26	75	1.26	75	1.26
	8	75	1.26	75	1.27	75	1.27
	9	75	1.27	75	1.27	75	1.27
	เฉลี่ย	**78.56**	**1.40**	**76.11**	**1.33**	**76.11**	**1.33**
	SD	**4.48**	**0.16**	**9.27**	**0.26**	**9.27**	**0.26**

จากตาราง พบว่า คะแนนจุดตัดที่กำหนดด้วยวิธีบุ๊คมาร์ค ระดับ1/2 รอบที่ 1 เฉลี่ยเท่ากับ 20.56 คะแนน เมื่อพิจารณาเป็นรายบุคคล พบว่า ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดสูงที่สุด มีค่าเท่ากับ 35 คะแนน ส่วนคะแนนจุดตัดต่ำที่สุด มีค่าเท่ากับ 9 คะแนน รอบที่ 2 เฉลี่ยเท่ากับ 25.00 คะแนน เมื่อพิจารณาเป็นรายบุคคล พบว่า ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดสูงที่สุด มีค่าเท่ากับ 35 คะแนน ส่วนคะแนนจุดตัดต่ำที่สุด มีค่าเท่ากับ 9 คะแนน รอบที่ 3 เฉลี่ยเท่ากับ 25.00 คะแนน เมื่อพิจารณาเป็นรายบุคคล พบว่า ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดสูงที่สุด มีค่าเท่ากับ 35 คะแนน ส่วนคะแนนจุดตัดต่ำที่สุด มีค่าเท่ากับ 9 คะแนน ระดับ 2/3 รอบที่ 1 เฉลี่ยเท่ากับ 54.67 คะแนน เมื่อพิจารณาเป็นรายบุคคล พบว่า ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดสูงที่สุด มีค่าเท่ากับ 64 คะแนน ส่วนคะแนนจุดตัดต่ำที่สุด มีค่าเท่ากับ 47 คะแนน รอบที่ 2 เฉลี่ยเท่ากับ 52.89 คะแนน เมื่อพิจารณาเป็นรายบุคคล พบว่า ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดสูงที่สุด มีค่าเท่ากับ 64 คะแนน ส่วนคะแนนจุดตัดต่ำที่สุด มีค่าเท่ากับ 47 คะแนน รอบที่ 3 เฉลี่ยเท่ากับ 52.89 คะแนน เมื่อพิจารณาเป็นรายบุคคล พบว่า ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดสูงที่สุด มีค่าเท่ากับ 64 คะแนน ส่วนคะแนนจุดตัดต่ำที่สุด มีค่าเท่ากับ 47 คะแนน ระดับ 3/4 รอบที่ 1 เฉลี่ยเท่ากับ 78.56 คะแนน เมื่อพิจารณาเป็นรายบุคคล พบว่า ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดสูงที่สุด มีค่าเท่ากับ 83 คะแนน ส่วนคะแนนจุดตัดต่ำที่สุด มีค่าเท่ากับ 74 คะแนน รอบที่ 2 เฉลี่ยเท่ากับ 76.11 คะแนน เมื่อพิจารณาเป็นรายบุคคล พบว่า ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดสูงที่สุด มีค่าเท่ากับ 83 คะแนน ส่วนคะแนนจุดตัดต่ำที่สุด มีค่าเท่ากับ 54 คะแนน รอบที่ 3 เฉลี่ยเท่ากับ 76.11 คะแนน เมื่อพิจารณาเป็นรายบุคคล พบว่า ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดสูงที่สุด มีค่าเท่ากับ 83 คะแนน ส่วนคะแนนจุดตัดต่ำที่สุด มีค่าเท่ากับ 54 คะแนน

2.1.7 ผลการกำหนดคะแนนจุดตัดวิชาวิทยาศาสตร์ด้วยวิธีเอบีซี ดังแสดงในตารางต่อไปนี้

ตารางที่ 4.11 ผลการกำหนดคะแนนจุดตัดวิชาวิทยาศาสตร์ด้วยวิธีเอบีซี

จุดตัด	ผู้เชี่ยวชาญ	ผลการพิจารณา					
		รอบที่ 1		รอบที่ 2		รอบที่ 3	
		Score	Theta	Score	Theta	Score	Theta
ระดับ1/2	1	21.5	-0.24	19	-0.38	19	-0.38
	2	30	0.13	21	-0.26	19	-0.38
	3	28	0.06	21	-0.26	26.5	0
	4	11	-1.1	16.5	-0.56	21	-0.26
	5	36	0.31	31	0.16	15.5	-0.64
	6	36.5	0.32	22.5	-0.18	20.5	-0.29
	7	28	0.06	24.5	-0.09	21	-0.26
	8	14	-0.78	18	-0.45	24.5	-0.09
	9	21.5	-0.24	22.5	-0.18	24.5	-0.09
	10	18.5	-0.42	18.5	-0.42	18	-0.45
	11	38	0.36	27.5	0.04	18.5	-0.42
	12	15.5	-0.64	11	-1.1	11	-1.1
	เฉลี่ย	**24.88**	**-0.18**	**21.08**	**-0.31**	**19.92**	**-0.36**
	SD	**9.24**	**0.47**	**5.21**	**0.32**	**4.20**	**0.29**
ระดับ2/3	1	54.5	0.76	52	0.7	52	0.7
	2	75	1.25	66	1.03	64	0.98
	3	72.5	1.19	65.5	1.02	71	1.15
	4	56	0.79	61.5	0.92	66	1.03
	5	74	1.22	69	1.1	53.5	0.74
	6	79.5	1.39	65.5	1.02	63.5	0.97
	7	78.5	1.36	75	1.25	71.5	1.16
	8	76.5	1.29	71	1.15	77.5	1.32
	9	54.5	0.76	55.5	0.78	57.5	0.83
	10	73.5	1.21	73.5	1.21	73	1.2
	11	76	1.28	65.5	1.02	56.5	0.81
	12	49	0.63	44.5	0.53	44.5	0.53
	เฉลี่ย	**68.29**	**1.09**	**63.71**	**0.98**	**62.54**	**0.95**
	SD	**11.21**	**0.27**	**9.03**	**0.21**	**9.91**	**0.24**

ตารางที่ 4.11 ผลการกำหนดคะแนนจุดตัดวิชาวิทยาศาสตร์ด้วยวิธีเอบีซี

จุดตัด	ผู้เชี่ยวชาญ	ผลการพิจารณา					
		รอบที่ 1		รอบที่ 2		รอบที่ 3	
		Score	Theta	Score	Theta	Score	Theta
ระดับ3/4	1	80.5	1.43	75	1.25	75	1.25
	2	91	2.34	87	1.83	85	1.67
	3	83.5	1.58	86.5	1.79	92	2.54
	4	90	2.18	83.5	1.58	88	1.93
	5	93	2.79	87	1.83	71.5	1.16
	6	92.5	2.66	84.5	1.64	82.5	1.52
	7	95	3.65	92	2.54	88.5	1.52
	8	92.5	2.66	94	3.14	94.5	3.37
	9	80.5	1.43	77.5	1.32	78.5	1.36
	10	92.5	2.66	94.5	3.37	93	2.79
	11	94	3.14	85.5	1.71	76.5	1.29
	12	86	1.79	70.5	1.14	70.5	1.14
	เฉลี่ย	**89.25**	**2.36**	**84.79**	**1.93**	**82.96**	**1.80**
	SD	**5.24**	**0.70**	**7.38**	**0.72**	**8.46**	**0.73**

จากตาราง พบว่า คะแนนจุดตัดที่กำหนดด้วยวิธีเอบีซี ระดับ1/2 รอบที่ 1 เฉลี่ยเท่ากับ 24.88 คะแนน เมื่อพิจารณาเป็นรายบุคคล พบว่า ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดสูงที่สุด มีค่าเท่ากับ 38 คะแนน ส่วนคะแนนจุดตัดต่ำที่สุด มีค่าเท่ากับ 11 คะแนน รอบที่ 2 เฉลี่ยเท่ากับ 21.08 คะแนน เมื่อพิจารณาเป็นรายบุคคล พบว่า ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดสูงที่สุด มีค่าเท่ากับ 27.5 คะแนน ส่วนคะแนนจุดตัดต่ำที่สุด มีค่าเท่ากับ 11 คะแนน รอบที่ 3 เฉลี่ยเท่ากับ 19.92 คะแนน เมื่อพิจารณาเป็นรายบุคคล พบว่า ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดสูงที่สุด มีค่าเท่ากับ 26.5 คะแนน ส่วนคะแนนจุดตัดต่ำที่สุด มีค่าเท่ากับ 11 คะแนน

ระดับ 2/3 รอบที่ 1 เฉลี่ยเท่ากับ 68.29 คะแนน เมื่อพิจารณาเป็นรายบุคคล พบว่า ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดสูงที่สุด มีค่าเท่ากับ 79.5 คะแนน ส่วนคะแนนจุดตัดต่ำที่สุด มีค่าเท่ากับ 49 คะแนน รอบที่ 2 เฉลี่ยเท่ากับ 63.71 คะแนน เมื่อพิจารณาเป็นรายบุคคล พบว่า ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดสูงที่สุด มีค่าเท่ากับ 75 คะแนน ส่วนคะแนนจุดตัดต่ำที่สุด มีค่าเท่ากับ 44.5 คะแนน รอบที่ 3 เฉลี่ยเท่ากับ 62.54 คะแนน เมื่อพิจารณาเป็นรายบุคคล พบว่า

ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดสูงที่สุด มีค่าเท่ากับ 71 คะแนน ส่วนคะแนนจุดตัดต่ำที่สุด มีค่าเท่ากับ 44.5 คะแนน

ระดับ 3/4 รอบที่ 1 เฉลี่ยเท่ากับ 89.25 คะแนน เมื่อพิจารณาเป็นรายบุคคล พบว่าผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดสูงที่สุด มีค่าเท่ากับ 95 คะแนน ส่วนคะแนนจุดตัดต่ำที่สุด มีค่าเท่ากับ 86 คะแนน รอบที่ 2 เฉลี่ยเท่ากับ 84.79 คะแนน เมื่อพิจารณาเป็นรายบุคคล พบว่าผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดสูงที่สุด มีค่าเท่ากับ 94.5 คะแนน ส่วนคะแนนจุดตัดต่ำที่สุด มีค่าเท่ากับ 70.5 คะแนน รอบที่ 3 เฉลี่ยเท่ากับ 82.96 คะแนน เมื่อพิจารณาเป็นรายบุคคล พบว่าผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดสูงที่สุด มีค่าเท่ากับ 93 คะแนน ส่วนคะแนนจุดตัดต่ำที่สุด มีค่าเท่ากับ 70.5 คะแนน

ตอนที่ 4 ผลการวิเคราะห์คุณภาพของคะแนนจุดตัด

ผู้วิจัยขอนำเสนอผลการวิเคราะห์คุณภาพของคะแนนจุดตัด ตามลำดับดังนี้

4.1 ผลการวิเคราะห์คุณภาพด้านความตรงของคะแนนจุดตัด ประกอบด้วย

4.1.1 ผลการวิเคราะห์ความตรงของปัจจัยนำเข้า

4.1.2 ผลการวิเคราะห์ความตรงของกระบวนการตัดสินใจ

4.1.3 ผลการวิเคราะห์ความตรงของผลลัพธ์

4.1.4 ผลการวิเคราะห์ความตรงของผลที่ตามมา

4.2 ผลการวิเคราะห์คุณภาพด้านความเที่ยงของคะแนนจุดตัด ประกอบด้วย

4.2.1 ผลการวิเคราะห์ความเที่ยงซึ่งแทนด้วยค่าสัมประสิทธิ์การสรุปอ้างอิง และวิเคราะห์จำนวนผู้เชี่ยวชาญ และจำนวนกลุ่มผู้เชี่ยวชาญ ที่มีความเหมาะสม

4.2.2 ผลการวิเคราะห์ความเที่ยงภายในผู้เชี่ยวชาญและความเที่ยงระหว่างผู้เชี่ยวชาญ

4.1 ผลการวิเคราะห์ความตรงของคะแนนจุดตัด มีรายละเอียดดังต่อไปนี้

4.1.1 ผลการวิเคราะห์ความตรงของปัจจัยนำเข้า รายละเอียดดังแสดงในตารางต่อไปนี้

ตารางที่ 4.12 ผลการวิเคราะห์ความตรงของปัจจัยนำเข้าของการกำหนดคะแนนจุดตัดวิชาคณิตศาสตร์

	ข้อความ	ระดับความเห็นด้วย								
		วิธีแองกอฟแบบใช่/ไม่ใช่			วิธีบุ๊คมาร์ค			วิธีเอบีซี		
		M	SD	ระดับ	*M*	SD	ระดับ	*M*	SD	ระดับ
	ด้านความถูกต้อง									
1.	การกำหนดคะแนนจุดครั้งนี้มีวิธีดำเนินการที่ถูกต้อง	3.56	1.01	มาก	3.78	0.44	มาก	3.67	0.71	มาก
2.	คำอธิบายระดับความสามารถมีเนื้อหาที่ถูกต้อง สื่อความหมายชัดเจน	3.56	0.73	มาก	3.78	0.44	มาก	3.56	0.73	มาก
3.	สารสนเทศที่ให้เพื่อประกอบการกำหนดคะแนนจุดตัดมีเนื้อหาที่ถูกต้อง สื่อความหมายชัดเจน	3.78	0.83	มาก	4.00	0.50	มาก	4.00	0.87	มาก
	รวม	3.63	0.73	มาก	3.85	0.24	มาก	3.74	0.68	มาก
	ด้านความเป็นไปได้									
4.	ขั้นตอนการกำหนดคะแนนจุดตัดสามารถปฏิบัติได้จริง	3.67	0.71	มาก	3.56	0.73	มาก	3.67	1.00	มาก
5.	ให้เวลาสำหรับการกำหนดคะแนนจุดตัดอย่างเพียงพอ	3.67	0.87	มาก	3.89	0.60	มาก	3.78	0.83	มาก
6.	ผู้เชี่ยวชาญทุกท่านมีโอกาสแสดงความคิดเห็นได้อย่างเสมอภาค	3.11	0.60	ปานกลาง	3.56	0.73	มาก	3.78	0.97	มาก
7.	ขั้นตอนการกำหนดคะแนนจุดตัดไม่ยุ่งยากซับซ้อน	3.78	0.83	มาก	4.00	0.87	มาก	3.67	1.00	มาก
8.	การอภิปรายร่วมกันระหว่างผู้เชี่ยวชาญช่วยให้กำหนดคะแนนจุดตัดได้ง่ายขึ้น	3.22	0.67	ปานกลาง	3.44	0.53	ปานกลาง	3.67	0.87	มาก
9.	ให้เวลาสำหรับการอภิปรายกลุ่มอย่างเพียงพอ	3.78	0.44	มาก	3.67	0.87	มาก	3.56	0.73	มาก
	รวม	3.54	0.36	มาก	3.69	0.40	มาก	3.69	0.72	มาก

ตารางที่ 4.12 ผลการวิเคราะห์ความตรงของปัจจัยนำเข้าของการกำหนดคะแนนจุดตัดวิชาคณิตศาสตร์

	ข้อความ	ระดับความเห็นด้วย								
		วิธีแองกอฟแบบใช่/ไม่ใช่			วิธีบุ๊คมาร์ค			วิธีเอบีซี		
		M	SD	ระดับ	*M*	SD	ระดับ	*M*	SD	ระดับ
	ด้านความเหมาะสม									
10.	เป้าหมายและการดำเนินการกำหนดคะแนนจุดตัดมีความสอดคล้องและเชื่อมโยงกัน	3.56	0.73	มาก	4.11	0.33	มาก	3.67	1.00	มาก
11.	เอกสารประกอบการดำเนินการสอดคล้องและเชื่อมโยงกันกับเป้าหมายของการกำหนดคะแนนจุดตัด	3.56	0.73	มาก	3.78	0.67	มาก	3.56	1.01	มาก
	รวม	3.56	0.46	มาก	3.94	0.17	มาก	3.61	0.86	มาก
	ด้านความเข้าใจและความมั่นใจของผู้เชี่ยวชาญ									
12.	ฉันเข้าใจความหมายของผู้มีความสามารถคาบเส้นเป็นอย่างดี	3.67	0.71	มาก	4.11	0.33	มาก	3.89	0.60	มาก
13.	ฉันเข้าใจถึงเป้าหมายของการกำหนดคะแนนจุดตัดเป็นอย่างดี	3.56	0.73	มาก	3.78	0.83	มาก	3.89	0.78	มาก
14.	ฉันเข้าใจถึงความแตกต่างระหว่างผู้ที่มีความสามารถที่แตกต่างกัน	3.56	0.53	มาก	3.89	0.78	มาก	3.67	0.87	มาก
15.	ฉันเข้าใจวิธีการและขั้นตอนในการกำหนดคะแนนจุดตัดเป็นอย่างดี	3.67	0.71	มาก	3.56	0.88	มาก	3.78	1.09	มาก
16.	ฉันไม่รู้สึกกดดันเมื่อคะแนนจุดตัดที่กำหนดไม่ใกล้เคียงกับผู้เชี่ยวชาญท่านอื่น ๆ	3.78	0.44	มาก	3.44	0.73	ปานกลาง	3.67	0.71	มาก
17.	ฉันมั่นใจว่าคะแนนจุดตัดที่กำหนดมีความเหมาะสม	3.56	1.13	มาก	3.67	0.87	มาก	3.67	1.22	มาก
18.	ฉันเห็นความสำคัญของการกำหนดคะแนนจุดตัดครั้งนี้	3.78	0.83	มาก	4.11	0.33	มาก	3.89	0.93	มาก
	รวม	3.65	0.31	มาก	3.79	0.36	มาก	3.78	0.55	มาก

ตารางที่ 4.12 ผลการวิเคราะห์ความตรงของปัจจัยนำเข้าของการกำหนดคะแนนจุดตัดวิชาคณิตศาสตร์

	ข้อความ	ระดับความเห็นด้วย								
		วิธีแองกอฟแบบใช่/ไม่ใช่			วิธีบุ๊คมาร์ค			วิธีเอบีซี		
		M	SD	ระดับ	*M*	SD	ระดับ	*M*	SD	ระดับ
	ด้านความเป็นประโยชน์ของเอกสาร									
19.	รายงานผลการกำหนดคะแนนจุดตัดในแต่ละรอบเป็นประโยชน์ต่อการกำหนดคะแนนจุดตัดในรอบต่อ ๆ ไป	3.44	0.73	ปานกลาง	3.44	0.88	ปานกลาง	3.33	0.71	ปานกลาง
20.	สารสนเทศที่ให้เพื่อประกอบการกำหนดคะแนนจุดตัดเป็นประโยชน์ต่อการกำหนดคะแนนจุดตัดอย่างยิ่ง	3.33	0.87	ปานกลาง	3.89	0.78	มาก	4.00	1.00	มาก
	รวม	3.39	0.65	ปานกลาง	3.67	0.66	มาก	3.67	0.71	มาก
	รวมทั้งฉบับ	**3.58**	**0.38**	**มาก**	**3.77**	**0.28**	**มาก**	**3.72**	**0.58**	**มาก**

จากตาราง พบว่า

1. ผู้เชี่ยวชาญมีความคิดเห็นว่าการกำหนดคะแนนจุดตัดด้วยวิธีแองกอฟแบบใช่/ไม่ใช่มีความตรงของปัจจัยนำเข้าโดยภาพรวมอยู่ในระดับมาก (M = 3.58) เมื่อพิจารณาเป็นรายด้านพบว่าปัจจัยนำเข้าในแต่ละด้านมีค่าต่างกัน โดยสามารถเรียงตามลำดับได้เป็นดังนี้

1.1 ด้านความเข้าใจและความมั่นใจของผู้เชี่ยวชาญสูงที่สุด ซึ่งอยู่ในระดับมาก (M = 3.65) โดยรายการที่มีระดับสูงที่สุดได้แก่ ผู้เชี่ยวชาญไม่รู้สึกกดดันเมื่อคะแนนจุดตัดที่กำหนดไม่ใกล้เคียงกับผู้เชี่ยวชาญท่านอื่น ๆและผู้เชี่ยวชาญเห็นความสำคัญของการกำหนดคะแนนจุดตัดครั้งนี้ ซึ่งอยู่ในระดับมาก (M = 3.78) รองลงมาคือผู้เชี่ยวชาญข้าใจความหมายของผู้มีความสามารถคาบเส้นเป็นอย่างดี และผู้เชี่ยวชาญเข้าใจวิธีการและขั้นตอนในการกำหนดคะแนนจุดตัดเป็นอย่างดีอยู่ในระดับมาก (M = 3.67) ส่วนผู้เชี่ยวชาญเข้าใจถึงเป้าหมายของการกำหนดคะแนนจุดตัดเป็นอย่างดี ผู้เชี่ยวชาญเข้าใจถึงความแตกต่างระหว่างผู้ที่มีความสามารถที่แตกต่างกัน และผู้เชี่ยวชาญมั่นใจว่าคะแนนจุดตัดที่กำหนดมีความเหมาะสม อยู่ในระดับมาก (M = 3.56) ตามลำดับ

1.2 ด้านความถูกต้อง อยู่ในระดับมาก (M = 3.63) รายการที่มีระดับสูงที่สุดได้แก่ สารสนเทศที่ให้เพื่อประกอบการกำหนดคะแนนจุดตัดมีเนื้อหาที่ถูกต้องสื่อความหมายชัดเจน อยู่ในระดับมาก (M = 3.78) รองลงมาคือการกำหนดคะแนนจุดครั้งนี้มีวิธีดำเนินการที่ถูกต้อง และคำอธิบายระดับความสามารถมีเนื้อหาที่ถูกต้อง สื่อความหมายชัดเจน อยู่ในระดับมาก (M = 3.56) ตามลำดับ

1.3 ด้านความเหมาะสม อยู่ในระดับมาก (M = 3.56) โดยรายการที่มีระดับสูงที่สุดได้แก่ เป้าหมายและการดำเนินการกำหนดคะแนนจุดตัดมีความสอดคล้องและเชื่อมโยงกัน และเอกสารประกอบการดำเนินการสอดคล้องและเชื่อมโยงกันกับเป้าหมายของการกำหนดคะแนนจุดตัด อยู่ในระดับมาก (M = 3.56)

1.4 ด้านความเป็นไปได้ อยู่ในระดับมาก (M = 3.54) โดยรายการที่มีระดับสูงที่สุดได้แก่ ขั้นตอนการกำหนดคะแนนจุดตัดไม่ยุ่งยากซับซ้อนและให้เวลาสำหรับการอภิปรายกลุ่มอย่างเพียงพอ อยู่ในระดับมาก (M = 3.78) รองลงมาคือขั้นตอนการกำหนดคะแนนจุดตัดสามารถปฏิบัติได้จริง และให้เวลาสำหรับการกำหนดคะแนนจุดตัดอย่างเพียงพออยู่ในระดับมาก (M = 3.67) การอภิปรายร่วมกันระหว่างผู้เชี่ยวชาญช่วยให้กำหนดคะแนนจุดตัดได้ง่ายขึ้น อยู่ในระดับมาก

(M =3.22) และผู้เชี่ยวชาญทุกท่านมีโอกาสแสดงความคิดเห็นได้อย่างเสมอภาค อยู่ในระดับมาก (M = 3.11) ตามลำดับ

1.5 ด้านความเป็นประโยชน์ของเอกสาร อยู่ในระดับที่ต่ำที่สุด โดยอยู่ในระดับปานกลาง (M = 3.39) โดยรายการที่มีระดับสูงที่สุดได้แก่ รายงานผลการกำหนดคะแนนจุดตัดในแต่ละรอบเป็นประโยชน์ต่อการกำหนดคะแนนจุดตัดในรอบต่อ ๆ ไป อยู่ในระดับปานกลาง (M = 3.44) รองลงมาคือสารสนเทศที่ให้เพื่อประกอบการกำหนดคะแนนจุดตัดเป็นประโยชน์ต่อการกำหนดคะแนนจุดตัดอย่างยิ่ง อยู่ในระดับปานกลาง (M = 3.33) ตามลำดับ

2. ผู้เชี่ยวชาญมีความคิดเห็นว่าการกำหนดคะแนนจุดตัดด้วยวิธีบุ๊คมาร์คมีความตรงของปัจจัยนำเข้าโดยภาพรวมอยู่ในระดับมาก (M = 3.77) เมื่อพิจารณาเป็นรายด้านพบว่า ปัจจัยนำเข้าในแต่ละด้านมีค่าต่างกัน โดยสามารถเรียงตามลำดับได้เป็นดังนี้

2.1 ด้านความเหมาะสมสูงที่สุด ซึ่งอยู่ในระดับมาก (M = 3.94) โดยรายการที่มีระดับสูงที่สุดได้แก่ เป้าหมายและการดำเนินการกำหนดคะแนนจุดตัดมีความสอดคล้องและเชื่อมโยงกัน อยู่ในระดับมาก (M = 4.11) รองลงมาคือเอกสารประกอบการดำเนินการสอดคล้องและเชื่อมโยงกันกับเป้าหมายของการกำหนดคะแนนจุดตัด อยู่ในระดับมาก (M = 3.78) ตามลำดับ

2.2 ด้านความถูกต้อง อยู่ในระดับมาก (M = 3.85) โดยรายการที่มีระดับสูงที่สุดได้แก่ สารสนเทศที่ให้เพื่อประกอบการกำหนดคะแนนจุดตัดมีเนื้อหาที่ถูกต้อง สื่อความหมายชัดเจน อยู่ในระดับมาก (M = 4.00) รองลงมาคือการกำหนดคะแนนจุดครั้งนี้มีวิธีดำเนินการที่ถูกต้อง และคำอธิบายระดับความสามารถมีเนื้อหาที่ถูกต้อง สื่อความหมายชัดเจน อยู่ในระดับมาก (M = 3.78)

2.3 ด้านความเข้าใจและความมั่นใจของผู้เชี่ยวชาญ อยู่ในระดับมาก (M = 3.79) โดยรายการที่มีระดับสูงที่สุดได้แก่ ผู้เชี่ยวชาญเข้าใจความหมายของผู้มีความสามารถคาบเส้นเป็นอย่างดี และผู้เห็นเชี่ยวชาญความสำคัญของการกำหนดคะแนนจุดตัดครั้งนี้ อยู่ในระดับมาก (M = 4.11) รองลงมาคือผู้เชี่ยวชาญเข้าใจถึงความแตกต่างระหว่างผู้ที่มีความสามารถที่แตกต่างกัน อยู่ในระดับมาก (M = 3.89) ผู้เชี่ยวชาญเข้าใจถึงเป้าหมายของการกำหนดคะแนนจุดตัดเป็นอย่างดี อยู่ในระดับมาก (M = 3.78) ผู้เชี่ยวชาญมั่นใจว่าคะแนนจุดตัดที่กำหนดมีความเหมาะสม อยู่ในระดับมาก (M = 3.67) ผู้เชี่ยวชาญเข้าใจวิธีการและขั้นตอนในการกำหนดคะแนนจุดตัดเป็นอย่างดี อยู่ในระดับ

มาก (M = 3.56) และผู้เชี่ยวชาญไม่รู้สึกกดดันเมื่อคะแนนจุดตัดที่กำหนดไม่ใกล้เคียงกับผู้เชี่ยวชาญท่านอื่น ๆ อยู่ในระดับปานกลาง (M = 3.44) ตามลำดับ

2.4 ด้านความเป็นไปได้ อยู่ในระดับมาก (M = 3.69) โดยรายการที่มีระดับสูงที่สุดได้แก่ ขั้นตอนการกำหนดคะแนนจุดตัดไม่ยุ่งยากซับซ้อน อยู่ในระดับมาก (M = 4.00) รองลงมาคือให้เวลาสำหรับการกำหนดคะแนนจุดตัดอย่างเพียงพอ อยู่ในระดับมาก (M = 3.89) ให้เวลาสำหรับการอภิปรายกลุ่มอย่างเพียงพอ อยู่ในระดับมาก (M = 3.67) ขั้นตอนการกำหนดคะแนนจุดตัดสามารถปฏิบัติได้จริง และผู้เชี่ยวชาญทุกท่านมีโอกาสแสดงความคิดเห็นได้อย่างเสมอภาค อยู่ในระดับมาก (M = 3.56) และการอภิปรายร่วมกันระหว่างผู้เชี่ยวชาญช่วยให้กำหนดคะแนนจุดตัดได้ง่ายขึ้น อยู่ในระดับปานกลาง (M = 3.44)

2.5 ด้านความเป็นประโยชน์ของเอกสาร อยู่ในระดับที่ต่ำที่สุด โดยอยู่ในระดับมาก (M = 3.67) โดยรายการที่มีระดับสูงที่สุดได้แก่ สารสนเทศที่ให้เพื่อประกอบการกำหนดคะแนนจุดตัดเป็นประโยชน์ต่อการกำหนดคะแนนจุดตัดอย่างยิ่ง อยู่ในระดับมาก (M = 3.89) รองลงมาคือรายงานผลการกำหนดคะแนนจุดตัดในแต่ละรอบเป็นประโยชน์ต่อการกำหนดคะแนนจุดตัดในรอบต่อ ๆ ไป อยู่ในระดับปานกลาง (M = 3.44) ตามลำดับ

3. ผู้เชี่ยวชาญมีความคิดเห็นว่าการกำหนดคะแนนจุดตัดด้วยวิธีเอบีซีมีความตรงของปัจจัยนำเข้าโดยภาพรวมอยู่ในระดับมาก (M = 3.72) เมื่อพิจารณาเป็นรายด้านพบว่า ปัจจัยนำเข้าในแต่ละด้านมีค่าต่างกัน โดยสามารถเรียงตามลำดับได้เป็นดังนี้

3.1 ด้านความเข้าใจและความมั่นใจของผู้เชี่ยวชาญสูงที่สุด ซึ่งอยู่ในระดับมาก (M = 3.78) โดยรายการที่มีระดับสูงที่สุดได้แก่ ผู้เชี่ยวชาญเข้าใจความหมายของผู้มีความสามารถคาบเส้นเป็นอย่างดี ผู้เชี่ยวชาญเข้าใจถึงเป้าหมายของการกำหนดคะแนนจุดตัดเป็นอย่างดี และเห็นความสำคัญของการกำหนดคะแนนจุดตัดครั้งนี้อยู่ในระดับมาก (M = 3.89) รองลงมาคือผู้เชี่ยวชาญเข้าใจวิธีการและขั้นตอนในการกำหนดคะแนนจุดตัดเป็นอย่างดี อยู่ในระดับมาก (M = 3.78) ผู้เชี่ยวชาญเข้าใจถึงความแตกต่างระหว่างผู้ที่มีความสามารถที่แตกต่างกัน ผู้เชี่ยวชาญไม่รู้สึกกดดันเมื่อคะแนนจุดตัดที่กำหนดไม่ใกล้เคียงกับผู้เชี่ยวชาญท่านอื่น ๆ และผู้เชี่ยวชาญมั่นใจว่าคะแนนจุดตัดที่กำหนดมีความเหมาะสม อยู่ในระดับมาก (M = 3.67)

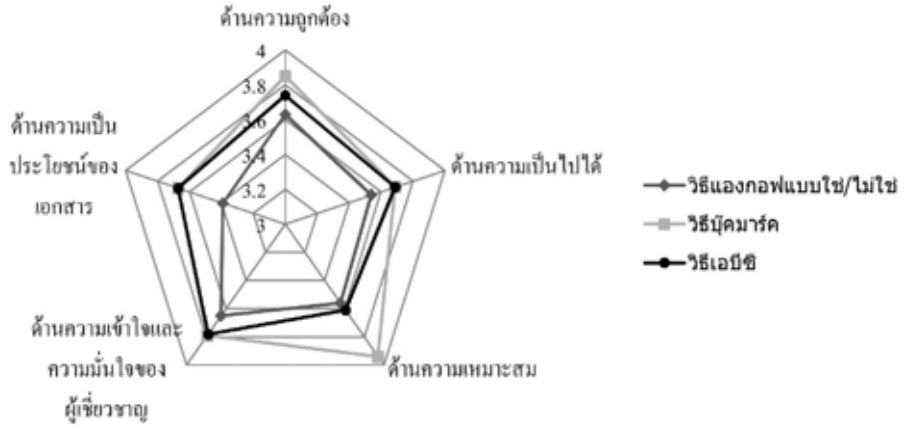
3.2 ด้านความถูกต้อง อยู่ในระดับมาก (M = 3.74) โดยรายการที่มีระดับสูงที่สุดได้แก่ สารสนเทศที่ให้เพื่อประกอบการกำหนดคะแนนจุดตัดมีเนื้อหาที่ถูกต้อง สื่อความหมายชัดเจน อยู่ในระดับมาก (M = 4.00) รองลงมาคือการกำหนดคะแนนจุดครั้งนี้มีวิธีดำเนินการที่ถูกต้อง อยู่ในระดับมาก (M = 3.67) คำอธิบายระดับความสามารถมีเนื้อหาที่ถูกต้อง สื่อความหมายชัดเจน อยู่ในระดับมาก (M = 3.56)

3.3 ด้านความเป็นไปได้ อยู่ในระดับมาก (M = 3.69) รายการที่มีระดับสูงที่สุดได้แก่ ให้เวลาสำหรับการกำหนดคะแนนจุดตัดอย่างเพียงพอและผู้เชี่ยวชาญทุกท่านมีโอกาสแสดงความคิดเห็นได้อย่างเสมอภาค อยู่ในระดับมาก (M = 3.78) รองลงมาคือขั้นตอนการกำหนดคะแนนจุดตัดสามารถปฏิบัติได้จริง ขั้นตอนการกำหนดคะแนนจุดตัดไม่ยุ่งยากซับซ้อน และการอภิปรายร่วมกันระหว่างผู้เชี่ยวชาญช่วยให้กำหนดคะแนนจุดตัดได้ง่ายขึ้น อยู่ในระดับมาก (M = 3.67) ให้เวลาสำหรับการอภิปรายกลุ่มอย่างเพียงพอ อยู่ในระดับมาก (M = 3.56) ตามลำดับ

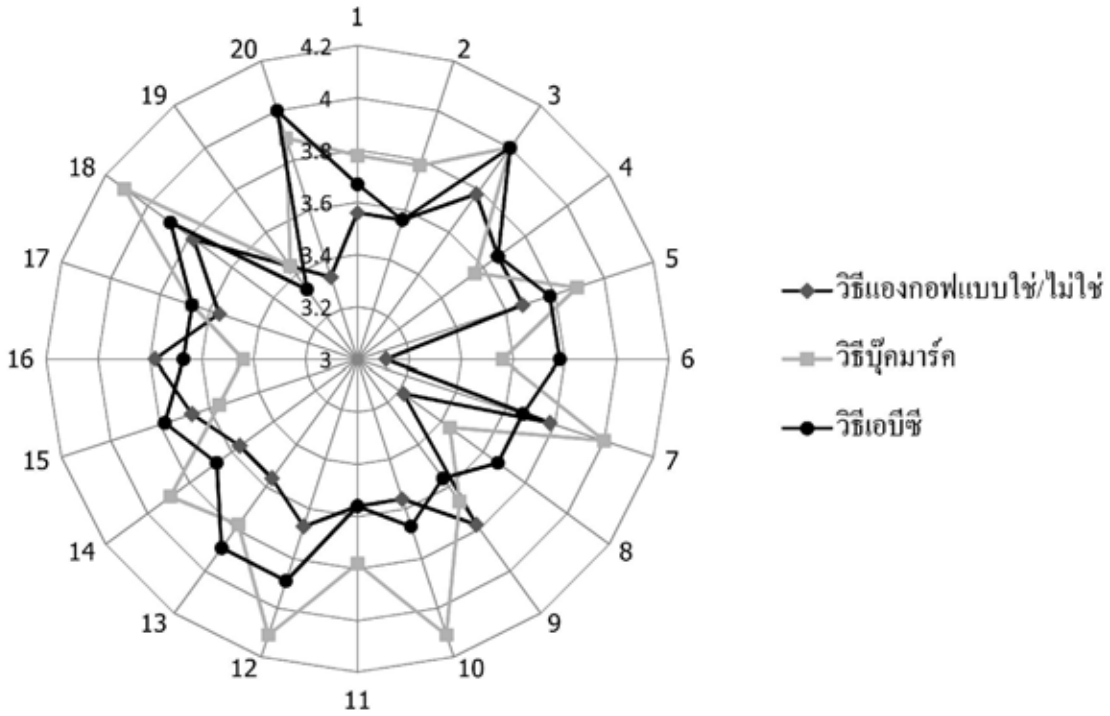
3.4 ด้านความเป็นประโยชน์ของเอกสาร อยู่ในระดับมาก (M = 3.67) รายการที่มีระดับสูงที่สุดได้แก่ สารสนเทศที่ให้เพื่อประกอบการกำหนดคะแนนจุดตัดเป็นประโยชน์ต่อการกำหนดคะแนนจุดตัดอย่างยิ่ง อยู่ในระดับมาก (M = 4.00) รองลงมารายงานผลการกำหนดคะแนนจุดตัดในแต่ละรอบเป็นประโยชน์ต่อการกำหนดคะแนนจุดตัดในรอบต่อ ๆ ไป อยู่ในระดับปานกลาง ($\overline{X}$ = 3.33)

3.5 ด้านความเหมาะสม อยู่ในระดับที่ต่ำที่สุด อยู่ในระดับมาก (M = 3.61) โดยรายการที่มีระดับสูงที่สุดได้แก่ เป้าหมายและการดำเนินการกำหนดคะแนนจุดตัดมีความสอดคล้องและเชื่อมโยงกัน อยู่ในระดับมาก (M = 3.67) รองลงมาคือเอกสารประกอบการดำเนินการสอดคล้องและเชื่อมโยงกันกับเป้าหมายของการกำหนดคะแนนจุดตัด อยู่ในระดับมาก (M = 3.56) ตามลำดับ

เพื่อให้ง่ายต่อการพิจารณาผลการวิเคราะห์ความตรงของปัจจัยนำเข้าของการกำหนดคะแนนจุดตัด ผู้วิจัยขอนำเสนอผลเป็นกราฟเส้น ดังแสดงในภาพต่อไปนี้



ภาพที่ 4.5 ผลการวิเคราะห์ความตรงของปัจจัยนำเข้าของการกำหนดคะแนนจุดตัดรายด้านวิชาคณิตศาสตร์



ภาพที่ 4.6 ผลการวิเคราะห์ความตรงของปัจจัยนำเข้าของการกำหนดคะแนนจุดตัดรายข้อวิชาคณิตศาสตร์

ตารางที่ 4.13 ผลการวิเคราะห์ความตรงของปัจจัยนำเข้าของการกำหนดคะแนนจุดตัดวิชาวิทยาศาสตร์

	ข้อความ	ระดับความเห็นด้วย								
		วิธีแองกอฟแบบใช่/ไม่ใช่			วิธีบุ๊คมาร์ค			วิธีเอบีซี		
		M	SD	ระดับ	*M*	SD	ระดับ	*M*	SD	ระดับ
	ด้านความถูกต้อง									
1.	การกำหนดคะแนนจุดครั้งนี้มีวิธีดำเนินการที่ถูกต้อง	3.67	0.50	มาก	4.11	0.93	มาก	3.92	0.67	มาก
2.	คำอธิบายระดับความสามารถมีเนื้อหาที่ถูกต้อง สื่อความหมายชัดเจน	3.67	0.50	มาก	3.89	0.93	มาก	3.83	0.58	มาก
3.	สารสนเทศที่ให้เพื่อประกอบการกำหนดคะแนนจุดตัดมีเนื้อหาที่ถูกต้อง สื่อความหมายชัดเจน	3.78	0.83	มาก	3.67	1.12	มาก	3.83	0.72	มาก
	รวม	3.70	0.26	มาก	3.89	0.82	มาก	3.86	0.46	มาก
	ด้านความเป็นไปได้									
4.	ขั้นตอนการกำหนดคะแนนจุดตัดสามารถปฏิบัติได้จริง	3.44	0.88	ปานกลาง	3.89	1.36	มาก	3.58	0.79	มาก
5.	ให้เวลาสำหรับการกำหนดคะแนนจุดตัดอย่างเพียงพอ	3.44	1.01	ปานกลาง	3.78	1.20	มาก	3.58	0.90	มาก
6.	ผู้เชี่ยวชาญทุกท่านมีโอกาสแสดงความคิดเห็นได้อย่างเสมอภาค	3.78	0.83	มาก	3.67	1.22	มาก	3.58	1.00	มาก
7.	ขั้นตอนการกำหนดคะแนนจุดตัดไม่ยุ่งยากซับซ้อน	4.22	0.44	มาก	3.78	1.20	มาก	3.75	0.62	มาก
8.	การอภิปรายร่วมกันระหว่างผู้เชี่ยวชาญช่วยให้กำหนดคะแนนจุดตัดได้ง่ายขึ้น	3.56	0.53	มาก	3.67	0.87	มาก	3.50	0.90	มาก
9.	ให้เวลาสำหรับการอภิปรายกลุ่มอย่างเพียงพอ	3.67	0.87	มาก	3.56	1.33	มาก	3.67	0.78	มาก
	รวม	3.69	0.29	มาก	3.72	0.79	มาก	3.61	0.51	มาก

ตารางที่ 4.13 ผลการวิเคราะห์ความตรงของปัจจัยนำเข้าของการกำหนดคะแนนจุดตัดวิชาวิทยาศาสตร์

	ข้อความ	ระดับความเห็นด้วย								
		วิธีแองกอฟแบบใช่/ไม่ใช่			วิธีบุ๊คมาร์ค			วิธีเอบีซี		
		M	SD	ระดับ	*M*	SD	ระดับ	*M*	SD	ระดับ
	ด้านความเหมาะสม									
10.	เป้าหมายและการดำเนินการกำหนดคะแนนจุดตัดมีความสอดคล้องและเชื่อมโยงกัน	3.89	0.93	มาก	3.67	1.22	มาก	4.00	0.95	มาก
11.	เอกสารประกอบการดำเนินการสอดคล้องและเชื่อมโยงกันกับเป้าหมายของการกำหนดคะแนนจุดตัด	4.00	0.71	มาก	3.67	1.12	มาก	3.50	0.80	มาก
	รวม	3.94	0.53	มาก	3.67	0.71	มาก	3.75	0.66	มาก
	ด้านความเข้าใจและความมั่นใจของผู้เชี่ยวชาญ									
12.	ฉันเข้าใจความหมายของผู้มีความสามารถคาบเส้นเป็นอย่างดี	3.89	0.78	มาก	3.67	1.12	มาก	3.83	1.03	มาก
13.	ฉันเข้าใจถึงเป้าหมายของการกำหนดคะแนนจุดตัดเป็นอย่างดี	3.67	0.71	มาก	4.00	0.50	มาก	3.92	0.67	มาก
14.	ฉันเข้าใจถึงความแตกต่างระหว่างผู้ที่มีความสามารถที่แตกต่างกัน	4.11	0.60	มาก	3.78	1.20	มาก	3.75	0.97	มาก
15.	ฉันเข้าใจวิธีการและขั้นตอนในการกำหนดคะแนนจุดตัดเป็นอย่างดี	3.89	0.93	มาก	4.00	0.87	มาก	4.00	1.04	มาก
16.	ฉันไม่รู้สึกกดดันเมื่อคะแนนจุดตัดที่กำหนดไม่ใกล้เคียงกับผู้เชี่ยวชาญท่านอื่น ๆ	3.89	0.93	มาก	4.00	0.50	มาก	3.75	0.97	มาก
17.	ฉันมั่นใจว่าคะแนนจุดตัดที่กำหนดมีความเหมาะสม	3.56	0.88	มาก	3.56	1.33	มาก	4.00	0.95	มาก
18.	ฉันเห็นความสำคัญของการกำหนดคะแนนจุดตัดครั้งนี้	4.22	0.44	มาก	3.78	1.09	มาก	4.08	0.67	มาก
	รวม	3.89	0.36	มาก	3.83	0.50	มาก	3.90	0.59	มาก

ตารางที่ 4.13 ผลการวิเคราะห์ความตรงของปัจจัยนำเข้าของการกำหนดคะแนนจุดตัดวิชาวิทยาศาสตร์

	ข้อความ	ระดับความเห็นด้วย								
		วิธีแองกอฟแบบใช่/ไม่ใช่			วิธีบุ๊คมาร์ค			วิธีเอบีซี		
		M	SD	ระดับ	*M*	SD	ระดับ	*M*	SD	ระดับ
	ด้านความเป็นประโยชน์ของเอกสาร									
19.	รายงานผลการกำหนดคะแนนจุดตัดในแต่ละรอบเป็นประโยชน์ต่อการกำหนดคะแนนจุดตัดในรอบต่อ ๆ ไป	3.22	0.67	ปานกลาง	3.44	1.01	ปานกลาง	3.50	0.67	มาก
20.	สารสนเทศที่ให้เพื่อประกอบการกำหนดคะแนนจุดตัดเป็นประโยชน์ต่อการกำหนดคะแนนจุดตัดอย่างยิ่ง	3.33	0.50	ปานกลาง	3.78	1.09	ปานกลาง	3.92	0.79	มาก
	รวม	3.28	0.44	ปานกลาง	3.61	0.89	มาก	3.71	0.54	มาก
	รวมทั้งฉบับ	**3.74**	**0.16**	**มาก**	**3.77**	**0.51**	**มาก**	**3.78**	**0.44**	**มาก**

จากตาราง พบว่า

1. ผู้เชี่ยวชาญมีความคิดเห็นว่าการกำหนดคะแนนจุดตัดด้วยวิธีแองกอฟแบบใช่/ไม่ใช่มีความตรงของปัจจัยนำเข้าโดยภาพรวมอยู่ในระดับมาก (M = 3.74) เมื่อพิจารณาเป็นรายด้านพบว่าปัจจัยนำเข้าในแต่ละด้านมีค่าต่างกัน โดยสามารถเรียงตามลำดับได้เป็นดังนี้

1.1 ด้านความเหมาะสม อยู่ในระดับมาก (M = 3.94) โดยรายการที่มีระดับสูงที่สุดได้แก่ เอกสารประกอบการดำเนินการสอดคล้องและเชื่อมโยงกันกับเป้าหมายของการกำหนดคะแนนจุดตัด อยู่ในระดับมาก (M = 4.00) รองลงมาคือเป้าหมายและการดำเนินการกำหนดคะแนนจุดตัดมีความสอดคล้องและเชื่อมโยงกัน อยู่ในระดับมาก (M = 3.89) ตามลำดับ

1.2 ด้านความเข้าใจและความมั่นใจของผู้เชี่ยวชาญ อยู่ในระดับมาก (M = 3.89) โดยรายการที่มีระดับสูงที่สุดได้แก่ ผู้เชี่ยวชาญเห็นความสำคัญของการกำหนดคะแนนจุดตัดครั้งนี้ อยู่ในระดับมาก (M = 4.22) รองลงมาคือผู้เชี่ยวชาญเข้าใจถึงความแตกต่างระหว่างผู้ที่มีความสามารถที่แตกต่างกัน อยู่ในระดับมาก (M = 4.11) ผู้เชี่ยวชาญเข้าใจความหมายของผู้มีความสามารถคาบเส้นเป็นอย่างดี ผู้เชี่ยวชาญเข้าใจวิธีการและขั้นตอนในการกำหนดคะแนนจุดตัดเป็นอย่างดี ผู้เชี่ยวชาญไม่รู้สึกกดดันเมื่อคะแนนจุดตัดที่กำหนดไม่ใกล้เคียงกับผู้เชี่ยวชาญท่านอื่น ๆ อยู่ในระดับมาก (M = 3.89) ผู้เชี่ยวชาญเข้าใจถึงเป้าหมายของการกำหนดคะแนนจุดตัดเป็นอย่างดี อยู่ในระดับมาก (M = 3.67) และผู้เชี่ยวชาญมั่นใจว่าคะแนนจุดตัดที่กำหนดมีความเหมาะสม อยู่ในระดับมาก (M =3.56) ตามลำดับ

1.3 ด้านความถูกต้อง อยู่ในระดับมาก (M = 3.70) โดยรายการที่มีระดับสูงที่สุดได้แก่ สารสนเทศที่ให้เพื่อประกอบการกำหนดคะแนนจุดตัดมีเนื้อหาที่ถูกต้อง สื่อความหมายชัดเจน อยู่ในระดับมาก (M = 3.78) รองลงมาคือการกำหนดคะแนนจุดครั้งนี้มีวิธีดำเนินการที่ถูกต้อง และคำอธิบายระดับความสามารถมีเนื้อหาที่ถูกต้อง สื่อความหมายชัดเจน อยู่ในระดับมาก (M =3.67)

1.4 ด้านความเป็นไปได้ อยู่ในระดับมาก (M = 3.69) โดยรายการที่มีระดับสูงที่สุด ได้แก่ ขั้นตอนการกำหนดคะแนนจุดตัดไม่ยุ่งยากซับซ้อน อยู่ในระดับมาก (M = 4.22) รองลงมาคือ ผู้เชี่ยวชาญทุกท่านมีโอกาสแสดงความคิดเห็นได้อย่างเสมอภาค อยู่ในระดับมาก (M = 3.78) ให้เวลาสำหรับการอภิปรายกลุ่มอย่างเพียงพอ อยู่ในระดับมาก (M = 3.67) การอภิปรายร่วมกันระหว่างผู้เชี่ยวชาญช่วยให้กำหนดคะแนนจุดตัดได้ง่ายขึ้น อยู่ในระดับมาก (M = 3.56) ขั้นตอนการ

กำหนดคะแนนจุดตัดสามารถปฏิบัติได้จริง และให้เวลาสำหรับการกำหนดคะแนนจุดตัดอย่างเพียงพอ อยู่ในระดับปานกลาง (M = 3.44)

1.5 ด้านความเป็นประโยชน์ของเอกสาร อยู่ในระดับที่ต่ำที่สุด อยู่ในระดับปานกลาง (M = 3.28) โดยรายการที่มีระดับสูงที่สุด ได้แก่ สารสนเทศที่ให้เพื่อประกอบการกำหนดคะแนนจุดตัดเป็นประโยชน์ต่อการกำหนดคะแนนจุดตัดอย่างยิ่ง อยู่ในระดับปานกลาง (M = 3.33) รองลงมาคือรายงานผลการกำหนดคะแนนจุดตัดในแต่ละรอบเป็นประโยชน์ต่อการกำหนดคะแนนจุดตัดในรอบต่อ ๆ ไป อยู่ในระดับปานกลาง (M = 3.22) ตามลำดับ

2. ผู้เชี่ยวชาญมีความคิดเห็นว่าการกำหนดคะแนนจุดตัดด้วยวิธีบุ๊คมาร์คมีความตรงของปัจจัยนำเข้าโดยภาพรวมอยู่ในระดับมาก (M = 3.77) เมื่อพิจารณาเป็นรายด้านพบว่า ปัจจัยนำเข้าในแต่ละด้านมีค่าต่างกัน โดยสามารถเรียงตามลำดับได้เป็นดังนี้

2.1 ด้านความถูกต้อง อยู่ในระดับมาก (M = 3.89) โดยรายการที่มีระดับสูงที่สุดได้แก่ การกำหนดคะแนนจุดครั้งนี้มีวิธีดำเนินการที่ถูกต้อง อยู่ในระดับมาก (M = 4.11) รองลงมาคือคำอธิบายระดับความสามารถมีเนื้อหาที่ถูกต้อง สื่อความหมายชัดเจน อยู่ในระดับมาก (M = 3.89) สารสนเทศที่ให้เพื่อประกอบการกำหนดคะแนนจุดตัดมีเนื้อหาที่ถูกต้อง สื่อความหมายชัดเจน อยู่ในระดับมาก (M = 3.67) ตามลำดับ

2.2 ด้านความเข้าใจและความมั่นใจของผู้เชี่ยวชาญ อยู่ในระดับมาก (M = 3.83) โดยรายการที่มีระดับสูงที่สุดได้แก่ ผู้เชี่ยวชาญเข้าใจถึงเป้าหมายของการกำหนดคะแนนจุดตัดเป็นอย่างดี ผู้เชี่ยวชาญเข้าใจวิธีการและขั้นตอนในการกำหนดคะแนนจุดตัดเป็นอย่างดี และผู้เชี่ยวชาญไม่รู้สึกกดดันเมื่อคะแนนจุดตัดที่กำหนดไม่ใกล้เคียงกับผู้เชี่ยวชาญท่านอื่น ๆ อยู่ในระดับมาก (M = 4.00) รองลงมาคือผู้เชี่ยวชาญเข้าใจถึงความแตกต่างระหว่างผู้ที่มีความสามารถที่แตกต่างกันและผู้เชี่ยวชาญเห็นความสำคัญของการกำหนดคะแนนจุดตัดครั้งนี้ อยู่ในระดับมาก (M = 3.78) ผู้เชี่ยวชาญเข้าใจความหมายของผู้มีความสามารถคาบเส้นเป็นอย่างดี อยู่ในระดับมาก (M = 3.67) และผู้เชี่ยวชาญมั่นใจว่าคะแนนจุดตัดที่กำหนดมีความเหมาะสม อยู่ในระดับมาก (M = 3.56) ตามลำดับ

2.3 ด้านความเป็นไปได้ อยู่ในระดับมาก (M = 3.72) โดยรายการที่มีระดับสูงที่สุด ได้แก่ ขั้นตอนการกำหนดคะแนนจุดตัดสามารถปฏิบัติได้จริง อยู่ในระดับมาก (M = 3.89) รองลงมาคือให้

เวลาสำหรับการกำหนดคะแนนจุดตัดอย่างเพียงพอ และขั้นตอนการกำหนดคะแนนจุดตัดไม่ยุ่งยากซับซ้อน อยู่ในระดับมาก (M = 3.78) ผู้เชี่ยวชาญทุกท่านมีโอกาสแสดงความคิดเห็นได้อย่างเสมอภาค และการอภิปรายร่วมกันระหว่างผู้เชี่ยวชาญช่วยให้กำหนดคะแนนจุดตัดได้ง่ายขึ้น อยู่ในระดับมาก (M = 3.67) และให้เวลาสำหรับการอภิปรายกลุ่มอย่างเพียงพอ อยู่ในระดับมาก (M = 3.56)

2.4 ด้านความเหมาะสม อยู่ในระดับมาก (M = 3.67) โดยรายการที่มีระดับสูงที่สุด ได้แก่ เป้าหมายและการดำเนินการกำหนดคะแนนจุดตัดมีความสอดคล้องและเชื่อมโยงกัน และเอกสารประกอบการดำเนินการสอดคล้องและเชื่อมโยงกันกับเป้าหมายของการกำหนดคะแนนจุดตัด อยู่ในระดับมาก (M = 3.67)

2.5 ด้านความเป็นประโยชน์ของเอกสาร อยู่ในระดับที่ต่ำที่สุด โดยอยู่ในระดับมาก (M = 3.61) โดยรายการที่มีระดับสูงที่สุดได้แก่ สารสนเทศที่ให้เพื่อประกอบการกำหนดคะแนนจุดตัดเป็นประโยชน์ต่อการกำหนดคะแนนจุดตัดอย่างยิ่ง อยู่ในระดับมาก (M = 3.78) รองลงมาคือ รายงานผลการกำหนดคะแนนจุดตัดในแต่ละรอบเป็นประโยชน์ต่อการกำหนดคะแนนจุดตัดในรอบต่อ ๆ ไป อยู่ในระดับปานกลาง (M = 3.44) ตามลำดับ

3. ผู้เชี่ยวชาญมีความคิดเห็นว่าการกำหนดคะแนนจุดตัดด้วยวิธีเอบีซีมีความตรงของปัจจัยนำเข้าโดยภาพรวมอยู่ในระดับมาก (M = 3.78) เมื่อพิจารณาเป็นรายด้านพบว่า ปัจจัยนำเข้าในแต่ละด้านมีค่าต่างกัน โดยสามารถเรียงตามลำดับได้เป็นดังนี้

3.1 ด้านความเข้าใจและความมั่นใจของผู้เชี่ยวชาญสูงที่สุด ซึ่งอยู่ในระดับมาก (M = 3.90) โดยรายการที่มีระดับสูงที่สุดได้แก่ ผู้เชี่ยวชาญเห็นความสำคัญของการกำหนดคะแนนจุดตัดครั้งนี้ อยู่ในระดับมาก (M = 4.08) รองลงมาคือ ผู้เชี่ยวชาญเข้าใจวิธีการและขั้นตอนในการกำหนดคะแนนจุดตัดเป็นอย่างดี และผู้เชี่ยวชาญมั่นใจว่าคะแนนจุดตัดที่กำหนดมีความเหมาะสมอยู่ในระดับมาก (M = 4.00) ผู้เชี่ยวชาญเข้าใจถึงเป้าหมายของการกำหนดคะแนนจุดตัดเป็นอย่างดีอยู่ในระดับมาก (M = 3.92) ผู้เชี่ยวชาญเข้าใจความหมายของผู้มีความสามารถคาบเส้นเป็นอย่างดีอยู่ในระดับมาก (M = 3.83) ผู้เชี่ยวชาญเข้าใจถึงความแตกต่างระหว่างผู้ที่มีความสามารถที่แตกต่างกัน และผู้เชี่ยวชาญไม่รู้สึกกดดันเมื่อคะแนนจุดตัดที่กำหนดไม่ใกล้เคียงกับผู้เชี่ยวชาญท่านอื่น ๆ อยู่ในระดับมาก (M = 3.75)

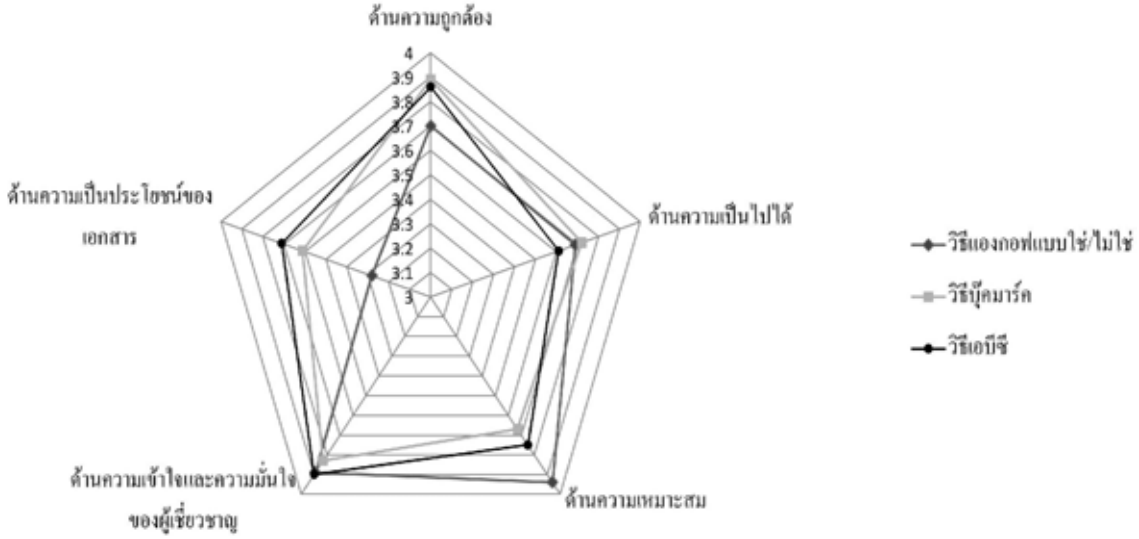
3.2 ด้านความถูกต้อง อยู่ในระดับมาก (M = 3.86) โดยรายการที่มีระดับสูงที่สุด ได้แก่ การกำหนดคะแนนจุดครั้งนี้มีวิธีดำเนินการที่ถูกต้อง อยู่ในระดับมาก (M = 3.92) รองลงมาคือ คำอธิบายระดับความสามารถมีเนื้อหาที่ถูกต้อง สื่อความหมายชัดเจน และสารสนเทศที่ให้เพื่อประกอบการกำหนดคะแนนจุดตัดมีเนื้อหาที่ถูกต้อง สื่อความหมายชัดเจน อยู่ในระดับมาก (M = 3.83)

3.3 ด้านความเหมาะสม อยู่ในระดับมาก (M = 3.75) โดยรายการที่มีระดับสูงที่สุด ได้แก่ เป้าหมายและการดำเนินการกำหนดคะแนนจุดตัดมีความสอดคล้องและเชื่อมโยงกัน อยู่ในระดับมาก (M = 4.00) รองลงมาคือเอกสารประกอบการดำเนินการสอดคล้องและเชื่อมโยงกันกับเป้าหมายของการกำหนดคะแนนจุดตัด อยู่ในระดับมาก (M = 3.50)

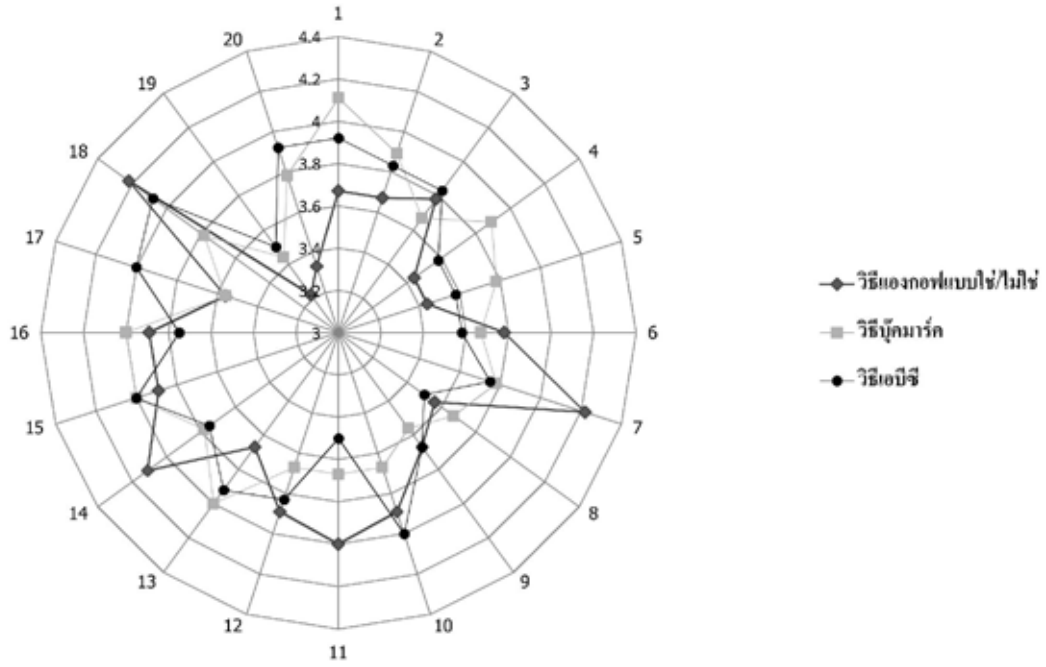
3.4 ด้านความเป็นประโยชน์ของเอกสาร อยู่ในระดับมาก (M = 3.71) โดยรายการที่มีระดับสูงที่สุด ได้แก่ สารสนเทศที่ให้เพื่อประกอบการกำหนดคะแนนจุดตัดเป็นประโยชน์ต่อการกำหนดคะแนนจุดตัดอย่างยิ่ง อยู่ในระดับมาก (M = 3.92) รองลงมาคือรายงานผลการกำหนดคะแนนจุดตัดในแต่ละรอบเป็นประโยชน์ต่อการกำหนดคะแนนจุดตัดในรอบต่อ ๆ ไป อยู่ในระดับมาก (M = 3.50) ตามลำดับ

3.5 ด้านความเป็นไปได้ อยู่ในระดับที่ต่ำที่สุด อยู่ในระดับมาก (M = 3.61) โดยรายการที่มีระดับสูงที่สุด ได้แก่ ขั้นตอนการกำหนดคะแนนจุดตัดไม่ยุ่งยากซับซ้อน อยู่ในระดับมาก (M = 3.75) รองลงมาคือให้เวลาสำหรับการอภิปรายกลุ่มอย่างเพียงพอ อยู่ในระดับมาก (M = 3.67) ขั้นตอนการกำหนดคะแนนจุดตัดสามารถปฏิบัติได้จริง ให้เวลาสำหรับการกำหนดคะแนนจุดตัดอย่างเพียงพอ และผู้เชี่ยวชาญทุกท่านมีโอกาสแสดงความคิดเห็นได้อย่างเสมอภาค อยู่ในระดับมาก (M = 3.58) การอภิปรายร่วมกันระหว่างผู้เชี่ยวชาญช่วยให้กำหนดคะแนนจุดตัดได้ง่ายขึ้น อยู่ในระดับมาก (M = 3.50) ตามลำดับ

เพื่อให้ง่ายต่อการพิจารณาผลการวิเคราะห์ความตรงของปัจจัยนำเข้าของการกำหนดคะแนนจุดตัด ผู้วิจัยขอนำเสนอผลเป็นกราฟเส้น ดังแสดงในภาพต่อไปนี้



ภาพที่ 4.7 ผลการวิเคราะห์ความตรงของปัจจัยนำเข้าของการกำหนดคะแนนจุดตัดรายด้านวิชาวิทยาศาสตร์



ภาพที่ 4.8 ผลการวิเคราะห์ความตรงของปัจจัยนำเข้าของการกำหนดคะแนนจุดตัดรายข้อวิชาวิทยาศาสตร์

ตารางที่ 4.14 ผลการวิเคราะห์วิชาและวิธีกำหนดคะแนนจุดตัดที่มีผลต่อความตรงของปัจจัยนำเข้าของการกำหนดคะแนนจุดตัด

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	ด้านความถูกต้อง	4.510	5	.902	.306	.907
	ด้านความเป็นไปได้	7.673	5	1.535	.145	.981
	ด้านความเหมาะสม	5.023	5	1.005	.681	.640
	ด้านความเข้าใจและความมั่นใจของผู้เชี่ยวชาญ	19.871	5	3.974	.371	.866
	ด้านความเป็นประโยชน์ของเอกสาร	5.890	5	1.178	.681	.640
Intercept	ด้านความถูกต้อง	7243.479	1	7243.479	2455.054	.000
	ด้านความเป็นไปได้	27089.005	1	27089.005	2554.202	.000
	ด้านความเหมาะสม	3161.744	1	3161.744	2143.634	.000
	ด้านความเข้าใจและความมั่นใจของผู้เชี่ยวชาญ	40013.932	1	40013.932	3734.525	.000
	ด้านความเป็นประโยชน์ของเอกสาร	2845.682	1	2845.682	1644.530	.000
วิชา	ด้านความถูกต้อง	.755	1	.755	.256	.615
	ด้านความเป็นไปได้	.696	1	.696	.066	.799
	ด้านความเหมาะสม	.391	1	.391	.265	.609
	ด้านความเข้าใจและความมั่นใจของผู้เชี่ยวชาญ	12.077	1	12.077	1.127	.293
	ด้านความเป็นประโยชน์ของเอกสาร	.098	1	.098	.057	.813
วิธี	ด้านความถูกต้อง	3.485	2	1.742	.591	.558
	ด้านความเป็นไปได้	2.818	2	1.409	.133	.876
	ด้านความเหมาะสม	.606	2	.303	.205	.815
	ด้านความเข้าใจและความมั่นใจของผู้เชี่ยวชาญ	2.402	2	1.201	.112	.894
	ด้านความเป็นประโยชน์ของเอกสาร	5.485	2	2.742	1.585	.215

ตารางที่ 4.14 ผลการวิเคราะห์วิชาและวิธีกำหนดคะแนนจุดตัดที่มีผลต่อความตรงของปัจจัยนำเข้าของการกำหนดคะแนนจุดตัด

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.
วิชา * วิธี	ด้านความถูกต้อง	.152	2	.076	.026	.975
	ด้านความเป็นไปได้	4.273	2	2.136	.201	.818
	ด้านความเหมาะสม	4.091	2	2.045	1.387	.259
	ด้านความเข้าใจและความมั่นใจของผู้เชี่ยวชาญ	4.705	2	2.352	.220	.804
	ด้านความเป็นประโยชน์ของเอกสาร	.232	2	.116	.067	.935
Error	ด้านความถูกต้อง	150.472	51	2.950		
	ด้านความเป็นไปได้	540.889	51	10.606		
	ด้านความเหมาะสม	75.222	51	1.475		
	ด้านความเข้าใจและความมั่นใจของผู้เชี่ยวชาญ	546.444	51	10.715		
	ด้านความเป็นประโยชน์ของเอกสาร	88.250	51	1.730		
Total	ด้านความถูกต้อง	7499.000	57			
	ด้านความเป็นไปได้	27917.000	57			
	ด้านความเหมาะสม	3279.000	57			
	ด้านความเข้าใจและความมั่นใจของผู้เชี่ยวชาญ	41153.000	57			
	ด้านความเป็นประโยชน์ของเอกสาร	2986.000	57			
Corrected Total	ด้านความถูกต้อง	154.982	56			
	ด้านความเป็นไปได้	548.561	56			
	ด้านความเหมาะสม	80.246	56			
	ด้านความเข้าใจและความมั่นใจของผู้เชี่ยวชาญ	566.316	56			
	ด้านความเป็นประโยชน์ของเอกสาร	94.140	56			

จากตาราง พบว่า วิชาต่างกันมีผลต่อความตรงของปัจจัยนำเข้าแตกต่างกันอย่างไม่มีนัยสำคัญทุกด้าน แสดงว่าวิชาต่างกันไม่มีผลต่อความตรงของปัจจัยนำเข้าทุกด้าน

วิธีการกำหนดคะแนนจุดตัดต่างกันมีผลต่อความตรงของปัจจัยนำเข้าแตกต่างกันอย่างไม่มีนัยสำคัญทุกด้าน แสดงว่าวิธีต่างกันไม่มีผลต่อความตรงของปัจจัยนำเข้าทุกด้าน

ปฏิสัมพันธ์ระหว่างวิชาและวิธีการกำหนดคะแนนจุดตัดมีผลต่อความตรงของปัจจัยนำเข้าแตกต่างกันอย่างไม่มีนัยสำคัญทุกด้าน แสดงว่าไม่มีปฏิสัมพันธ์ระหว่างวิชาและวิธีการกำหนดคะแนนจุดตัดที่ส่งผลต่อความตรงของปัจจัยนำเข้าทุกด้าน

4.1.2 ผลการวิเคราะห์ความตรงของกระบวนการตัดสินใจของผู้เชี่ยวชาญ

การเก็บข้อมูลเพื่อตรวจสอบความตรงของกระบวนการตัดสินใจของผู้เชี่ยวชาญ ได้ดำเนินการโดยใช้เทคนิคการคิดออกเสียง การสังเกตพฤติกรรม การบันทึกข้อสนทนา และการสนทนากลุ่ม ซึ่งเป็นวิธีการเก็บข้อมูลเชิงคุณภาพ ผู้วิจัยจึงสรุปผลการวิเคราะห์ข้อมูลเชิงคุณภาพซึ่งสามารถแบ่งได้เป็น 3 ประเด็น ได้แก่ ประเด็นที่เกี่ยวกับหลักการที่ใช้ในการพิจารณาเพื่อกำหนดคะแนนจุดตัด ปัญหา และข้อเสนอแนะสำหรับการกำหนดคะแนนจุดตัด ซึ่งมีรายละเอียดดังนี้

4.1.2.1 หลักการที่ใช้ในการพิจารณาเพื่อกำหนดคะแนนจุดตัด

4.1.2.1.1 ผู้เชี่ยวชาญกำหนดคะแนนจุดตัดโดยพิจารณาระดับพฤติกรรมด้านพุทธิพิสัยที่ข้อสอบต้องการวัด

ข้อมูลจากการใช้เทคนิคการคิดออกเสียง แสดงให้เห็นว่า ผู้เชี่ยวชาญจะพิจารณาข้อสอบโดยพิจารณาจากระดับพฤติกรรมด้านพุทธิพิสัยที่วัด โดยระดับที่พิจารณาส่วนใหญ่ได้แก่ ความรู้ความจำ ความเข้าใจ การนำไปประยุกต์ใช้ และการวิเคราะห์ ดังตัวออย่างต่อไปนี้

ข้อ 15.
นาย ก และ นาย ข ยืนอยู่บนพื้นราบซึ่งห่างจากกำแพงเป็นระยะ 10 เมตร และ 40 เมตร ตามลำดับ ถ้านาย ก มองหลอดไฟบนกำแพงด้วยมุมเงย α องศา ในขณะที่นาย ข มองหลอดไฟดวงเดียวกันด้วยมุมเงย $90 - \alpha$ องศา ถ้าไม่คิดความสูงของนาย ก และนาย ข แล้วหลอดไฟอยู่สูงกว่าพื้นราบกี่เมตร

1. 10
2. $10\sqrt{2}$
3. $10\sqrt{3}$
4. 20

"ข้อนี้ ยากนะ ต้องวิเคราะห์โจทย์ก่อน มันซับซ้อนมาก.....อืม...ต้องประยุกต์ทฤษฎีพีทาโกรัสหาคำตอบ คือไม่ใช่แค่จำกับเข้าใจนะ มันต้องมากกว่านั้น ข้อนี้ระดับ 4 นะถึงทำได้ มันอยู่ระดับการประยุกต์แล้ว" (ข้อมูลจากการใช้เทคนิคการคิดออกเสียง)

"ข้อนี้ต้องตีโจทย์ให้แตกก่อน แล้วใช้ทฤษฎีพีทาโกรัสหาคำตอบ มันซับซ้อนอยู่นะ คือแค่เข้าใจไม่พอ ต้องคนเก่ง" (ข้อมูลจากการใช้เทคนิคการคิดออกเสียง)

ข้อ 14.
กำหนดให้สามเหลี่ยมมุมฉาก ABC มี $\hat{C} = 90^o$
ให้ D เป็นจุดบนด้าน AB ซึ่งทำให้ CD ตั้งฉากกับ AB
ถ้า AB ยาว 20 หน่วย และ CD ยาว 8 หน่วย
แล้ว AD มีความยาวมากที่สุดกี่หน่วย

1. 10
2. 12
3. 14
4. 16

"ข้อนี้ เหมือนจะยากแต่ไม่มากนะ เหมือนข้อ 13 เลย คือถ้าเข้าใจเรื่องคุณสมบัติของสามเหลี่ยม...แต่..มีสูตรลัดนะ ถ้าจำสูตรของทฤษฎีพีทากอรัสได้นะ เร็วเลย" (ข้อมูลจากการใช้เทคนิคการคิดออกเสียง)

“ข้อนี้ต้องวาดรูปดู แล้วใช้ความรู้เกี่ยวกับรูปสามเหลี่ยมมุมฉาก แล้วหาความยาวของสามเหลี่ยมมุมฉาก คือมันซับซ้อนหน่อย แต่ไม่เท่าข้อ 15 ถ้าค่อย ๆ คิดเป็นระบบแล้วเข้าใจเรื่องรูปสามเหลี่ยมมุมฉาก ก็ทำได้เลยนะ” (ข้อมูลจากการใช้เทคนิคการคิดออกเสียง)

ข้อ 27.
สัญลักษณ์ต่อไปนี้มีความหมายว่าอย่างไร

1. สามารถรีไซเคิลได้อีก 5 ครั้ง
2. สามารถรีไซเคิลได้ทั้งหมด 5 ครั้ง
3. ผ่านการรีไซเคิลมาได้ 5 ครั้งแล้ว
4. เป็นพลาสติกรีไซเคิลประเภทที่ 5

“สัญลักษณ์นี้ง่าย ๆ เรียนมาแล้ว จำได้ก็ตอบได้ ง่ายมากเลยข้อนี้ ถ้าทำไม่ได้แสดงว่าไม่จำ คือเรียนแล้ว” (ข้อมูลจากการใช้เทคนิคการคิดออกเสียง)

“ข้อนี้วัดความจำ ล้วน ๆ ต้องจำสัญลักษณ์นี้ได้ถึงจะตอบได้ ไม่ต้องประยุกต์อะไร ทั้งนั้น จำมาตอบเอา ข้อนี้ระดับที่ 1 ก็ตอบได้แล้ว” (ข้อมูลจากการใช้เทคนิคการคิดออกเสียง)

4.1.2.1.2 ผู้เชี่ยวชาญกำหนดคะแนนจุดตัดโดยจะพิจารณาเนื้อหาที่สอบวัด รูปแบบของข้อสอบ และขั้นตอนของการได้มาซึ่งคำตอบที่ถูกต้อง

ข้อมูลจากการสังเกตพฤติกรรม พบว่า ผู้เชี่ยวชาญส่วนใหญ่จะลงมือทำข้อสอบจนได้คำตอบ ซึ่งผู้วิจัยได้แจ้งแก่ผู้เชี่ยวชาญว่าไม่จำเป็นต้องลงมือทำข้อสอบก็ได้ แต่ก็ยังมีผู้เชี่ยวชาญหลายท่านลงมือทำข้อสอบที่เป็นข้อสอบแบบเติมคำอยู่ ผู้วิจัยจึงเดินแจ้งให้นักศึกษาช่วยงานสอบถามถึงเหตุผลที่ลงมือทำข้อสอบ ก็ได้คำตอบว่าผู้เชี่ยวชาญส่วนใหญ่ต้องการพิจารณาทุกรายละเอียดของการได้มาซึ่งคำตอบที่ถูกต้อง อาทิ เวลาที่ใช้ในการทำข้อสอบ จุดที่ผู้สอบมีโอกาสผิดพลาดหากไม่รอบคอบพอ และความซับซ้อนของขั้นตอนในการหาคำตอบ เป็นต้น ซึ่งข้อมูลเหล่านี้จะใช้ในการตัดสินใจกำหนดคะแนนจุดตัดต่อไป

นอกจากนี้ข้อมูลจากการใช้เทคนิคการคิดออกเสียง พบว่า ผู้เชี่ยวชาญมีความคิดเห็นว่าจะไม่พิจารณาเฉพาะเนื้อหาที่วัดหรือว่าระดับพฤติกรรมด้านพุทธิพิสัยที่วัดเท่านั้นแต่ต้องพิจารณารูปแบบข้อสอบด้วย กล่าวคือ ข้อสอบบางข้อวัดเนื้อหาที่ไม่ยาก หรือระดับพฤติกรรมที่ไม่สูงนัก แต่อาจจะเป็นข้อที่ยากสำหรับผู้สอบได้ เนื่องจากรูปแบบข้อสอบอาจส่งผลให้ข้อสอบยากขึ้น เช่น ข้อสอบวิชาคณิตศาสตร์ซึ่งมีข้อสอบส่วนหนึ่งเป็นข้อสอบแบบเติมคำตอบ ผลการวิเคราะห์ค่าความยากของข้อสอบพบว่าเป็นข้อสอบในกลุ่มที่มีความยากปานกลางไปจนถึงยากมาก ทั้ง ๆ ที่เนื้อหาที่สอบวัดไม่ยาก ดังตัวอย่างต่อไปนี้

ข้อ 23.

ถ้า $x=\frac{\sqrt{2}+\sqrt{3}}{\sqrt{2}-\sqrt{3}}$ และ $y=\frac{\sqrt{2}-\sqrt{3}}{\sqrt{2}+\sqrt{3}}$ แล้ว $x^2-4xy+y^2$ เท่ากับเท่าใด

"ข้อนี้ต้องคิดตรงสมการ $x^2-4xy+y^2$ ก่อน ไม่งั้นค่าเพี้ยน คำนวณหาค่า x y เลยไม่ได้ ต้องทั้งหมดมาแทนในสูตร ต้องใช้ความรู้เรื่องสังยุค แต่ถ้าคิดค่า x แล้วมาแทนค่า คำตอบมันไม่เป๊ะนะ คือเติมคำตอบผิด...ผิดเลย เด็กต้องรอบคอบนะ" (ข้อมูลจากการใช้เทคนิคการคิดออกเสียง)

ข้อ 37.

ข้อมูลชุดหนึ่งเรียงลำดับจากน้อยไปมากดังนี้

2 3 3 x 4 y 7

ค่าเฉลี่ยเลขคณิตและส่วนเบี่ยงเบนมาตรฐานของข้อมูลชุดนี้เท่ากับ 4 และ $\frac{4}{\sqrt{7}}$ ตามลำดับ แล้ว $y-x$ มีค่าเท่าใด

"ข้อ 37 (ลำดับที่ 31) ง่ายกว่าข้อ 16 นะ (ลำดับที่ 6) มันวัดแค่ความรู้ ความจำในเรื่องสถิติเบื้องต้น แต่ขั้นตอนมันเยอะหน่อย แต่ไม่ยาก ถ้าเด็กทำไม่ได้ น่าจะเป็นเพราะไม่ทำเพราะมันใช้เวลานานอยู่ หรือไม่ก็คิดเลขผิด" (ข้อมูลจากการใช้เทคนิคการคิดออกเสียง)

"ข้อนี้ต้องแก้สมการเป็น จำสูตรหาค่าเฉลี่ย S.D ได้ แต่ก็ไม่แน่นะ จำสูตรได้แต่คำนวณผิดก็ทำผิดเลย เด็กครูนะไม่ทำแน่ คือเด็กเดี๋ยวนี้ไม่ค่อยตั้งใจนะ ข้อไหนต้องทำนานเขาข้ามเลย ข้อนี้เลยอยู่ลำดับนี้ไง ถ้าเป็น ก ข ค ง น่าจะง่ายกว่านี้ ข้อนี้ต้องระดับเก่งหน่อยนะถึงทำเป๊ะ"
(ข้อมูลจากการใช้เทคนิคการคิดออกเสียง)

4.1.2.2 ปัญหา ได้แก่

4.1.2.2.1 ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดด้วยวิธีบุ๊คมาร์คมีความคิดเห็นว่าข้อสอบบางข้อที่ถูกเรียงไว้ในคู่มือเรียงข้อสอบ มีลำดับที่ไม่เหมาะสม

ข้อมูลจากการใช้เทคนิคการคิดออกเสียง และการสังเกตพฤติกรรม พบว่าผู้เชี่ยวชาญหลายท่านตั้งข้อสังเกตเกี่ยวกับลำดับความยากของข้อสอบบางข้อว่ามีลำดับที่ไม่เหมาะสม โดยเฉพาะข้อสอบวิชาคณิตศาสตร์ มีข้อสอบจำนวน 40 ข้อ ซึ่งประกอบไปด้วยข้อสอบ 2 แบบ ได้แก่ ข้อสอบแบบเลือกตอบ และข้อสอบแบบเติมคำตอบ โดยประเด็นที่ผู้เชี่ยวชาญตั้งข้อสังเกตคือข้อสอบแบบเติมคำตอบนั้น เมื่อพิจารณาโดยเนื้อหาที่สอบวัดแล้ว เป็นเนื้อหาที่ไม่ซับซ้อนและไม่ยากมากนักแต่กลับอยู่ในลำดับที่ยากเกินไป ตัวอย่างเช่น ข้อ 26 อยู่ในลำดับที่ 23 ข้อ 37 อยู่ในลำดับที่ 26 และข้อ 36 อยู่ในลำดับที่ 25 เป็นต้น

ลำดับที่ 23

ข้อ 26. ถ้า $(p-2)^2 = 25$ และ $(q+1)^2 = 81$ แล้ว ค่ามากที่สุดที่เป็นไปได้ของ p - 2q เท่ากับเท่าใด

ผู้สอบที่มีความน่าจะเป็น 0.67 ที่จะตอบข้อสอบข้อนี้ได้ถูกต้อง คือผู้สอบที่มีความสามารถ (θ) เท่ากับ 1.03

ลำดับที่ 25

ข้อ 36. ในการสำรวจน้ำหนักตัวของนักเรียนในชั้นเรียนที่มีนักเรียน 30 คน เป็นดังนี้

น้ำหนัก (กิโลกรัม)	ความถี่สะสม (คน)
30 – 49	10
50 – 69	26
70 - 89	30

ค่าเฉลี่ยเลขคณิตของน้ำหนักตัวของนักเรียนในชั้นนี้เท่ากับกี่กิโลกรัม

ผู้สอบที่มีความน่าจะเป็น 0.67 ที่จะตอบข้อสอบข้อนี้ได้ถูกต้อง คือผู้สอบที่มีความสามารถ (θ) เท่ากับ 1.29

ลำดับที่ 26

ข้อ 37. ข้อมูลชุดหนึ่งเรียงลำดับจากน้อยไปมากดังนี้

2 3 3 x 4 y 7

ค่าเฉลี่ยเลขคณิตและส่วนเบี่ยงเบนมาตรฐานของข้อมูลชุดนี้เท่ากับ 4 และ $\frac{4}{\sqrt{7}}$ ตามลำดับ แล้ว $y-x$ มีค่าเท่าใด

ผู้สอบที่มีความน่าจะเป็น 0.67 ที่จะตอบข้อสอบข้อนี้ได้ถูกต้อง คือผู้สอบที่มีความสามารถ (θ) เท่ากับ 1.37

ผู้เชี่ยวชาญส่วนใหญ่มีความคิดเห็นว่าข้อ 26 ข้อ 37 และข้อ 36 อยู่ในตำแหน่งที่ไม่เหมาะสม กล่าวคือ ข้อเหล่านี้ควรอยู่ในลำดับที่ต่ำกว่านี้ ดังตัวอย่างต่อไปนี้

"ครูว่า ข้อ 26 ยากพอ ๆ กับข้อ 24 ซึ่งอยู่ในลำดับที่ 7 นะ" (ข้อมูลจากการใช้เทคนิคการคิดออกเสียง)

"ข้อ 26 แทนค่าแล้วได้คำตอบเลยนะ เขาบอกค่า p, q มาแล้ว หาคำตอบง่ายมาก ไม่น่าจะอยู่ลำดับนี้หรือเปล่า เด็กสามารถแทนค่าได้เลย" (ข้อมูลจากการใช้เทคนิคการคิดออกเสียง)

"ครูว่าข้อ 26 แก้สมการแล้วหาค่าตัวแปรเพียงอย่างเดียว ไม่น่ายากนะ" (ข้อมูลจากการใช้เทคนิคการคิดออกเสียง)

"ข้อ 37 เขาให้มาหมดเลยนะ เพียงแต่ต้องหาคำตอบหลายขั้นหน่อย แต่ก็ไม่ยากมากนะ" (ข้อมูลจากการใช้เทคนิคการคิดออกเสียง)

"ข้อ 36 ง่ายกว่าข้อ 21 (ลำดับที่ 10), 22 (ลำดับที่ 19) อีกนะ ครูว่าไม่ต้องคิดซับซ้อนเลย" (ข้อมูลจากการใช้เทคนิคการคิดออกเสียง)

นอกจากนี้ข้อสอบแบบเลือกตอบบางข้อซึ่งผู้เชี่ยวชาญมองว่าเป็นข้อที่ยากกว่าข้อสอบแบบเติมคำตอบบางข้อ แต่กลับถูกเรียงไว้ในตำแหน่งที่ง่ายกว่า ตัวอย่างเช่น ข้อ 7 อยู่ในตำแหน่งที่ 13 ดังตัวอย่างต่อไปนี้

ลำดับที่ 19

ข้อ 7. ถ้าสมการ $\left(x^2+1\right)\left(2x^2-6x+c\right)=0$ มีรากที่เป็นจำนวนจริงเพียง 1 ราก ค่าของ c จะอยู่ในช่วงใดต่อไปนี้

1. $(0,3)$
2. $(3,6)$
3. $(6,9)$
4. $(9,12)$

ผู้สอบที่มีความน่าจะเป็น 0.67 ที่จะตอบข้อสอบข้อนี้ได้ถูกต้อง คือผู้สอบที่มีความสามารถ (θ) เท่ากับ 0.94

"ครูว่า ข้อ 7 ยากอยู่นะ ครูว่ามันยากกว่าข้อเติมคำบางข้ออีกนะ ดูนะ....ดูความซับซ้อน มันต้องประยุกต์ ต้องเข้าใจเรื่องหารากและการแก้สมการ เป็นการหาตัวร่วมแล้วดึงออกมา" (ข้อมูลจากการใช้เทคนิคการคิดออกเสียง)

"ข้อ 7 ค่อนข้างยากมาก มันต้องใช้ความรู้เกี่ยวกับการหาราก เป็นการแก้สมการ ครูว่ายากกว่าข้อเมื่อกี้อีก" (ข้อมูลจากการใช้เทคนิคการคิดออกเสียง)

ในประเด็นความไม่เหมาะสมของลำดับของข้อสอบบางข้อตามความคิดเห็นของผู้เชี่ยวชาญนั้น ผู้วิจัยได้เดินสอบถามจากนักศึกษาช่วยงานเกี่ยวกับประเด็นนี้เนื่องจากไม่ต้องการรบกวนผู้เชี่ยวชาญขณะกำหนดคะแนนจุดตัด พบว่า มีผู้เชี่ยวชาญหลายท่านตั้งข้อสังเกตเกี่ยวกับเรื่องนี้ ซึ่งทั้งนี้ผู้วิจัยเห็นว่าหากไม่พูดถึงประเด็นนี้ก่อนอาจเป็นปัญหาต่อการกำหนดคะแนนจุดตัดได้ ผู้วิจัยจึงอธิบายเกี่ยวกับลำดับของข้อสอบอีกครั้งว่าเกิดจากการเรียงตามลำดับความยากของข้อสอบ ซึ่งวิเคราะห์จากผลการตอบของผู้สอบ ดังนั้นอาจมีข้อสอบบางข้อที่

ผู้เชี่ยวชาญมองว่าเนื้อหาที่สอบวัดง่ายแต่ผู้สอบอาจจะทำไม่ได้ จึงทำให้ผลการวิเคราะห์สรุปว่าข้อสอบเป็นข้อยาก

จากนั้นผู้วิจัยจึงเปิดโอกาสให้ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดด้วยวิธีอื่น ๆ ร่วมกันอภิปรายในประเด็นนี้เพื่อหาข้อสรุป จากการอภิปราย พบว่า ผู้เชี่ยวชาญหลายท่านตั้งข้อสังเกตว่า รูปแบบของข้อสอบที่ใช้อาจมีผลต่อความยากของข้อสอบ กล่าวคือผู้เชี่ยวชาญให้คำอธิบายว่าข้อสอบแบบเติมคำส่วนใหญ่ เน้นวัดเนื้อหาที่ไม่ยากนัก และไม่ค่อยซับซ้อนเท่าไหร่ แต่ผลการตอบข้อสอบพบว่าผู้สอบกลับทำข้อสอบแบบนี้ได้ค่อนข้างน้อย แต่เมื่อพิจารณาข้อสอบแบบเลือกตอบซึ่งมีเนื้อหาที่สอบวัดค่อนข้างยากรวมถึงบางข้อต้องมีการประยุกต์ใช้ความรู้ แต่ผู้สอบกลับตอบถูกมากกว่าข้อสอบแบบเติมคำ ซึ่งประเด็นนี้เป็นที่มาของข้อสรุปในประเด็นที่อภิปรายว่า ผู้สอบส่วนใหญ่ทำข้อสอบแบบเติมคำตอบไม่ค่อยได้ ทั้งนี้อาจจะเป็นเพราะรูปแบบข้อสอบที่ผู้สอบไม่คุ้นเคย ผู้เชี่ยวชาญหลายท่านให้ความเห็นสนับสนุนว่า ก่อนที่นักเรียนจะสอบครูผู้สอนก็ได้นำข้อสอบเก่า ๆ มาให้นักเรียนทำ ซึ่งส่วนใหญ่เป็นข้อสอบแบบเลือกตอบ ดังนั้นนักเรียนจะถนัดกับข้อสอบแบบนี้ นอกจากนี้ผู้เชี่ยวชาญหลายท่านกล่าวว่า นักเรียนของเราส่วนใหญ่จะไม่ค่อยตั้งใจทำข้อสอบเท่าไหร่ ดังนั้นพฤติกรรมการทำข้อสอบจะไม่ค่อยรอบคอบมากนักจึงอาจทำให้คำนวณผิดพลาดบางขั้นตอนได้ หากเป็นข้อสอบแบบเลือกตอบนักเรียนก็จะเลือกคำตอบที่ใกล้เคียงกับคำตอบที่ทำได้ ซึ่งก็อาจมีโอกาสตอบถูก แต่ในกรณีที่ข้อสอบเป็นแบบเติมคำนั้น คำตอบที่ผู้สอบต้องตอบจะไม่สามารถหาคำตอบที่ใกล้เคียงได้ จึงส่งผลให้การทำข้อสอบแบบเติมคำมีผู้สอบทำถูกไม่มากนัก ซึ่งจากข้อสรุปนี้จึงนำไปสู่การที่ผู้เชี่ยวชาญกำหนดคะแนนจุดตัดโดยจะไม่พิจารณาเฉพาะเนื้อหาที่สอบวัดเท่านั้นแต่จะพิจารณาถึงรูปแบบของข้อสอบด้วย ซึ่งได้กล่าวไปแล้วก่อนหน้านี้

4.1.2.2.3 ผู้เชี่ยวชาญบางท่านมีแนวโน้มที่จะกำหนดคะแนนจุดตัดให้ใกล้เคียงกับผู้เชี่ยวชาญที่ตนเห็นว่าเก่งและอธิบาย หรือแสดงความคิดเห็นต่าง ๆ ได้มีหลักการดี

ข้อมูลจากการสังเกตพฤติกรรม และการบันทึกข้อสนทนา พบว่า ในการอภิปรายร่วมกันระหว่างผู้เชี่ยวชาญมักจะมีผู้เชี่ยวชาญ 2 - 3 ท่าน ที่จะแสดงความคิดเห็นก่อนท่านอื่นเสมอ และผู้เชี่ยวชาญท่านอื่นๆ จะให้ความเคารพและตั้งใจฟังมาก จากการสังเกตก็พบว่าผู้เชี่ยวชาญที่มักจะแสดงความคิดเห็นก่อนจะเป็นผู้ที่มีวัยวุฒิสูงกว่าผู้เชี่ยวชาญท่านอื่นๆ และเมื่อ

พิจารณาจากประวัติก็พบว่าเป็นครูที่มาจากโรงเรียนประจำจังหวัด ส่วนผู้เชี่ยวชาญที่อายุน้อยมักจะไม่ค่อยแสดงความคิดเห็นเท่าที่ควร

นอกจากนี้ข้อมูลจากการบันทึกข้อสนทนา พบว่า ผู้เชี่ยวชาญบางท่านเมื่อฟังหลักการและเหตุผลของผู้เชี่ยวชาญที่แสดงความคิดเห็นบ่อย ๆ แล้วก็รู้สึกคล้อยตามไปด้วย ดังตัวอย่างต่อไปนี้

"ครูคนนี้ดังนะ มาจากโรงเรียนราชสีมาวิทยาลัย อย่างเก่ง....อธิบายดี๊ดีนะ พูดแล้วเราเข้าใจเลยอ่ะ รู้สึกว่าข้อนี้ง่ายไปเลยจริงๆ" (ข้อมูลจากการบันทึกข้อสนทนา)

"คะแนนรอบแรกน้องให้เท่าไหร่ ของพี่ไม่ค่อยใกล้คนอื่นเลย พี่คิดไม่ถี่ถ้วนแน่เลย พี่เอาตัวพี่เองเป็นตัวตั้งหรือป่าว" (ข้อมูลจากการบันทึกข้อสนทนา)

"ตอนแรกพี่ว่าข้อสอบบางข้อมันง่ายนะ แต่พอฟังคนอื่นพูดแล้ว ก็รู้สึกว่ามันก็ยากสำหรับเด็กคนอื่นนะ คือเด็กเรา เราเน้นเรื่องนี้ไง เพราะเรารู้ว่าน่าจะออก" (ข้อมูลจากการบันทึกข้อสนทนา)

4.1.2.2.3 การอภิปรายร่วมกันระหว่างผู้เชี่ยวชาญหากดำเนินการไม่ดีอาจทำให้ผู้เชี่ยวชาญเบื่อหน่ายได้

ข้อมูลจากการใช้เทคนิคการคิดออกเสียง และการสังเกตพฤติกรรม พบว่า มีผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดด้วยวิธีแองกอฟแบบใช่/ไม่ใช่ และวิธีบุ๊คมาร์คบางท่านเกิดอาการเบื่อหน่ายในระหว่างที่อภิปรายร่วมกัน ซึ่งข้อมูลจากการสังเกตพฤติกรรม พบว่า ผู้เชี่ยวชาญบางท่านแสดงกริยาอาการที่บ่งบอกว่าไม่สนใจการอภิปราย และเบื่อหน่าย ดังตัวอย่างพฤติกรรม ได้แก่ นำโทรศัพท์ออกมากดเล่น หาวเป็นระยะ ๆ แคะเล็บ คุยกับเพื่อนนอกเหนือจากประเด็นที่กำลังอภิปราย เช่น ชวนกันไปกินบุฟเฟต์เป็นต้น และเมื่อผู้เชี่ยวชาญกำหนดคะแนนจุดตัดรอบที่ 2 ผู้วิจัยจึงให้นักศึกษาช่วยงานลองถามในประเด็นนี้ พบว่า ผู้เชี่ยวชาญบางท่านรู้สึกว่าการอภิปรายแบบให้ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดสูงที่สุดและต่ำที่สุดอธิบายเหตุผลที่กำหนดคะแนนจุดตัดนั้น ๆ ทำให้ใช้เวลาในการอภิปรายมาก และรู้สึกว่าไม่มีประโยชน์ต่อการกำหนดคะแนนจุดตัดของ

ตนเอง นอกจากนี้ผู้เชี่ยวชาญบางท่านก็พูดในประเด็นที่ไม่เกี่ยวกับกับการกำหนดคะแนนจุดตัดแต่เป็นเรื่องปัญหาต่าง ๆ ซ้ำ ๆ อาทิ ปัญหางานสอน การบริหารของผู้บริหาร เป็นต้น ซึ่งปัญหานี้ ผู้วิจัยได้พยายามปรับปรุงให้ดีขึ้นในการอภิปรายกลุ่มใหญ่ เพื่อให้การดำเนินการมีประสิทธิภาพมากขึ้น โดยกำหนดประเด็นในการอภิปรายสำหรับวิธีแองกอฟแบบใช่/ไม่ใช่และวิธีบุ๊คมาร์คใหม่ ให้เหมือนกับวิธีเอบีซี กล่าวคือให้ผู้เชี่ยวชาญทุกท่านอภิปรายเกี่ยวกับข้อสอบ หรือประเด็นที่ยังลังเลที่จะตัดสินใจ เพื่อให้ผู้เชี่ยวชาญท่านอื่นได้แสดงความคิดเห็นในประเด็นนั้น เพื่อให้ผู้เชี่ยวชาญได้สารสนเทศประกอบการตัดสินใจมากขึ้น และใช้เวลาในการอภิปรายน้อยลง อันเป็นการลดปัญหาความเบื่อหน่ายที่อาจจะเกิดขึ้นได้

4.1.2.3 ข้อเสนอแนะ

4.1.2.3.1 ผู้เชี่ยวชาญคิดว่าการอภิปรายกลุ่มช่วยให้ได้แลกเปลี่ยนความรู้ซึ่งกันและกัน ซึ่งเป็นประโยชน์อย่างยิ่งต่อการกำหนดคะแนนจุดตัด จึงควรให้เวลาสำหรับการอภิปรายให้มาก

ข้อมูลจากการสนทนากลุ่ม พบว่า ผู้เชี่ยวชาญส่วนใหญ่มีความคิดเห็นว่ากิจกรรมการอภิปรายกลุ่มไม่ว่าจะเป็นการอภิปรายกลุ่มย่อยหรือกลุ่มใหญ่ล้วนเป็นประโยชน์ต่อการกำหนดคะแนนจุดตัดทั้งสิ้น เนื่องจากเป็นการแลกเปลี่ยนความรู้ซึ่งกันและกัน อันเป็นการสร้างความมั่นใจในการกำหนดคะแนนจุดตัดให้กับผู้เชี่ยวชาญ นอกจากนี้ผู้เชี่ยวชาญเสนอว่าการอภิปรายกลุ่มไม่ควรเน้นให้ผู้ที่กำหนดคะแนนจุดตัดสูงที่สุดและต่ำที่สุดอธิบายเหตุผลของการกำหนดคะแนนจุดตัดนั้น ๆ แต่ควรเน้นให้ทุกคนได้สอบถามหรือตั้งประเด็นที่ตนเองไม่เข้าใจหรือไม่แน่ใจ เพื่อให้ผู้เชี่ยวชาญท่านอื่น ๆ ได้ร่วมแสดงความคิดเห็นหรือแลกเปลี่ยนความรู้เพื่อสร้างความกระจ่างให้มากขึ้น

4.1.2.3.2 ผู้เชี่ยวชาญเสนอแนะว่าการพิจารณาข้อสอบที่ถูกเรียงตามลำดับความยากซึ่งถูกแบ่งเป็นฉบับย่อย ๆ ตามสาระการเรียนรู้ ควรจะแบ่งในแต่ละฉบับย่อยตามระดับพฤติกรรมด้านพุทธิพิสัยที่วัด เช่น ความรู้ความจำ ความเข้าใจ การนำไปประยุกต์ใช้ และการวิเคราะห์ เป็นต้น ซึ่งจะช่วยให้การพิจารณากำหนดคะแนนจุดตัดทำได้ง่ายเร็วขึ้น

เนื่องจากข้อมูลที่เก็บมาจากการใช้เทคนิคการคิดออกเสียง และการสอบถามจากนักศึกษาช่วยงาน พบว่า ผู้เชี่ยวชาญส่วนใหญ่จะกำหนดคะแนนจุดตัดโดยพิจารณาระดับ

พฤติกรรมด้านพุทธิพิสัยที่ข้อสอบต้องการวัด ดังนั้นผู้วิจัยจึงตั้งประเด็นการสนทนาเกี่ยวกับการจัดกลุ่มข้อสอบตามระดับพฤติกรรมด้านพุทธิพิสัยที่วัด เช่น ความรู้ความจำ ความเข้าใจ การนำไปประยุกต์ใช้ และการวิเคราะห์ เป็นต้น จะดีหรือมีประโยชน์ต่อการกำหนดคะแนนจุดตัดของผู้เชี่ยวชาญหรือไม่ ซึ่งข้อมูลจากการสนทนากลุ่ม พบว่า ผู้เชี่ยวชาญส่วนใหญ่เห็นด้วยว่าการพิจารณากำหนดคะแนนจุดตัดจากข้อสอบที่มีการจัดไว้เป็นกลุ่มตามระดับพฤติกรรมด้านพุทธิพิสัยน่าจะช่วยให้การกำหนดคะแนนจุดตัดทำได้เร็วขึ้น โดยผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดด้วยวิธีเอบีซีส่วนใหญ่เสนอว่า หากข้อสอบแต่ละฉบับย่อย ๆ ตามสาระการเรียนรู้ที่ถูกเรียงไว้ตามลำดับความยากตามหลักการของวิธีเอบีซี มีการแบ่งกลุ่มข้อสอบในแต่ละฉบับย่อยตามระดับพฤติกรรมด้านพุทธิพิสัยที่วัดน่าจะช่วยให้การดำเนินการได้เร็วขึ้นกว่าเดิม โดยผู้เชี่ยวชาญคิดว่าน่าจะใช้เวลาน้อยกว่าวิธีบุ๊คมาร์ค

4.1.2.3.3 ผู้เชี่ยวชาญเสนอแนะให้นักเรียนได้มีส่วนร่วมในการกำหนดคะแนนจุดตัดด้วย

ข้อมูลจากการสนทนากลุ่ม พบว่า ผู้เชี่ยวส่วนใหญ่มีความคิดเห็นว่าควรให้นักเรียนได้มีส่วนร่วมในการกำหนดคะแนนจุดตัดด้วย เนื่องจากว่าบางครั้งผู้เชี่ยวชาญอาจจะกำหนดคะแนนจุดตัดโดยยึดเอาตนเองเป็นหลักได้ กล่าวคือข้อที่ผู้เชี่ยวชาญคิดว่าง่ายอาจจะยากสำหรับนักเรียนได้ ซึ่งความคิดเห็นเกี่ยวกับข้อสอบของนักเรียนเป็นสิ่งสำคัญที่จะทำให้ผู้เชี่ยวได้แลกเปลี่ยนมุมมองกับนักเรียนซึ่งน่าจะช่วยให้การกำหนดคะแนนจุดตัดมีความเหมาะสมและเป็นธรรมกับทุกฝ่ายมากขึ้น

4.1.3 ผลการวิเคราะห์ความตรงของผลลัพธ์

ผลการวิเคราะห์ความสอดคล้องของคะแนนจุดตัดที่กำหนดโดยวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซี โดยการวิเคราะห์ค่าสัมประสิทธิ์แคปป้า (Kappa coefficient: K)

4.1.3.1 ผลการวิเคราะห์ความสอดคล้องของคะแนนจุดตัดที่กำหนดโดยวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซี วิชาคณิตศาสตร์ รายละเอียดดังแสดงในตารางต่อไปนี้

ตารางที่ 4.15 ผลการวิเคราะห์ความสอดคล้องของคะแนนจุดตัดที่กำหนดโดยวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซี วิชาคณิตศาสตร์

รอบ	วิธี	วิธีแองกอฟแบบใช่/ไม่ใช่	วิธีบุ๊คมาร์ค	วิธีเอบีซี
รอบที่ 1	วิธีแองกอฟแบบใช่/ไม่ใช่	-	-	-
	วิธีบุ๊คมาร์ค	0.415	-	-
	วิธีเอบีซี	0.189	0.059	-
รอบที่ 2	วิธีแองกอฟแบบใช่/ไม่ใช่	-	-	-
	วิธีบุ๊คมาร์ค	0.293	-	-
	วิธีเอบีซี	0.106	0.026	-
รอบที่ 3	วิธีแองกอฟแบบใช่/ไม่ใช่	-	-	-
	วิธีบุ๊คมาร์ค	0.041	-	-
	วิธีเอบีซี	0.324	0.028	
เฉลี่ย	วิธีแองกอฟแบบใช่/ไม่ใช่	-	-	-
	วิธีบุ๊คมาร์ค	0.189	-	-
	วิธีเอบีซี	0.191	0.038	-

จากตาราง พบว่า คะแนนจุดตัดที่กำหนดโดยวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซี มีความสอดคล้องกันน้อย

4.1.3.2 ผลการวิเคราะห์ความสอดคล้องของคะแนนจุดตัดที่กำหนดโดยวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซี วิชาวิทยาศาสตร์ รายละเอียดดังแสดงในตารางต่อไปนี้

ตารางที่ 4.16 ผลการวิเคราะห์ความสอดคล้องของคะแนนจุดตัดที่กำหนดโดยวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซี วิชาวิทยาศาสตร์

รอบ	วิธี	วิธีแองกอฟแบบใช่/ไม่ใช่	วิธีบุ๊คมาร์ค	วิธีเอบีซี
รอบที่ 1	วิธีแองกอฟแบบใช่/ไม่ใช่	-	-	-
	วิธีบุ๊คมาร์ค	0.017	-	-
	วิธีเอบีซี	0.148	0.343	-
รอบที่ 2	วิธีแองกอฟแบบใช่/ไม่ใช่	-	-	-
	วิธีบุ๊คมาร์ค	0.002	-	-
	วิธีเอบีซี	0.003	0.381	-
รอบที่ 3	วิธีแองกอฟแบบใช่/ไม่ใช่	-	-	-
	วิธีบุ๊คมาร์ค	0.004	-	-
	วิธีเอบีซี	0.004	0.245	-
เฉลี่ย	วิธีแองกอฟแบบใช่/ไม่ใช่	-	-	-
	วิธีบุ๊คมาร์ค	0.003	-	-
	วิธีเอบีซี	0.015	0.600	-

จากตาราง พบว่า คะแนนจุดตัดที่กำหนดโดยวิธีแองกอฟแบบใช่/ไม่ใช่ มีความสอดคล้องน้อยกับคะแนนจุดตัดที่กำหนดโดยวิธีบุ๊คมาร์ค และวิธีเอบีซี ส่วนคะแนนจุดตัดคะแนนจุดตัดที่กำหนดโดยวิธีบุ๊คมาร์ค และวิธีเอบีซี เฉลี่ย มีความสอดคล้องกันในระดับปานกลาง

4.1.4 ผลการวิเคราะห์เพื่อตรวจสอบความตรงของผลที่ตามมา

ผลการวิเคราะห์ความถูกต้องของการกำหนดกลุ่ม (Classification accuracy: $\hat{\gamma}$) และความคลาดเคลื่อนของการกำหนดกลุ่ม (Classification errors) ได้แก่ ความคลาดเคลื่อนแบบบวก (false positive: $\hat{\gamma}^{+}$) และความคลาดเคลื่อนแบบลบ (false negative : $\hat{\gamma}^{-}$) ของคะแนนจุดตัดวิชาคณิตศาสตร์ และวิทยาศาสตร์

4.1.4.1 ผลการวิเคราะห์ความถูกต้องของการกำหนดกลุ่ม ความคลาดเคลื่อนของการกำหนดกลุ่มแบบบวก และความคลาดเคลื่อนของการกำหนดกลุ่มแบบลบของคะแนนจุดตัดวิชาคณิตศาสตร์ ดังแสดงในตารางต่อไปนี้

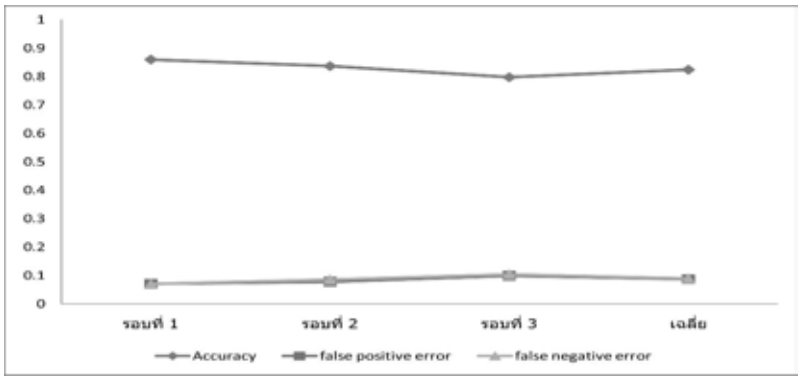
ตารางที่ 4.17 ผลการวิเคราะห์ความถูกต้องของการกำหนดกลุ่ม ความคลาดเคลื่อนของการกำหนดกลุ่มแบบบวก และความคลาดเคลื่อนของการกำหนดกลุ่มแบบลบของคะแนนจุดตัดวิชาคณิตศาสตร์

วิธีกำหนดคะแนนจุดตัด	$\hat{\gamma}$	$\hat{\gamma}^{+}$	$\hat{\gamma}^{-}$
วิธีแองกอฟแบบใช่/ไม่ใช่			
รอบที่ 1	0.8597	0.0706	0.0697
รอบที่ 2	0.8367	0.0781	0.0852
รอบที่ 3	0.7976	0.0988	0.1036
เฉลี่ย	0.8240	0.0870	0.0890
วิธีบุ๊คมาร์ค			
รอบที่ 1	0.9137	0.0440	0.0423
รอบที่ 2	0.9103	0.0479	0.0418
รอบที่ 3	0.9103	0.0479	0.0418
เฉลี่ย	0.9120	0.0490	0.0390
วิธีเอบีซี			
รอบที่ 1	0.8353	0.0771	0.0876
รอบที่ 2	0.8830	0.0382	0.0788
รอบที่ 3	0.8865	0.0432	0.0703
เฉลี่ย	0.8520	0.0410	0.1070

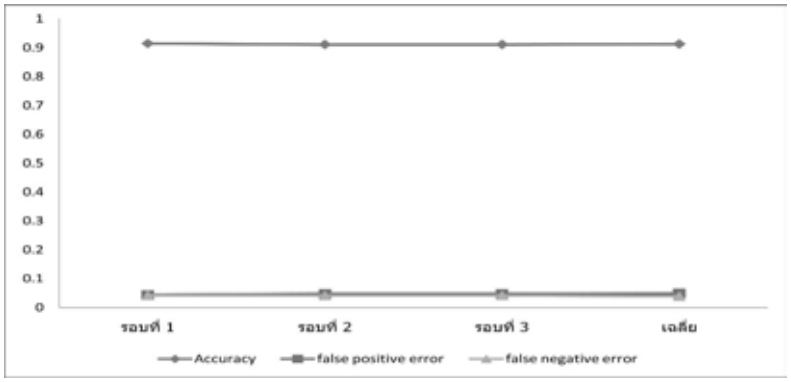
จากตาราง พบว่า ความถูกต้องของการกำหนดกลุ่มของคะแนนจุดตัดที่กำหนดขึ้นจาก 3 วิธีมีค่าแตกต่างกัน โดยวิธีบุ๊คมาร์คมีค่าความถูกต้องของการกำหนดกลุ่มเฉลี่ยสูงที่สุด มีค่า 0.9120 รองลงมาคือวิธีเอบีซี มีค่า 0.8520 ส่วนวิธีแองกอฟแบบใช่/ไม่ใช่มีค่าความถูกต้องของการกำหนดกลุ่มเฉลี่ยต่ำที่สุด มีค่า 0.8240 เมื่อพิจารณาความคลาดเคลื่อนในการกำหนดกลุ่มของคะแนนจุดตัด โดยพิจารณาจากความคลาดเคลื่อนแบบบวก พบว่า วิธีแองกอฟแบบใช่/ไม่ใช่มีความคลาดเคลื่อนแบบบวกเฉลี่ยสูงที่สุด มีค่า 0.0870 รองลงมาคือวิธีบุ๊คมาร์ค มีค่า 0.0490 ส่วนวิธีเอบีซีมีความคลาดเคลื่อนแบบบวกเฉลี่ยต่ำที่สุด มีค่า 0.0410 และเมื่อพิจารณาความคลาดเคลื่อนแบบลบ พบว่า วิธีเอบีซีมีความคลาดเคลื่อนแบบลบสูงที่สุด มีค่า 0.1070 รองลงมาคือวิธีแองกอฟแบบใช่/ไม่ใช่ มีค่า 0.0890 ส่วนวิธีบุ๊คมาร์ค มีความคลาดเคลื่อนแบบลบต่ำที่สุด มีค่า 0.0390

เพื่อแสดงให้เห็นความแตกต่างของความถูกต้องของการกำหนดกลุ่ม ความคลาดเคลื่อนของการกำหนดกลุ่มแบบบวก และความคลาดเคลื่อนของการกำหนดกลุ่มแบบลบของคะแนนจุดตัดวิชาคณิตศาสตร์แต่ละวิธี ผู้วิจัยขอนำเสนอผลเป็นกราฟเส้น ดังแสดงในภาพต่อไปนี้

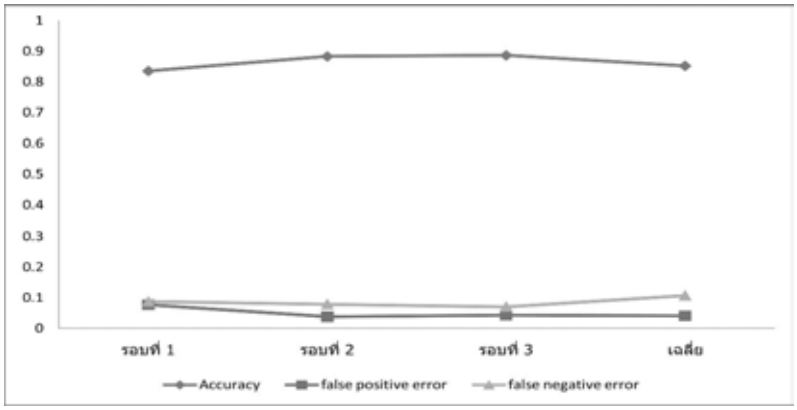
วิธีแองกอฟแบบใช่/ไม่ใช่



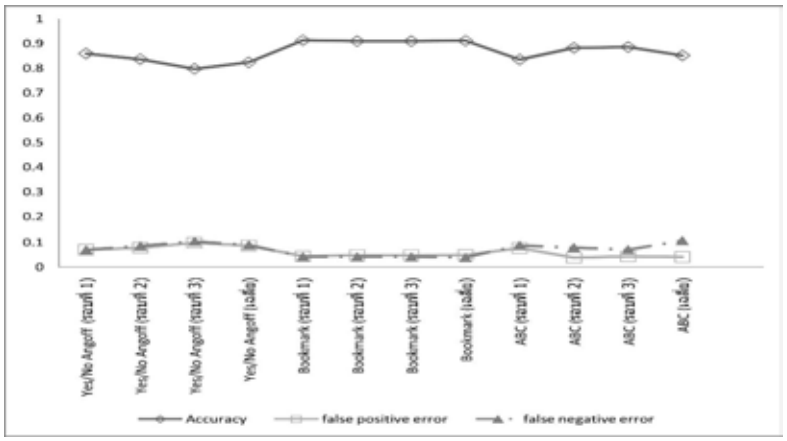
วิธีบุ๊คมาร์ค



วิธีเอบีซี



วิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซี



ภาพที่ 4.9 ความถูกต้องของการกำหนดกลุ่ม ความคลาดเคลื่อนของการกำหนดกลุ่มแบบบวก และความคลาดเคลื่อนของการกำหนดกลุ่มแบบลบของคะแนนจุดตัดวิชาคณิตศาสตร์

4.1.4.2 ผลการวิเคราะห์ความถูกต้องของการกำหนดกลุ่ม ความคลาดเคลื่อนของการกำหนดกลุ่มแบบบวก และความคลาดเคลื่อนของการกำหนดกลุ่มแบบลบของคะแนนจุดตัดวิชาวิทยาศาสตร์ ดังแสดงในตารางต่อไปนี้

ตารางที่ 4.18 ผลการวิเคราะห์ความถูกต้องของการกำหนดกลุ่ม ความคลาดเคลื่อนของการกำหนดกลุ่มแบบบวก และความคลาดเคลื่อนของการกำหนดกลุ่มแบบลบของคะแนนจุดตัดวิชาวิทยาศาสตร์

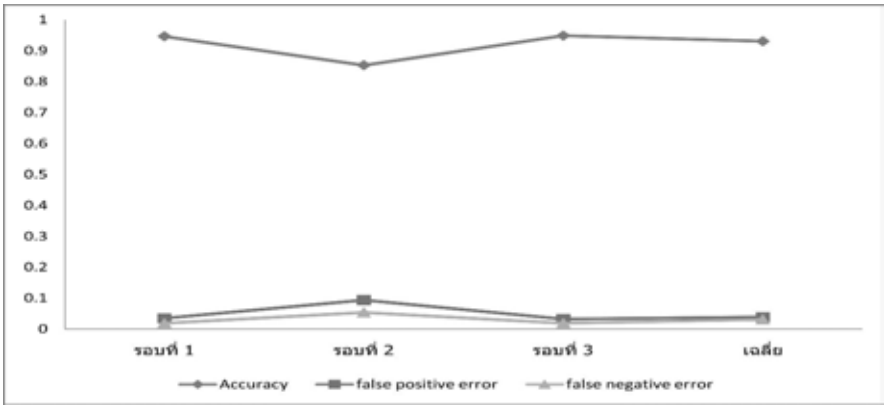
วิธีกำหนดคะแนนจุดตัด	$\hat{\gamma}$	$\hat{\gamma}^{+}$	$\hat{\gamma}^{-}$
วิธีแองกอฟแบบใช่/ไม่ใช่			
รอบที่ 1	0.9465	0.0345	0.0190
รอบที่ 2	0.8530	0.0934	0.0536
รอบที่ 3	0.9490	0.0321	0.0190
เฉลี่ย	0.9307	0.0386	0.0308
วิธีบุ๊คมาร์ค			
รอบที่ 1	0.9234	0.0304	0.0462
รอบที่ 2	0.8705	0.0569	0.0726
รอบที่ 3	0.8705	0.0569	0.0726
เฉลี่ย	0.8907	0.0363	0.0731
วิธีเอบีซี			
รอบที่ 1	0.8957	0.0409	0.0635
รอบที่ 2	0.9338	0.0153	0.0509
รอบที่ 3	0.9534	0.0165	0.0301
เฉลี่ย	0.9359	0.0188	0.0453

จากตาราง พบว่า ความถูกต้องของการกำหนดกลุ่มของคะแนนจุดตัดที่กำหนดขึ้นจาก 3 วิธีมีค่าแตกต่างกัน โดยวิธีเอบีซีมีความถูกต้องของการกำหนดกลุ่มเฉลี่ยสูงที่สุด มีค่า 0.9359 รองลงมาคือวิธีแองกอฟแบบใช่/ไม่ใช่มีค่า 0.9307 ส่วนวิธีบุ๊คมาร์คมีความถูกต้องของการกำหนดกลุ่มเฉลี่ยต่ำที่สุด มีค่า 0.8907 เมื่อพิจารณาความคลาดเคลื่อนของการกำหนดกลุ่มของคะแนนจุดตัด โดยพิจารณาจากความคลาดเคลื่อนแบบบวก พบว่า วิธีแองกอฟแบบใช่/ไม่ใช่มีความคลาดเคลื่อนแบบบวกเฉลี่ยสูงที่สุด มีค่า 0.0870 รองลงมาคือวิธีบุ๊คมาร์ค มีค่า 0.0490 ส่วนวิธีเอบีซีมีความคลาดเคลื่อนแบบบวกเฉลี่ยต่ำที่สุด มีค่า 0.0188 ส่วนความคลาดเคลื่อนแบบลบ พบว่า วิธีบุ๊คมาร์คมี

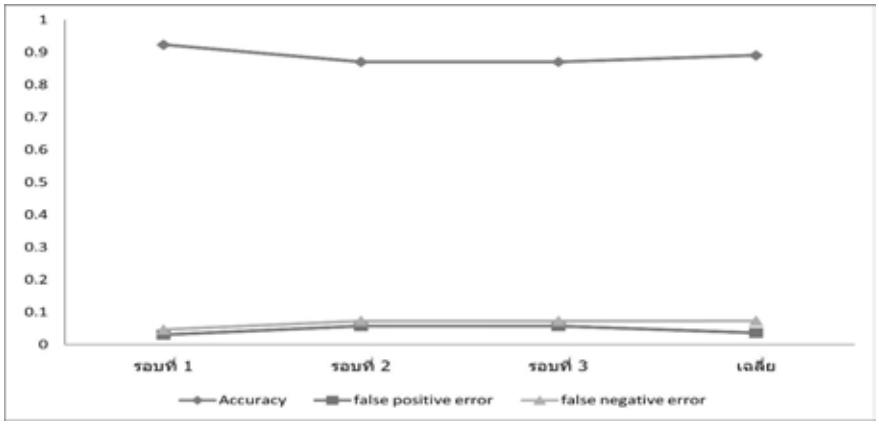
ความคลาดเคลื่อนแบบลบสูงที่สุด มีค่า 0.0731 รองลงมาคือวิธีเอบีซี มีค่า 0.0453 ส่วนวิธีแองกอฟแบบใช่/ไม่ใช่ มีความคลาดเคลื่อนแบบลบต่ำที่สุด มีค่า 0.0390

เพื่อแสดงให้เห็นความแตกต่างของความถูกต้องของการกำหนดกลุ่ม ความคลาดเคลื่อนของการกำหนดกลุ่มแบบบวก และความคลาดเคลื่อนของการกำหนดกลุ่มแบบลบของคะแนนจุดตัดวิชาวิทยาศาสตร์แต่ละวิธี ผู้วิจัยขอนำเสนอผลเป็นกราฟเส้น ดังแสดงในภาพต่อไปนี้

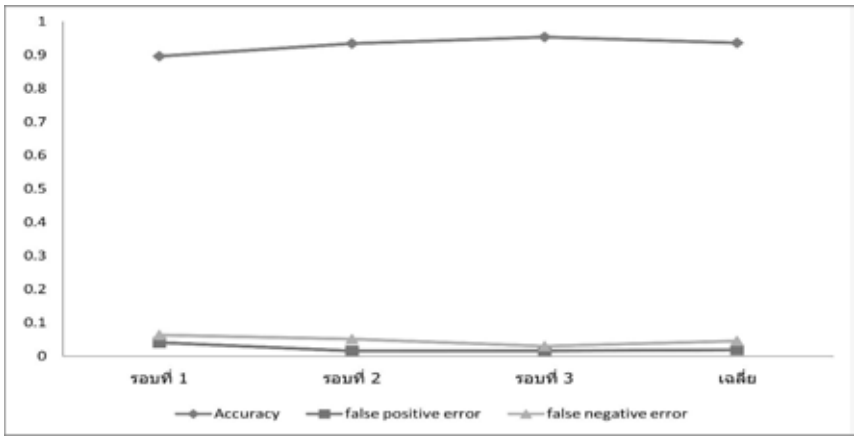
วิธีแองกอฟแบบใช่/ไม่ใช่



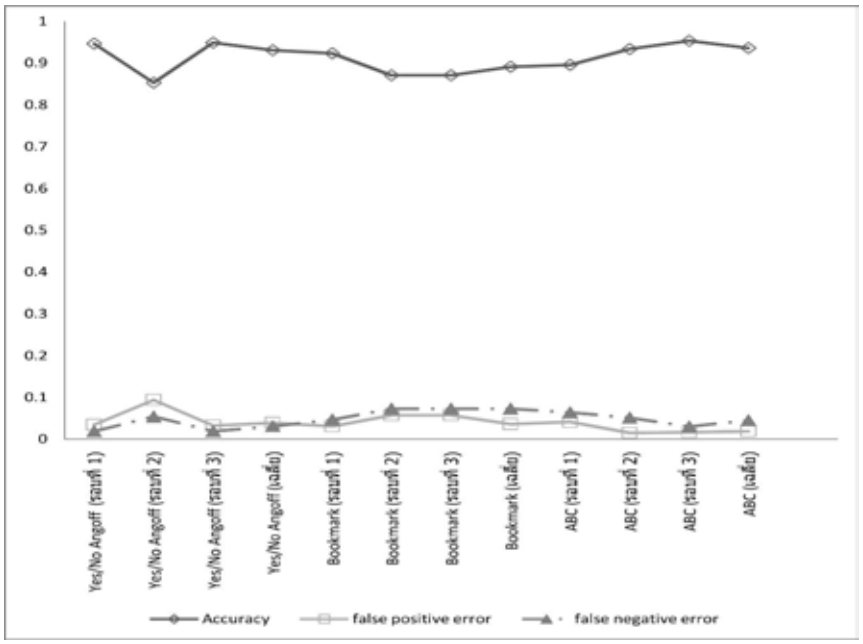
วิธีบุ๊คมาร์ค



วิธีเอบีซี



วิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซี



ภาพที่ 4.10 ความถูกต้องของการกำหนดกลุ่ม ความคลาดเคลื่อนของการกำหนดกลุ่มแบบบวก และความคลาดเคลื่อนของการกำหนดกลุ่มแบบลบของคะแนนจุดตัดวิชาวิทยาศาสตร์

4.2 ผลการวิเคราะห์ความเที่ยงของคะแนนจุดตัด

4.2.1 ผลการวิเคราะห์ความเที่ยงซึ่งแทนด้วยค่าสัมประสิทธิ์การสรุปอ้างอิง และวิเคราะห์จำนวนผู้เชี่ยวชาญ และจำนวนกลุ่มผู้เชี่ยวชาญ ที่มีความเหมาะสม

4.2.1.1 ผลการศึกษาขนาดขององค์ประกอบความแปรปรวนจากแหล่งความคลาดเคลื่อนต่าง ๆ และค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์ของการกำหนดคะแนนจุดตัดวิชาคณิตศาสตร์ด้วยวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซี มีรายละเอียดดังต่อไปนี้

4.2.1.1.1 ผลการศึกษาขนาดขององค์ประกอบความแปรปรวนจากแหล่งความคลาดเคลื่อนต่าง ๆ ของการกำหนดคะแนนจุดตัดด้วยวิธีแองกอฟแบบใช่/ไม่ใช่

ตารางที่ 4.19 ผลการศึกษาขนาดขององค์ประกอบความแปรปรวนจากแหล่งความคลาดเคลื่อนต่าง ๆ ของการกำหนดคะแนนจุดตัดด้วยวิธีแองกอฟแบบใช่/ไม่ใช่

Source of Variation	df	Sum of square (SS)	Mean Square (MS)	Estimated Variance Component	Percent of Total Variability
			Level 1/2		
g	2	196.2222	98.1111	0	0
p:g	6	933.1111	155.5185	30.5741	24.42
r	2	609.5556	304.7778	24.7469	19.77
gr	4	328.2222	82.0556	6.0864	4.86
pr:g,e	12	765.5556	63.7963	63.7963	50.95
Total	26	2832.6667		125.2037	100.00
			Level 2/3		
g	2	102.7407	51.3704	0	0
p:g	6	1008.2222	168.0370	6.6296	1.45
r	2	5068.0741	2534.0370	247.1296	54.22
gr	4	1239.4815	309.8704	53.9074	11.83
pr:g,e	12	1777.7778	148.1482	148.1481	32.50
Total	26	9196.2963		455.8147	100
			Level 3/4		
g	2	213.4074	106.7037	7.6420	23.38
p:g	6	210.4444	35.0741	5.8519	17.91
r	2	53.6296	26.8148	0.7160	2.19
gr	4	81.4815	20.3704	0.9506	2.91
pr:g,e	12	210.2222	17.5185	17.5185	53.61
Total	26	769.1852		32.6790	100.00

จากตาราง พบว่า ขนาดขององค์ประกอบความแปรปรวนจากแหล่งความคลาดเคลื่อนต่าง ๆ ของการกำหนดคะแนนจุดตัดที่ระดับ 1/2 เรียงจากมากไปหาน้อย ได้แก่ องค์ประกอบความแปรปรวนส่วนที่เหลือ (*pr:g, e*) มีค่า 63.7963 คิดเป็นร้อยละ 50.95 รองลงมาคือ

องค์ประกอบความแปรปรวนของผู้เชี่ยวชาญในแต่ละกลุ่ม (*p:g*) มีค่า 30.5741 คิดเป็นร้อยละ 24.42 องค์ประกอบความแปรปรวนของรอบในการตัดสิน (*r*) มีค่า 24.7469 คิดเป็นร้อยละ 19.77 องค์ประกอบความแปรปรวนของปฏิสัมพันธ์ระหว่างกลุ่มผู้เชี่ยวชาญและรอบในการตัดสิน (*gr*) มีค่า 6.0864 คิดเป็นร้อยละ 4.86 และองค์ประกอบความแปรปรวนของกลุ่มผู้เชี่ยวชาญ(*g*) มีค่า 0 แสดงว่าความแตกต่างของคะแนนจุดตัดที่ระดับ 1/2 ส่วนใหญ่เนื่องมาจากองค์ประกอบความแปรปรวนส่วนที่เหลือ รองลงมาเกิดจากความแตกต่างกันของผู้เชี่ยวชาญในแต่ละกลุ่ม ส่วนกลุ่มผู้เชี่ยวชาญไม่มีความแตกต่างกันในการกำหนดคะแนนจุดตัดที่ระดับ 1/2

ขนาดขององค์ประกอบความแปรปรวนจากแหล่งความคลาดเคลื่อนต่าง ๆ ของการกำหนดคะแนนจุดตัดที่ระดับ 2/3 เรียงจากมากไปหาน้อย ได้แก่องค์ประกอบความแปรปรวนของรอบในการตัดสิน (*r*) มีค่า 247.1296 คิดเป็นร้อยละ 54.22 รองลงมาคือองค์ประกอบความแปรปรวนส่วนที่เหลือ (*pr:g, e*) มีค่า 148.1481 คิดเป็นร้อยละ 32.50 องค์ประกอบความแปรปรวนของปฏิสัมพันธ์ระหว่างกลุ่มผู้เชี่ยวชาญและรอบในการตัดสิน (*gr*) มีค่า 53.9074 คิดเป็นร้อยละ 11.83 องค์ประกอบความแปรปรวนของผู้เชี่ยวชาญในแต่ละกลุ่ม (*p:g*) มีค่า 6.6296 คิดเป็นร้อยละ 1.45 และองค์ประกอบความแปรปรวนของกลุ่มผู้เชี่ยวชาญ (*g*) มีค่า 0 แสดงว่าความแตกต่างของคะแนนจุดตัดที่ระดับ 2/3 ส่วนใหญ่เนื่องมาจากความแตกต่างของรอบในการตัดสิน รองลงมาคือองค์ประกอบความแปรปรวนส่วนที่เหลือ ส่วนกลุ่มผู้เชี่ยวชาญไม่มีความแตกต่างกันในการกำหนดคะแนนจุดตัดที่ระดับ 2/3

ขนาดขององค์ประกอบความแปรปรวนจากแหล่งความคลาดเคลื่อนต่าง ๆ ของการกำหนดคะแนนจุดตัดที่ระดับ 3/4 เรียงจากมากไปหาน้อย ได้แก่องค์ประกอบความแปรปรวนส่วนที่เหลือ (*pr:g, e*) มีค่า 17.5185 คิดเป็นร้อยละ 53.61 รองลงมาคือ องค์ประกอบความแปรปรวนของกลุ่มผู้เชี่ยวชาญ (*g*) มีค่า 7.6420 คิดเป็นร้อยละ 23.38 องค์ประกอบความแปรปรวนของผู้เชี่ยวชาญในแต่ละกลุ่ม (*p:g*) มีค่า 5.8519 คิดเป็นร้อยละ 17.91 องค์ประกอบความแปรปรวนของปฏิสัมพันธ์ระหว่างกลุ่มผู้เชี่ยวชาญและรอบในการตัดสิน (*gr*) มีค่า 0.9506 คิดเป็นร้อยละ 2.91 และองค์ประกอบความแปรปรวนของรอบในการตัดสิน (*r*) มีค่า 0.7160 คิดเป็นร้อยละ 2.19 แสดงว่าความแตกต่างของคะแนนจุดตัดที่ระดับ 3/4 ส่วนใหญ่เนื่องมาจากองค์ประกอบความแปรปรวน

ส่วนที่เหลือ รองลงมาคือความแตกต่างกันของกลุ่มผู้เชี่ยวชาญ และคะแนนจุดตัดมีความแตกต่างกันน้อยในแต่ละรอบของการตัดสิน

4.2.1.1.2 ผลการศึกษาขนาดขององค์ประกอบความแปรปรวนจากแหล่งความคลาดเคลื่อนต่าง ๆ ของการกำหนดคะแนนจุดตัดวิชาคณิตศาสตร์ด้วยวิธีบุ๊คมาร์ค

ตารางที่ 4.20 ผลการศึกษาขนาดขององค์ประกอบความแปรปรวนจากแหล่งความคลาดเคลื่อนต่าง ๆ ของการกำหนดคะแนนจุดตัดวิชาคณิตศาสตร์ด้วยวิธีบุ๊คมาร์ค

Source of Variation	df	Sum of square (SS)	Mean Square (MS)	Estimated Variance Component	Percent of Total Variability
			Level 1/2		
g	2	231.1852	115.5926	0	0
p:g	6	1365.1111	227.5185	71.7407	80.56
r	2	101.4074	50.7037	5.0123	5.63
gr	4	22.3704	5.5926	0	0
pr:g,e	12	147.5556	12.2963	12.2963	13.81
Total	26	1867.6296		89.0493	100.00
			Level 2/3		
g	2	24.0741	12.0370	0	0
p:g	6	308.2222	51.3703	15.3704	69.51
r	2	26.7407	13.3703	1.4815	6.70
gr	4	0.1482	0.0370	0	0
pr:g,e	12	63.1111	5.2592	5.2593	23.79
Total	26	422.2963		22.1112	100.00
			Level 2/3		
g	2	87.6296	43.8148	0	0
p:g	6	512.8889	85.4815	26.6667	77.34
r	2	46.2963	23.1482	1.7778	5.16
gr	4	28.5926	7.1482	0.5556	1.61
pr:g,e	12	65.7778	5.4815	5.4815	15.90
Total	26	741.1852	165.0741	34.4816	100.00

จากตาราง พบว่า ขนาดขององค์ประกอบความแปรปรวนจากแหล่งความคลาดเคลื่อนต่าง ๆ ของการกำหนดคะแนนจุดตัดที่ระดับ 1/2 เรียงจากมากไปหาน้อย ได้แก่

องค์ประกอบความแปรปรวนของผู้เชี่ยวชาญในแต่ละกลุ่ม (*p:g*) มีค่า 71.7407 คิดเป็นร้อยละ 80.56 รองลงมาคือองค์ประกอบความแปรปรวนส่วนที่เหลือ (*pr:g, e*) มีค่า 12.2963 คิดเป็นร้อยละ 13.81 องค์ประกอบความแปรปรวนของรอบในการตัดสิน (*r*) มีค่า 5.0123 คิดเป็นร้อยละ 5.63 ส่วนองค์ประกอบความแปรปรวนของกลุ่มผู้เชี่ยวชาญ(*g*) และองค์ประกอบความแปรปรวนของปฏิสัมพันธ์ระหว่างกลุ่มผู้เชี่ยวชาญและรอบในการตัดสิน (*gr*) มีค่า 0 แสดงว่าความแตกต่างของคะแนนจุดตัดที่ระดับ 1/2 ส่วนใหญ่เนื่องมาจากความแตกต่างกันของผู้เชี่ยวชาญในแต่ละกลุ่ม รองลงมาคือองค์ประกอบความแปรปรวนส่วนที่เหลือ ส่วนกลุ่มผู้เชี่ยวชาญ และผู้เชี่ยวชาญแต่ละกลุ่มไม่มีความแตกต่างกันในการกำหนดคะแนนจุดตัดที่ระดับ 1/2

ขนาดขององค์ประกอบความแปรปรวนจากแหล่งความคลาดเคลื่อนต่าง ๆ ของการกำหนดคะแนนจุดตัดที่ระดับ 2/3 เรียงจากมากไปหาน้อย ได้แก่องค์ประกอบความแปรปรวนของผู้เชี่ยวชาญในแต่ละกลุ่ม (*p:g*) มีค่า 15.3704 คิดเป็นร้อยละ 69.51 รองลงมาคือองค์ประกอบความแปรปรวนส่วนที่เหลือ (*pr:g, e*) มีค่า 5.2593คิดเป็นร้อยละ 23.79 องค์ประกอบความแปรปรวนของรอบในการตัดสิน (*r*) มีค่า 1.4815 คิดเป็นร้อยละ 6.70 ส่วนองค์ประกอบความแปรปรวนของกลุ่มผู้เชี่ยวชาญ (*g*) และองค์ประกอบความแปรปรวนของปฏิสัมพันธ์ระหว่างกลุ่มผู้เชี่ยวชาญและรอบในการตัดสิน (*gr*) มีค่า 0 แสดงว่าความแตกต่างของคะแนนจุดตัดที่ระดับ 2/3 ส่วนใหญ่เนื่องมาจากความแตกต่างกันของผู้เชี่ยวชาญในแต่ละกลุ่ม รองลงมาคือองค์ประกอบความแปรปรวนส่วนที่เหลือ ส่วนกลุ่มผู้เชี่ยวชาญ และผู้เชี่ยวชาญในแต่ละกลุ่มไม่มีความแตกต่างกันในการกำหนดคะแนนจุดตัดที่ระดับ 2/3

ขนาดขององค์ประกอบความแปรปรวนจากแหล่งความคลาดเคลื่อนต่าง ๆ ของการกำหนดคะแนนจุดตัดระดับ 3/4 เรียงจากมากไปหาน้อย ได้แก่องค์ประกอบความแปรปรวนของผู้เชี่ยวชาญในแต่ละกลุ่ม (*p:g*) มีค่า 26.6667 คิดเป็นร้อยละ 77.34 รองลงมาคือองค์ประกอบความแปรปรวนส่วนที่เหลือ (p*r:g, e*) มีค่า 5.4815 คิดเป็นร้อยละ 15.90 องค์ประกอบความแปรปรวนของรอบในการตัดสิน (*r*) มีค่า 1.7778 คิดเป็นร้อยละ 5.16 และองค์ประกอบความแปรปรวนของปฏิสัมพันธ์ระหว่างกลุ่มผู้เชี่ยวชาญและรอบในการตัดสิน (*gr*) มีค่า 0.5556 คิดเป็นร้อยละ 1.61 ส่วนองค์ประกอบความแปรปรวนของกลุ่มผู้เชี่ยวชาญ (*g*) มีค่า 0 แสดงว่า ความแตกต่างของคะแนนจุดตัดที่ระดับ 3/4 ส่วนใหญ่เนื่องมาจากความแตกต่างกันของผู้เชี่ยวชาญในแต่ละกลุ่ม

รองลงมาคือองค์ประกอบความแปรปรวนส่วนที่เหลือ ส่วนกลุ่มผู้เชี่ยวชาญไม่มีความแตกต่างกันในการกำหนดคะแนนจุดตัดที่ระดับ 3/4

4.2.1.1.3 ผลการศึกษาขนาดขององค์ประกอบความแปรปรวนจากแหล่งความคลาดเคลื่อนต่าง ๆ ของการกำหนดคะแนนจุดตัดวิชาคณิตศาสตร์ด้วยวิธีเอบีซี

ตารางที่ 4.21 ผลการศึกษาขนาดขององค์ประกอบความแปรปรวนจากแหล่งความคลาดเคลื่อนต่าง ๆ ของการกำหนดคะแนนจุดตัดวิชาคณิตศาสตร์ด้วยวิธีเอบีซี

Source of Variation	df	Sum of square (SS)	Mean Square (MS)	Estimated Variance Component	Percent of Total Variability
			Level 1/2		
g	2	10.6667	5.3333	0.6543	5.60
p:g	6	22.6667	3.7778	0	0
r	2	84.2222	42.1111	4.4259	37.86
gr	4	9.1111	2.2778	0	0
pr:g,e	12	79.3333	6.6111	6.6111	56.55
Total	26	206.0000		11.6913	
			Level 2/3		
g	2	200.0741	100.03704	8.8086	12.44
p:g	6	399.1111	66.51852	6.5000	9.18
r	2	154.7407	77.37037	8.4568	11.95
gr	4	5.0370	1.25926	0	0
pr:g,e	12	564.2222	47.01852	47.0185	66.43
Total	26	1323.1852		70.7839	
			Level 2/3		
g	2	9.1852	4.59259	0	0
p:g	6	649.7778	108.2963	7.1667	6.80
r	2	289.8519	144.92593	11.4938	10.90
gr	4	165.9259	41.48148	0	0
pr:g,e	12	1041.5556	86.7963	86.7963	82.30
Total	26	2156.2963		105.4568	

จากตาราง พบว่า ขนาดขององค์ประกอบความแปรปรวนจากแหล่งความคลาดเคลื่อนต่าง ๆ ของการกำหนดคะแนนจุดตัดระดับ 1/2 เรียงจากมากไปหาน้อยได้แก่

องค์ประกอบความแปรปรวนส่วนที่เหลือ (*pr:g, e*) มีค่า 6.6111คิดเป็นร้อยละ 56.55 รองลงมาคือองค์ประกอบความแปรปรวนของรอบในการตัดสิน (*r*) มีค่า 4.4259 คิดเป็นร้อยละ 37.86 และองค์ประกอบความแปรปรวนของกลุ่มผู้เชี่ยวชาญ (*g*) มีค่า 0.6543 คิดเป็นร้อยละ 5.60 ส่วนองค์ประกอบความแปรปรวนของผู้เชี่ยวชาญในแต่ละกลุ่ม (*p:g*) และองค์ประกอบความแปรปรวนของปฏิสัมพันธ์ระหว่างกลุ่มผู้เชี่ยวชาญและรอบในการตัดสิน (*gr*) มีค่า 0 แสดงว่าความแตกต่างของคะแนนจุดตัดที่ระดับ 1/2 ส่วนใหญ่เนื่องมาจากองค์ประกอบความแปรปรวนส่วนที่เหลือ รองลงมาคือความแตกต่างของรอบในการตัดสิน ส่วนผู้เชี่ยวชาญในแต่ละกลุ่มกำหนดคะแนนจุดตัดไม่แตกต่างกัน และผู้เชี่ยวชาญแต่ละกลุ่มมีความคงเส้นคงวาในการกำหนดคะแนนจุดตัดในแต่ละรอบ

ขนาดขององค์ประกอบความแปรปรวนจากแหล่งความคลาดเคลื่อนต่าง ๆ ของการกำหนดคะแนนจุดตัดที่ระดับ 2/3 เรียงจากมากไปหาน้อย ได้แก่องค์ประกอบความแปรปรวนส่วนที่เหลือ (*pr:g, e*) มีค่า 47.0185 คิดเป็นร้อยละ 66.43 รองลงมาคือองค์ประกอบความแปรปรวนของกลุ่มผู้เชี่ยวชาญ (*g*) มีค่า 8.8086 คิดเป็นร้อยละ 12.44 องค์ประกอบความแปรปรวนของรอบในการตัดสิน (*r*) มีค่า 8.4568 คิดเป็นร้อยละ 11.95 และองค์ประกอบความแปรปรวนของผู้เชี่ยวชาญในแต่ละกลุ่ม (*p:g*) มีค่า 6.5000 คิดเป็นร้อยละ 9.18 ส่วนองค์ประกอบความแปรปรวนของปฏิสัมพันธ์ระหว่างกลุ่มผู้เชี่ยวชาญและรอบในการตัดสิน (*gr*) มีค่า 0 แสดงว่าความแตกต่างของคะแนนจุดตัดที่ระดับ 2/3 ส่วนใหญ่เนื่องมาจากองค์ประกอบความแปรปรวนส่วนที่เหลือ รองลงมาคือความแตกต่างกันของกลุ่มผู้เชี่ยวชาญ ส่วนผู้เชี่ยวชาญแต่ละกลุ่มมีความคงเส้นคงวาในการกำหนดคะแนนจุดตัดในแต่ละรอบ

ขนาดขององค์ประกอบความแปรปรวนจากแหล่งความคลาดเคลื่อนต่าง ๆ ของการกำหนดคะแนนจุดตัดที่ระดับ 3/4 เรียงจากมากไปหาน้อยได้แก่ องค์ประกอบความแปรปรวนส่วนที่เหลือ (*pr:g, e*) มีค่า 86.7963 คิดเป็นร้อยละ 82.30 รองลงมาคือองค์ประกอบความแปรปรวนของรอบในการตัดสิน (*r*) มีค่า 11.4938 คิดเป็นร้อยละ 10.90 และองค์ประกอบความแปรปรวนของผู้เชี่ยวชาญในแต่ละกลุ่ม (*p:g*) มีค่า 7.1667 คิดเป็นร้อยละ 6.80 ส่วนองค์ประกอบความแปรปรวนของกลุ่มผู้เชี่ยวชาญ (*g*) และองค์ประกอบความแปรปรวนของปฏิสัมพันธ์ระหว่างกลุ่มผู้เชี่ยวชาญและรอบในการตัดสิน (*gr*) มีค่า 0 แสดงว่าความแตกต่างของคะแนนจุดตัดระดับ 3/4 ส่วนใหญ่

เนื่องมาจากองค์ประกอบความแปรปรวนส่วนที่เหลือ รองลงมาคือความแตกต่างกันของผู้เชี่ยวชาญในแต่ละกลุ่ม รองลงมาคือความแตกต่างของรอบในการตัดสิน ส่วนผู้เชี่ยวชาญในแต่ละกลุ่มกำหนดคะแนนจุดตัดไม่แตกต่างกัน และผู้เชี่ยวชาญแต่ละกลุ่มมีความคงเส้นคงวาในการกำหนดคะแนนจุดตัดแต่ละรอบ

4.2.1.1.4 ผลการเปรียบเทียบขนาดขององค์ประกอบความแปรปรวนจากแหล่งความคลาดเคลื่อนต่าง ๆ ของการกำหนดคะแนนจุดตัดวิชาคณิตศาสตร์แต่ละวิธี

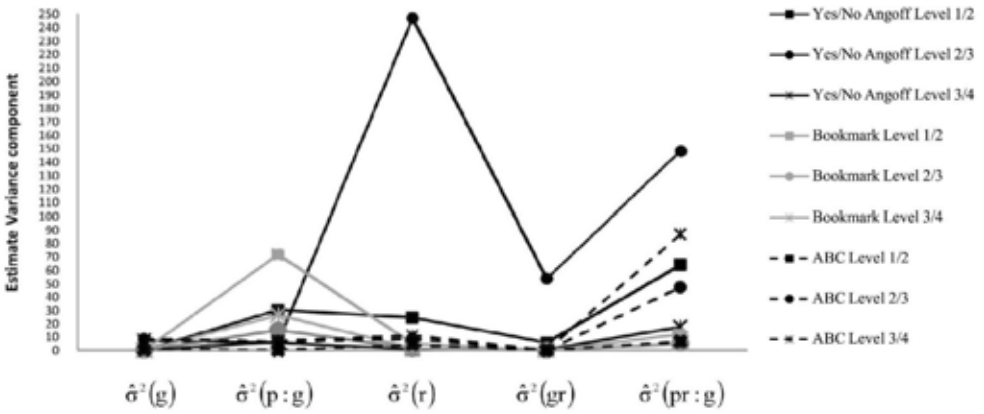
ตารางที่ 4.22 ผลการเปรียบเทียบขนาดขององค์ประกอบความแปรปรวนจากแหล่งความคลาดเคลื่อนต่าง ๆ ของการกำหนดคะแนนจุดตัดวิชาคณิตศาสตร์แต่ละวิธี

Source of Variation	วิธีแองกอฟแบบใช่/ไม่ใช่		วิธีบุ๊คมาร์ค		วิธีเอบีซี	
	Estimated Variance Component	Percent of Total Variability	Estimated Variance Component	Percent of Total Variability	Estimated Variance Component	Percent of Total Variability
			Level 1/2			
g	0	0	0	0	0.6543	5.60
p:g	30.5741	24.42	71.7407	80.56	0	0.00
r	24.7469	19.77	5.0123	5.63	4.4259	37.85
gr	6.0864	4.86	0	0	0	0
pr:g,e	63.7963	50.95	12.2963	13.81	6.6111	56.55
Total	125.2037	100.00	89.0493	100.00	11.6913	100.00
			Level 2/3			
g	0	0	0	0	8.8086	12.44
p:g	6.6296	1.45	15.3704	69.51	6.5	9.18
r	247.1296	54.22	1.4815	6.70	8.4568	11.95
gr	53.9074	11.83	0	0	0	0.00
pr:g,e	148.1481	32.50	5.2593	23.79	47.0185	66.43
Total	455.8147	100	22.1112	100.00	70.7839	100.00
			Level 3/4			
g	7.6420	23.38	0	0	0	0.00
p:g	5.8519	17.91	26.6667	77.34	7.1667	6.80
r	0.7160	2.19	1.7778	5.16	11.4938	10.90
gr	0.9506	2.91	0.5556	1.61	0	0.00
pr:g,e	17.5185	53.61	5.4815	15.90	86.7963	82.30
Total	32.6790	100.00	34.4816	100.00	105.4568	100.00

จากตาราง พบว่า แต่ละวิธีมีขนาดองค์ประกอบความแปรปรวนแตกต่างกัน ดังนี้องค์ประกอบความแปรปรวนของผู้เชี่ยวชาญในแต่ละกลุ่ม (*p:g*) ส่งผลต่อความแตกต่างกัน

ของคะแนนจุดตัดที่กำหนดโดยวิธีแองกอฟแบบใช่/ไม่ใช่ และบุ๊คมาร์คมากกว่าวิธีเอบีซี องค์ประกอบความแปรปรวนของรอบในการตัดสิน (*r*) ส่งผลต่อความแตกต่างกันของคะแนนจุดตัดที่กำหนดโดยวิธีแองกอฟแบบใช่/ไม่ใช่สูงกว่าวิธีบุ๊คมาร์คและวิธีเอบีซี นอกจากนี้ องค์ประกอบความแปรปรวนของปฏิสัมพันธ์ระหว่างกลุ่มผู้เชี่ยวชาญและรอบในการตัดสิน (*gr*) ส่งผลต่อความแตกต่างกันของคะแนนจุดตัดที่กำหนดโดยวิธีแองกอฟแบบใช่/ไม่ใช่สูงกว่าวิธีบุ๊คมาร์คและวิธีเอบีซี ส่วนองค์ประกอบความแปรปรวนส่วนที่เหลือ (*pr:g*, *e*) ส่งผลต่อความแตกต่างกันของคะแนนจุดตัดที่กำหนดโดยทุกวิธีค่อนข้างมาก

เพื่อแสดงให้เห็นความแตกต่างของขนาดองค์ประกอบความแปรปรวนของการกำหนดคะแนนจุดตัดวิชาคณิตศาสตร์แต่ละวิธี ผู้วิจัยขอนำเสนอผลเป็นกราฟเส้น ดังแสดงในภาพต่อไปนี้



ภาพที่ 4.11 ขนาดองค์ประกอบความแปรปรวนของการกำหนดคะแนนจุดตัดวิชาคณิตศาสตร์แต่ละวิธี

4.2.1.1.5 ผลวิเคราะห์ค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์ของการกำหนดคะแนนจุดตัดวิชาคณิตศาสตร์ ด้วยวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซี เมื่อจำนวนผู้เชี่ยวชาญ และจำนวนกลุ่มผู้เชี่ยวชาญต่างกัน

ตารางที่ 4.23 ผลวิเคราะห์ค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์ของการกำหนดคะแนนจุดตัดวิชาคณิตศาสตร์ด้วยวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซี เมื่อจำนวนผู้เชี่ยวชาญ และจำนวนกลุ่มผู้เชี่ยวชาญต่างกัน

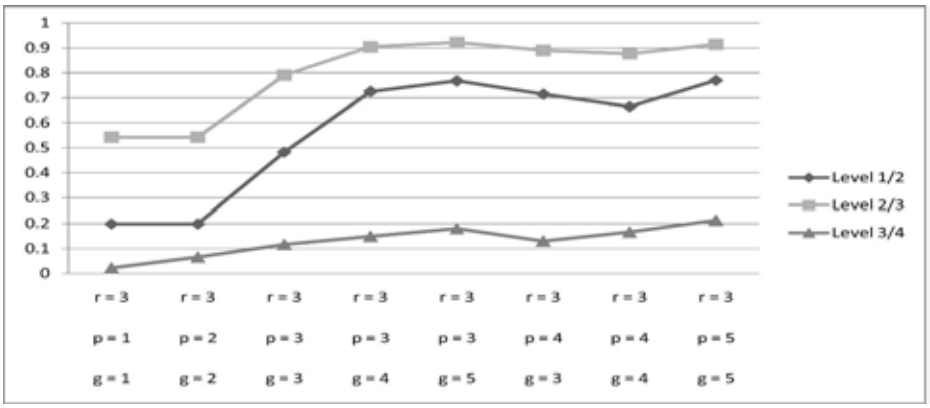
Dependability coefficient ($\hat{\phi}$)	**Size of decision study**														
	$g=1$ $p=1$ $r=3$	$g=2$ $p=2$ $r=3$	$g=3$ $p=3$ $r=3$	$g=4$ $p=3$ $r=3$	$g=5$ $p=3$ $r=3$	$g=3$ $p=4$ $r=3$	$g=4$ $p=4$ $r=3$	$g=4$ $p=5$ $r=3$	$g=4$ $p=6$ $r=3$	$g=4$ $p=7$ $r=3$	$g=5$ $p=4$ $r=3$	$g=5$ $p=5$ $r=3$	$g=5$ $p=6$ $r=3$	$g=6$ $p=6$ $r=3$	$g=7$ $p=7$ $r=3$
Yes/No Angoff method															
Level 1/2	0.198	0.198	0.482	0.725	0.767	0.714	0.664	0.799	0.819	0.835	0.832	0.850	0.863	0.872	0.899
Level 2/3	0.542	0.542	0.790	0.904	0.921	0.889	0.875	0.921	0.925	0.929	0.930	0.936	0.939	0.949	0.958
Level 3/4	0.022	0.066	0.116	0.149	0.179	0.130	0.166	0.178	0.187	0.194	0.199	0.213	0.223	0.256	0.296
เฉลี่ย	0.254	0.269	0.463	0.592	0.623	0.578	0.568	0.632	0.644	0.652	0.654	0.666	0.675	0.692	**0.717**
Bookmark method															
Level 1/2	0.048	0.169	0.314	0.379	0.432	0.379	0.448	0.504	0.549	0.587	0.504	0.559	0.604	0.646	0.713
Level 2/3	0.042	0.149	0.282	0.344	0.396	0.344	0.411	0.466	0.512	0.550	0.466	0.522	0.567	0.611	0.682
Level 3/4	0.052	0.176	0.321	0.387	0.441	0.383	0.453	0.504	0.546	0.580	0.508	0.560	0.601	0.643	0.707
เฉลี่ย	0.047	0.165	0.306	0.370	0.423	0.369	0.438	0.492	0.536	0.572	0.493	0.547	0.590	0.634	**0.701**
ABC method															
Level 1/2	0.374	0.705	0.843	0.878	0.900	0.878	0.905	0.923	0.935	0.944	0.923	0.937	0.947	**0.956**	0.967
Level 2/3	0.056	0.181	0.319	0.385	0.439	0.372	0.441	0.483	0.516	0.543	0.496	0.539	0.571	**0.615**	0.675
Level 3/4	0.064	0.216	0.382	0.452	0.508	0.452	0.524	0.579	0.623	0.658	0.579	0.632	0.673	**0.712**	0.771
เฉลี่ย	0.165	0.367	0.515	0.571	0.615	0.567	0.623	0.662	0.691	0.715	0.666	**0.703**	**0.731**	0.761	**0.804**

จากตาราง พบว่า ค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์ของการกำหนดคะแนนจุดตัดแต่ละวิธีมีความแตกต่าง โดยทุกวิธีจะมีค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์เพิ่มขึ้นเมื่อเพิ่มจำนวนผู้เชี่ยวชาญให้มากขึ้น เมื่อพิจารณาวิธีการเพิ่มจำนวนผู้เชี่ยวชาญสามารถทำได้ 2 วิธี คือ เพิ่มจำนวนผู้เชี่ยวชาญในแต่ละกลุ่มให้มากขึ้นและเพิ่มจำนวนกลุ่มผู้เชี่ยวชาญให้มากขึ้น เมื่อพิจารณาค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์ของการกำหนดคะแนนจุดตัดด้วยวิธีแองกอฟแบบใช่/ไม่ใช่ พบว่า เมื่อกำหนดจำนวนกลุ่มเป็น 3 กลุ่มและจำนวนผู้เชี่ยวชาญแต่ละกลุ่มเป็น 4 คน (รวมมีผู้เชี่ยวชาญจำนวน 12 คน) พบว่า ค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์ 0.71 และ 0.89 ตามลำดับ แต่เมื่อกำหนดจำนวนกลุ่มเป็น 4 กลุ่มและจำนวนผู้เชี่ยวชาญแต่ละกลุ่มเป็น 3 คน (รวมมีผู้เชี่ยวชาญจำนวน 12 คน) พบว่า ค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์ 0.73 และ 0.90 ตามลำดับ แสดงว่าการเพิ่มจำนวนกลุ่มให้มากขึ้นทำให้ค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์ของการกำหนดคะแนนจุดตัดด้วยวิธีแองกอฟแบบใช่/ไม่ใช่สูงกว่าการเพิ่มจำนวนผู้เชี่ยวชาญในแต่ละกลุ่มให้มากขึ้น ส่วนวิธีบุ๊คมาร์ค และวิธีเอบีซี พบว่า การเพิ่มจำนวนกลุ่มให้มากขึ้นหรือการเพิ่มจำนวนผู้เชี่ยวชาญในแต่ละกลุ่มให้มากขึ้นไม่ได้ทำให้ค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์แตกต่างกันนัก

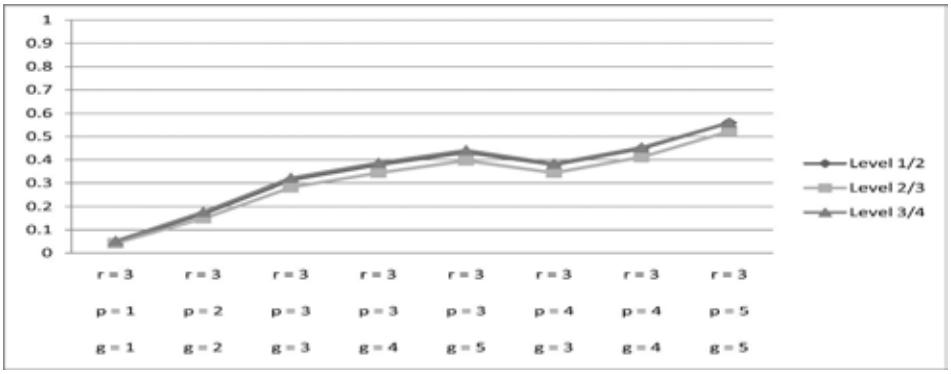
การกำหนดคะแนนจุดตัดให้มีความน่าเชื่อถือสำหรับนำไปใช้ในการตัดสินใจเชิงสัมบูรณ์เฉลี่ยตั้งแต่ 0.70 ขึ้นไปนั้น วิธีแองกอฟแบบใช่/ไม่ใช่และวิธีบุ๊คมาร์คควรกำหนดด้วยผู้เชี่ยวชาญจำนวนตั้งแต่ 49 คนขึ้นไป โดยแบ่งเป็น 7 กลุ่ม กลุ่มละ 7 คน สำหรับวิธีเอบีซีควรกำหนดด้วยผู้เชี่ยวชาญจำนวน 25 คนขึ้นไป โดยแบ่งเป็น 5 กลุ่ม กลุ่มละ 5 คน

เพื่อแสดงให้เห็นความแตกต่างของค่าค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์ของการกำหนดคะแนนจุดตัดวิชาคณิตศาสตร์แต่ละวิธี ผู้วิจัยขอนำเสนอผลการวิเคราะห์ในบางสถานการณ์เป็นกราฟเส้น ดังแสดงในภาพต่อไปนี้

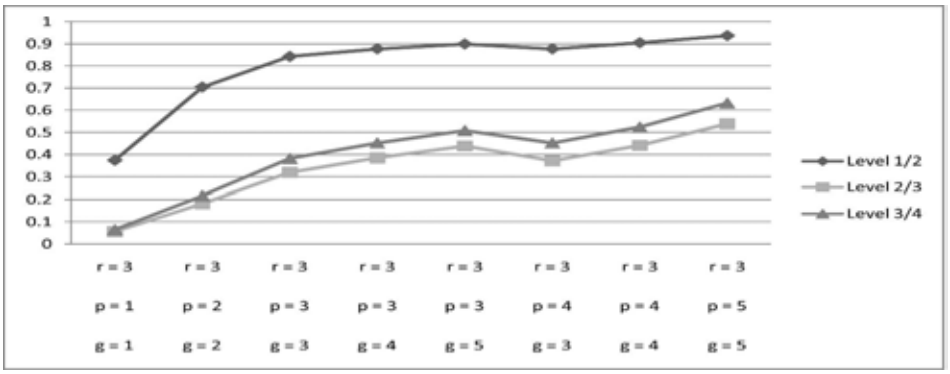
วิธีแองกอฟแบบใช่/ไม่ใช่



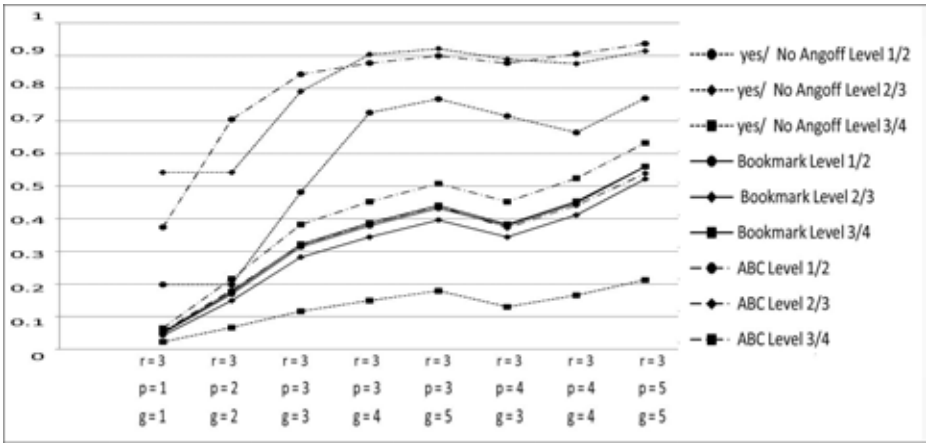
วิธีบุ๊คมาร์ค



วิธีเอบีซี



วิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซี



ภาพที่ 4.12 ค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์ของการกำหนดคะแนนจุดตัดวิชาคณิตศาสตร์แต่ละวิธี

4.2.1.1 ผลการศึกษาขนาดขององค์ประกอบความแปรปรวนจากแหล่งความคลาดเคลื่อนต่าง ๆ และค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์ของการกำหนดคะแนนจุดตัดวิชาวิทยาศาสตร์ด้วยวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซี มีรายละเอียดดังต่อไปนี้

4.2.1.1.1 ผลการศึกษาขนาดขององค์ประกอบความแปรปรวนจากแหล่งความคลาดเคลื่อนต่าง ๆ ของการกำหนดคะแนนจุดตัดวิชาวิทยาศาสตร์ด้วยวิธีแองกอฟแบบใช่/ไม่ใช่

ตารางที่ 4.24 ผลการศึกษาขนาดขององค์ประกอบความแปรปรวนจากแหล่งความคลาดเคลื่อนต่าง ๆ ของการกำหนดคะแนนจุดตัดวิชาวิทยาศาสตร์ด้วยวิธีแองกอฟแบบใช่/ไม่ใช่

Source of Variation	DF	Sum of square (SS)	Mean Square (MS)	Estimated Variance Component	Percent of Total Variability
			Level 1/2		
g	2	459.5000	229.7500	0	0
p:g	6	1488.5000	248.0833	73.8102	50.82
r	2	841.5556	420.7778	43.3056	29.82
gr	4	124.1111	31.0278	1.4583	1.00
pr:g,e	12	319.8333	26.6528	26.6528	18.35
Total	26	3233.5000	956.2917	145.2269	100.00
			Level 2/3		
g	2	1683.7222	841.8611	64.2047	36.08
p:g	6	1584.1111	264.0185	83.8241	47.11
r	2	337.7222	168.8611	17.3683	9.76
gr	4	48.5556	12.1389	0	0
pr:g,e	12	150.5556	12.5463	12.5463	7.05
Total	26	3804.6667	1299.4259	177.9434	100.00
			Level 3/4		
g	2	181.6296	90.8148	5.1883	17.00
p:g	6	159.2778	26.5463	5.1667	16.93
r	2	115.7963	57.8982	3.2531	10.66
gr	4	114.4815	28.6204	5.8580	19.20
pr:g,e	12	132.5556	11.0463	11.0463	36.20
Total	26	703.7408	214.9259	30.5123	100.00

จากตาราง พบว่า ขนาดขององค์ประกอบความแปรปรวนจากแหล่งความคลาดเคลื่อนต่าง ๆ ของการกำหนดคะแนนจุดตัดที่ระดับ 1/2 เรียงจากมากไปหาน้อย ได้แก่ องค์ประกอบความแปรปรวนของผู้เชี่ยวชาญในแต่ละกลุ่ม (*p:g*) มีค่า 73.8102 คิดเป็นร้อยละ 50.82

รองลงมาคือองค์ประกอบความแปรปรวนของรอบในการตัดสิน (*r*) มีค่า 43.3056 คิดเป็นร้อยละ 29.82 องค์ประกอบความแปรปรวนส่วนที่เหลือ (*pr:g, e*) มีค่า 26.6528 คิดเป็นร้อยละ 18.35 องค์ประกอบความแปรปรวนของปฏิสัมพันธ์ระหว่างกลุ่มผู้เชี่ยวชาญและรอบในการตัดสิน (*gr*) มีค่า 1.4583 คิดเป็นร้อยละ 1.00 และองค์ประกอบความแปรปรวนของกลุ่มผู้เชี่ยวชาญ(*g*) มีค่า 0 แสดงว่าความแตกต่างของคะแนนจุดตัดที่ระดับ 1/2 ส่วนใหญ่เนื่องมาจากความแตกต่างของผู้เชี่ยวชาญในแต่ละกลุ่ม รองลงมาเกิดจากความแตกต่างกันของรอบในการตัดสิน ส่วนส่วนกลุ่มผู้เชี่ยวชาญไม่มีความแตกต่างกันในการกำหนดคะแนนจุดตัดที่ระดับ 1/2

ขนาดขององค์ประกอบความแปรปรวนจากแหล่งความคลาดเคลื่อนต่าง ๆ ของการกำหนดคะแนนจุดตัดที่ระดับ 2/3 เรียงจากมากไปหาน้อย ได้แก่องค์ประกอบความแปรปรวนของผู้เชี่ยวชาญในแต่ละกลุ่ม (*p:g*) มีค่า 83.8241 คิดเป็นร้อยละ 47.11 รองลงมาคือองค์ประกอบความแปรปรวนของกลุ่มผู้เชี่ยวชาญ(*g*) มีค่า 64.2047 คิดเป็นร้อยละ 36.08 องค์ประกอบความแปรปรวนของรอบในการตัดสิน (*r*) มีค่า 17.3683คิดเป็นร้อยละ 9.76 องค์ประกอบความแปรปรวนส่วนที่เหลือ (*pr:g, e*) มีค่า 12.5463 คิดเป็นร้อยละ 7.05 และองค์ประกอบความแปรปรวนของปฏิสัมพันธ์ระหว่างกลุ่มผู้เชี่ยวชาญและรอบในการตัดสิน (*gr*) มีค่า 0 แสดงว่าความแตกต่างของคะแนนจุดตัดที่ระดับ 2/3 ส่วนใหญ่เนื่องมาจากความแตกต่างของผู้เชี่ยวชาญในแต่ละกลุ่ม รองลงมาคือความแตกต่างของกลุ่มผู้เชี่ยวชาญ ส่วนผู้เชี่ยวชาญแต่ละกลุ่มมีความคงเส้นคงวาในการกำหนดคะแนนจุดตัดในแต่ละรอบ

ขนาดขององค์ประกอบความแปรปรวนจากแหล่งความคลาดเคลื่อนต่าง ๆ ของการกำหนดคะแนนจุดตัดที่ระดับ 3/4 เรียงจากมากไปหาน้อย ได้แก่องค์ประกอบความแปรปรวนส่วนที่เหลือ (*pr:g, e*) มีค่า 11.0463 คิดเป็นร้อยละ 36.20 รองลงมาคือองค์ประกอบความแปรปรวนของปฏิสัมพันธ์ระหว่างกลุ่มผู้เชี่ยวชาญและรอบในการตัดสิน (*gr*) มีค่า 5.8580 คิดเป็นร้อยละ 19.20 องค์ประกอบความแปรปรวนของกลุ่มผู้เชี่ยวชาญ (*g*) มีค่า 5.1883 คิดเป็นร้อยละ 17.00 องค์ประกอบความแปรปรวนของผู้เชี่ยวชาญในแต่ละกลุ่ม (*p:g*) มีค่า 5.1667 คิดเป็นร้อยละ 16.93 องค์ประกอบความแปรปรวนของรอบในการตัดสิน (*r*) มีค่า 3.2531 คิดเป็นร้อยละ 10.66 แสดงว่าความแตกต่างของคะแนนจุดตัดที่ระดับ 3/4 ส่วนใหญ่เนื่องมาจากองค์ประกอบความแปรปรวน

ส่วนที่เหลือ รองลงมาคือความแตกต่างกันของกลุ่มผู้เชี่ยวชาญ ส่วนผู้เชี่ยวชาญแต่ละกลุ่มมีการกำหนดคะแนนจุดตัดในแต่ละรอบแตกต่างกันน้อย

4.2.1.1.2 ผลการศึกษาขนาดขององค์ประกอบความแปรปรวนจากแหล่งความคลาดเคลื่อนต่าง ๆ ของการกำหนดคะแนนจุดตัดด้วยวิธีบุ๊คมาร์ค

ตารางที่ 4.25 ผลการศึกษาขนาดขององค์ประกอบความแปรปรวนจากแหล่งความคลาดเคลื่อนต่าง ๆ ของการกำหนดคะแนนจุดตัดวิชาวิทยาศาสตร์ด้วยวิธีบุ๊คมาร์ค

Source of Variation	DF	Sum of square (SS)	Mean Square (MS)	Estimated Variance Component	Percent of Total Variability
			Level 1/2		
g	2	526.5185	263.2593	0	0
p:g	6	1662.2222	277.0370	86.8148	80.27
r	2	118.5185	59.2593	4.7407	4.38
gr	4	66.3704	16.5926	0.0000	0.00
pr:g,e	12	199.1111	16.5926	16.5926	15.34
Total	26	2572.7407	632.7407	108.1481	100.00
			Level 2/3		
g	2	56.5185	28.2593	0	0
p:g	6	536.2222	89.3704	28.1111	83.56
r	2	18.9630	9.4815	0.4938	1.47
gr	4	0.5926	0.1482	0.0000	0.00
pr:g,e	12	60.4444	5.0370	5.0370	14.97
Total	26	672.7407	132.2963	33.6420	100.00
			Level 2/3		
g	2	148.0741	74.0370	0.0000	0.00
p:g	6	1153.7778	192.2963	59.1481	79.56
r	2	35.8519	17.9259	0.3416	0.46
gr	4	53.9259	13.4815	0.0000	0.00
pr:g,e	12	178.2222	14.8519	14.8519	19.98
Total	26	1569.8519	312.5926	74.3416	100.00

จากตาราง พบว่า ขนาดขององค์ประกอบความแปรปรวนจากแหล่งความคลาดเคลื่อนต่าง ๆ ของการกำหนดคะแนนจุดตัดที่ระดับ 1/2 เรียงจากมากไปหาน้อย ได้แก่ องค์ประกอบความแปรปรวนของผู้เชี่ยวชาญในแต่ละกลุ่ม (*p:g*) มีค่า 86.8148 คิดเป็นร้อยละ 80.27 รองลงมาคือองค์ประกอบความแปรปรวนส่วนที่เหลือ (*pr:g, e*) มีค่า 16.5926 คิดเป็นร้อยละ 15.34 องค์ประกอบความแปรปรวนของรอบในการตัดสิน (*r*) มีค่า 4.7407 คิดเป็นร้อยละ 4.38 ส่วน

องค์ประกอบความแปรปรวนของกลุ่มผู้เชี่ยวชาญ (*g*) และองค์ประกอบความแปรปรวนของปฏิสัมพันธ์ระหว่างกลุ่มผู้เชี่ยวชาญและรอบในการตัดสิน (*gr*) มีค่า 0 แสดงว่าความแตกต่างของคะแนนจุดตัดที่ระดับ 1/2 ส่วนใหญ่เนื่องมาจากความแตกต่างกันของผู้เชี่ยวชาญในแต่ละกลุ่ม รองลงมาคือองค์ประกอบความแปรปรวนส่วนที่เหลือ ส่วนกลุ่มผู้เชี่ยวชาญ และผู้เชี่ยวชาญแต่ละกลุ่มไม่มีความแตกต่างกันในการกำหนดคะแนนจุดตัดที่ระดับ 1/2

ขนาดขององค์ประกอบความแปรปรวนจากแหล่งความคลาดเคลื่อนต่าง ๆ ของการกำหนดคะแนนจุดตัดที่ระดับ 2/3 เรียงจากมากไปหาน้อย ได้แก่องค์ประกอบความแปรปรวนของผู้เชี่ยวชาญในแต่ละกลุ่ม (*p:g*) มีค่า 28.1111 คิดเป็นร้อยละ 83.56 รองลงมาคือองค์ประกอบความแปรปรวนส่วนที่เหลือ (*pr:g, e*) มีค่า 5.0370 คิดเป็นร้อยละ 14.97 องค์ประกอบความแปรปรวนของรอบในการตัดสิน (*r*) มีค่า 0.4938 คิดเป็นร้อยละ 1.47 ส่วนองค์ประกอบความแปรปรวนของกลุ่มผู้เชี่ยวชาญ (*g*) และองค์ประกอบความแปรปรวนของปฏิสัมพันธ์ระหว่างกลุ่มผู้เชี่ยวชาญและรอบในการตัดสิน (*gr*) มีค่า 0 แสดงว่าความแตกต่างของคะแนนจุดตัดที่ระดับ 2/3 ส่วนใหญ่เนื่องมาจากความแตกต่างกันของผู้เชี่ยวชาญในแต่ละกลุ่ม รองลงมาคือองค์ประกอบความแปรปรวนส่วนที่เหลือ ส่วนกลุ่มผู้เชี่ยวชาญ และผู้เชี่ยวชาญในแต่ละกลุ่มไม่มีความแตกต่างกันในการกำหนดคะแนนจุดตัดที่ระดับ 2/3

ขนาดขององค์ประกอบความแปรปรวนจากแหล่งความคลาดเคลื่อนต่าง ๆ ของการกำหนดคะแนนจุดตัดระดับ 3/4 เรียงจากมากไปหาน้อย ได้แก่องค์ประกอบความแปรปรวนของผู้เชี่ยวชาญในแต่ละกลุ่ม (*p:g*) มีค่า 59.1481 คิดเป็นร้อยละ 79.56 รองลงมาคือองค์ประกอบความแปรปรวนส่วนที่เหลือ (p*r:g, e*) มีค่า 14.8519 คิดเป็นร้อยละ 19.98 องค์ประกอบความแปรปรวนของรอบในการตัดสิน (*r*) มีค่า 0.3416 คิดเป็นร้อยละ 0.46 ส่วนองค์ประกอบความแปรปรวนของกลุ่มผู้เชี่ยวชาญ (*g*) และองค์ประกอบความแปรปรวนของปฏิสัมพันธ์ระหว่างกลุ่มผู้เชี่ยวชาญและรอบในการตัดสิน (*gr*) มีค่า 0 แสดงว่าความแตกต่างของคะแนนจุดตัดที่ระดับ 3/4 ส่วนใหญ่เนื่องมาจากความแตกต่างกันของผู้เชี่ยวชาญในแต่ละกลุ่ม รองลงมาคือองค์ประกอบความแปรปรวนส่วนที่เหลือ ส่วนกลุ่มผู้เชี่ยวชาญ และผู้เชี่ยวชาญในแต่ละกลุ่มไม่มีความแตกต่างกันในการกำหนดคะแนนจุดตัดที่ระดับ 3/4

4.2.1.1.3 ผลการศึกษาขนาดขององค์ประกอบความแปรปรวนจากแหล่งความคลาดเคลื่อนต่าง ๆ ของการกำหนดคะแนนจุดตัดวิชาวิทยาศาสตร์ด้วยด้วยวิธีเอบีซี

ตารางที่ 4.26 ผลการศึกษาขนาดขององค์ประกอบความแปรปรวนจากแหล่งความคลาดเคลื่อน ต่าง ๆ ของการกำหนดคะแนนจุดตัดวิชาวิทยาศาสตร์ด้วยวิธีเอบีซี

Source of Variation	DF	Sum of square (SS)	Mean Square (MS)	Estimated Variance Component	Percent of Total Variability
			Level 1/2		
g	2	104.5417	52.2708	0	0
p:g	9	668.4792	74.2755	13.6042	26.68
r	2	161.2917	80.6458	3.9319	7.71
gr	4	56.5417	14.1354	0.0000	0.00
pr:g,e	18	602.3333	33.4630	33.4630	65.61
Total	35	1593.1875	254.7905	50.9990	100.00
			Level 2/3		
g	2	781.9306	390.9653	13.9510	12.34
p:g	9	2011.9792	223.5532	65.7384	58.14
r	2	221.7222	110.8611	7.0436	6.23
gr	4	92.6944	23.1736	0.0000	0.00
pr:g,e	18	474.0833	26.3380	26.3380	23.29
Total	35	3582.4097	774.8912	113.0710	100.00
			Level 2/3		
g	2	207.7917	103.8958	0	0
p:g	9	988.3750	109.8194	29.1898	48.50
r	2	251.2917	125.6458	8.5530	14.21
gr	4	92.0417	23.0104	0.1901	0.32
pr:g,e	18	400.5000	22.2500	22.2500	36.97
Total	35	1940.0000	384.6215	60.1829	100.00

จากตาราง พบว่า ขนาดขององค์ประกอบความแปรปรวนจากแหล่งความคลาดเคลื่อนต่าง ๆ ของการกำหนดคะแนนจุดตัดระดับ 1/2 เรียงจากมากไปหาน้อยได้แก่ องค์ประกอบความแปรปรวนส่วนที่เหลือ (*pr:g, e*) มีค่า 33.4630 คิดเป็นร้อยละ 65.61 รองลงมาคือ องค์ประกอบความแปรปรวนของผู้เชี่ยวชาญในแต่ละกลุ่ม (*p:g*) มีค่า 13.6042 คิดเป็นร้อยละ 26.68 องค์ประกอบความแปรปรวนของรอบในการตัดสิน (*r*) มีค่า 3.9319 คิดเป็นร้อยละ 7.71 ส่วนองค์ประกอบความแปรปรวนของกลุ่มผู้เชี่ยวชาญ (*g*) และองค์ประกอบความแปรปรวนของปฏิสัมพันธ์ระหว่างกลุ่มผู้เชี่ยวชาญและรอบในการตัดสิน (*gr*) มีค่า 0 แสดงว่าความแตกต่างของ

คะแนนจุดตัดที่ระดับ 1/2 ส่วนใหญ่เนื่องมาจากองค์ประกอบความแปรปรวนส่วนที่เหลือ รองลงมาคือความแตกต่างกันของผู้เชี่ยวชาญในแต่ละกลุ่ม ส่วนผู้เชี่ยวชาญในแต่ละกลุ่มกำหนดคะแนนจุดตัดไม่แตกต่างกัน และผู้เชี่ยวชาญแต่ละกลุ่มมีความคงเส้นคงวาในการกำหนดคะแนนจุดตัดในแต่ละรอบ

ขนาดขององค์ประกอบความแปรปรวนจากแหล่งความคลาดเคลื่อนต่าง ๆ ของการกำหนดคะแนนจุดตัดที่ระดับ 2/3 เรียงจากมากไปหาน้อย ได้แก่องค์ประกอบความแปรปรวนของผู้เชี่ยวชาญในแต่ละกลุ่ม (*p:g*) มีค่า 65.7384 คิดเป็นร้อยละ 58.14 รองลงมาคือองค์ประกอบความแปรปรวนส่วนที่เหลือ (*pr:g, e*) มีค่า 26.3380 คิดเป็นร้อยละ 23.29 องค์ประกอบความแปรปรวนของกลุ่มผู้เชี่ยวชาญ (*g*) มีค่า 13.9510 คิดเป็นร้อยละ 12.34 องค์ประกอบความแปรปรวนของรอบในการตัดสิน (*r*) มีค่า 7.0436 คิดเป็นร้อยละ 6.23 ส่วนองค์ประกอบความแปรปรวนของปฏิสัมพันธ์ระหว่างกลุ่มผู้เชี่ยวชาญและรอบในการตัดสิน (*gr*) มีค่า 0 แสดงว่าความแตกต่างของคะแนนจุดตัดที่ระดับ 2/3 ส่วนใหญ่เนื่องมาจากความแตกต่างกันของผู้เชี่ยวชาญในแต่ละกลุ่ม รองลงมาคือองค์ประกอบความแปรปรวนส่วนที่เหลือ ส่วนส่วนผู้เชี่ยวชาญแต่ละกลุ่มมีความคงเส้นคงวาในการกำหนดคะแนนจุดตัดในแต่ละรอบ

ขนาดขององค์ประกอบความแปรปรวนจากแหล่งความคลาดเคลื่อนต่าง ๆ ของการกำหนดคะแนนจุดตัดที่ระดับ 3/4 เรียงจากมากไปหาน้อยได้แก่องค์ประกอบความแปรปรวนของผู้เชี่ยวชาญในแต่ละกลุ่ม (*p:g*) มีค่า 29.1898 คิดเป็นร้อยละ 48.50 รองลงมาคือองค์ประกอบความแปรปรวนส่วนที่เหลือ (*pr:g, e*) มีค่า 22.2500 คิดเป็นร้อยละ 36.97 องค์ประกอบความแปรปรวนของรอบในการตัดสิน (*r*) มีค่า 8.5530 คิดเป็นร้อยละ 14.21 องค์ประกอบความแปรปรวนของปฏิสัมพันธ์ระหว่างกลุ่มผู้เชี่ยวชาญและรอบในการตัดสิน (*gr*) มีค่า 0.1901 คิดเป็นร้อยละ 48.50 ส่วนองค์ประกอบความแปรปรวนของกลุ่มผู้เชี่ยวชาญ(*g*) มีค่า 0 แสดงว่าความแตกต่างของคะแนนจุดตัดระดับ 3/4 ส่วนใหญ่เนื่องมาจากความแตกต่างกันของผู้เชี่ยวชาญในแต่ละกลุ่ม รองลงมาคือองค์ประกอบความแปรปรวนส่วนที่เหลือ ส่วนผู้เชี่ยวชาญในแต่ละกลุ่มกำหนดคะแนนจุดตัดไม่แตกต่างกัน

4.2.1.1.4 ผลการเปรียบเทียบขนาดขององค์ประกอบความแปรปรวนจากแหล่งความคลาดเคลื่อนต่าง ๆ ของการกำหนดคะแนนจุดตัดวิชาวิทยาศาสตร์แต่ละวิธี

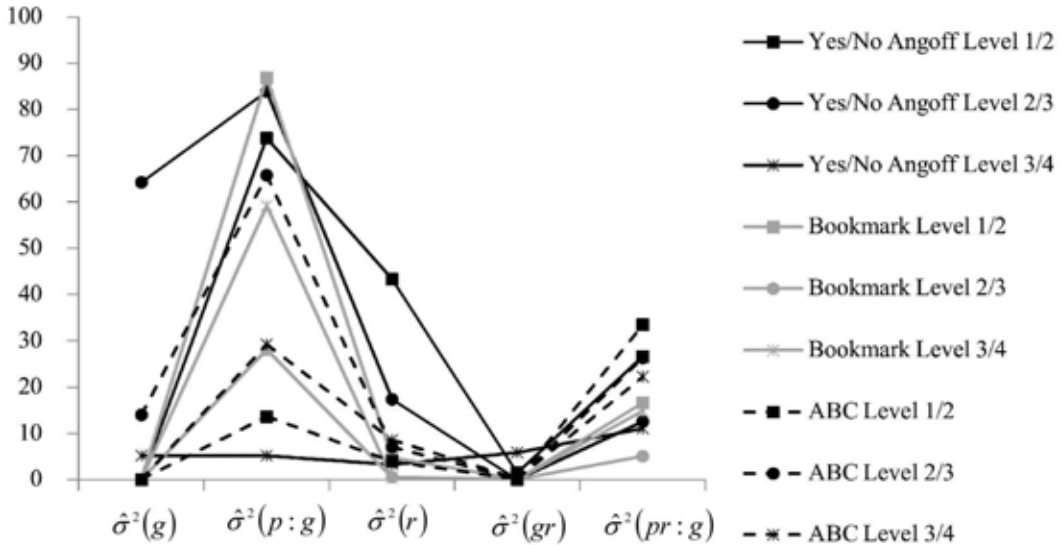
ตารางที่ 4.27 ผลการเปรียบเทียบขนาดขององค์ประกอบความแปรปรวนจากแหล่งความคลาดเคลื่อนต่าง ๆ ของการกำหนดคะแนนจุดตัดวิชาวิทยาศาสตร์แต่ละวิธี

Source of Variation	วิธีแองกอฟแบบใช่/ไม่ใช่		วิธีบุ๊คมาร์ค		วิธีเอบีซี	
	Estimated Variance Component	Percent of Total Variability	Estimated Variance Component	Percent of Total Variability	Estimated Variance Component	Percent of Total Variability
			Level 1/2			
g	0	0	0	0	0	0
p:g	73.8102	50.82	86.8148	80.27	13.6042	26.68
r	43.3056	29.82	4.7407	4.38	3.9319	7.71
gr	1.4583	1.00	0.0000	0.00	0.0000	0.00
pr:g,e	26.6528	18.35	16.5926	15.34	33.4630	65.61
Total	145.2269	100.00	108.1481	100.00	50.9990	100.00
			Level 2/3			
g	64.2047	36.08	0	0	13.9510	12.34
p:g	83.8241	47.11	28.1111	83.56	65.7384	58.14
r	17.3683	9.76	0.4938	1.47	7.0436	6.23
gr	0	0	0.0000	0.00	0.0000	0.00
pr:g,e	12.5463	7.05	5.0370	14.97	26.3380	23.29
Total	177.9434	100.00	33.6420	100.00	113.0710	100.00
			Level 3/4			
g	5.1883	17.00	0	0	0	0
p:g	5.1667	16.93	59.1481	79.56	29.1898	48.50
r	3.2531	10.66	0.3416	0.46	8.5530	14.21
gr	5.8580	19.20	0	0	0.1901	0.32
pr:g,e	11.0463	36.20	14.8519	19.98	22.2500	36.97
Total	30.5123	100.00	74.3416	100.00	60.1829	100.00

จากตาราง พบว่า แต่ละวิธีมีขนาดองค์ประกอบความแปรปรวนแตกต่างกัน ดังนี้องค์ประกอบความแปรปรวนของผู้เชี่ยวชาญในแต่ละกลุ่ม (*p:g*) ส่งผลต่อความแตกต่างกันของคะแนนจุดตัดที่กำหนดโดยวิธีแองกอฟแบบใช่/ไม่ใช่ และบุ๊คมาร์คและวิธีเอบีซีมาก องค์ประกอบความแปรปรวนของรอบในการตัดสิน (*r*) ส่งผลต่อความแตกต่างกันของคะแนนจุดตัดที่กำหนดโดยวิธีแองกอฟแบบใช่/ไม่ใช่สูงกว่าวิธีบุ๊คมาร์คและวิธีเอบีซี นอกจากนี้

องค์ประกอบความแปรปรวนของปฏิสัมพันธ์ระหว่างกลุ่มผู้เชี่ยวชาญและรอบในการตัดสิน (*gr*) ส่งผลต่อความแตกต่างกันของคะแนนจุดตัดที่กำหนดโดยวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซีค่อนข้างน้อย ส่วนองค์ประกอบความแปรปรวนส่วนที่เหลือ (*pr:g*, *e*) ส่งผลต่อความแตกต่างกันของคะแนนจุดตัดที่กำหนดโดยทุกวิธีค่อนข้างมาก

เพื่อแสดงให้เห็นความแตกต่างของขนาดองค์ประกอบความแปรปรวนของการกำหนดคะแนนจุดตัดวิชาวิทยาศาสตร์แต่ละวิธี ผู้วิจัยขอนำเสนอผลเป็นกราฟเส้น ดังแสดงในภาพต่อไปนี้



ภาพที่ 4.13 ขนาดองค์ประกอบความแปรปรวนของการกำหนดคะแนนจุดตัดวิชาวิทยาศาสตร์แต่ละวิธี

4.2.1.1.5 ผลวิเคราะห์ค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์ของการกำหนดคะแนนจุดตัดวิชาวิทยาศาสตร์ด้วยวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซี เมื่อจำนวนผู้เชี่ยวชาญ และจำนวนกลุ่มผู้เชี่ยวชาญต่างกัน

ตารางที่ 4.28 ผลวิเคราะห์ค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์ของการกำหนดคะแนนจุดตัดวิชาวิทยาศาสตร์ด้วยวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซี เมื่อจำนวนผู้เชี่ยวชาญ และจำนวนกลุ่มผู้เชี่ยวชาญต่างกัน

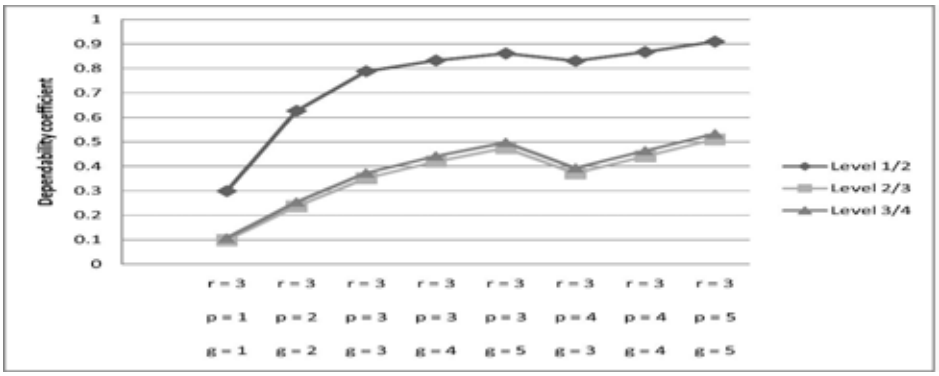
Dependability coefficient ($\hat{\phi}$)	**Size of decision study**														
	$g=1$ $p=1$ $r=3$	$g=2$ $p=2$ $r=3$	$g=3$ $p=3$ $r=3$	$g=4$ $p=3$ $r=3$	$g=5$ $p=3$ $r=3$	$g=3$ $p=4$ $r=3$	$g=4$ $p=4$ $r=3$	$g=4$ $p=5$ $r=3$	$g=4$ $p=6$ $r=3$	$g=4$ $p=7$ $r=3$	$g=5$ $p=4$ $r=3$	$g=5$ $p=5$ $r=3$	$g=5$ $p=6$ $r=3$	$g=6$ $p=6$ $r=3$	$g=7$ $p=7$ $r=3$
Yes/No Angoff method															
Level 1/2	0.30	0.63	0.79	0.83	0.86	0.83	0.87	0.89	0.91	0.92	0.89	0.91	0.92	0.94	0.95
Level 2/3	0.10	0.24	0.35	0.42	0.47	0.37	0.44	0.45	0.46	0.47	0.50	0.51	0.52	0.57	0.61
Level 3/4	0.11	0.25	0.37	0.44	0.50	0.39	0.46	0.48	0.49	0.49	0.52	0.53	0.54	0.59	0.63
เฉลี่ย	0.17	0.37	0.50	0.56	0.61	0.53	0.59	0.61	0.62	0.63	0.64	0.65	0.66	**0.70**	0.73
Bookmark method															
Level 1/2	0.04	0.16	0.29	0.36	0.41	0.36	0.42	0.48	0.52	0.56	0.48	0.53	0.58	0.62	0.69
Level 2/3	0.02	0.06	0.12	0.15	0.18	0.15	0.19	0.23	0.26	0.29	0.23	0.27	0.31	0.35	0.42
Level 3/4	0.01	0.02	0.04	0.05	0.07	0.04	0.07	0.09	0.10	0.11	0.10	0.12	0.14	0.12	0.14
เฉลี่ย	0.02	0.08	0.15	0.19	0.22	0.18	0.23	0.26	0.30	0.32	0.27	0.31	0.34	0.36	0.42
ABC method															
Level 1/2	0.08	0.25	0.43	0.50	0.56	0.50	0.57	0.63	0.67	0.70	0.63	0.68	0.72	0.75	0.80
Level 2/3	0.06	0.19	0.32	0.39	0.44	0.36	0.43	0.47	0.49	0.51	0.49	0.52	0.55	0.59	0.65
Level 3/4	0.14	0.40	0.60	0.66	0.71	0.66	0.72	0.77	0.80	0.82	0.77	0.80	0.83	0.85	0.89
เฉลี่ย	0.09	0.28	0.45	0.52	0.57	0.51	0.58	0.63	0.65	0.68	0.63	0.67	**0.70**	0.73	0.78

จากตาราง พบว่า ค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์ของการกำหนดคะแนนจุดตัดแต่ละวิธีมีความแตกต่าง โดยทุกวิธีจะมีค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์เพิ่มขึ้นเมื่อเพิ่มจำนวนผู้เชี่ยวชาญมากขึ้น โดยการเพิ่มจำนวนกลุ่มให้มากขึ้นจะทำให้ค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์ของการกำหนดคะแนนจุดตัดด้วยวิธีแองกอฟแบบใช่/ไม่ใช่สูงกว่าการเพิ่มจำนวนผู้เชี่ยวชาญในแต่ละกลุ่มให้มากขึ้น ส่วนวิธีบุ๊คมาร์ค และวิธีเอบีซี พบว่า การเพิ่มจำนวนกลุ่มให้มากขึ้นหรือการเพิ่มจำนวนผู้เชี่ยวชาญในแต่ละกลุ่มให้มากขึ้นไม่ได้ทำให้ค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์แตกต่างกันนัก

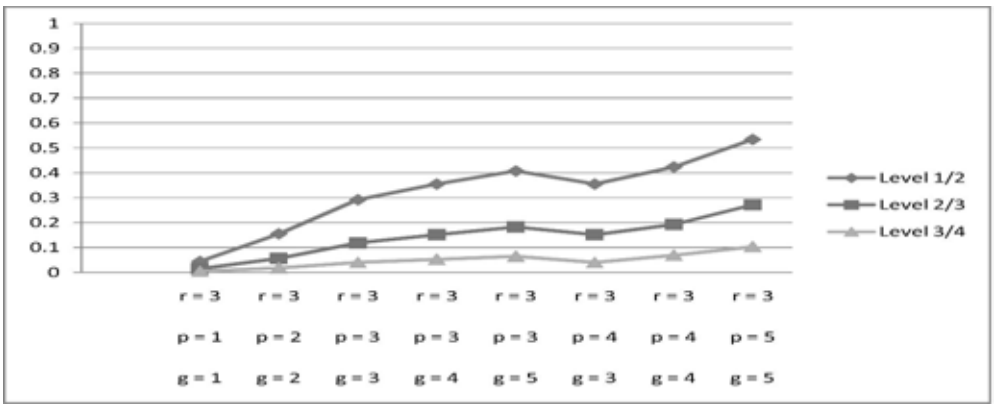
การกำหนดคะแนนจุดตัดให้มีความน่าเชื่อถือสำหรับนำไปใช้ในการตัดสินใจเชิงสัมบูรณ์เฉลี่ยตั้งแต่ 0.70 ขึ้นไปนั้น วิธีแองกอฟแบบใช่/ไม่ใช่ควรกำหนดด้วยผู้เชี่ยวชาญจำนวนตั้งแต่ 36 คนขึ้นไป โดยแบ่งเป็น 6 กลุ่ม กลุ่มละ 6 คน สำหรับวิธีเอบีซีควรกำหนดด้วยผู้เชี่ยวชาญจำนวน 30 คนขึ้นไป โดยแบ่งเป็น 5 กลุ่ม กลุ่มละ 6 คน ส่วนวิธีบุ๊คมาร์คนั้นจะต้องกำหนดคะแนนจุดตัดด้วยผู้เชี่ยวชาญจำนวนมากกว่าวิธีแองกอฟแบบใช่/ไม่ใช่ และวิธีเอบีซีจึงจะทำให้ได้คะแนนจุดตัดที่มีความน่าเชื่อถือสำหรับนำไปใช้ในการตัดสินใจเชิงสัมบูรณ์สูง

เพื่อแสดงให้เห็นความแตกต่างของค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์ของการกำหนดคะแนนจุดตัดวิชาวิทยาศาสตร์แต่ละวิธี ผู้วิจัยขอนำเสนอผลการวิเคราะห์ในบางสถานการณ์เป็นกราฟเส้น ดังแสดงในภาพต่อไปนี้

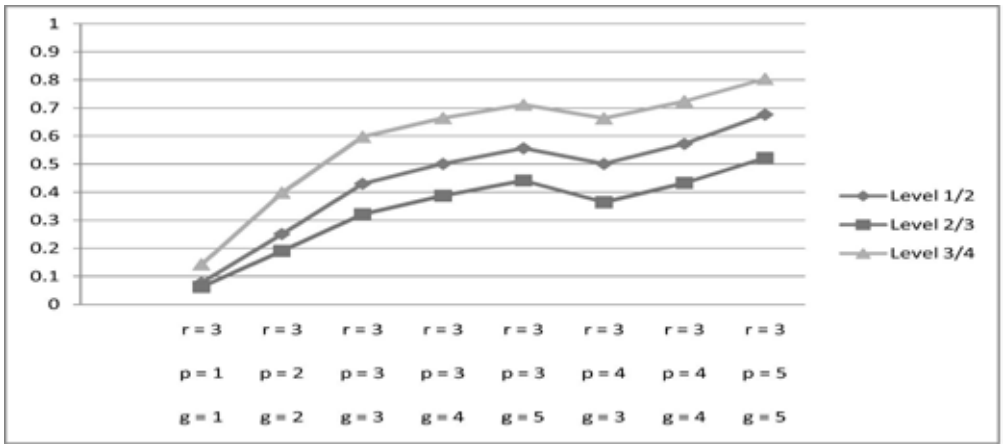
วิธีแองกอฟแบบใช่/ไม่ใช่



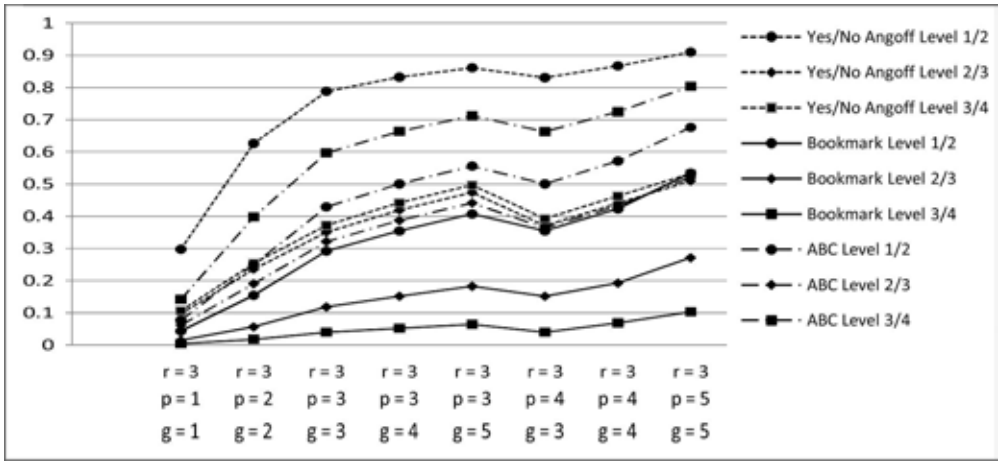
วิธีบุ๊คมาร์ค



วิธีเอบีซี



วิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซี



ภาพที่ 4.14 ค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์ของการกำหนดคะแนนจุดตัดวิชาวิทยาศาสตร์แต่ละวิธี

4.2.2 ผลการวิเคราะห์ความเที่ยงภายในผู้เชี่ยวชาญ (Intra-rater reliability) และความเที่ยงระหว่างผู้เชี่ยวชาญ (Inter-rater reliability)

4.2.2.1 ผลการวิเคราะห์ความเที่ยงภายในผู้เชี่ยวชาญ และความเที่ยงระหว่างผู้เชี่ยวชาญ วิชาคณิตศาสตร์ ดังแสดงในตารางต่อไปนี้

ตารางที่ 4.29 ผลการวิเคราะห์ความเที่ยงภายในผู้เชี่ยวชาญ และความเที่ยงระหว่างผู้เชี่ยวชาญ ของการกำหนดคะแนนจุดตัดวิชาคณิตศาสตร์ และวิชาวิทยาศาสตร์

วิธีกำหนดคะแนนจุดตัด		**Intra-rater reliability**		**Inter-rater reliability**	
วิธีแองกอฟแบบใช่/ไม่ใช่		คณิตศาสตร์	วิทยาศาสตร์	คณิตศาสตร์	วิทยาศาสตร์
	ระดับ 1/2	0.5157	0.8861	0.7757	0.9341
	ระดับ 2/3	0.3579	0.9695	0.9256	0.9263
	ระดับ 3/4	0.6559	0.6377	0.3201	0.7333
	เฉลี่ย	0.5098	0.8311	0.6738	0.8646
วิธีบุ๊คมาร์ค					
	ระดับ 1/2	0.9449	0.9394	0.7865	0.7200
	ระดับ 2/3	0.9114	0.9485	0.6875	0.5977
	ระดับ 3/4	0.9270	0.9108	0.7401	0.1906
	เฉลี่ย	0.9278	0.9329	0.7380	0.5028
วิธีเอบีซี					
	ระดับ 1/2	0.3267	0.5738	0.8687	0.6286
	ระดับ 2/3	0.5250	0.8986	0.5402	0.7676
	ระดับ 3/4	0.8380	0.7941	0.4793	0.8218
	เฉลี่ย	0.5632	0.7555	0.6294	0.7393

จากตาราง พบว่า

1. การกำหนดคะแนนจุดตัดวิชาคณิตศาสตร์มีความเที่ยงภายในผู้เชี่ยวชาญและความเที่ยงระหว่างผู้เชี่ยวชาญของการกำหนดคะแนนจุดตัดแต่ละวิธีแตกต่างกัน ดังมีรายละเอียดต่อไปนี้

1.1 ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดด้วยวิธีแองกอฟแบบใช่/ไม่ใช่ กำหนดคะแนนจุดตัดในแต่ละรอบได้สอดคล้องกันในระดับปานกลาง (ค่าความเที่ยงภายใน

ผู้เชี่ยวชาญเฉลี่ย เท่ากับ 0.5098) ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดด้วยวิธีบุ๊คมาร์ค กำหนดคะแนนจุดตัดในแต่ละรอบได้สอดคล้องกันอย่างยิ่ง (ค่าความเที่ยงภายในผู้เชี่ยวชาญเฉลี่ย เท่ากับ 0.9278) และผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดด้วยวิธีเอบีซี กำหนดคะแนนจุดตัดในแต่ละรอบได้สอดคล้องกันในระดับปานกลาง (ค่าความเที่ยงภายในผู้เชี่ยวชาญเฉลี่ย เท่ากับ 0.5632)

1.2 ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดด้วยวิธีแองกอฟแบบใช่/ไม่ใช่ กำหนดคะแนนจุดตัดได้สอดคล้องกันในระดับมาก (ค่าความเที่ยงระหว่างผู้เชี่ยวชาญเฉลี่ย เท่ากับ 0.6738) ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดด้วยวิธีบุ๊คมาร์ค กำหนดคะแนนจุดตัดได้สอดคล้องกันในระดับมาก (ค่าความเที่ยงระหว่างผู้เชี่ยวชาญเฉลี่ย เท่ากับ 0.7380) และผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดด้วยวิธีเอบีซี กำหนดคะแนนจุดตัดได้สอดคล้องกันในระดับมาก (ค่าความเที่ยงระหว่างผู้เชี่ยวชาญเฉลี่ย เท่ากับ 0.6294)

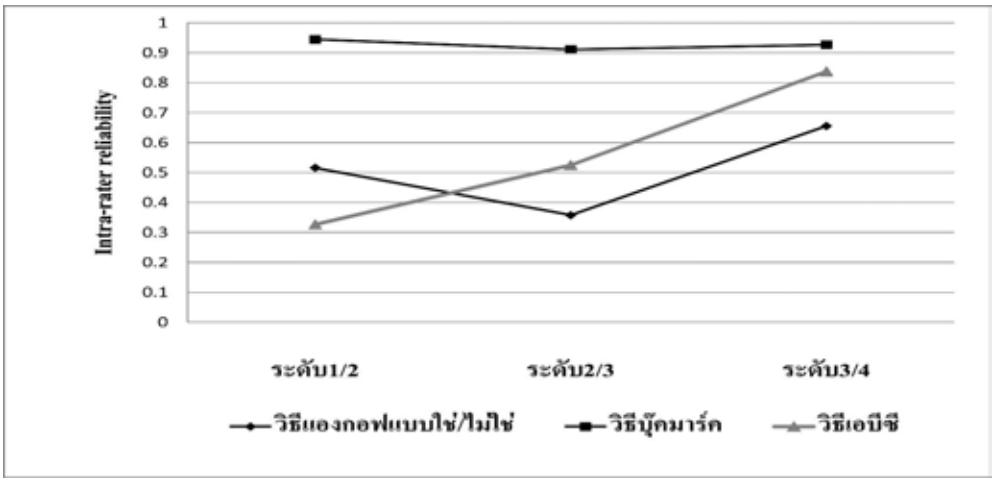
2. การกำหนดคะแนนจุดตัดวิชาวิทยาศาสตร์มีความเที่ยงภายในผู้เชี่ยวชาญและความเที่ยงระหว่างผู้เชี่ยวชาญของการกำหนดคะแนนจุดตัดแต่ละวิธีแตกต่างกัน ดังมีรายละเอียดต่อไปนี้

2.1 ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดด้วยวิธีแองกอฟแบบใช่/ไม่ใช่ กำหนดคะแนนจุดตัดในแต่ละรอบได้สอดคล้องกันอย่างยิ่ง (ค่าความเที่ยงภายในผู้เชี่ยวชาญเฉลี่ย เท่ากับ 0.8311) ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดด้วยวิธีบุ๊คมาร์ค กำหนดคะแนนจุดตัดในแต่ละรอบได้สอดคล้องกันอย่างยิ่ง (ค่าความเที่ยงภายในผู้เชี่ยวชาญเฉลี่ย เท่ากับ 0.9329) และผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดด้วยวิธีเอบีซี กำหนดคะแนนจุดตัดในแต่ละรอบได้สอดคล้องกันมาก (ค่าความเที่ยงภายในผู้เชี่ยวชาญเฉลี่ย เท่ากับ 0.7555)

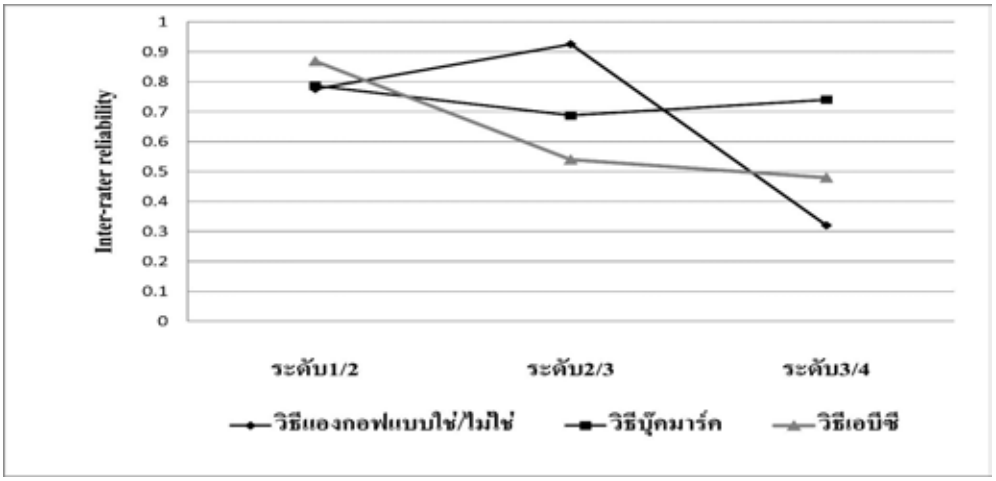
2.2 ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดด้วยวิธีแองกอฟแบบใช่/ไม่ใช่ กำหนดคะแนนจุดตัดได้สอดคล้องกันอย่างยิ่ง (ค่าความเที่ยงระหว่างผู้เชี่ยวชาญเฉลี่ย เท่ากับ 0.8646) ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดด้วยวิธีบุ๊คมาร์ค กำหนดคะแนนจุดตัดได้สอดคล้องกันในระดับปานกลาง (ค่าความเที่ยงระหว่างผู้เชี่ยวชาญเฉลี่ย เท่ากับ 0.5028) และผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดด้วยวิธีเอบีซี กำหนดคะแนนจุดตัดได้สอดคล้องกันมาก (ค่าความเที่ยงระหว่างผู้เชี่ยวชาญเฉลี่ย เท่ากับ 0.7393)

เพื่อแสดงให้เห็นความแตกต่างของความเที่ยงภายในผู้เชี่ยวชาญ และความเที่ยงระหว่างผู้เชี่ยวชาญของคะแนนจุดตัดที่กำหนดขึ้นด้วยวิธีการที่ต่างกัน ผู้วิจัยขอนำเสนอผลเป็นกราฟเส้น ดังแสดงในภาพต่อไปนี้

ความเที่ยงภายในผู้เชี่ยวชาญ

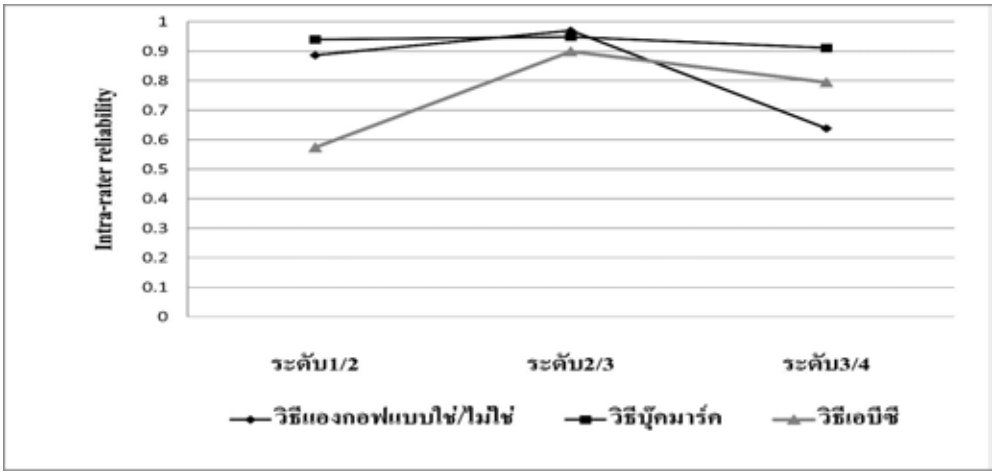


ความเที่ยงระหว่างผู้เชี่ยวชาญ

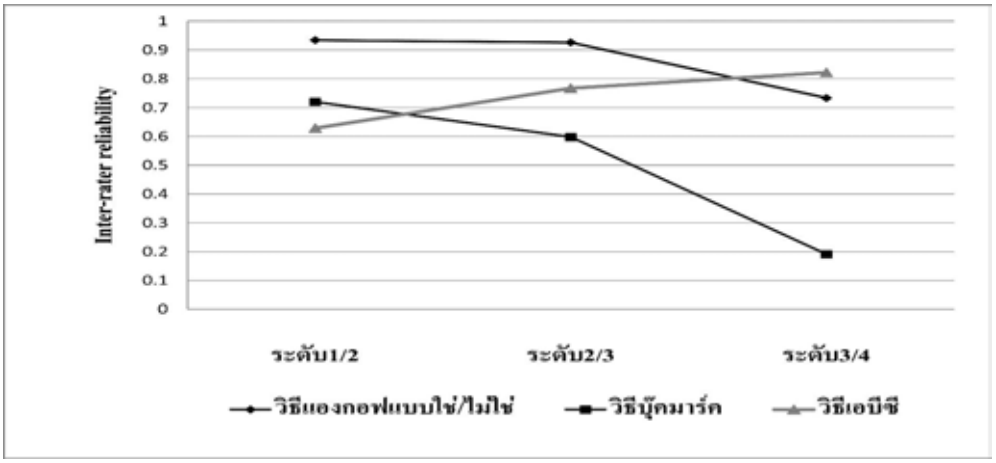


ภาพที่ 4.15 ความเที่ยงภายในผู้เชี่ยวชาญ และความเที่ยงระหว่างผู้เชี่ยวชาญของคะแนนจุดตัดวิชาคณิตศาสตร์

ความเที่ยงภายในผู้เชี่ยวชาญ



ความเที่ยงระหว่างผู้เชี่ยวชาญ



ภาพที่ 4.16 ความเที่ยงภายในผู้เชี่ยวชาญ และความเที่ยงระหว่างผู้เชี่ยวชาญของคะแนนจุดตัดวิชาวิทยาศาสตร์

บทที่ 5

สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ

การศึกษาครั้งนี้มีวัตถุประสงค์ 1) เพื่อพัฒนาวิธีการกำหนดคะแนนจุดตัดสำหรับการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐานวิชาคณิตศาสตร์ และวิชาวิทยาศาสตร์ ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 2) เพื่อกำหนดคะแนนจุดตัดด้วยวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซี 3) เพื่อตรวจสอบคุณภาพของคะแนนจุดตัด 4) เพื่อวิเคราะห์จำนวนผู้เชี่ยวชาญ และจำนวนกลุ่มผู้เชี่ยวชาญ ที่มีความเหมาะสม กลุ่มตัวอย่างผู้เชี่ยวชาญเป็นครูผู้สอนวิชาคณิตศาสตร์จำนวน 27 คน และเป็นครูผู้สอนวิชาฟิสิกส์ จำนวน 10 คน วิชาเคมี จำนวน 10 คน และวิชาชีววิทยา จำนวน 10 คน ข้อสอบที่ใช้ในการกำหนดคะแนนจุดตัดเป็นข้อสอบสำหรับการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน ช่วงชั้นที่ 4 (มัธยมศึกษาปีที่ 6) ปีการศึกษา 2553 วิชาคณิตศาสตร์ จำนวน 40 ข้อ แบ่งเป็นข้อสอบแบบเลือกตอบ 4 ตัวเลือก จำนวน 20 ข้อ และ ข้อสอบแบบเติมคำตอบ จำนวน 20 ข้อ และวิชาวิทยาศาสตร์ มีจำนวน 90 ข้อ แบ่งเป็นข้อสอบแบบเลือกตอบที่มีคำตอบถูกเพียงข้อเดียว จำนวน 80 ข้อ และข้อสอบแบบเลือกตอบที่มีคำตอบถูก 2 ข้อ จำนวน 10 ข้อ

ขั้นตอนการดำเนินการมี 3 ระยะ โดยระยะที่ 1 ผู้วิจัยศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องกับการกำหนดคะแนนจุดตัดเพื่อศึกษาแนวคิด และวิธีการในการกำหนดคะแนนจุดตัด รวมทั้งพิจารณาถึงจุดเด่นและจุดด้อยของวิธีการต่างๆ โดยผู้วิจัยใช้วิธีการสืบค้นข้อมูลจากอินเตอร์เน็ตเป็นหลัก นอกจากนี้ได้สอบถามความรู้จากผู้เชี่ยวชาญในต่างประเทศ ผ่านทาง E-mail และพัฒนาวิธีการกำหนดคะแนนจุดตัดขึ้นเรียกว่า “วิธีเอบีซี” ในระยะที่ 2 ผู้วิจัยวิเคราะห์เนื้อหาจากเอกสารและงานวิจัยเพื่อศึกษาแนวคิดและหลักการในการกำหนดจำนวนของระดับ ชื่อของระดับ และคำอธิบายระดับความสามารถ รวมถึงศึกษาความคิดเห็นของผู้บริหารระดับสูงในหน่วยงานที่รับผิดชอบการจัดการศึกษาขั้นพื้นฐาน จากนั้นจึงร่างจำนวนของระดับ ชื่อของระดับ และคำอธิบายระดับความสามารถแล้วนำร่างดังกล่าวไปให้ผู้เชี่ยวชาญพิจารณา เพื่อขอข้อเสนอแนะในการปรับปรุงให้มีความสมบูรณ์ยิ่งขึ้น ในระยะที่ 3 ผู้วิจัยจัดการกำหนดคะแนนจุดตัดตามวิธีการที่ได้กำหนดไว้

สรุปผลการวิจัย

1. ผลการพัฒนาวิธีการกำหนดคะแนนจุดตัด

วิธีเอบีซีถูกพัฒนาขึ้นจากการผสมผสานระหว่างวิธีแองกอฟแบบใช่/ไม่ใช่ และวิธีบุ๊คมาร์คเพื่อดึงเอาจุดเด่นและลดข้อด้อยของแต่ละวิธีลง โดยวิธีเอบีซีมีขั้นตอนการดำเนินการ 3 รอบ ตามลำดับดังนี้

รอบที่ 1 ผู้เชี่ยวชาญกำหนดคะแนนจุดตัดจากคู่มือเรียงข้อสอบรายสาระการเรียนรู้ (Domain–Ordered Item Booklet: D-OIB) ซึ่งเป็นคู่มือที่จัดทำขึ้นโดยแบ่งข้อสอบเป็นกลุ่มตามสาระการเรียนรู้ กล่าวคือคู่มือเรียงข้อสอบรายสาระการเรียนรู้จะมีหลายฉบับขึ้นอยู่กับจำนวนสาระการเรียนรู้ที่สอบวัด ข้อสอบในแต่ละสาระการเรียนรู้จะถูกจัดเรียงจากข้อง่ายไปข้อยาก ข้อละ 1 หน้า โดยใช้ทฤษฎีการตอบสนองข้อสอบ โมเดล 3 พารามิเตอร์ ในการวิเคราะห์ค่าความยากของข้อสอบ (b) นอกจากนี้วิธีเอบีซีจะเสนอสารสนเทศที่มากกว่าค่าความยากของข้อสอบ โดยจะเสนอสารสนเทศเกี่ยวกับผลการวิเคราะห์กลุ่ม สารสนเทศนี้จะนำเสนอเป็นผลคะแนนเฉลี่ยรายสาระการเรียนรู้ของแต่ละกลุ่ม และช่วงของคะแนนสอบที่คาบเกี่ยวกันของแต่ละกลุ่ม

หลักการพิจารณาเพื่อกำหนดคะแนนจุดตัดคือให้ผู้เชี่ยวชาญพิจารณาข้อสอบตามลำดับความยาก และตัดสินว่าผู้ที่สามารถตอบข้อสอบแต่ละข้อได้ถูกต้องควรมีความสามารถอย่างน้อยระดับใด นอกจากนี้วิธีเอบีซีมีหลักการพิจารณาเพิ่มเติมคือหากข้อใดที่ผู้เชี่ยวชาญไม่มั่นใจก็ให้ระบุไว้ในข้อสอบ ซึ่งสำหรับข้อสอบที่ผู้เชี่ยวชาญยังไม่มั่นใจนั้น ผู้เชี่ยวชาญสามารถนำประเด็นที่ไม่มั่นใจมาอภิปรายร่วมกันกับผู้เชี่ยวชาญท่านอื่น ๆ ในรอบที่ 2 และ 3 ได้

รอบที่ 2 ผู้เชี่ยวชาญร่วมกันอภิปรายกลุ่มย่อยเกี่ยวกับข้อสอบที่ผู้เชี่ยวชาญยังไม่มั่นใจในรอบที่ 1 เพื่อให้การอภิปรายกระชับและตรงประเด็นมากขึ้น และกำหนดคะแนนจุดตัดอีกครั้ง โดยอาศัยหลักการเดียวกับรอบที่ 1

รอบที่ 3 ผู้ตัดสินร่วมกันอภิปรายกลุ่มใหญ่เกี่ยวกับข้อสอบที่ผู้ตัดสินยังไม่มั่นใจในรอบที่ 2 และกำหนดคะแนนจุดตัดอีกครั้ง โดยอาศัยหลักการเดียวกับรอบที่ 1

2. ผลการกำหนดคะแนนจุดตัด

2.1 คะแนนจุดตัดวิชาคณิตศาสตร์ระดับ 1/2 เฉลี่ยทุกวิธีเท่ากับ 20.22 คะแนน เมื่อพิจารณาแต่ละรอบ พบว่า คะแนนจุดตัดระดับ 1/2 ในรอบที่ 1 เฉลี่ยเท่ากับ 23.63 คะแนน รอบที่ 2 เฉลี่ยเท่ากับ 19.93 คะแนน รอบที่ 3 เฉลี่ยเท่ากับ 17.11 คะแนน สำหรับคะแนนจุดตัดระดับ 2/3 เฉลี่ยทุกวิธีเท่ากับ 49.58 คะแนน เมื่อพิจารณาแต่ละรอบ พบว่า คะแนนจุดตัดระดับ 2/3 ในรอบที่ 1 เฉลี่ยเท่ากับ 55.82 คะแนน รอบที่ 2 เฉลี่ยเท่ากับ 50.44 คะแนน รอบที่ 3 เฉลี่ยเท่ากับ 42.48 คะแนน ส่วนคะแนนจุดตัดระดับ 3/4 เฉลี่ยทุกวิธีเท่ากับ 77.30 คะแนน เมื่อพิจารณาแต่ละรอบ พบว่า คะแนนจุดตัดระดับ 2/3 ในรอบที่ 1 เฉลี่ยเท่ากับ 79.89 คะแนน รอบที่ 2 เฉลี่ยเท่ากับ 75.29 คะแนน รอบที่ 3 เฉลี่ยเท่ากับ 76.70 คะแนน

คะแนนจุดตัดวิชาคณิตศาสตร์ระดับ 1/2, ระดับ 2/3 และระดับ 3/4 ที่กำหนดด้วยวิธีแองกอฟแบบใช่/ไม่ใช่ มีค่าเท่ากับ 15.11, 42.63 และ81.26 ตามลำดับ วิธีบุ๊คมาร์ค มีค่า 39.22, 55.37 และ 70.26 ตามลำดับ และวิธีเอบีซี มีค่า 6.34, 50.74 และ 80.37 ตามลำดับ

2.2 คะแนนจุดตัดวิชาวิทยาศาสตร์ระดับ 1/2 เฉลี่ยทุกวิธีเท่ากับ 29.88 เมื่อพิจารณาแต่ละรอบ พบว่า คะแนนจุดตัดระดับ 1/2 ในรอบที่ 1 เฉลี่ยเท่ากับ 27.24 รอบที่ 2 เฉลี่ยเท่ากับ 31.30 รอบที่ 3 เฉลี่ยเท่ากับ 31.10 สำหรับคะแนนจุดตัดระดับ 2/3 เฉลี่ยทุกวิธีเท่ากับ 63.59 เมื่อพิจารณาแต่ละรอบ พบว่า คะแนนจุดตัดระดับ 2/3 ในรอบที่ 1 เฉลี่ยเท่ากับ 63.47 รอบที่ 2 เฉลี่ยเท่ากับ 63.89 รอบที่ 3 เฉลี่ยเท่ากับ 63.59 ส่วนคะแนนจุดตัดระดับ 3/4 เฉลี่ยทุกวิธีเท่ากับ 83.70 เมื่อพิจารณาแต่ละรอบ พบว่า คะแนนจุดตัดระดับ 3/4 ในรอบที่ 1 เฉลี่ยเท่ากับ 79.95 รอบที่ 2 เฉลี่ยเท่ากับ 84.92 รอบที่ 3 เฉลี่ยเท่ากับ 83.50

คะแนนจุดตัดวิชาวิทยาศาสตร์ระดับ 1/2 ระดับ 2/3 และระดับ 3/4 ที่กำหนดด้วยวิธีแองกอฟแบบใช่/ไม่ใช่ มีค่าเท่ากับ 44.17, 72.44 และ88.52 ตามลำดับ วิธีบุ๊คมาร์ค มีค่า 23.52, 53.48 และ 76.93 ตามลำดับ และวิธีเอบีซี มีค่า 21.96, 64.85 และ 85.67 ตามลำดับ

3. ผลวิเคราะห์คุณภาพด้านความตรง

3.1 ผลวิเคราะห์ความตรงของปัจจัยนำเข้า

3.1.1 ผลวิเคราะห์ความตรงของปัจจัยนำเข้าของการกำหนดคะแนนจุดตัดวิชาคณิตศาสตร์ พบว่า ผู้เชี่ยวชาญมีความคิดเห็นว่าการกำหนดคะแนนจุดตัดด้วยวิธีแองกอฟแบบใช่/ไม่ใช่มีความตรงของปัจจัยนำเข้าโดยภาพรวมอยู่ในระดับมาก เมื่อพิจารณาเป็นรายด้าน พบว่า ผู้เชี่ยวชาญมีความคิดเห็นว่าการกำหนดคะแนนจุดตัดด้วยวิธีแองกอฟแบบใช่/ไม่ใช่มีความตรงของปัจจัยนำเข้าด้านความเข้าใจและความมั่นใจของผู้เชี่ยวชาญสูงที่สุด ซึ่งอยู่ในระดับมาก รองลงมาคือด้านความถูกต้อง อยู่ในระดับมาก ส่วนด้านความเป็นประโยชน์ของเอกสาร อยู่ในระดับที่ต่ำที่สุด โดยอยู่ในระดับปานกลาง สำหรับวิธีบุ๊คมาร์คผู้เชี่ยวชาญมีความคิดเห็นว่าการกำหนดคะแนนจุดตัดมีความตรงของปัจจัยนำเข้าโดยภาพรวมอยู่ในระดับมาก เมื่อพิจารณาเป็นรายด้าน พบว่า ผู้เชี่ยวชาญมีความคิดเห็นว่าการกำหนดคะแนนจุดตัดด้วยวิธีบุ๊คมาร์คมีความตรงของปัจจัยนำเข้าด้านความเหมาะสมสูงที่สุด ซึ่งอยู่ในระดับมาก รองลงมาคือด้านความถูกต้อง อยู่ในระดับมาก ส่วนด้านความเป็นประโยชน์ของเอกสาร อยู่ในระดับที่ต่ำที่สุด โดยอยู่ในระดับมาก นอกจากนี้ผู้เชี่ยวชาญมีความคิดเห็นว่าการกำหนดคะแนนจุดตัดด้วยวิธีเอบีซีมีความตรงของปัจจัยนำเข้าโดยภาพรวมอยู่ในระดับมาก เมื่อพิจารณาเป็นรายด้าน พบว่า ผู้เชี่ยวชาญมีความคิดเห็นว่าการกำหนดคะแนนจุดตัดด้วยวิธีเอบีซีมีความตรงของปัจจัยนำเข้าด้านความเข้าใจและความมั่นใจของผู้เชี่ยวชาญสูงที่สุด ซึ่งอยู่ในระดับมาก รองลงมาคือด้านความถูกต้อง อยู่ในระดับมาก ส่วนด้านความเหมาะสม อยู่ในระดับที่ต่ำที่สุด อยู่ในระดับมาก

3.1.2 ผลวิเคราะห์ความตรงของปัจจัยนำเข้าของการกำหนดคะแนนจุดตัดวิชาวิทยาศาสตร์ พบว่า ผู้เชี่ยวชาญมีความคิดเห็นว่าการกำหนดคะแนนจุดตัดด้วยวิธีแองกอฟแบบใช่/ไม่ใช่มีความตรงของปัจจัยนำเข้าโดยภาพรวมอยู่ในระดับมาก เมื่อพิจารณาเป็นรายด้าน พบว่า ผู้เชี่ยวชาญมีความคิดเห็นว่าการกำหนดคะแนนจุดตัดด้วยวิธีแองกอฟแบบใช่/ไม่ใช่มีความตรงของปัจจัยนำเข้าด้านความเหมาะสมสูงที่สุด อยู่ในระดับมาก รองลงมาคือด้านความเข้าใจและความมั่นใจของผู้เชี่ยวชาญ อยู่ในระดับมาก ส่วนด้านความเป็นประโยชน์ของเอกสาร อยู่ในระดับที่ต่ำที่สุด อยู่ในระดับปานกลาง สำหรับวิธีบุ๊คมาร์ค ผู้เชี่ยวชาญมีความคิดเห็นว่าการกำหนดคะแนนจุดตัดมีความตรงของปัจจัยนำเข้าโดยภาพรวมอยู่ในระดับมาก เมื่อพิจารณาเป็นรายด้าน พบว่า

ผู้เชี่ยวชาญมีความคิดเห็นว่าการกำหนดคะแนนจุดตัดด้วยวิธีบุ๊คมาร์คมีความตรงของปัจจัยนำเข้าด้านความถูกต้อง อยู่ในระดับมาก รองลงมาคือด้านความเข้าใจและความมั่นใจของผู้เชี่ยวชาญอยู่ในระดับมาก ส่วนด้านความเป็นประโยชน์ของเอกสาร อยู่ในระดับที่ต่ำที่สุด โดยอยู่ในระดับมาก ส่วนวิธีเอบีซี ผู้เชี่ยวชาญมีความคิดเห็นว่าการกำหนดคะแนนจุดตัดมีความตรงของปัจจัยนำเข้าโดยภาพรวมอยู่ในระดับมาก เมื่อพิจารณาเป็นรายด้าน พบว่า ผู้เชี่ยวชาญมีความคิดเห็นว่าการกำหนดคะแนนจุดตัดด้วยวิธีเอบีซีมีความตรงของปัจจัยนำเข้าด้านความเข้าใจและความมั่นใจของผู้เชี่ยวชาญสูงที่สุด ซึ่งอยู่ในระดับมาก รองลงมาคือด้านความถูกต้อง อยู่ในระดับมาก ส่วนด้านความเป็นไปได้ อยู่ในระดับที่ต่ำที่สุด อยู่ในระดับมาก

3.2 ผลวิเคราะห์ความตรงของกระบวนการตัดสินใจ

3.2.1 หลักการที่ผู้เชี่ยวชาญใช้ในการพิจารณาเพื่อกำหนดคะแนนจุดตัด

3.2.1.1 ผู้เชี่ยวชาญกำหนดคะแนนจุดตัดโดยพิจารณาระดับพฤติกรรมด้านพุทธิพิสัยที่ข้อสอบต้องการวัด โดยระดับที่พิจารณาส่วนใหญ่ได้แก่ ความรู้ความจำ ความเข้าใจ การนำไปประยุกต์ใช้ และการวิเคราะห์

3.2.1.2 ผู้เชี่ยวชาญกำหนดคะแนนจุดตัดโดยพิจารณาเนื้อหาที่สอบวัด รูปแบบของข้อสอบ และขั้นตอนของการได้มาซึ่งคำตอบที่ถูกต้อง โดยผู้เชี่ยวชาญส่วนใหญ่จะลงมือทำข้อสอบเพื่อพิจารณารายละเอียดของการได้มาซึ่งคำตอบที่ถูกต้อง อาทิ เวลาที่ใช้ในการทำข้อสอบ จุดที่ผู้สอบมีโอกาสผิดพลาดหากไม่รอบคอบพอ และความซับซ้อนของขั้นตอนในการหาคำตอบ เป็นต้น

3.2.2 ปัญหา

2.2.2.1 ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดด้วยวิธีบุ๊คมาร์คมีความคิดเห็นว่าข้อสอบบางข้อที่เรียงไว้ในคู่มือการจัดเรียงข้อสอบมีลำดับที่ไม่เหมาะสม โดยเฉพาะข้อสอบวิชาคณิตศาสตร์ ซึ่งประกอบด้วยข้อสอบ 2 แบบ ได้แก่ ข้อสอบแบบเลือกตอบ และข้อสอบแบบเติมคำตอบ โดยประเด็นที่ผู้เชี่ยวชาญตั้งข้อสังเกตคือข้อสอบแบบเติมคำตอบนั้น เมื่อพิจารณาโดยเนื้อหาที่สอบวัดแล้ว เป็นเนื้อหาที่ไม่ซับซ้อนและไม่ยากมากนักแต่กลับอยู่ในลำดับที่ยากเกินไป นอกจากนี้ข้อสอบแบบเลือกตอบบางข้อซึ่งผู้เชี่ยวชาญมองว่าเป็นข้อที่ยากกว่าข้อสอบแบบเติมคำตอบบางข้อ แต่กลับถูกเรียงไว้ในตำแหน่งที่ง่าย หลังจากที่ผู้เชี่ยวชาญร่วมกันอภิปรายในประเด็น

นี้เพื่อหาข้อสรุป พบว่า ผู้เชี่ยวชาญหลายท่านตั้งข้อสังเกตว่า รูปแบบของข้อสอบที่ใช้อาจมีผลต่อความยากของข้อสอบ กล่าวคือข้อสอบแบบเติมคำส่วนใหญ่ เน้นวัดเนื้อหาที่ไม่ยากและไม่ซับซ้อนนัก แต่ผู้สอบกลับทำข้อสอบแบบนี้ได้ค่อนข้างน้อย เมื่อพิจารณาข้อสอบแบบเลือกตอบซึ่งมีเนื้อหาที่สอบวัดค่อนข้างยาก รวมถึงบางข้อต้องมีการประยุกต์ใช้ความรู้ แต่ผู้สอบกลับตอบถูกมากกว่าข้อสอบแบบเติมคำ ซึ่งประเด็นนี้เป็นที่มาของข้อสรุปในประเด็นที่อภิปรายว่า ผู้สอบส่วนใหญ่ทำข้อสอบแบบเติมคำตอบได้ค่อนข้างน้อย อาจเป็นผลจากรูปแบบข้อสอบที่ผู้สอบไม่คุ้นเคย ซึ่งจากข้อสรุปนี้จึงนำไปสู่การที่ผู้เชี่ยวชาญกำหนดคะแนนจุดตัดโดยจะไม่พิจารณาเฉพาะเนื้อหาที่สอบวัดเท่านั้นแต่จะพิจารณาถึงรูปแบบของข้อสอบด้วย

2.2.2.2 ผู้เชี่ยวชาญบางท่านมีแนวโน้มที่จะกำหนดคะแนนจุดตัดให้ใกล้เคียงกับผู้เชี่ยวชาญที่ตนเห็นว่าเก่งและสามารถอธิบาย หรือแสดงความคิดเห็นต่าง ๆ ได้อย่างมีหลักการ

2.2.2.3 การอภิปรายร่วมกันระหว่างผู้เชี่ยวชาญหากดำเนินการไม่ดีอาจทำให้ผู้เชี่ยวชาญเบื่อหน่ายได้ จึงควรอภิปรายเกี่ยวกับข้อสอบ หรือประเด็นที่ผู้เชี่ยวชาญยังลังเลที่จะตัดสินใจ เพื่อให้ผู้เชี่ยวชาญท่านอื่นได้แสดงความคิดเห็นในประเด็นนั้น ซึ่งจะทำให้ผู้เชี่ยวชาญได้สารสนเทศประกอบการตัดสินใจมากขึ้น และใช้เวลาในการอภิปรายน้อยลง อันเป็นการลดปัญหาความเบื่อหน่ายที่อาจจะเกิดขึ้นได้

2.2.3 ข้อเสนอแนะ

2.2.3.1 ผู้เชี่ยวชาญเสนอแนะว่าการอภิปรายกลุ่มช่วยให้ได้แลกเปลี่ยนความรู้ซึ่งกันและกัน ซึ่งเป็นประโยชน์อย่างยิ่งต่อการกำหนดคะแนนจุดตัด เนื่องจากสามารถสร้างความมั่นใจในการกำหนดคะแนนจุดตัดให้กับผู้เชี่ยวชาญ นอกจากนี้ผู้เชี่ยวชาญเสนอว่าการอภิปรายกลุ่มไม่ควรเน้นให้ผู้ที่กำหนดคะแนนจุดตัดสูงที่สุดและต่ำที่สุดอธิบายเหตุผลของการกำหนดคะแนนจุดตัดนั้น ๆ แต่ควรเน้นให้ทุกคนได้สอบถามหรือตั้งประเด็นที่ตนเองไม่เข้าใจหรือไม่แน่ใจ เพื่อให้ผู้เชี่ยวชาญท่านอื่น ๆ ได้ร่วมแสดงความคิดเห็นหรือแลกเปลี่ยนความรู้เพื่อสร้างความกระจ่างให้มากขึ้น

2.2.3.2 ผู้เชี่ยวชาญเสนอแนะว่าควรแบ่งกลุ่มข้อสอบในแต่สาระการเรียนรู้ตามระดับพฤติกรรมด้านพุทธิพิสัยที่วัด เพื่อให้การกำหนดคะแนนจุดตัดทำได้เร็วขึ้น

2.2.3.3 ผู้เชี่ยวชาญเสนอแนะให้นักเรียนได้มีส่วนร่วมในการกำหนดคะแนนจุดตัดด้วย ซึ่งจะทำให้ผู้เชี่ยวได้แลกเปลี่ยนมุมมองกับนักเรียนน่าจะช่วยให้การกำหนดคะแนนจุดตัดมีความเหมาะสมและเป็นธรรมกับทุกฝ่ายมากขึ้น

2.3 ผลวิเคราะห์ความตรงของผลลัพธ์

2.3.1 คะแนนจุดตัดวิชาคณิตศาสตร์ที่กำหนดโดยวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซี มีความสอดคล้องกันน้อย

2.3.2 คะแนนจุดตัดวิชาวิทยาศาสตร์ที่กำหนดโดยวิธีแองกอฟแบบใช่/ไม่ใช่ มีความสอดคล้องน้อยกับคะแนนจุดตัดที่กำหนดโดยวิธีบุ๊คมาร์ค และวิธีเอบีซี ส่วนคะแนนจุดตัดที่กำหนดโดยวิธีบุ๊คมาร์ค และวิธีเอบีซีมีความสอดคล้องกันในระดับปานกลาง

2.4 ผลการวิเคราะห์เพื่อตรวจสอบความตรงของผลที่ตามมา

2.4.1 ความถูกต้องของการกำหนดกลุ่มของคะแนนจุดตัดวิชาคณิตศาสตร์ที่กำหนดขึ้นด้วยวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซี มีค่าแตกต่างกัน โดยวิธีบุ๊คมาร์คมีค่าความถูกต้องของการกำหนดกลุ่มเฉลี่ยสูงที่สุด รองลงมาคือวิธีเอบีซี ส่วนวิธีแองกอฟแบบใช่/ไม่ใช่มีค่าความถูกต้องของการกำหนดกลุ่มเฉลี่ยต่ำที่สุด เมื่อพิจารณาความคลาดเคลื่อนในการกำหนดกลุ่มของคะแนนจุดตัด โดยพิจารณาจากความคลาดเคลื่อนแบบบวก พบว่า วิธีแองกอฟแบบใช่/ไม่ใช่มีความคลาดเคลื่อนแบบบวกเฉลี่ยสูงที่สุด รองลงมาคือวิธีบุ๊คมาร์ค ส่วนวิธีเอบีซีมีความคลาดเคลื่อนแบบบวกเฉลี่ยต่ำที่สุด และเมื่อพิจารณาความคลาดเคลื่อนแบบลบ พบว่า วิธีเอบีซีมีความคลาดเคลื่อนแบบลบสูงที่สุด รองลงมาคือวิธีแองกอฟแบบใช่/ไม่ใช่ ส่วนวิธีบุ๊คมาร์คมีความคลาดเคลื่อนแบบลบต่ำที่สุด

2.4.2 ความถูกต้องของการกำหนดกลุ่มของคะแนนจุดตัดวิชาวิทยาศาสตร์ที่กำหนดขึ้นด้วยวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซี มีค่าแตกต่างกัน โดยวิธีเอบีซีมีความถูกต้องของการกำหนดกลุ่มเฉลี่ยสูงที่สุด รองลงมาคือวิธีแองกอฟแบบใช่/ไม่ใช่ ส่วนวิธีบุ๊คมาร์คมีความถูกต้องของการกำหนดกลุ่มเฉลี่ยต่ำที่สุด เมื่อพิจารณาความคลาดเคลื่อนของการกำหนดกลุ่มของคะแนนจุดตัด โดยพิจารณาจากความคลาดเคลื่อนแบบบวก พบว่า วิธีแองกอฟแบบใช่/ไม่ใช่มีความคลาดเคลื่อนแบบบวกเฉลี่ยสูงที่สุด รองลงมาคือวิธีบุ๊คมาร์ค ส่วนวิธีเอบีซีมีความคลาดเคลื่อนแบบบวกเฉลี่ยต่ำที่สุด ส่วนความคลาดเคลื่อนแบบลบ พบว่า วิธีบุ๊คมาร์คมีความ

คลาดเคลื่อนแบบลบสูงที่สุด รองลงมาคือวิธีเอบีซี ส่วนวิธีแองกอฟแบบใช่/ไม่ใช่ มีความคลาดเคลื่อนแบบลบต่ำที่สุด

4. ผลวิเคราะห์คุณภาพด้านความเที่ยง

4.1 ผลวิเคราะห์ค่าสัมประสิทธิ์การสรุปอ้างอิง

ค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์ของการกำหนดคะแนนจุดตัดทั้งวิชาคณิตศาสตร์และวิชาวิทยาศาสตร์ด้วยวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซี มีค่าแตกต่างกัน ทุกวิธีจะมีค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์เพิ่มขึ้นเมื่อเพิ่มจำนวนผู้เชี่ยวชาญให้มากขึ้น ซึ่งการเพิ่มจำนวนกลุ่มให้มากขึ้นจะทำให้ค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์ของการกำหนดคะแนนจุดตัดด้วยวิธีแองกอฟแบบใช่/ไม่ใช่สูงกว่าการเพิ่มจำนวนผู้เชี่ยวชาญในแต่ละกลุ่มให้มากขึ้น ส่วนวิธีบุ๊คมาร์ค และวิธีเอบีซี พบว่า การเพิ่มจำนวนกลุ่มให้มากขึ้นหรือการเพิ่มจำนวนผู้เชี่ยวชาญในแต่ละกลุ่มให้มากขึ้นทำให้ค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์ไม่แตกต่างกันนัก

4.2 ผลวิเคราะห์ความเที่ยงภายในผู้เชี่ยวชาญ และความเที่ยงระหว่างผู้เชี่ยวชาญ

4.2.1 ความเที่ยงภายในผู้เชี่ยวชาญ และความเที่ยงระหว่างผู้เชี่ยวชาญของการกำหนดคะแนนจุดตัดวิชาคณิตศาสตร์แต่ละวิธีแตกต่างกัน โดยมีรายละเอียดดังนี้

4.2.1.1 ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดด้วยวิธีแองกอฟแบบใช่/ไม่ใช่ กำหนดคะแนนจุดตัดในแต่ละรอบได้สอดคล้องกันในระดับปานกลาง ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดด้วยวิธีบุ๊คมาร์ค กำหนดคะแนนจุดตัดในแต่ละรอบได้สอดคล้องกันอย่างยิ่ง และผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดด้วยวิธีเอบีซี กำหนดคะแนนจุดตัดในแต่ละรอบได้สอดคล้องกันในระดับปานกลาง

4.2.1.2 ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดด้วยวิธีแองกอฟแบบใช่/ไม่ใช่ กำหนดคะแนนจุดตัดได้สอดคล้องกันในระดับมาก ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดด้วยวิธีบุ๊คมาร์ค กำหนดคะแนนจุดตัดได้สอดคล้องกันในระดับมาก และผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดด้วยวิธีเอบีซี กำหนดคะแนนจุดตัดได้สอดคล้องกันในระดับมาก

4.2.2 ความเที่ยงภายในผู้เชี่ยวชาญ และความเที่ยงระหว่างผู้เชี่ยวชาญของการกำหนดคะแนนจุดตัดวิชาวิทยาศาสตร์แต่ละวิธีแตกต่างกัน โดยมีรายละเอียดดังนี้

4.2.2.1 ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดด้วยวิธีแองกอฟแบบใช่/ไม่ใช่ กำหนดคะแนนจุดตัดในแต่ละรอบได้สอดคล้องกันอย่างยิ่ง ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดด้วยวิธีบุ๊คมาร์ค กำหนดคะแนนจุดตัดในแต่ละรอบได้สอดคล้องกันอย่างยิ่ง และผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดด้วยวิธีเอบีซี กำหนดคะแนนจุดตัดในแต่ละรอบได้สอดคล้องกันมาก

4.2.2.2 ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดด้วยวิธีแองกอฟแบบใช่/ไม่ใช่ กำหนดคะแนนจุดตัดได้สอดคล้องกันอย่างยิ่ง ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดด้วยวิธีบุ๊คมาร์ค กำหนดคะแนนจุดตัดได้สอดคล้องกันในระดับปานกลาง และผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดด้วยวิธีเอบีซี กำหนดคะแนนจุดตัดได้สอดคล้องกันมาก

อภิปรายผล

1. จากผลการกำหนดคะแนนจุดตัด ผู้วิจัยมีประเด็นที่จะอภิปรายดังนี้

1.1 คะแนนจุดตัดที่กำหนดด้วยวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซี มีค่าแตกต่างกัน กล่าวคือ คะแนนจุดตัดวิชาคณิตศาสตร์ที่กำหนดโดยวิธีบุ๊คมาร์คมีค่าสูงที่สุด รองลงมาคือคะแนนจุดตัดที่กำหนดโดยวิธีแองกอฟแบบใช่/ไม่ใช่ และคะแนนจุดตัดที่ต่ำที่สุดเป็นคะแนนจุดตัดที่กำหนดโดยวิธีเอบีซี ส่วนคะแนนจุดตัดวิชาวิทยาศาสตร์ที่กำหนดด้วยวิธีแองกอฟแบบใช่/ไม่ใช่มีค่าสูงที่สุด รองลงมาคือคะแนนจุดตัดที่กำหนดโดยวิธีบุ๊คมาร์ค และคะแนนจุดตัดที่ต่ำที่สุดเป็นคะแนนจุดตัดที่กำหนดโดยวิธีเอบีซี ตามลำดับ ซึ่งสอดคล้องกับผลการวิจัยของ Yin และ Schulz (2005 cited in Yin & Sconing, 2008) ที่ศึกษาเพื่อเปรียบเทียบผลการกำหนดคะแนนจุดตัดระหว่างวิธีบุ๊คมาร์คที่ได้รับการปรับปรุง ชื่อว่าวิธีแมปมาร์ค กับวิธีแองกอฟ ผลการศึกษา พบว่า คะแนนจุดตัดที่กำหนดโดยวิธีแมปมาร์คมีค่าต่ำกว่าคะแนนจุดตัดที่กำหนดโดยวิธีแองกอฟ สอดคล้องกับผลการวิจัยของ Jaeger (1989 cited in Kaftandjieva , 2010) ที่ศึกษาเปรียบเทียบคะแนนจุดตัดซึ่งกำหนดด้วยวิธีการที่แตกต่างกันจำนวน 32 คู่ พบว่า คะแนนจุดตัดแต่ละคู่มีความแตกต่างกัน โดยช่วงคะแนนที่ต่างกันมีตั้งแต่ 1 – 52 คะแนน นอกจากนี้มีคะแนนจุดตัดจำนวน 15 คู่ที่คะแนนจุดตัดค่าหนึ่งมีค่ามากกว่าอีกค่าหนึ่งถึง 1.5 เท่า และสอดคล้อง

กับผลการวิจัยของ Bontempo, Marks และKarabatsos (1998) ที่ได้วิเคราะห์อภิมาน (meta-analysis) เพื่อเปรียบเทียบผลการกำหนดคะแนนจุดตัดจากงานวิจัยจำนวน 9 เรื่อง พบว่า ไม่ว่าจะกำหนดคะแนนจุดตัดด้วยวิธีการเดียวกันสองครั้ง หรือกำหนดคะแนนจุดตัดด้วยวิธีการที่ต่างกัน คะแนนจุดตัดที่ได้ล้วนแตกต่างกัน

1.2 เมื่อพิจารณาคะแนนจุดตัดที่กำหนดด้วยวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซีในแต่ละรอบ พบว่า ในรอบที่ 1 ผู้เชี่ยวชาญกำหนดคะแนนจุดตัดแตกต่างกันมาก ส่วนในรอบที่ 2 และรอบที่ 3 ผู้เชี่ยวชาญกำหนดคะแนนจุดตัดได้ใกล้เคียงกันมากขึ้น เห็นได้จากค่าความเบี่ยงเบนมาตรฐานในรอบที่ 2 และรอบที่ 3 ที่มีค่าลดลงจากรอบที่ 1 ทั้งนี้ในรอบที่ 1 ผู้เชี่ยวชาญแต่ละท่านกำหนดคะแนนจุดตัดโดยอาศัยแนวคิด และหลักการของตนเองเป็นหลัก สำหรับการกำหนดคะแนนจุดตัดในรอบที่ 2 นั้น ผู้เชี่ยวชาญได้ร่วมกันอภิปรายกลุ่มย่อย และในรอบที่ 3 ผู้เชี่ยวชาญได้ร่วมกันอภิปรายกลุ่มใหญ่ก่อนที่จะกำหนดคะแนนจุดตัด ซึ่งการอภิปรายร่วมกันของผู้เชี่ยวชาญทำให้ผู้เชี่ยวชาญได้แลกเปลี่ยนแนวคิด ความคิดเห็น ตลอดจนข้อสังเกตต่างๆ ร่วมกัน ดังนั้นการกำหนดคะแนนจุดตัดของผู้เชี่ยวชาญแต่ละท่านในรอบที่ 2 และรอบที่ 3 อาจได้รับอิทธิพลจากแนวคิด ความคิดเห็นของผู้เชี่ยวชาญท่านอื่น ๆด้วย กล่าวคือ หลังจากการอภิปรายร่วมกันอาจมีผู้เชี่ยวชาญบางท่านปรับเปลี่ยนคะแนนจุดตัดที่ได้กำหนดไว้ในรอบก่อน ๆ ซึ่งเห็นได้จากคะแนนจุดตัดของผู้เชี่ยวชาญที่แตกต่างกันในแต่ละรอบ ซึ่งการปรับเปลี่ยนคะแนนจุดตัดของผู้เชี่ยวชาญทำให้คะแนนจุดตัดของผู้เชี่ยวชาญแต่ละท่านมีค่าใกล้เคียงกันมากขึ้น

1.3 คะแนนจุดตัดวิชาคณิตศาสตร์ที่กำหนดด้วยวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซี ระดับ 1/2 มีความแตกต่างกันมาก คือมีค่าเฉลี่ยเท่ากับ 15.11, 39.22 และ 6.34 ซึ่งจะเห็นว่าคะแนนจุดตัดที่กำหนดด้วยวิธีบุ๊คมารค์จะสูงกว่าคะแนนจุดตัดกำหนดด้วยวิธีแองกอฟแบบใช่/ไม่ใช่ และวิธีเอบีซีมาก ทั้งนี้อาจเป็นผลจากอิทธิพลเพดาน (Ceiling effect) ซึ่งการกำหนดคะแนนจุดตัดด้วยวิธีบุ๊คมาร์คจะได้รับอิทธิพลนี้เมื่อข้อสอบที่ใช้ในการกำหนดคะแนนจุดตัดมีจำนวนน้อย ๆ และข้อสอบแต่ละข้อมีความยากใกล้เคียงกัน ซึ่งข้อสอบวิธีคณิตศาสตร์ที่ใช้ในการกำหนดคะแนนจุดตัดครั้งนี้มีข้อสอบเพียง 40 ข้อและข้อสอบส่วนใหญ่เป็นข้อสอบที่มีค่าความยากอยู่ระหว่าง 0.17 – 1.86

การที่วิธีบุ๊คมาร์คได้รับอิทธิพลนี้เมื่อจำนวนข้อสอบที่ใช้ในการกำหนดคะแนนจุดตัดมีไม่มาก และเป็นแบบสอบที่ข้อสอบแต่ละข้อมีความยากใกล้เคียงกัน เนื่องมาจากว่า วิธีบุ๊คมาร์คจะคิดคะแนนจุดตัดแต่ละค่าโดยการเปลี่ยนค่าความสามารถของผู้สอบ (θ) ที่มีความน่าจะเป็น 0.67 ที่จะตอบข้อสอบข้อที่อยู่ก่อนหน้าที่คั่นหนังสือได้ถูกต้องไปเป็นคะแนนสอบ ซึ่งเป็นการคิดคะแนนจุดตัดจากการคำนวณความน่าจะเป็นของการตอบข้อสอบได้ถูกต้องจากทุก ๆ ข้อ แนวคิดนี้เป็นแนวคิดที่ถือว่าผู้สอบแต่ละระดับความสามารถล้วนมีความน่าจะเป็นที่จะตอบข้อสอบได้ถูกต้องเพียงแต่มีระดับความน่าจะเป็นที่จะตอบข้อสอบได้ถูกต้องแตกต่างกันไป เมื่อลองพิจารณาข้อสอบที่ง่ายที่สุด คือข้อ 8 ซึ่งเป็นข้อสอบที่ผู้สอบที่มีความน่าจะเป็น 0.67 ที่จะตอบข้อสอบข้อนี้ได้ถูกต้องควรจะเป็นผู้สอบที่มีความสามารถ (θ) เท่ากับ 0.32 และเมื่อเปลี่ยนค่าความสามารถของผู้สอบ (θ) เป็นคะแนนสอบ จะได้เท่ากับ 23.829 คะแนน ซึ่งจะเห็นว่าหากข้อ 8 คือข้อสอบที่เป็นจุดตัด การคิดคะแนนจุดตัดตามแบบวิธีบุ๊คมาร์คจะได้คะแนนจุดตัดเท่ากับ 23.829 คะแนน จากกรณีนี้จะเห็นได้ว่าคะแนนจุดตัดในระดับต่ำ ๆ ดังเช่นในการวิจัยครั้งนี้คือระดับที่ 1/2 คะแนนจุดตัดที่ได้จะมีค่าสูงมาก

2. ผลจากการวิเคราะห์คุณภาพ

2.1 ผลจากการวิเคราะห์ความตรง

2.1.1 ผลจากการวิเคราะห์ความตรงของปัจจัยนำเข้า ผู้วิจัยมีประเด็นที่จะอภิปรายดังนี้

ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดด้วยวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซีมีความคิดเห็นว่ารายงานผลการกำหนดคะแนนจุดตัดในแต่ละรอบเป็นประโยชน์ต่อการกำหนดคะแนนจุดตัดในรอบต่อ ๆ ไป อยู่ในระดับปานกลาง กล่าวได้ว่าผู้เชี่ยวชาญมีความคิดเห็นว่าผลการกำหนดคะแนนจุดตัดของผู้เชี่ยวชาญท่านอื่น ๆ มีประโยชน์ต่อการตัดสินใจกำหนดคะแนนจุดตัดของตนเองไม่มากนัก นอกจากนี้ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดด้วยวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซีมีความคิดเห็นว่าการอภิปรายร่วมกันระหว่างผู้เชี่ยวชาญช่วยให้กำหนดคะแนนจุดตัดได้ง่ายขึ้นอยู่ในระดับมาก กล่าวคือ ผู้เชี่ยวชาญเห็นว่าการอภิปรายร่วมกันระหว่างผู้เชี่ยวชาญเป็นกระบวนการที่เป็นประโยชน์ต่อการกำหนดคะแนนจุดตัด

จากข้อมูลดังกล่าวพอจะสรุปได้ว่าในการกำหนดคะแนนจุดตัดด้วยวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซี ควรเปิดโอกาสให้ผู้เชี่ยวชาญได้อภิปรายร่วมกัน ซึ่งสอดคล้องกับผลการศึกษาของ Clauser และคณะ (2009) เป็นการศึกษาผลกระทบของการอภิปรายกลุ่มต่อการพิจารณากำหนดคะแนนจุดตัด พบว่า การอภิปรายกลุ่มช่วยให้ผลการตัดสินของผู้เชี่ยวชาญแต่ละท่านใกล้เคียงกันมากขึ้น ส่วนในแต่ละรอบของการกำหนดคะแนนจุดตัดอาจจะไม่จำเป็นต้องรายงานผลการกำหนดคะแนนจุดตัดของผู้เชี่ยวชาญแต่ละท่านที่กำหนดขึ้นในรอบก่อนให้ผู้เชี่ยวชาญท่านอื่น ๆ ทราบก็ได้

2.1.2 ผลจากการวิเคราะห์ความตรงของกระบวนการตัดสินใจ ผู้วิจัยมีประเด็นที่จะอภิปรายดังนี้

ผู้เชี่ยวชาญส่วนใหญ่จะกำหนดคะแนนจุดตัดโดยพิจารณาระดับพฤติกรรมด้านพุทธิพิสัยที่ข้อสอบต้องการวัด เช่น ความรู้ความจำ ความเข้าใจ การนำไปประยุกต์ใช้ และการวิเคราะห์ เป็นต้น นอกจากนี้ผู้เชี่ยวชาญส่วนใหญ่เห็นด้วยว่าการพิจารณากำหนดคะแนนจุดตัดจากข้อสอบที่มีการจัดไว้เป็นกลุ่มตามตามระดับพฤติกรรมด้านพุทธิพิสัยน่าจะช่วยให้การกำหนดคะแนนจุดตัดทำได้เร็วขึ้น โดยผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดด้วยวิธีเอบีซีส่วนใหญ่เสนอว่าหากข้อสอบแต่ละฉบับย่อย ๆ ที่ถูกแบ่งไว้ตามสาระการเรียนรู้ มีการแบ่งกลุ่มข้อสอบในแต่ละฉบับย่อยตามระดับพฤติกรรมด้านพุทธิพิสัยที่วัดน่าจะช่วยให้การกำหนดคะแนนจุดตัดทำได้เร็วขึ้น โดยผู้เชี่ยวชาญคิดว่าน่าจะใช้เวลาน้อยกว่าวิธีบุ๊คมาร์ค ซึ่งเมื่อกล่าวถึงความเร็วของการกำหนดคะแนนจุดตัดผู้วิจัยขออธิบายถึงความเร็วของการกำหนดคะแนนจุดตัดแต่ละวิธี จากการสังเกต พบว่าวิธีบุ๊คมาร์คใช้เวลาในการดำเนินการน้อยที่สุด รองลงมาเป็นวิธีเอบีซี และที่ใช้เวลามากที่สุดเป็นวิธีแองกอฟแบบใช่/ไม่ใช่ ดังนั้นหากมีการแบ่งกลุ่มข้อสอบตามระดับพฤติกรรมด้านพุทธิพิสัยที่วัดไว้ให้ผู้เชี่ยวชาญน่าจะช่วยให้การดำเนินการกำหนดคะแนนจุดตัดทุกวิธีทำได้เร็วขึ้นกว่าเดิม

2.1.3 ผลจากการวิเคราะห์ความตรงของผลลัพธ์ ผู้วิจัยมีประเด็นที่จะอภิปรายดังนี้

ผลจากการวิเคราะห์ความสอดคล้องของคะแนนจุดตัดที่กำหนดโดยวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซี พบว่า คะแนนจุดตัดมีความสอดคล้องกันน้อย Downing Tekian และ Yudkowsky (2006) Green และคณะ (2003) และ Berk (1996) กล่าวว่าการ

กำหนดคะแนนจุดตัดแต่ละวิธีมีหลายอย่างที่ต่างกัน เช่น วิธีการดำเนินการ สารสนเทศที่ใช้ประกอบการกำหนดคะแนนจุดตัด และผู้เชี่ยวชาญ เป็นต้น คะแนนจุดตัดที่กำหนดขึ้นจึงเป็นคะแนนที่ได้จากการตัดสินของผู้เชี่ยวชาญกลุ่มนั้น ๆ ซึ่งตัดสินใจโดยอาศัยความรู้และประสบการณ์ในขณะนั้นเท่านั้น จึงเป็นเรื่องปกติที่คะแนนจุดตัดที่กำหนดจะแตกต่างกันได้

นอกจากนี้ Browers (1989) ได้ศึกษาเพื่อเปรียบเทียบผลการกำหนดคะแนนจุดตัดด้วยวิธีแองกอฟโดยในรอบแรกให้ผู้เชี่ยวชาญกำหนดคะแนนจุดตัดโดยพิจารณาจากข้อสอบเพียงอย่างเดียว ส่วนรอบที่ 2 ให้สารสนเทศเพิ่มเติมเกี่ยวกับค่าความยากของข้อสอบ ผลจากการศึกษาพบว่า ผลการกำหนดคะแนนจุดตัดทั้งสองรอบมีความสัมพันธ์กันน้อยมาก ซึ่งกล่าวได้ว่าแม้ว่าผู้เชี่ยวชาญกลุ่มเดิมจะกำหนดคะแนนจุดตัดสองครั้งด้วยวิธีการเดิมแต่ได้สารสนเทศในแต่ละรอบต่างจากเดิม คะแนนจุดตัดที่ได้ก็ไม่เท่ากัน

จากการที่คะแนนจุดตัดที่กำหนดด้วยวิธีที่ต่างกันอาจจะแตกต่างกันได้ จึงเกิดแนวคิดเกี่ยวกับการเลือกวิธีการกำหนดคะแนนจุดตัดขึ้นสองแนวคิด กล่าวไว้โดย Zieky และคณะ (2008) แนวคิดแรกเสนอให้เลือกวิธีการกำหนดคะแนนจุดตัดเพียงวิธีเดียวแล้วดำเนินการตามวิธีนั้นอย่างเคร่งครัดและดีที่สุดเท่าที่จะทำได้ ส่วนแนวคิดที่สองเสนอให้เลือกหลาย ๆ วิธี หรืออย่างน้อย 2 วิธี และหากเกิดกรณีที่คะแนนจุดตัดที่กำหนดด้วยวิธีที่ต่างกันมีค่าแตกต่างกัน Green และคณะ(2003) ก็ได้เสนอแนวทางในการแก้ปัญหาดังกล่าวไว้ 3 แนวทาง โดยแนวทางแรกให้พิจารณาเลือกคะแนนจุดตัดที่กำหนดจากวิธีที่ผู้วิจัยมั่นใจว่าได้มีการดำเนินการตามขั้นตอนอย่างเคร่งครัดและถูกต้องที่สุด ซึ่งอาจพิจารณาจากคุณภาพของคะแนนจุดตัดเป็นต้น แนวทางที่สองคือการหาค่าเฉลี่ยของคะแนนจุดตัดจากทุก ๆ วิธีที่กำหนดได้ และแนวทางที่สาม คือการให้ผู้เชี่ยวชาญที่ไม่ได้เข้าร่วมกำหนดคะแนนจุดตัดวิธีใดวิธีหนึ่งที่กำหนดไว้ พิจารณาคะแนนจุดตัดทั้งหมดและให้ตัดสินใจเลือกคะแนนจุดตัดที่เหมาะสมที่สุด หรืออาจกำหนดคะแนนจุดตัดอีกครั้งก็ได้

2.2 ผลจากการวิเคราะห์ความเที่ยงของคะแนนจุดตัด ผู้วิจัยมีประเด็นที่จะอภิปรายดังนี้

2.2.1 ผลจากการวิเคราะห์ขนาดองค์ประกอบความแปรปรวนของการกำหนดคะแนนจุดตัด ค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์ จำนวนผู้เชี่ยวชาญและจำนวนกลุ่มผู้เชี่ยวชาญที่เหมาะสม ผู้วิจัยมีประเด็นที่จะอภิปรายดังนี้

2.2.2.1 ผลจากการศึกษา พบว่า องค์ประกอบความแปรปรวนหลักของการกำหนดคะแนนจุดตัดได้แก่ ผู้เชี่ยวชาญ กล่าวคือผู้เชี่ยวชาญแต่ละคนกำหนดคะแนนจุดตัดไม่ใกล้เคียงกันโดยเฉพาะในวิชาวิทยาศาสตร์ ซึ่งสอดคล้องกับงานวิจัยของ Yin และ Sconing (2008) และ Lee และ Lewis (2008) ที่พบว่าแหล่งความแปรปรวนหลักของการกำหนดคะแนนจุดตัดมาจากกลุ่มผู้เชี่ยวชาญและได้ให้ข้อเสนอแนะว่าการจะลดความคลาดเคลื่อนของการกำหนดคะแนนจุดตัดให้น้อยลงควรเพิ่มจำนวนผู้เชี่ยวชาญให้มีจำนวนมาก ๆ ทั้งนี้สาเหตุที่ผู้เชี่ยวชาญแต่ละคนกำหนดคะแนนจุดตัดได้ไม่ใกล้เคียงกันโดยเฉพาะในวิชาวิทยาศาสตร์นั้น อาจเนื่องมาจากว่าผู้เชี่ยวชาญที่เข้าร่วมการกำหนดคะแนนจุดตัดวิชาวิทยาศาสตร์ประกอบด้วยครูผู้สอนวิชาฟิสิกส์ วิชาเคมี และวิชาชีววิทยา กล่าวคือเป็นครูที่มีความถนัดและมีประสบการณ์ในการสอนในวิชาที่ต่างกัน จึงอาจทำให้ผลการกำหนดคะแนนจุดตัดมีความแตกต่างกันมาก

2.2.2.2 จากผลการวิเคราะห์ค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์ พบว่า การเพิ่มจำนวนกลุ่มผู้เชี่ยวชาญในการกำหนดคะแนนจุดตัดด้วยวิธีแองกอฟแบบใช่/ไม่ใช่ให้มากขึ้นจะทำให้ค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์สูงกว่าการเพิ่มจำนวนผู้เชี่ยวชาญในแต่ละกลุ่มให้มากขึ้น ส่วนการกำหนดคะแนนจุดตัดด้วยวิธีบุ๊คมาร์คและวิธีเอบีซีนั้น หากต้องการให้ค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์เพิ่มขึ้น ควรเพิ่มจำนวนผู้เชี่ยวชาญให้มากขึ้นโดยอาจเพิ่มจำนวนกลุ่มให้มากขึ้น หรือเพิ่มผู้เชี่ยวชาญในแต่ละกลุ่มให้มากขึ้นก็ได้ เนื่องจากการเพิ่มจำนวนกลุ่มให้มากขึ้นหรือการเพิ่มจำนวนผู้เชี่ยวชาญในแต่ละกลุ่มให้มากขึ้นไม่ได้ทำให้ค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์แตกต่างกัน

2.2.2.3 องค์ประกอบความแปรปรวนส่วนที่เหลือ (*pr:g*, *e*) ส่งผลต่อความแตกต่างกันของคะแนนจุดตัดที่กำหนดโดยทุกวิธีค่อนข้างมาก ซึ่งสอดคล้องกับงานวิจัยของ Lee และ Lewis (2008) และ Yin และ Sconing (2008) ที่ว่าคะแนนจุดตัดที่ต่างกันเป็นผลมาจากองค์ประกอบความแปรปรวนส่วนที่เหลือ กล่าวคือ ยังมีองค์ประกอบความแปรปรวนอื่น ๆ ที่ไม่ได้ทำการศึกษาซึ่งส่งผลต่อความต่างกันของคะแนนจุดตัด

2.2.2 ผลจากการวิเคราะห์ความเที่ยงภายในผู้เชี่ยวชาญ ผู้วิจัยมีประเด็นที่จะอภิปรายดังนี้

ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดด้วยวิธีบุ๊คมาร์ค กำหนดคะแนนจุดตัดในแต่ละรอบได้สอดคล้องกันมากกว่าผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดด้วยวิธีแองกอฟแบบใช่/ไม่ใช่ และวิธีเอบีซี แสดงว่าผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดด้วยวิธีบุ๊คมาร์คในแต่ละรอบมีความคงเส้นคงวาสูงกว่าวิธีแองกอฟแบบใช่/ไม่ใช่ และวิธีเอบีซี ทั้งนี้อาจเนื่องมาจากว่าการกำหนดคะแนนจุดตัดในรอบที่ 2 และ 3 ของวิธีบุ๊คมาร์คทำได้โดยการเปลี่ยนตำแหน่งที่คั่นหนังสือที่เคยกำหนดไว้ในรอบก่อน ๆ ซึ่งทำได้ง่ายโดยไม่ต้องพิจารณาข้อสอบทุกข้อ เพียงแค่พิจารณาข้อสอบไม่กี่ข้อที่อยู่ก่อนและหลังที่คั่นหนังสือเดิมเท่านั้น ส่วนการกำหนดคะแนนจุดตัดในรอบที่ 2 และ 3 ของวิธีแองกอฟแบบใช่/ไม่ใช่ และวิธีเอบีซีนั้น ผู้เชี่ยวชาญจะต้องพิจารณาข้อสอบทุกข้อเพื่อกำหนดคะแนนจุดตัดแต่ละรอบ ดังนั้นผลจากการอภิปรายร่วมกันระหว่างผู้เชี่ยวชาญและสารสนเทศที่ให้เพิ่มเติมล้วนสามารถส่งผลต่อการตัดสินใจในทุก ๆ ข้อ จึงอาจทำให้คะแนนจุดตัดที่กำหนดในแต่ละรอบต่างกันได้

ผลจากการศึกษาครั้งนี้สอดคล้องกับ Yin และ Sconing (2008) ที่พบว่า ผู้เชี่ยวชาญที่กำหนดคะแนนจุดตัดด้วยวิธี Item rating มีความคงเส้นคงวาของคะแนนจุดตัดในแต่ละรอบน้อย ทั้งนี้วิธี Item rating และวิธีแองกอฟแบบใช่/ไม่ใช่ เป็นวิธีซึ่งปรับปรุงมาจากวิธีแองกอฟ โดยทั้งสองวิธีมีหลักการกำหนดคะแนนจุดตัดในรอบที่ 2 และ 3 โดยการให้ผู้เชี่ยวชาญพิจารณาข้อสอบทุกข้อเหมือนกัน จากที่กล่าวมาพอจะสรุปได้ว่าการพิจารณาเพื่อกำหนดคะแนนจุดตัดตามแนวคิดของวิธีบุ๊คมาร์คที่ให้การกำหนดคะแนนจุดตัดในรอบที่ 2 และ 3 ทำได้ง่ายโดยไม่ต้องพิจารณาข้อสอบทุกข้อ ทำให้ผลการกำหนดคะแนนจุดตัดของผู้เชี่ยวชาญในแต่ละรอบมีความใกล้เคียงกันมากกว่าการให้ผู้เชี่ยวชาญต้องพิจารณาข้อสอบทุกข้อเพื่อกำหนดคะแนนจุดตัดแต่ละรอบ

ข้อเสนอแนะ

ข้อเสนอแนะในการนำผลวิจัยไปใช้

1. การกำหนดคะแนนจุดตัดให้มีค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์เฉลี่ยตั้งแต่ 0.70 ขึ้นไปนั้น สำหรับวิธีบุ๊คมาร์คควรกำหนดคะแนนจุดตัดด้วยผู้เชี่ยวชาญจำนวนตั้งแต่ 49 คนขึ้นไป โดยแบ่งเป็น 7 กลุ่ม กลุ่มละ 7 คน วิธีแองกอฟแบบใช่/ไม่ใช่ควรกำหนดคะแนนจุดตัดด้วยผู้เชี่ยวชาญจำนวนตั้งแต่ 36 คนขึ้นไป โดยแบ่งเป็น 6 กลุ่ม กลุ่มละ 6 คน สำหรับวิธีเอบีซีควรกำหนดด้วยผู้เชี่ยวชาญจำนวน 25 คนขึ้นไป โดยแบ่งเป็น 5 กลุ่ม กลุ่มละ 5 คน ซึ่งวิธีเอบีซีเป็นวิธีที่อาศัยผู้เชี่ยวชาญจำนวนน้อยที่สุดเมื่อเปรียบเทียบกับวิธีแองกอฟแบบใช่/ไม่ใช่และวิธีบุ๊คมาร์ค

2. การเพิ่มค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์ของคะแนนจุดตัดนั้น ทั้งวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซี สามารถทำได้โดยการเพิ่มจำนวนผู้เชี่ยวชาญให้มากขึ้น ในการเพิ่มจำนวนผู้เชี่ยวชาญนั้น สำหรับวิธีแองกอฟแบบใช่/ไม่ใช่ควรเพิ่มจำนวนกลุ่มให้มากขึ้นดีกว่าเพิ่มจำนวนผู้เชี่ยวชาญในแต่ละกลุ่มให้มากขึ้น ส่วนวิธีบุ๊คมาร์ค และวิธีเอบีซีควรเพิ่มจำนวนผู้เชี่ยวชาญในแต่ละกลุ่มให้มากขึ้นจะดีกว่า เนื่องจากการเพิ่มจำนวนกลุ่มให้มากขึ้นหรือเพิ่มผู้เชี่ยวชาญในแต่ละกลุ่มให้มากขึ้นไม่ได้ทำให้ค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์แตกต่างกัน จากประสบการณ์ที่ผู้วิจัยได้ดำเนินการเพื่อกำหนดคะแนนจุดตัดมา พบว่า การเพิ่มจำนวนผู้เชี่ยวชาญในแต่ละกลุ่มทำได้ง่ายและสะดวกกว่าการเพิ่มจำนวนกลุ่มผู้เชี่ยวชาญ เนื่องจากหากเพิ่มจำนวนกลุ่มให้มากขึ้นจะต้องเพิ่มผู้อำนวยความสะดวกและผู้ดำเนินรายการประจำกลุ่มย่อยมากขึ้น ซึ่งทำให้เสียค่าใช้จ่ายในการดำเนินการมากขึ้นด้วย

3. การกำหนดคะแนนจุดตัดด้วยวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซี ควรเปิดโอกาสให้ผู้เชี่ยวชาญได้อภิปรายร่วมกัน เพื่อให้ผู้เชี่ยวชาญให้แลกเปลี่ยนแนวคิด และหลักการในการกำหนดคะแนนจุดตัดร่วมกันซึ่งจะทำให้คะแนนจุดตัดของผู้เชี่ยวชาญแต่ละท่านใกล้เคียงกันมากขึ้น ส่วนในแต่ละรอบของการกำหนดคะแนนจุดตัดอาจจะไม่จำเป็นต้องรายงานผลการกำหนดคะแนนจุดตัดของผู้เชี่ยวชาญแต่ละท่านที่กำหนดขึ้นในรอบก่อน ๆ ให้ผู้เชี่ยวชาญท่านอื่น ๆ ทราบก็ได้ เนื่องจากเป็นข้อมูลที่ไม่มีประโยชน์ต่อการกำหนดคะแนนจุดตัดของผู้เชี่ยวชาญมากนัก

4. เนื่องจากคะแนนจุดตัดที่กำหนดโดยวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซีมีความแตกต่างกัน ดังนั้นสำหรับผู้ที่ต้องการนำผลการกำหนดคะแนนจุดตัดไปใช้สามารถดำเนินการได้ 3 แนวทาง โดยแนวทางแรกให้ผู้ที่ต้องการใช้คะแนนจุดตัดพิจารณาเลือกคะแนนจุดตัดที่มีคุณภาพสูง ซึ่งคุณภาพของคะแนนจุดตัดสามารถพิจารณาได้จากคุณภาพด้านความตรงและความเที่ยง ซึ่งผู้วิจัยได้นำเสนอคุณภาพของคะแนนจุดตัดไว้แล้ว แนวทางที่สอง คือการเฉลี่ย

คะแนนจุดตัดจากทุก ๆ วิธีที่กำหนดไว้ เพื่อเป็นตัวแทนของทุกวิธี และแนวทางที่สาม คือการเชิญผู้เชี่ยวชาญที่ไม่ได้เข้าร่วมกำหนดคะแนนจุดตัดครั้งนี้ มาร่วมกันพิจารณาคะแนนจุดตัดทั้งหมดอีกครั้ง และให้ผู้เชี่ยวชาญร่วมกันพิจารณาคะแนนจุดตัดที่ได้กำหนดไว้ รวมทั้งพิจารณาคุณภาพของคะแนนจุดตัด จากนั้นให้ผู้เชี่ยวชาญร่วมกันตัดสินใจเลือกคะแนนจุดตัดที่เหมาะสมที่สุด หรืออาจกำหนดคะแนนจุดตัดอีกครั้งก็ได้

5. การเลือกวิธีการกำหนดคะแนนจุดตัดควรเลือกให้เหมาะสมกับบริบทที่จะใช้ โดยวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซี เหมาะที่จะใช้ในบริบทต่อไปนี้

ตารางที่ 5.1 บริบทที่เหมาะสมกับวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซี

บริบท	วิธีแองกอฟแบบใช่/ไม่ใช่	วิธีบุ๊คมาร์ค	วิธีเอบีซี
จำนวนข้อสอบ	น้อย	มาก	น้อยหรือมากก็ได้
จำนวนสาระที่วัด	สาระเดียว	สาระเดียว	หลายสาระ
ลักษณะการให้คะแนน	ให้คะแนนแบบ 2 ค่า หรือให้คะแนนมากกว่า 2 ค่า	ให้คะแนนแบบ 2 ค่า หรือให้คะแนนมากกว่า 2 ค่า	ให้คะแนนแบบ 2 ค่า หรือให้คะแนนมากกว่า 2 ค่า
ความรู้ทางสถิติ	ค่าเฉลี่ย	IRT	IRT Cluster analysis
ระยะเวลา	มาก	น้อย	ปานกลาง

นอกจากนี้ข้อจำกัดหรือจุดด้อยของแต่ละวิธีก็มีส่วนสำคัญต่อการตัดสินใจเลือกวิธีการกำหนดคะแนนจุดตัด โดยวิธีเอบีซีที่ผู้วิจัยพัฒนาขึ้นยังมีจุดด้อยคือต้องใช้เวลาในการเตรียมการเพื่อกำหนดคะแนนจุดตัดมาก เนื่องจากวิธีเอบีซีมีสารสนเทศที่เสนอต่อผู้เชี่ยวชาญมาก อาทิ คู่มือเรียงข้อสอบรายสาระการเรียนรู้ และผลการวิเคราะห์กลุ่ม เป็นต้น นอกจากนี้ยังต้องอาศัยการวิเคราะห์ข้อมูลด้วยสถิติขั้นสูง ดังนั้นในการเตรียมเอกสารเหล่านี้และการวิเคราะห์ข้อมูลจึงต้องอาศัยเวลาที่มากพอสมควร นอกจากนี้วิธีเอบีซียังมีจุดด้อยอยู่ที่ระยะเวลาในการกำหนดคะแนนจุดตัดแต่ละรอบที่ต้องใช้เวลามากพอสมควร

ข้อเสนอแนะสำหรับการวิจัยครั้งต่อไป

1. เนื่องจากองค์ประกอบความแปรปรวนส่วนที่เหลือ (*pr:g*, *e*) ส่งผลต่อความแตกต่างกันของคะแนนจุดตัดที่กำหนดด้วยวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซีค่อนข้างมาก กล่าวคือ ยังมีองค์ประกอบความแปรปรวนอื่น ๆ ที่ไม่ได้ทำการศึกษาซึ่งส่งผลต่อความต่างกันของคะแนนจุดตัด ดังนั้นควรมีการศึกษาเพิ่มเติมเกี่ยวกับองค์ประกอบความแปรปรวนอื่น ๆ ที่น่าจะส่งผลต่อความต่างกันของคะแนนจุดตัด อาทิ จำนวนคะแนนจุดตัดที่กำหนด จำนวนแหล่งของผู้เชี่ยวชาญ เป็นต้น

2. จากผลการวิจัยที่พบว่าผู้เชี่ยวชาญส่วนใหญ่จะกำหนดคะแนนจุดตัดโดยพิจารณาระดับพฤติกรรมด้านพุทธิพิสัยที่ข้อสอบต้องการวัด นอกจากนี้ผู้เชี่ยวชาญส่วนใหญ่เห็นด้วยว่าการกำหนดคะแนนจุดตัดจากข้อสอบที่มีการจัดไว้เป็นกลุ่มตามระดับพฤติกรรมด้านพุทธิพิสัยน่าจะช่วยให้การกำหนดคะแนนจุดตัดทำได้เร็วขึ้น จึงควรมีการศึกษาว่าหากให้ผู้เชี่ยวชาญกำหนดคะแนนจุดตัดโดยพิจารณาจากข้อสอบที่แบ่งเป็นฉบับย่อย ๆ ตามสาระการเรียนรู้และแบ่งกลุ่มข้อสอบในแต่ละฉบับย่อยตามระดับพฤติกรรมด้านพุทธิพิสัยที่วัดแล้วเรียงข้อสอบตามความยาก จะทำให้การกำหนดคะแนนจุดตัดทำได้ง่ายและเร็วขึ้นกว่าวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซีหรือไม่ และควรศึกษาว่าคุณภาพของคะแนนจุดตัดเป็นอย่างไรด้วย เพื่อให้ได้ข้อมูลในการพัฒนาวิธีการกำหนดคะแนนจุดตัดที่มีประสิทธิภาพต่อไป

3. การศึกษาวิจัยครั้งนี้ผู้วิจัยดำเนินการเก็บข้อมูลเพื่อตรวจสอบความตรงของกระบวนการตัดสินใจของผู้เชี่ยวชาญด้วยการสังเกตพฤติกรรม การใช้เทคนิคการคิดออกเสียง การบันทึกข้อสนทนา และการสนทนากลุ่ม จากการดำเนินการ พบว่า นอกจากข้อมูลที่ได้จะมีประโยชน์ต่อการตรวจสอบความตรงของคะแนนจุดตัดแล้ว ยังมีประโยชน์ต่อการดำเนินการแต่ละขั้นตอนด้วย กล่าวคือ ข้อมูลที่เก็บรวบรวมมาเป็นข้อมูลเกี่ยวกับความคิด ความกังวลขณะตัดสินใจ และปัญหาที่เกิดขึ้นในระหว่างที่ผู้เชี่ยวชาญกำหนดคะแนนจุดตัด ซึ่งข้อมูลเหล่านี้ช่วยให้ผู้วิจัยเห็นปัญหา จึงทำให้สามารถดำเนินการแก้ไขปัญหาที่เกิดขึ้นได้อย่างทันท่วงที เพื่อไม่ให้ปัญหาเหล่านั้นส่งผลกระทบต่อคุณภาพของคะแนนจุดตัด ดังนั้นสำหรับผู้ที่จะจัดการกำหนดคะแนนจุดตัดควรให้ความสำคัญกับการเก็บรวบรวมข้อมูลเกี่ยวกับความคิด ความกังวลขณะตัดสินใจ และปัญหาที่เกิด

ขึ้นกับผู้เชี่ยวชาญ เพื่อใช้เป็นข้อมูลในการปรับปรุงวิธีการในระหว่างที่ดำเนินการกำหนดคะแนนจุดตัดให้มีคุณภาพมากขึ้น

4. การวิจัยครั้งนี้เป็นการกำหนดคะแนนจุดตัดสำหรับการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน ช่วงชั้นที่ 4 (มัธยมศึกษาปีที่ 6) ปีการศึกษา 2553 วิชาคณิตศาสตร์และวิชาวิทยาศาสตร์ ซึ่งยังมีอีกหลายวิชาที่ยังไม่ได้กำหนดคะแนนจุดตัด จึงควรมีการศึกษาในวิชาอื่น ๆ ด้วยเพื่อเป็นแนวทางสำหรับการกำหนดคะแนนจุดตัดสำหรับหน่วยงานที่เกี่ยวข้องกับการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐานต่อไป

5. การวิจัยครั้งนี้เป็นการกำหนดคะแนนจุดตัดโดยผู้เชี่ยวชาญที่เป็นครูผู้สอนในระดับมัธยมศึกษาตอนปลายในวิชาคณิตศาสตร์ วิชาฟิสิกส์ วิชาเคมี และวิชาชีววิทยา ซึ่งคะแนนจุดตัดที่ได้ก็เป็นผลจากการตัดสินของบุคคลกลุ่มนี้เท่านั้น จึงควรมีการวิจัยเพื่อกำหนดคะแนนจุดตัดโดยผู้มีส่วนได้ส่วนเสียกลุ่มอื่น ๆ ด้วย อาทิ นักเรียน อาจารย์ผู้สอนในระดับมหาวิทยาลัย เป็นต้น ซึ่งจะทำให้ได้ข้อมูลเกี่ยวกับคะแนนจุดตัดที่กำหนดโดยกลุ่มต่างๆ เพื่อนำไปสู่การพิจารณากำหนดคะแนนจุดตัดที่เหมาะสมต่อไป

รายการอ้างอิง

ภาษาไทย

กัลยา วานิชบัญชา. 2550. *การใช้ SPSS for Windows ในการวิเคราะห์ข้อมูล.* พิมพ์ครั้งที่ 10. กรุงเทพฯ: โรงพิมพ์จุฬาลงกรณ์มหาวิทยาลัย.

ศิริชัย กาญจนวาสี. 2548. *ทฤษฎีการทดสอบแบบดั้งเดิม.* พิมพ์ครั้งที่ 5. กรุงเทพฯ: โรงพิมพ์จุฬาลงกรณ์มหาวิทยาลัย.

ศิริชัย กาญจนวาสี. 2550. *ทฤษฎีการทดสอบแนวใหม่.* พิมพ์ครั้งที่ 3. กรุงเทพฯ: โรงพิมพ์จุฬาลงกรณ์มหาวิทยาลัย.

สุรีพร อนุศาสนนันท์. 2550. *การเปรียบเทียบคุณภาพของการกำหนดมาตรฐานระหว่างวิธีแองกอฟที่ได้รับการปรับปรุงกับวิธีบุ๊คมาร์ค.* วิทยานิพนธ์ปริญญาดุษฎีบัณฑิต สาขาวิชาการวัดและประเมินผลทางการศึกษา ภาควิชาวิจัยและจิตวิทยาการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

สถาบันทดสอบทางการศึกษาแห่งชาติ. 2550. *คู่มือสำหรับผู้บริหารในการนำผลการสอบ O-NET ไปวางแผนปรับปรุงการจัดการเรียนการสอนและพัฒนาครู-นักเรียน.* สถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน).

ภาษาอังกฤษ

Berk, R. A. 1980. *Criterion-referenced testing: State of the art*. Baltimore: John HopKins University Press.

Berk, R. A. 1986. A consumer's guide to setting performance standard on criterion-referenced tests. *Review of Educational Research*, 56(1): 137-172.

Berk, R. A. 1996. Standard setting: The next generation (Where few psychometricians have gone before!). *Applied Measurement in Education,* 9(3): 215-235.

Beretvas, S. N. 2004. Comparison of bookmark difficulty locations under different item response models. *Applied Psychological Measurement,* 28(1): 25-47.

Bejar, I. I. 2008. *Standard setting: What is it? Why is it important?* [Online]. 2008. Available from : http://www.ets.org/Media/Research/pdf/RD_Connections7.pdf [2008, April 5]

Branch, J. L. 2000. Investigating the information-seeking processes of adolescents: The value of using think alouds and think afters. *Library & Information Science Research*, 22(4), 371-392.

Brandon, P. R. 2004. Conclusions about frequently studied modified angoff standard-setting topics. *Applied Measurement in Education,* 17(1):59-88.

Brennan, R. L. 2001. *Generalizability theory*. lowa: Springer-Verlag New York.

Brennan, R. L. 2006. *Educational measurement (4th ed.).* Westport, CT: American Council on Education/Praeger.

Bontempo, B. D., Marks C. M. & Karabatsos G.1998. A meta-analytic assessment of empirical differences in standard setting procedures. Paper presented at the annual meeting of the American Educational Research Association, San Diego.

Boursicot, K. .A., Roberts, T. E. & Pell G. 2006. Standard setting for clinical competence at graduation from medical school: a comparison of passing scores across five medical schools. *Advance in Health Science Education,* 11(2): 173-183.

Bowen, C. W. 1994. Think-aloud methods in chemistry education: understanding student thinking. *Journal of Chemical Education*, 71(3), 183-190.

Bowers, J. J. & Shindoll, R. R. 1989. A comparison of the Angoff, Beuk and Hofstee methods for setting a passing score. *ACT Research Report Series 89-2*. Iowa City, Iowa: ACT; 1–16.

Busch, J. C. & Jaeger, R. M. 1990. Influence of type of judge, normative information, and discussion on standards recommended for the National Teacher Examinations. *Journal of Educational Measurement,* 27: 145-163.

Buckendahl, C. W., Smith, R. W., Impara, J. *C.* & Plak, B. S. 2002. A comparison of angoff and bookmark standard setting methods. *Journal of Educational Measuremen,* 39(3): 253-263.

Clauser, B. E., Harik, P., Margolis, M. J., McManus, I. C., Mollon, J., Chis, L. & Williams, S. 2009. An empirical examination of the impact of group discussion and examinee performance information on judgments made in the angoff standard-setting procedure. *Applied Measurement in Education*, 22: 1-21.

Cizek, G. J. 2005. Adapting testing technology to serve accountability aims: The case of vertically moderated standard setting. *Applied Measurement in Education*, 18 (1): 1-9.

Cizek, G. J., Bunch, M. B. & Koons, H. 2004. Setting performance standards: Contemporary methods. *Educational Measurement: Issues and Practice*, 23(4): 31-50.

Cizek, G. J. & Bunch, M. B. 2007. *Standard setting: A guide to establishing and evaluating performance standards on tests.* Thousand Oaks, CA: Sage.

Charters, E. 2003. The use of think-aloud methods in qualitative research: An introduction to think-aloud methods. *Brock Education*, 12(2), 68-82.

Chinn, R. N. & Hertz, N. R. 2002. Alternative approaches to standard setting for licensing and certification examinations. *Applied Measurement in Education*, 15: 1-14.

Downing S, Tekian A,Yudkowsky R. 2006. Procedures for establishing defensible absolute passing scores on performance examinations in health professions education. *Teaching and Learning in Medicine,* 18(1):50-7.

Fehrmann, M. L. et al. 1991. An empirical comparison of cutoff score methods for content-related and criterion-related validity settings. *Educational and Psychological Measurement,* 51: 1029-1039.

Ferdous, A. A. & Plake, B. S. 2005. The use of subsets of test questions in an angoff standard - setting method. *Educational and Psychological Measurement,* 65(2): 185-201.

Ferdous, A. A. & Plake, B. S. 2005. Understanding the factors that influence decisions of panelists in a standard-setting study. *Applied Measurement in Education,* 18(3): 257-267.

Ferdous, A. A. & Plake, B. S. 2007. Item selection strategy for reducing the number of items rated in an angoff standard setting study. *Educational and Psychological Measurement*, 67 (2): 193-206.

Ferdous, A. A. & Plake, B. S. 2008. Item response theory–based approaches for computing minimum passing scores from an angoff-based standard - setting study. *Educational and Psychological Measurement,* 68(5): 778-796.

Fowell, L. et al. 2008. Estimating the minimum number of judges required for test-centred standard setting on written assessments: Do discussion and iteration have an influence? *Advances in Health Sciences Education,* 13: 11-24.

Giraud, G., Impara, J. C. & Plake, B. S. 2005. Teachers' conceptions of the target examinee in angoff standard setting. *Applied Measurement in Education,* 18(3): 223-232.

Glass, G. 1978. Standards and criteria. *Journal of Educational Measurement*, 15(4), 237–261.

Gomez, P. G. 2007. Proficiency descriptors based on a scale-anchoring study of the new TOEFL iBT reading test. *Language Testing, 24*: 417-444.

Green, B. F. 1995. *Setting performance standard: content, goals, individual differences*. Princeton, NJ: Educational Testing Service.

Green, D. R; Trimble, C. S. & Lewis, D. *M* . 2003. Interpreting the results of three different standard- setting procedures. *Educational Measurement: Issues and Practice,* 22 (1), 22-32.

Halpin, G. 1987. An analysis of the reliability and validity of procedures for setting minimum competency standards. *Educational and Psychological Measurement,* 47(4): 977-983.

Hambleton, R. K. & Plake, B. S. 1995. Using an extended angoff procedure to set standards on complex performance assessments. *Applied Measurement in Education,* 8: 1-55.

Hambleton, R. K., Jaeger, R. M., Plake, B. S. & Mills, C. 2000. Setting performance standards on complex educational assessments. *Applied Psychological Measurement,* 24: 355-366.

Hess, B., Subhiyah, R. G.& Giordano, C. 2007. Convergence between cluster analysis and the angoff method for setting minimum passing scores on credentialing examinations. *Evaluation & the Health Professions,* 30(4): 362-375.

Hurtz, G.M. & Hertz, N.R. 1999. How many raters should be used for establishing cutoff scores with the Angoff Method? A generalizability theory study. *Educational and Psychological Measurement ,*59: 885–897.

Hurtz, G. M., Jones, J. P. & Jones, C. N. 2008 .Conversion of proportion-correct standard-setting judgments to cutoff scores on the item response theory θ scale. *Applied Psychological Measurement,* 32 (5): 385-406.

Impara, J. C. & Plake, B. S. 1997. Standard setting: An alternative approach. *Journal of Educational Measurement,* 34: 353-366.

Johnstone, C. J., Bottsford-Miller, N. A. & Thompson, S. J. 2006. *Using the think aloud method (cognitive labs) to evaulate test design for students with disabilities and English language learners* (Technical Report 44). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes [Online]. 2011. Available from: http://education.umn.edu/NCEO/OnlinePubs/Tech44/ [2011, January 10]

Kane, M. 1994. Validating the performance standards associated with passing score. *Review of Educational Research,* 64(3): 425-461.

Kaftandjieva, F. Methods for Setting Cut Scores in Criterion-referenced Achievement Tests: A comparative analysis of six recent methods with an application to tests of reading in EFL. European Association for Language Testing and Assessment [Online]. 2010. Available from: http://www.ealta.eu.org/documents/resources/FK_second_doctorate.pdf [2012, January 10]

Karantonis, A. & Sireci, S. G. 2006. The bookmark standard-setting method: A literature review. *Educational Measurement: Issues and Practice,* 25(1): 4-12.

Katalin, E. 2000. "Please, keep talking": The 'think-aloud' method in second language reading research. *Feature Articles*. 7(3) [Online]. 2011. Available from: http://ludens.elte.hu/~deal/pages/novelty/htm2/vol73/elekes.htm [2011, January 10]

Kurpuis, S. E. R. 2006. *Testing and measurement: a user-friendly guide*. California: Sage Publications.

Landis J. R. & Koch G.G., 1977. The measurement of observer agreement for categorical data. *Biometrics*, 33: 159-74.

Lee, W. C. 2008. *Program MULT-CLASS (version 3) for multinomial and compound-multinomial classification consistency and accuracy*. Iowa City, IA: Center for Advanced Studies in Measurment and Assessment, University of Lowa.

Lee, G. & Lewis, D. M. 2008. A generalizability theory approach to standard error estimates for bookmark standard settings. *Educational and Psychological Measurement,* 68 (4): 603-620.

Lewis, D. M. 2006. *Standard setting for grades 3-8 mathematics and english language arts* [Online]. 2008. Available from: http://www.nysut.org/files/k12_071210_assessments03_standards.pdf [2008, April 11]

Lissitz, R. W. & Kroopnick, M. H. 2008. *Running ahead: An adaptive procedure for standard setting* [Online]. 2008. Available from: http://www.education.umd.edu/EDMS/MARCES/presentation/Lissitz_KroopnickFNLNCME07.pdf [2008, April 11]

Lin, J. 2008. *The bookmark standard setting procedure: Strengths and weaknesses* [Online]. 2008. Available from: http://www.education.ualberta.ca/educ/psych/crame/files/standard_setting.pdf [2008, April 11]

Linn, R. L. 1989. *Educational measurement (3rd edition).* American Council on Education. Macmillan Publishing Company, New York, NY.

Livingston, S. & Zieky, M. 1982. *Passing scores: A manual for setting standards of performance on educational and occupational tests.* Princeton, NJ: Educational Testing Service.

MacCann, R. G. & Stanley G. 2006. The use of rash modeling to improve standard setting. *Practical Assessment, Research & Evaluation*, 11: 1-17.

MacCann, R .G. 2008. A modification to angoff and bookmarking cutscores to account for the imperfect reliability of test scores. *Educational and Psychological Measurement*, 68: 197-214.

Marion, S. & Pellegrino, J. 2006. A validity framework for evaluating the technical quality of alternate assessments. *Educational Measurement: Issues and Practice,* 25(4): 47-57.

McGinty, D. 2005. Illuminating the "black box" of standard setting: An exploratory qualitative study. *Applied Measurement in Education,* 18(3), 265-287.

Morgan, D. L. & Michalis, P. M. 2005. *Setting cut scores for college placement.* The College Board NewYork [Online]. 2011. Available from: http://professionals.collegeboard.com/profdownload/accuplacer-setting-cut-scores.pdf [2011, April 11]

Näsström, G. & Nyström, P. 2008. A comparison of two different methods for setting performance standards for a test with constructed-response items. *Practical Assessment, Research & Evaluation,* 13(9): 1-12.

Nellhaus, J. M. 2000. *States with NAEP like performance standard*. Retrieved 11/5/2008 from http://www.nagb.org/pubs/studentperfstandard.pdf.

Nijlen, D. V. & Janssen, R. 2008. Modeling judgments in the angoff and contrasting-groups method of standard setting. *Journal of Educational Measurement,* 45: 45-63.

Norcini, J. J., Shea, J. A. & Kanya, D. T. 1988. The effect of various factors on standard setting. *Journal of Educational Measurement,* 25: 57-65.

Norcini, J. J. 2003. Setting standards on educational tests. *Medical Education, 37*(5)*:*464-469.

Papageorgiou, S. 2010. *Setting Cut Scores on the Common European Framework of Reference for the Michigan English Test* [Online]. 2010. Available from:

http://www.lsa.umich.edu/UMICH/eli/Home/Test%20Programs/MET/Researchers/METStandardSetting.pdf [2010, April 02]

Perie, M. 2006. *Standard setting.*presentation at the inclusive assessment seminar. Denver CO / Alexandria, VA [Online]. 2008. Available from: http://www.naacpartners.org/products/workshops/presentations/16210.pdf [2008, April 11]

Perie, M. 2008. A guide to understanding and developing performance-level descriptors. *Educational Measurement: Issues and Practice,* 27(4) 15-29.

Plake, B. S. 2008. Standard setters: Stand up and take a stand! *Education Measurement: Issues and Practice,* 27(1): 3-9.

Reckase, M. D. 2006. Evaluating standard setting methods using error models proposed by Schulz. *Education Measurement: Issues and Practice,* 25(3): 14-17.

Sampat M. P., Whitman G. J., Stephens T. W., Broemeling L. D., Heger N. A., Bovik A. C., & Markey M. K. 2006. The reliability of measuring physical characteristics of speculated masses on mammography. *The British Journal of Radiology*, 79: 134–140.

Sireci, S. G. 1995, August. *Using cluster analysis to solve the problem of standard setting*. Paper presented at the annual meeting of the American Psychological Association, NewYork, NY.

Sireci, S. G.,& Robin, F. 1999. Using cluster analysis to facilitate standard setting. *Applied Measurement in Education,* 12(3): 301-325.

Skaggs G., Hein S. F. & Awuor R. 2007 . Setting passing scores on passage-based tests: A comparison of traditional and single-passage bookmark methods. *Applied Measurement in Education,* 20(4): 405-426.

Skorupski W. P., Hambleton R. K. 2005. What are panelists thinking when they participate in standard-setting studies? *Applied Measurement in Education,* 18(3), 233-256.

Towles-Reeves, E. 2008. *Standard setting approaches for alternate assessment*. National Alternate Assessment Center [Online]. 2008. Available from: http://www.ccsso.org/content/PDFs/190_Towles-Reeves.pdf [2008, April 11]

van der Linden, W. J. 1984. Some thoughts on the use of decision theory to set cutoff scores: Comment on de Gruijter and Hambleton. *Applied Psychological Measurement,* 8: 9-17.

van Someren, M. W., Barnard, Y. F. & Sandberg, J. A.C. 1994. *The think aloud method: A practical guide to modeling cognitive Processes*. Department of Social Science Informatics University of Amsterdam. Academic Press, London.

Viera A.J. & Garrett J.M. 2005.Understanding interobserver agreement: the kappa statistic. *Family Medicine,* 37:360-363.

Violato, C., Marini, A. & Lee, C. 2003. A validity study of expert judgment procedures for setting cutoff scores on high - stakes credentialing examinations using cluster analysis. *Evaluation & the Health Profession,* 26 (1): 59-72.

Whitting, M. S., López, J., Schley, E. & Fisher, K. 2000. *Using think aloud protocols to compare cognitive levels of students and professors in college classrooms*. Proceedings of the 27th Annual National Agricultural Education Research Conference [online]. 2011. Available from: http://aaae.okstate.edu/proceedings/2000/web/L2.pdf [2011, January 10]

Yin, P. & Sconing, J. 2008. Estimating standard errors of cut scores for item rating and mapmark procedures a generalizability theory approach. *Educational and Psychological Measurement,* 68: 25-41.

Yudkowsky, R. 2008. Simpler standards for local performance examinations: The yes/no angoff and whole-test ebel. *Teaching and Learning in Medicine,* 20(3): 212-217.

Zieky, M. J. & Perie, M. 2006. *A primer on setting cut scores on tests of educational achievement*. Princeton, NJ: Educational Testing Service.

Zieky, M. J., Perie, M. & Livingston, S. 2008. *Cutscores: A manual for setting standards of performance on educational and occupational tests*. Prinnceton, NJ: Educational Testing Service.

ภาคผนวก

ภาคผนวก ก

จดหมายขอความร่วมมือในการวิจัย

ที่ ศธ 0512.6(2771)/ 54- 2396

คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ถนนพญาไท กรุงเทพมหานคร 10330

เมษายน 2555

เรื่อง ขอเชิญเป็นผู้ทรงคุณวุฒิตรวจเครื่องมือวิจัย

เรียน ผู้ช่วยศาสตราจารย์ ธนวัฒน์ ธิติธนานันท์

สิ่งที่ส่งมาด้วย เครื่องมือที่ใช้ในการวิจัย

ด้วยนางศิริพันธ์ ดิยะวงศ์สุวรรณ นิสิตหลักสูตรครุศาสตรดุษฎีบัณฑิต สาขาวิชาการวัดและประเมินผลการศึกษา ภาควิชาวิจัยและจิตวิทยาการศึกษา อยู่ระหว่างการดำเนินงานวิจัยวิทยานิพนธ์เรื่อง "การพัฒนาวิธีการกำหนดคะแนนจุดตัดสำหรับการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน วิชาวิทยาศาสตร์และคณิตศาสตร์ของนักเรียนมัธยมศึกษาปีที่ 6" โดยมีผู้ช่วยศาสตราจารย์ ดร. ณัฏฐภรณ์ หลาวทอง และศาสตราจารย์ ดร.ศิริชัย กาญจนวาสี เป็นอาจารย์ที่ปรึกษา ในการนี้ใคร่ขอเชิญท่านเป็นผู้ทรงคุณวุฒิตรวจเครื่องมือวิจัย ทั้งนี้นิสิตผู้วิจัยจะได้ประสานงานในรายละเอียดต่อไป

จึงเรียนมาเพื่อขอความอนุเคราะห์จากท่านโปรดเป็นผู้ทรงคุณวุฒิดังกล่าวเพื่อประโยชน์ทางวิชาการต่อไป และขอขอบคุณมาในโอกาสนี้

ขอแสดงความนับถือ

(รองศาสตราจารย์ ดร.อาชัญญา รัตนอุบล)
รองคณบดี
ปฏิบัติการแทนคณบดี

งานหลักสูตรและการจัดการเรียนการสอน ฝ่ายวิชาการ
โทร. 0-2218-2681-82 ต่อ 612

ที่ ศธ 0512.6(2771)/ 54-2393

คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ถนนพญาไท กรุงเทพมหานคร 10330

เมษายน 2555

เรื่อง ขอเชิญเป็นผู้ทรงคุณวุฒิกำหนดคะแนนจุดตัด

เรียน

สิ่งที่ส่งมาด้วย เครื่องมือที่ใช้ในการวิจัย

ด้วยนางศิริพันธ์ ดิยะวงศ์สุวรรณ นิสิตหลักสูตรครุศาสตรดุษฎีบัณฑิต สาขาวิชาการวัดและประเมินผลการศึกษา ภาควิชาวิจัยและจิตวิทยาการศึกษา อยู่ระหว่างการดำเนินงานวิจัยวิทยานิพนธ์เรื่อง "การพัฒนาวิธีการกำหนดคะแนนจุดตัดสำหรับการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน วิชาวิทยาศาสตร์และคณิตศาสตร์ของนักเรียนมัธยมศึกษาปีที่ 6" โดยมีผู้ช่วยศาสตราจารย์ ดร. ณัฏฐภรณ์ หลาวทอง และศาสตราจารย์ ดร.ศิริชัย กาญจนวาสี เป็นอาจารย์ที่ปรึกษา ในการนี้ใคร่ขอเชิญท่านเป็นผู้ทรงคุณวุฒิกำหนดคะแนนจุดตัด ทั้งนี้นิสิตผู้วิจัยจะได้ประสานงานในรายละเอียดต่อไป

จึงเรียนมาเพื่อขอความอนุเคราะห์จากท่านโปรดเป็นผู้ทรงคุณวุฒิดังกล่าวเพื่อประโยชน์ทางวิชาการต่อไป และขอขอบคุณมาในโอกาสนี้

ขอแสดงความนับถือ

(รองศาสตราจารย์ ดร.อาชัญญา รัตนอุบล)
รองคณบดี
ปฏิบัติการแทนคณบดี

งานหลักสูตรและการจัดการเรียนการสอน ฝ่ายวิชาการ
โทร. 0-2218-2681-82 ต่อ 612

ผอ.

โรงเรียนเตรียมอุดมศึกษา
รับที่ 718
วันที่ 3 พ.ค. 54
เวลา 13.00 น.

ที่ ศธ 0512.6(2771)/54- 2167

คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ถนนพญาไท กรุงเทพมหานคร 10330

เมษายน 2554

เรื่อง ขอเชิญเป็นผู้ทรงคุณวุฒิเพื่อกำหนดจำนวนของระดับ ชื่อและนิยามเชิงนโยบายของระดับความสามารถ

เรียน คุณวิศรุต สนธิชัย

สิ่งที่ส่งมาด้วย เครื่องมือที่ใช้ในการวิจัย

ด้วยนางศิริพันธ์ ดิยะวงศ์สุวรรณ นิสิตหลักสูตรครุศาสตรดุษฎีบัณฑิต สาขาวิชาการวัดและประเมินผลการศึกษา ภาควิชาวิจัยและจิตวิทยาการศึกษา อยู่ระหว่างการดำเนินงานวิจัยวิทยานิพนธ์เรื่อง "การพัฒนาวิธีการกำหนดคะแนนจุดตัดสำหรับการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐานวิชาวิทยาศาสตร์และคณิตศาสตร์ของนักเรียนมัธยมศึกษาปีที่ 6" โดยมีผู้ช่วยศาสตราจารย์ ดร.ณัฏฐภรณ์ หลาวทอง และศาสตราจารย์ ดร.ศิริชัย กาญจนวาสี เป็นอาจารย์ที่ปรึกษา ในการนี้ใคร่ขอเชิญท่านเป็นผู้ทรงคุณวุฒิเพื่อกำหนดจำนวนของระดับ ชื่อและนิยามเชิงนโยบายของระดับความสามารถ ทั้งนี้นิสิตผู้วิจัยจะได้ประสานงานในรายละเอียดต่อไป

จึงเรียนมาเพื่อขอความอนุเคราะห์จากท่านโปรดเป็นผู้ทรงคุณวุฒิดังกล่าวเพื่อประโยชน์ทางวิชาการต่อไป และขอขอบคุณมาในโอกาสนี้

ขอแสดงความนับถือ

(รองศาสตราจารย์ ดร.อาชัญญา รัตนอุบล)
รองคณบดี
ปฏิบัติการแทนคณบดี

(นายวิศรุต สนธิชัย)
ผู้อำนวยการโรงเรียนเตรียมอุดมศึกษา

เรียน ผู้อำนวยการโรงเรียนเตรียมอุดมศึกษา
เพื่อโปรด
☑ 1. ทราบ
☑ 2. พิจารณา เห็นชอบให้ ผศ.ดร.ศิริชัย กาญจนวาสี เป็นผู้ทรงคุณวุฒิ ตามที่เสนอ
ตามที่เสนอ
ผู้อำนวยการกลุ่มวิชาระดับมัธยมศึกษา
3 / ต.ค. / 54

ฝ่ายวิชาการ หลักสูตรและการสอน
โทร. 0-2218-2681-82 ต่อ 612

ที่ ศธ 0512.6(2771)/54- 2166

คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ถนนพญาไท กรุงเทพมหานคร 10330

เมษายน 2554

เรื่อง ขอเชิญเป็นผู้ทรงคุณวุฒิเพื่อกำหนดจำนวนของระดับ ชื่อและนิยามเชิงนโยบายของระดับความสามารถ

เรียน คุณอ่องจิต เมธยะประกาส

สิ่งที่ส่งมาด้วย เครื่องมือที่ใช้ในการวิจัย

ด้วยนางศิริพันธ์ ติยะวงศ์สุวรรณ นิสิตหลักสูตรครุศาสตรดุษฎีบัณฑิต สาขาวิชาการวัดและประเมินผลการศึกษา ภาควิชาวิจัยและจิตวิทยาการศึกษา อยู่ระหว่างการดำเนินงานวิจัยวิทยานิพนธ์เรื่อง "การพัฒนาวิธีการกำหนดคะแนนจุดตัดสำหรับการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐานวิชาวิทยาศาสตร์และคณิตศาสตร์ของนักเรียนมัธยมศึกษาปีที่ 6" โดยมีผู้ช่วยศาสตราจารย์ ดร.ณัฏฐกรณ์ หลาวทอง และศาสตราจารย์ ดร.ศิริชัย กาญจนวาสี เป็นอาจารย์ที่ปรึกษา ในการนี้ใคร่ขอเชิญท่านเป็นผู้ทรงคุณวุฒิเพื่อกำหนดจำนวนของระดับ ชื่อและนิยามเชิงนโยบายของระดับความสามารถ ทั้งนี้นิสิตผู้วิจัยจะได้ประสานงานในรายละเอียดต่อไป

จึงเรียนมาเพื่อขอความอนุเคราะห์จากท่านโปรดเป็นผู้ทรงคุณวุฒิดังกล่าวเพื่อประโยชน์ทางวิชาการต่อไป และขอขอบคุณมาในโอกาสนี้

ขอแสดงความนับถือ

(รองศาสตราจารย์ ดร.อาชัญญา รัตนอุบล)
รองคณบดี
ปฏิบัติการแทนคณบดี

ฝ่ายวิชาการ หลักสูตรและการสอน
โทร. 0-2218-2681-82 ต่อ 612
ศิริพันธ์ 081-5485102 ภายใน 1 สาขา
pooky551@hotmail.com

ภาคผนวก ข

1. รายนามผู้บริหารระดับสูงของหน่วยงานที่รับผิดชอบการจัดการศึกษาขั้นพื้นฐาน
2. รายนามผู้เชี่ยวชาญพิจารณาร่างจำนวนของระดับ ชื่อและคำอธิบายระดับความสามารถ
3. รายนามผู้เชี่ยวชาญตรวจสอบคุณภาพเครื่องมือ
4. รายนามผู้เชี่ยวชาญเข้าร่วมการกำหนดคะแนนจุดตัด

รายนามผู้บริหารระดับสูงของหน่วยงานที่รับผิดชอบการจัดการศึกษาขั้นพื้นฐาน

1.	นางสุราง โพธิ์พฤกษาวงษ์	ที่ปรึกษาด้านวิจัยและประเมินผลการศึกษา สำนักงานเลขาธิการสภาการศึกษา
2.	นางอ่องจิต เมธยะประภาส	ที่ปรึกษาด้านนโยบายและแผน สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน (เป็นผู้แทนของนายชินภัทร ภูมิรัตน เลขาธิการคณะกรรมการการศึกษาขั้นพื้นฐาน)
3.	นางวาทินี ธีระตระกูล	ที่ปรึกษาด้านมาตรฐานการศึกษา สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน
4.	นางสาวไพรวัลย์ พิทักษ์สาลี	ผู้อำนวยการสำนักทดสอบทางการศึกษา คณะกรรมการการศึกษาขั้นพื้นฐาน
5.	นายวิศรุต สนธิชัย	ผู้อำนวยการโรงเรียนเตรียมอุดมศึกษา กรุงเทพมหานคร

รายนามผู้เชี่ยวชาญพิจารณาร่างจำนวนของระดับ ชื่อและคำอธิบายระดับความสามารถ

	ชื่อ-สกุล	ตำแหน่ง/ วิทยฐานะ	หน่วยงาน
1.	ธีรศักดิ์ เกิดทอง	ครูเชี่ยวชาญ	โรงเรียนอุตรดิตถ์ดรุณี
2.	ดิเรก ดาสี	ครูชำนาญการ	โรงเรียนอุตรดิตถ์ดรุณี
3.	จันทนา เกตุกอ	ครูชำนาญการพิเศษ	โรงเรียนอุตรดิตถ์ดรุณี
4.	ดาราวดี พุ่มจันทร์	ครูชำนาญการ	โรงเรียนอุตรดิตถ์ดรุณี
5.	สุธารัตน์ จุลญาติ	ครูชำนาญการ	โรงเรียนอุตรดิตถ์ดรุณี
6.	วิไลลักษณ์ วังวลสินธุ์	ครูชำนาญการ	โรงเรียนอุตรดิตถ์ดรุณี
7.	พชรพรรณ อมรนพรัตนกุล	ครูชำนาญการ	โรงเรียนอุตรดิตถ์ดรุณี
8.	สุรเชษฐ์ บุญรักษ์	อาจารย์ประจำสาขาคณิตศาสตร์	คณะครุศาสตร์ มหาวิทยาลัยราชภัฏอุตรดิตถ์
9.	พูนทรัพย์ โนราช	อาจารย์ประจำสาขาคณิตศาสตร์	คณะครุศาสตร์ มหาวิทยาลัยราชภัฏอุตรดิตถ์

รายนามผู้เชี่ยวชาญตรวจสอบคุณภาพเครื่องมือ

1. ผู้ช่วยศาสตราจารย์ ธนวัฒน์ ธิติธนานันท์ อาจารย์ประจำโปรแกรมวิชาวิจัยและประเมินผลการศึกษา มหาวิทยาลัยราชภัฏนครราชสีมา
2. ดร.วราพร เอราวรรณ์ อาจารย์ประจำภาควิชาวิจัยและพัฒนาการศึกษา มหาวิทยาลัยมหาสารคาม
3. ดร.อัญชลี ศรีกลชาญ อาจารย์ประจำภาควิชาสังคมวิทยา คณะสังคมศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ

รายนามผู้เชี่ยวชาญเข้าร่วมการกำหนดคะแนนจุดตัดวิชาคณิตศาสตร์

ณ ห้องประชุมคุรุเวทย์ อาคาร 17 มหาวิทยาลัยราชภัฏนครราชสีมา

วันที่ 15 พฤศจิกายน 2554 เวลา 08.00 – 16.00 น.

	ชื่อ-สกุล	ตำแหน่ง/ วิทยฐานะ	โรงเรียน	ประสบการณ์การสอน
วิธีแองกอฟแบบใช่/ไม่ใช่				
1.	นางสาวกัญชญาณิศร์ วิบูลธนังกูล	ครู	ปากช่อง	7
2.	นายสันติสุข คำมูล	ครู	ปากช่อง	2
3.	นายจินตน์ ฟองสันเทียะ	ครู คศ.3	ราชสีมาวิทยาลัย	29
4.	นางเบญจวรรณ ยศกลาง	ครู ชำนาญการพิเศษ	เมืองคง	33
5.	นางปราณี ธีระพิทยาตระกูล	ครู ชำนาญการ	เมืองคง	15
6.	นางสาวธีรดา บุรินทรสุชาติ	ครู คศ.3	ราชสีมาวิทยาลัย	35
7.	นางจารุวรรณ คนึงหาญ	ครู ชำนาญการ	ขามสะแกแสง	15
8.	นางสาวบัวริม กวนสำโรง	ครู ชำนาญการ	ขามสะแกแสง	13
9.	นางจงรักษ์ ขอบเขต	ครู ชำนาญการพิเศษ	ราชสีมาวิทยาลัย	32
วิธีบุ๊คมาร์ค				
10.	นางสาวปวีณา เดชขุนทด	ครู	ขามสะแกแสง	2
11.	นางสุวิชญา ดังไธสง	ครู	โนนสูงศรีธานี	13
12.	นายวิชิรวรรธน์ วัฒนสันติพงศ์	ครู	โนนสูงศรีธานี	2
13.	นางทิพพา แก้วหนองแสง	ครู ชำนาญการ	พิมายวิทยา	18
14.	นายศุภโชค มีวารา	ครู	ปากช่อง	7
15.	นายสมชาย คล่องชอบ	ครู	พิมายวิทยา	17
16.	นายชิษณุชา พระสว่าง	ครู ชำนาญการพิเศษ	สุรนารีวิทยา	22
17.	นางสาวสุภาวดี กิติวิศิษฐ์	ครู ชำนาญการพิเศษ	พิมายวิทยา	25
18.	นายนรินทร์ โอภาสพันธ์สิน	ครู ชำนาญการพิเศษ	บุญวัฒนา	34
วิธีเอบีซี				
19.	นางทัศนี ชื่นยง	ครู คศ.3	สุรนารีวิทยา	28
20.	นางจันทร์เพ็ญ แสงสว่าง	ครู	ปักธงชัยประชานิรมิตร	30
21.	นางสมพร ผายสระน้อย	ครู คศ.3	ปักธงชัยประชานิรมิตร	20
22.	นางศศิญา ประพิณ	ครู ชำนาญการพิเศษ	ปักธงชัยประชานิรมิตร	35
23.	นายไพฑูรย์ สมสอางค์	ครู คศ.3	ชุมพวงศึกษา	26
24.	นางชูศิลป์ นาราหนองแวง	ครู ชำนาญการพิเศษ	ชุมพวงศึกษา	32
25.	นายยุทธศาสตร์ พลนา	ครู ชำนาญการพิเศษ	บุญวัฒนา	25
26.	นางสาวปรารถนา คิดถูก	ครู	ชุมพวงศึกษา	3
27.	นางสาวศิริลักษณ์ กุโบลา	ครู คศ.3 ชำนาญการพิเศษ	สุรนารีวิทยา	37

รายนามผู้เชี่ยวชาญเข้าร่วมการกำหนดคะแนนจุดตัดวิชาวิทยาศาสตร์

ณ ห้องประชุมคุรุเวทย์ อาคาร 17 มหาวิทยาลัยราชภัฏนครราชสีมา

วันที่ 16 พฤศจิกายน 2554 เวลา 08.00 – 16.00 น.

	ชื่อ-สกุล	ตำแหน่ง/ วิทยฐานะ	โรงเรียน	ประสบการณ์การสอน
วิธีแองกอฟแบบใช่/ไม่ใช่				
1.	นายสมศักดิ์ ตุ่นไธสง	ครู คศ. 3	ราชสีมาวิทยาลัย	30
2.	นางปิ่นแก้ว สระแก้ว	ครู ชำนาญการพิเศษ	เมืองคง	24
3.	นางรัตนา ห้าวหาญ	ครู ชำนาญการพิเศษ	พิมายพิทยา	30
4.	นางสรินทิพย์ เพียงพานิช	ครู ชำนาญการพิเศษ	สุรนารีวิทยา	28
5.	นายวิโรจน์ ดวนสันเทียะ	ครูชำนาญการ	ขามสะแกแสง	12
6.	นางสาวอัญชลี ขันตี	ครูชำนาญการพิเศษ	ปักธงชัยประชานิรมิตร	18
7.	นายประกอบ ชิวพิมาย	ครู คศ. 3	พิมายพิทยา	33
8.	นางสาวโสภา อรุณใหม่	ครู	ชุมพวงศึกษา	7
9.	นางสาวยุพาพร ทัศน์ไพร	ครู	ชุมพวงศึกษา	4
วิธีบุ๊คมาร์ค				
10.	นางสุรางค์ พุฒเกลาง	ครู ชำนาญการพิเศษ	สุรนารีวิทยา	30
11.	นางสาวจิราภรณ์ บุญสงค์	ครูชำนาญการ	บุญวัฒนา	15
12.	นายวรวัฒน์ ใยระย้า	ครู	ชุมพวงศึกษา	4
13.	นางสาวณัฏกันย์ ดอกสันเทียะ	ครู คศ.2	ขามสะแกแสง	12
14.	นายคณิศร ท้าวนอก	ครู คศ.1	เมืองคง	10
15.	นางนราพร แป้นจำ	ครู	โนนสูงศรีธานี	14
16.	นายเทพวุธ ห้าวหาญ	ครู คศ.3	พิมายพิทยา	34
17.	นางจุฑามาศ ชูสกุล	ครูชำนาญการ	ปากช่อง	13
18.	นายสกล ดุริยศาสตร์	ครู ชำนาญการพิเศษ	สุรนารีวิทยา	25
วิธีเอบีซี				
19.	นางอัจฉรา โทแหล่ง	ครู ชำนาญการพิเศษ	ราชสีมาวิทยาลัย	17
20.	นายเกียรติศักดิ์ ลิ้มสุเวช	ชำนาญการ	ราชสีมาวิทยาลัย	25
21.	ภารดี พูนแก้ว	ครู	โนนสูงศรีธานี	9
22.	นายอมรวัฒน์ ชนะภักดิ์	ครู ชำนาญการ	ขามสะแกแสง	14
23.	นายแสงชัย การงาน	ครู	ปากช่อง	3
24.	นางสาวกิ่งกาญจน์ กลิ่นจันทร์	ครู	บุญวัฒนา	3
25.	นางนภัชนันท์ ภาคินศุภเศรษฐ์	ครูชำนาญการ	ปากช่อง	14
26.	นางพิมพ์ละม้าย สาสุวงษ์	ครู ชำนาญการพิเศษ	ปักธงชัยประชานิรมิตร	18
27.	นางดวงใจ บานศิริ	ครู คศ.3	ปักธงชัยประชานิรมิตร	20
28.	นางจันทร์เพ็ญ อุทุมทอง	ครู	โนนสูงศรีธานี	2
29.	นายกิตติพงษ์ เวียงสมุทร	ครู	บุญวัฒนา	2
30.	นางบุษราคัม บุญกลาง	ครู ชำนาญการพิเศษ	เมืองคง	20

ภาคผนวก ค

1. แบบสอบถามความคิดเห็นเกี่ยวกับจำนวนของระดับ ชื่อของระดับ และคำอธิบายระดับความสามารถ
2. แบบสอบถามความคิดเห็นเกี่ยวกับร่างคำอธิบายระดับความสามารถสำหรับการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน วิชาคณิตศาสตร์และวิทยาศาสตร์ ชั้นมัธยมศึกษาปีที่ 6
3. ตัวอย่างจำนวนของระดับ ชื่อของระดับ และคำอธิบายระดับความสามารถ ที่นิยมกำหนดในต่างประเทศ

แบบสอบถามความคิดเห็น

เกี่ยวกับจำนวนของระดับ ชื่อของระดับ และคำอธิบายระดับความสามารถ

คำชี้แจง

แบบสอบถามชุดนี้เป็นแบบสอบถามความคิดเห็นของผู้บริหารระดับสูงในหน่วยงานที่เกี่ยวกับการศึกษาขั้นพื้นฐานเกี่ยวกับจำนวนของระดับ ชื่อของระดับ และคำอธิบายระดับความสามารถทางคณิตศาสตร์และวิทยาศาสตร์ของนักเรียนมัธยมศึกษาปีที่ 6 ที่เหมาะสมและเป็นไปได้

คำจำกัดความ

ความเหมาะสม หมายถึง ความสอดคล้องของจำนวนของระดับ ชื่อของระดับ และคำอธิบายระดับความสามารถทางคณิตศาสตร์และวิทยาศาสตร์กับวัตถุประสงค์ของการกำหนดคะแนนจุดตัด ซึ่งการกำหนดคะแนนจุดตัดครั้งนี้มีวัตถุประสงค์เพื่อใช้ในการรายงานผลการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน (O-NET) วิชาวิทยาศาสตร์และวิชาคณิตศาสตร์ ของนักเรียนชั้นมัธยมศึกษาปีที่ 6

ความเป็นไปได้ หมายถึง การเป็นที่ยอมรับได้ของผู้ที่เกี่ยวข้อง ของจำนวนของระดับ ชื่อของระดับ และคำอธิบายระดับความสามารถทางคณิตศาสตร์และวิทยาศาสตร์

ประเด็นคำถาม

จำนวนของระดับ ชื่อของระดับ และคำอธิบายระดับความสามารถทางคณิตศาสตร์และวิทยาศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ที่เหมาะสมและเป็นไปได้ควรมีกี่ระดับ แต่ละระดับควรชื่อว่าอะไรและแต่ละระดับควรจะมีคำอธิบายระดับความสามารถว่าอย่างไร

.....

.....

.....

.....

.....

.....

แบบสอบถามความคิดเห็นเกี่ยวกับร่างคำอธิบายระดับความสามารถสำหรับการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน วิชาคณิตศาสตร์และวิทยาศาสตร์ ชั้นมัธยมศึกษาปีที่ 6

ชื่อผู้ทรงคุณวุฒิ..... ตำแหน่ง/วิทยฐานะ.....
โรงเรียน.....ประสบการณ์การสอน.....ปี

คำชี้แจง

จำนวนของระดับ ชื่อของแต่ละระดับ และคำอธิบายระดับความสามารถที่ท่านจะได้พิจารณาต่อไปนี้เป็นข้อมูลที่ผู้วิจัยสรุปจากการสอบถามความคิดเห็นของผู้บริหารในหน่วยงานที่รับผิดชอบการจัดการศึกษาขั้นพื้นฐาน เพื่อใช้สำหรับการกำหนดคะแนนจุดตัดสำหรับการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐานวิชาคณิตศาสตร์ และวิชาวิทยาศาสตร์ ของนักเรียนชั้นมัธยมศึกษาปีที่ 6

ขอให้ท่านพิจารณาจำนวนของระดับ ชื่อของระดับ และคำอธิบายระดับความสามารถที่ร่างขึ้น ว่าจำนวนของระดับ ชื่อของระดับ และคำอธิบายระดับความสามารถทางคณิตศาสตร์และวิทยาศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ที่กำหนดขึ้นมีความเหมาะสมและเป็นไปได้หรือไม่

ความเหมาะสม หมายถึง ความสอดคล้องของจำนวนของระดับ ชื่อของระดับ และคำอธิบายระดับความสามารถทางคณิตศาสตร์และวิทยาศาสตร์กับวัตถุประสงค์ของการกำหนดคะแนนจุดตัด ซึ่งการกำหนดคะแนนจุดตัดครั้งนี้มีวัตถุประสงค์เพื่อใช้ในการรายงานผลการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน (O-NET) วิชาวิทยาศาสตร์และวิชาคณิตศาสตร์ ของนักเรียนชั้นมัธยมศึกษาปีที่ 6

ความเป็นไปได้ หมายถึง การเป็นที่ยอมรับได้ของผู้ที่เกี่ยวข้อง ของจำนวนของระดับ ชื่อของระดับ และคำอธิบายระดับความสามารถทางคณิตศาสตร์และวิทยาศาสตร์

กรุณาให้ข้อเสนอแนะสำหรับการนำไปพัฒนาร่างจำนวนของระดับ ชื่อของระดับ และคำอธิบายระดับความสามารถให้มีความสมบูรณ์ยิ่งขึ้น

จำนวนของระดับ ชื่อของระดับ และคำอธิบายระดับความสามารถ
สำหรับการกำหนดคะแนนจุดตัดสำหรับการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน
วิชาคณิตศาสตร์ และวิชาวิทยาศาสตร์ ของนักเรียนชั้นมัธยมศึกษาปีที่ 6

ชื่อของระดับ	**คำอธิบายระดับความสามารถ**	**ข้อเสนอแนะเพื่อการแก้ไขปรับปรุง**
ระดับที่ 1	มีความรู้ ความสามารถทางวิชาการ และมีทักษะพื้นฐานในชั้นที่เรียนอยู่เล็กน้อย	
ระดับที่ 2	มีความรู้ ความสามารถทางวิชาการ และมีทักษะพื้นฐานในชั้นที่เรียนอยู่อย่างดี	
ระดับที่ 3	มีความรู้ ความสามารถทางวิชาการ และมีทักษะพื้นฐานในชั้นที่เรียนอยู่ดีเยี่ยม	
ระดับที่ 4	มีความรู้ ความสามารถทางวิชาการ และมีทักษะพื้นฐานในชั้นที่เรียนอยู่ดีเยี่ยม สามารถนำความรู้ที่มีไปประยุกต์ใช้ได้อย่างดี	

ตัวอย่างจำนวนของระดับ ชื่อของระดับ และคำอธิบายระดับความสามารถที่นิยมกำหนดในต่างประเทศ

รัฐ/หน่วยงาน	ระดับความสามารถ	ชื่อระดับความสามารถ	คำอธิบายระดับความสามารถ
The National Assessment Educational (Zieky, et. al., 2008)	3	ความสามารถระดับพื้นฐาน (Basic)	มีความรู้บางส่วนจากความรู้ทั้งหมดที่ควรรู้ และมีทักษะพื้นฐานสำหรับระดับชั้นที่เรียนอยู่
		ความสามารถระดับเชี่ยวชาญ (Proficient)	มีความรู้ความสามารถในเชิงวิชาการที่ควรรู้ทั้งหมดในระดับชั้นนั้น ๆ และแสดงความสามารถในสถานการณ์การทดสอบที่ท้าทายโดยสามารถนำความรู้และความสามารถที่มีมาใช้ในการวิเคราะห์สถานการณ์และแก้ปัญหาได้อย่างเหมาะสม
		ความสามารถระดับเชี่ยวชาญพิเศษ (Advance)	มีความสามารถยอดเยี่ยม
รัฐเพนซิวาเนีย (Pennsylvania) (Perie, 2008)	4	ความสามารถระดับเชี่ยวชาญพิเศษ (Advance)	มีความรู้ความสามารถในเชิงวิชาการยอดเยี่ยม คือมีความเข้าใจในเนื้อหาอย่างลึกซึ้ง และมีทักษะยอดเยี่ยม
		ความสามารถระดับเชี่ยวชาญ (Proficient)	มีความรู้ความสามารถในเชิงวิชาการเพียงพอ คือเข้าในเนื้อหาทั้งหมดและมีทักษะที่เหมาะสม
		ความสามารถระดับพื้นฐาน (Basic)	มีความรู้ความสามารถในเชิงวิชาการเล็กน้อย คือเข้าใจเนื้อหาเพียงบางส่วน และมีทักษะที่จำกัด ไม่สามารถปฏิบัติงานที่ได้รับมอบหมายได้สำเร็จ
		ความสามารถต่ำกว่าระดับพื้นฐาน (Below Basic)	มีความรู้ความสามารถในเชิงวิชาการไม่ดีพอ คือไม่เข้าใจในเนื้อหาและมีทักษะเพียงเล็กน้อย
รัฐ อัลบามา (Alabama) (Perie, 2008)	4	ระดับที่ 4 (Level IV)	มีความรู้ความสามารถในเชิงวิชาการเหนือกว่ามาตรฐาน
		ระดับที่ 3 (Level III)	มีความรู้ความสามารถในเชิงวิชาการตามมาตรฐาน
		ระดับที่ 2 (Level II)	มีความสามารถทางวิชาการถึงระดับมาตรฐานเพียงบางส่วน
		ระดับที่ 1 (Level I)	มีความสามารถทางวิชาการไม่ถึงระดับมาตรฐาน

รัฐ/หน่วยงาน	ระดับความสามารถ	ชื่อระดับความสามารถ	คำอธิบายระดับความสามารถ
รัฐอริโซนา (Arizona) (Perie, 2008)	4	เหนือกว่ามาตรฐาน (Exceeds the Standard)	มีความสามารถทางวิชาการระดับยอดเยี่ยม
		ถึงมาตรฐาน (Meets the Standard)	มีความรู้ความสามารถในเชิงวิชาการทั้งหมด สามารถประยุกต์ใช้ความรู้ในสถานการณ์จริง ได้
		เข้าใกล้มาตรฐาน (Approaches the Standard)	มีความเข้าใจในเนื้อหา และสามารถประยุกต์ ทักษะพื้นฐานต่างๆ ได้
		ต่ำกว่ามาตรฐานมาก (Falls Far Below the Standard)	มีความรู้และทักษะเบื้องต้นที่ต้องมี

ภาคผนวก ง

1. กำหนดการในการกำหนดคะแนนจุดตัด
2. คู่มือการกำหนดคะแนนจุดตัด
3. แบบกรอกผลการพิจารณาคะแนนจุดตัดด้วยวิธีแองกอฟแบบใช่/ไม่ใช่
4. แบบกรอกผลการพิจารณาคะแนนจุดตัดด้วยวิธีบุ๊คมาร์ค
5. แบบกรอกผลการพิจารณาคะแนนจุดตัดด้วยวิธีเอบีซี

คู่มือ

การกำหนดคะแนนจุดตัด

วิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซี

จัดทำโดย

นางศิริพันธ์ ติยะวงศ์สุวรรณ

อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก ผู้ช่วยศาสตราจารย์ ดร.ณัฏฐภรณ์ หลาวทอง

อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม ศาสตราจารย์ ดร. ศิริชัย กาญจนวาสี

ภาควิชาวิจัยและจิตวิทยาการศึกษา คณะครุศาสตร์

จุฬาลงกรณ์มหาวิทยาลัย

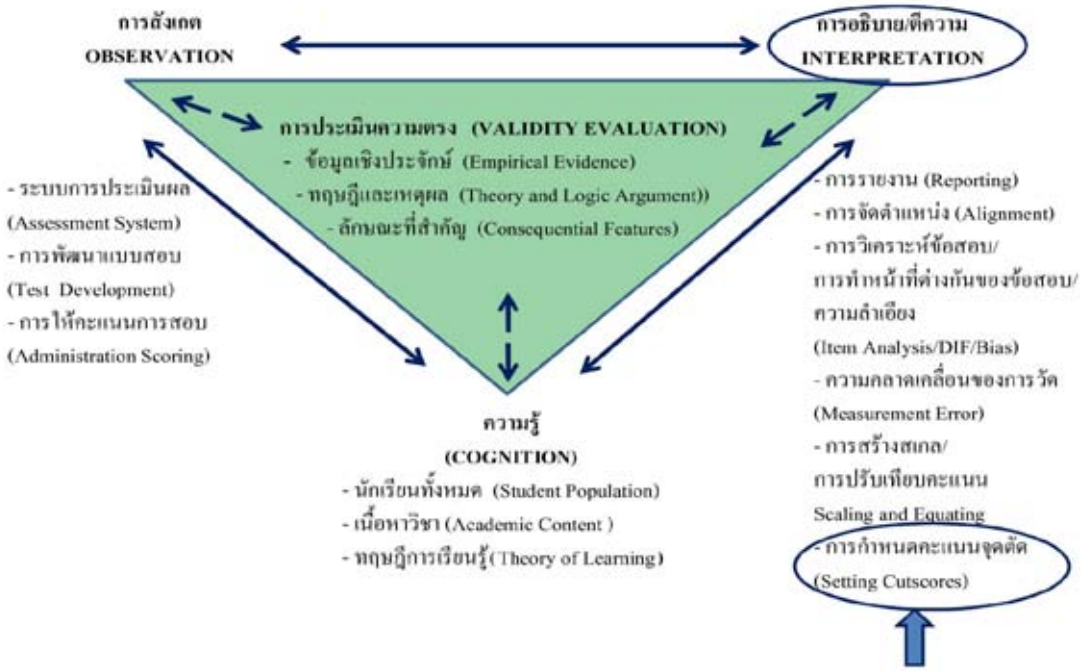
คำชี้แจง คู่มือการกำหนดคะแนนจุดตัดประกอบด้วย รายละเอียดเกี่ยวกับความเป็นมาของการกำหนดคะแนนจุดตัด ความเป็นมาและขั้นตอนการกำหนดคะแนนจุดตัดทั้ง 3 วิธี ได้แก่ วิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีเอบีซี

ความเป็นมาของการกำหนดคะแนนจุดตัด

การกำหนดคะแนนจุดตัด (Setting Cutscores) เป็นการดำเนินการกำหนดคะแนนเพื่อใช้เป็นขีดจำกัด (Threshold) ในการแบ่งคะแนนสอบตามระดับความสามารถของผู้สอบ (Towles-Reeves, 2008) ผู้เชี่ยวชาญบางท่านใช้คำว่า *มาตรฐาน (Standard)* ในความหมายเดียวกับคำว่าคะแนนจุดตัด (Cutscores) (Zieky et. al., 2008: 1) ทำให้มีการเรียกกระบวนการกำหนดคะแนนจุดตัดว่าการกำหนดมาตรฐาน (Standard Setting) คะแนนจุดตัด คือ คะแนนบนการแจกแจงของคะแนนทั้งหมดที่ใช้ในการแบ่งคะแนนออกเป็น 2 กลุ่มหรือมากกว่าสองกลุ่ม (Berk, 1980; Cizek & Bunch, 2007; Zieky et. al., 2008) ซึ่งแนวคิดเรื่องการกำหนดคะแนนจุดตัด มีวิวัฒนาการมากว่า 60 ปี ในปี ค.ศ.1970 การกำหนดคะแนนจุดตัดถูกนำมาใช้เพื่อการตีความหมายและอธิบายคะแนนที่ได้จากการสอบโดยเฉพาะการประเมินผลสัมฤทธิ์แบบอิงเกณฑ์ ในวงการศึกษาของสหรัฐอเมริกานิยมวัดผลสัมฤทธิ์ทางการเรียนโดยใช้ข้อสอบแบบอิงเกณฑ์ (criterion referenced testing) ซึ่งเป็นการนำคะแนนของแต่ละคนไปเปรียบเทียบกับคะแนนที่กำหนดไว้

ในปี ค.ศ. 2001 คะแนนจุดตัดได้ถูกใช้อย่างกว้างขวางมากขึ้นในสหรัฐอเมริกา เนื่องจากรัฐบาลกลางของสหรัฐอเมริกาออกกฎหมายปฏิรูปการศึกษาที่ชื่อว่า No Child Left Behind ที่มีผลบังคับทั้งโรงเรียนประถมศึกษาและมัธยมศึกษาทั่วสหรัฐอเมริกา โดยมีสาระสำคัญว่านักเรียน เกรด 3–8 จะต้องมีพัฒนาการทางการเรียนด้านการอ่านและการคำนวณ จึงมีการทดสอบเพื่อวัดผลของพัฒนาการ ทางการเรียนเป็นรายปี เรียกว่า Adequate Yearly Progress ดังนั้นจึงมีกำหนดคะแนนจุดตัดเพื่อใช้เป็นขีดจำกัด (Threshold) ในการตัดสินเกี่ยวกับระดับความสามารถทางด้านการอ่านและการคำนวณ (Zieky & Perie, 2006) จากความสำคัญของคะแนนจุดตัดที่กล่าวมา ทำให้นักวิชาการพยายามพัฒนาวิธีการกำหนดคะแนนจุดตัดเพื่อให้ได้คะแนนจุดตัดที่มีความแม่นยำในการจำแนกผู้สอบตามระดับความสามารถและมีความน่าเชื่อถือ ในปี ค.ศ. 1986 มีวิธีที่ใช้ในการกำหนดคะแนนจุดตัดทั้งสิ้น 38 วิธี (Berk, 1986) ต่อมาในปี ค.ศ.1996 พบว่ามีวิธีการกำหนด

คะแนนจุดตัด 50 วิธี (Berk, 1996) ในปัจจุบันการกำหนดคะแนนจุดตัดมีเป้าหมายเพื่อการรายงานผลการจัดการศึกษาของระบบการศึกษาต่างๆ ทั่วโลกอย่างกว้างขวาง เช่น สหรัฐอเมริกา อังกฤษ ออสเตรเลียเป็นต้น โดยวิธีที่นิยมใช้มากที่สุดคือวิธีแองกอฟและวิธีบุ๊คมาร์ค (Zieky et. al., 2008)



ภาพที่ 1 บทบาทของการกำหนดคะแนนจุดตัดในกระบวนการวัดและประเมินผลการเรียนรู้

ที่มา : Marion และ Pellegrino (2006)

จากภาพที่ 1 การกำหนดคะแนนจุดตัด มีบทบาทเพื่อการตีความคะแนนสอบในกระบวนการวัดและประเมินผลการเรียนรู้ โดยใช้เป็นขีดจำกัด (Threshold) ในการแบ่งคะแนนสอบเป็นกลุ่มๆ ซึ่งคะแนนแต่ละกลุ่มก็จะสามารถอธิบายความหมายตามเป้าหมายของการกำหนดคะแนนจุดตัด

วิธีกำหนดคะแนนจุดตัด

1. วิธีแองกอฟแบบใช่/ไม่ใช่

วิธีแองกอฟ (Angoff method) ถูกคิดขึ้นโดย William H. Angoff ในปีค.ศ.1970 เพื่อใช้ในการกำหนดคะแนนจุดตัดสำหรับข้อสอบแบบเลือกตอบ (Multiple choice tests) ซึ่งวิธีแองกอฟถูกนำไปใช้ในการกำหนดคะแนนจุดตัดอย่างกว้างขวางสำหรับการทดสอบที่มีความสำคัญ (high stakes Testing) (Angoff, 1971, cited in Kane, 1994) และยังได้รับความนิยมอย่างต่อเนื่องนับจนถึงปัจจุบัน (Zieky et. al., 2008) ทั้งนี้เนื่องจากว่าวิธีแองกอฟเป็นวิธีที่ง่ายต่อการปฏิบัติและเป็นการกำหนดคะแนนจุดตัดโดยการพิจารณาจากข้อสอบ จึงสามารถดำเนินการได้ทั้งก่อนและหลังจากที่มีการจัดสอบแล้ว โดยวิธีการดำเนินการกำหนดคะแนนจุดตัดนั้นกระทำโดยให้ผู้เชี่ยวชาญพิจารณาความน่าจะเป็นที่ผู้มีความสามารถแต่ละระดับจะสามารถตอบข้อสอบแต่ละข้อได้ถูกต้อง

วิธีแองกอฟแบบใช่/ไม่ใช่ เป็นวิธีการที่พัฒนามาจากวิธีแองกอฟ ซึ่งการดำเนินการจะต่างจากวิธีแองกอฟตรงที่วิธีแองกอฟแบบใช่/ไม่ใช่จะให้ผู้เชี่ยวชาญพิจารณาว่าผู้ที่มีความสามารถแต่ละระดับสามารถตอบข้อสอบได้ถูกต้องใช่หรือไม่ ซึ่งผู้เชี่ยวชาญจะตอบเพียงว่าใช่ หรือไม่ใช่

ขั้นตอนการกำหนดคะแนนจุดตัด

การกำหนดคะแนนจุดตัดโดยวิธีแองกอฟแบบใช่/ไม่ใช่ ดำเนินการกำหนดคะแนนจุดตัด 3 รอบตามลำดับดังต่อไปนี้

รอบที่ 1 ผู้เชี่ยวชาญแต่ละท่านพิจารณาข้อสอบทีละข้อและตัดสินว่าผู้มีความสามารถคาบเส้นของแต่ละระดับความสามารถจะสามารถตอบข้อสอบแต่ละข้อได้ถูกต้องใช่หรือไม่ หากผู้เชี่ยวชาญพิจารณาว่าผู้มีความสามารถคาบเส้นของระดับความสามารถที่กำลังพิจารณาอยู่สามารถตอบข้อสอบข้อนั้นได้ถูกต้องก็ให้ใส่เลข 1 ลงในช่องผลการพิจารณาของข้อนั้น แต่หากผู้เชี่ยวชาญพิจารณาว่าผู้มีความสามารถคาบเส้นของระดับความสามารถที่กำลังพิจารณาอยู่ไม่สามารถตอบข้อสอบข้อนั้นได้ถูกต้องก็ให้ใส่เลข 0 ลงในช่องผลการพิจารณาของข้อนั้น โดยผู้เชี่ยวชาญจะต้องพิจารณาทีละระดับความสามารถจนครบทุกระดับ

รอบที่ 2 ผู้เชี่ยวชาญร่วมกันอภิปรายกลุ่มย่อยเกี่ยวกับคะแนนจุดตัดที่กำหนดในรอบที่ 1 และกำหนดคะแนนจุดตัดอีกครั้งในรอบนี้ โดยใช้หลักการเดียวกันกับรอบที่ 1

รอบที่ 3 ผู้เชี่ยวชาญร่วมกันอภิปรายกลุ่มใหญ่เกี่ยวกับคะแนนจุดตัดที่กำหนดในรอบที่ 2 และกำหนดคะแนนจุดตัดครั้งสุดท้าย โดยใช้หลักการเดียวกันกับรอบที่ 1

วิธีการคิดคะแนนจุดตัด

การคิดคะแนนจุดตัดที่กำหนดโดยผู้เชี่ยวชาญแต่ละท่าน สามารถดำเนินการได้ดังนี้

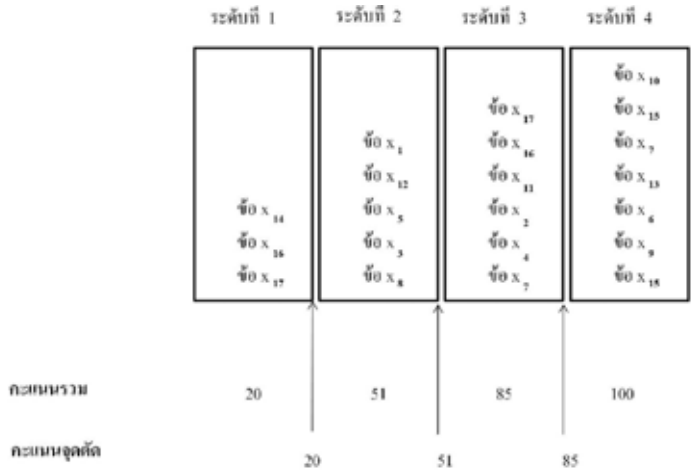
การศึกษาครั้งนี้เป็นการกำหนดคะแนนจุดตัดสำหรับใช้เป็นเกณฑ์ในการแบ่งผู้สอบออกเป็น 4 ระดับความสามารถ ได้แก่ ระดับที่ 1 ระดับที่ 2 ระดับที่ 3 และระดับที่ 4 ดังนั้นจะต้องกำหนดคะแนนจุดตัดทั้งสิ้น 3 ค่า โดยค่าแรกเป็นคะแนนจุดตัดระหว่างความสามารถระดับที่ 1 กับระดับที่ 2 ค่าที่สองเป็นคะแนนจุดตัดระหว่างความสามารถระดับที่ 2 กับระดับที่ 3 และค่าที่สามเป็นคะแนนจุดตัดระหว่างความสามารถระดับที่ 3 กับระดับที่ 4 ตามลำดับ หลักในการคิดคะแนนจุดตัดแต่และค่าเป็นดังนี้

คะแนนจุดตัดระหว่างความสามารถระดับที่ 1 กับระดับที่ 2 คิดจากผลคะแนนรวมของข้อสอบข้อที่เชี่ยวชาญคิดว่าผู้มีความสามารถคาบเส้นของระดับที่ 1 สามารถตอบได้ถูกต้อง

คะแนนจุดตัดระหว่างความสามารถระดับที่ 2 กับระดับที่ 3 คิดจากผลคะแนนรวมของข้อสอบข้อที่ผู้เชี่ยวชาญคิดว่าผู้มีความสามารถคาบเส้นของระดับที่ 2 สามารถตอบได้ถูกต้อง

คะแนนจุดตัดระหว่างความสามารถระดับที่ 3 กับระดับที่ 4 คิดจากผลคะแนนรวมของข้อสอบข้อที่ผู้เชี่ยวชาญคิดว่าผู้มีความสามารถคาบเส้นของระดับที่ 4 สามารถตอบได้ถูกต้อง

ผู้วิจัยขอเสนอภาพประกอบคำอธิบายวิธีการคิดคะแนนจุดตัดของวิธีแองกอฟแบบใช่/ไม่ใช่ โดย “ข้อ x” แสดงถึงข้อสอบข้อที่ผู้เชี่ยวชาญคิดว่าผู้มีความสามารถคาบเส้นของแต่ละระดับจะสามารถตอบได้ถูกต้อง รายละเอียดต่าง ๆ ดังแสดงในภาพต่อไปนี้



ภาพที่ 2 ตัวอย่างการคิดคะแนนจุดตัดของวิธีแองกอฟแบบใช่/ไม่ใช่

2. วิธีบุ๊คมาร์คถูกเสนอขึ้นครั้งแรกในปี 1996 โดย Lewis และ Green (1996 cited in Zieky et. al., 2008) และได้รับความนิยมอย่างรวดเร็ว โดยเฉพาะในปี ค.ศ. 2000 วิธีบุ๊คมาร์คถูกนำไปใช้กันอย่างแพร่หลายในสหรัฐอเมริกา โดยจากการสำรวจพบว่าวิธีบุ๊คมาร์คถูกนำไปใช้ใน 18 รัฐ ซึ่งถือว่าถูกใช้มากที่สุด (Nellhaus, 2000) ต่อมาในปี ค.ศ. 2001 วิธีนี้ก็ได้ถูกนำมาใช้เพื่อกำหนดคะแนนจุดตัดมากกว่า 28 รัฐ (Egan, 2001 cited in Lin, 2008) และเมื่อปี 2005 มี 31 รัฐ ที่ใช้นี้เพื่อกำหนดคะแนนจุดตัด และใช้ในการกำหนดคะแนนจุดตัดสำหรับการสอบที่สำคัญ (Perie, 2005 cited in Lee & Lewis, 2008) วิธีบุ๊คมาร์คเป็นวิธีที่สามารถใช้ได้ทั้งข้อสอบที่เป็นแบบเลือกตอบที่ให้คะแนน 2 ค่า และข้อสอบแบบให้คะแนนมากกว่า 2 ค่า นอกจากนี้วิธีบุ๊คมาร์คยังมีจุดเด่นอยู่ที่การจัดทำคู่มือเรียงข้อสอบ (Ordered Item Booklet: OIB) ซึ่งจัดทำโดยการเรียงข้อสอบจากข้อง่ายไปข้อยาก โดยใช้ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory) วิเคราะห์ค่าความยากของข้อสอบ และเป็นการทำให้ความสามารถของนักเรียนอยู่ในระดับเดียวกัน (same continuum) ซึ่งช่วยให้ผู้เชี่ยวชาญกำหนดจุดตัดได้ง่ายขึ้น

วิธีบุ๊คมาร์คจะดำเนินการกำหนดคะแนนจุดตัด 3 รอบ ดังนี้

รอบที่ 1 ผู้เชี่ยวชาญแต่ละท่านพิจารณาคู่มือเรียงข้อสอบในแต่ละหน้า และพิจารณาในประเด็นต่อไปนี้ 1) ผู้สอบต้องมีความรู้ความสามารถในเรื่องใดจึงจะสามารถตอบข้อสอบแต่ละข้อได้ถูกต้อง และ 2) เหตุใดข้อสอบข้อนี้จึงยากกว่าข้อก่อนๆ จากนั้นผู้เชี่ยวชาญกำหนดคะแนนจุดตัดอย่างอิสระโดยนำที่คั่นหนังสือมาคั่นที่ข้อสอบที่ผู้เชี่ยวชาญคิดว่ามีความเหมาะสมที่จะเป็นจุดตัด โดยข้อที่อยู่ก่อนหน้าที่คั่นหนังสือจะเป็นข้อที่ผู้มีความสามารถคาบเส้นของระดับความสามารถที่กำลังพิจารณาอยู่ มีความน่าจะเป็น 0.67 ที่จะตอบข้อสอบได้ถูกต้อง โดยผู้เชี่ยวชาญจะพิจารณาจากระดับความสามารถต่ำที่สุดไปจนถึงระดับความสามารถที่สูงที่สุด

รอบที่ 2 ผู้เชี่ยวชาญร่วมกันอภิปรายกลุ่มย่อยเกี่ยวกับคะแนนจุดตัดที่กำหนดในรอบที่ 1 และกำหนดคะแนนจุดตัดอีกครั้ง โดยอาศัยหลักการเดียวกับรอบที่ 1

รอบที่ 3 ผู้เชี่ยวชาญร่วมกันอภิปรายกลุ่มใหญ่เกี่ยวกับคะแนนจุดตัดที่กำหนดในรอบที่ 2 และกำหนดคะแนนจุดตัดอีกครั้ง โดยอาศัยหลักการเดียวกับรอบที่ 1

การจัดทำคู่มือเรียงข้อสอบ

คู่มือเรียงข้อสอบจัดทำโดยแสดงรายละเอียดของข้อสอบไว้หน้าละ 1 ข้อ แต่ละหน้าจะถูกจัดเรียงตามค่าความยากของข้อสอบโดยเรียงลำดับหน้าจากข้อสอบง่ายที่ง่ายที่สุดไปหาข้อที่ยากที่สุด โดยค่าความยากของข้อสอบได้มาจากการวิเคราะห์โดยใช้ทฤษฎีการตอบสนองข้อสอบโมเดล 3 พารามิเตอร์ ตัวอย่างรายละเอียดคู่มือเรียงข้อสอบในแต่ละหน้าดังแสดงในภาพต่อไปนี้

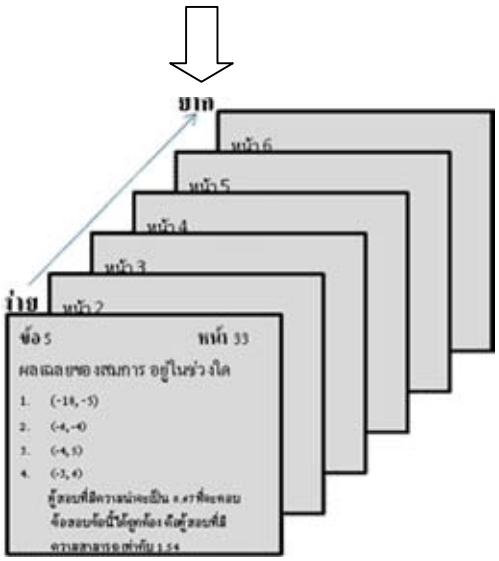
หน้า 33

ข้อ 5

ผลเฉลยของสมการ $2|5-x|=1$ อยู่ในช่วงใด

1. (-10, -5)
2. (-6, -4)
3. (-4, 5)
4. (-3, 6)

ผู้สอบที่มีความน่าจะเป็น 0.67 ที่จะตอบข้อสอบข้อนี้ได้ถูกต้อง คือผู้สอบที่มีความสามารถ (θ) เท่ากับ 1.54



ภาพที่ 3 ตัวอย่างคู่มือเรียงข้อสอบ

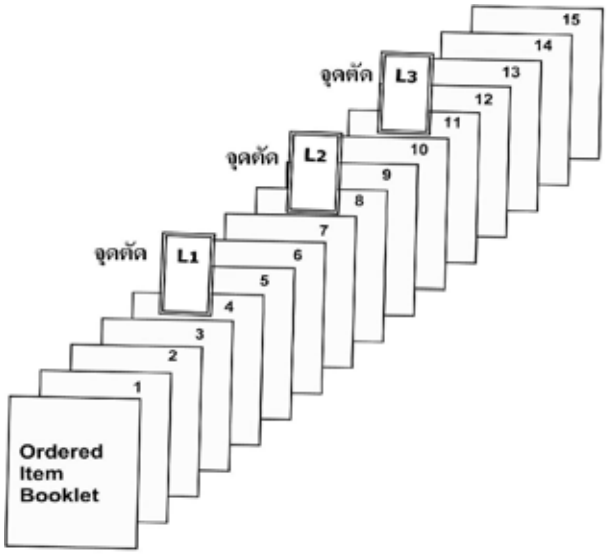
จากตัวอย่างคู่มือการจัดเรียงข้อสอบแต่ละหน้า มีรายละเอียดที่แสดงดังนี้

1) แสดงลำดับข้อสอบในแบบสอบไว้ที่มุมซ้ายมือ

2) แสดงลำดับหน้าในคู่มือเรียงข้อสอบ ไว้ที่มุมขวามือ

3) แสดงข้อคำถาม และตัวเลือก ไว้ในย่อหน้าแรก

4) แสดงค่าความสามารถของผู้สอบ (θ) ที่มีความน่าจะเป็น 0.67 ที่จะตอบข้อสอบข้อนั้นได้ถูกต้อง ไว้ในบรรทัดสุดท้าย



ภาพที่ 4 ตัวอย่างการกำหนดคะแนนจุดตัดของวิธีบุ๊คมาร์ค

วิธีการคิดคะแนนจุดตัด

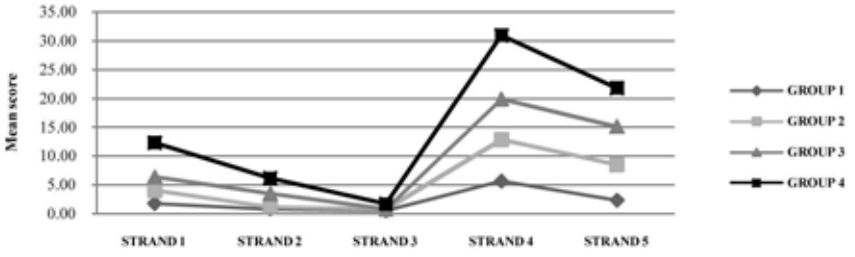
การคิดคะแนนจุดตัดที่กำหนดโดยผู้เชี่ยวชาญแต่ละท่าน สามารถดำเนินการได้ดังนี้ คะแนนจุดตัดแต่ละค่าหาได้จากการเปลี่ยนค่าความสามารถของผู้สอบ (θ) ที่มีความน่าจะเป็น 0.67 ที่จะตอบข้อสอบข้อที่อยู่ก่อนหน้าที่คั่นหนังสือได้ถูกต้อง ไปเป็นคะแนนสอบ (Expected score) โดยสามารถดูคะแนนสอบที่ระดับความสามารถต่าง ๆ ได้จากสารสนเทศเกี่ยวกับการแปลงค่าความสามารถเป็นคะแนนจุดตัด

3. วิธีเอบีซีเป็นวิธีที่ผสมผสานระหว่างวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีวิเคราะห์กลุ่ม ซึ่งวิธีแองกอฟแบบใช่/ไม่ใช่ และวิธีบุ๊คมาร์คเป็นวิธีแบบสอบเป็นศูนย์กลาง (test-center methods) (Zieky et al., 2008) กล่าวคือเป็นการกำหนดคะแนนจุดตัดโดยเน้นการพิจารณาจากข้อสอบเป็นหลัก ส่วนวิธีการวิเคราะห์กลุ่มเป็นวิธีผู้สอบเป็นศูนย์กลาง (Examinee-centered methods) ซึ่งเป็นการกำหนดคะแนนจุดตัดที่เน้นการพิจารณาจากผลการตอบของผู้สอบเป็นหลัก (Berk, 1996; Plake, 2008; Lee & Lewis, 2008; Näsström & Nyström, 2008) วิธีเอบีซีจะดึงเอาจุดเด่นและขณะเดียวกันก็ลดจุดด้อยของแต่ละวิธีลง ซึ่งวิธีแองกอฟแบบใช่/ไม่ใช่ วิธีบุ๊คมาร์ค และวิธีการวิเคราะห์กลุ่มต่างก็มีจุดเด่นที่ต่างกัน โดยวิธีแองกอฟแบบใช่/ไม่ใช่มีจุดเด่นคือการให้ความสำคัญกับการกำหนดคะแนนจุดตัดของผู้เชี่ยวชาญโดยปราศจากกรอบเรื่องค่าความยากง่ายของข้อสอบ กล่าวคือเป็นการกำหนดคะแนนจุดตัดที่ให้ความสำคัญกับการตัดสินใจของผู้เชี่ยวชาญที่พิจารณาจากเนื้อหาที่สอบวัดเป็นหลัก สำหรับวิธีบุ๊คมาร์คมีจุดเด่นอยู่ที่คู่มือการจัดเรียงข้อสอบ ที่ช่วยให้ผู้เชี่ยวชาญกำหนดจุดตัดได้ง่ายและเร็ว ส่วนวิธีการวิเคราะห์กลุ่มเป็นการวิเคราะห์ทางสถิติเพื่อจัดให้กลุ่มผู้สอบที่มีลักษณะที่สนใจเหมือนหรือคล้ายกันอยู่กลุ่มเดียวกัน และแต่ละกลุ่มจะต้องมีลักษณะที่สนใจต่างกันให้มากที่สุด ซึ่งกล่าวได้ว่าเป็นวิเคราะห์เพื่อการกำหนดกลุ่มให้ผู้สอบจากลักษณะที่สนใจโดยปราศจากความลำเอียงใดๆ

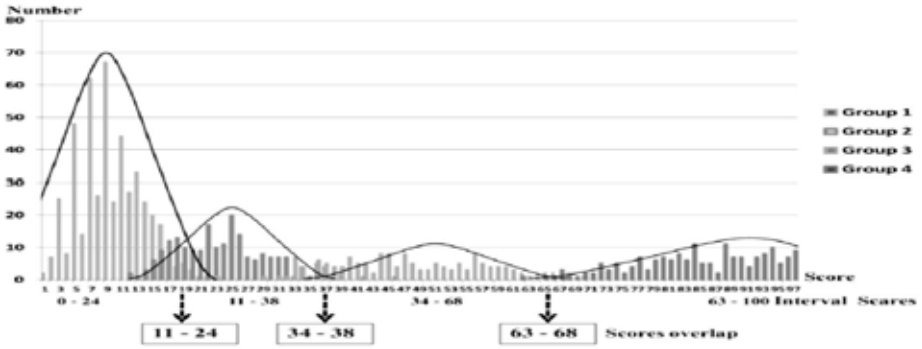
สำหรับข้อด้อยของทั้ง 3 วิธีข้างต้น คือวิธีแองกอฟแบบใช่/ไม่ใช่นั้นผู้เชี่ยวชาญจะต้องใช้กระบวนการคิดอย่างมาก (high cognitive load) ในการพิจารณาว่าผู้มีความสามารถคาบเส้นจะสามารถตอบข้อสอบได้ถูกต้องใช่หรือไม่ โดยเฉพาะในกรณีที่มีจำนวนข้อสอบที่ต้องพิจารณามาก ๆ และต้องกำหนดคะแนนจุดตัดหลายรอบ อาจทำให้ผู้เชี่ยวชาญเกิดความอ่อนล้าในขณะที่กำหนดคะแนนจุดตัดซึ่งอาจส่งผลให้คะแนนจุดตัดที่กำหนดเกิดความคลาดเคลื่อนได้ (Clauser, et al., 2009) ส่วนวิธีบุ๊คมาร์คมีจุดด้อย คือหากเกิดกรณีที่ผู้เชี่ยวชาญรู้สึกว่าข้อสอบที่ถูกเรียงไว้ตามลำดับความยากง่ายนั้นบางข้อมีตำแหน่งที่ไม่เหมาะสม ซึ่งกรณีเช่นนี้อาจส่งผลต่อคุณภาพของคะแนนจุดตัดที่กำหนดได้ ส่วนวิธีการวิเคราะห์กลุ่มก็มีจุดด้อยคือไม่สามารถระบุได้ว่าแต่ละกลุ่มที่แบ่งมีความสามารถระดับใดและไม่สามารถกำหนดคะแนนจุดตัดที่สูงกว่าหรือต่ำกว่าผลการตอบข้อสอบของผู้สอบกลุ่มที่เลือกมาวิเคราะห์ได้ (Violato, Marini & Lee, 2003; Hess, Subhiyah & Giordano, 2007)

จากที่กล่าวมาข้างต้น วิธีเอบีซีจึงเป็นวิธีที่ถูกพัฒนาขึ้นโดยมีเป้าหมายที่จะทำให้ผู้เชี่ยวชาญสามารถกำหนดคะแนนจุดตัดได้ง่ายและเร็ว โดยให้ผู้เชี่ยวชาญกำหนดคะแนนจุดตัดจากคู่มือเรียงข้อสอบซึ่งถือเป็นจุดเด่นของบุ๊คมาร์ค แต่วิธีเอบีซีจะแบ่งคู่มือเรียงข้อสอบเป็นหลาย ๆ ฉบับตามสาระการเรียนรู้เรียกว่า คู่มือเรียงข้อสอบรายสาระการเรียนรู้ (Domain–Ordered Item Booklet: D-OIB) ซึ่งเป็นคู่มือที่จัดทำขึ้นโดยแบ่งข้อสอบเป็นกลุ่มตามสาระการเรียนรู้ กล่าวคือคู่มือเรียงข้อสอบรายสาระการเรียนรู้จะมีหลายฉบับขึ้นอยู่กับจำนวนสาระการเรียนรู้ที่สอบวัด ข้อสอบในแต่ละสาระการเรียนรู้จะถูกจัดเรียงจากข้อง่ายไปข้อยาก ข้อละ 1 หน้า โดยใช้ทฤษฎีการตอบสนองข้อสอบ โมเดล 3 พารามิเตอร์ (Three parameter model) ในการวิเคราะห์ค่าความยากของข้อสอบ (b) ค่าอำนาจจำแนกของข้อสอบ (a) และค่าโอกาสในการเดาข้อสอบ (c) ทั้งนี้ก็เพื่อให้การพิจารณากำหนดคะแนนจุดตัดทำได้ง่ายขึ้น เนื่องจากคู่มือเรียงข้อสอบรายสาระการเรียนรู้แต่ละฉบับจะประกอบด้วยข้อสอบที่วัดสาระการเรียนรู้เดียวกัน ส่วนหลักการพิจารณาเพื่อกำหนดคะแนนจุดตัดจะดำเนินคล้ายกับวิธีแองกอฟแบบใช่/ไม่ใช่ กล่าวคือให้ผู้เชี่ยวพิจารณาข้อสอบตามลำดับความยากง่าย และตัดสินว่าผู้ที่สามารถตอบข้อสอบแต่ละข้อได้ถูกต้องควรมีความสามารถอย่างน้อยระดับใดแทนที่จะให้ผู้เชี่ยวชาญนำที่คั่นหนังสือมาขั้นตรงข้อที่ผู้เชี่ยวชาญคิดว่าเป็นจุดตัด ซึ่งการพิจารณาแบบนี้น่าจะช่วยลดปัญหาที่เกิดจากการที่ผู้เชี่ยวชาญคิดว่าข้อสอบที่ถูกจัดเรียงไว้ตามความยากง่ายมีลำดับที่ไม่เหมาะสม เพราะวิธีเอบีซีใช้คู่มือการจัดเรียงข้อสอบเป็นเพียงสารสนเทศประกอบการกำหนดคะแนนจุดตัดที่เสนอเกี่ยวกับค่าความยากง่ายของข้อสอบที่ได้จากการวิเคราะห์จากผลการตอบของผู้สอบเท่านั้น กล่าวได้ว่าวิธีเอบีซียังคงให้ความสำคัญกับความคิดเห็นของผู้เชี่ยวชาญที่พิจารณาจากเนื้อหาที่สอบเป็นหลักแทนที่จะต้องพิจารณาคะแนนจุดตัดที่ถูกจำกัดไว้กับลำดับของความยากง่ายของข้อสอบจนเกินไป ซึ่งถือเป็นจุดเด่นของวิธีแองกอฟแบบใช่/ไม่ใช่ นอกจากนี้วิธีเอบีซีจะเสนอสารสนเทศที่มากกว่าค่าความยากง่ายของข้อสอบ โดยจะเสนอสารสนเทศเกี่ยวกับผลการวิเคราะห์กลุ่ม (Clustering – Result: CR) ซึ่งเป็นผลการจัดกลุ่มที่มิใช่เป็นการกำหนดโดยอาศัยความคิดเห็นของผู้เชี่ยวชาญแต่เป็นการวิเคราะห์ข้อมูลทางสถิติที่ใช้ข้อมูลจากผลการสอบของผู้สอบมาทำการวิเคราะห์ โดยตัวแปรที่ใช้ในการแบ่งกลุ่ม ได้แก่ คะแนนรวมของข้อสอบของแต่ละรายสาระการเรียนรู้ สารสนเทศนี้จะนำเสนอเป็นผลคะแนนเฉลี่ยรายสาระการ

เรียนรู้ของแต่ละกลุ่ม และช่วงของคะแนนสอบที่คาบเกี่ยวกันของแต่ละระดับความสามารถ ตัวอย่างดังแสดงในภาพที่ 1 และภาพที่ 2



ภาพที่ 1 แสดงตัวอย่างผลคะแนนเฉลี่ยรายสาระการเรียนรู้ของแต่ละกลุ่ม



ภาพที่ 2 แสดงตัวอย่างช่วงของคะแนนสอบที่คาบเกี่ยวกันของแต่ละกลุ่ม

วิธีเอบีซีจะดำเนินการกำหนดคะแนนจุดตัด 3 รอบ ดังนี้

รอบที่ 1 ผู้เชี่ยวชาญพิจารณาคู่มือเรียงข้อสอบรายสาระการเรียนรู้ และผลการวิเคราะห์กลุ่ม จากนั้นก็ให้ผู้เชี่ยวชาญพิจารณาว่าผู้ที่สามารถตอบข้อสอบแต่ละข้อได้ถูกต้องควรมีความสามารถอย่างน้อยระดับใด โดยมีหลักการพิจารณาเพิ่มเติมคือหากข้อใดที่ผู้เชี่ยวชาญไม่มั่นใจก็ให้ระบุไว้ในข้อสอบ ดังนั้นข้อสอบที่ผู้เชี่ยวชาญกำหนดไว้ในแต่ละระดับความสามารถก็จะมีสองกลุ่ม คือกลุ่มที่ผู้เชี่ยวชาญมั่นใจและกลุ่มที่ผู้เชี่ยวชาญไม่มั่นใจ

รอบที่ 2 ผู้เชี่ยวชาญร่วมกันอภิปรายกลุ่มย่อยเกี่ยวกับข้อสอบที่ผู้เชี่ยวชาญยังไม่มั่นใจในรอบที่ 1 เพื่อให้การอภิปรายกระชับและตรงประเด็นมากขึ้น และกำหนดคะแนนจุดตัดอีกครั้งโดยอาศัยหลักการเดียวกับรอบที่ 1

รอบที่ 3 ผู้เชี่ยวชาญร่วมกันอภิปรายกลุ่มใหญ่เกี่ยวกับข้อสอบที่ผู้เชี่ยวชาญยังไม่มั่นใจในรอบที่ 2 และกำหนดคะแนนจุดตัดอีกครั้ง โดยอาศัยหลักการเดียวกับรอบที่ 1

วิธีการคิดคะแนนจุดตัด

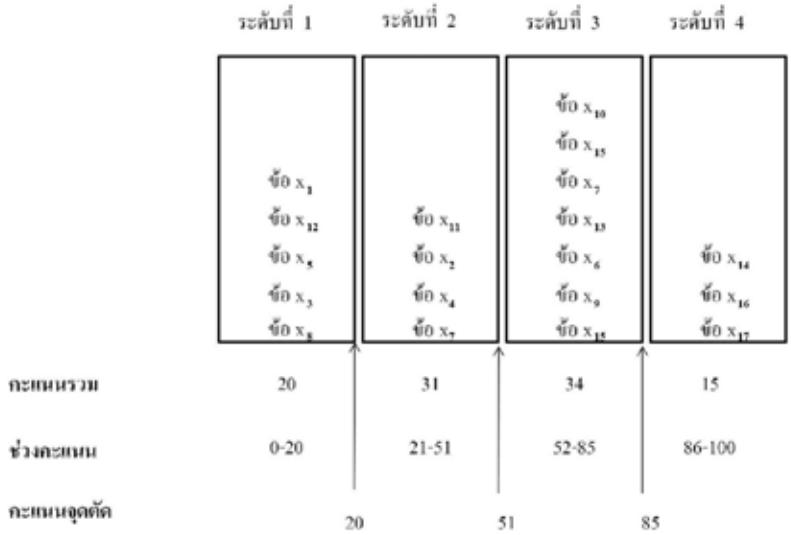
การคิดคะแนนจุดตัดที่กำหนดโดยผู้เชี่ยวชาญแต่ละท่าน สามารถดำเนินการได้ดังนี้

คะแนนจุดตัดระหว่างความสามารถระดับที่ 1 กับระดับที่ 2 คิดจากผลคะแนนรวมของข้อสอบที่ผู้เชี่ยวชาญคิดว่าผู้ที่สามารถตอบได้ถูกต้องควรมีความสามารถอย่างน้อยระดับที่ 1

คะแนนจุดตัดระหว่างความสามารถระดับที่ 2 กับระดับที่ 3 คิดจากผลคะแนนรวมของข้อสอบที่ผู้เชี่ยวชาญคิดว่าผู้ที่สามารถตอบได้ถูกต้องควรมีความสามารถอย่างน้อยระดับที่ 1 และข้อสอบที่ผู้เชี่ยวชาญคิดว่าผู้ที่สามารถตอบได้ถูกต้องควรมีความสามารถอย่างน้อยระดับที่ 2

คะแนนจุดตัดระหว่างความสามารถระดับที่ 3 กับระดับที่ 4 คิดจากผลคะแนนรวมของข้อสอบที่ผู้เชี่ยวชาญคิดว่าผู้ที่สามารถตอบได้ถูกต้องควรมีความสามารถอย่างน้อยระดับที่ 1 ข้อสอบที่ผู้เชี่ยวชาญคิดว่าผู้ที่สามารถตอบได้ถูกต้องควรมีความสามารถอย่างน้อยระดับที่ 2 และข้อสอบที่ผู้เชี่ยวชาญคิดว่าผู้ที่สามารถตอบได้ถูกต้องควรมีความสามารถอย่างน้อยระดับที่ 3

ผู้วิจัยขอเสนอภาพประกอบคำอธิบายวิธีการคิดคะแนนจุดตัดของวิธีเอบีซีโดย “ข้อ x” แสดงถึงข้อสอบที่ผู้เชี่ยวชาญคิดว่าผู้ที่สามารถตอบได้ถูกต้องควรมีความสามารถอย่างน้อยระดับนั้น ๆ รายละเอียดต่าง ๆ ดังแสดงในภาพต่อไปนี้



ภาพที่ 5 ตัวอย่างการคิดคะแนนจุดตัดของวิธีเอบีซี

แบบกรอกผลการพิจารณาคะแนนจุดตัดด้วยวิธีแองกอฟแบบใช่/ไม่ใช่

ชื่อผู้ทรงคุณวุฒิ.....

ตำแหน่ง/วิทยฐานะ..... โรงเรียน.....
ประสบการณ์การสอน.....ปี

คำชี้แจง กรุณาพิจารณาข้อสอบทีละข้อ และพิจารณาว่าผู้ที่มีความสามารถแต่ละระดับจะสามารถตอบข้อสอบแต่ละข้อได้ถูกต้องหรือไม่ หากผู้เชี่ยวชาญพิจารณาว่าผู้ที่มีความสามารถแต่ละระดับความสามารถที่กำลังพิจารณาอยู่สามารถตอบข้อสอบข้อนั้นได้ถูกต้องก็ให้ใส่เลข 1 ลงในช่องผลการพิจารณา แต่หากผู้เชี่ยวชาญพิจารณาว่าผู้ที่มีความสามารถแต่ละระดับความสามารถที่กำลังพิจารณาอยู่ไม่สามารถตอบข้อสอบข้อนั้นได้ถูกต้องก็ให้ใส่เลข 0 ลงในช่องผลการพิจารณา โดยผู้เชี่ยวชาญจะต้องพิจารณาทีละระดับความสามารถจนครบทุกระดับความสามารถ

ข้อ	รอบที่ 1				รอบที่ 2				รอบที่ 3			
	ระดับที่ 1	ระดับที่ 2	ระดับที่ 3	ระดับที่ 4	ระดับที่ 1	ระดับที่ 2	ระดับที่ 3	ระดับที่ 4	ระดับที่ 1	ระดับที่ 2	ระดับที่ 3	ระดับที่ 4
1.												
2.												
3.												
4.												
5.												
6.												
7.												
8.												
9.												
10												
11.												
12.												
13.												
14.												
15.												
16.												
17.												
18.												
19												
20.												
รวม												

ข้อ	รอบที่ 1				รอบที่ 2				รอบที่ 3			
	ระดับที่ 1	ระดับที่ 2	ระดับที่ 3	ระดับที่ 4	ระดับที่ 1	ระดับที่ 2	ระดับที่ 3	ระดับที่ 4	ระดับที่ 1	ระดับที่ 2	ระดับที่ 3	ระดับที่ 4
21.*												
22. *												
23. *												
24. *												
25. *												
26. *												
27. *												
28. *												
29. *												
30. *												
31. *												
32. *												
33. *												
34. *												
35. *												
36. *												
37. *												
38. *												
39*												
40. *												
รวม												
รวมทั้งสิ้น												

แบบกรอกผลการพิจารณาคะแนนจุดตัดด้วยวิธีบุ๊คมาร์ค

ชื่อผู้ทรงคุณวุฒิ.....

ตำแหน่ง/วิทยฐานะ..... โรงเรียน.....
ประสบการณ์การสอน.....ปี

คำชี้แจง กรุณาพิจารณาคู่มือการจัดเรียงข้อสอบ และตัดสินว่าข้อสอบข้อใดที่เป็นจุดตัดระหว่างแต่ละระดับความสามารถ โดยข้อสอบข้อนั้นจะเป็นข้อที่ผู้มีความสามารถคาบเส้นของระดับความสามารถที่กำลังพิจารณาอยู่ มีความน่าจะเป็น 0.67 ที่จะตอบข้อสอบได้ถูกต้อง โดยผู้เชี่ยวชาญจะพิจารณาจากระดับความสามารถต่ำที่สุดไปจนถึงระดับความสามารถที่สูงที่สุด

ระดับความสามารถ	รอบที่ 1		รอบที่ 2		รอบที่ 3	
	หน้า	ข้อที่	หน้า	ข้อที่	หน้า	ข้อที่
ระดับที่ 1/2						
ระดับที่ 2/3						
ระดับที่ 3/4						

แบบกรอกผลการพิจารณาคะแนนจุดตัดด้วยวิธีเอบีซี

วิชาคณิตศาสตร์

ชื่อผู้ทรงคุณวุฒิ.....

ตำแหน่ง/วิทยฐานะ..... โรงเรียน.....
ประสบการณ์การสอน.....ปี

คำชี้แจง กรุณาพิจารณาคู่มือเรียงข้อสอบรายสาระการเรียนรู้ทีละข้อ และพิจารณาว่าผู้ที่สามารถตอบข้อสอบแต่ละข้อได้ถูกต้องควรมีความสามารถอย่างน้อยระดับใด โดยให้ใส่เครื่องหมาย ✓ลงในช่องระดับความสามารถที่ท่านเลือก หากข้อใดที่ท่านไม่มั่นใจให้ใส่เครื่องหมาย * กำกับข้อสอบข้อนั้นไว้

สาระที่	ข้อ	รอบที่ 1				รอบที่ 2				รอบที่ 3			
		ระดับที่ 1	ระดับที่ 2	ระดับที่ 3	ระดับที่ 4	ระดับที่ 1	ระดับที่ 2	ระดับที่ 3	ระดับที่ 4	ระดับที่ 1	ระดับที่ 2	ระดับที่ 3	ระดับที่ 4
1	24												
	23												
	7												
	6												
	25												
	27												
	4												
2	29												
	28												
	30												
	15												
3	13												
	14												
4	8												
	32												
	2												

สาระที่	ข้อ	รอบที่ 1				รอบที่ 2				รอบที่ 3			
		ระดับที่ 1	ระดับที่ 2	ระดับที่ 3	ระดับที่ 4	ระดับที่ 1	ระดับที่ 2	ระดับที่ 3	ระดับที่ 4	ระดับที่ 1	ระดับที่ 2	ระดับที่ 3	ระดับที่ 4
	31												
	1												
	16												
	17												
	21												
	11												
	22												
	10												
	26												
	9												
	12												
	5												
	3												
5	35												
	20												
	40												

สาระที่	ข้อ	รอบที่ 1				รอบที่ 2				รอบที่ 3			
		ระดับที่ 1	ระดับที่ 2	ระดับที่ 3	ระดับที่ 4	ระดับที่ 1	ระดับที่ 2	ระดับที่ 3	ระดับที่ 4	ระดับที่ 1	ระดับที่ 2	ระดับที่ 3	ระดับที่ 4
	18												
	39												
	34												
	36												
	37												
	38												
	33												
	19												
รวม													

แบบกรอกผลการพิจารณาคะแนนจุดตัดด้วยวิธีเอบีซี

วิชาวิทยาศาสตร์

ชื่อผู้ทรงคุณวุฒิ.....

ตำแหน่ง/วิทยฐานะ..... โรงเรียน.....
ประสบการณ์การสอน.....ปี

คำชี้แจง กรุณาพิจารณาคู่มือเรียงข้อสอบรายสาระการเรียนรู้ทีละข้อ และพิจารณาว่าผู้ที่สามารถตอบข้อสอบแต่ละข้อได้ถูกต้องควรมีความสามารถอย่างน้อยระดับใด โดยให้ใส่เครื่องหมาย ✓ลงในช่องระดับความสามารถที่ท่านเลือก หากข้อใดที่ท่านไม่มั่นใจให้ใส่เครื่องหมาย * กำกับข้อสอบข้อนั้นไว้

สาระที่	ข้อ	รอบที่ 1				รอบที่ 2				รอบที่ 3			
		ระดับที่ 1	ระดับที่ 2	ระดับที่ 3	ระดับที่ 4	ระดับที่ 1	ระดับที่ 2	ระดับที่ 3	ระดับที่ 4	ระดับที่ 1	ระดับที่ 2	ระดับที่ 3	ระดับที่ 4
1	4												
	88.1												
	6												
	9												
	3												
	5												
	2												
	1												
	7												
	8												
	88.2												
2	87.1												
	15												
	13												
	16												
	10												

สาระที่	ข้อ	รอบที่ 1				รอบที่ 2				รอบที่ 3			
		ระดับที่ 1	ระดับที่ 2	ระดับที่ 3	ระดับที่ 4	ระดับที่ 1	ระดับที่ 2	ระดับที่ 3	ระดับที่ 4	ระดับที่ 1	ระดับที่ 2	ระดับที่ 3	ระดับที่ 4
	12												
	87.2												
	17												
	11												
	14												
	18												
3	27												
	23												
	32												
	34												
	25												
	35												
	20												
	26												
	36												
	28												

สาระที่	ข้อ	รอบที่ 1				รอบที่ 2				รอบที่ 3			
		ระดับที่ 1	ระดับที่ 2	ระดับที่ 3	ระดับที่ 4	ระดับที่ 1	ระดับที่ 2	ระดับที่ 3	ระดับที่ 4	ระดับที่ 1	ระดับที่ 2	ระดับที่ 3	ระดับที่ 4
	29												
	19												
	22												
	24												
	30												
	33												
	21												
	31												
4	90.1												
	41												
	44												
	37												
	42												
	90.2												
	45												
	43												

สาระที่	ข้อ	รอบที่ 1				รอบที่ 2				รอบที่ 3			
		ระดับที่ 1	ระดับที่ 2	ระดับที่ 3	ระดับที่ 4	ระดับที่ 1	ระดับที่ 2	ระดับที่ 3	ระดับที่ 4	ระดับที่ 1	ระดับที่ 2	ระดับที่ 3	ระดับที่ 4
	38												
	40												
	39												
5	89.1												
	49												
	50												
	46												
	53												
	48												
	47												
	52												
	54												
	89.2												
	51												
6	81.1												
	82.1												

สาระที่	ข้อ	รอบที่ 1				รอบที่ 2				รอบที่ 3			
		ระดับที่ 1	ระดับที่ 2	ระดับที่ 3	ระดับที่ 4	ระดับที่ 1	ระดับที่ 2	ระดับที่ 3	ระดับที่ 4	ระดับที่ 1	ระดับที่ 2	ระดับที่ 3	ระดับที่ 4
	62												
	58												
	57												
	59												
	63												
	61												
	55												
	56												
	82.2												
	60												
	81.2												
7	84.1												
	83.1												
	67												
	70												
	72												

สาระที่	ข้อ	รอบที่ 1				รอบที่ 2				รอบที่ 3			
		ระดับที่ 1	ระดับที่ 2	ระดับที่ 3	ระดับที่ 4	ระดับที่ 1	ระดับที่ 2	ระดับที่ 3	ระดับที่ 4	ระดับที่ 1	ระดับที่ 2	ระดับที่ 3	ระดับที่ 4
	65												
	68												
	71												
	64												
	66												
	84.2												
	69												
	83.2												
8	80												
	86.1												
	85.1												
	78												
	76												
	79												
	86.2												
	73												

สาระที่	ข้อ	รอบที่ 1				รอบที่ 2				รอบที่ 3			
		ระดับที่ 1	ระดับที่ 2	ระดับที่ 3	ระดับที่ 4	ระดับที่ 1	ระดับที่ 2	ระดับที่ 3	ระดับที่ 4	ระดับที่ 1	ระดับที่ 2	ระดับที่ 3	ระดับที่ 4
	77												
	75												
	85.2												
	74												
รวม													

ภาคผนวก จ

1. แบบสอบถามความคิดเห็นเกี่ยวกับความตรงของปัจจัยนำเข้า
2. แบบสังเกตพฤติกรรมของผู้เข้าร่วมการกำหนดคะแนนจุดตัด

แบบสอบถามความคิดเห็นเกี่ยวกับความตรงของปัจจัยนำเข้า

คำชี้แจง

แบบสอบถามชุดนี้เป็นแบบสอบถามเพื่อรวบรวมความคิดเห็นของผู้เชี่ยวชาญที่เข้าร่วมการกำหนดคะแนนจุดตัด เกี่ยวกับความตรงของปัจจัยนำเข้า แบ่งออกเป็น 2 ตอน ได้แก่

ตอนที่ 1 ข้อมูลเบื้องต้นของผู้เชี่ยวชาญ

ตอนที่ 2 ความคิดเห็นของผู้เชี่ยวชาญต่อปัจจัยนำเข้าของการกำหนดคะแนนจุดตัด

ตอนที่ 1 ข้อมูลเบื้องต้นของผู้เชี่ยวชาญ

ชื่อผู้เชี่ยวชาญ.....

ตำแหน่ง.....

สถานที่ทำงาน.....

ตอนที่ 2 ความคิดเห็นของผู้เชี่ยวชาญต่อปัจจัยนำเข้าของการกำหนดคะแนนจุดตัด

คำชี้แจง

กรุณาขีดเครื่องหมาย ✓ในช่องระดับความคิดเห็นเพื่อแสดงระดับความเห็นด้วยของท่านต่อข้อความต่อไปนี้

ความหมายของระดับความเห็นด้วย

หมายเลข 5	หมายถึง	เห็นด้วยมากที่สุด
หมายเลข 4	หมายถึง	เห็นด้วยมาก
หมายเลข 3	หมายถึง	เห็นด้วยปานกลาง
หมายเลข 2	หมายถึง	เห็นด้วยน้อย
หมายเลข 1	หมายถึง	ไม่เห็นด้วย

ข้อความ	ระดับความเห็นด้วย				
	1	2	3	4	5
ด้านความถูกต้อง					
1. การกำหนดคะแนนจุดครั้งนี้มีวิธีดำเนินการที่ถูกต้อง					
2. คำอธิบายระดับความสามารถมีเนื้อหาที่ถูกต้อง สื่อความหมายชัดเจน					
3. สารสนเทศที่ให้เพื่อประกอบการกำหนดคะแนนจุดตัดมีเนื้อหาที่ถูกต้อง สื่อความหมายชัดเจน					
ด้านความเป็นไปได้					
4. ขั้นตอนการกำหนดคะแนนจุดตัดสามารถปฏิบัติได้จริง					
5. ให้เวลาสำหรับการกำหนดคะแนนจุดตัดอย่างเพียงพอ					
6. ผู้เชี่ยวชาญทุกท่านมีโอกาสแสดงความคิดเห็นได้อย่างเสมอภาค					
7. ขั้นตอนการกำหนดคะแนนจุดตัดไม่ยุ่งยากซับซ้อน					
8. การอภิปรายร่วมกันระหว่างผู้เชี่ยวชาญช่วยให้กำหนดคะแนนจุดตัดได้ง่ายขึ้น					
9. ให้เวลาสำหรับการอภิปรายกลุ่มอย่างเพียงพอ					
ด้านความเหมาะสม					
10. เป้าหมายและการดำเนินการกำหนดคะแนนจุดตัดมีความสอดคล้องและเชื่อมโยงกัน					
11. เอกสารประกอบการดำเนินการสอดคล้องและเชื่อมโยงกันกับเป้าหมายของการกำหนดคะแนนจุดตัด					
ด้านความเข้าใจและความมั่นใจของผู้เชี่ยวชาญ					
12. ฉันเข้าใจความหมายของผู้มีความสามารถคาบเส้นเป็นอย่างดี					
13. ฉันเข้าใจถึงเป้าหมายของการกำหนดคะแนนจุดตัดเป็นอย่างดี					
14. ฉันเข้าใจถึงความแตกต่างระหว่างผู้ที่มีความสามารถที่แตกต่างกัน					
15. ฉันเข้าใจวิธีการและขั้นตอนในการกำหนดคะแนนจุดตัดเป็นอย่างดี					
16. ฉันรู้สึกกดดันเมื่อคะแนนจุดตัดที่กำหนดไม่ใกล้เคียงกับผู้เชี่ยวชาญท่านอื่น ๆ					
17. ฉันมั่นใจว่าคะแนนจุดตัดที่กำหนดมีความเหมาะสม					
18. ฉันเห็นความสำคัญของการกำหนดคะแนนจุดตัดครั้งนี้					
ด้านความเป็นประโยชน์ของเอกสาร					
19. รายงานผลการกำหนดคะแนนจุดตัดในแต่ละรอบเป็นประโยชน์ต่อการกำหนดคะแนนจุดตัดในรอบต่อ ๆ ไป					
20. สารสนเทศที่ให้เพื่อประกอบการกำหนดคะแนนจุดตัดเป็นประโยชน์ต่อการกำหนดคะแนนจุดตัดอย่างยิ่ง					

แบบสังเกตพฤติกรรมของผู้เข้าร่วมการกำหนดคะแนนจุดตัด

วิธี.....

ชื่อผู้เชี่ยวชาญที่ถูกสังเกต.....

ผู้สังเกต.....

คำชี้แจง: บันทึกพฤติกรรมของครูที่เข้าร่วมการกำหนดคะแนนจุดตัดตามประเด็นที่กำหนด

ตัวอย่างคำอธิบายหรือตัวอย่างพฤติกรรมที่ต้องการให้สังเกต

ที่	พฤติกรรม	สิ่งที่ต้องบันทึก
1.	ให้ความสนใจในการฝึกอบรมเพื่อกำหนดคะแนนจุดตัด (สนใจ) ผู้เข้าร่วมหมายเลข 5 จ้องมองการอธิบายอย่างตั้งใจ พยักหน้ารับและ จดตามเนื้อหาที่ผู้อบรมอธิบาย (ไม่สนใจ) ผู้เข้าร่วมหมายเลข 5 ไม่มองผู้อธิบาย ไม่พยักหน้ารับและ ไม่จดตามเนื้อหาที่ผู้อบรมอธิบาย	- ระบุพฤติกรรมที่แสดงว่าให้ความสนใจ - ระบุพฤติกรรมที่แสดงว่าไม่สนใจ

ขั้นตอนการฝึกอบรม

ที่	พฤติกรรม	สิ่งที่ต้องบันทึก
1.	ให้ความสนใจในการฝึกอบรมเพื่อกำหนดคะแนนจุดตัด	- ระบุพฤติกรรมที่แสดงว่าให้ความสนใจ - ระบุพฤติกรรมที่แสดงว่าไม่สนใจ
2.	สอบถามประเด็นที่ไม่เข้าใจจากผู้บรรยายหรือเพื่อนครูด้วยกัน	- ระบุคำถาม โดยระบุคำพูดทุกคนที่ถามคำถามและตอบคำถาม
3.	แสดงความคิดเห็น และข้อเสนอแนะ	- ระบุคำพูดที่ได้แสดงความคิดเห็น โดยระบุคำพูดทุกคนที่แสดงความคิดเห็นร่วมกับครูที่เราสังเกต
4.	วิพากษ์ข้อบกพร่องของวิธีการกำหนดคะแนนจุดตัด	- ระบุคำพูดที่ได้วิพากษ์ โดยระบุคำพูดทุกคนที่วิพากษ์ร่วมกับครูที่เราสังเกต

ที่	พฤติกรรม	สิ่งที่ต้องบันทึก
5.	ไม่กล้าแสดงความคิดเห็น ไม่กล้าแสดงออก	- ระบุพฤติกรรม
6.	แสดงความเหนื่อยล้า	- ระบุพฤติกรรม
7.	อื่นๆ	- ระบุ พฤติกรรม

ขั้นตอนการกำหนดจุดตัด

ที่	พฤติกรรม	สิ่งที่ต้องบันทึก
1.	ให้ความสนใจในการกำหนดคะแนนจุดตัด	- ระบุพฤติกรรมที่แสดงว่าให้ความสนใจ - ระบุพฤติกรรมที่แสดงว่าไม่สนใจ
2.	สอบถามประเด็นที่ไม่เข้าใจจากผู้บรรยายหรือเพื่อนครูด้วยกัน	- ระบุคำถาม โดยระบุคำพูดทุกคนที่ถามคำถามและตอบคำถาม
3.	แสดงความคิดเห็น และข้อเสนอแนะ	- ระบุคำพูดที่ได้แสดงความคิดเห็น โดยระบุคำพูดทุกคนที่แสดงความคิดเห็นร่วมกับครูที่เราสังเกต
4.	วิพากษ์ข้อบกพร่องของวิธีการกำหนดคะแนนจุดตัด	- ระบุคำพูดที่ได้วิพากษ์ โดยระบุคำพูดทุกคนที่วิพากษ์ร่วมกับครูที่เราสังเกต

ที่	พฤติกรรม	สิ่งที่ต้องบันทึก
5.	ไม่กล้าแสดงความคิดเห็น ไม่กล้าแสดงออก	- ระบุพฤติกรรม
6.	แสดงความเหนื่อยล้า	- ระบุพฤติกรรม
7.	อื่นๆ	- ระบุ พฤติกรรม

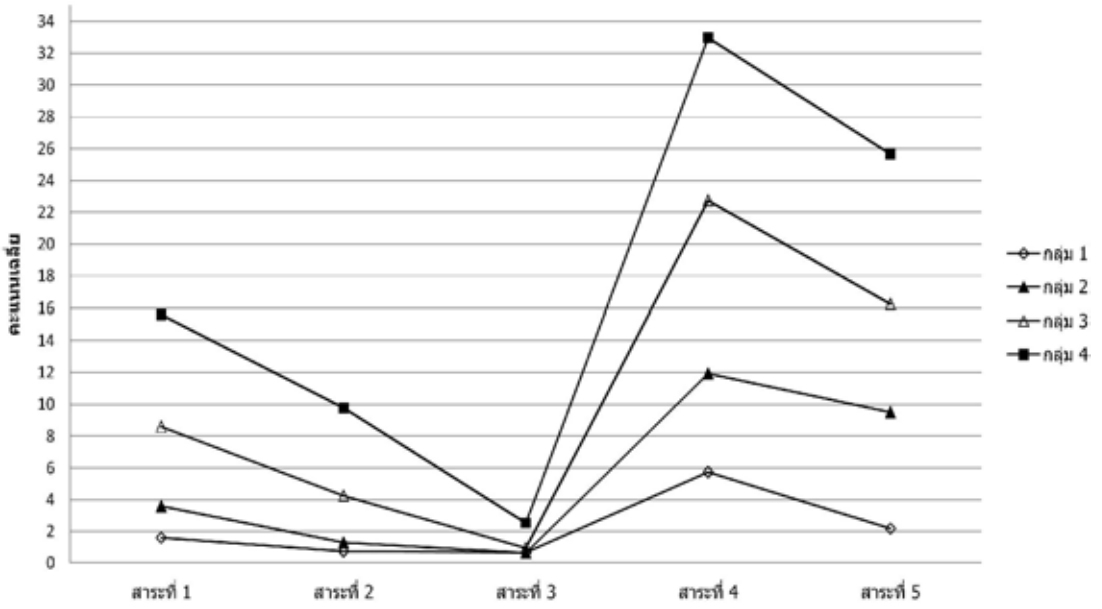
ภาคผนวก ฉ

1. สารสนเทศเกี่ยวกับผลการวิเคราะห์กลุ่ม (Clustering – Result: CR)
2. สารสนเทศเกี่ยวกับการแปลงค่า Theta เป็นคะแนนจุดตัด

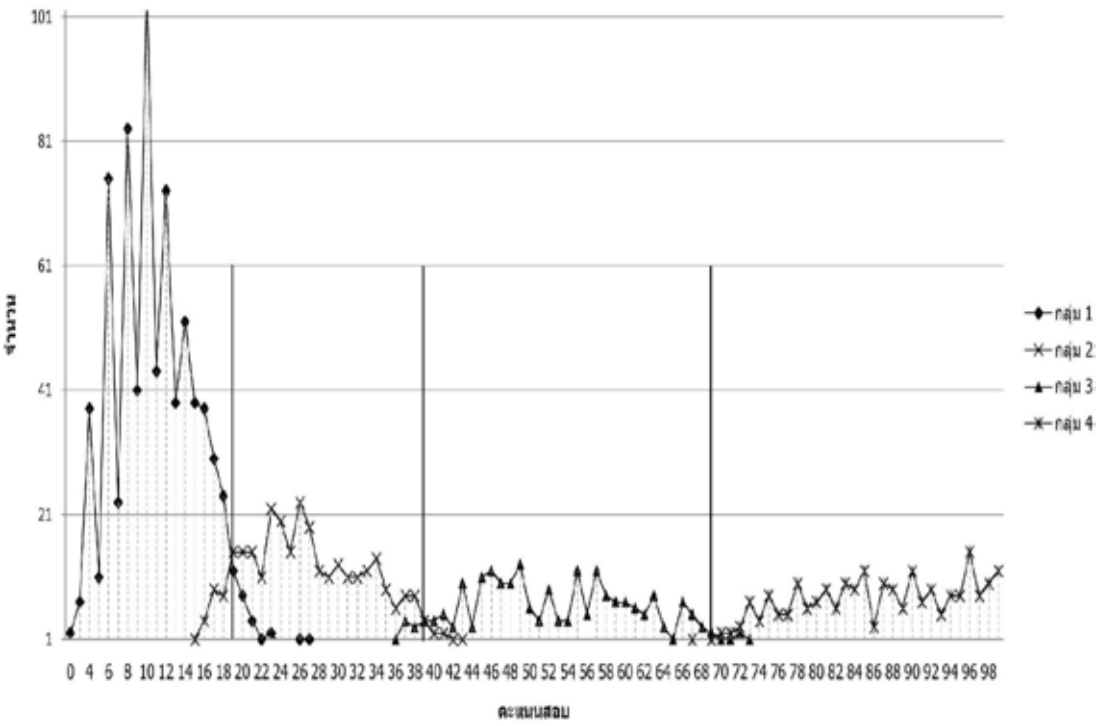
สารสนเทศเกี่ยวกับผลการวิเคราะห์กลุ่ม (Clustering – Result: CR)

วิชาคณิตศาสตร์

1. ผลคะแนนเฉลี่ยรายสาระการเรียนรู้ของแต่ละกลุ่ม



2. ช่วงของคะแนนสอบที่คาบเกี่ยวกันของแต่ละกลุ่ม



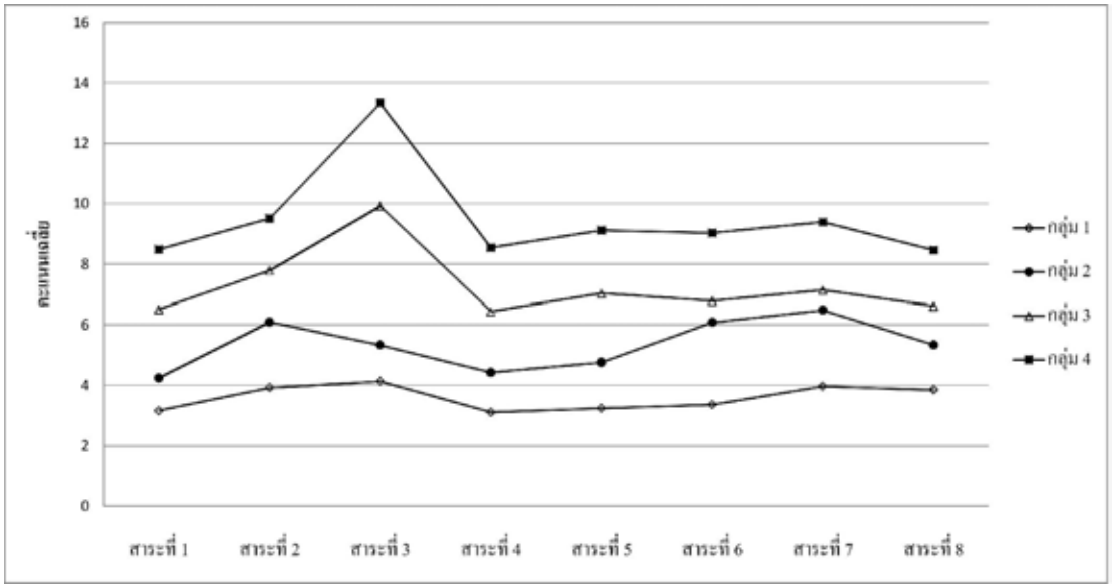
3. ผลการแจกแจงความถี่คะแนนแต่ละกลุ่ม

score	G 1	G 2	G 3	G4	total	score	G 1	G 2	G 3	G 4	total	score	G1	G2	G3	G4	total
0	2	0	0	0	2	35	0	9	0	0	9	68	0	0	3	0	3
2	7	0	0	0	7	36	0	6	1	0	7	69	0	0	2	1	3
4	38	0	0	0	38	37	0	8	4	0	12	70	0	0	1	2	3
5	11	0	0	0	11	38	0	8	3	0	11	71	0	0	1	2	3
6	75	0	0	0	75	39	0	4	4	0	8	72	0	0	2	3	5
7	23	0	0	0	23	40	0	2	4	0	6	73	0	0	1	7	8
8	83	0	0	0	83	41	0	2	5	0	7	74	0	0	0	4	4
9	41	0	0	0	41	42	0	1	3	0	4	75	0	0	0	8	8
10	105	0	0	0	105	43	0	1	10	0	11	76	0	0	0	5	5
11	44	0	0	0	44	44	0	0	3	0	3	77	0	0	0	5	5
12	73	0	0	0	73	45	0	0	11	0	11	78	0	0	0	10	10
13	39	0	0	0	39	46	0	0	12	0	12	79	0	0	0	6	6
14	52	0	0	0	52	47	0	0	10	0	10	80	0	0	0	7	7
15	39	1	0	0	40	48	0	0	10	0	10	81	0	0	0	9	9
16	38	4	0	0	42	49	0	0	13	0	13	82	0	0	0	6	6
17	30	9	0	0	39	50	0	0	6	0	6	83	0	0	0	10	10
18	24	8	0	0	32	51	0	0	4	0	4	84	0	0	0	9	9
19	12	15	0	0	27	52	0	0	9	0	9	85	0	0	0	12	12
20	8	15	0	0	23	53	0	0	4	0	4	86	0	0	0	3	3
21	4	15	0	0	19	54	0	0	4	0	4	87	0	0	0	10	10
22	1	11	0	0	12	55	0	0	12	0	12	88	0	0	0	9	9
23	2	22	0	0	24	56	0	0	5	0	5	89	0	0	0	6	6
24	0	20	0	0	20	57	0	0	12	0	12	90	0	0	0	12	12
25	0	15	0	0	15	58	0	0	8	0	8	91	0	0	0	7	7
26	1	23	0	0	24	59	0	0	7	0	7	92	0	0	0	9	9
27	1	19	0	0	20	60	0	0	7	0	7	93	0	0	0	5	5
28	0	12	0	0	12	61	0	0	6	0	6	94	0	0	0	8	8
29	0	11	0	0	11	62	0	0	5	0	5	95	0	0	0	8	8
30	0	13	0	0	13	63	0	0	8	0	8	96	0	0	0	15	15
31	0	11	0	0	11	64	0	0	3	0	3	97	0	0	0	8	8
32	0	11	0	0	11	65	0	0	1	0	1	98	0	0	0	10	10
33	0	12	0	0	12	66	0	0	7	0	7	100	0	0	0	12	12
34	0	14	0	0	14	67	0	0	5	1	6	sum	753	302	216	229	1500

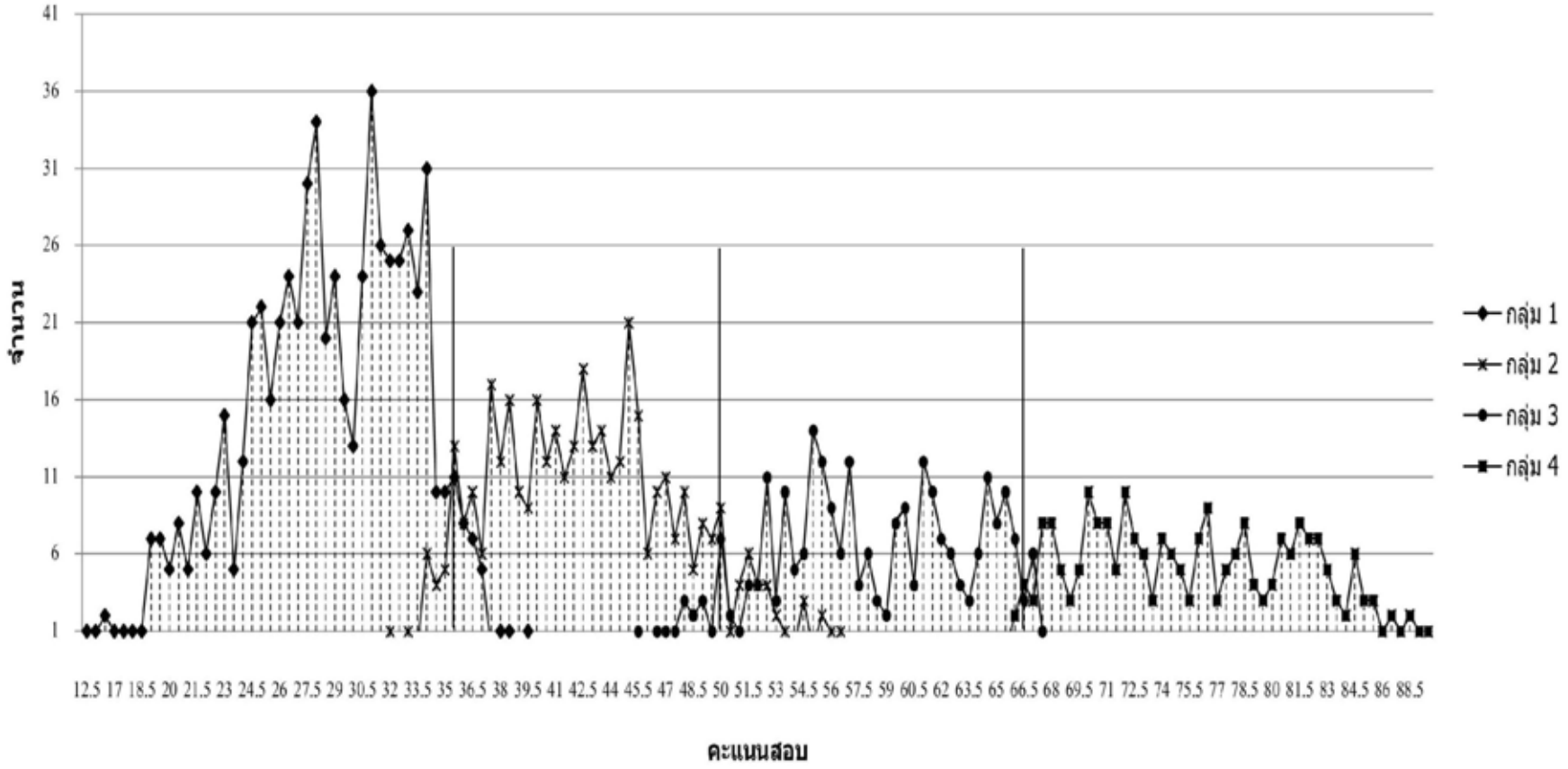
สารสนเทศเกี่ยวกับผลการวิเคราะห์กลุ่ม (Clustering – Result: CR)

วิชาวิทยาศาสตร์

1. ผลคะแนนเฉลี่ยรายสาระการเรียนรู้ของแต่ละกลุ่ม



2. ช่วงของคะแนนสอบที่คาบเกี่ยวกันของแต่ละกลุ่ม



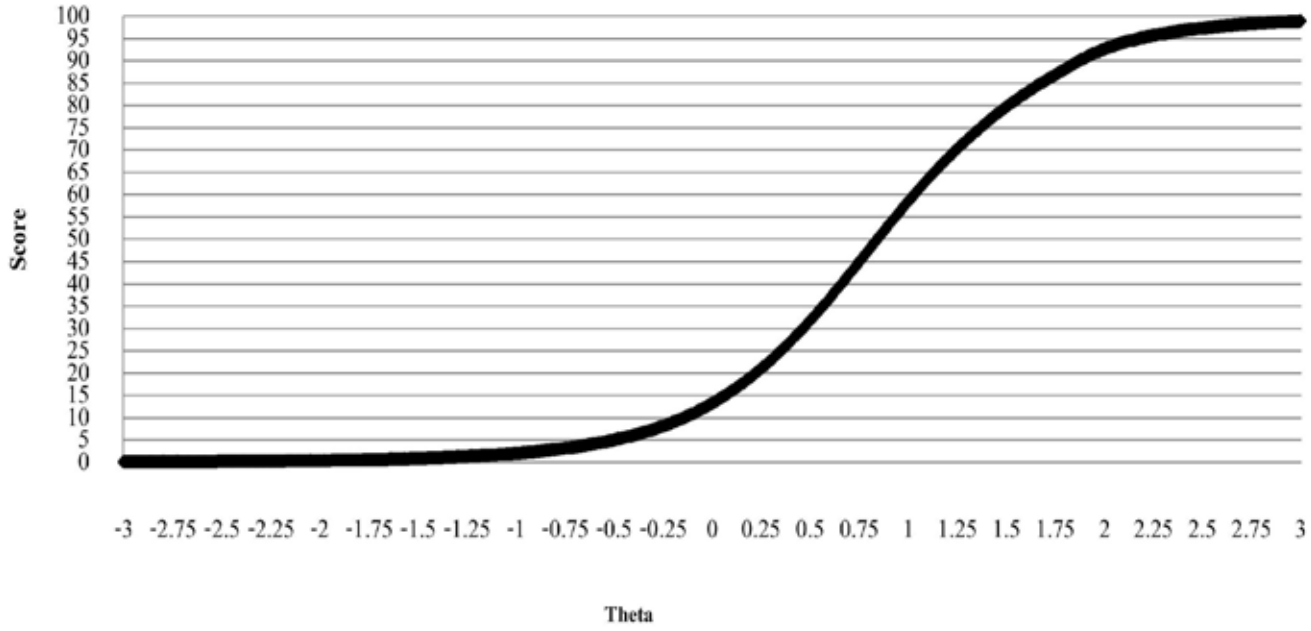
3. ผลการแจกแจงความถี่คะแนนแต่ละกลุ่ม

score	G 1	G 2	G 3	G4	total	scor	G 1	G 2	G 3	G 4	total	scor	G1	G2	G3	G4	total
12.5	1	0	0	0	1	40.5	0	12	0	0	12	65.5	0	0	10	0	10
15.5	1	0	0	0	1	41	0	14	0	0	14	66	0	0	7	2	9
16	2	0	0	0	2	41.5	0	11	0	0	11	66.5	0	0	3	4	7
17	1	0	0	0	1	42	0	13	0	0	13	67	0	0	6	3	9
17.5	1	0	0	0	1	42.5	0	18	0	0	18	67.5	0	0	1	8	9
18	1	0	0	0	1	43	0	13	0	0	13	68	0	0	0	8	8
18.5	1	0	0	0	1	43.5	0	14	0	0	14	68.5	0	0	0	5	5
19	7	0	0	0	7	44	0	11	0	0	11	69	0	0	0	3	3
19.5	7	0	0	0	7	44.5	0	12	0	0	12	69.5	0	0	0	5	5
20	5	0	0	0	5	45	0	21	0	0	21	70	0	0	0	10	10
20.5	8	0	0	0	8	45.5	0	15	1	0	16	70.5	0	0	0	8	8
21	5	0	0	0	5	46	0	6	0	0	6	71	0	0	0	8	8
21.5	10	0	0	0	10	46.5	0	10	1	0	11	71.5	0	0	0	5	5
22	6	0	0	0	6	47	0	11	1	0	12	72	0	0	0	10	10
22.5	10	0	0	0	10	47.5	0	7	1	0	8	72.5	0	0	0	7	7
23	15	0	0	0	15	48	0	10	3	0	13	73	0	0	0	6	6
23.5	5	0	0	0	5	48.5	0	5	2	0	7	73.5	0	0	0	3	3
24	12	0	0	0	12	49	0	8	3	0	11	74	0	0	0	7	7
24.5	21	0	0	0	21	49.5	0	7	1	0	8	74.5	0	0	0	6	6
25	22	0	0	0	22	50	0	9	7	0	16	75	0	0	0	5	5
25.5	16	0	0	0	16	50.5	0	1	2	0	3	75.5	0	0	0	3	3
26	21	0	0	0	21	51	0	4	1	0	5	76	0	0	0	7	7
26.5	24	0	0	0	24	51.5	0	6	4	0	10	76.5	0	0	0	9	9
27	21	0	0	0	21	52	0	4	4	0	8	77	0	0	0	3	3
27.5	30	0	0	0	30	52.5	0	4	11	0	15	77.5	0	0	0	5	5
28	34	0	0	0	34	53	0	2	3	0	5	78	0	0	0	6	6
28.5	20	0	0	0	20	53.5	0	1	10	0	11	78.5	0	0	0	8	8
29	24	0	0	0	24	54	0	0	5	0	5	79	0	0	0	4	4
29.5	16	0	0	0	16	54.5	0	3	6	0	9	79.5	0	0	0	3	3
30	13	0	0	0	13	55	0	0	14	0	14	80	0	0	0	4	4
30.5	24	0	0	0	24	55.5	0	2	12	0	14	80.5	0	0	0	7	7
31	36	0	0	0	36	56	0	1	9	0	10	81	0	0	0	6	6
31.5	26	0	0	0	26	56.5	0	1	6	0	7	81.5	0	0	0	8	8
32	25	1	0	0	26	57	0	0	12	0	12	82	0	0	0	7	7
32.5	25	0	0	0	25	57.5	0	0	4	0	4	82.5	0	0	0	7	7
33	27	1	0	0	28	58	0	0	6	0	6	83	0	0	0	5	5
33.5	23	0	0	0	23	58.5	0	0	3	0	3	83.5	0	0	0	3	3
34	31	6	0	0	37	59	0	0	2	0	2	84	0	0	0	2	2
34.5	10	4	0	0	14	59.5	0	0	8	0	8	84.5	0	0	0	6	6
35	10	5	0	0	15	60	0	0	9	0	9	85	0	0	0	3	3
35.5	11	13	0	0	24	60.5	0	0	4	0	4	85.5	0	0	0	3	3
36	8	8	0	0	16	61	0	0	12	0	12	86	0	0	0	1	1
36.5	7	10	0	0	17	61.5	0	0	10	0	10	87.5	0	0	0	2	2
37	5	6	0	0	11	62	0	0	7	0	7	88	0	0	0	1	1
37.5	0	17	0	0	17	62.5	0	0	6	0	6	88.5	0	0	0	2	2
38	1	12	0	0	13	63	0	0	4	0	4	90	0	0	0	1	1
38.5	1	16	0	0	17	63.5	0	0	3	0	3	92	0	0	0	1	1
39	0	10	0	0	10	64	0	0	6	0	6	sum	631	390	249	230	1500
39.5	1	9	0	0	10	64.5	0	0	11	0	11						
40	0	16	0	0	16	65	0	0	8	0	8						

สารสนเทศเกี่ยวกับการแปลงค่าความสามารถ (Theta) เป็นคะแนนจุดตัด

วิชาคณิตศาสตร์

Test Characteristic Curve



ตารางแสดงการแปลงค่าความสามารถ (Theta) เป็นคะแนนจุดตัดวิชาคณิตศาสตร์

Theta	score	Theta	score	Theta	score	Theta	score	Theta	score	Theta	score	Theta	score	Theta	score	Theta	score	Theta	score
-3	0.106	-2.75	0.149	-2.5	0.210	-2.25	0.299	-2.16	0.340	-1.91	0.489	-1.66	0.711	-1.41	1.049	-1.16	1.572	-0.91	2.405
-2.99	0.107	-2.74	0.151	-2.49	0.213	-2.24	0.303	-2.15	0.345	-1.9	0.496	-1.65	0.722	-1.4	1.066	-1.15	1.599	-0.9	2.448
-2.98	0.109	-2.73	0.153	-2.48	0.216	-2.23	0.308	-2.14	0.350	-1.89	0.504	-1.64	0.733	-1.39	1.083	-1.14	1.626	-0.89	2.491
-2.97	0.110	-2.72	0.155	-2.47	0.219	-2.22	0.312	-2.13	0.355	-1.88	0.511	-1.63	0.745	-1.38	1.100	-1.13	1.653	-0.88	2.535
-2.96	0.112	-2.71	0.157	-2.46	0.222	-2.21	0.317	-2.12	0.360	-1.87	0.519	-1.62	0.756	-1.37	1.118	-1.12	1.681	-0.87	2.580
-2.95	0.113	-2.7	0.160	-2.45	0.226	-2.2	0.321	-2.11	0.366	-1.86	0.527	-1.61	0.768	-1.36	1.136	-1.11	1.709	-0.86	2.626
-2.94	0.115	-2.69	0.162	-2.44	0.229	-2.19	0.326	-2.1	0.371	-1.85	0.535	-1.6	0.780	-1.35	1.154	-1.1	1.738	-0.85	2.672
-2.93	0.117	-2.68	0.164	-2.43	0.232	-2.18	0.331	-2.09	0.376	-1.84	0.543	-1.59	0.792	-1.34	1.173	-1.09	1.767	-0.84	2.720
-2.92	0.118	-2.67	0.166	-2.42	0.235	-2.17	0.335	-2.08	0.382	-1.83	0.551	-1.58	0.804	-1.33	1.192	-1.08	1.797	-0.83	2.768
-2.91	0.120	-2.66	0.169	-2.41	0.239	-2.16	0.340	-2.07	0.387	-1.82	0.559	-1.57	0.817	-1.32	1.211	-1.07	1.828	-0.82	2.818
-2.9	0.121	-2.65	0.171	-2.4	0.242	-2.15	0.345	-2.06	0.393	-1.81	0.567	-1.56	0.829	-1.31	1.231	-1.06	1.859	-0.81	2.869
-2.89	0.123	-2.64	0.173	-2.39	0.245	-2.14	0.350	-2.05	0.399	-1.8	0.576	-1.55	0.842	-1.3	1.251	-1.05	1.891	-0.8	2.920
-2.88	0.125	-2.63	0.176	-2.38	0.249	-2.13	0.355	-2.04	0.404	-1.79	0.585	-1.54	0.855	-1.29	1.271	-1.04	1.923	-0.79	2.973
-2.87	0.126	-2.62	0.178	-2.37	0.252	-2.12	0.360	-2.03	0.410	-1.78	0.593	-1.53	0.869	-1.28	1.292	-1.03	1.956	-0.78	3.026
-2.86	0.128	-2.61	0.181	-2.36	0.256	-2.11	0.366	-2.02	0.416	-1.77	0.602	-1.52	0.882	-1.27	1.313	-1.02	1.990	-0.77	3.081
-2.85	0.130	-2.6	0.183	-2.35	0.260	-2.1	0.371	-2.01	0.423	-1.76	0.612	-1.51	0.896	-1.26	1.334	-1.01	2.024	-0.76	3.137
-2.84	0.132	-2.59	0.186	-2.34	0.263	-2.25	0.299	-2	0.429	-1.75	0.621	-1.5	0.910	-1.25	1.356	-1	2.059	-0.75	3.194
-2.83	0.134	-2.58	0.188	-2.33	0.267	-2.24	0.303	-1.99	0.435	-1.74	0.630	-1.49	0.925	-1.24	1.379	-0.99	2.094	-0.74	3.252
-2.82	0.135	-2.57	0.191	-2.32	0.271	-2.23	0.308	-1.98	0.441	-1.73	0.640	-1.48	0.939	-1.23	1.401	-0.98	2.131	-0.73	3.312
-2.81	0.137	-2.56	0.194	-2.31	0.275	-2.22	0.312	-1.97	0.448	-1.72	0.650	-1.47	0.954	-1.22	1.424	-0.97	2.168	-0.72	3.372
-2.8	0.139	-2.55	0.196	-2.3	0.279	-2.21	0.317	-1.96	0.455	-1.71	0.659	-1.46	0.969	-1.21	1.448	-0.96	2.205	-0.71	3.434
-2.79	0.141	-2.54	0.199	-2.29	0.283	-2.2	0.321	-1.95	0.461	-1.7	0.669	-1.45	0.985	-1.2	1.472	-0.95	2.244	-0.7	3.497
-2.78	0.143	-2.53	0.202	-2.28	0.287	-2.19	0.326	-1.94	0.468	-1.69	0.680	-1.44	1.000	-1.19	1.496	-0.94	2.283	-0.69	3.561
-2.77	0.145	-2.52	0.205	-2.27	0.291	-2.18	0.331	-1.93	0.475	-1.68	0.690	-1.43	1.016	-1.18	1.521	-0.93	2.323	-0.68	3.627
-2.76	0.147	-2.51	0.207	-2.26	0.295	-2.17	0.335	-1.92	0.482	-1.67	0.701	-1.42	1.032	-1.17	1.547	-0.92	2.364	-0.67	3.694

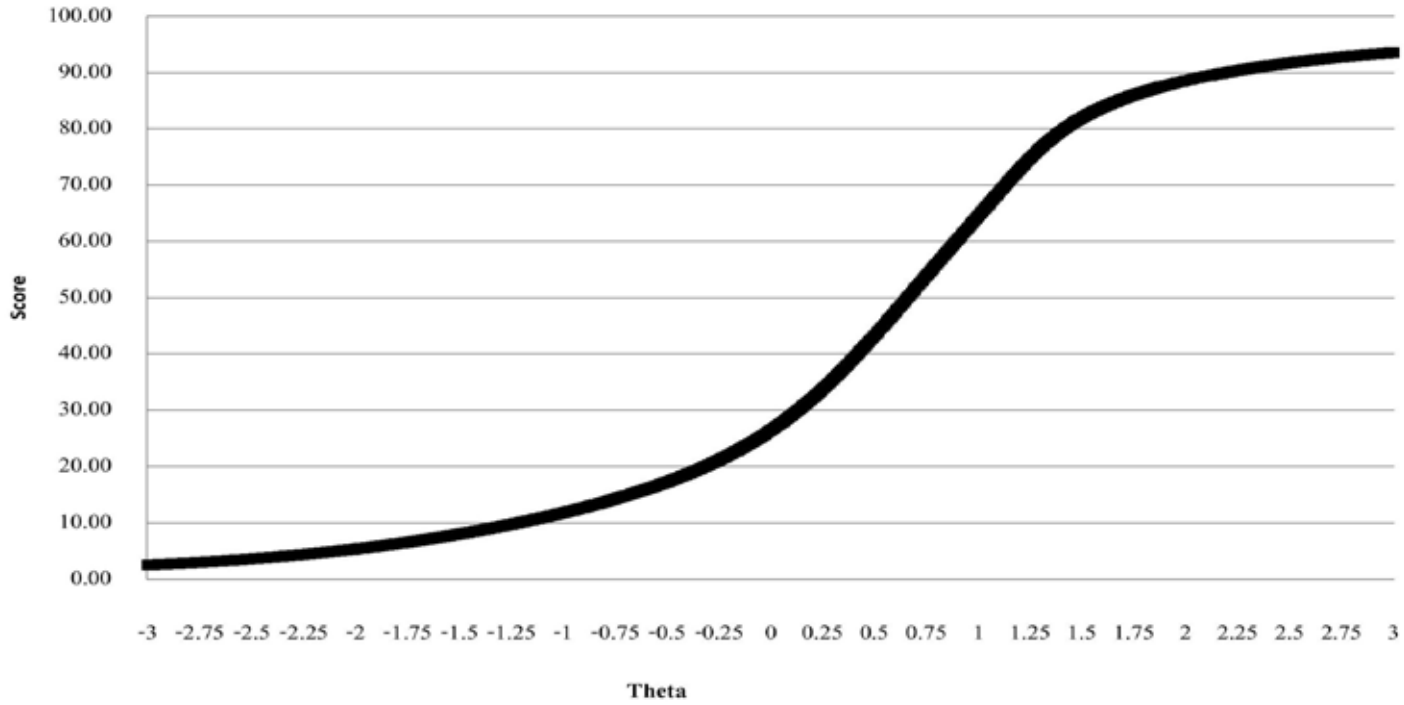
Theta	score	Theta	score	Theta	score	Theta	score	Theta	score	Theta	score	Theta	score	Theta	score	Theta	score	Theta	score
-0.66	3.762	-0.37	6.496	-0.08	11.407	0.21	19.664	0.5	31.868	0.79	47.246	1.08	62.625	1.37	75.187	1.66	84.571	1.95	91.790
-0.65	3.832	-0.36	6.622	-0.07	11.630	0.22	20.019	0.51	32.357	0.8	47.800	1.09	63.112	1.38	75.562	1.67	84.847	1.96	91.985
-0.64	3.903	-0.35	6.751	-0.06	11.857	0.23	20.378	0.52	32.849	0.81	48.353	1.1	63.596	1.39	75.933	1.68	85.122	1.97	92.175
-0.63	3.976	-0.34	6.883	-0.05	12.088	0.24	20.742	0.53	33.345	0.82	48.906	1.11	64.075	1.4	76.301	1.69	85.395	1.98	92.359
-0.62	4.050	-0.33	7.017	-0.04	12.324	0.25	21.111	0.54	33.846	0.83	49.459	1.12	64.551	1.41	76.664	1.7	85.666	1.99	92.538
-0.61	4.126	-0.32	7.154	-0.03	12.563	0.26	21.485	0.55	34.350	0.84	50.011	1.13	65.023	1.42	77.024	1.71	85.935	2	92.711
-0.6	4.203	-0.31	7.294	-0.02	12.807	0.27	21.864	0.56	34.857	0.85	50.561	1.14	65.492	1.43	77.379	1.72	86.203	2.01	92.880
-0.59	4.282	-0.3	7.437	-0.01	13.055	0.28	22.247	0.57	35.368	0.86	51.111	1.15	65.956	1.44	77.731	1.73	86.470	2.02	93.043
-0.58	4.362	-0.29	7.582	0	13.308	0.29	22.635	0.58	35.883	0.87	51.659	1.16	66.417	1.45	78.079	1.74	86.735	2.03	93.201
-0.57	4.444	-0.28	7.731	0.01	13.565	0.3	23.028	0.59	36.401	0.88	52.206	1.17	66.873	1.46	78.423	1.75	86.999	2.04	93.354
-0.56	4.528	-0.27	7.883	0.02	13.826	0.31	23.426	0.6	36.922	0.89	52.751	1.18	67.326	1.47	78.764	1.76	87.262	2.05	93.503
-0.55	4.614	-0.26	8.037	0.03	14.092	0.32	23.829	0.61	37.447	0.9	53.295	1.19	67.775	1.48	79.101	1.77	87.523	2.06	93.647
-0.54	4.701	-0.25	8.195	0.04	14.362	0.33	24.236	0.62	37.974	0.91	53.836	1.2	68.220	1.49	79.433	1.78	87.784	2.07	93.787
-0.53	4.790	-0.24	8.356	0.05	14.637	0.34	24.648	0.63	38.504	0.92	54.376	1.21	68.661	1.5	79.763	1.79	88.043	2.08	93.923
-0.52	4.881	-0.23	8.520	0.06	14.916	0.35	25.065	0.64	39.037	0.93	54.913	1.22	69.098	1.51	80.088	1.8	88.300	2.09	94.055
-0.51	4.974	-0.22	8.688	0.07	15.200	0.36	25.487	0.65	39.573	0.94	55.448	1.23	69.531	1.52	80.410	1.81	88.556	2.1	94.184
-0.5	5.069	-0.21	8.859	0.08	15.488	0.37	25.913	0.66	40.110	0.95	55.980	1.24	69.960	1.53	80.728	1.82	88.810	2.11	94.308
-0.49	5.165	-0.2	9.033	0.09	15.781	0.38	26.344	0.67	40.651	0.96	56.510	1.25	70.386	1.54	81.043	1.83	89.061	2.12	94.429
-0.48	5.264	-0.19	9.211	0.1	16.079	0.39	26.779	0.68	41.193	0.97	57.037	1.26	70.807	1.55	81.354	1.84	89.310	2.13	94.547
-0.47	5.365	-0.18	9.392	0.11	16.381	0.4	27.220	0.69	41.737	0.98	57.561	1.27	71.225	1.56	81.662	1.85	89.556	2.14	94.662
-0.46	5.468	-0.17	9.576	0.12	16.688	0.41	27.664	0.7	42.283	0.99	58.082	1.28	71.639	1.57	81.966	1.86	89.799	2.15	94.774
-0.45	5.573	-0.16	9.765	0.13	17.000	0.42	28.114	0.71	42.830	1	58.600	1.29	72.048	1.58	82.267	1.87	90.038	2.16	94.882
-0.44	5.680	-0.15	9.957	0.14	17.316	0.43	28.568	0.72	43.379	1.01	59.115	1.3	72.454	1.59	82.565	1.88	90.274	2.17	94.988
-0.43	5.790	-0.14	10.152	0.15	17.638	0.44	29.026	0.73	43.929	1.02	59.627	1.31	72.856	1.6	82.860	1.89	90.505	2.18	95.091
-0.42	5.902	-0.13	10.352	0.16	17.963	0.45	29.489	0.74	44.481	1.03	60.135	1.32	73.255	1.61	83.152	1.9	90.732	2.19	95.192
-0.41	6.016	-0.12	10.555	0.17	18.294	0.46	29.956	0.75	45.033	1.04	60.640	1.33	73.649	1.62	83.441	1.91	90.954	2.2	95.290
-0.4	6.132	-0.11	10.762	0.18	18.630	0.47	30.428	0.76	45.585	1.05	61.142	1.34	74.039	1.63	83.727	1.92	91.171	2.21	95.386
-0.39	6.251	-0.1	10.973	0.19	18.970	0.48	30.904	0.77	46.139	1.06	61.640	1.35	74.426	1.64	84.011	1.93	91.383	2.22	95.479
-0.38	6.372	-0.09	11.188	0.2	19.315	0.49	31.384	0.78	46.692	1.07	62.134	1.36	74.809	1.65	84.292	1.94	91.589	2.23	95.571

Theta	score	Theta	score	Theta	score	Theta	score	Theta	score	Theta	score	Theta	score	Theta	score	Theta	score	Theta	score
2.25	95.747	2.52	97.491	2.79	98.483														
2.26	95.831	2.53	97.538	2.8	98.510														
2.27	95.914	2.54	97.584	2.81	98.537														
2.28	95.996	2.55	97.630	2.82	98.563														
2.29	96.075	2.56	97.674	2.83	98.589														
2.3	96.152	2.57	97.718	2.84	98.614														
2.31	96.228	2.58	97.760	2.85	98.639														
2.32	96.302	2.59	97.802	2.86	98.664														
2.33	96.374	2.6	97.843	2.87	98.687														
2.34	96.445	2.61	97.883	2.88	98.711														
2.35	96.515	2.62	97.922	2.89	98.734														
2.36	96.582	2.63	97.961	2.9	98.756														
2.37	96.649	2.64	97.999	2.91	98.778														
2.38	96.714	2.65	98.036	2.92	98.800														
2.39	96.777	2.66	98.072	2.93	98.821														
2.4	96.839	2.67	98.107	2.94	98.842														
2.41	96.900	2.68	98.142	2.95	98.863														
2.42	96.960	2.69	98.176	2.96	98.883														
2.43	97.018	2.7	98.210	2.97	98.903														
2.44	97.075	2.71	98.243	2.98	98.922														
2.45	97.131	2.72	98.275	2.99	98.941														
2.46	97.186	2.73	98.306	3	98.959														
2.47	97.239	2.74	98.337																
2.48	97.292	2.75	98.367																
2.49	97.343	2.76	98.397																
2.5	97.393	2.77	98.426																
2.51	97.443	2.78	98.455																

สารสนเทศเกี่ยวกับการแปลงค่าความสามารถ (Theta) เป็นคะแนนจุดตัด

วิชาวิทยาศาสตร์

Test Characteristic Curve



ตารางแสดงการแปลงค่า ความสามารถ (Theta) เป็นคะแนนจุดตัดวิชาวิทยาศาสตร์

Theta	score	Theta	score	Theta	score	Theta	score	Theta	score	Theta	score	Theta	score	Theta	score	Theta	score	Theta	score
-3	2.57	-2.75	3.04	-2.5	3.65	-2.25	4.42	-2	5.40	-1.75	6.59	-1.5	8.05	-1.25	9.77	-1	11.82	-0.75	14.25
-2.99	2.59	-2.74	3.07	-2.49	3.68	-2.24	4.46	-1.99	5.44	-1.74	6.65	-1.49	8.11	-1.24	9.85	-0.99	11.90	-0.74	14.36
-2.98	2.60	-2.73	3.09	-2.48	3.71	-2.23	4.49	-1.98	5.48	-1.73	6.70	-1.48	8.17	-1.23	9.93	-0.98	11.99	-0.73	14.47
-2.97	2.62	-2.72	3.11	-2.47	3.74	-2.22	4.53	-1.97	5.53	-1.72	6.76	-1.47	8.24	-1.22	10.00	-0.97	12.09	-0.72	14.58
-2.96	2.64	-2.71	3.13	-2.46	3.76	-2.21	4.57	-1.96	5.57	-1.71	6.81	-1.46	8.30	-1.21	10.08	-0.96	12.18	-0.71	14.69
-2.95	2.65	-2.7	3.15	-2.45	3.79	-2.2	4.60	-1.95	5.62	-1.7	6.86	-1.45	8.37	-1.2	10.16	-0.95	12.27	-0.7	14.80
-2.94	2.67	-2.69	3.18	-2.44	3.82	-2.19	4.64	-1.94	5.66	-1.69	6.92	-1.44	8.43	-1.19	10.23	-0.94	12.36	-0.69	14.92
-2.93	2.69	-2.68	3.20	-2.43	3.85	-2.18	4.68	-1.93	5.71	-1.68	6.97	-1.43	8.50	-1.18	10.31	-0.93	12.45	-0.68	15.03
-2.92	2.71	-2.67	3.22	-2.42	3.88	-2.17	4.71	-1.92	5.75	-1.67	7.03	-1.42	8.57	-1.17	10.39	-0.92	12.55	-0.67	15.14
-2.91	2.73	-2.66	3.25	-2.41	3.91	-2.16	4.75	-1.91	5.80	-1.66	7.09	-1.41	8.64	-1.16	10.47	-0.91	12.64	-0.66	15.26
-2.9	2.74	-2.65	3.27	-2.4	3.94	-2.15	4.79	-1.9	5.85	-1.65	7.14	-1.4	8.70	-1.15	10.55	-0.9	12.74	-0.65	15.38
-2.89	2.76	-2.64	3.29	-2.39	3.97	-2.14	4.83	-1.89	5.89	-1.64	7.20	-1.39	8.77	-1.14	10.63	-0.89	12.83	-0.64	15.49
-2.88	2.78	-2.63	3.32	-2.38	4.00	-2.13	4.86	-1.88	5.94	-1.63	7.26	-1.38	8.84	-1.13	10.71	-0.88	12.93	-0.63	15.61
-2.87	2.80	-2.62	3.34	-2.37	4.03	-2.12	4.90	-1.87	5.99	-1.62	7.32	-1.37	8.91	-1.12	10.79	-0.87	13.03	-0.62	15.73
-2.86	2.82	-2.61	3.37	-2.36	4.06	-2.11	4.94	-1.86	6.04	-1.61	7.38	-1.36	8.98	-1.11	10.88	-0.86	13.12	-0.61	15.85
-2.85	2.84	-2.6	3.39	-2.35	4.09	-2.1	4.98	-1.85	6.09	-1.6	7.43	-1.35	9.05	-1.1	10.96	-0.85	13.22	-0.6	15.97
-2.84	2.86	-2.59	3.42	-2.34	4.13	-2.09	5.02	-1.84	6.14	-1.59	7.49	-1.34	9.12	-1.09	11.04	-0.84	13.32	-0.59	16.10
-2.83	2.88	-2.58	3.44	-2.33	4.16	-2.08	5.06	-1.83	6.18	-1.58	7.55	-1.33	9.19	-1.08	11.12	-0.83	13.42	-0.58	16.22
-2.82	2.90	-2.57	3.47	-2.32	4.19	-2.07	5.10	-1.82	6.23	-1.57	7.61	-1.32	9.26	-1.07	11.21	-0.82	13.52	-0.57	16.35
-2.81	2.92	-2.56	3.49	-2.31	4.22	-2.06	5.14	-1.81	6.28	-1.56	7.67	-1.31	9.33	-1.06	11.29	-0.81	13.63	-0.56	16.47
-2.8	2.94	-2.55	3.52	-2.3	4.26	-2.05	5.18	-1.8	6.34	-1.55	7.73	-1.3	9.41	-1.05	11.38	-0.8	13.73	-0.55	16.60
-2.79	2.96	-2.54	3.54	-2.29	4.29	-2.04	5.23	-1.79	6.39	-1.54	7.80	-1.29	9.48	-1.04	11.47	-0.79	13.83	-0.54	16.73
-2.78	2.98	-2.53	3.57	-2.28	4.32	-2.03	5.27	-1.78	6.44	-1.53	7.86	-1.28	9.55	-1.03	11.55	-0.78	13.94	-0.53	16.86
-2.77	3.00	-2.52	3.60	-2.27	4.36	-2.02	5.31	-1.77	6.49	-1.52	7.92	-1.27	9.63	-1.02	11.64	-0.77	14.04	-0.52	16.99
-2.76	3.02	-2.51	3.62	-2.26	4.39	-2.01	5.35	-1.76	6.54	-1.51	7.98	-1.26	9.70	-1.01	11.73	-0.76	14.15	-0.51	17.13

Theta	score	Theta	score	Theta	score	Theta	score	Theta	score	Theta	score	Theta	score	Theta	score	Theta	score	Theta	score
-0.5	17.26	-0.21	21.88	0.08	28.51	0.37	38.04	0.66	50.04	0.95	62.38	1.24	74.39	1.53	82.56	1.82	86.82	2.11	89.47
-0.49	17.40	-0.2	22.07	0.09	28.79	0.38	38.42	0.67	50.47	0.96	62.81	1.25	74.76	1.54	82.75	1.83	86.93	2.12	89.55
-0.48	17.53	-0.19	22.26	0.1	29.07	0.39	38.81	0.68	50.90	0.97	63.23	1.26	75.13	1.55	82.94	1.84	87.04	2.13	89.62
-0.47	17.67	-0.18	22.46	0.11	29.35	0.4	39.20	0.69	51.33	0.98	63.65	1.27	75.49	1.56	83.12	1.85	87.15	2.14	89.69
-0.46	17.81	-0.17	22.66	0.12	29.64	0.41	39.59	0.7	51.76	0.99	64.07	1.28	75.85	1.57	83.30	1.86	87.25	2.15	89.76
-0.45	17.96	-0.16	22.85	0.13	29.93	0.42	39.98	0.71	52.19	1	64.50	1.29	76.20	1.58	83.48	1.87	87.36	2.16	89.83
-0.44	18.10	-0.15	23.06	0.14	30.23	0.43	40.38	0.72	52.62	1.01	64.92	1.3	76.54	1.59	83.65	1.88	87.46	2.17	89.90
-0.43	18.24	-0.14	23.26	0.15	30.53	0.44	40.78	0.73	53.05	1.02	65.34	1.31	76.88	1.6	83.82	1.89	87.56	2.18	89.97
-0.42	18.39	-0.13	23.47	0.16	30.83	0.45	41.18	0.74	53.48	1.03	65.77	1.32	77.20	1.61	83.98	1.9	87.66	2.19	90.04
-0.41	18.54	-0.12	23.68	0.17	31.13	0.46	41.59	0.75	53.91	1.04	66.19	1.33	77.53	1.62	84.14	1.91	87.76	2.2	90.11
-0.4	18.69	-0.11	23.89	0.18	31.44	0.47	42.00	0.76	54.34	1.05	66.61	1.34	77.84	1.63	84.30	1.92	87.86	2.21	90.18
-0.39	18.84	-0.1	24.11	0.19	31.76	0.48	42.41	0.77	54.76	1.06	67.04	1.35	78.15	1.64	84.46	1.93	87.95	2.22	90.24
-0.38	18.99	-0.09	24.32	0.2	32.08	0.49	42.82	0.78	55.19	1.07	67.46	1.36	78.45	1.65	84.61	1.94	88.05	2.23	90.31
-0.37	19.15	-0.08	24.55	0.21	32.40	0.5	43.23	0.79	55.62	1.08	67.88	1.37	78.74	1.66	84.76	1.95	88.14	2.24	90.37
-0.36	19.30	-0.07	24.77	0.22	32.72	0.51	43.65	0.8	56.04	1.09	68.30	1.38	79.02	1.67	84.91	1.96	88.23	2.25	90.43
-0.35	19.46	-0.06	25.00	0.23	33.05	0.52	44.07	0.81	56.47	1.1	68.73	1.39	79.30	1.68	85.05	1.97	88.32	2.26	90.50
-0.34	19.62	-0.05	25.23	0.24	33.38	0.53	44.49	0.82	56.89	1.11	69.15	1.4	79.58	1.69	85.19	1.98	88.41	2.27	90.56
-0.33	19.78	-0.04	25.46	0.25	33.72	0.54	44.91	0.83	57.32	1.12	69.57	1.41	79.84	1.7	85.33	1.99	88.50	2.28	90.62
-0.32	19.95	-0.03	25.70	0.26	34.06	0.55	45.33	0.84	57.74	1.13	69.98	1.42	80.10	1.71	85.47	2	88.59	2.29	90.68
-0.31	20.11	-0.02	25.94	0.27	34.40	0.56	45.75	0.85	58.16	1.14	70.40	1.43	80.35	1.72	85.60	2.01	88.67	2.3	90.74
-0.3	20.28	-0.01	26.18	0.28	34.75	0.57	46.18	0.86	58.59	1.15	70.81	1.44	80.60	1.73	85.73	2.02	88.76	2.31	90.80
-0.29	20.45	0	26.42	0.29	35.10	0.58	46.60	0.87	59.01	1.16	71.23	1.45	80.84	1.74	85.86	2.03	88.84	2.32	90.86
-0.28	20.62	0.01	26.67	0.3	35.46	0.59	47.03	0.88	59.43	1.17	71.64	1.46	81.07	1.75	85.99	2.04	88.92	2.33	90.91
-0.27	20.79	0.02	26.93	0.31	35.82	0.6	47.46	0.89	59.85	1.18	72.04	1.47	81.30	1.76	86.11	2.05	89.01	2.34	90.97
-0.26	20.97	0.03	27.18	0.32	36.18	0.61	47.89	0.9	60.28	1.19	72.44	1.48	81.52	1.77	86.23	2.06	89.09	2.35	91.03
-0.25	21.15	0.04	27.44	0.33	36.54	0.62	48.32	0.91	60.70	1.2	72.84	1.49	81.74	1.78	86.35	2.07	89.17	2.36	91.08
-0.24	21.33	0.05	27.70	0.34	36.91	0.63	48.75	0.92	61.12	1.21	73.24	1.5	81.95	1.79	86.47	2.08	89.24	2.37	91.14
-0.23	21.51	0.06	27.97	0.35	37.29	0.64	49.18	0.93	61.54	1.22	73.63	1.51	82.16	1.8	86.59	2.09	89.32	2.38	91.19
-0.22	21.70	0.07	28.24	0.36	37.66	0.65	49.61	0.94	61.96	1.23	74.01	1.52	82.36	1.81	86.70	2.1	89.40	2.39	91.24

Theta	score	Theta	score	Theta	score	Theta	score	Theta	score	Theta	score	Theta	score	Theta	score	Theta	score	Theta	score
2.4	91.30	2.67	92.53	2.94	93.45														
2.41	91.35	2.68	92.57	2.95	93.48														
2.42	91.40	2.69	92.61	2.96	93.51														
2.43	91.45	2.7	92.64	2.97	93.54														
2.44	91.50	2.71	92.68	2.98	93.57														
2.45	91.55	2.72	92.72	2.99	93.59														
2.46	91.60	2.73	92.76	3	93.62														
2.47	91.65	2.74	92.79																
2.48	91.70	2.75	92.83																
2.49	91.75	2.76	92.86																
2.5	91.80	2.77	92.90																
2.51	91.84	2.78	92.93																
2.52	91.89	2.79	92.97																
2.53	91.94	2.8	93.00																
2.54	91.98	2.81	93.04																
2.55	92.03	2.82	93.07																
2.56	92.07	2.83	93.10																
2.57	92.11	2.84	93.14																
2.58	92.16	2.85	93.17																
2.59	92.20	2.86	93.20																
2.6	92.24	2.87	93.23																
2.61	92.28	2.88	93.27																
2.62	92.33	2.89	93.30																
2.63	92.37	2.9	93.33																
2.64	92.41	2.91	93.36																
2.65	92.45	2.92	93.39																
2.66	92.49	2.93	93.42																

ภาคผนวก ช

1. รายละเอียดเกี่ยวกับข้อสอบที่วัดในแต่ละสาระการเรียนรู้ วิชาคณิตศาสตร์
2. รายละเอียดเกี่ยวกับข้อสอบที่วัดในแต่ละสาระการเรียนรู้ วิชาวิทยาศาสตร์

รายละเอียดเกี่ยวกับข้อสอบที่วัดในแต่ละสาระการเรียนรู้

วิชาคณิตศาสตร์

ข้อที่	รหัสมาตรฐาน	รหัสสาระ	ลำดับที่	Theta (RP=50)	Theta (RP=0.67)
1	ค 4.2	สาระที่ 4	8	0.53	0.93
2	ค 4.1	สาระที่ 4	5	0.43	0.98
3	ค 4.2	สาระที่ 4	38	1.57	1.82
4	ค 1.3	สาระที่ 1	40	1.86	1.91
5	ค 4.2	สาระที่ 4	33	1.23	1.54
6	ค 1.4	สาระที่ 1	21	0.81	1.14
7	ค 1.4	สาระที่ 1	19	0.79	0.94
8	ค 4.1	สาระที่ 4	1	0.17	0.32
9	ค 4.1	สาระที่ 4	29	1.03	1.28
10	ค 4.1	สาระที่ 4	22	0.81	1.02
11	ค 4.2	สาระที่ 4	18	0.78	0.92
12	ค 4.2	สาระที่ 4	30	1.04	1.35
13	ค 3.2	สาระที่ 3	36	1.4	1.69
14	ค 3.2	สาระที่ 3	39	1.72	1.87
15	ค 2.3	สาระที่ 2	37	1.44	1.56
16	ค 4.2	สาระที่ 4	9	0.59	0.85
17	ค 4.2	สาระที่ 4	12	0.68	0.89
18	ค 5.1	สาระที่ 5	11	0.65	1.15
19	ค 5.2	สาระที่ 5	34	1.24	1.71
20	ค 5.2	สาระที่ 5	3	0.21	0.8
21	ค 4.2	สาระที่ 4	13	0.7	0.89
22	ค 4.2	สาระที่ 4	20	0.79	1.02
23	ค 1.1	สาระที่ 1	17	0.77	1.00
24	ค 1.2	สาระที่ 1	10	0.6	0.83
25	ค 1.2	สาระที่ 1	24	0.86	1.11

ข้อที่	รหัสมาตรฐาน	รหัสสาระ	ลำดับที่	Theta (RP=50)	Theta (RP=0.67)
26	ค 4.2	สาระที่ 4	23	0.82	1.03
27	ค 1.1	สาระที่ 1	32	1.22	1.67
28	ค 2.2	สาระที่ 2	27	0.95	1.21
29	ค 2.3	สาระที่ 2	15	0.74	0.98
30	ค 2.2	สาระที่ 2	35	1.26	1.53
31	ค 4.1	สาระที่ 4	7	0.48	0.79
32	ค 4.1	สาระที่ 4	4	0.41	0.70
33	ค 5.2	สาระที่ 5	31	1.08	1.36
34	ค 5.2	สาระที่ 5	16	0.76	1.03
35	ค 5.2	สาระที่ 5	2	0.2	0.42
36	ค 5.1	สาระที่ 5	25	0.89	1.29
37	ค 5.1	สาระที่ 5	26	0.94	1.36
38	ค 5.1	สาระที่ 5	28	0.96	1.18
39	ค 5.1	สาระที่ 5	14	0.71	0.97
40	ค 5.3	สาระที่ 5	6	0.44	0.71

รายละเอียดเกี่ยวกับข้อสอบที่วัดในแต่ละสาระการเรียนรู้

วิชาวิทยาศาสตร์

ข้อที่	รหัสมาตรฐาน	รหัสสาระ	ลำดับที่	Theta (RP=50)	Theta (RP=0.67)
1	ว 1.1	สาระที่ 1	51	0.58	0.76
2	ว 1.1	สาระที่ 1	44	0.57	0.71
3	ว 1.1	สาระที่ 1	45	0.52	0.71
4	ว 1.1	สาระที่ 1	12	-0.71	-0.23
5	ว 1.1	สาระที่ 1	46	0.56	0.71
6	ว 1.2	สาระที่ 1	11	-0.61	-0.25
7	ว 1.2	สาระที่ 1	49	0.58	0.73
8	ว 1.2	สาระที่ 1	52	0.64	0.76
9	ว 1.2	สาระที่ 1	41	0.37	0.67
10	ว 2.1	สาระที่ 2	22	0.05	0.25
11	ว 2.1	สาระที่ 2	47	0.56	0.71
12	ว 2.1	สาระที่ 2	30	0.3	0.57
13	ว 2.1	สาระที่ 2	13	-0.59	-0.22
14	ว 2.1	สาระที่ 2	54	0.6	0.85
15	ว 2.2	สาระที่ 2	7	-1.91	-1.31
16	ว 2.2	สาระที่ 2	25	-0.31	0.38
17	ว 2.2	สาระที่ 2	43	0.48	0.69
18	ว 2.2	สาระที่ 2	60	0.84	1
19	ว 3.2	สาระที่ 3	69	1	1.19
20	ว 3.2	สาระที่ 3	42	0.52	0.68
21	ว 3.1	สาระที่ 3	94	1.32	2.06
22	ว 3.2	สาระที่ 3	83	1.06	1.49
23	ว 3.2	สาระที่ 3	21	-0.35	0.22
24	ว 3.1	สาระที่ 3	84	1.09	1.55
25	ว 3.1	สาระที่ 3	37	0.48	0.63
26	ว 3.1	สาระที่ 3	50	0.55	0.73
27	ว 3.1	สาระที่ 3	15	-0.57	0
28	ว 3.1	สาระที่ 3	64	0.87	1.07

ข้อที่	รหัสมาตรฐาน	รหัสสาระ	ลำดับที่	Theta (RP=50)	Theta (RP=0.67)
29	ว 3.2	สาระที่ 3	68	0.95	1.16
30	ว 3.2	สาระที่ 3	82	1.14	1.46
31	ว 3.2	สาระที่ 3	95	1.66	2.16
32	ว 3.2	สาระที่ 3	26	0.21	0.4
33	ว 3.2	สาระที่ 3	78	1.23	1.3
34	ว 3.2	สาระที่ 3	38	0.46	0.64
35	ว 3.2	สาระที่ 3	61	0.48	1.01
36	ว 3.2	สาระที่ 3	53	0.68	0.84
37	ว 4.1	สาระที่ 4	59	0.19	0.97
38	ว 4.1	สาระที่ 4	85	1.01	1.61
39	ว 4.1	สาระที่ 4	80	1.15	1.34
40	ว 4.1	สาระที่ 4	73	1.09	1.21
41	ว 4.2	สาระที่ 4	23	-0.24	0.29
42	ว 4.2	สาระที่ 4	27	0.28	0.43
43	ว 4.2	สาระที่ 4	57	0.8	0.94
44	ว 4.2	สาระที่ 4	24	-0.16	0.34
45	ว 4.2	สาระที่ 4	58	0.72	0.95
46	ว 5.1	สาระที่ 5	28	0.28	0.55
47	ว 5.1	สาระที่ 5	34	0.43	0.6
48	ว 5.1	สาระที่ 5	32	0.4	0.59
49	ว 5.1	สาระที่ 5	20	-0.14	0.21
50	ว 5.1	สาระที่ 5	56	-0.04	0.86
51	ว 5.1	สาระที่ 5	70	1.12	1.19
52	ว 5.1	สาระที่ 5	55	0.62	0.85
53	ว 5.1	สาระที่ 5	31	0.34	0.58
54	ว 5.1	สาระที่ 5	62	0.82	1.03
55	ว 6.1	สาระที่ 6	71	1.04	1.19
56	ว 6.1	สาระที่ 6	77	1.19	1.27
57	ว 6.1	สาระที่ 6	48	0.07	0.71
58	ว 6.1	สาระที่ 6	16	-0.31	0.07
59	ว 6.1	สาระที่ 6	92	0.5	1.84

ข้อที่	รหัสมาตรฐาน	รหัสสาระ	ลำดับที่	Theta (RP=50)	Theta (RP=0.67)
60	ว 6.1	สาระที่ 6	90	1.46	1.75
61	ว 6.1	สาระที่ 6	67	0.75	1.12
62	ว 6.1	สาระที่ 6	14	-0.51	-0.05
63	ว 6.1	สาระที่ 6	65	0.6	1.07
64	ว 7.1	สาระที่ 7	76	1.15	1.26
65	ว 7.1	สาระที่ 7	33	0.31	0.59
66	ว 7.1	สาระที่ 7	79	1.24	1.32
67	ว 7.1	สาระที่ 7	19	-0.86	0.18
68	ว 7.1	สาระที่ 7	72	0.77	1.19
69	ว 7.1	สาระที่ 7	88	1.33	1.66
70	ว 7.2	สาระที่ 7	18	-0.19	0.09
71	ว 7.2	สาระที่ 7	74	1.13	1.24
72	ว 7.2	สาระที่ 7	29	-0.11	0.56
73	ว 8.1	สาระที่ 8	75	1.12	1.24
74	ว 8.1	สาระที่ 8	98	2.88	3.47
75	ว 8.1	สาระที่ 8	86	1.2	1.62
76	ว 8.1	สาระที่ 8	35	-0.3	0.6
77	ว 8.1	สาระที่ 8	81	1.19	1.39
78	ว 8.1	สาระที่ 8	17	-0.58	0.08
79	ว 8.1	สาระที่ 8	63	0.84	1.04
80	ว 8.1	สาระที่ 8	1	-3.91	-5
81.1	ว 6.1	สาระที่ 6	97	-1.39	2.77
81.2	ว 6.1	สาระที่ 6	100	18.45	22.61
82.1	ว 6.1	สาระที่ 6	10	-1.15	-0.85
82.2	ว 6.1	สาระที่ 6	89	1.39	1.69
83.1	ว 7.1	สาระที่ 7	9	-1.39	-0.91
83.2	ว 7.1	สาระที่ 7	93	1.41	1.89
84.1	ว 7.1	สาระที่ 7	5	-1.94	-1.41
84.2	ว 7.1	สาระที่ 7	91	1.28	1.81
85.1	ว 8.1	สาระที่ 8	6	-1.95	-1.33
85.2	ว 8.1	สาระที่ 8	96	1.55	2.17

ข้อที่	รหัสมาตรฐาน	รหัสสาระ	ลำดับที่	Theta (RP=50)	Theta (RP=0.67)
86.1	ว 8.1	สาระที่ 8	2	-2.78	-2.24
86.2	ว 8.1	สาระที่ 8	87	1.1	1.64
87.1	ว 2.1	สาระที่ 2	3	-1.98	-1.67
87.2	ว 2.1	สาระที่ 2	39	0.33	0.64
88.1	ว 1.2	สาระที่ 1	36	-0.66	0.6
88.2	ว 1.2	สาระที่ 1	99	5.5	6.76
89.1	ว 5.1	สาระที่ 5	8	-1.32	-1.08
89.2	ว 5.1	สาระที่ 5	66	0.87	1.11
90.1	ว 4.1	สาระที่ 4	4	-1.69	-1.42
90.2	ว 4.1	สาระที่ 4	40	0.39	0.66

ภาคผนวก ซ

1. ตัวอย่างผลวิเคราะห์ค่าความยากของข้อสอบด้วยโปรแกรม MULTILIG
2. ตัวอย่างผลการวิเคราะห์ค่าสัมประสิทธิ์การสรุปอ้างอิงด้วยโปรแกรม GENOVA
3. ตัวอย่างผลวิเคราะห์ความถูกต้อง ความคลาดเคลื่อนแบบบวก และความคลาดเคลื่อนแบบลบด้วยโปรแกรม MULT-CLASS

ตัวอย่างผลวิเคราะห์ค่าความยากของข้อสอบด้วยโปรแกรม MULTILIG

```
ITEM   1:      2 NOMINAL CATEGORIES,  2 HIGH
 TRADITIONAL 3PL, NORMAL METRIC:    A         B         C
                                                1.05      0.53      0.07
            CONTRAST-COEFFICIENTS (STANDARD ERRORS)
 FOR:           A                 C                 D
 CONTRAST P(#)  COEFF.[ DEV.]  P(#)  COEFF.[ DEV.]  P(#)  COEFF.[ DEV.]
    1      1   1.78 (0.87)     2  -0.95 (0.34)     3  -2.62 (0.67)
@THETA:     INFORMATION:  (Theta values increase in steps of 0.2)
-3.0 - -1.6  0.000  0.000  0.001  0.001  0.002  0.004  0.009  0.016
-1.4 -  0.0  0.029  0.050  0.084  0.134  0.203  0.292  0.396  0.504
 0.2 -  1.6  0.601  0.668  0.692  0.668  0.605  0.517  0.422  0.331
 1.8 -  3.0  0.252  0.187  0.137  0.099  0.071  0.051  0.036

 OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN
 CATEGORY(K):  1      2
 OBS. FREQ.    345    135
 OBS. PROP.  0.7188 0.2812
 EXP. PROP.  0.6055 0.3945

ITEM   2:      2 NOMINAL CATEGORIES,  2 HIGH
 TRADITIONAL 3PL, NORMAL METRIC:    A         B         C
                                                0.75      0.43      0.00
            CONTRAST-COEFFICIENTS (STANDARD ERRORS)
 FOR:           A                 C                 D
 CONTRAST P(#)  COEFF.[ DEV.]  P(#)  COEFF.[ DEV.]  P(#)  COEFF.[ DEV.]
    1      4   1.27 (0.30)     5  -0.54 (0.26)     6 -11.56 (****)
@THETA:     INFORMATION:  (Theta values increase in steps of 0.2)
-3.0 - -1.6  0.020  0.026  0.033  0.042  0.053  0.068  0.085  0.106
-1.4 -  0.0  0.132  0.161  0.195  0.233  0.272  0.311  0.347  0.377
 0.2 -  1.6  0.397  0.405  0.400  0.383  0.355  0.321  0.282  0.242
 1.8 -  3.0  0.204  0.170  0.139  0.112  0.090  0.072  0.057
 OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN
 CATEGORY(K):  1      2
 OBS. FREQ.    338    142
 OBS. PROP.  0.7042 0.2958
 EXP. PROP.  0.6015 0.3985
```

ITEM 3: 2 NOMINAL CATEGORIES, 2 HIGH

TRADITIONAL 3PL, NORMAL METRIC: A B C

1.61 1.57 0.13

CONTRAST-COEFFICIENTS (STANDARD ERRORS)

FOR:	A	C	D
CONTRAST	P(#) COEFF.[DEV.]	P(#) COEFF.[DEV.]	P(#) COEFF.[DEV.]
1	7 2.74 (1.91)	8 -4.30 (2.79)	9 -1.86 (0.28)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-1.4 - 0.0	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.003	0.008
0.2 - 1.6	0.022	0.057	0.137	0.298	0.568	0.926	1.266	1.440	
1.8 - 3.0	1.371	1.120	0.812	0.541	0.341	0.208	0.124		

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K):	1	2
OBS. FREQ.	404	76
OBS. PROP.	0.8417	0.1583
EXP. PROP.	0.7840	0.2160

ITEM 4: 2 NOMINAL CATEGORIES, 2 HIGH

TRADITIONAL 3PL, NORMAL METRIC: A B C

6.82 1.86 0.20

CONTRAST-COEFFICIENTS (STANDARD ERRORS)

FOR:	A	C	D
CONTRAST	P(#) COEFF.[DEV.]	P(#) COEFF.[DEV.]	P(#) COEFF.[DEV.]
1	10 11.60 (****)	11 -21.53 (****)	12 -1.37 (0.29)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-1.4 - 0.0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.2 - 1.6	0.000	0.000	0.000	0.000	0.000	0.000	0.013	1.007	
1.8 - 3.0	17.428	13.773	1.904	0.194	0.019	0.002	0.000		

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K):	1	2
OBS. FREQ.	382	98
OBS. PROP.	0.7958	0.2042
EXP. PROP.	0.7692	0.2308

ITEM 5: 2 NOMINAL CATEGORIES, 2 HIGH

TRADITIONAL 3PL, NORMAL METRIC: A B C

1.32 1.23 0.19

CONTRAST-COEFFICIENTS (STANDARD ERRORS)

FOR:	A	C	D

CONTRAST P(#) COEFF.[DEV.] P(#) COEFF.[DEV.] P(#) COEFF.[DEV.]

1 13 2.24 (1.19) 14 -2.75 (1.17) 15 -1.44 (0.28)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

-3.0 - -1.6 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000

-1.4 - 0.0 0.000 0.000 0.001 0.002 0.005 0.012 0.027 0.057

0.2 - 1.6 0.115 0.213 0.360 0.541 0.720 0.841 0.865 0.791

1.8 - 3.0 0.655 0.502 0.363 0.252 0.170 0.113 0.074

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN

CATEGORY(K): 1 2

OBS. FREQ. 360 120

OBS. PROP. 0.7500 0.2500

EXP. PROP. 0.6740 0.3260

ตัวอย่างผลการวิเคราะห์ค่าสัมประสิทธิ์การสรุปอ้างอิงด้วยโปรแกรม GENOVA

```
GGGGGGGGGG  EEEEEEEEEEE NN      NN  OOOOOOOO  VV      VV  AAAAAAAAAA
GGGGGGGGGGGG EEEEEEEEEEE NNN     NN  OOOOOOOOOO  VV      VV
AAAAAAAAAAAA
GG      G EE          NNNN     NN OO      OO VV      VV AA      AA
GG        EE          NN NN    NN OO      OO  VV    VV  AA      AA
GG        EE          NN  NN   NN OO      OO  VV    VV  AA      AA
GG        EE          NN   NN  NN OO      OO  VV    VV  AA      AA
GG        EE          NN    NN NN OO      OO  VV    VV  AA      AA
GG        EEEEEEEE    NN     NN NN OO      OO   VV   VV   AA      AA
GG    GGGG EEEEEEEE    NN     NNNN OO      OO   VV   VV   AAAAAAAAAAAA
GG    GGGG EE          NN      NNN OO      OO    VV VV    AAAAAAAAAAAA
GG      GG EE          NN      NN OO      OO    VV VV    AA      AA
GG      GG EE          NN      NN OO      OO    VV VV    AA      AA
GG      GG EE          NN      NN OO      OO     V V     AA      AA
GG      GG EE          NN      NN OO      OO     VVV     AA      AA
GGGGGGGGGGGG EEEEEEEEEEE NN      NN  OOOOOOOOOO     VVV     AA      AA
GGGGGGGGGG  EEEEEEEEEEE NN      NN  OOOOOOOO      VV      AA      AA
```

A GENERAL PURPOSE ANALYSIS OF VARIANCE SYSTEM

GENOVA IS A FORTRAN 77 PROGRAM FOR ANALYSIS OF VARIANCE

AND GENERALIZABILITY ANALYSES WITH BALANCED DESIGNS

AUTHORS

Joe E. Crick, Ed.D.

Chief Technology & Information Officer

Vice President Applications and Database Services

National Board of Medical Examiners

Philadelphia, PA 19104

Robert L. Brennan, Ed.D.

Director, Iowa Testing Program

University of Iowa

Iowa City, Iowa 52242

VERSION 3.1

January, 2001

GENOVA has been checked for accuracy of output, however the authors

can make no assurances that the program is totally without error.

GENOVA was developed in part under contract No. N00123-78-C-1206 with the Navy Personnel Research and Development Center (NPRDC); Robert L. Brennan Principal Investigator. GENOVA does not necessarily reflect NPRDC positions or policy, and no official endorsement should be inferred

```
GENOVA  VERSION 3.1                                                    PAGE      1
                                CONTROL CARD INPUT LISTING
COLUMN   1111111111222222222233333333334444444444555555555566666666667777777778
1234567890123456789012345678901234567890123456789012345678901234567890123456789
GSTUDY     (P:G)XR DESIGN -- RANDOM MODEL
OPTIONS    RECORDS 2
EFFECT     * G 3 0
EFFECT     + P:G 3 0
EFFECT     + R 3 0
FORMAT     (3F2.0)
PROCESS
GENOVA VERSION 3.1                                                     PAGE      2
     G STUDY                        (P:G)XR DESIGN -- RANDOM MODEL
                              EXPANDED MAIN AND INTERACTION EFFECT TABLE
     (** = INFINITE)   G       P       R              TOTAL   DEGREES
     SAMPLE SIZE       3       3       3      PRIMARY  NUMBER    OF
     UNIVERSE SIZE   ****    ****    ****      INDICES  INDICES  FREEDOM
     *********************************************
     *         *       *       *       *
     * G       *   1   *   0   *   0   *    1      1       2
     * P:G     *   2   *   1   *   0   *    1      2       6
     * R       *   0   *   0   *   1   *    1      1       2
     *         *       *       *       *
     *********************************************
     *         *       *       *       *
     * GR      *   1   *   0   *   1   *    2      2       4
     * PR:G    *   2   *   1   *   1   *    2      3       12
     *         *       *       *       *
     *********************************************
GENOVA VERSION 3.1                                                     PAGE      3
     G STUDY                        (P:G)XR DESIGN -- RANDOM MODEL
                              INPUT RECORD LISTING WITH RECORD MEANS
RECORD #     1    7.00000   12.00000    8.00000   44.00000   38.00000    8.00000   18.00000   24.00000
            11.00000   18.88889
RECORD #     2   21.00000   18.00000   14.00000    7.00000    7.00000    4.00000   14.00000   17.00000
            13.00000   12.77778
RECORD #     3   16.00000   16.00000   11.00000   17.00000    7.00000    8.00000   38.00000    8.00000
             2.00000   13.66667
GENOVA VERSION 3.1                                                     PAGE      4
     G STUDY                        (P:G)XR DESIGN -- RANDOM MODEL
```

CELL MEAN SCORES

****

*** GRAND MEAN = 15.1111111 ***

****

MEAN SCORES FOR EFFECT: R SUBSCRIPT NOTATION: (R)

(1) = 20.222222 (2) = 16.333333 (3) = 8.777778

GENOVA VERSION 3.1 PAGE 5

G STUDY (P:G)XR DESIGN -- RANDOM MODEL

ANOVA TABLE

(** = INFINITE) G P R

SAMPLE SIZE 3 3 3

UNIVERSE SIZE **** **** ****

EFFECT	DEGREES OF FREEDOM	SUMS OF SQUARES FOR MEAN SCORES	SUMS OF SQUARES FOR SCORE EFFECTS	MEAN SQUARES	F STATISTIC	(QF = QUASI F RATIO) F-TEST DEGREES OF FREEDOM NUMERATOR	DENOMINATOR
G	2	6361.55556	196.22222	98.11111	0.56458 QF	2 QF	5 QF
P:G	6	7294.66667	933.11111	155.51852	2.43774	6	12
R	2	6774.88889	609.55556	304.77778	3.71429	2	4
GR	4	7299.33333	328.22222	82.05556	1.28621	4	12
PR:G	12	8998.00000	765.55556	63.79630			
MEAN		6165.33333					
TOTAL	26	2832.66667					

NOTE: FOR GENERALIZABILITY ANALYSES, F-STATISTICS SHOULD BE IGNORED

GENOVA VERSION 3.1 PAGE 6

G STUDY (P:G)XR DESIGN -- RANDOM MODEL

G STUDY RESULTS

(** = INFINITE) G P R

SAMPLE SIZE 3 3 3

UNIVERSE SIZE **** **** **** QFM = QUADRATIC FORM

M O D E L V A R I A N C E C O M P O N E N T S

DEGREES OF	----- USING	-------- USING EMS	---------- STANDARD

```
     EFFECT      FREEDOM     ALGORITHM         EQUATIONS         ERROR
     ------------------------------------------------------------------------------
     G              2       (0.0)           (0.0)           12.9981921
     P:G            6      30.5740741       30.5740741       27.1373591
     R              2      24.7469136       24.7469136       24.5173454
     ------------------------------------------------------------------------------
     GR             4       6.0864198        6.0864198       17.7194034
     PR:G          12      63.7962963       63.7962963       24.1127335
     ------------------------------------------------------------------------------
       NOTE: THE "ALGORITHM" AND "EMS" ESTIMATED VARIANCE COMPONENTS WILL BE
          IDENTICAL IF THERE ARE NO NEGATIVE ESTIMATES
 GENOVA  VERSION 3.1                                                    PAGE      7
      G STUDY                    (P:G)XR DESIGN -- RANDOM MODEL
                                 EXPECTED MEAN SQUARE EQUATIONS
     (** = INFINITE)   G       P       R
     SAMPLE SIZE       3       3       3
     UNIVERSE SIZE   ****    ****    ****
EMS(G)        = 1.00*VC(PR:G) + 3.00*VC(GR) + 3.00*VC(P:G) + 9.00*VC(G)
EMS(R)        = 1.00*VC(PR:G) + 3.00*VC(GR) + 9.00*VC(R)
EMS(P:G)      = 1.00*VC(PR:G) + 3.00*VC(P:G)
EMS(GR)       = 1.00*VC(PR:G) + 3.00*VC(GR)
EMS(PR:G)     = 1.00*VC(PR:G)
 GENOVA  VERSION 3.1                                                    PAGE      8
      G STUDY                    (P:G)XR DESIGN -- RANDOM MODEL
                  VARIANCE - COVARIANCE MATRIX FOR ESTIMATED VARIANCE COMPONENTS (V)
           G          R          P:G        GR         PR:G
G          168.9529973
R           27.7082889   601.1002261
P:G       -245.4787525     0.0000000   736.4362575
GR        -104.6590858   -83.1248666    64.6026575   313.9772574
PR:G        64.6026575     0.0000000  -193.8079724  -193.8079724   581.4239173

STANDARD
                           STANDARD ERROR OF
                      VARIANCE DEVIATION VARIANCE
         UNIVERSE SCORE    24.74691   4.97463  24.51735
   EXPECTED OBSERVED SCORE    28.51605   5.34004  24.04659
       LOWER CASE DELTA     3.76914   1.94143   3.22311        GENERALIZABILITY COEFFICIENT =  0.86782
( 6.56567)
       UPPER CASE DELTA     4.99210   2.23430   2.73342                    PHI =  0.83214  ( 4.95722)
```

MEAN 10.72831 3.27541

NOTE: SIGNAL/NOISE RATIOS ARE IN PARENTHESES

GENOVA VERSION 3.1 PAGE 19

D STUDY -- P X (R:T) DESIGN -- R AND T RANDOM

D STUDY DESIGN NUMBER 001-005

VARIANCE - COVARIANCE MATRIX FOR ESTIMATED VARIANCE COMPONENTS (FOR MEAN SCORES) IN UNIVERSE OF GENERALIZATION (W)

	G	R	P:G	GR	PR:G
G	6.7581199				
R	5.5416578	601.1002261			
P:G	-1.9638300	0.0000000	1.1782980		
GR	-4.1863634	-16.6249733	0.5168213	12.5590903	
PR:G	0.5168213	0.0000000	-0.3100928	-1.5504638	0.9302783

GENOVA VERSION 3.1 PAGE 20

D STUDY -- P X (R:T) DESIGN -- R AND T RANDOM

SUMMARY OF D STUDY RESULTS FOR SET OF CONTROL CARDS NO. 001

V A R I A N C E S

D STUDY DESIGN NO	INDEX= UNIV.=	SAMPLE SIZES G INF.	P INF.	$R INF.	UNIVERSE SCORE	OBSERVED SCORE	CASE DELTA	CASE DELTA	EXPECTED LOWER GEN. MEAN	UPPER COEF.	PHI
001-001	1	1	3		24.74691	94.62963	69.88272	100.45679	62.11728	0.26151	0.19765
001-002	2	2	3		24.74691	43.73920	18.99228	26.63580	22.22325	0.56578	0.48162
001-003	3	3	3		24.74691	33.86420	9.11728	12.51440	14.68519	0.73077	0.66414
001-004	4	4	3		24.74691	30.25579	5.50887	7.41975	11.99614	0.81792	0.76933
001-005	5	5	3		24.74691	28.51605	3.76914	4.99210	10.72831	0.86782	0.83214

GENOVA VERSION 3.1 PAGE 21

ตัวอย่างผลวิเคราะห์ความถูกต้อง ความคลาดเคลื่อนแบบบวก และความคลาดเคลื่อนแบบลบ ด้วยโปรแกรม MULT-CLASS

```
*****************************************************************
***      MULT-CLASS: Multinomial & Compound-Multinomial       ***
***        Classification Consistency and Accuracy            ***
***                                                           ***
***                      Version 3.0                          ***
***                                                           ***
***                    Won-Chan Lee                           ***
***                      CASMA                                ***
***                  University of Iowa                       ***
***                                                           ***
***                     July 2008                             ***
***                                                           ***
***                  All Rights Reserved                      ***
*****************************************************************
*****************************************************************
                      List of Control Cards Options
                  ---------------------------------------------------------------
$Number of item sets
 2
$Set1: weight, #items, #score points, score points, data file
 2.00
20
 2
 0.0 1.0
 dich1.dat
$Set2: weight, #items, #score points, score points, data file
 3.00
20
 2
 0.0 1.0
 dich2.dat
$#categories, observed cut scores, true cut scores
 4
  15.110   42.630   81.260
  15.110   42.630   81.260
$Conditional results (Yes=1, No=0)
 1
```

$Output file

mathangaverate.out

$Bias correction (Yes=1, No=0)

1

****

****

Error Variance and Reliability for Total Raw Scores

Sample variance = 952.17617

Avg error variance = 33.78336

Reliability = 0.96452

****

****

Overall Classification Consistency

---

consistency (phi) = 0.75684

1-phi = 0.24316

chance probability = 0.27571

kappa = 0.66427

****

****

Overall Classification Accuracy

---

accuracy (gamma) = 0.82427

false positive error = 0.08697

false negative error = 0.08

ภาคผนวก ฌ
ประมวลภาพบรรยากาศการกำหนดคะแนนจุดตัด

ภาพบรรยากาศการศึกษานำร่อง
กำหนดคะแนนจุดตัดโดยนักศึกษา

ภาพบรรยากาศการกำหนดคะแนนจุดตัดโดยผู้เชี่ยวชาญ

ประวัติผู้เขียนวิทยานิพนธ์

นางศิริพันธ์ ติยะวงศ์สุวรรณ เกิดเมื่อวันที่ 1 เมษายน 2521 ปีการศึกษา 2544 สำเร็จการศึกษาปริญญาครุศาสตรบัณฑิต วิชาเอกการสอนวิทยาศาสตร์-เคมี จากคณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ปีการศึกษา 2546 สำเร็จการศึกษาปริญญาศึกษาศาสตร์มหาบัณฑิต สาขาวิชาการวิจัยและสถิติทางการศึกษา มหาวิทยาลัยศรีนครินทร์วิโรฒ ประสานมิตร ปีการศึกษา 2551 เข้าศึกษาต่อหลักสูตรครุศาสตรดุษฎีบัณฑิต สาขาวิชาการวัดและประเมินผลการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย โดยรับทุนพัฒนาบุคลากรจากมหาวิทยาลัยราชภัฏนครราชสีมา