การตรวจสอบคุณภาพของการตัดเกรด
โดยประยุกต์ใช้ทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ

Investigating the Quality of Grading
Through Multidimensional Item Response Theory

นายเจนรบ โกรธา

ตำแหน่งครู วิทยฐานะครูชำนาญการ

กลุ่มสาระการเรียนรู้วิทยาศาสตร์และเทคโนโลยี

ปีการศึกษา 2567

โรงเรียนบ้านฉางกาญจนกุลวิทยา อำเภอบ้านฉาง จังหวัดระยอง

สำนักงานเขตพื้นที่การศึกษามัธยมศึกษาชลบุรี ระยอง

บทคัดย่อ

การวิจัยเรื่องการตรวจสอบคุณภาพของการตัดเกรด โดยประยุกต์ใช้ทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ มีวัตถุประสงค์เพื่อตรวจสอบคุณภาพของการตัดเกรดในแต่ละรายวิชา และเพื่อตรวจสอบความสอดคล้องของระดับความสามารถที่แท้จริงของนักเรียนระดับชั้นมัธยมศึกษาปีที่ 6 กับผลการเรียนเฉลี่ย 6 ภาคเรียน กลุ่มตัวอย่างในการวิจัยในครั้งนี้คือ เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2566 โรงเรียนบ้านฉางกาญจนกุลวิทยา จำนวน 160 คน (จำนวน 8 ห้อง ห้องละ 20 คน) ซึ่งได้โดยวิธีการสุ่มตัวอย่างแบบแบ่งพวก (Stratified Sampling) วิเคราะห์และตีความข้อมูลตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ โดยใช้โมเดลสัมประสิทธิ์การสุ่มแบบโลจิท (Multidimensional Random Coefficients Multinomial Logit: MRCML) ผลการวิจัยสรุปได้ดังนี้

1) ผลการวิเคราะห์ข้อมูล สะท้อนให้เห็นคุณภาพของการตัดเกรด พบว่ามี 4 ลักษณะ คือ (I) วิชาที่ตัดเกรดได้เหมาะสม เป็นวิชาที่สะท้อนให้เห็นถึงระดับความสามารถของผู้เรียนได้อย่างชัดเจนสามารถจำแนกการตัดเกรดได้ตรงกับระดับความสามารถของผู้เรียน ได้แก่ วิชาที่ 1, 3, 8, 12, 16 (II) วิชาที่ยากต่อการได้เกรดสูงมีแนวโน้มกดเกรด เป็นวิชาที่สะท้อนให้เห็นว่ามีบางรายวิชาที่ผู้เรียนต้องมีความสามารถสูงมากจึงจะได้เกรด 4 ได้แก่ วิชาที่ 2, 4, 7, 9, 10, 11, 17 (III) วิชาที่ง่ายต่อการได้เกรดสูงมีแนวโน้มปล่อยเกรด สะท้อนให้เห็นว่ามีบางรายวิชาที่ผู้เรียนต้องมีความสามารถปานกลางค่อยมาทางอ่อนมีโอกาสได้เกรด 4 สูง ได้แก่ วิชาที่ 13 และ (IV) วิชาที่ไม่สามารถแบ่งแยกความแตกต่างระหว่างเกรด สะท้อนการตัดเกรดบางรายวิชาไม่สามารถแบ่งแยกความแตกต่างระหว่างเกรดได้ เนื่องจากอยู่ในตำแหน่งเดียวกัน ได้แก่ รายวิชาที่ 5, 6, 14, 15 จะเห็นว่าส่วนใหญ่ตัดเกรดแบบกดเกรด (41.176%) รองลงมาคือ ตัดเกรดเหมาะสม (29.411%) ตัดเกรดแยกเกรดไม่ได้ (29.411%) และ ตัดเกรดปล่อยเกรด (5.882%) ตามลำดับ

2) ผลการตรวจสอบความสอดคล้องของระดับความสามารถที่แท้จริงของนักเรียนระดับชั้นมัธยมศึกษาปีที่ 6 กับผลการเรียนเฉลี่ย 6 ภาคเรียน จะพบว่า (I) คะแนนมาตรฐานของผลการเรียน 6 เทอม ซึ่งไม่สอดคล้องกับข้อมูลผลการวิเคราะห์ทิศทางการตัดเกรด เนื่องจากทิศทางการตัดเกรดเป็นการวิเคราะห์ผลการเรียนเฉพาะรายวิชาพื้นฐานเท่านั้น (II) คะแนนมาตรฐานของความสามารถที่แท้จริง (θ) เป็นไปในทิศทางเดียวกันกับข้อมูลผลการวิเคราะห์ทิศทางการตัดเกรด เพราะทั้งสองจะใช้ข้อมูลชุดเดียว และสอดคล้องกับคุณภาพของการตัดเกรดที่มีการกดเกรดเป็นส่วนใหญ่ทำให้ระดับความสามารถต่ำกว่าความเป็นจริงและมีคะแนนมาตรฐานของความสามารถที่แท้จริง (θ) ห่างจากคะแนนมาตรฐานของผลการเรียน 6 เทอมค่อนข้างสูง (III) คะแนนมาตรฐานของผลการเรียน 6 เทอมและคะแนนมาตรฐานของความสามารถที่แท้จริง (θ) สะท้อนให้เห็นว่ายังคงมีทิศทางคะแนนไปในทิศทางเดียวกัน

สารบัญ

	หน้า
บทคัดย่อ	ก
สารบัญตาราง	ง
สารบัญภาพ	จ
บทที่ 1 บทนำ	
1.1 ความเป็นมาและความสำคัญของปัญหา	1
1.2 คำถามการวิจัย	2
1.3 วัตถุประสงค์การวิจัย	2
1.4 สมมติฐานการวิจัย	2
1.5 ขอบเขตการวิจัย	2
1.6 นิยามศัพท์เฉพาะ	3
1.7 ประโยชน์ที่คาดว่าจะได้รับ	5
บทที่ 2 เอกสารและงานวิจัยที่เกี่ยวข้อง	
2.1 ระดับความสามารถของผู้เรียน (Student's proficiency level)	6
2.2 การตัดสินผลการเรียนและการให้ระดับผลการเรียน	8
2.3 ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT)	10
2.4 โมเดลการตอบสนองข้อสอบแบบพหุมิติ	16
2.5 กรอบแนวคิดการวิจัย	22
บทที่ 3 วิธีดำเนินการวิจัย	
3.1 รวบรวมและตรวจสอบข้อมูลผลการเรียน	24
3.2 กำหนดระดับของคะแนนตามเกณฑ์การตัดเกรด	25
3.3 วิเคราะห์ข้อมูลเพื่อตรวจสอบคุณภาพการตัดเกรดและคุณภาพผู้เรียน	26
3.4 ตรวจสอบหลักฐานการประเมินด้านความตรงตามเกณฑ์มาตรฐานการประเมิน	44
บทที่ 4 ผลการวิจัย	
4.1 ผลการตรวจสอบคุณภาพการตัดเกรด	46
4.2 ผลการตรวจสอบคุณภาพผู้เรียน	49
4.3 ผลการตรวจสอบหลักฐานการประเมินด้านความตรงตามเกณฑ์มาตรฐานการประเมิน	51

สารบัญ (ต่อ)

หน้า

บทที่ 5 สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ

5.1 ผลการตรวจสอบคุณภาพของการตัดเกรดในแต่ละรายวิชา 52

5.2 ผลการตรวจสอบความสอดคล้องของระดับความสามารถที่แท้จริงของนักเรียนระดับชั้นมัธยมศึกษาปีที่ 6 กับผลการเรียนเฉลี่ย 6 ภาคเรียน 54

บรรณานุกรม 56

ภาคผนวก 58

ประวัติผู้วิจัย 61

สารบัญตาราง

	หน้า
ตารางที่ 1 การให้ระดับผลการเรียน 8 ระดับ และความหมายของแต่ละระดับ	10
ตารางที่ 2 ตัวอย่างผลการเรียนทำการกรอง	25
ตารางที่ 3 กำหนดระดับของคะแนนตามเกณฑ์โมเดล MRCLM	26
ตารางที่ 4 กำหนดชื่อรายวิชาให้มีความเหมาะสมกับการประมวลผล	27
ตารางที่ 5 การพิจารณาความเหมาะสมของแต่ละรายวิชา	28
ตารางที่ 6 ค่า Threshold ในแต่ละระดับขั้นของการตอบของแต่ละข้อใน Wright Map	29
ตารางที่ 7 การวิเคราะห์คุณภาพผู้เรียน	37

สารบัญภาพ

หน้า

ภาพที่ 1 ลักษณะข้อสอบ (ICC) สำหรับโมเดลการตอบสนองข้อสอบแบบ 1 พารามิเตอร์ 13
ภาพที่ 2 ลักษณะข้อสอบ (ICC) สำหรับโมเดลการตอบสนองข้อสอบแบบ 2 พารามิเตอร์ 14
ภาพที่ 3 ลักษณะข้อสอบ (ICC) สำหรับโมเดลการตอบสนองข้อสอบแบบ 3 พารามิเตอร์ 14
ภาพที่ 4 ลักษณะกราฟฟังก์ชันสารสนเทศของข้อสอบ (Item Information) 15
ภาพที่ 5 ลักษณะกราฟฟังก์ชันสารสนเทศของแบบสอบ (Test Information) 15
ภาพที่ 6 กราฟความน่าจะเป็นของการตอบสนองข้อสอบข้อที่ถูกและข้อที่ผิด 19
ภาพที่ 7กราฟความน่าจะเป็นของการตอบสนองข้อสอบข้อที่ถูกและข้อที่ผิด ใน 2 มิติ 20
ภาพที่ 8 กรอบแนวคิดในการวิจัย 23
ภาพที่ 9 Test Information 29
ภาพที่ 10 Wright Map แสดงคุณภาพของการตัดเกรดและความสามารถของนักเรียน 30
ภาพที่ 11 Wright Map ระดับ Thresholds ของข้อสอบแต่ละรายวิชา 34
ภาพที่ 12 Wright Map คุณภาพของการตัดเกรดและความสามารถของนักเรียน 35
ภาพที่ 13 การกำหนดคะแนนจุดตัดและจำนวนนักเรียนในแต่ละความสามารถ 43
ภาพที่ 14 เปรียบเทียบคะแนนมาตรฐานเกรด 6 เทอมกับความสามารถที่แท้จริง (θ) 44
ภาพที่ 15 Wright Map แสดงคุณภาพของการตัดเกรดและความสามารถของนักเรียน 47
ภาพที่ 16 Wright Map ระดับ Thresholds ของข้อสอบแต่ละรายวิชา 48
ภาพที่ 17 Wright Map คุณภาพของการตัดเกรดและความสามารถของนักเรียน 49
ภาพที่ 18 เปรียบเทียบคะแนนมาตรฐานเกรด 6 เทอมกับความสามารถที่แท้จริง (θ) 50

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

TCAS ย่อมาจาก Thai University Central Admission System คือ ระบบการคัดเลือกสอบเข้ามหาวิทยาลัย โดยเปรียบเสมือนพื้นที่กลางที่มหาวิทยาลัยเข้าระบบมาเพื่อรับสมัครนักเรียน และนักเรียนก็สามารถเข้ามาหามหาวิทยาลัยที่ต้องการได้ แบ่งการรับออกเป็น 4 รอบ ดังนี้ รอบที่ 1 รอบ Portfolio (ใช้แฟ้มสะสมผลงาน) รอบที่ 2 รอบ Quota (รอบโควตา) รอบที่ 3 รอบ Admission (รอบแอดมิชชั่น) และรอบที่ 4 รอบ Direct Admission (รอบรับตรงอิสระ) โดยเมื่อพิจารณาจากองค์ประกอบ/คะแนนที่ใช้คัดเลือกทุกรอบแล้วจะพบว่ามีการใช้คะแนนผลการเรียนเฉลี่ย (GPA, GPAX) เป็นองค์ประกอบ

GPAX เป็นคำที่คุ้นเคยสำหรับนักเรียนทั้งในระดับมัธยมศึกษาปีที่ 3 และ มัธยมศึกษาปีที่ 6 ใช้ในการประเมินผลการเรียนในช่วงเวลาที่กำหนด เป็นคะแนนที่หลายคนรู้ว่า ถูกนำมาใช้ในเกณฑ์การเข้าเรียนต่อในระดับ ม.4 หรือระดับมหาวิทยาลัย เพราะหลายโรงเรียน หรือมหาวิทยาลัยที่คัดเลือกผ่านระบบ TCAS ก็มีการกำหนดเกรดขั้นต่ำหรือใช้ GPAX เป็นเกณฑ์การคัดเลือกด้วย ดังนั้นการรักษาเกรดให้ดีตั้งแต่ตอนนี้จะยิ่งช่วยเพิ่มโอกาสในการสอบติดมหาวิทยาลัยที่อยากเข้า นอกจากคะแนน GPAX ขั้นต่ำที่กำหนดไว้ จะถูกนำมาคัดเลือกผู้สมัครแล้ว การมีคะแนนส่วนนี้เยอะ จะเพิ่มโอกาสในการสอบติดมากขึ้น เพราะยิ่งมีคะแนนสูงเท่าไหร่ โอกาสยิ่งมากขึ้นเท่านั้น และในระบบ TCAS ยังใช้คะแนน GPAX เข้าไปอยู่ในเกณฑ์คัดเลือกหลายรอบ รวมทั้งยังเป็นปัจจัยสำคัญในการเพิ่มโอกาสสมัครโครงการนักเรียนทุน เพราะโครงการทุนส่วนใหญ่ พิจารณาจากคะแนน GPAX ที่เฟ้นหานักเรียนที่เรียนดี มีความสามารถ เกรดเฉลี่ยที่ดีจึงมีผลในการคัดเลือกคนสอบติด

การประเมินผลการเรียนรู้ของนักเรียนในระดับการศึกษาขั้นพื้นฐานของการให้เกรดในแต่ละรายวิชาถือว่าเป็นส่วนสำคัญมากในระบบการศึกษาในประเทศ เนื่องจากมีผลต่อการได้รับการคัดเลือกให้เข้าเรียนในระดับที่สูงขึ้นในอนาคต หรือมีผลต่อการให้ทุนการศึกษา ดังนั้น GPAX ในแต่ละภาคเรียนหรือตลอดทั้งหลักสูตรถือเป็นดัชนีชี้วัดถึงคุณภาพของนักเรียน ครูประจำรายวิชา และโรงเรียน ว่าเป็นไปตามผลลัพธ์การเรียนรู้ที่คาดหวังหรือไม่ อย่างไร ตลอดจนมีผลต่อการได้รับการยอมรับในสังคม แต่อย่างไรก็ตามมักมีประเด็นที่กล่าวถึงคุณภาพของการตัดเกรดในแต่ละรายวิชามีมาตรฐานเป็นอย่างไร และมีความน่าเชื่อถือมากน้อยเพียงใด นักเรียนมีระดับความสามารถที่แท้จริงสอดคล้องกับผลการเรียนเฉลี่ย 6 ภาคเรียน

ดังนั้นผู้วิจัยจึงได้สนใจที่จะทำการวิจัยเรื่อง การตรวจสอบคุณภาพของการตัดเกรด โดยประยุกต์ใช้ทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ เพื่อเป็นการศึกษาและนำเสนอแนวทางการตรวจสอบคุณภาพของ

การตัดเกรดในการศึกษาครั้งนี้นำเสนอในเชิงของการศึกษานำร่อง (Pilot Study) เน้นการศึกษาเฉพาะกรณี (Case Study) กับนักเรียนระดับชั้นมัธยมศึกษาปีที่ 6 เพื่อศึกษาถึงความเหมาะสมและความเป็นไปได้ในการปฏิบัติจริง โดยสุ่มผลการตัดเกรดของนักเรียนระดับชั้นมัธยมศึกษาปีที่ 6 จำนวน 160 คน ในทุกแผน

1.2 คำถามการวิจัย

1.2.1 คุณภาพของการตัดเกรดในแต่ละรายวิชาเป็นอย่างไร เมื่อพิจารณาจากหลักฐานด้านความตรงและความเที่ยง ตลอดจนทิศทางการตัดเกรดในแต่ละรายวิชามีลักษณะอย่างไร

1.2.2 นักเรียนมีระดับความสามารถที่แท้จริงสอดคล้องกับผลการเรียนเฉลี่ย 6 ภาคเรียน เป็นอย่างไร มีจุดเด่นจุด ที่ควรปรับปรุงอย่างไรบ้าง

1.3 วัตถุประสงค์การวิจัย

1.3.1 เพื่อตรวจสอบคุณภาพของการตัดเกรดในแต่ละรายวิชา

1.3.2 เพื่อตรวจสอบความสอดคล้องของระดับความสามารถที่แท้จริงของนักเรียนระดับชั้นมัธยมศึกษาปีที่ 6 กับผลการเรียนเฉลี่ย 6 ภาคเรียน

1.4 สมมติฐานการวิจัย

1.4.1 คุณภาพของการตัดเกรดในแต่ละรายวิชามีความเหมาะสม

1.4.2 ระดับความสามารถที่แท้จริงของนักเรียนระดับชั้นมัธยมศึกษาปีที่ 6 มีความสอดคล้องกับผลการเรียนเฉลี่ย 6 ภาคเรียน

1.5 ขอบเขตการวิจัย

ในการวิจัยครั้งนี้ ผู้วิจัยได้กำหนดขอบเขตของการวิจัย ดังนี้

1.5.1 ด้านเนื้อหา

เมื่อพิจารณาโครงสร้างหลักสูตรของนักเรียนระดับมัธยมศึกษาตอนปลาย โรงเรียนบ้านฉางกาญจนกุลวิทยา จะประกอบไปด้วย 2 กลุ่มรายวิชา คือ (1) กลุ่มรายวิชาพื้นฐาน คือ รายวิชาที่นักเรียนในทุกห้องเรียนและทุกแผนการเรียนจะต้องได้เรียน (2) กลุ่มรายวิชาเพิ่มเติม คือ รายวิชาที่นักเรียนในทุกห้องเรียนและทุกแผนการเรียนมีรายวิชาที่ไม่เหมือนกัน แตกต่างกันออกไปตามแผนการเรียน

จากรายละเอียดของโครงสร้างหลักสูตรโรงเรียน พบว่า กลุ่มรายวิชาที่เหมาะสมในการตรวจสอบคุณภาพของการตัดเกรดในแต่ละรายวิชาที่เหมาะสมที่สุด คือ กลุ่มรายวิชาพื้นฐาน ดังรายละเอียดต่อไปนี้

1) สำหรับการตรวจสอบคุณภาพการตัดเกรด จะใช้ผลการเรียนรายวิชาพื้นฐาน ภาคเรียนที่ 1 และ 2 ปีการศึกษา 2566 ของระดับชั้นมัธยมศึกษาปีที่ 6 ซึ่งมีทั้งหมด 17 รหัสวิชา ได้แก่

ภาษาไทย คณิตศาสตร์ วิทยาศาสตร์ วิทยาการคำนวณ การออกแบบและเทคโนโลยี สังคมศึกษา ภาษาอังกฤษ สุขศึกษาและพลศึกษา ดนตรีและนาฏศิลป์ การงานอาชีพ

2) สำหรับตรวจสอบความสอดคล้องของระดับความสามารถที่แท้จริงของนักเรียนระดับชั้นมัธยมศึกษาปีที่ 6 กับผลการเรียนเฉลี่ย 6 ภาคเรียน

1.5.2 ประชากรและกลุ่มตัวอย่าง

1) ประชากร คือ นักเรียนระดับชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2566

2) กลุ่มตัวอย่างในการวิจัยในครั้งนี้ คือ เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2566 โรงเรียนบ้านฉางกาญจนกุลวิทยา จำนวน 160 คน (จำนวน 8 ห้อง ห้องละ 20 คน) ซึ่งได้โดยวิธีการสุ่มตัวอย่างแบบแบ่งพวก (Stratified Sampling)

1.5.3 ตัวแปรที่ศึกษา

ผู้วิจัยได้กำหนดตัวแปรที่ต้องการศึกษา ดังรายละเอียดต่อไปนี้

1) รายวิชาพื้นฐานทั้ง 17 รหัสวิชา เป็นรายการประเมิน (Item)

2) นักเรียนกลุ่มตัวอย่างทั้ง 160 คน กำหนดเป็นผู้เข้าสอบ (Examinees)

1.6 นิยามศัพท์เฉพาะ

1.6.1 โมเดลการตอบสนองข้อสอบแบบพหุมิติ (Multidimensional Random Coefficients Multinomial Logit: MRCML) (Adams, Wilson & Wang, 1997) หมายถึง โมเดลทางการวัดที่ใช้ในการตรวจสอบคุณภาพของแบบทดสอบได้ทั้งคุณภาพของข้อสอบรายข้อ ค่าประสิทธิภาพของตัวลวง ค่าฟังก์ชันสารสนเทศของข้อสอบและแบบทดสอบ คุณภาพของแบบทดสอบทั้งฉบับประเภทต่าง ๆ และความสามารถในการตรวจสอบความตรงเชิงโครงร้างภายใน (Internal Construct Validity) โดยสะท้อนออกมาเป็น Wright Map ซึ่งเป็นหัวใจที่สำคัญที่สุดของการตรวจสอบคุณภาพการออกแบบการประเมินที่สามารถนาผลการประเมินที่ได้ไปสู่การตีความของผู้เรียน

1.6.2 การวิเคราะห์เพื่อตรวจสอบคุณภาพของการตัดเกรดและคุณภาพผู้เรียน (Measurement Model/Wright Map) หมายถึง การตรวจสอบความตรงของเครื่องมือที่ใช้ในการประเมินระดับความสามารถที่แท้จริงของผู้เรียน โดยใช้โปรแกรม ConQuest 2.0 (Wu, Adums, Wilson, & Haldane, 2007) เมื่อพิจารณาจาก Wright Map ที่แสดงถึงลักษณะการกระจายความสามารถที่แท้จริงของผู้สอบที่มีความสัมพันธ์กับความยากของข้อสอบแต่ละข้อ พร้อมทั้งเป็นดัชนีบ่งชี้คุณภาพของโมเดลการวัด

1.6.3 คุณภาพการตัดเกรดของแต่ละรายวิชาเพื่อประเมินระดับความสามารถที่แท้จริงของนักเรียน ตรวจสอบหลักฐานการประเมินด้านความตรง 5 ด้านตามเกณฑ์มาตรฐานการประเมินทางด้านการศึกษาและจิตวิทยาที่มีคุณภาพ (AERA, APA, & NCME, 2014) เพื่อพิจารณาคุณภาพของการตัดเกรด

ประกอบด้วย (1) หลักฐานความตรงด้านเนื้อหาของรายการประเมิน (2) หลักฐานความตรงด้านกระบวนการตอบ (3) หลักฐานความตรงด้านโครงสร้างภายใน (4) หลักฐานความตรงที่สัมพันธ์กับเกณฑ์อื่น (5) หลักฐานความตรงที่เป็นผลมาจากการตัดเกรด ซึ่งในการวิจัยครั้งนี้ได้ปรับประยุกต์ใช้เพียง 3 หลักฐานความตรงที่สอดคล้องกับการตรวจสอบคุณภาพ โดยมีรายละเอียดแต่ละด้าน ดังนี้

(1) หลักฐานความตรงด้านโครงสร้างภายใน (Validity Evidence Based on Internal Structure) หมายถึง กระบวนการที่ใช้ในการตรวจสอบโครงสร้างภายในของรายการที่ประเมินกับมิติที่ต้องการศึกษา โดยพิจารณาจากดัชนีที่สะท้อนให้เห็นถึงความถูกต้องแม่นยำ เช่น ความเที่ยง ความตรง สัมประสิทธิ์ของการสรุปอ้างอิง ซึ่งในการวิจัยครั้งนี้ ผู้วิจัยได้ทำการตรวจสอบความตรงโดยใช้โมเดลการตอบสนองข้อสอบแบบพหุมิติ (MRCML) โดยมีการนำเสนอในรูปของ Wright Map พร้อมทั้งดัชนีที่สะท้อนถึงความตรงเชิงโครงสร้างภายใน โดยใช้โปรแกรม ACER Conquest 2.0 (Wu, Adams, Wilson, & Haldane, 2007)

(2) หลักฐานความตรงที่สัมพันธ์กับเกณฑ์อื่น (Evidence Based on Relations to Other Variables) หมายถึง การพิจารณาค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนน GPA ทั้ง 11 รายวิชา กับคะแนนหรือผลการประเมินอื่น ๆ ซึ่งเป็นเกณฑ์ภายนอกมาพิจารณา เพื่อเป็นหลักฐานสนับสนุนที่ถูกต้องในการตัดเกรดซึ่งเป็นเกณฑ์ภายนอก โดยในการวิจัยครั้งนี้ ผู้วิจัยมีการพิจารณาจากเมทริกซ์ความสัมพันธ์เพื่อเป็นหลักฐานยืนยันของคุณภาพการตัดเกรด เพื่อตรวจสอบคุณภาพด้านความตรงตามสภาพและความตรงเชิงทำนาย

- ตรวจสอบความตรงตามสภาพ โดยพิจารณาจากสัมประสิทธิ์สหสัมพันธ์กับตัวแปรภายนอก คือ ผลการเรียนเฉลี่ย 6 เทอม กับคะแนนความสามารถที่แท้จริงที่ประมาณค่าได้จากโปรแกรม เพื่อเป็นหลักฐานยืนยันถึงคุณภาพการตัดเกรดสอดคล้องกับคุณลักษณะที่แท้จริงของผู้เรียน

- ตรวจสอบความตรงเชิงทำนาย โดยพิจารณาจากสัมประสิทธิ์สหสัมพันธ์กับตัวแปรภายนอกที่สะท้อนถึงความสามารถของนักเรียน คือ คะแนน GPA ทั้ง 17 รายวิชา กับผลการเรียนเฉลี่ย 6 ภาคเรียน

(3) หลักฐานความตรงที่เป็นผลมาจากการตัดเกรด (Evidence Based on Consequences of Testing) หมายถึง ผลการประเมินส่งผลกระทบทั้งทางบวกและทางลบ สำหรับการศึกษาครั้งนี้จะมีการศึกษาทั้งผลกระทบทางบวกและทางลบจากการตัดเกรดเพื่อดูแนวโน้มว่าเกิดผลอย่างไรภายหลังการตัดเกรดทั้งกับผู้เรียน ผู้สอน และหลักสูตร

1.7 ประโยชน์ที่คาดว่าจะได้รับ

1.7.1 ประโยชน์ทางด้านวิชาการ

ได้แนวทาง (Pilot Study) ในการวิเคราะห์คุณภาพของการตัดเกรด ที่สะท้อนให้เห็นระดับความสามารถที่แท้จริงของนักเรียน การจัดการเรียนรู้และการวัดประเมินผลของครู และโรงเรียนได้เห็นถึงมีจุดเด่น จุดที่ควรปรับปรุงพัฒนา

1.7.2 ประโยชน์ด้านการนำไปใช้

1) ครูได้ทราบผลการวิเคราะห์คุณภาพของการตัดเกรดของรายวิชาที่รับผิดชอบเพื่อนำไปใช้ในการพัฒนา แก้ไขรูปแบบการจัดการเรียนการสอนและการวัดประเมินผลให้สอดคล้องกับความสมารถที่แท้จริงของนักเรียน เพื่อให้ครูสามารถตัดเกรดนักเรียนได้อย่างเหมาะสม

2) ผู้บริหารสถานศึกษา สามารถใช้ผลการวิเคราะห์คุณภาพของการตัดเกรดของแต่ละรายวิชา เป็นส่วนหนึ่งในการประเมินผล ติดตาม ส่งเสริม หรือนิเทศ กระบวนการจัดการเรียนการสอน การวัดประเมินผลและการตัดเกรดของครู เพื่อให้มีรูปแบบวิธีการวัดประเมินผลและการตัดเกรดที่เหมาะสม

1.7.3 ประโยชน์เชิงนโยบาย

ผู้บริหารสถานศึกษา สามารถใช้สารสนเทศจากการวิเคราะห์คุณภาพของการตัดเกรดของรายวิชาใช้ในการกำหนดแผนพัฒนาหรือประเมินความเหมาะสมการจัดเกรดของครูผู้สอนในสถานศึกษา

บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

การตรวจสอบคุณภาพของการตัดเกรดโรงเรียนบ้านฉางกาญจนกุลวิทยา โดยประยุกต์ใช้ทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ ได้ศึกษาค้นคว้าข้อมูลจากเอกสาร บทความ และงานวิจัย แล้วนำมาสรุปรวบรวม จำแนกเป็น 5 ตอน ได้แก่

1. ระดับความสามารถของผู้เรียน (Student's proficiency level)
2. การตัดสินผลการเรียนและการให้ระดับผลการเรียน
3. ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT)
4. โมเดลการตอบสนองข้อสอบแบบพหุมิติ
5. กรอบแนวคิดการวิจัย

แต่ละตอนมีรายละเอียดดังต่อไปนี้

2.1 ระดับความสามารถของผู้เรียน (Student's proficiency level)

การทดสอบทางการศึกษาส่วนใหญ่ ไม่ว่าจะเป็นการทดสอบในระดับใด หรือเพื่อวัตถุประสงค์ใดก็ล้วนแต่เพื่อเป็นการตรวจสอบความสามารถของผู้เรียนอันเป็นผลมาจากการเรียนรู้ที่ผ่านมา สำหรับการอธิบายความสามารถของผู้เรียนที่เกิดขึ้นในปัจจุบัน หรือเพื่อการทำนายความสามารถ ความพร้อม ความต้องการ ในการเรียนรู้ในอนาคต ความสามารถของบุคคลนั้นเป็นคุณลักษณะภายใน ที่ไม่สามารถวัดได้โดยตรง ซึ่งต้องอาศัยวิธีการวัดทางอ้อมโดยการตอบสนองต่อสิ่งเร้าที่จัดขึ้น แล้วนำผลการตอบสนองที่ได้มาทำการวิเคราะห์ ทำการสรุปอ้างอิงไปยังคุณลักษณะภายในที่มุ่งวัดนั้น ๆ ผู้วิจัยได้นำเสนอความหมายของความสามารถของผู้เรียน ดังรายละเอียดต่อไปนี้

2.1.1 ความหมายของความสามารถของผู้เรียน (Student's proficiency)

จากการศึกษาเกี่ยวกับความสามารถ (Proficiency) จะพบว่ามีลักษณะความเกี่ยวข้องกันโดยตรงกับเรื่องของสมรรถนะ (Competency) พบว่าผู้เขียนมักให้คำจำกัดความที่แตกต่างกันออกไปตามความเข้าใจของนักวิชาการศึกษาสมรรถนะแต่ละคน หลายท่านต่างก็ให้คำแปลและคำจำกัดความว่า "ขีดความสามารถ" บางท่านแปลว่า "สมรรถนะ" หรือบางท่านแปลว่า "ศักยภาพ" ซึ่ง Competency มีความหมายตามพจนานุกรมว่า "ความสามารถ" หรือ "สมรรถนะ" ในภาษาอังกฤษมีคำที่มีความหมายคล้ายกันอยู่หลายคำ ได้แก่ Capability, Ability, Proficiency, Expertise, Skill Fitness และ Aptitude

ซึ่งเมื่อพิจารณาถึงความหมายของคำว่า "ความสามารถ" สามารถอธิบายได้ว่าคือความเชี่ยวชาญซึ่งได้ให้ความหมายว่าหมายถึง ความสามารถในการประยุกต์ความรู้ที่มีอยู่เมื่อเผชิญและเกี่ยวข้องกับสถานการณ์ต่าง ๆ (เอกชัย อาชาพิพัฒน์, 2560)

หากลองพิจารณาคำว่า Proficiency และคำว่า Competency นั่นจะมีความสัมพันธ์กันอยู่ ซึ่งคำว่า Competency หมายถึง ความรู้ (Knowledge) ทักษะ (Skill) และคุณลักษณะเชิงพฤติกรรม (Attributes) ของบุคคล ซึ่งจำเป็นในการปฏิบัติงานตำแหน่งต่าง ๆ ให้ประสบความสำเร็จ โดยได้ผลงานที่สูงกว่ามาตรฐาน และโดดเด่นกว่าเพื่อนร่วมงานในสถานการณ์ที่หลากหลายกว่า

สมรรถนะที่จัดระดับพฤติกรรม (Leveled Competency) หมายถึง สมรรถนะที่มีความแตกต่างกันตามประสบการณ์การทำงานของบุคคล

สมรรถนะที่ไม่มีการจัดระดับ (Non-leveled Competency) หมายถึง สมรรถนะที่บุคคลทุกคนควรมีเหมือนกันในสมรรถนะนั้น โดยไม่ขึ้นอยู่กับประสบการณ์ในการทำงาน

จะสังเกตได้ว่า คำว่า Proficiency เป็นส่วนหนึ่งของ Competency เพื่อไว้ใช้ในการระบุและอธิบายระดับความสามารถที่เกิดขึ้น หรือในบางกรณีก็สามารถระบุได้ถึงพฤติกรรม ความรู้ ความสามารถที่อยู่ในปัจจุบัน หรือที่ยังขาดหรือบกพร่องไป ซึ่งระดับความสามารถ (Proficiency Level) นั่นได้มีผู้อธิบายไว้หลายท่าน ดังนี้

ขจรศักดิ์ ศิริมัย และคณะ (2554) ได้อธิบายไว้ว่า ระดับความสามารถ (Proficiency Level) หมายถึง ระดับพฤติกรรมที่แสดงออกให้เห็นถึงความสามารถ ความชำนาญในแต่ละสมรรถนะ ซึ่งขีดความสามารถของกลุ่มสมรรถนะหลัก (Core Competency) ได้แบ่งออกเป็นระดับความสามารถ 5 ระดับ ได้แก่

ระดับที่ 1 ขั้นพื้นฐาน (Novice)

ระดับที่ 2 ขั้นปฏิบัติงาน (Adequate)

ระดับที่ 3 ขั้นประยุกต์ (Develop)

ระดับที่ 4 ขั้นก้าวหน้า (Advance)

ระดับที่ 5 ขั้นเชี่ยวชาญ (Expert)

ราวดี ปฏิวัติวงศ์ (2552) ได้กล่าวว่าระดับความสามารถ (Proficiency Level) เป็นการบ่งบอกถึงพฤติกรรมที่คาดหวังหรือต้องการให้เกิดขึ้น ซึ่งจะแยกตามระดับที่แตกต่างกันไป โดยจะแบ่งออกเป็น 5 ระดับ ได้แก่

ขั้นเรียนรู้ (Basic Level)

ขั้นปฏิบัติ (Doing Level)

ขั้นพัฒนา (Developing Level)

ขั้นก้าวหน้า (Advanced Level)

ขั้นผู้เชี่ยวชาญ (Expert Level)

จากแนวคิดของความสามารถของผู้เรียนจะพบว่าความสามารถคือความรู้ความสามารถ ทักษะ

และคุณลักษณะเชิงพฤติกรรมของบุคคลที่แสดงออกให้เห็นถึงความสามารถ ความชำนาญในการเผชิญกับสถานการณ์ต่าง ๆ ซึ่งเป็นปัจจัยที่ใช้อธิบายในการประเมินเพื่อการเรียนรู้อย่างมีประสิทธิภาพจำเป็นอย่างยิ่งที่จะต้องรู้จักความรู้ ความสามารถ ทักษะผู้เรียนรายบุคคลอย่างละเอียด ถึงพฤติกรรมที่สามารถระบุสิ่งที่ผู้เรียนสามารถกระทำได้หรือเข้าใจอยู่ในปัจจุบัน รวมทั้งสิ่งที่ผู้เรียนบกพร่องหรือยังมีความเข้าใจที่คาดเคลื่อนไป สำหรับผลการประเมินความเข้าใจ การให้ข้อมูลสารสนเทศสิ่งที่ผู้เรียนมีความรู้ สิ่งที่จำเป็นต้องพัฒนาและปรับปรุง และแนวทางที่จะนำผู้เรียนไปสู่เป้าหมาย

2.2 การตัดสินผลการเรียนและการให้ระดับผลการเรียน

2.2.1 แนวปฏิบัติการวัดและประเมินผลการเรียนรู้ ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551

การวัดและประเมินผลการเรียนรู้ของผู้เรียนต้องอยู่บนจุดมุ่งหมายพื้นฐานสองประการ ประการแรกคือ การวัดและประเมินผลเพื่อพัฒนาผู้เรียน (Formative Assessment) และจุดมุ่งหมายประการที่สอง คือ การวัดและประเมินผลเพื่อตัดสินผลการเรียน เป็นการประเมินสรุปผลการเรียนรู้ (Summative Assessment) ซึ่งมีหลายระดับ ได้แก่ เมื่อเรียนจบหน่วยการเรียน จบรายวิชาเพื่อตัดสินให้คะแนน หรือให้ระดับผลการเรียน ให้การรับรองความรู้ความสามารถของผู้เรียนว่าผ่านรายวิชาหรือไม่ ควรได้รับการเลื่อนชั้นหรือไม่ หรือสามารถจบหลักสูตรหรือไม่ ในการประเมินเพื่อตัดสินผลการเรียนที่ดีต้องให้โอกาสผู้เรียนแสดงความรู้ความสามารถด้วยวิธีการที่หลากหลายและพิจารณาตัดสินบนพื้นฐานของเกณฑ์ผลการปฏิบัติมากกว่าใช้เปรียบเทียบระหว่างผู้เรียน

หลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 กำหนดว่า การวัดและประเมินผลการเรียนรู้เป็นกระบวนการพัฒนาปรับปรุงการเรียนรู้ของผู้เรียน และตัดสินว่าผู้เรียนมีความรู้ทักษะความสามารถ คุณลักษณะอันพึงประสงค์ อันเป็นผลมาจากการเรียนการสอนบรรลุตามมาตรฐานการเรียนรู้/ตัวชี้วัดในระดับใด สามารถที่จะได้รับการเลื่อนชั้นหรือจบการศึกษาได้หรือไม่

2.2.2 การตัดสินผลการเรียน

หลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 กำหนดหลักเกณฑ์การวัดและประเมินผลการเรียนรู้ เพื่อตัดสินผลการเรียนของผู้เรียน ดังนี้

1) ตัดสินผลการเรียนเป็นรายวิชา ผู้เรียนต้องมีเวลาเรียนตลอดภาคเรียนไม่น้อยกว่าร้อยละ 80 ของเวลาเรียนทั้งหมดในรายวิชานั้น ๆ

2) ผู้เรียนต้องได้รับการประเมินทุกตัวชี้วัดและผ่านตามเกณฑ์ที่สถานศึกษากำหนด

3) ผู้เรียนต้องได้รับการตัดสินผลการเรียนทุกรายวิชา

4) ผู้เรียนต้องได้รับการประเมินและมีผลการประเมินผ่านตามเกณฑ์ที่สถานศึกษากำหนดในการอ่าน คิดวิเคราะห์ และเขียน คุณลักษณะอันพึงประสงค์ และกิจกรรมพัฒนาผู้เรียน

เพื่อให้การจัดการเรียนการสอนมีประสิทธิภาพ ผู้สอนต้องตรวจสอบความรู้ความสามารถที่แสดงพัฒนาการของผู้เรียนอย่างสม่ำเสมอและต่อเนื่อง อีกทั้งต้องสร้างให้ผู้เรียนรับผิดชอบการเรียนรู้ของตนด้วยการตรวจสอบความก้าวหน้าในการเรียนของตนเองอย่างสม่ำเสมอเช่นกัน ตัวชี้วัดซึ่งมีความสำคัญในการนำมาใช้ออกแบบหน่วยการเรียนรู้นั้นยังเป็นแนวทางสำหรับผู้สอนและผู้เรียนใช้ในการตรวจสอบย้อนกลับว่าผู้เรียนเกิดการเรียนรู้หรือยัง การประเมินในชั้นเรียนซึ่งต้องอาศัยทั้งการประเมินเพื่อการพัฒนาและการประเมินเพื่อสรุปการเรียนรู้จะเป็นเครื่องมือสำคัญในการตรวจสอบความก้าวหน้าในการเรียนรู้ของผู้เรียน สถานศึกษา โดยผู้สอนกำหนดเกณฑ์ที่ยอมรับได้ในการผ่านตัวชี้วัดทุกตัวให้เหมาะสมกับบริบทของสถานศึกษา กล่าวคือ ให้ท้าทายการเรียนรู้ ไม่ยากหรือง่ายเกินไป เพื่อใช้เป็นเกณฑ์ในการประเมินว่าสิ่งที่ผู้เรียนรู้ เข้าใจทำได้นั้น เป็นที่น่าพอใจ บรรลุตามเกณฑ์ที่ยอมรับได้ หากยังไม่บรรลุจะต้องหาวิธีการช่วยเหลือ เพื่อให้ผู้เรียนได้รับการพัฒนาสูงสุด การกำหนดเกณฑ์นี้ผู้สอนสามารถให้ผู้เรียนร่วมกำหนดด้วยได้ เพื่อให้เกิดความรับผิดชอบร่วมกันและสร้างแรงจูงใจในการเรียน การประเมินเพื่อการพัฒนาส่วนมากเป็นการประเมินอย่างไม่เป็นทางการ เช่น สังเกต หรือซักถาม หรือการทดสอบย่อย ในการประเมินเพื่อการพัฒนานี้ ควรให้ผู้เรียนได้รับการพัฒนาจนผ่านเกณฑ์ที่ยอมรับได้ ผู้เรียนแต่ละคนอาจใช้เวลาเรียนและวิธีการเรียนที่แตกต่างกัน ฉะนั้น ผู้สอนควรนำข้อมูลที่ได้มาใช้ปรับวิธีการสอนเพื่อให้ผู้เรียนได้รับการพัฒนาเต็มศักยภาพ อันจะนำไปสู่การบรรลุมาตรฐานการเรียนรู้ในท้ายที่สุดอย่างมีคุณภาพ การประเมินเพื่อการพัฒนาจึงไม่จำเป็นต้องตัดสินให้คะแนนเสมอไป การตัดสินให้คะแนนหรือให้เป็นระดับคุณภาพควรดำเนินการโดยใช้การประเมินสรุปผลรวมเมื่อจบหน่วยการเรียนรู้และจบรายวิชา การตัดสินผลการเรียนตัดสินเป็นรายวิชา โดยใช้ผลการประเมินระหว่างภาคและปลายภาคตามสัดส่วนที่สถานศึกษากำหนด ทุกรายวิชาต้องได้รับการตัดสินและให้ระดับผลการเรียน ทั้งนี้ ผู้เรียนต้องผ่านทุกรายวิชาพื้นฐาน

2.2.3 การให้ระดับผลการเรียน

หลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 กำหนดการตัดสินเพื่อให้ระดับผลการเรียนรายวิชาของกลุ่มสาระการเรียนรู้ ให้ใช้ตัวเลขแสดงผลการเรียนเป็น 8 ระดับ

การตัดสินผลการเรียนในระดับการศึกษาขั้นพื้นฐานใช้ระบบผ่านและไม่ผ่าน โดยกำหนดเกณฑ์การตัดสินผ่านแต่ละรายวิชาที่ร้อยละ 50 จากนั้นจึงให้ระดับผลการเรียนที่ผ่าน สำหรับระดับมัธยมศึกษาตอนต้นและตอนปลาย ใช้ตัวเลขแสดงระดับผลการเรียนเป็น 8 ระดับ แนวการให้ระดับผลการเรียน 8 ระดับ และความหมายของแต่ละระดับดังแสดงในตาราง ดังนี้

ระดับผลการเรียน	ความหมาย	ช่วงคะแนนเป็นร้อยละ
4	ดีเยี่ยม	80 – 100
3.5	ดีมาก	75 – 79
3	ดี	70 – 74
2.5	ค่อนข้างดี	65 – 69
2	ปานกลาง	60 – 64
1.5	พอใช้	55 – 59
1	ผ่านเกณฑ์ขั้นต่ำ	50 – 54
0	ต่ำกว่าเกณฑ์	0 – 49

ตารางที่ 1 การให้ระดับผลการเรียน 8 ระดับ และความหมายของแต่ละระดับ

จากการศึกษาการตัดสินผลการเรียนและการให้ระดับผลการเรียนตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 จะพบว่า บทบาทที่สำคัญในการดำเนินการระหว่างทางจนกระทั่งถึงขั้นตอนในการตัดสินผลการเรียนและการให้ระดับผลการเรียนนั้นจะอยู่ที่ครูผู้สอนเป็นสำคัญ ตั้งแต่พัฒนาผู้เรียนให้เรียนรู้ รับผิดชอบการเรียน การตรวจสอบความก้าวหน้าในการเรียนรู้ของนักเรียน การประเมินผลทั้งเพื่อพัฒนา และการประเมินผลเพื่อสรุปผลที่มีความหลากหลาย เพื่อให้ผู้เรียนสามารถเป็นไปตามที่ตัวชี้วัดกำหนดและตามที่สถานศึกษากำหนด ดังนั้นหากครูผู้สอนหากไม่ได้ปฏิบัติให้สอดคล้องตามระเบียบวัดและประเมินผลที่กำหนด นอกจากอาจจะส่งผลต่อการเรียนรู้ของนักเรียน ที่ไม่ได้รับการติดตามและพัฒนาให้สอดคล้อง ยังส่งผลต่อการติดสินผลการเรียนที่ไม่สอดคล้องความสามารถที่แท้จริงของนักเรียน และอาจส่งผลเสียต่อนักเรียน ทั้งการได้รับรู้คุณภาพการเรียนรู้ของตนเองไม่ถูกต้อง การใช้ผลการเรียนสมัครเข้าเรียนในระดับที่สูงขึ้น อาจเป็นการตัดโอกาสในการสมัครเข้าเรียนหรือเรียนของนักเรียนได้

2.3 ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT)

การจัดการศึกษามีความจำเป็นที่จะต้องกำหนดจุดมุ่งหมายที่มีความชัดเจน เพื่อช่วยในการกำหนดคุณลักษณะที่ต้องการของผู้เรียน เพื่อใช้เป็นแนวทางในการวางแผนเพื่อใช้การจัดการเรียนรู้ จุดมุ่งหมายยังมีส่วนช่วยในการกำหนดทิศทางของการศึกษา การจัดประสบการณ์การเรียนรู้หรือกิจกรรมการจัดการเรียนรู้ที่เหมาะสม การวัดและการประเมินผลจึงเข้ามามีบทบาทสำคัญในการตัดสินผลการเรียนรู้ของผู้เรียนว่าสำเร็จตรงตามจุดมุ่งหมายที่กำหนดไว้หรือไม่ รวมทั้งยังช่วยให้สารสนเทศเพื่อช่วยในการปรับปรุงละพัฒนาการเรียนรู้ของผู้เรียนและการจัดกิจกรรมการเรียนรู้ของผู้สอน เมื่อการวัดและการประเมินผลเข้ามามีบทบาทสำคัญ ทำให้ทฤษฎีการทดสอบเข้ามามีบทบาทสำคัญในการวัดและประเมินผล เนื่องจากการการวัดและประเมินผลทางการศึกษานั้นเป็นสิ่งที่อยู่ภายในไม่สามารถวัดออกมาได้โดยตรง เพื่อทำความเข้าใจในสิ่งที่ต้องการวัดตามคุณลักษณะที่มุ่งวัด โครงสร้างของการวัด และการพัฒนาเครื่องมือ จึงจำเป็นต้อง

อาศัยทฤษฎีการทดสอบทดสอบเข้ามาช่วย โดยทฤษฎีการทดสอบนั้นหากแบ่งประเภทเป็น 2 ประเภทหลักใหญ่ๆ ได้แก่ ทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory) และทฤษฎีการทดสอบแนวใหม่ (Modern Test Theory)

ทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory) จะเน้นไปทางการตรวจสอบความสัมพันธ์ระหว่างคะแนนที่สังเกตได้กับคะแนนที่แท้จริง และวิเคราะห์คุณภาพโดยส่วนรวมของข้อสอบและแบบสอบ โดยใช้การวิเคราะห์คุณภาพของข้อสอบและแบบทดสอบ โดยใช้โมเดลคะแนนจริงแบบดั้งเดิม การใช้ทฤษฎีความเที่ยงและความคลาดเคลื่อน การใช้ความตรงเพื่อตรวจสอบความตรงต่างๆ หลักการการสร้างแบบทดสอบ และใช้หลักการวิเคราะห์ข้อสอบ ทฤษฎีการทดสอบแบบดั้งเดิมยังมีข้อด้อยของข้อตกลงเบื้องต้น มีข้อจำกัดของฐานความเชื่อเกี่ยวกับคะแนนความคลาดเคลื่อนและค่าพารามิเตอร์ของข้อสอบและแบบทดสอบ จึงส่งผลให้มีนักทฤษฎีการทดสอบหลายท่านได้ก่อการปฏิรูปแนวคิดเกิดขึ้นสู่ทฤษฎีการทดสอบแนวใหม่ (Modern Test Theory) เพื่อคลายข้อตกลงเบื้องต้น และแก้ไขจุดอ่อนบางประการ

ทฤษฎีคะแนนจริงแบบดั้งเดิมนี้จะตั้งอยู่บนพื้นฐานของ โมเดลการวัดและข้อตกลงเบื้องต้นที่สำคัญหลายประการ โดยทฤษฎีนี้หลัก ๆ จะเน้นเพื่อการพัฒนาหลักการสร้างข้อสอบ การวิเคราะห์คุณภาพของแบบสอบ และการวิเคราะห์คุณภาพของข้อสอบ โมเดลคะแนนจริงแบบดั้งเดิม (X) เกิดจากองค์ประกอบที่สังเกตไม่ได้ 2 ส่วน คือคะแนนจริง (T) และคะแนนความคลาดเคลื่อน (E) ซึ่งมีความสัมพันธ์ดังสมการ $X = T + E$

ทฤษฎีการทดสอบแนวใหม่ (Modern Test Theory) จะมุ่งเน้นการขยายแนวคิดของทฤษฎีการทดสอบแบบดั้งเดิมให้มีขอบเขตที่กว้างขวาง และมีความน่าเชื่อถือมากยิ่งขึ้น โดยปรับข้อตกลงให้มีความสมเหตุสมผล ตามความจริงมากขึ้น เกี่ยวกับความคลาดเคลื่อนของการวัด ยอมให้มีความคลาดเคลื่อนจากการวัดนั้นมีความแตกต่างกันออกไปตามสถานการณ์การวัด เพื่อศึกษาความเที่ยงทั่วไป ภายใต้เงื่อนไขของการวัดตามแนวของ Generalizability Theory: G-Theory รวมทั้งศึกษาความคลาดเคลื่อนจากการวัดที่มีความแตกต่างกันไปตามลักษณะของข้อสอบและแบบสอบ ขึ้นกับระดับความสามารถของผู้สอบ และของข้อสอบและแบบสอบ ทั้งยังพยายามวัดคุณลักษณะภายในหรือความสามารถที่แท้จริงของบุคคลตามแนว Item Response Theory (IRT) นอกจากนี้ความคลาดเคลื่อนจากการวัดของบุคคลเองก็ส่งผลต่อการตอบข้อสอบมากกว่า 1 องค์ประกอบ ดังนั้นเพื่อวัดความสามารถของบุคคลจึงมี 2 องค์ประกอบขึ้นไป เพื่อให้ความสามารถหลายมิติของบุคคลสอดคล้องกับข้อมูลได้ดียิ่งขึ้นและส่งผลต่อความน่าจะเป็นของการตอบถูกตามแนวของ Multidimensional Item Response Theory (MIRT). ทั้งทฤษฎีการทดสอบแบบดั้งเดิม ทฤษฎีการทดสอบแนวใหม่ ต่างก็ต้องการวัดที่เน้นผลผลิต และพฤติกรรมที่เกิดจากการนำความรู้และการนำทักษะไปใช้ในบริบทหรือสถานการณ์ที่เป็นชีวิตประจำวันจริงและด้วยการใช้สถานการณ์ที่เลียนแบบชีวิตจริง

ทฤษฎีการทดสอบแบบดั้งเดิม Classical Test Theory (CTT) ถูกใช้เป็นแนวคิดพื้นฐานสำหรับการพัฒนาแบบทดสอบทางการศึกษาและจิตวิทยาเป็นส่วนใหญ่ ตลอดช่วงศตวรรษที่ 20 ที่ผ่านมา

ในปัจจุบันก็ยังมีการใช้กันอย่างกว้างขวาง แต่ด้วยข้อจำกัดหลายประการของการวัดตามทฤษฎีการทดสอบแบบดั้งเดิม ทำให้มีนักทฤษฎีทางการทดสอบหลายท่านได้ก่อกระแสการวัดคุณลักษณะภายในแนวใหม่เพื่อที่จะแก้ไขจุดอ่อนเกี่ยวกับข้อตกลงเบื้องต้นของทฤษฎีการทดสอบแบบดั้งเดิม เริ่มต้นด้วยเทอร์สโตน (L.L. Thurston, 1927, 1928) เป็นผู้บุกเบิกการวัดคุณลักษณะภายในบุคคล และพัฒนาเทคนิคการวิเคราะห์ตัวประกอบสำหรับศึกษาคุณลักษณะทางจิตวิทยา อันเป็นการวางรากฐานความคิดที่สำคัญเกี่ยวกับ ทฤษฎีการทดสอบแนวใหม่ (Modern Test Theory: MTT) จากผลงานของ ครอนบาคและคณะ (Cronbach et.al., 1963, 1972) ได้เขียนบทความเรื่อง “The Dependability of Behavioral Measurement: Theory of Generalizability of Score and Profiles” ได้เสนอแนวคิดเกี่ยวกับโมเดลความเที่ยงทั่วไปภายใต้เงื่อนไขแบบต่างๆของการทดสอบ รวมทั้งผลงานของ ลอร์ดและโนวิค (Load and Novice, 1968) เรื่อง “Statistical Theories of Mental Test Score” ซึ่งได้นำเสนอหลักการและแนวคิดสำคัญที่ช่วยผลักดันกระแสการปฏิรูประบบความคิดของการวัดสู่ทฤษฎีการทดสอบแนวใหม่ รวมทั้งการมีส่วนร่วมของนักทฤษฎีการทดสอบในยุคใหม่อีกหลายท่าน จนได้พัฒนาการทดสอบแนวใหม่ซึ่งสามารถจำแนกออกเป็น 2 แนวทางสำคัญๆ ได้แก่ ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของการวัด (Generalizability Theory) หรือ G-Theory และทฤษฎีการตอบสนองข้อสอบ (Item Response Theory) หรือ IRT

2.3.1 ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT)

ทฤษฎีการตอบสนองข้อสอบ หรือ IRT สามารถจำแนกออกได้เป็น 2 ลักษณะ ได้แก่ ทฤษฎีการตอบสนองข้อสอบแบบตรวจให้คะแนน 2 ค่า (Dichotomous IRT) และทฤษฎีการตอบสนองข้อสอบแบบตรวจให้คะแนนมากกว่า 2 ค่า (Polytomous IRT)

โมเดลการตอบสนองข้อสอบ (IRT Models) เป็นระบบความสัมพันธ์ระหว่างตัวแปรอิสระที่รวมกันสำหรับทำนายตัวแปรตาม สำหรับโมเดลการตอบสนองข้อสอบ ตัวแปรอิสระจะประกอบด้วย ตัวแปรแฝง คือความสามารถที่แท้จริงของผู้สอบ ("θ") และคุณลักษณะของผู้สอบ (B) sหรือค่าพารามิเตอร์ข้อสอบ (a, b, c) ส่วนตัวแปรอิสระเป็นตัวแปรที่สังเกตได้ คือ โอกาสการตอบข้อสอบได้ถูกต้อง โมเดลการตอบสนองข้อสอบจะประกอบด้วยค่าพารามิเตอร์และค่าคงที่ดังนี้

1. พารามิเตอร์ของผู้สอบ

θ = ระดับความสามารถผู้สอบ โดยที่ค่า θ มีพิสัยอยู่ระหว่าง $-\infty$ ถึง $+\infty$

2. พารามิเตอร์ของข้อสอบ

a_i = ค่าพารามิเตอร์อำนาจจำแนกของข้อสอบข้อที่ i (Discrimination parameter) ในทางทฤษฎีแล้วมีค่าระหว่าง $-\infty$ ถึง $+\infty$แต่ในทางปฏิบัตินิยมใช้ a_i ระหว่าง +0.50 ถึง +2.50

b_i = ค่าพารามิเตอร์ความยากของข้อสอบข้อที่ i (Difficulty parameter) ในทางทฤษฎี

แล้วมีค่าระหว่าง $-\infty$ ถึง $+\infty$ แต่ในทางปฏิบัตินิยมใช้ b_i ระหว่าง -2.50 ถึง +2.50

c_i = ค่าพารามิเตอร์โอกาสในการเดาข้อสอบถูก (Guessing parameter) ในทางทฤษฎีแล้วมีค่าระหว่าง 0 ถึง 1 แต่ในทางปฏิบัตินิยมใช้ c_i ไม่เกิน 0.30

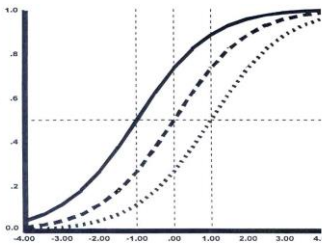
3. ค่าคงที่ e = 2.71828, D = 1.70

ข้อตกลงของทฤษฎีการตอบข้อสอบ

1. ความเป็นมิติเดียว (Unidimensionality: One trait)
2. ความเป็นอิสระ (Independence: Local Independent)
3. โค้งคุณลักษณะข้อสอบ (Item Characteristic Curve: Item Response Model)
4. ข้อสอบที่ใช้ต้องไม่เป็นข้อสอบประเภทความเร็ว

(1) โมเดลการตอบสนองข้อสอบแบบ 1 พารามิเตอร์ (One parameter model or Rasch Model) โมเดลแรกที่มีความสำคัญที่สุดที่นำมาใช้ในโมเดลทฤษฎีการตอบข้อสอบ โค้งลักษณะข้อสอบสำหรับโมเดลโลจิสติกแบบ 1 พารามิเตอร์ แสดงในรูปฟังก์ชันทางคณิตศาสตร์ ดังนี้

$$P_i = \frac{e^{(\theta - b_i)}}{1 + e^{(\theta - b_i)}} \; ; \; I = 1,2,\ldots,n$$

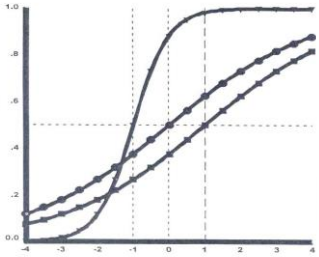


ภาพที่ 1 ลักษณะข้อสอบ (ICC) สำหรับโมเดลการตอบสนองข้อสอบแบบ 1 พารามิเตอร์

โค้งลักษณะข้อสอบ (ICC) สำหรับโมเดลการตอบสนองข้อสอบแบบ 1 พารามิเตอร์ มีลักษณะที่สำคัญคือ b_i มีค่าแปรเปลี่ยนตามลักษณะข้อสอบแต่ละข้อ a_i มีค่าคงที่ และ c_i เท่ากับ 0

(2) โมเดลการตอบสนองข้อสอบแบบ 2 พารามิเตอร์ (Two parameter model) โมเดลแรกที่มีความสำคัญที่สุดที่นำมาใช้ในโมเดลทฤษฎีการตอบข้อสอบ โค้งลักษณะข้อสอบสำหรับโมเดลโลจิสติกแบบ 2 พารามิเตอร์ แสดงในรูปฟังก์ชันทางคณิตศาสตร์ ดังนี้

$$P_i = \frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}} \; ; \; I = 1,2,\ldots,n$$



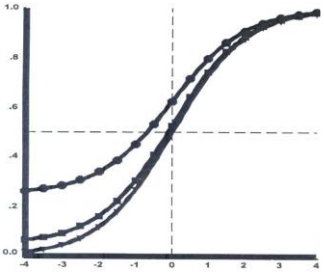
ภาพที่ 2 ลักษณะข้อสอบ (ICC) สำหรับโมเดลการตอบสนองข้อสอบแบบ 2 พารามิเตอร์

โค้งลักษณะข้อสอบ (ICC) สำหรับโมเดลการตอบสนองข้อสอบแบบ 2 พารามิเตอร์ มีลักษณะที่สำคัญคือ b_i เป็นค่าที่แสดงตำแหน่งของ ICC ณ จุด θ ที่มีโอกาสตอบข้อสอบถูก 0.50 ส่วน a_i เป็นค่าความชันของ ICC ณ จุด b_i ส่วน c_i เท่ากับ 0 และ e = 2.718, D = 1.7

(3) โมเดลการตอบสนองข้อสอบแบบ 3 พารามิเตอร์ (Three parameter model)

โมเดลแรกที่มีความสำคัญที่สุดที่นำมาใช้ในโมเดลทฤษฎีการตอบข้อสอบ โค้งลักษณะข้อสอบสำหรับโมเดลโลจิสติกแบบ 3 พารามิเตอร์ แสดงในรูปฟังก์ชันทางคณิตศาสตร์ ดังนี้

$$P_i = \boldsymbol{C_i} + (1-c_i)\frac{e^{Da_i(\theta-b_i)}}{1+e^{Da_i(\theta-b_i)}} \; ; \; I = 1,2,\ldots,n$$

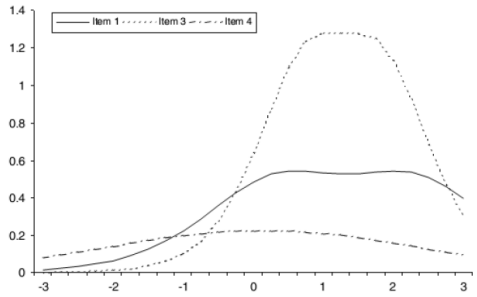


ภาพที่ 3 ลักษณะข้อสอบ (ICC) สำหรับโมเดลการตอบสนองข้อสอบแบบ 3 พารามิเตอร์

โค้งลักษณะข้อสอบ (ICC) สำหรับโมเดลการตอบสนองข้อสอบแบบ 1 พารามิเตอร์ มีลักษณะที่สำคัญคือ b_i เป็นค่าที่แสดงตำแหน่งของ ICC ณ จุด θ ที่มีโอกาสตอบข้อสอบถูก $\frac{1+c_i}{2}$ ส่วน a_i เป็นค่าความชันของ ICC ณ จุด b_i ส่วน C_i คือค่าพารามิเตอร์โอกาสการเดาข้อสอบได้ถูก e =2.718, D =1.7

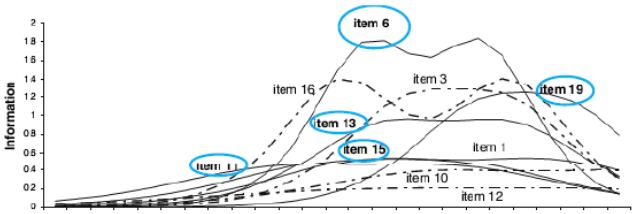
2.3.2 ฟังก์ชันสารสนเทศของข้อสอบและแบบสอบ (Item and Test Information)

(1) ฟังก์ชันสารสนเทศของข้อสอบ (Item Information) เป็นดัชนีผสมที่สร้างจากดัชนีคุณลักษณะของข้อสอบหลายลักษณะ ประกอบด้วย ค่าพารามิเตอร์อำนาจจำแนก ค่าพารามิเตอร์ความยาก ค่าความแปรปรวนคะแนนรายข้อ เพื่อใช้บ่งชี้คุณภาพของข้อสอบ



ภาพที่ 4 ลักษณะกราฟฟังก์ชันสารสนเทศของข้อสอบ (Item Information)

(2) ฟังก์ชันสารสนเทศของแบบสอบ (Test Information) เป็นค่าฟังก์ชันสารสนเทศของแบบสอบ อันเกิดจากผลรวมเชิงพีชคณิตของฟังก์ชันสารสนเทศของข้อสอบแต่ละข้อเข้าด้วยกันทั้งฉบับ ณ ตำแหน่ง "θ" เดียวกัน



ภาพที่ 5 ลักษณะกราฟฟังก์ชันสารสนเทศของแบบสอบ (Test Information)

(3) ความคลาดเคลื่อนมาตรฐานของการประมาณค่า (Standard Error of Estimation) ความคลาดเคลื่อนมาตรฐานของการประมาณค่า [SE("θ")] เป็นส่วนเบี่ยงเบนมาตรฐานของการแจกแจงความน่าจะเป็น ของค่าประมาณความสามารถที่แท้จริง ("θ") ซึ่งเป็นค่าสัดส่วนผกผันกับความถูกต้องแม่นยำของการประมาณค่าความสามารถ หรือค่าฟังก์ชันสารสนเทศของแบบสอบ

จากการศึกษาเอกสารงานวิจัยที่ได้กล่าวไว้ข้างต้นแล้ว การวัดทางการศึกษาและจิตวิทยาเป็นการวัด คุณลักษณะภายในของมนุษย์ ซึ่งไม่สามารถสังเกตได้โดยตรง แต่มีความสำคัญและจำเป็นต้องศึกษา เพราะการวัดคุณลักษณะภายในจะทำให้เข้าใจการเกิดพฤติกรรมภายนอกของมนุษย์ที่สามารถสังเกตได้โดยตรง อันจะนำไปสู่การทำนายควบคุม และพัฒนาพฤติกรรมมนุษย์ การวัดคุณลักษณะภายในของมนุษย์จำเป็นต้องอาศัย ทฤษฎีการทดสอบ เพื่อทำความเข้าใจคุณลักษณะของสิ่งที่มุ่งวัด โครงสร้างของการวัด และการพัฒนาเครื่องมือสำหรับทดสอบ

ทฤษฎีการทดสอบแนวใหม่จึงน่าจะให้ผลการวัดที่ชัดเจนตรงประเด็นมากขึ้น อันจะเป็นประโยชน์ต่อการพัฒนาข้อสอบและแบบสอบให้สามารถนำไปประยุกต์ใช้ในสถานการณ์ต่าง ๆ ของการทดสอบได้อย่างมีประสิทธิภาพ สามารถระบุนัยทั่วไปของคะแนนจริง คุณของแบบสอบตามเงื่อนไขของการทดสอบ รวมทั้งการประเมินความสามารถที่แท้จริงของผู้สอบ และการบรรยายพฤติกรรมการตอบสนองข้อสอบของผู้สอบได้เป็นอย่างดี

2.4 โมเดลการตอบสนองข้อสอบแบบพหุมิติ

แนวคิดของทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ (MIRT) เกิดมาจากแนวคิดพื้นฐานสำคัญ 2 กลุ่มแนวคิด โดยกลุ่มแรกเป็นการศึกษาโดย Spearman (1927) และ Thurstone (1947) ที่พบว่าโมเดล MIRT มีที่มาจากการวิเคราะห์องค์ประกอบ (Factor analysis: FA) เนื่องจากลักษณะของการวิเคราะห์องค์ประกอบจะมีความสัมพันธ์โดยตรงกับการวิเคราะห์ MIRT ส่วนอีกกลุ่มหนึ่งเป็นการศึกษาของ Lazarsfeld (1950) กล่าวว่า MIRT ได้รับอิทธิพลมาจาก UIRT เช่นเดียวกับการวิเคราะห์องค์ประกอบ แต่แนวคิดเกี่ยวกับ UIRT ในช่วงแรกยังไม่เป็นที่นิยมเหมือนเช่นปัจจุบัน (พัชรี จันทร์เพ็ง, 2550) จากแนวคิดทั้งสองแนวคิดอาจกล่าวได้ว่า MIRT เป็นกรณีพิเศษของการวิเคราะห์เชิงสถิติหลายตัวแปรโดยเฉพาะกรณีของการวิเคราะห์องค์ประกอบหรือโมเดลสมการเชิงโครงสร้าง (Structural equation modelling) หรืออาจจะเป็นแนวคิดที่แผ่ขยายมาจากทฤษฎีตอบสนองข้อสอบแบบมิติเดียว (Unidimensional Item Response Theory: UIRT) (Reckase, 2009)

โมเดล MIRT ถือเป็นแนวคิดที่แผ่ขยายมาจากโมเดล UIRT หรือเป็นกรณีเฉพาะของการวิเคราะห์องค์ประกอบหรือโมเดลสมการเชิงโครงสร้างนั่นเองซึ่งจะแตกต่างจากแนวคิดแบบเดิมในส่วนของระเบียบวิธีของ MIRT ที่นำมาประยุกต์และการแปลความหมายของผลที่ได้ โดยการวิเคราะห์ประกอบมุ่งไปที่เทคนิคการสกัดข้อมูลและสังเคราะห์ข้อมูลใหม่ให้มีขนาดเล็กลง (Data Reduction Technique) เพื่อให้ได้มิติที่น้อยที่สุดของข้อมูล และง่ายต่อความเข้าใจสำหรับการบ่งชี้คุณลักษณะต่างๆ ส่วนโมเดล MIRT ไม่ได้มุ่งไปที่เทคนิคการสกัดข้อมูลและสังเคราะห์ข้อมูลใหม่ให้มีขนาดเล็กลง แต่มุ่งไปที่ปฏิสัมพันธ์ระหว่างผู้สอบและข้อสอบ ซึ่งพบว่าในช่วงปลายปี ค.ศ. 1971-1979 ถึงช่วงต้นปี ค.ศ. 1981-1989 เริ่มมีการพัฒนาแนวคิดเกี่ยวกับ MIRT มากขึ้นเพื่อให้เกิดผลในทางปฏิบัติ สิ่งสำคัญที่สุดในการที่จะบอกว่าโมเดลใดเหมาะสมที่สุด คือ กระบวนการในการประมาณค่าพารามิเตอร์ จากการศึกษาของ Mulaik (1972) และ Spray et al. (1990) มีข้อสังเกตว่า โมเดลขาดความชัดเจนในกระบวนการประมาณค่าพารามิเตอร์ จึงยากต่อการพัฒนาแนวคิดต่อไป แสดงให้เห็นว่ากระบวนการประมาณค่าพารามิเตอร์ตามโมเดล MIRT ถือเป็นสิ่งสำคัญมากที่สุดและมีความท้าทายในการพัฒนา เนื่องจากการศึกษาที่ผ่านมาจนถึงปัจจุบันค่าสถิติที่สามารถสังเกตได้ยังไม่เพียงพอต่อ การประมาณค่าพารามิเตอร์ของผู้สอบและข้อสอบได้อย่างเป็นอิสระ ซึ่งประสบปัญหาเช่นเดียวกับการวิเคราะห์องค์ประกอบและโมเดลสมการเชิงโครงสร้าง ในเรื่องของระเบียบวิธีในการประมาณค่าเช่นเดียวกัน ดังนั้น การวิจัยครั้งนี้ ใช้การวิเคราะห์ตามโมเดลการตอบสนองข้อสอบแบบพหุมิติ (MIRT)

โมเดลคุณลักษณะแฝงองค์ประกอบเดียวสามารถพัฒนาสู่โมเดลคุณลักษณะแฝงมากกว่า 1 องค์ประกอบหรือพหุมิติ (Multidimensional IRT Models) ได้ โดยขยายโมเดลให้สามารถรองรับ θ หลายองค์ประกอบ โมเดลการตอบสนองข้อสอบพหุมิติ ถือว่าคุณลักษณะแฝงของบุคคลมากกว่า 1 องค์ประกอบ ส่งผลต่อการตอบข้อสอบ ดังนั้นพารามิเตอร์ความสามารถของผู้สอบจึงมีตั้งแต่ 2 พารามิเตอร์ขึ้นไป การพิจารณาถึงความสามารถหลายมิติของบุคคลน่าจะช่วยทำให้โมเดลมีความ สอดคล้องกับข้อมูล

ได้ดียิ่งขึ้น ถ้าผู้สอบใช้ความสามารถหลายมิติที่ต่างกันอย่างเป็นระบบในการตอบข้อสอบย่อมมีความสัมพันธ์กับระดับความยากและอำนาจจำแนกของข้อสอบรายข้อ และมีผลต่อความน่าจะเป็นในการตอบข้อสอบได้ถูกต้อง (ศิริชัย กาญจนวาสี, 2555)

เมื่อพิจารณาการวิเคราะห์โมเดลการตอบสนองข้อสอบแบบพหุมิติแล้วจะพบว่า สามารถทำได้ 2 วิธีการหลัก ๆ คือ การสังเคราะห์โดยการใช้โมเดลเชิงสำรวจ (Exploratory model) ซึ่งเป็นการประมาณค่าพารามิเตอร์ทั้งข้อสอบและความสามารถของผู้สอบด้วย ในลักษณะที่มากกว่า 1 มิติ เพื่อให้โมเดลและข้อมูลเชิงประจักษ์มีความสอดคล้องกันมากยิ่งขึ้น สำหรับลักษณะการวิเคราะห์นั่นจะไม่ได้กำหนดจำนวนองค์ประกอบคุณลักษณะแฝง แต่เป็นการค้นหาจำนวนองค์ประกอบหรือมิติแฝงของบุคคลที่มีผลต่อการตอบสนองข้อสอบ ซึ่งอาจเรียกได้ว่าเป็น Exploratory MIRT model และอีกวิธีการคือ การวิเคราะห์โดยใช้โมเดลเชิงยืนยัน (Confirmatory model) จะมีลักษณะตรงกันข้ามกับ Exploratory MIRT Model เพราะเป็นการประมาณค่าพารามิเตอร์ของข้อสอบและความสามารถของผู้ตอบข้อสอบ ซึ่งจะประมาณค่าตามจำนวนองค์ประกอบของคุณลักษณะแฝง ซึ่งมีการกำหนดไว้แล้วล่วงหน้า จากโมเดลการตอบสนองข้อสอบแบบพหุมิติของ Embretson and Reise (2000) จากการศึกษาความเป็นมาและพัฒนาการของโมเดลตอบสนองข้อสอบแบบพหุมิติ ถือกว่าจุดสำคัญที่ทำให้ค้นพบแนวคิดที่นำไปสู่การพัฒนาวิธีการศึกษาและพัฒนาแบบพหุมิติ และนำไปสู่การประยุกต์ใช้มากขึ้นซึ่ง (Ackerman, 1994; Adam, Wilson and Wang, 1997) ได้เสนอแนวคิดเกี่ยวกับการพัฒนาโมเดลตอบสนองข้อสอบแบบพหุมิติ และแบ่งประเภทของการตอบสนองข้อสอบแบบพหุมิติออกเป็น 2 ประเภทหลัก ๆ คือ ประเภทคะแนนที่ได้จากการวัดมี 2 ค่า (Dichotomously scored) และประเภทคะแนนที่ได้จากการวัดมีมากกว่า 2 ค่า (Polytomously scored)

ประเภทคะแนนที่ได้จากการวัดมีมากกว่า 2 ค่า (Polytomously scored) เครื่องมือทางการศึกษาและจิตวิทยาหลายประเภท แต่ละข้อคำถามมักมีหลายรายการคำตอบที่กำหนดลำดับหรือ น้ำหนักคะแนนต่างกัน เพราะน่าจะให้สารสนเทศและความเที่ยงจากการตอบที่สูงกว่าการตรวจให้คะแนนแบบ 2 ค่า จึงได้มีการพัฒนาโมเดลการตอบสนอง ข้อสอบสำหรับใช้กับการตรวจให้คะแนนรายข้อมากกว่า 2 ค่า ซึ่งปัจจุบันมีรูปแบบของเครื่องมือ ที่นิยมใช้กัน เช่น มาตรประมาณค่า (Rating scale) การตรวจให้คะแนนความรู้บางส่วน (Partial credit) การตรวจให้คะแนนตามลำดับขั้นของรายการหลายคำตอบ (Ordered-response categories) เป็นต้น (คิริชัย กาญจนวาสี, 2550)

Multidimensional Random Coefficients Multinomial Logit Model: MRCML Model

เป็นโมเดลที่ได้มีการปรับขยายมาจากโมเดลการตอบสนองข้อสอบของราชส์ (Rasch family of items response model) การประยุกต์ใช้โมเดล ดังกล่าวเริ่มต้นจากนักวิชาการหลายคนที่ทำการศึกษาโมเดลนี้ เช่น Adams (1997), Wang (1999), Wang, Wilson and Adams (1997) มีแนวคิดพื้นฐานการพัฒนาโมเดลจากการสร้างโครงร่าง (Building-block) หมายความว่า สำหรับข้อสอบข้อที่ i ซึ่งเป็นข้อสอบที่สามารถตอบได้หลายคำตอบและสามารถให้คะแนนได้หลายค่า (k ค่า) หรืออาจเรียกว่ามิติที่สามารถ

เป็นไปได้ ทั้งหมดซึ่งมีลักษณะเป็นกลุ่มขนาดใหญ่ โดยกำหนดให้เป็น d (d=1,...,D) ส่วนผู้สอบที่ตอบข้อสอบกำหนดให้เป็น p (p=1,...,p) แล้ว ความน่าจะเป็นของผู้สอบที่ตอบข้อสอบข้อที่ i (P_{ik}) ซึ่ง มีอยู่ k กลุ่ม เมื่อเปรียบเทียบกับ k-1 กลุ่ม (P_{ik} - 1) แล้วจะได้เป็นฟังก์ชันเชิงเส้นตรงของความสามารถแฝงบนมิติ (θ_d) และความยากเชิงสัมพัทธ์ของกลุ่ม k (Briggs and Wilson, 2003) ดังนี้

$$K(\delta_{ik}) : \log\left[\frac{P_{ik}}{P_{ik-1}}\right] = \theta_d - \delta_{ik}$$

MRCML Model สามารถเขียนในรูปสมการได้ดังนี้ คือ (Wilson, Adam and Wang, 1997 cited in Briggs and Wilson, 2003)

$$P(X_{ik} = 1;A, B,\xi|\theta) = \frac{\exp(b_{ik}\theta + \acute{a}_{ik}\xi)}{\sum_{k=1}^{k_i}\exp(b_{ik}\theta + \acute{a}_{ik}\xi)}$$

เมื่อ $\acute{a}_{ik}$ ค่าพารามิเตอร์อำนาจจำแนกของข้อสอบ

b_{ik} ค่าพารามิเตอร์ความยากของข้อสอบข้อที่ i

θ พารามิเตอร์ความสามารถของผู้สอบ

จากการศึกษาประเภทของโมเดลการตอบสนองข้อสอบแบบพหุมิติ สรุปได้ว่าโมเดลการตอบสนองข้อสอบแบบพหุมิติสามารถแบ่งออกได้เป็น 2 ประเภท ขึ้นอยู่กับคะแนนที่ได้จากการวัดเป็นเกณฑ์คือ คือ ประเภทคะแนนที่ได้จากการวัดมี 2 ค่า (Dichotomously scored) และประเภทคะแนนที่ได้จากการวัดมีมากกว่า 2 ค่า (Polytomously scored) สำหรับในการวิจัยครั้งนี้ผู้วิจัยได้ใช้โมเดลพหุมิติ (Multidimensional Random Coefficients Multinomial Logit Model: MRCML) ในการประเมินความสามารถของผู้เรียนที่มีความสามารถหลายมิติ โดยได้ทำการศึกษาและนำเสนอดังหัวข้อต่อไปนี้

2.4.1 โมเดลพหุมิติ (Multidimensional Random Coefficients Multinomial Logit Model: MRCML)

การคำนวณความน่าจะเป็นของคะแนนผลสัมฤทธิ์ จะมีค่าไม่น้อยกว่า 0 และมีค่าไม่มากกว่า 1 สำหรับข้อสอบข้อที่ i ซึ่งทำให้ได้ค่าพารามิเตอร์ความยากของข้อสอบ (δi) และทำให้ได้ระดับความสามรถในโครงสร้าง นั่นคือค่า θ ในกรณีที่เป็นการวัดที่เป็นลักษณะมิติเดียว โดยใช้สูตรในการคำนวณ (Rasch, 1960) ดังนี้

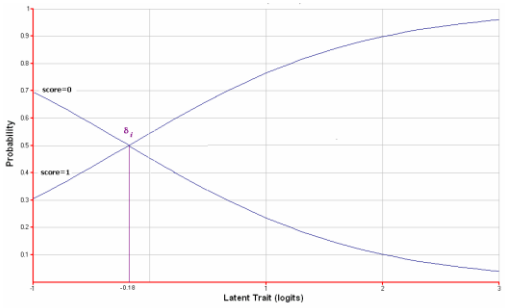
$$P(x_i = 1 \mid \theta, \delta_i) = \frac{P(x=1)}{P(x=0)+P(x=1)} = \frac{\exp(\theta - \delta_i)}{1+\exp(\theta - \delta_i)}$$

สำหรับกรณีที่เป็นข้อสอบแบบให้คะแนนสองค่า โดยจะมีค่าความน่าจะเป็น 2 ค่า จากสมการต่อไปนี้

$$P(x=0) = \frac{1}{1+\exp(\theta - \delta_1)}$$

$$P(x=1) = \frac{\exp(\theta - \delta_1)}{1+\exp(\theta - \delta_1)}$$

เมื่อการตอบข้อสอบสามารถเป็นไปได้เพียงสองค่าเท่านั้น คือ ตอบถูกหรือไม่ถูก ตัวอย่างเช่นบริบททางการศึกษา ค่าความยากของข้อสอบจะแสดงให้เห็นถึงความสามารถที่ผู้สอบสามารตอบถูกในข้อที่ถูกต้องอยู่ในระดับใด ซึ่งสามารถอธิบายค่าความยากของข้อสอบได้ว่าเป็นระดับความสามรถเมื่อโอกาสในการตอบถูกหรือตอบไม่ถูกต้องมีโอกาสเกิดขึ้นเท่า ๆ กัน นั่นคือความน่าจะเป็นของทั้งสองมีค่าอยู่ที่ 0 จากภาพที่ 1 ค่าความยากของข้อสอบมีค่าเท่ากับ -0.18 ซึ่งจุดนี้คือจุดที่เส้นโค้งความน่าจะเป็นของการตอบข้อที่ถูกและข้อที่ไม่ถูกตัดกัน แสดงดังภาพที่ 2



ภาพที่ 6 กราฟความน่าจะเป็นของการตอบสนองข้อสอบข้อที่ถูกและข้อที่ผิด ที่มา: Kennedy (2005)

สำหรับกรณีที่เป็นข้อสอบประเภทให้คะแนนมากกว่า 2 ค่า โมเดลการให้ คะแนนบางส่วน (Masters, 1981) สมการต่อไปนี้จะแสดงให้เห็นถึงความน่าจะเป็นที่ผู้สอบที่มีความสามรถของ θ จะตอบสนองข้อสอบในประเภท C มากว่าประเภทอื่น ๆ สำหรับข้อสอบแต่ละข้อ i ทำให้ได้ค่าพารามิเตอร์ความยากของข้อสอบ

$\xi_i = (\delta_{i1}, \delta_{i2}, \ldots \delta_{im})$

$$P(x_i = c \mid \theta, \xi_i) = \frac{\exp \sum_{j=0}^{c} (\theta - \delta_{ij})}{\sum_{k=0}^{m} \exp \sum_{j=0}^{k} (\theta - \delta_{ij})}$$

โดยที่ m คือ จำนวนขั้นตอน (จำนวนประเภท – 1 สำหรับข้อสอบ)

ตัวอย่าง สำหรับข้อสอบที่มี ประเภท 2 3 ขั้นตอน

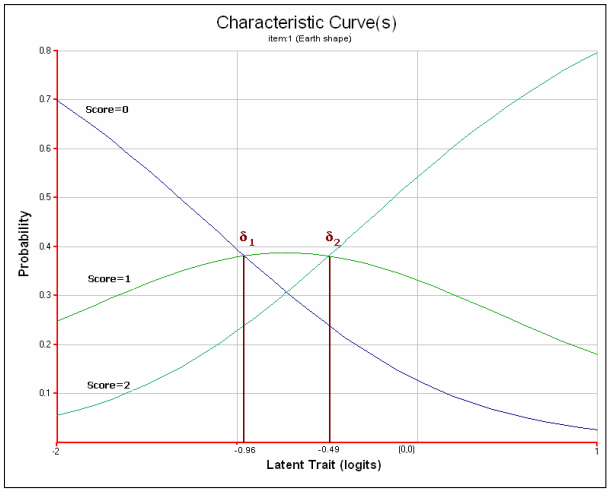
$$P(x_i = 0) = \frac{1}{\sum_{k=0}^{2} \exp \sum_{j=0}^{k} (\theta - \delta_{ij})} = \frac{1}{1 + \exp(\theta - \delta_{i1}) + \exp(2\theta - (\delta_{i1} + \delta_{i2}))}$$

$$P(x_i = 1) = \frac{\exp \sum_{j=0}^{1} (\theta - \delta_{ij})}{\sum_{k=0}^{2} \exp \sum_{j=0}^{k} (\theta - \delta_{ij})} = \frac{\exp(\theta - \delta_{i1})}{1 + \exp(\theta - \delta_{i1}) + \exp(2\theta - (\delta_{i1} + \delta_{i2}))}$$

$$P(x_i = 2) = \frac{\exp \sum_{j=0}^{2} (\theta - \delta_{ij})}{\sum_{k=0}^{2} \exp \sum_{j=0}^{k} (\theta - \delta_{ij})} = \frac{\exp(\theta - \delta_{i1} + \theta - \delta_{i2})}{1 + \exp(\theta - \delta_{i1}) + \exp(2\theta - (\delta_{i1} + \delta_{i2}))} = \frac{\exp(2\theta - (\delta_{i1} + \delta_{i2}))}{1 + \exp(\theta - \delta_{i1}) + \exp(2\theta - (\delta_{i1} + \delta_{i2}))}$$

เมื่อข้อสอบสามารถตอบได้มากกว่า 1 ค่า จำเป็นที่ต้องมีสารสนเทศที่มากกว่าค่าความยากของข้อสอบ และจำเป็นจะต้องรู้วิธีการอื่นๆ ของโครงสร้างให้มากขึ้น ความรู้ พฤติกรรม หรือจิตพิสัย ซึ่งถือว่า

เป็นสิ่งสำคัญมากเพื่อให้ได้ความเป็นไปได้ของคะแนนในแต่ละข้อ กรณีการให้คะแนนแบบบางส่วนถือว่าเป็นส่วนขยายของกรณีที่เป็นการให้คะแนนแบบสองค่า กล่าวคือ การเปลี่ยนจากประเภทหนึ่งไปสู่ประเภทอื่น ๆ หมายถึงว่า ข้อสอบแบบตัวเลือกที่เป็นประเภทให้คะแนนสองค่าจะมีการเปลี่ยนแปลงอยู่ 2 ระดับ ตัวอย่างเช่น เมื่อลองพิจารณา ข้อสอบหนึ่งข้อกับจำนวนประเภท 3 ประเภท ที่มีการให้คะแนนเป็น 0, 1 หรือ 2 มีค่าความยากสำหรับขั้นที่ 1 คือ δ_1 ซึ่งเป็นตำแหน่งที่ตั้งอยู่ของค่าความยาก ถ้าหากพิจารณาเพียงแค่ประเภท 0 และ 1 จะมีค่าเท่ากันกับข้อสอบที่ให้คะแนนเพียงบางส่วน ที่ถ้าตอบถูก x = 1 หรือตอบไม่ถูก x = 0 จากภาพที่ 2 แสดงให้เห็นว่าเป็นโค้งของการตอบข้อที่มีคะแนนเท่ากับ 0 ตัดกับโค้งของการตอบข้อที่มีคะแนนเท่ากับ 1 ซึ่งก็จะทำให้ได้ค่าความยากสำหรับการอธิบายเช่นเดียวกัน ค่าความยากสำหรับขั้นที่ 2, δ_2 คือความสามารถที่ต้องการ ซึ่งจะมีค่าเท่ากับความน่าจะเป็นของการตอบได้คะแนนเท่ากับ 2 หรือเท่ากับ 1 สำหรับข้อสอบข้อนั้น ดังตัวอย่างปรากฏในภาพ



ภาพที่ 7 กราฟความน่าจะเป็นของการตอบสนองข้อสอบข้อที่ถูกและข้อที่ผิด ใน 2 มิติ

ที่มา: Kennedy (2005)

เมื่อใช้โมเดลแบบให้คะแนนบางส่วน โดยปกติแล้วค่าพารามิเตอร์ความยากของคะแนนด้าน j สำหรับข้อสอบข้อที่ i เขียนแทนด้วย δ_{ij} ซึ่ง δ_{ij} เป็นระดับความสามารถที่ความหวัง ซึ่งมีโอกาสในการที่จะเกิดขึ้นที่เท่าเทียมกัน ในประเภท j หรือประเภท j – 1 สำหรับข้อสอบข้อที่ i อีกวิธีหนึ่งเราอาจจะใช้ค่าเฉลี่ยของ δ_{ij} เป็นค่าความยากโดยรวมของข้อสอบ และสำหรับการหาค่าความยากสำหรับแต่ละขั้นตอนก็สามารถหาได้จากค่าเฉลี่ยในแต่ละขั้นตอน ในการหาค่าความยากของข้อสอบด้วยวิธีนี้ โดยจะหาในแต่ละ δ_{ij} สามารถคำนวณได้จาก $\delta_{ij} + \tau_{ij}$ เมื่อ τ_{ij} คือ ส่วนเบี่ยงเบนจากค่าเฉลี่ยความยากของข้อสอบข้อที่ i ขั้นตอนที่ j สำหรับในกรณีนี้

2.4.2 การประยุกต์ใช้โมเดล MRCML (Adams, Wilson, & Wang, 1997) โดยผสมผสานร่วมกันระหว่าง Rasch Model (Rasch, 1960) และ Partial Credit Model (Masters, 1982) โดยใช้โปรแกรม ACER ConQuest Version 2.0 (Wu, Adam, Wilson, & Haldane, 2007) ประกอบด้วยข้อมูลเชิงปริมาณที่ได้จากการตรวจสอบคุณภาพของกรอบการประเมิน ซึ่งพิจารณารายละเอียดดังต่อไปนี้

(1) คุณภาพของข้อสอบรายข้อ โดยนำผลที่ได้สู่การปรับปรุงและพัฒนาข้อสอบรายข้อ ประกอบด้วย

(1.1) ค่าความยาก (Estimate: b) ซึ่งโดยทั่วไปค่าที่ยอมรับได้อยู่ในช่วงความยาก -3 ถึง +3

(1.2) โค้งคุณลักษณะของข้อสอบแต่ละข้อ (ICC) เพื่อบ่งชี้ว่าข้อสอบแต่ละข้อเหมาะสมกับผู้สอบที่มีความสามารถอยู่ในระดับใด

(1.3) อำนาจจำแนกตามทฤษฎีการทดสอบแบบดั้งเดิม (CTT) โดยพิจารณาจากค่า Point Biserial (r) โดยปกติควรมีค่าตั้งแต่ 0.2 ขึ้นไป และมีนัยสำคัญ

(1.4) ประสิทธิภาพตัวลวงในกรณีที่เป็นข้อสอบแบบปรนัยตามทฤษฎีการทดสอบแบบดั้งเดิม (CTT) โดยพิจารณาจากจำนวนของผู้สอบในการเลือกตัวลวงแต่ละค่า และพิจารณาจากโค้ง ICC ของตัวลวง ซึ่งปรากฏอยู่ในไฟล์ .itn เพื่อพิจารณาว่าตัวลวงแต่ละตัวเหมาะสำหรับผู้สอบที่ความสามารถในระดับใด โดยพิจารณาจากกราฟ Modelled and Empirical Category Characteristics Curves (MCC)

(1.5) ค่าสารสนเทศของข้อสอบรายข้อ

(1.6) ค่าความคลาดเคลื่อนในการประมาณค่า โดยพิจารณาจากค่า MNSQ ใน Weight Fit โดยมีค่าอยู่ในช่วง CI โดยควรมีค่าความเชื่อมั่นที่ยอมรับในระดับร้อยละ 95 ของแต่ละข้อ (ค่า MNSQ อยู่ในช่วง CI)

(1.7) พิจารณาค่าขนาดอิทธิพล (Effect Size) โดยพิจารณาจากค่า MNSQ ควรอยู่ในช่วง 0.75-1.33 ซึ่งเป็นช่วงที่ยอมรับได้

(1.8) พิจารณาค่า T ซึ่งเป็นค่าสะท้อนถึงความคลาดเคลื่อนที่ยอมรับได้ โดยเปรียบเทียบกับโค้ง ICC ที่คาดหวังกับ โค้ง ICC ที่เก็บข้อมูลได้จริงจากข้อมูลเชิงประจักษ์ โดยค่าที่ยอมรับได้อยู่ในช่วง -2 ถึง +2

(2) คุณภาพของแบบทดสอบทั้งฉบับ ประกอบด้วย

(2.1) ค่าความเที่ยงแบบสอดคล้องภายใน (สัมประสิทธิ์แอลฟาของครอนบาค; α) ตามทฤษฎีการทดสอบแบบดั้งเดิมมีค่าตั้งแต่ 0.7

(2.2) ค่าฟังก์ชันสารสนเทศของแบบทดสอบ (TCC) เป็นค่าที่สะท้อนถึงความคงเส้นคงวาของการประมาณค่าของแบบทดสอบทั้งสอบทั้งฉบับ ว่าแบบทดสอบทั้งฉบับเหมาะสำหรับผู้สอบที่มีความสามารถที่แท้จริง (θ) ที่ระดับใดจึงจะเหมาะสม ซึ่งตำแหน่งดังกล่าวจะเป็น

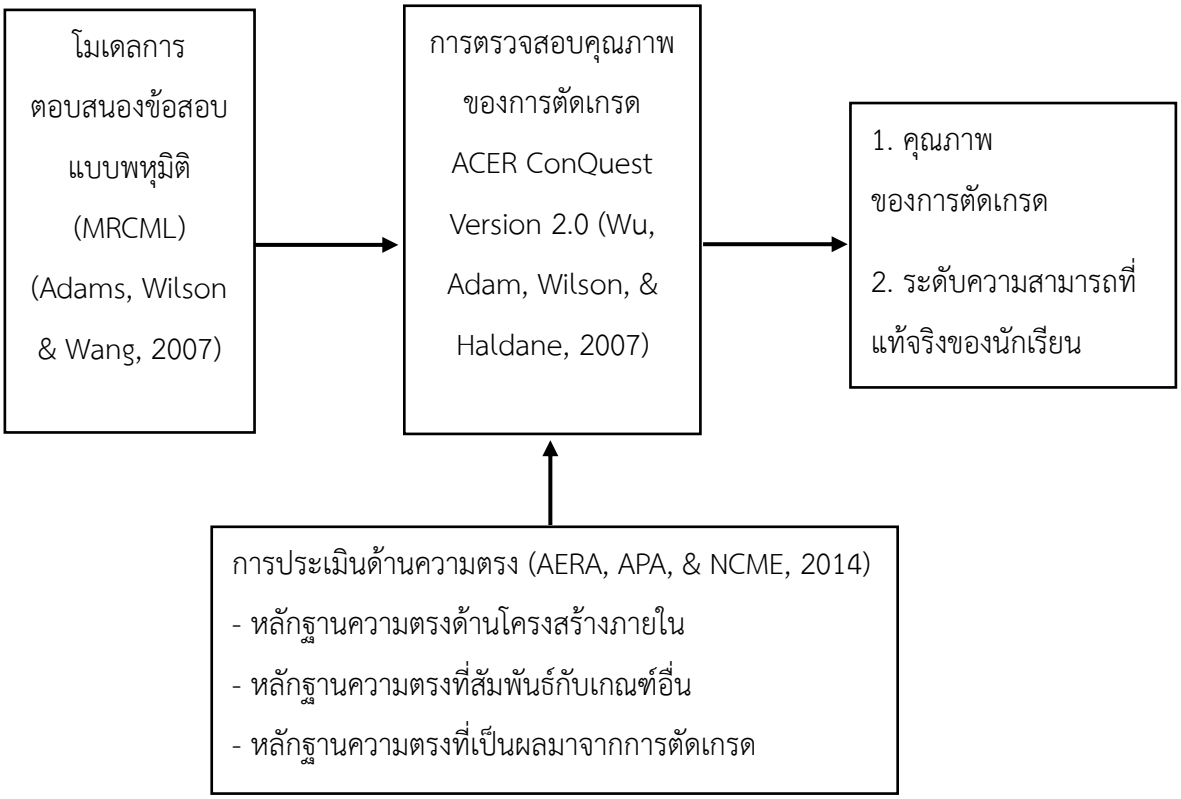
ตำแหน่งที่มีความคลาดเคลื่อนในการประมาณค่า (SE) ต่ำที่สุด โดยค่าฟังก์ชันสารสนเทศของแบบทดสอบจะสะท้อนถึงค่าความเที่ยง (Reliability)

(2.3) ค่าความตรงเชิงโครงสร้าง โดยพิจารณาจาก Wright Map สะท้อนถึงความตรงเชิงโครงสร้างภายในของข้อสอบแต่ละข้อที่สัมพันธ์กับผลการตอบของนักเรียนและแผนที่ความก้าวหน้าที่พัฒนาขึ้นสอดคล้องกับข้อมูลเชิงประจักษ์หรือไม่ โดยแสดงให้เห็นถึงความน่าจะเป็นในการตอบข้อสอบจากระดับขั้นหนึ่งไปยังระดับขั้นที่สูงกว่า (ระดับ k-1 ไปยังระดับ k) ของข้อสอบข้อที่ i ซึ่งในที่นี้ เรียกว่า “Item Thresholds” ในกรณีที่มีการตรวจให้คะแนนมากกว่า 1 ค่า เช่น ข้อสอบอัตนัย ซึ่งในการวิจัยครั้งนี้เลือกใช้โมเดลที่เรียกว่า “Partial Credit Model” ซึ่งระดับขั้นของการตอบ (Step) มีค่าเท่ากับ k-1 เมื่อ k คือระดับ (Category) ของแผนภาพความก้าวหน้าในการเรียนรู้ (Progress Map)

2.5 กรอบแนวคิดการวิจัย

การวิจัยครั้งนี้มีจุดประสงค์เพื่อตรวจสอบคุณภาพของการตัดเกรดในแต่ละรายวิชา และตรวจสอบความสอดคล้องของระดับความสามารถที่แท้จริงของนักเรียนระดับชั้นมัธยมศึกษาปีที่ 6 กับการเรียน 6 ภาคเรียน

ในการเรียนรู้ในแต่ละรายวิชาไม่ได้เรียนรู้เพียงด้านใดด้านหนึ่งแต่ประกอบไปด้วยหลากหลายด้าน การวัดเพียงมิติใดมิติหนึ่ง อาจไม่เพียงพอต่อสถานการณ์ในปัจจุบัน จึงจำเป็นต้องนำเทคนิคการวิเคราะห์ข้อมูลโดยใช้โมเดลการตอบสนองข้อสอบแบบพหุมิติ (Multidimensional Random Coefficients Multinomial Logit: MRCML) (Adams, Wilson & Wang, 1997) มาใช้เป็นแนวทางในการประเมินความสามารถของผู้เรียน เพื่อให้ผลการประเมินของผู้เรียนมีความถูกต้องยิ่งขึ้น โดยใช้โปรแกรม ACER ConQuest Version 2.0 (Wu, Adam, Wilson, & Haldane, 2007) มีการตรวจสอบความถูกต้องจากหลักฐานการประเมินด้านความตรง ตามเกณฑ์มาตรฐานการประเมินทางด้านการศึกษาและจิตวิทยาที่มีคุณภาพ (AERA, APA, & NCME, 2014) ได้แก่ (1) หลักฐานความตรงด้านโครงสร้างภายใน (Validity Evidence Based on Internal Structure) (2) หลักฐานความตรงที่สัมพันธ์กับเกณฑ์อื่น (Evidence Based on Relations to Other Variables) และ (3) หลักฐานความตรงที่เป็นผลมาจากการตัดเกรด (Evidence Based on Consequences of Testing) โดยสามารถแสดงกรอบแนวคิดในการวิจัย ดังภาพที่ 8



ภาพที่ 8 กรอบแนวคิดในการวิจัย

บทที่ 3

วิธีดำเนินการวิจัย

การวิจัยครั้งนี้ผู้วิจัยประยุกต์ใช้ทฤษฎีการตอบสนองข้อสอบเป็นฐานสำคัญในการตรวจสอบคุณภาพ ภายใต้การออกแบบการวิจัยผสมผสานระหว่างการออกแบบการวิจัยเชิงปริมาณและเชิงคุณภาพประเภทอธิบายตามกาลเวลา (Explanatory Sequential Mixed Method Design) (Creswell & Clark, 2010) โดยช่วงแรกเน้นการเก็บรวบรวมข้อมูลและวิเคราะห์ข้อมูลเชิงปริมาณ จากนั้นนำผลที่ได้ไปสู่การเก็บรวบรวมข้อมูลเชิงปริมาณ จากนั้นนำผลที่ได้ไปสู่การเก็บรวบรวมข้อมูลเชิงคุณภาพ เพื่อสนับสนุนข้อมูลที่ได้จากการวิเคราะห์ข้อมูลเชิงปริมาณ แบ่งการดำเนินการวิจัยเป็น 2 ระยะ แต่อย่างไรก็ตามในการวิจัยครั้งนี้ผู้วิจัยขอนำเสนอเฉพาะกรณีระยะที่ 1 ซึ่งมุ่งเน้นการวิเคราะห์ข้อมูลโดยการพิจารณาจาก Wright Map มาใช้ในการวิจัยเพื่อตอบคำถามในการวิจัยข้อที่ 1 และ 2 ในคราวเดียวกัน โดยมีขั้นตอนการวิจัยดังนี้

1. รวบรวมและตรวจสอบข้อมูลผลการเรียน
2. กำหนดระดับของคะแนนตามเกณฑ์การตัดเกรด
3. วิเคราะห์ข้อมูลเพื่อตรวจสอบคุณภาพการตัดเกรดและคุณภาพผู้เรียน
4. ตรวจสอบหลักฐานการประเมินด้านความตรงตามเกณฑ์มาตรฐานการประเมิน

รายละเอียดเกี่ยวกับวิธีดำเนินการวิจัย มีดังนี้

1. รวบรวมและตรวจสอบข้อมูลผลการเรียน

ขั้นตอนการรวบรวมข้อมูลและการตรวจสอบสอบผลการเรียน มีขั้นตอนรายละเอียดดังนี้

1.1 ศึกษาและพิจารณาโครงสร้างหลักสูตรสถานศึกษาของโรงเรียนบ้านฉางกาญจนกุลวิทยา ระดับชั้นมัธยมศึกษาตอนปลาย จะพบว่า รายวิชาเรียนจะแบ่งออกเป็น 2 กลุ่มหลัก คือ (1) กลุ่มรายวิชาพื้นฐาน คือ รายวิชาที่นักเรียนในทุกห้องเรียนและทุกแผนการเรียนจะต้องได้เรียน และ (2) กลุ่มรายวิชาเพิ่มเติม คือ รายวิชาที่นักเรียนในทุกห้องเรียนและทุกแผนการเรียนมีรายวิชาที่ไม่เหมือนกัน แตกต่างกันออกไปตามแผนการเรียน ดังนั้นเพื่อให้สามารถตรวจสอบคุณภาพการตัดเกรด และคุณภาพผู้เรียนมีความเหมาะสมมากที่สุด ต้องใช้รายวิชาที่นักเรียนทุกห้องเรียนได้เรียนเพื่อเป็นตัวแทนคุณภาพเกรดที่เหมาะสมในการวิเคราะห์ข้อมูล ผู้วิจัยจึงได้เลือกใช้ กลุ่มรายวิชาพื้นฐาน ในภาคเรียนที่ 1 และ 2 ปีการศึกษา 2566 ของระดับชั้นมัธยมศึกษาปีที่ 6 ซึ่งจำนวนทั้งหมด 17 รหัสวิชา ได้แก่ ภาษาไทย คณิตศาสตร์ วิทยาศาสตร์ วิทยาการคำนวณ การออกแบบและเทคโนโลยี สังคมศึกษา ภาษาอังกฤษ สุขศึกษาและพลศึกษา ดนตรีและนาฏศิลป์ การงานอาชีพ

1.2 การเก็บข้อมูลผลการเรียน ในการรวบรวมข้อมูลผู้วิจัยต้องการข้อมูลที่เป็นตัวแทนความสามารถของผู้เรียนให้ครอบคลุมทุกแผนการเรียน เพื่อให้ได้ผลลัพธ์ที่มีความตรงที่สุด มีความเชื่อถือได้

และความยุติธรรม (High Quality Evidence) จึงได้มีสุ่มกลุ่มตัวอย่างโดยวิธีการสุ่มตัวอย่างแบบแบ่งพวก (Stratified Sampling) คือ การสุ่มโดยแบ่งประชากรออกเป็นแต่ละชั้น (ห้อง) ซึ่งนักเรียนชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2566 มีจำนวนทั้งหมด 8 ห้องเรียนและใช้วิธีการสุ่มตัวอย่างแบบง่าย (Simple random sampling) จากแต่ละชั้น (ห้อง) โดยเลือกมาห้องเรียนละ 20 คน จะได้กลุ่มตัวอย่างทั้งหมด 160 คน

1.3 การตรวจสอบข้อมูลผลการเรียน เนื่องจากการวิเคราะห์ข้อมูลโดยใช้โมเดลโลจิสตามทฤษฎีการตอบสนองแบบพหุมิติ (Multidimensional Random Coefficients Multinomial Logit Model) จะต้องใช้ข้อมูลเป็นตัวเลขในการวิเคราะห์ ทางผู้วิจัยจึงได้ทำการกรองข้อมูลที่มีผลการเรียนเป็น ร และ มส ทิ้ง และเลือกใช้เฉพาะข้อมูลที่เป็นตัวเลข ตั้งแต่ 0 ถึง 4 เท่านั้น ดังตารางที่ 2

คนที่	**รายวิชาที่**																
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**
01	2.5	4	3.5	2.5	2	3.5	4	4	2	4	2.5	3	3.5	4	3.5	3.5	2
02	4	4	4	3.5	4	4	2	4	3.5	4	4	4	4	4	4	3.5	4
03	3	1	3.5	3	4	4	2	3	2	2.5	3	3	4	4	3	3.5	2
04	4	3	3.5	3.5	4	4	3.5	4	3	3.5	2.5	4	4	4	4	4	3
05	4	4	4	3	4	4	4	4	3.5	4	4	4	4	4	4	3.5	2.5
06	2.5	3	3.5	3	3.5	4	3	3	2.5	3	3.5	4	3.5	4	4	3.5	2.5
07	4	4	4	3.5	4	4	4	4	4	4	4	4	4	4	4	4	4
08	4	4	4	4	4	4	4	4	4	3.5	4	4	4	4	4	4	4
09	4	4	4	3.5	4	4	3.5	3.5	4	3.5	4	4	4	4	4	3.5	4
10	3.5	3	3.5	3	4	4	2.5	3.5	3	2.5	2	4	4	4	4	4	2.5
11	1.5	2.5	3.5	3.5	4	4	1	3	1.5	2.5	2	2.5	4	4	3	3	2
12	2.5	3	4	3.5	4	4	2.5	2.5	2	2	1.5	2	4	4	4	3.5	2
13	3.5	2.5	4	4	4	4	3	3.5	3.5	3	3	4	4	4	4	3.5	2.5
14	4	3	4	3.5	4	4	4	4	3.5	3	3.5	4	4	4	4	4	2.5
15	3.5	4	4	3.5	4	4	3.5	4	4	3.5	3.5	4	4	4	4	3.5	3
16	3.5	3.5	4	4	4	4	3.5	4	3	4	3	4	4	4	4	3.5	3
17	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
18	1	1	1	3	4	4	1	3	2	2.5	2	1	4	4	1	3.5	1.5
19	2.5	3	4	3	4	4	3	3.5	2.5	3	2.5	3	4	4	3.5	3.5	3
20	3	2.5	4	3	4	4	3	4	2.5	2.5	2	4	4	4	4	3.5	2.5

ตารางที่ 2 ตัวอย่างผลการเรียนทำการกรอง

2. กำหนดระดับของคะแนนตามเกณฑ์การตัดเกรด

ขั้นตอนการกำหนดระดับของคะแนนตามเกณฑ์การตัดเกรด มีขั้นตอนรายละเอียดดังนี้

2.1 กำหนดระดับของคะแนนตามเกณฑ์การตัดเกรด ผู้วิจัยได้ศึกษาการให้ระดับผลการเรียนตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 กำหนดการตัดสินเพื่อให้ระดับผลการเรียนรายวิชาของกลุ่มสาระการเรียนรู้ ให้ใช้ตัวเลขแสดงผลการเรียนเป็น 8 ระดับสำหรับระดับมัธยมศึกษาตอนต้นและตอนปลาย ใช้ตัวเลขแสดงระดับผลการเรียนเป็น 8 ระดับ แนวการให้ระดับผลการเรียน 8 ระดับ และความหมายของแต่ละระดับ

2.2 กำหนดระดับของคะแนนตามเกณฑ์การตัดเกรดให้สอดคล้องกับโมเดลโลจิสตามทฤษฎีการตอบสนองแบบพหุมิติ (Multidimensional Random Coefficients Multinomial Logit Model) ในที่นี้คณะผู้วิจัยใช้ชื่อย่อว่า “MRCLM” ดังตารางที่ 3

ระดับผลการเรียน	ความหมาย	ระดับความสามารถตาม MRCLM	STEP
4	ดีเยี่ยม	7	6
3.5	ดีมาก	6	5
3	ดี	5	4
2.5	ค่อนข้างดี	4	3
2	ปานกลาง	3	2
1.5	พอใช้	2	1
1	ผ่านเกณฑ์ขั้นต่ำ	1	
0	ต่ำกว่าเกณฑ์	0	

ตารางที่ 3 กำหนดระดับของคะแนนตามเกณฑ์โมเดล MRCLM

3. วิเคราะห์ข้อมูลเพื่อตรวจสอบคุณภาพการตัดเกรดและคุณภาพผู้เรียน

ขั้นตอนวิเคราะห์ข้อมูลเพื่อตรวจสอบคุณภาพการตัดเกรดและคุณภาพผู้เรียน มีขั้นตอนรายละเอียดดังนี้

3.1 กำหนดตัวแปรตามโมเดล MRCLM ซึ่งซึ่งจะประกอบไปด้วย 2 เมตริกซ์ คือ เมตริกซ์การออกแบบ (Design Metrix) และเมตริกซ์การให้คะแนน (Scoring Metrix)

สำหรับเมตริกซ์การให้คะแนน (Scoring Metrix) แทนด้วยสัญลักษณ์ $\mathbf{b}_{ik}=(b_{ik1},\ldots,b_{ikV})'$ และทักจะแทนด้วยสัญลักษณ์ว่า B (Scoring Matrix **B**) เป็นเมตริกซ์แสดงความสัมพันธ์ระหว่างข้อสอบและมิติความสามารถที่แท้จริงซึ่งแสดงเป็นฟังก์ชันของการให้คะแนน เช่น การตอบสนองในระดับที่ j ในแผนที่ความก้าวหน้าในข้อที่ i ซึ่งถูกแทนด้วย เวคเตอร์ V ซึ่งเป็นเวคเตอร์แสดงความสามารถของผู้สอบในแต่ละมิติ แทนด้วยสัญลักษณ์ vector $\boldsymbol{\theta} = (\boldsymbol{\theta_1}, \ldots, \boldsymbol{\theta_V})$. เช่น ด.ช. บอย มีความสามารถในการแก้ปัญหาทางคณิตศาสตร์ 2 ด้าน คือ ด้านกระบวนการแก้ปัญหาในระดับปานกลาง และด้านการใช้กลยุทธ์

ในการแก้ปัญหาระดับสูง แสดงว่า ความสามารถในการแก้ปัญหาครั้งนี้ประกอบด้วย 2 มิติ ซึ่งจะมีความสามารถที่แท้จริง 2 ค่า คือ θ_1 และ θ_2 ซึ่งจำนวนของค่าความสามารถที่แท้จริงของผู้สอบแต่ละคนจะขึ้นอยู่กับจำนวนมิติหรือจำนวนด้านที่ต้องการศึกษานั่นเอง

จำนวนค่าพาราเตอร์ของข้อสอบ แทนด้วย p แทนด้วยเวคเตอร์ $\boldsymbol{\xi}=(\xi_1, \ldots, \xi_p)$ ซึ่งแสดงความสัมพันธ์ระหว่างการตอบสนองข้อสอบกับค่าพารามิเตอร์ ซึ่งในที่นี้เรียกว่าเมตริกซ์การออกแบบ A (Design Matrix **A**) แสดงด้วยสัญลักษณ์ $\mathbf{a}_{ij}$ ในแต่ละค่าพารามิเตอร์ซึ่งเมื่อนำมาคูณด้วยเวคเตอร์ $\boldsymbol{\xi}$ ซึ่งเป็นการดำเนินการรวมกันเชิงเส้นตรงของค่าพารามิเตอร์สำหรับการตอบสนองข้อสอบระดับ j ข้อที่ i

ความน่าจะเป็นสำหรับการตอบสนองระดับที่ j ของข้อที่ i สามารถแสดงได้ดังสมการที่ 1

$$\mathrm{P}(X_{ij}=1|A,B,\xi|\theta) = \frac{\exp\left(b_{ij}\theta+a'_{ij}\xi\right)}{\sum_{k=1}^{J_i}\exp\left(b_{ik}\theta+a'_{ik}\xi\right)}$$

ซึ่งในงานวิจัยครั้งนี้ผู้วิจัยได้กำหนดตัวแปรที่ต้องการศึกษาดังรายละเอียดต่อไปนี้

1) รายวิชาพื้นฐานทั้ง 17 รหัสวิชา เป็นรายการประเมิน (Item)

2) นักเรียนกลุ่มตัวอย่างทั้ง 160 คน กำหนดเป็นผู้เข้าสอบ (Examinees)

3.2 กำหนดชื่อรายวิชาให้มีความเหมาะสมกับการประมวลผลโดยใช้โปรแกรมConQuest 2.0 (Wu, Adums, Wilson, & Haldane, 2007) ดังนี้

ที่	รหัส	ความหมาย	ที่	รหัส	ความหมาย
1	THAI1	ภาษาไทย ภาคเรียนที่ 1	10	THAI2	ภาษาไทย ภาคเรียนที่ 2
2	MATH1	คณิตศาสตร์ ภาคเรียนที่ 1	11	MATH2	คณิตศาสตร์ ภาคเรียนที่ 2
3	SCI1	วิทยาศาสตร์ ภาคเรียนที่ 1	12	COM2	วิทยาการคำนวณ ภาคเรียนที่ 2
4	COM1	โครงงานคอมฯ ภาคเรียนที่ 1	13	SOC2	สังคมศึกษา ภาคเรียนที่ 2
5	SOC1	สังคมศึกษา ภาคเรียนที่ 1	14	PHE2	สุขและพลศึกษา ภาคเรียนที่ 2
6	PHE1	สุขและพลศึกษา ภาคเรียนที่ 1	15	ART2	ศิลปะ ภาคเรียนที่ 2
7	ART1	ศิลปะ ภาคเรียนที่ 1	16	WORK2	การงานอาชีพ ภาคเรียนที่ 2
8	WORK1	การงานอาชีพ ภาคเรียนที่ 1	17	ENG2	ภาษาอังกฤษ ภาคเรียนที่ 2
9	ENG1	ภาษาอังกฤษ ภาคเรียนที่ 1			

ตารางที่ 4 กำหนดชื่อรายวิชาให้มีความเหมาะสมกับการประมวลผล

3.2 การวิเคราะห์เพื่อตรวจสอบคุณภาพของการตัดเกรดและคุณภาพผู้เรียน (Measurement Model/Wright Map) โดยใช้โปรแกรม ConQuest 2.0 (Wu, Adums, Wilson, & Haldane, 2007) เมื่อพิจารณาจาก Wright Map ที่แสดงถึงลักษณะการกระจายความสามารถที่แท้จริงของผู้สอบที่มีความสัมพันธ์กับความยากของแต่ละรายวิชา พร้อมทั้งเป็นดัชนีบ่งชี้คุณภาพของโมเดลการวัด

3.2.1 การพิจารณาความเหมาะสมของแต่ละรายวิชา โดยมีการพิจารณาดังนี้

1) ค่าความยาก (Estimate: b) หมายถึง ความยากง่ายในการได้เกรดสูง ซึ่งโดยทั่วไปค่าที่ยอมรับได้อยู่ในช่วงความยาก -3 ถึง +3

2) โค้งคุณลักษณะของข้อสอบแต่ละข้อ (ICC) เพื่อบ่งชี้ว่าแต่ละรายวิชามีการตัดเกรดเหมาะสมกับผู้สอบที่มีความสามารถอยู่ในระดับใด

3) ค่าความคลาดเคลื่อนในการประมาณค่า โดยพิจารณาจากค่า MNSQ ใน Weight Fit โดยมีค่าอยู่ในช่วง CI โดยควรมีค่าความเชื่อมั่นที่ยอมรับในระดับร้อยละ 95 ของแต่ละวิชา

4) ค่า T ซึ่งเป็นค่าสะท้อนถึงความคลาดเคลื่อนที่ยอมรับได้โดยเปรียบเทียบกับโค้ง ICC ที่คาดหวังกับ โค้ง ICC ที่เก็บข้อมูลได้จริงจากข้อมูลเชิงประจักษ์ โดยค่าที่ยอมรับได้อยู่ในช่วง -2 ถึง +2

ซึ่งการพิจารณาความเหมาะสมของแต่ละรายวิชา ดังแสดงในตารางที่ 4

ที่	วิชา	Estimate	Weight Fit			แปลผล
			MNSQ	CI	T	
1	THAI1	0.406	0.78	(0.78, 1.22)	-2.2	ค่า MNSQ อยู่ในช่วงที่ยอมรับได้
2	MATH1	0.863	1.07	(0.78, 1.22)	0.6	ค่า MNSQ อยู่ในช่วงที่ยอมรับได้
3	SCI1	-0.259	0.75	(0.72, 1.28)	-1.9	ค่า MNSQ อยู่ในช่วงที่ยอมรับได้
4	COM1	0.081	1.23	(0.78, 1.23)	1.9	ค่า MNSQ อยู่ในช่วงที่ยอมรับได้
5	SOC1	-1.242	1.15	(0.54, 1.46)	0.7	ค่า MNSQ อยู่ในช่วงที่ยอมรับได้
6	PHE1	-1.621	1.42	(0.15, 1.85)	0.9	ค่า MNSQ อยู่ในช่วงที่ยอมรับได้
7	ART1	0.960	1.24	(0.76, 1.24)	1.9	ค่า MNSQ อยู่ในช่วงที่ยอมรับได้
8	WORK1	0.009	0.81	(0.76, 1.24)	-1.6	ค่า MNSQ อยู่ในช่วงที่ยอมรับได้
9	ENG1	0.718	0.94	(0.78, 1.22)	-0.5	ค่า MNSQ อยู่ในช่วงที่ยอมรับได้
10	THAI2	0.193	0.67	(0.67, 1.22)	-3.3	ค่า MNSQ อยู่ในช่วงที่ยอมรับได้
11	MATH2	0.981	1.16	(0.77, 1.23)	1.4	ค่า MNSQ อยู่ในช่วงที่ยอมรับได้
12	COM2	0.348	1.24	(0.78, 1.24)	2.0	ค่า MNSQ อยู่ในช่วงที่ยอมรับได้
13	SOC2	-1.027	1.27	(0.68, 1.32)	1.6	ค่า MNSQ อยู่ในช่วงที่ยอมรับได้

ที่	วิชา	Estimate	Weight Fit			แปลผล
			MNSQ	CI	T	
14	PHE2	-1.699	1.05	(0.18, 1.82)	0.2	ค่า MNSQ อยู่ในช่วงที่ยอมรับได้
15	ART2	-0.157	1.33	(0.73, 1.33)	2.2	ค่า MNSQ อยู่ในช่วงที่ยอมรับได้
16	WORK2	0.305	0.95	(0.73, 1.27)	-0.3	ค่า MNSQ อยู่ในช่วงที่ยอมรับได้
17	ENG2	1.142	0.89	(0.78, 1.22)	-1.0	ค่า MNSQ อยู่ในช่วงที่ยอมรับได้

ตารางที่ 5 การพิจารณาความเหมาะสมของแต่ละรายวิชา

3.2.2 การพิจารณาความตรงของแต่ละรายวิชา

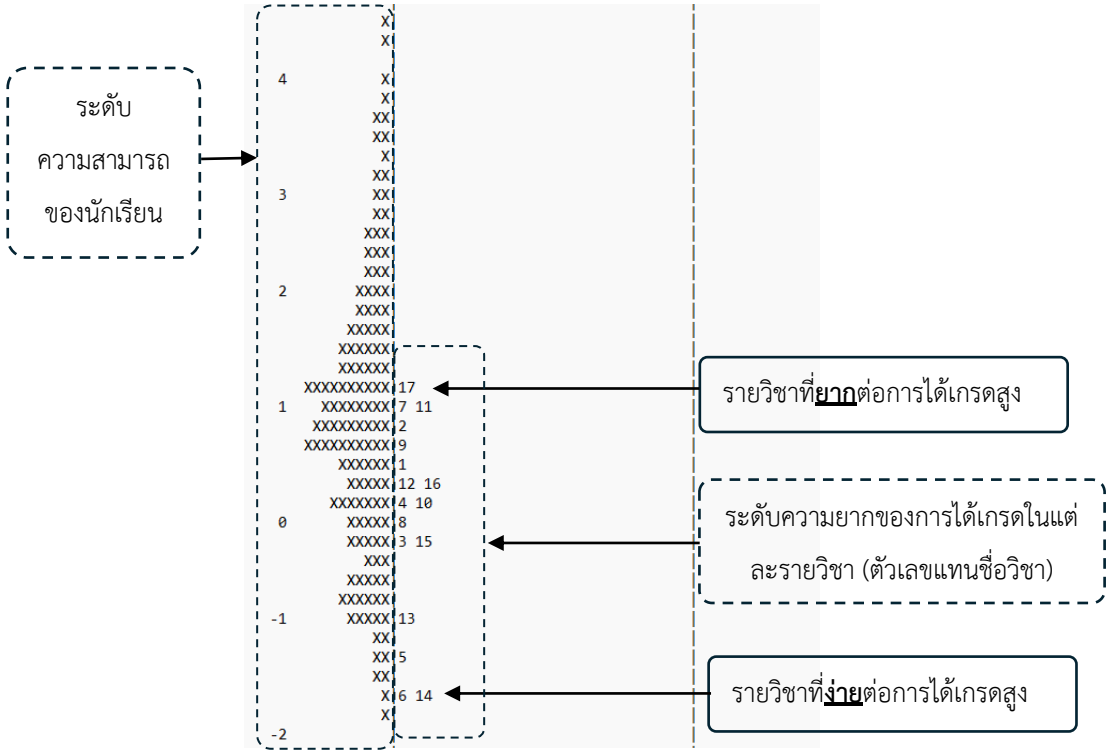
1) ค่าความเที่ยงแบบสอดคล้องภายใน (สัมประสิทธิ์แอลฟาของครอนบาค; α) ตามทฤษฎีการทดสอบแบบดั้งเดิม ควรมีค่าตั้งแต่ 0.7 ขึ้นไป

ภาพที่ 9 Test Information

EAP/PV RELIABILITY: 0.955

Coefficient Alpha: 0.95

2) ค่าความตรงเชิงโครงสร้าง โดยพิจารณาจาก Wright Map สะท้อนถึงความตรงเชิงโครงสร้างภายในของแต่ละรายวิชาที่สัมพันธ์กับผลการเรียนของนักเรียน



ภาพที่ 10 Wright Map แสดงคุณภาพของการตัดเกรดและความสามารถของนักเรียน

จากภาพที่ 10 จะแสดงให้เห็นจำนวนนักเรียนในแต่ละระดับความสามารถของนักเรียน (สัญลักษณ์ X) และแสดงระดับความยากง่ายของการได้เกรดในแต่ละรายวิชา (ตัวเลขแทนชื่อวิชาตามตารางที่ 4) ซึ่งจากภาพจะพบว่านักเรียนส่วนใหญ่มีระดับความสามารถปานกลางค่อนไปทางสูง และรายวิชาที่มีความยากในการได้เกรดสูง เช่น รายวิชาที่ 17, 7, 11 เป็นต้น และรายวิชาที่ง่ายต่อการได้เกรดสูง เช่น 6 และ 14 เป็นต้น เพื่อให้สามารถพิจารณาถึงทิศทางการตัดเกรดจะต้องพิจารณาจาก Wright Map ที่กำหนดค่าขั้นของความยาก (Threshold) ดังภาพที่ 11

3.2.3 การพิจารณาทิศทางการตัดเกรด (เหมาะสม/ปล่อยเกรด/กดเกรด) โดยการนำค่าขั้นของความยาก (Threshold) ในแต่ละขั้นของการตอบเดียวกันรวมกันแล้วนำมาหาค่าเฉลี่ยเพื่อกำหนดจุดตัดของแต่ละเกรดที่เหมาะสม

ระดับขั้นของการตอบที่แท้จริง (Step)	รายวิชาที่อยู่ในระดับขั้นของการตอบที่แท้จริงเดียวกัน	ค่า Threshold (Estimate) ในแต่ละข้อที่อยู่ใน step เดียวกัน	จำนวนวิชา	ค่าเฉลี่ยของ Threshold-จุดตัด
1 (ขั้นเกรด 1 ไป 1.5)	รายวิชาที่ 1 – 17	-0.497 -1.574 0.665 -0.736 5.306	17	0.705

ระดับขั้นของการตอบที่แท้จริง (Step)	รายวิชาที่อยู่ในระดับขั้นของการตอบที่แท้จริงเดียวกัน	ค่า Threshold (Estimate) ในแต่ละข้อที่อยู่ใน step เดียวกัน	จำนวนวิชา	ค่าเฉลี่ยของ Threshold-จุดตัด
		11.566 0.501 -0.79 -2.427 -2.847 -0.576 -0.808 -0.235 6.447 0.482 -0.501 -1.999		
2 (ขั้นเกรด 1.5 ไป 2)	รายวิชาที่ 1 – 17	-1.481 -1.353 -1.219 -2.17 -3.684 -0.392 -1.459 -0.959 -1.294 -0.842 -1.397 -1.011 -0.373 -1.259 -0.235 -1.207 -0.999	17	-1.255
3 (ขั้นเกรด 2 ไป 2.5)	รายวิชาที่ 1 – 17	-0.379 -0.372 -0.517 -0.81 -0.574	17	-0.834

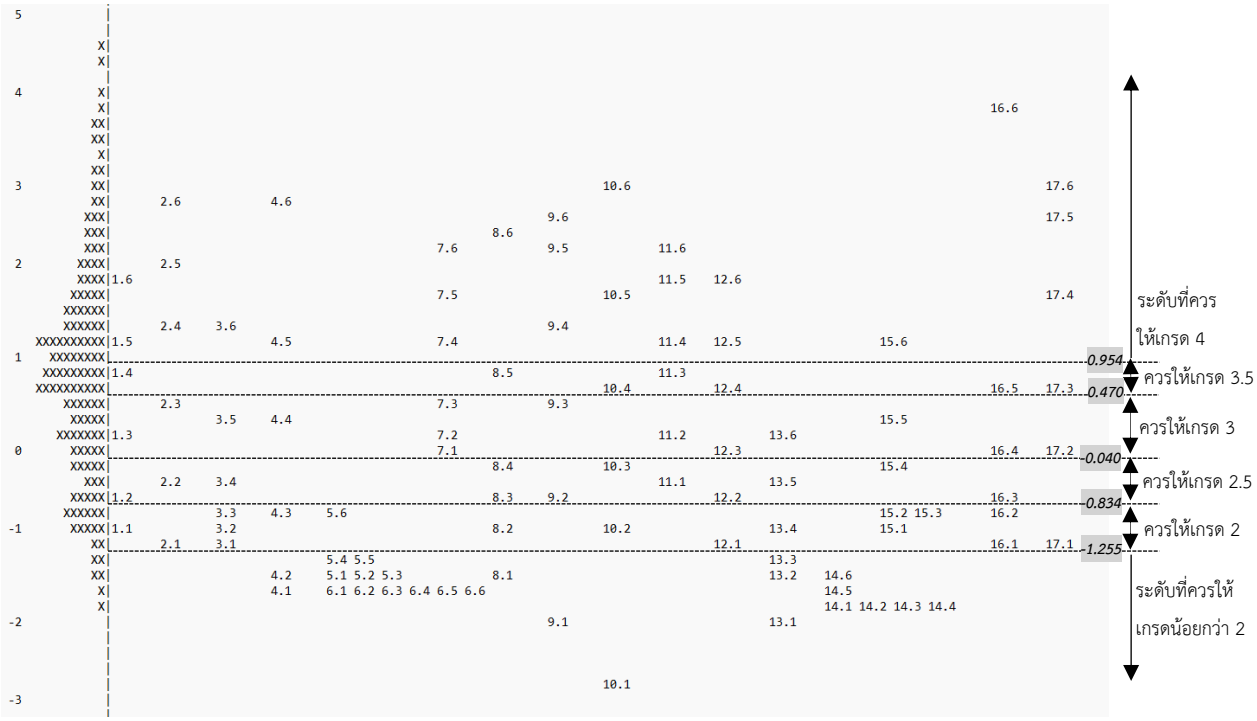
ระดับขั้นของการตอบที่แท้จริง (Step)	รายวิชาที่อยู่ในระดับขั้นของการตอบที่แท้จริงเดียวกัน	ค่า Threshold (Estimate) ในแต่ละข้อที่อยู่ใน step เดียวกัน	จำนวนวิชา	ค่าเฉลี่ยของ Threshold-จุดตัด
		-0.664 -0.954 -0.781 -0.006 -0.563 0.174 -0.687 -0.435 -3.961 -1.884 -1.128 -0.63		
4 (ขั้นเกรด 2.5 ไป 3)	รายวิชาที่ 1 – 17	0.608 0.416 -0.671 0.119 -0.063 -1.462 0.446 -0.584 0.518 0.345 -0.11 0.459 -0.335 -0.731 0.236 -0.511 0.647	17	-0.040
5 (ขั้นเกรด 3 ไป 3.5)	รายวิชาที่ 1 – 17	0.681 1.565 0.458 1.036 -1.19	17	0.470

ระดับขั้นของการตอบที่แท้จริง (Step)	รายวิชาที่อยู่ในระดับขั้นของการตอบที่แท้จริงเดียวกัน	ค่า Threshold (Estimate) ในแต่ละข้อที่อยู่ใน step เดียวกัน	จำนวนวิชา	ค่าเฉลี่ยของ Threshold-จุดตัด
		-5.826 0.626 0.899 2.115 1.464 1.487 0.967 0.886 0.508 0.336 -0.145 2.119		
6 (ขั้นเกรด 3.5 ไป 4)	รายวิชาที่ 1 – 17	1.068 1.317 1.284 2.56 0.206 -3.222 0.84 2.215 1.094 2.442 0.422 1.08 0.492 -1.005 1.065 3.492 0.862	17	0.954

ตารางที่ 6 ค่า Threshold ในแต่ละระดับขั้นของการตอบของแต่ละข้อใน Wright Map

จากตารางที่ 6 แสดงค่า Threshold (Estimate) ในแต่ละข้อที่อยู่ใน step เดียวกัน เพื่อคำนวณหาจุดตัดระหว่างเกรด โดยการหาค่าเฉลี่ยของ Threshold เช่น ระดับความสามารถของนักเรียนที่

ควรได้เกรด 3.5 และเกรด 4 จะมีจุดแบ่งระดับความสามารถอยู่ที่ 0.954 หมายถึง นักเรียนที่ควรได้เกรด 4 ต้องมีระดับความสามารถตั้งแต่ 0.954 ขึ้นไป จากตารางจะเห็นว่าเส้นจุดแบ่งระดับความสามารถของนักเรียนที่ควรได้เกรด 1 กับ เกรด 1.5 มีค่าเท่ากับ 0.705 และเส้นจุดแบ่งระดับความสามารถของนักเรียนที่ควรได้เกรด 1.5 กับ เกรด 2 มีค่าเท่ากับ -1.255 มีค่าน้อยกว่าซึ่งเป็นไปไม่ได้ แสดงว่าในระดับความสามารถ 1 1.5 และ 2 ไม่สามารถที่แยกความสามารถของนักเรียนหรือกำหนดจุดตัดแบ่งระดับนักเรียนที่ควรได้เกรด 1 เกรด 1.5 หรือ เกรด 2 ได้ ดังนั้นในการวิจัยครั้งนี้จึงยุบระดับขั้นของการตอบที่แท้จริง (Step) ที่ 1 คือ เกรด 1 ไปเกรด 1.5 ลงไว้ที่ผลการเรียนที่ต่ำกว่า 1.5 ไปเกรด 2 แทนเพื่อความเหมาะสมตามผลการวิเคราะห์ทางสถิติที่ได้ เมื่อนำค่าเฉลี่ยของ Threshold ที่ได้จากระดับขั้นของการตอบที่แท้จริง (Step) ของแต่ละขั้นมาเปรียบเทียบกับ Wright Map ดังแสดงในภาพที่ 11



ภาพที่ 11 Wright Map ระดับ Thresholds ของข้อสอบแต่ละรายวิชา
เพื่อตรวจสอบความตรงเชิงโครงสร้างตามโมเดล PCM

จากภาพที่ 11 แสดงระดับจุดตัดแบ่งระดับเกรดที่ได้จากการหาค่าเฉลี่ยของ Threshold ในแผนภาพ Wright Map เพื่อเปรียบเทียบการให้เกรดในแต่ละรายวิชาทั้งหมด 17 รายวิชา พบว่า

ระดับจุดตัดระดับความสามารถนักเรียนที่ควรได้เกรดน้อยกว่า 2 คือ น้อยกว่า -1.255 รายวิชาที่ให้เกรดน้อยกว่า 2 และอยู่ในช่วงที่คำนวณได้ เช่น วิชาที่ 4, 5, 6, 8, 9, 10 ตามลำดับ

ระดับจุดตัดระดับความสามารถนักเรียนที่ควรได้เกรด 2 คือ -1.255 ถึง -0.834 รายวิชาที่ให้เกรด 2 และอยู่ในช่วงที่คำนวณได้ เช่น วิชาที่ 3, 8, 10, 15, 16 ตามลำดับ

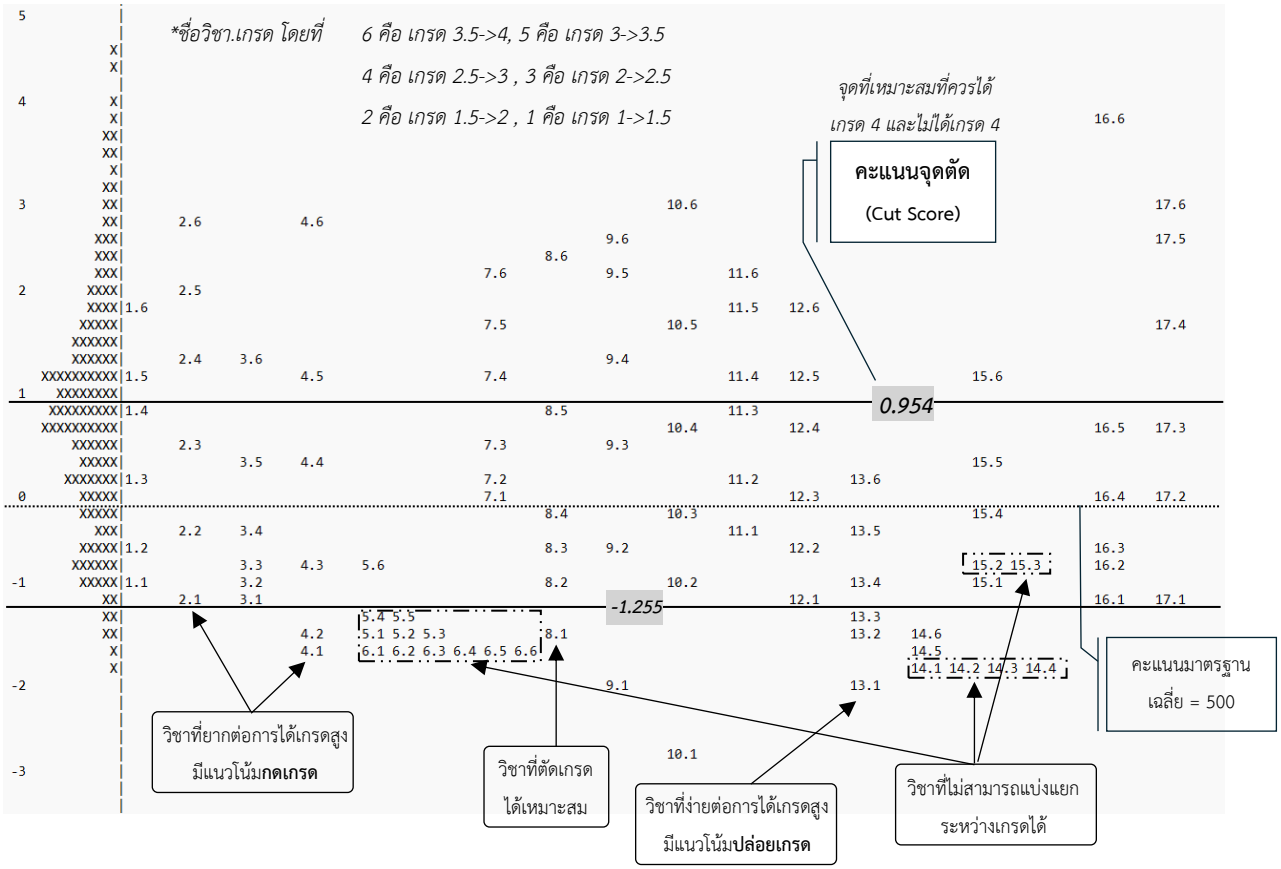
ระดับจุดตัดระดับความสามารถนักเรียนที่ควรได้เกรด 2.5 คือ -0.835 ถึง -0.040 รายวิชาที่ให้เกรด 2.5 และอยู่ในช่วงที่คำนวณได้ เช่น วิชาที่ 8, 10, 16 ตามลำดับ

ระดับจุดตัดระดับความสามารถนักเรียนที่ควรได้เกรด 3 คือ -0.041ถึง 0.470 รายวิชาที่ให้เกรด 3 และอยู่ในช่วงที่คำนวณได้ เช่น วิชาที่ 4, 16 ตามลำดับ

ระดับจุดตัดระดับความสามารถนักเรียนที่ควรได้เกรด 3.5 คือ 0.471 ถึง 0.954 รายวิชาที่ให้เกรด 3.5 และอยู่ในช่วงที่คำนวณได้ เช่น วิชาที่ 8, 16 ตามลำดับ

ระดับจุดตัดระดับความสามารถนักเรียนที่ควรได้เกรด 4 คือ 0.955 ขึ้นไป รายวิชาที่ให้เกรด 4 และอยู่ในช่วงที่คำนวณได้ เช่น วิชาที่ 1, 2, 3, 4, 7, 8, 9, 10, 11, 12, 15, 16, 17 ตามลำดับ

ถึงแม้การตัดเกรดของแต่ละรายวิชาจะตัดเกรดสอดคล้องกับระดับจุดตัดระดับความสามารถนักเรียนและอยู่ในช่วงที่คำนวณได้ก็ยังไม่สามารถระบุได้ถึงคุณภาพของการตัดเกรดได้ทั้งหมด เนื่องจากมีบางรายวิชาที่ไม่สามารถแบ่งแยกระดับผลการเรียนได้ หมายความว่า นักเรียนในทุกระดับความสามารถสามารถมีสิทธิ์ได้ทั้งเกรดสูงและเกรดต่ำ โดยไม่ได้สอดคล้องกับความสามารถที่แท้จริงของนักเรียน เช่น วิชาที่ 5 วิชาที่ 6 วิชาที่ 14 และ วิชาที่ 15 ดังแสดงในแผนภาพที่ 12



ภาพที่ 12 Wright Map คุณภาพของการตัดเกรดและความสามารถของนักเรียน

จากภาพที่ 12 จะแสดงลักษณะของการตัดเกรด ผู้วิจัยได้แสดงบน Wright Map เพื่อสะท้อนให้เห็นแนวทางในการพิจารณาคุณภาพของการตัดเกรดบน Wright Map โดยการพิจารณาคุณภาพของการตัดเกรดบน Wright Map โดยการพิจารณาเบื้องต้นคือการกำหนดจุดตัดที่เหมาะสมระหว่างสองเกรดที่ติดกัน เช่น จุดตัดแบ่งระหว่าง 4 กับ 3.5 3.5 กับ 3 3 กับ 2.5 เป็นต้น สำหรับการวิจัยครั้งนี้ผู้วิจัยใช้วิธีการกำหนดจุดตัดจากการนำค่าขั้นของความยาก (Threshold) ในแต่ละขั้นของการตอบเดียวกันรวมกันแล้วหาค่าเฉลี่ย สำหรับการวิจัยครั้งนี้ผู้วิจัยนำค่าขั้นความยากจากระดับที่ได้เกรด 3.5 ไปยังระดับที่ได้เกรด 4 (Step) ของทั้ง 17 รายวิชา มีค่าเท่ากับ 16.218 แล้วหาค่าเฉลี่ย (16.218/17) เพื่อกำหนดจุดตัดในการแบ่งระดับความสามารถของระดับที่ 1 และ 2 ซึ่งมีค่าเท่ากับ 0.954(θ) เพื่อให้เข้าใจง่ายขึ้นสามารถแปลงเป็นคะแนนสเกลที่มีค่าเฉลี่ยเป็น 500 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 100 โดยใช้สูตร 500 + 100(θ) = 595.4 เพื่อให้ง่ายต่อความเข้าใจและการนำไปใช้งานจึงกำหนดคะแนนจุดตัดเท่ากับ 595.4

ผลที่ได้จากการคำนวณดังกล่าวนำมาสู่การตีความได้ว่าผู้เรียนที่มีความสามารถตั้งแต่ 0.954 หรือได้คะแนนมาตรฐานตั้งแต่ 594.5 คะแนนมีโอกาสร้อยละ 50 ที่จะได้เกรด 4 ยิ่งผู้เรียนมีความสามารถสูงยิ่งมีโอกาสได้เกรด 4 มากขึ้น ในกรณีนี้การที่มีคะแนนจุดตัดมาตรฐานเท่ากับ 594.5 แสดงให้เห็นถึงจุดแบ่งที่เหมาะสมของผู้เรียนที่สมควรได้เกรด 4 และไม่ได้เกรด 4 ถ้าตัดเกรดแต่ละรายวิชาอยู่ในช่วงที่ตรงกับความสามารถผู้เรียนแสดงว่ารายวิชานั้นตัดเกรดได้เหมาะสม ตำแหน่งที่เหมาะสมในการตัดเกรด 4 ควรเหมาะกับผู้สอบที่มีความสามารถตั้งแต่ 0.954 หรือคะแนนจุดตัดมาตรฐานเท่ากับ 594.5 ขึ้นไป เช่น นักเรียน A มีความสามารถที่แท้จริงเท่ากับ 0 แสดงว่ามีคะแนนสเกลมาตรฐานเท่ากับ 500 ซึ่งอยู่ต่ำกว่าคะแนนจุดตัด ซึ่งควรมีโอกาสน้อยมากที่จะได้เกรด 4 แต่ผลการประเมินบางรายวิชามีโอกาสสูงมากที่จะได้เกรด 4 เช่น วิชาที่ 13 แสดงว่าวิชาดังกล่าวมีแนวโน้มในการประเมินความสามารถของนักเรียน A สูงกว่าความเป็นจริง เป็นต้น

เมื่อพิจารณาผลการวิเคราะห์ข้อมูลจากการศึกษานำร่องสะท้อนให้เห็นคุณภาพของการตัดเกรด 4 ลักษณะ ดังต่อไปนี้

(1) วิชาที่ตัดเกรดได้เหมาะสม เป็นวิชาที่สะท้อนให้เห็นถึงระดับความสามารถของผู้เรียนได้อย่างชัดเจนสามารถจำแนกการตัดเกรดได้ตรงกับระดับความสามารถของผู้เรียนได้โดยเฉพาะตำแหน่งที่สามารถจำแนกความแตกต่างระหว่างผู้เรียนที่ควรได้เกรด 4 และ ไม่ได้เกรด 4 เช่น วิชาที่ 8, 16 เป็นต้น

(2) วิชาที่ยากต่อการได้เกรดสูงมีแนวโน้มกดเกรด จาก Wright Map สะท้อนให้เห็นว่ามีบางรายวิชาที่ผู้เรียนต้องมีความสามารถสูงมากจึงจะได้เกรด 4 เช่น วิชาที่ 2 ผู้เรียนต้องมีความสามารถที่แท้จริง 2.8 ซึ่งเทียบเท่าคะแนนสเกล 780 คะแนนถึงจะมีโอกาสได้เกรด 4 นอกจากนี้รายวิชาดังกล่าวยังพบว่าผู้เรียนที่มีระดับความสามารถ 1.3 ซึ่งเทียบเท่าคะแนนสเกล 630 คะแนน แต่ได้เกรด 3 ซึ่งเป็นความสามารถที่ที่สูงมากเมื่อเทียบกับรายวิชาอื่นที่ตัดเกรด 3 แสดงว่ารายวิชาดังกล่าวมีแนวโน้มตัดเกรดต่ำ

กว่าความเป็นจริง จึงมีแนวโน้มในการตัดกดเกรด

(3) วิชาที่ง่ายต่อการได้เกรดสูงมีแนวโน้มปล่อยเกรด จาก Wright Map สะท้อนให้เห็นว่ามีบางรายวิชาที่ผู้เรียนต้องมีความสามารถปานกลางค่อยมาทางอ่อนมีโอกาสได้เกรด 4 สูง เช่น วิชาที่ 13 ผู้เรียนมีความสามารถที่แท้จริง 0.3 ซึ่ง เทียบเท่ากับคะแนนสเกลมาตรฐานต่ำกว่าค่าเฉลี่ย มีโอกาสได้เกรด 4 นอกจากนี้ พบว่าผู้เรียนมีความสามารถที่แท้จริง -2.10 มีโอกาสได้เกรด 1.5 ซึ่งไม่มีผู้เรียนที่อยู่ในระดับความสามารถดังกล่าวและผู้เรียนทั้งหมดมีโอกาสได้เกรด 4 เกือบร้อยละ 100 เป็นต้น

(4) วิชาที่ไม่สามารถแบ่งแยกความแตกต่างระหว่างเกรด จากผลการวิเคราะห์พบว่าบางรายวิชาตัดเกรดได้เหมาะสมระหว่างผู้เรียนที่ควรได้เกรด 4 และไม่ได้เกรด 4 แต่การตัดเกรดบางรายวิชาไม่สามารถแบ่งแยกความแตกต่างระหว่างเกรดได้ เนื่องจากอยู่ในตำแหน่งเดียวกัน เช่น รายวิชาที่ 5 รายวิชาที่ 6 รายวิชาที่ 14 และรายวิชาที่ 15 ไม่สามารถจำแนกความสามารถผู้เรียนในระดับเกรดที่ต่างกันได้

3.2.4 การวิเคราะห์คุณภาพผู้เรียน การวิจัยในครั้งนี้ได้นำระดับความสามารถของผู้เรียนที่ได้จากการวิเคราะห์ด้วยโปรแกรม ConQuest 2.0 มาแปลงเป็นคะแนนสเกลที่มีค่าเฉลี่ยเป็น 500 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 100 โดยใช้สูตร 500 + 100(θ) ดังตารางที่ 7

	ผลการเรียนรายวิชา		**ผลการเรียนเฉลี่ย 6 ภาคเรียน**	
คนที่	**ระดับความสามารถ (θ)**	**คะแนนสเกล**	**GPAX**	**คะแนนสเกล**
1	0.99495	599.495	2.56	756
2	2.72572	772.572	3.41	841
3	0.55735	555.735	2.19	719
4	2.08262	708.262	3.19	819
5	2.72572	772.572	3.39	839
6	1.24786	624.786	2.92	792
7	3.91282	891.282	3.80	880
8	3.91282	891.282	3.78	878
9	3.02615	802.615	3.59	859
10	1.37564	637.564	3.03	803
11	0.24859	524.859	2.58	758
12	0.61434	561.434	2.94	794
13	1.82584	682.584	3.36	836
14	2.36488	736.488	3.25	825
15	2.59506	759.506	3.38	838

	ผลการเรียนรายวิชา		ผลการเรียนเฉลี่ย 6 ภาคเรียน	
คนที่	ระดับความสามารถ (θ)	คะแนนสเกล	GPAX	คะแนนสเกล
16	2.36488	736.488	3.29	829
17	4.28805	928.805	3.99	899
18	-0.51844	448.156	2.19	719
19	1.24786	624.786	2.36	736
20	1.31091	631.091	3.19	819
21	-0.60809	439.191	2.23	723
22	0.86351	586.351	3.06	806
23	2.36488	736.488	3.49	849
24	2.17136	717.136	3.17	817
25	-0.12408	487.592	1.88	688
26	1.99598	699.598	3.21	821
27	1.91037	691.037	2.96	796
28	1.18557	618.557	3.07	807
29	0.44899	544.899	2.50	750
30	0.55735	555.735	3.20	820
31	1.82584	682.584	3.28	828
32	0.73492	573.492	2.93	793
33	1.91037	691.037	3.51	851
34	1.66312	666.312	3.21	821
35	1.91037	691.037	3.21	821
36	2.36488	736.488	3.60	860
37	1.51292	651.292	3.20	820
38	1.99598	699.598	3.26	826
39	1.31091	631.091	3.19	819
40	1.66312	666.312	3.05	805
41	0.99495	599.495	3.11	811
42	1.82584	682.584	3.44	844
43	0.34661	534.661	2.93	793
44	1.31091	631.091	3.52	852

	ผลการเรียนรายวิชา		ผลการเรียนเฉลี่ย 6 ภาคเรียน	
คนที่	ระดับความสามารถ (θ)	คะแนนสเกล	GPAX	คะแนนสเกล
45	1.24786	624.786	3.25	825
46	2.17136	717.136	3.58	858
47	0.06055	506.055	2.11	711
48	0.50231	550.231	2.04	704
49	3.40288	840.288	3.91	891
50	3.40288	840.288	3.94	894
51	0.99495	599.495	2.33	733
52	2.59506	759.506	3.58	858
53	3.02615	802.615	3.76	876
54	4.28805	928.805	3.74	874
55	3.40288	840.288	3.64	864
56	3.91282	891.282	3.67	867
57	1.74314	674.314	3.19	819
58	3.62933	862.933	3.18	818
59	1.58632	658.632	2.94	794
60	1.24786	624.786	2.79	779
61	-1.27793	372.207	1.37	637
62	1.12307	612.307	2.95	795
63	0.29714	529.714	2.61	761
64	1.37564	637.564	3.23	823
65	0.99495	599.495	2.35	735
66	-0.51844	448.156	1.86	686
67	-0.4204	457.960	1.88	688
68	0.15359	515.359	1.83	683
69	-0.03149	496.851	1.83	683
70	2.08262	708.262	3.26	826
71	1.31091	631.091	2.23	723
72	1.51292	651.292	3.27	827
73	1.12307	612.307	2.77	777

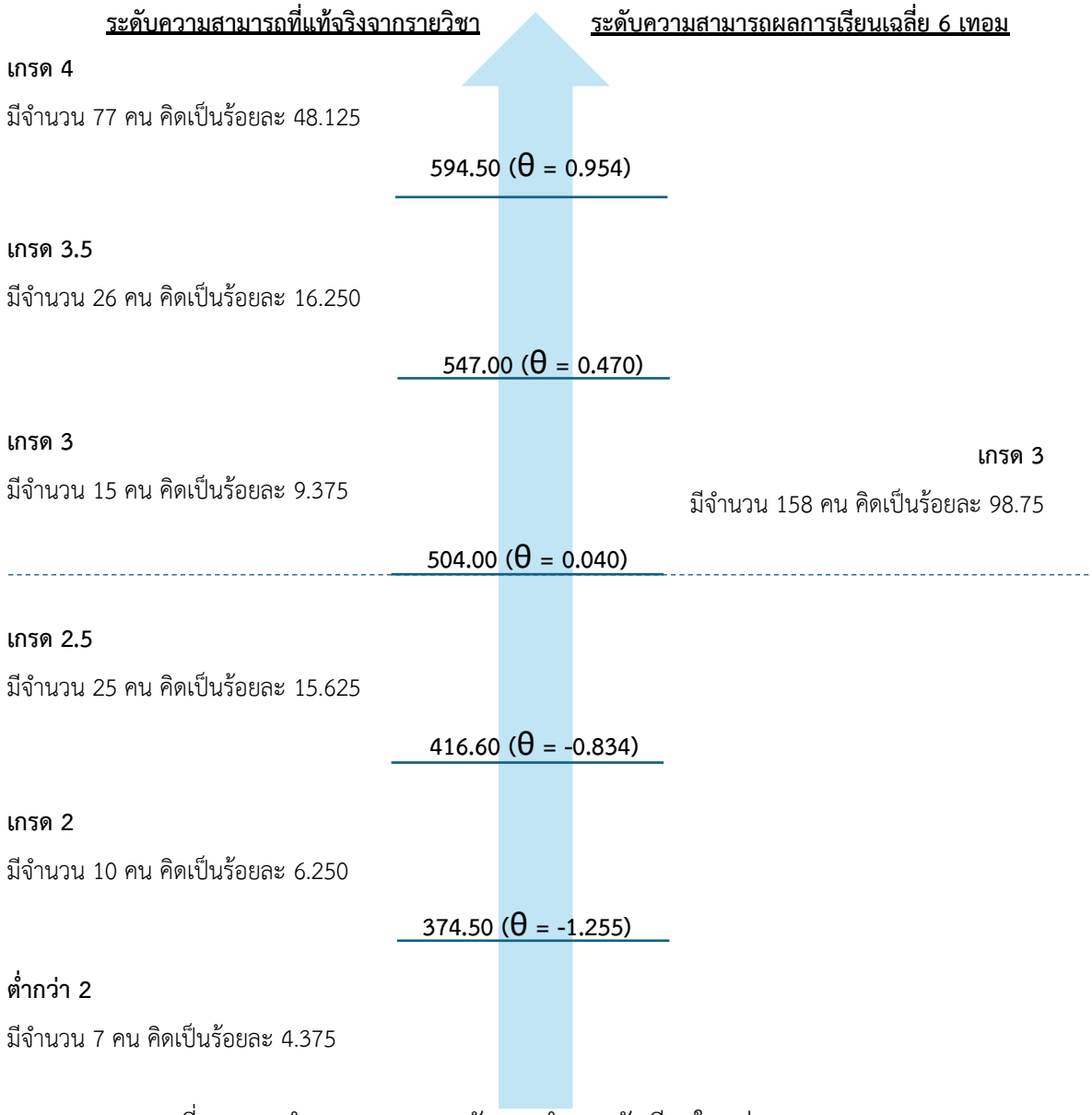
	ผลการเรียนรายวิชา		ผลการเรียนเฉลี่ย 6 ภาคเรียน	
คนที่	ระดับความสามารถ (θ)	คะแนนสเกล	GPAX	คะแนนสเกล
74	2.08262	708.262	3.23	823
75	1.74314	674.314	2.80	780
76	0.86351	586.351	2.29	729
77	-0.03149	496.851	1.78	678
78	0.86351	586.351	2.8	780
79	-0.64931	435.069	1.67	667
80	0.01448	501.448	2.52	752
81	-0.87066	412.934	1.74	674
82	-0.51844	448.156	2.36	736
83	-0.12408	487.592	2.34	734
84	-1.38875	361.125	1.65	665
85	-0.87066	412.934	2.63	763
86	-0.47023	452.977	2.06	706
87	-0.87066	412.934	2.04	704
88	0.86351	586.351	2.87	787
89	0.79843	579.843	3.10	810
90	0.44899	544.899	2.78	778
91	1.37564	637.564	3.6	860
92	1.18557	618.557	3.26	826
93	0.73492	573.492	2.43	743
94	1.82584	682.584	3.34	834
95	2.8681	786.810	3.49	849
96	-0.68845	431.155	2.12	712
97	-1.02642	397.358	1.94	694
98	-1.38875	361.125	1.53	653
99	3.20397	820.397	3.81	881
100	0.34661	534.661	2.68	768
101	0.61434	561.434	2.69	769
102	1.24786	624.786	2.87	787

	ผลการเรียนรายวิชา		ผลการเรียนเฉลี่ย 6 ภาคเรียน	
คนที่	ระดับความสามารถ (θ)	คะแนนสเกล	GPAX	คะแนนสเกล
103	0.73492	573.492	2.76	776
104	0.39716	539.716	3.23	823
105	1.12307	612.307	3.31	831
106	0.86351	586.351	3.13	813
107	3.40288	840.288	3.88	888
108	1.31091	631.091	3.13	813
109	0.6735	567.35	3.00	800
110	0.50231	550.231	3.08	808
111	0.6735	567.350	3.46	846
112	2.8681	786.810	3.65	865
113	1.12307	612.307	3.25	825
114	2.08262	708.262	3.68	868
115	0.79843	579.843	2.56	756
116	-0.87066	412.934	2.64	764
117	0.79843	579.843	2.77	777
118	0.99495	599.495	3.35	835
119	0.29714	529.714	3.14	814
120	1.4428	644.280	3.40	840
121	-0.21944	478.056	2.41	741
122	-1.15331	384.669	1.63	663
123	0.55735	555.735	3.15	815
124	0.44899	544.899	2.27	727
125	0.6735	567.350	2.56	756
126	-1.60931	339.069	1.72	672
127	-1.42194	357.806	1.43	643
128	1.31091	631.091	3.59	859
129	2.17136	717.136	3.65	865
130	0.79843	579.843	2.70	770
131	1.18557	618.557	3.54	854

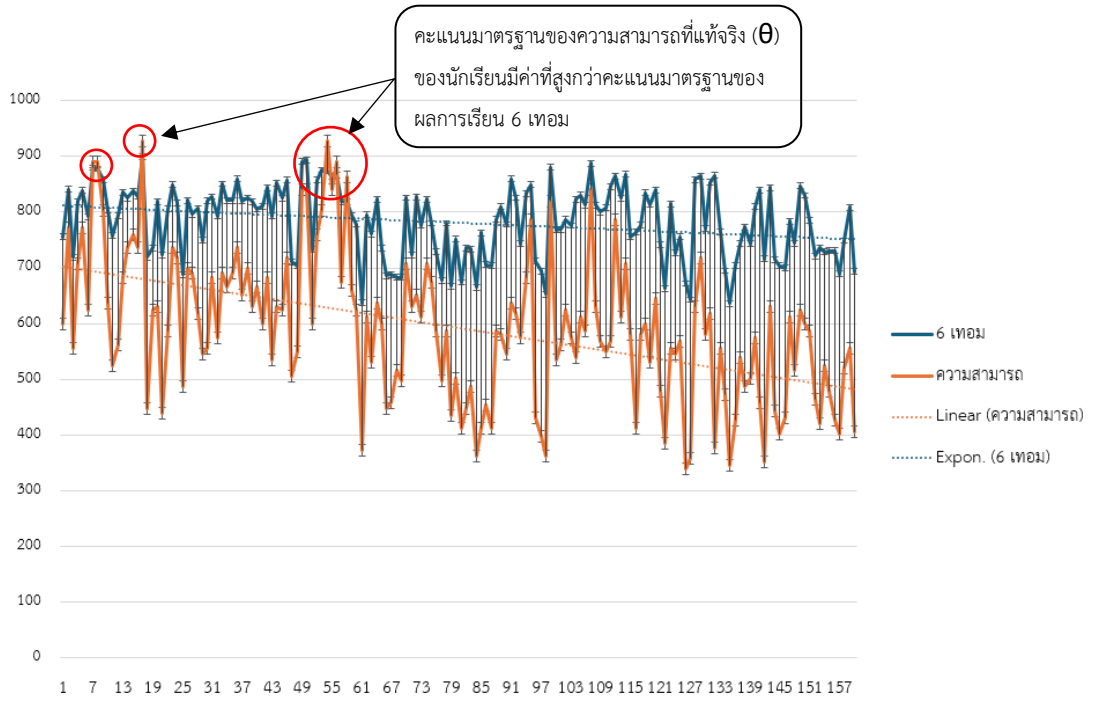
	ผลการเรียนรายวิชา		ผลการเรียนเฉลี่ย 6 ภาคเรียน	
คนที่	ระดับความสามารถ (θ)	คะแนนสเกล	GPAX	คะแนนสเกล
132	-1.23748	376.252	3.66	866
133	0.55735	555.735	2.65	765
134	-0.26869	473.131	1.95	695
135	-1.54446	345.554	1.36	636
136	-0.72597	427.403	2.05	705
137	0.39716	539.716	2.43	743
138	-0.12408	487.592	2.73	773
139	0.01448	501.448	2.43	743
140	0.73492	573.492	3.10	810
141	-0.4204	457.960	3.40	840
142	-1.48412	351.588	2.16	716
143	1.31091	631.091	3.44	844
144	-0.56448	443.552	2.15	715
145	-0.98571	401.429	2.04	704
146	-0.68845	431.155	2.00	700
147	1.12307	612.307	2.84	784
148	0.15359	515.359	2.44	744
149	1.24786	624.786	3.47	847
150	0.99495	599.495	3.30	830
151	0.86351	586.351	2.86	786
152	-0.36969	463.031	2.21	721
153	-0.79833	420.167	2.36	736
154	0.24859	524.859	2.29	729
155	-0.21944	478.056	2.31	731
156	-0.72597	427.403	2.31	731
157	-0.98571	401.429	1.89	689
158	0.20079	520.079	2.49	749
159	0.55735	555.735	3.09	809
160	-0.9462	405.380	1.93	693

ตารางที่ 7 การวิเคราะห์คุณภาพผู้เรียน

เพื่อสะท้อนให้เห็นถึงลักษณะของการกำหนดคะแนนจุดตัดที่เหมาะสมและจำนวนของนักเรียนที่อยู่ในระดับความสามารถดังกล่าว ดังภาพที่ 13 ที่พบว่าเมื่อทำการพัฒนาคะแนนจุดตัดที่เหมาะสมพบว่าสามารถแบ่งระดับความสามารถของนักเรียนได้เป็น 5 ระดับ ตั้งแต่ระดับเกรด 2 ถึงระดับเกรด 4 ซึ่งจะเห็นว่าโอกาสที่ผู้เรียนจะมีความสามารถในระดับเกรด 4 มากที่สุด รองลงมาคือระดับเกรด 3.5 และ เกรด 3



ภาพที่ 13 การกำหนดคะแนนจุดตัดและจำนวนนักเรียนในแต่ละความสามารถ



ภาพที่ 14 เปรียบเทียบคะแนนมาตรฐานเกรด 6 เทอมกับความสามารถที่แท้จริง (θ)

จากภาพจะแสดงให้แนวโน้มของคะแนนมาตรฐานของผลการเรียน 6 เทอม กับคะแนนมาตรฐานของความสามารถที่แท้จริง (θ) จะมีแนวโน้มไปในทิศทางเดียวกันเป็นส่วนใหญ่ หมายความว่า เมื่อคะแนนมาตรฐานของผลการเรียน 6 เทอมสูงขึ้นหรือลดลง จะส่งผลให้คะแนนมาตรฐานของความสามารถที่แท้จริง (θ) สูงขึ้นหรือลดลงด้วยเช่นกัน ตามลำดับ มีความสอดคล้องกันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 (X^2 = 69.104, df=1, Sig=.000) แต่มีบางคนที่คะแนนมาตรฐานของความสามารถที่แท้จริง (θ) ของนักเรียนมีค่าที่สูงกว่าคะแนนมาตรฐานของผลการเรียน 6 เทอม ทั้งนี้อาจเป็นผลมาจากการตัดเกรดของครูในแต่ละรายวิชาหรือคุณลักษณะภายในของตัวนักเรียน

4. ตรวจสอบหลักฐานการประเมินด้านความตรงตามเกณฑ์มาตรฐานการประเมิน

4.1 หลักฐานความตรงด้านโครงสร้างภายใน ผู้วิจัยได้ทำการตรวจสอบความตรงโดยใช้โมเดลการตอบสนองข้อสอบแบบพหุมิติ (MRCML) โดยมีการนำเสนอในรูปของ Wright Map พร้อมทั้งดัชนีที่สะท้อนถึงความตรงเชิงโครงสร้างภายใน โดยใช้โปรแกรม ACER Conquest 2.0 (Wu, Adams, Wilson, & Haldane, 2007) หลักฐานความเที่ยงด้วยค่าสัมประสิทธิ์สัมพันธ์ (Reliability evidence based on reliability coefficient) ได้แก่

1) ความเที่ยงแบบ Expected A Posteriori หรือ EAP/PV (EAP/PV Reliability) (Briggs & Wilson, 2003) เป็นค่าสัดส่วนของความแปรปรวนที่อธิบายได้จากโมเดล Partial credit model ต่อความแปรปรวนในภาพรวม (total variance) ของค่าความสามารถแฝงของนักเรียน

ผ่านเกณฑ์การยอมรับได้ที่ .70 (Cronbach, 1990) สามารถอธิบายความแปรปรวนของความสามารถผู้เรียนได้ผ่านเกณฑ์โดยมีค่าเท่ากับ .955 (EAP/PV RELIABILITY: 0.955)

2) ค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation coefficient) จากการหาความสัมพันธ์ มีค่าความเที่ยงเท่ากับ .95 (Coefficient Alpha: 0.95) ซึ่งมีความสัมพันธ์กัน

4.2 หลักฐานความตรงที่สัมพันธ์กับเกณฑ์อื่น

1) ตรวจสอบความตรงตามสภาพ โดยพิจารณาจากสัมประสิทธิ์สหสัมพันธ์กับตัวแปรภายนอก คือ ผลการเรียนเฉลี่ย 6 เทอม กับคะแนนความสามารถที่แท้จริงที่ประมาณค่าได้จากโปรแกรม จะพบว่าคุณภาพของการตัดเกรด 4 ลักษณะ (1) วิชาที่ตัดเกรดได้เหมาะสม (2) วิชาที่ยากต่อการได้เกรดสูงมีแนวโน้มกดเกรด (3) วิชาที่ง่ายต่อการได้เกรดสูงมีแนวโน้มปล่อยเกรด และ (4) วิชาที่ไม่สามารถแบ่งแยกความแตกต่างระหว่างเกรด ซึ่งสะท้อนออกมาให้เห็นดังภาพที่ 13 แสดงให้เห็นว่านักเรียนร้อยละ 98.75 มีระดับความสามารถ 0.040 ขึ้นไป (คะแนนมาตรฐาน 504) แต่ระดับความสามารถที่แท้จริงมีเพียงร้อยละ 73.75 ที่มีระดับความสามารถ 0.040 ขึ้นไป (คะแนนมาตรฐาน 504) แสดงให้เห็นว่าการตัดเกรดยังมีบางส่วนยังไม่ได้สะท้อนความสามารถที่แท้จริงของนักเรียนออกมาได้ตรงความสามารถที่แท้จริง

2) ตรวจสอบความตรงเชิงทำนาย โดยพิจารณาจากสัมประสิทธิ์สหสัมพันธ์กับตัวแปรภายนอกที่สะท้อนถึงความสามารถของนักเรียน คือ คะแนน GPA ทั้ง 17 รายวิชา กับผลการเรียนเฉลี่ย 6 ภาคเรียน มีความสอดคล้องกันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 (X^2 = 69.104, df=1, Sig=.000) คะแนนมาตรฐานของผลการเรียน 6 เทอม กับคะแนนมาตรฐานของความสามารถที่แท้จริง (θ) จะมีแนวโน้มไปในทิศทางเดียวกัน ดังแสดงในภาพที่ 14

บทที่ 4

ผลการวิจัย

การดำเนินการวิจัยครั้งนี้ เป็นการวิจัยเพื่อตรวจสอบคุณภาพของการตัดเกรดในแต่ละรายวิชาและตรวจสอบคุณภาพผู้เรียน รวมไปถึงเพื่อตรวจสอบความสอดคล้องของระดับความสามารถที่แท้จริงของนักเรียนระดับชั้นมัธยมศึกษาปีที่ 6 กับผลการเรียน 6 เทอม ผู้วิจัยได้ดำเนินการรวบรวมข้อมูลและวิเคราะห์ข้อมูลในการวิจัย จึงขอนำเสนอผลการวิจัยและอภิปรายผลการวิจัยตามลำดับดังนี้

1. ผลการตรวจสอบคุณภาพการตัดเกรด
2. ผลการตรวจสอบคุณภาพผู้เรียน
3. ผลการตรวจสอบหลักฐานการประเมินด้านความตรงตามเกณฑ์มาตรฐานการประเมิน

รายละเอียดเกี่ยวกับวิธีดำเนินการวิจัย มีดังนี้

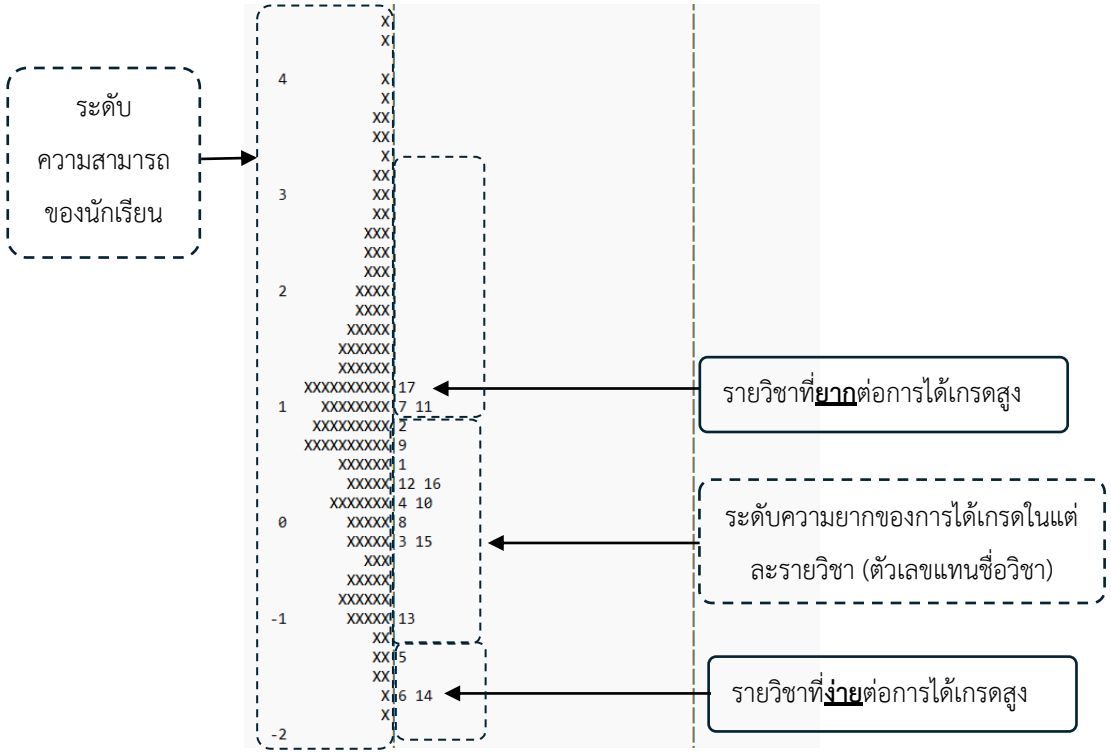
1. ผลการตรวจสอบคุณภาพการตัดเกรด

ผลการรวบรวมข้อมูลและการตรวจสอบสอบผลการเรียน มีรายละเอียดดังนี้

1.1 ผลการรวบรวมข้อมูล ผู้วิจัยเก็บข้อมูลที่เป็นตัวแทนความสามารถของผู้เรียนให้ครอบคลุมทุกแผนการเรียนเลือกใช้ กลุ่มรายวิชาพื้นฐาน ในภาคเรียนที่ 1 และ 2 ปีการศึกษา 2566 ของระดับชั้นมัธยมศึกษาปีที่ 6 ซึ่งจำนวนทั้งหมด 17 รหัสวิชา ได้แก่ ภาษาไทย คณิตศาสตร์ วิทยาศาสตร์ วิทยาการคำนวณ การออกแบบและเทคโนโลยี สังคมศึกษา ภาษาอังกฤษ สุขศึกษาและพลศึกษา ดนตรีและนาฏศิลป์ การงานอาชีพ

ผลการสุ่มโดยแบ่งประชากรออกเป็นแต่ละชั้น (ห้อง) ซึ่งนักเรียนชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2566 มีจำนวนทั้งหมด 8 ห้องเรียนและใช้วิธีการสุ่มตัวอย่างแบบง่าย (Simple random sampling) จากแต่ละชั้น (ห้อง) โดยเลือกมาห้องเรียนละ 20 คน จะได้กลุ่มตัวอย่างทั้งหมด 160 คน

1.2 ผลการตรวจสอบคุณภาพการตัดเกรด ผู้วิจัยได้วิเคราะห์ข้อมูลโดยใช้โมเดลโลจิสติกตามทฤษฎีการตอบสนองแบบพหุมิติ (Multidimensional Random Coefficients Multinomial Logit Model) ใช้ชื่อย่อว่า “MRCLM”) โดยใช้โปรแกรม ConQuest 2.0 (Wu, Adums, Wilson, & Haldane, 2007) ซึ่งจากการวิเคราะห์ผลการตรวจสอบคุณภาพการตัดเกรด จะพบว่า



ภาพที่ 15 Wright Map แสดงคุณภาพของการตัดเกรดและความสามารถของนักเรียน

1.2.1 นักเรียนมีระดับความสามารถของนักเรียนทั้งหมดมีความสามารถระดับปานกลางค่อนไปทางสูง จากภาพที่ 15 จะพบความยากง่ายในการตัดเกรดของแต่ละวิชา ดังนี้

- รายวิชาที่มีความยากในการได้เกรดสูง เช่น รายวิชาที่ 17, 7, 11
- รายวิชาที่ง่ายต่อการได้เกรดสูง เช่น 5, 6, 14
- รายวิชาที่มีโอกาสได้เกรดตามความสามารถ เช่น 2, 1, 12, 16, 4, 10, 8, 3, 15, 13

1.2.2 ผลการคำนวณหาจุดตัดระหว่างเกรด จะพบว่า

ระดับจุดตัดระดับความสามารถนักเรียนที่ควรได้เกรดน้อยกว่า 2 คือ น้อยกว่า -1.255 รายวิชาที่ให้เกรดน้อยกว่า 2 และอยู่ในช่วงที่คำนวณได้ เช่น วิชาที่ 4, 5, 6, 8, 9, 10 ตามลำดับ

ระดับจุดตัดระดับความสามารถนักเรียนที่ควรได้เกรด 2 คือ -1.255 ถึง -0.834 รายวิชาที่ให้เกรด 2 และอยู่ในช่วงที่คำนวณได้ เช่น วิชาที่ 3, 8, 10, 15, 16 ตามลำดับ

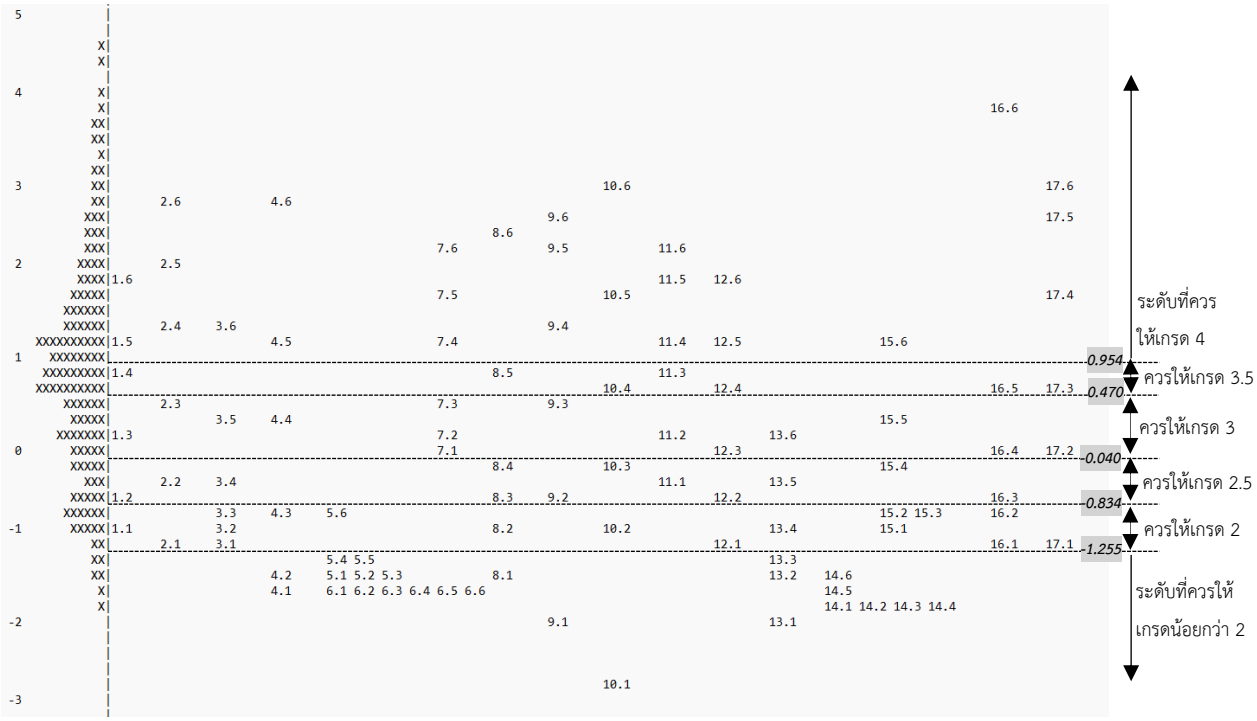
ระดับจุดตัดระดับความสามารถนักเรียนที่ควรได้เกรด 2.5 คือ -0.835 ถึง -0.040 รายวิชาที่ให้เกรด 2.5 และอยู่ในช่วงที่คำนวณได้ เช่น วิชาที่ 8, 10, 16 ตามลำดับ

ระดับจุดตัดระดับความสามารถนักเรียนที่ควรได้เกรด 3 คือ -0.041ถึง 0.470 รายวิชาที่ให้เกรด 3 และอยู่ในช่วงที่คำนวณได้ เช่น วิชาที่ 4, 16 ตามลำดับ

ระดับจุดตัดระดับความสามารถนักเรียนที่ควรได้เกรด 3.5 คือ 0.471 ถึง 0.954 รายวิชาที่ให้

เกรด 3.5 และอยู่ในช่วงที่คำนวณได้ เช่น วิชาที่ 8, 16 ตามลำดับ

ระดับจุดตัดระดับความสามารถนักเรียนที่ควรได้เกรด 4 คือ 0.955 ขึ้นไป รายวิชาที่ให้เกรด 4 และอยู่ในช่วงที่คำนวณได้ เช่น วิชาที่ 1, 2, 3, 4, 7, 8, 9, 10, 11, 12, 15, 16, 17 ตามลำดับดังแสดงในภาพที่ 16



ภาพที่ 16 Wright Map ระดับ Thresholds ของข้อสอบแต่ละรายวิชา
เพื่อตรวจสอบความตรงเชิงโครงสร้างตามโมเดล PCM

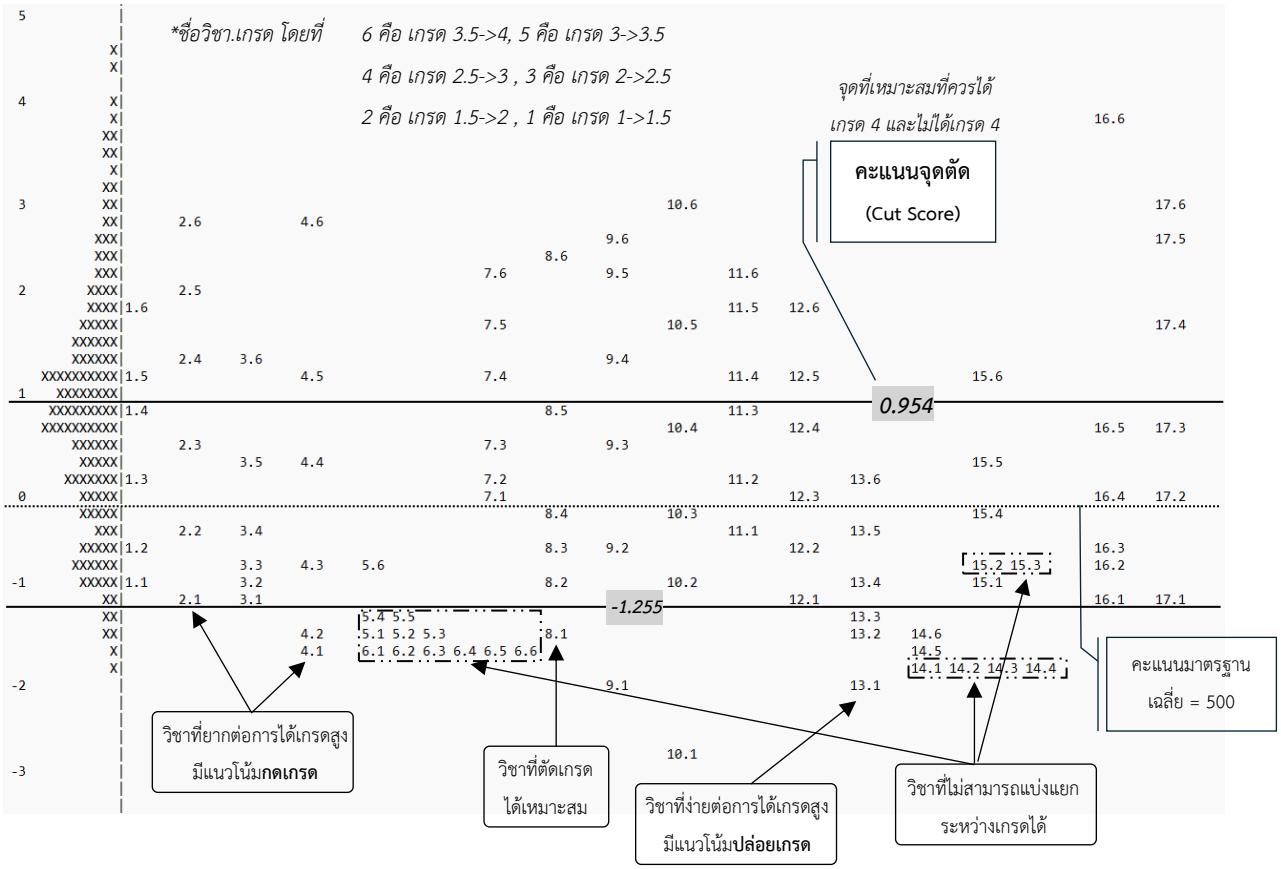
1.2.3 ผลการวิเคราะห์ทิศทางการตัดเกรด (เหมาะสม/ปล่อยเกรด/กดเกรด) โดยการหาค่าเฉลี่ยของ Threshold เมื่อพิจารณาผลการวิเคราะห์ข้อมูลจากการศึกษานำร่อง ตามภาพที่ 17 สะท้อนให้เห็นคุณภาพของการตัดเกรด 4 ลักษณะ ดังต่อไปนี้

(1) วิชาที่ตัดเกรดได้เหมาะสม เป็นวิชาที่สะท้อนให้เห็นถึงระดับความสามารถของผู้เรียนได้อย่างชัดเจนสามารถจำแนกการตัดเกรดได้ตรงกับระดับความสามารถของผู้เรียนได้โดยเฉพาะตำแหน่งที่สามารถจำแนกความแตกต่างระหว่างผู้เรียนที่ควรได้เกรด 4 และ ไม่ได้เกรด 4 เช่น วิชาที่ 1, 3, 8, 12, 16

(2) วิชาที่ยากต่อการได้เกรดสูงมีแนวโน้มกดเกรด จาก Wright Map สะท้อนให้เห็นว่ามีบางรายวิชาที่ผู้เรียนต้องมีความสามารถสูงมากจึงจะได้เกรด 4 ได้แก่ วิชาที่ 2, 4, 7, 9, 10, 11, 17 แสดงว่ารายวิชาดังกล่าวมีแนวโน้มตัดเกรดต่ำกว่าความเป็นจริง จึงมีแนวโน้มในการตัดกดเกรด

(3) วิชาที่ง่ายต่อการได้เกรดสูงมีแนวโน้มปล่อยเกรด จาก Wright Map สะท้อนให้เห็นว่ามีบางรายวิชาที่ผู้เรียนต้องมีความสามารถปานกลางค่อยมาทางอ่อนมีโอกาสได้เกรด 4 สูง เช่น วิชาที่ 13

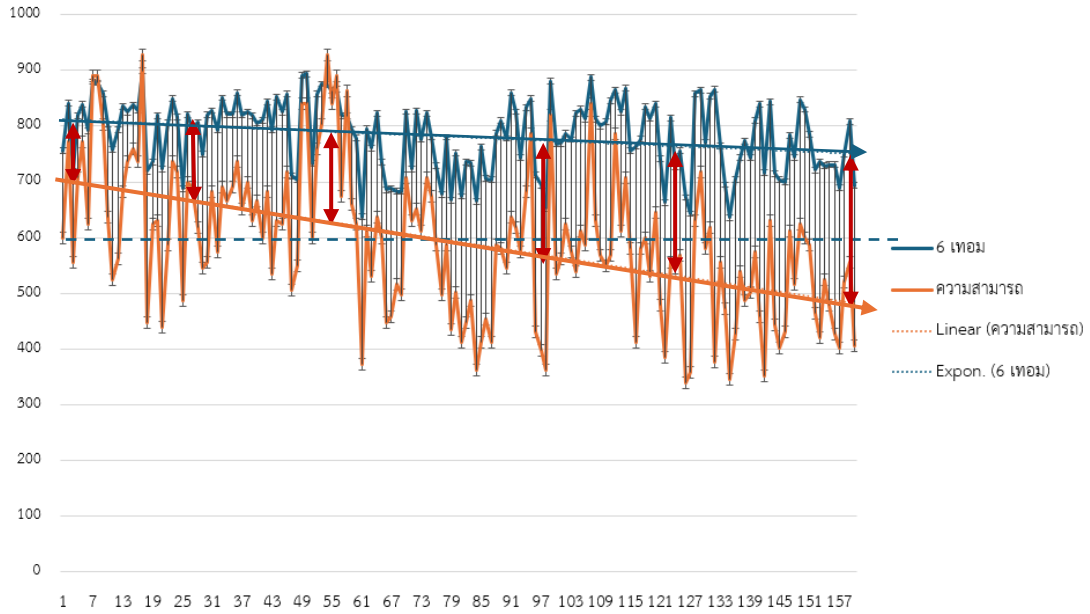
(4) วิชาที่ไม่สามารถแบ่งแยกความแตกต่างระหว่างเกรด จากผลการวิเคราะห์พบว่าบางรายวิชาตัดเกรดได้เหมาะสมระหว่างผู้เรียนที่ควรได้เกรด 4 และไม่ได้เกรด 4 แต่การตัดเกรดบางรายวิชาไม่สามารถแบ่งแยกความแตกต่างระหว่างเกรดได้ เนื่องจากอยู่ในตำแหน่งเดียวกัน เช่น รายวิชาที่ 5, 6, 14, 15 ไม่สามารถจำแนกความสามารถผู้เรียนในระดับเกรดที่ต่างกันได้



ภาพที่ 17 Wright Map คุณภาพของการตัดเกรดและความสามารถของนักเรียน

2. ผลการตรวจสอบคุณภาพผู้เรียน

ผลการตรวจสอบคุณภาพผู้เรียน เพื่อให้เข้าใจง่ายขึ้นผู้วิจัยได้แปลงเป็นคะแนนสเกลที่มีค่าเฉลี่ยเป็น 500 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 100 โดยใช้สูตร 500 + 100(θ) เทียบกับผลการเรียนเฉลี่ย 6 เทอม ดังภาพที่ 18



ภาพที่ 18 เปรียบเทียบคะแนนมาตรฐานเกรด 6 เทอมกับความสามารถที่แท้จริง (θ)

2.1 ผลการวิเคราะห์คะแนนมาตรฐานของผลการเรียน 6 เทอม จะพบว่านักเรียนส่วนใหญ่กว่าร้อยละ 98.75 มีความสามารถที่สูงมากมีระดับความสามารถ 0.040 ขึ้นไป (คะแนนมาตรฐาน 504) ซึ่งไม่สอดคล้องกับข้อมูลผลการวิเคราะห์ทิศทางการตัดเกรดซึ่งเป็นการวิเคราะห์ผลการเรียนเฉพาะรายวิชาพื้นฐาน แต่ผลการเรียนเฉลี่ยรวม 6 เทอม เป็นการวิเคราะห์รวมทั้งวิชาพื้นฐานและวิชาเฉพาะแผนการเรียน เป็นรายวิชาที่นักเรียนชอบหรือถนัด อาจเป็นผลทำให้คะแนนความสามารถ ส่งผลให้คะแนนมาตรฐานของผลการเรียน 6 เทอมมีคะแนนที่สูงมาก

2.2 ผลการวิเคราะห์คะแนนมาตรฐานของความสามารถที่แท้จริง (θ) จะพบว่านักเรียนร้อยละ 73.75 ที่มีระดับความสามารถ 0.040 ขึ้นไป (คะแนนมาตรฐาน 504) เป็นไปในทิศทางเดียวกันกับข้อมูลผลการวิเคราะห์ทิศทางการตัดเกรดที่มีรายวิชาที่กดเกรดค่อนข้างสูง 7 รายวิชา คิดเป็นร้อยละ 41.176 ของรายวิชาพื้นฐานที่ทำการวิเคราะห์ทั้งหมด อาจส่งผลให้คะแนนมาตรฐานของความสามารถที่แท้จริง (θ) ของผู้เรียนอาจจะต่ำกว่าความเป็นจริง หรือความสามารถที่แท้จริงได้

2.3 ผลการวิเคราะห์เชิงทำนาย จากภาพที่ 18 คะแนนมาตรฐานของผลการเรียน 6 เทอม และคะแนนมาตรฐานของความสามารถที่แท้จริง (θ) จะพบว่ามีผลคะแนนที่แตกต่างกันแต่ยังคงมีทิศทางคะแนนไปในทิศทางเดียวกัน เมื่อคะแนนมาตรฐานของความสามารถที่แท้จริง (θ) สูงขึ้นหรือลดลง จะส่งผลให้คะแนนมาตรฐานของผลการเรียน 6 เทอม สูงขึ้นหรือลดลงด้วยเช่นกัน แต่ยังพบว่ากรณีที่คะแนนมาตรฐานของความสามารถที่แท้จริง (θ) สูงกว่าคะแนนมาตรฐานของผลการเรียน 6 เทอม อาจเป็นอุปมาได้ว่าเป็นผลมาจากที่นักเรียนอาจจะเรียนในแผนการเรียนที่ไม่ถนัดหรือตรงความสามารถของนักเรียน

3. ผลการตรวจสอบหลักฐานการประเมินด้านความตรงตามเกณฑ์มาตรฐานการประเมิน

3.1 หลักฐานความตรงด้านโครงสร้างภายใน ผู้วิจัยได้ทำการตรวจสอบความตรงโดยใช้โมเดลการตอบสนองข้อสอบแบบพหุมิติ (MRCML) โดยมีการนำเสนอในรูปของ Wright Map พร้อมทั้งดัชนีที่สะท้อนถึงความตรงเชิงโครงสร้างภายใน โดยใช้โปรแกรม ACER Conquest 2.0 (Wu, Adams, Wilson, & Haldane, 2007) หลักฐานความเที่ยงด้วยค่าสัมประสิทธิ์สัมพันธ์ (Reliability evidence based on reliability coefficient) ได้แก่

1) ความเที่ยงแบบ Expected A Posteriori หรือ EAP/PV (EAP/PV Reliability) (Briggs & Wilson, 2003) ความแปรปรวนของความสามารถผู้เรียนได้ผ่านเกณฑ์โดยมีค่าเท่ากับ .955 (EAP/PV RELIABILITY: 0.955)

2) ค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation coefficient) จากการหาความสัมพันธ์ มีค่าความเที่ยงเท่ากับ .95 (Coefficient Alpha: 0.95) ซึ่งมีความสัมพันธ์กัน

3.2 หลักฐานความตรงที่สัมพันธ์กับเกณฑ์อื่น

1) ตรวจสอบความตรงตามสภาพ โดยพิจารณาจากสัมประสิทธิ์สหสัมพันธ์กับตัวแปรภายนอก คือ ผลการเรียนเฉลี่ย 6 เทอม กับคะแนนความสามารถที่แท้จริงที่ประมาณค่าได้จากโปรแกรม จะพบว่าคุณภาพของการตัดเกรด 4 ลักษณะ (1) วิชาที่ตัดเกรดได้เหมาะสม (2) วิชาที่ยากต่อการได้เกรดสูงมีแนวโน้มกดเกรด (3) วิชาที่ง่ายต่อการได้เกรดสูงมีแนวโน้มปล่อยเกรด และ (4) วิชาที่ไม่สามารถแบ่งแยกความแตกต่างระหว่างเกรด

2) ตรวจสอบความตรงเชิงทำนาย โดยพิจารณาจากสัมประสิทธิ์สหสัมพันธ์กับตัวแปรภายนอกที่สะท้อนถึงความสามารถของนักเรียน คือ คะแนน GPA ทั้ง 17 รายวิชา กับผลการเรียนเฉลี่ย 6 ภาคเรียน มีความสอดคล้องกันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 (X^2 = 69.104, df=1, Sig=.000) คะแนนมาตรฐานของความสามารถที่แท้จริง (θ) สามารถทำนายคะแนนมาตรฐานของผลการเรียน 6 เทอม

บทที่ 5

สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ

การดำเนินการวิจัยครั้งนี้ มีวัตถุประสงค์เพื่อตรวจสอบคุณภาพของการตัดเกรดในแต่ละรายวิชา และตรวจสอบคุณภาพผู้เรียน รวมไปถึงเพื่อตรวจสอบความสอดคล้องของระดับความสามารถที่แท้จริงของนักเรียนระดับชั้นมัธยมศึกษาปีที่ 6 กับผลการเรียนเฉลี่ย 6 เทอม ผู้วิจัยได้ดำเนินการรวบรวมข้อมูลและวิเคราะห์ข้อมูลในการวิจัย จึงขอนำเสนอผลการวิจัยและอภิปรายผลการวิจัยตามลำดับดังนี้

1. ผลการตรวจสอบคุณภาพของการตัดเกรดในแต่ละรายวิชา

2. ผลการตรวจสอบความสอดคล้องของระดับความสามารถที่แท้จริงของนักเรียนระดับชั้นมัธยมศึกษาปีที่ 6 กับผลการเรียนเฉลี่ย 6 ภาคเรียน

มีรายละเอียด ดังนี้

1. ผลการตรวจสอบคุณภาพของการตัดเกรดในแต่ละรายวิชา

การตรวจสอบคุณภาพการตัดเกรด โดยใช้โมเดลโลจิสตามทฤษฎีการตอบสนองแบบพหุมิติ (Multidimensional Random Coefficients Multinomial Logit Model) ใช้ชื่อย่อว่า “MRCLM”) โดยใช้โปรแกรม ConQuest 2.0 (Wu, Adums, Wilson, & Haldane, 2007) จากการวิเคราะห์ผลการตรวจสอบคุณภาพการตัดเกรด จะพบความยากง่ายในการตัดเกรดของแต่ละวิชา ดังนี้

- รายวิชาที่มีความยากในการได้เกรดสูง ได้แก่ รายวิชาที่ 17, 7, 11 ซึ่งเป็นรายวิชาที่เน้นการฝึกฝนทักษะเฉพาะเช่น ความสามารถทางด้านภาษา ความสามารถทางด้านศิลปะ

- รายวิชาที่ง่ายต่อการได้เกรดสูง ได้แก่ รายวิชาที่ 5, 6, 14 เป็นรายวิชาปฏิบัติเป็นส่วนใหญ่และเน้นทักษะการปฏิบัติซึ่งนักเรียนอาจจะมีความสามารถที่จะปฏิบัติได้จริง หรือเป็นผลจากการตัดเกรด

- รายวิชาที่มีโอกาสได้เกรดตามความสามารถหรือตัดเกรดได้เหมาะสมกับความสามารถที่แท้จริงของนักเรียน ได้แก่ รายวิชาที่ 2, 1, 12, 16, 4, 10, 8, 3, 15, 13

เมื่อพิจารณาผลการวิเคราะห์ข้อมูล สะท้อนให้เห็นคุณภาพของการตัดเกรด 4 ลักษณะ ดังต่อไปนี้

(1) วิชาที่ตัดเกรดได้เหมาะสม เป็นวิชาที่สะท้อนให้เห็นถึงระดับความสามารถของผู้เรียนได้อย่างชัดเจนสามารถจำแนกการตัดเกรดได้ตรงกับระดับความสามารถของผู้เรียนได้โดยเฉพาะตำแหน่งที่สามารถจำแนกความแตกต่างระหว่างผู้เรียนที่ควรได้เกรด 4 และ ไม่ได้เกรด 4 เช่น วิชาที่ 1, 3, 8, 12, 16 การตัดเกรดที่เหมาะสมกับความสามารถที่แท้จริงส่งผลให้ผู้เรียนได้รับผลการเรียนที่เหมาะสม รวมทั้งสะท้อนให้เห็นความถนัดหรือความเชี่ยวชาญของนักเรียนในการเรียนรู้ในแต่ละแผนการเรียนได้ด้วย

(2) วิชาที่ยากต่อการได้เกรดสูงมีแนวโน้มกดเกรด จาก Wright Map สะท้อนให้เห็นว่ามีบางรายวิชาที่ผู้เรียนต้องมีความสามารถสูงมากจึงจะได้เกรด 4 ได้แก่ วิชาที่ 2, 4, 7, 9, 10, 11, 17 แสดงว่า

รายวิชาดังกล่าวมีแนวโน้มตัดเกรดต่ำกว่าความเป็นจริง จึงมีแนวโน้มในการตัดกดเกรด สะท้อนให้เห็นว่าการตัดเกรดไม่ได้เป็นไปตามระดับความสามารถของผู้เรียนซึ่งจะต่ำกว่าความเป็นจริง นอกจากจะไม่ได้สะท้อนสารสนเทศย้อนกลับ (Feed back) เพื่อให้นักเรียนได้รับข้อมูลเพื่อปรับปรุง พัฒนาใด ๆ (Feed up, Feed Forward) ตามแนวคิดของ Black and William (2009), Chappuis (2015), Frey and Fisher (2011), Hattie and Timperley (2007) และยังแสดงให้เห็นคุณภาพของการจัดการเรียนรู้ และการวัดและประเมินผลของครูประจำรายวิชาอาจมีข้อมูลผิดพลาด หรือใช้รูปแบบที่ไม่เหมาะสมหรือไม่หลากหลายที่จะวัดและประเมินผลผู้เรียนให้ได้ระดับความสามารถที่แท้จริงได้ นอกจากนี้ยังส่งผลต่อภาพรวมของผลการเรียนเฉลี่ยของนักเรียนทั้ง 6 ภาคเรียน อาจจะส่งผลต่อการเข้าเรียนในระดับที่สูงขึ้น หรือแม้แต่การขอทุนการศึกษาที่มีการกำหนดเกรดเฉลี่ยขั้นต่ำไว้ด้วย

(3) วิชาที่ง่ายต่อการได้เกรดสูงมีแนวโน้มปล่อยเกรด จาก Wright Map สะท้อนให้เห็นว่ามีบางรายวิชาที่ผู้เรียนต้องมีความสามารถปานกลางค่อยมาทางอ่อนมีโอกาสได้เกรด 4 สูง เช่น วิชาที่ 13 จะแสดงให้เห็นว่าคุณภาพของการจัดการเรียนรู้ และการวัดและประเมินผลของครูประจำรายวิชาอาจมีข้อมูลผิดพลาด หรือใช้รูปแบบที่ไม่เหมาะสมหรือไม่หลากหลาย แต่เน้นต้องการให้นักเรียนได้รับผลการเรียนที่ดีหรือเกรดที่สูงโดยไม่ได้คำนึงถึงผลการประเมิน หรือคุณภาพของการประเมินเพื่อหาความสามารถที่แท้จริงของผู้เรียน การประเมินผลในลักษณะนี้ก็ส่งผลเสียทำให้ไม่ได้รับสารสนเทศย้อนกลับ (Feed back) เพื่อให้นักเรียนได้รับข้อมูลเพื่อปรับปรุง พัฒนาใด ๆ (Feed up, Feed Forward)

(4) วิชาที่ไม่สามารถแบ่งแยกความแตกต่างระหว่างเกรด จากผลการวิเคราะห์พบว่าบางรายวิชาตัดเกรดได้เหมาะสมระหว่างผู้เรียนที่ควรได้เกรด 4 และไม่ได้เกรด 4 แต่การตัดเกรดบางรายวิชาไม่สามารถแบ่งแยกความแตกต่างระหว่างเกรดได้ เนื่องจากอยู่ในตำแหน่งเดียวกัน เช่น รายวิชาที่ 5, 6, 14, 15 ไม่สามารถจำแนกความสามารถผู้เรียนในระดับเกรดที่ต่างกันได้ เป็นประเภทที่สะท้อนให้เห็นการตัดเกรดที่ไม่ได้มาตรฐานอย่างมาก หมายความว่า รูปแบบข้อที่ 2 กดเกรด และ ข้อที่ 3 ปล่อยเกรด แต่ยังคงสามารถแบ่งระดับความสามารถที่แท้จริงของนักเรียนได้ครบทุกระดับ แต่รูปแบบที่ 4 ไม่สามารถแบ่งแยกความแตกต่างระหว่างเกรด จะไม่สามารถระบุอะไรได้เลยเพราะ ไม่ว่านักเรียนจะมีความสามารถระดับใดก็มีสิทธิ์ที่จะได้ทุกเกรด เช่น นักเรียนมีความสามารถระดับต่ำ ก็มีสิทธิ์ได้เกรด 4 เท่ากันกับนักเรียนที่มีความสามารถระดับสูง ดังนั้นรูปแบบนี้จึงเป็นการตัดเกรดประเภทที่ไม่สามารถวัดอะไรตามหลักการวัดและประเมินผลได้เลย ซึ่งสะท้อนให้เห็นถึงคุณภาพของการวัดและการประเมินผล อาจจะไม่ได้คุณภาพหรือไม่มีการวัดและประเมินผลเลย

จะเห็นว่าส่วนใหญ่ตัดเกรดแบบกดเกรด (41.176%) รองลงมาคือ ตัดเกรดเหมาะสม (29.411%) ตัดเกรดแยกเกรดไม่ได้ (29.411%) และ ตัดเกรดปล่อยเกรด (5.882%) ตามลำดับ

2. ผลการตรวจสอบความสอดคล้องของระดับความสามารถที่แท้จริงของนักเรียนระดับชั้นมัธยมศึกษาปีที่ 6 กับผลการเรียนเฉลี่ย 6 ภาคเรียน

การตรวจสอบคุณภาพผู้เรียน เพื่อให้เข้าใจง่ายขึ้นผู้วิจัยได้แปลงเป็นคะแนนสเกลที่มีค่าเฉลี่ยเป็น 500 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 100 โดยใช้สูตร 500 + 100(θ) เทียบกับผลการเรียนเฉลี่ย 6 เทอม จะพบว่า

คะแนนมาตรฐานของผลการเรียน 6 เทอม ซึ่งไม่สอดคล้องกับข้อมูลผลการวิเคราะห์ทิศทางการตัดเกรด ซึ่งอาจเป็นผลมาจากการวิเคราะห์ทิศทางการตัดเกรดเป็นการวิเคราะห์ผลการเรียนเฉพาะรายวิชาพื้นฐาน ซึ่งไม่รวมผลการเรียนรายวิชาเพิ่มเติมอื่น ๆ ของแต่ละแผนการเรียน ซึ่งเป็นรายวิชาที่เป็นที่มีความถนัดเฉพาะของผู้เรียน แต่คะแนนมาตรฐานของผลการเรียน 6 เทอมรวมทั้งรายวิชาพื้นฐานและรายวิชาเพิ่มเติมเฉพาะของแต่ละแผนการเรียนอาจเป็นผลทำให้คะแนนความสามารถ ส่งผลให้คะแนนมาตรฐานของผลการเรียน 6 เทอมมีคะแนนที่สูงมาก

เมื่อพิจารณาคะแนนมาตรฐานของความสามารถที่แท้จริง (θ) เป็นไปในทิศทางเดียวกันกับข้อมูลผลการวิเคราะห์ทิศทางการตัดเกรด เพราะทั้งสองจะใช้ข้อมูลชุดเดียวกัน คือ ผลการเรียนเฉพาะรายวิชาพื้นฐานเท่านั้น ระดับความสามารถที่แท้จริงส่วนใหญ่อยู่ระดับปานกลาง ซึ่งก็สอดคล้องกับคุณภาพของการตัดเกรด ที่มีรายวิชากดเกรด 7 รายวิชา คิดเป็นร้อยละ 41.176 ของรายวิชาพื้นฐานที่ทำการวิเคราะห์ทั้งหมด ซึ่งอาจจะส่งผลให้ระดับความสามารถต่ำกว่าความเป็นจริง และมีคะแนนมาตรฐานของความสามารถที่แท้จริง (θ) ห่างจากคะแนนมาตรฐานของผลการเรียน 6 เทอมค่อนข้างสูง

ทั้งนี้คะแนนมาตรฐานของผลการเรียน 6 เทอมและคะแนนมาตรฐานของความสามารถที่แท้จริง (θ) จะพบว่ามีผลคะแนนที่แตกต่างกันค่อนข้างสูง แต่สะท้อนให้เห็นว่ายังคงมีทิศทางคะแนนไปในทิศทางเดียวกัน เมื่อคะแนนมาตรฐานของความสามารถที่แท้จริง (θ) สูงขึ้นหรือลดลง จะส่งผลให้คะแนนมาตรฐานของผลการเรียน 6 เทอม สูงขึ้นหรือลดลงด้วยเช่นกัน

จุดที่น่าสังเกตคือมีบางชุดข้อมูลที่คะแนนมาตรฐานของความสามารถที่แท้จริง (θ) สูงกว่าคะแนนมาตรฐานของผลการเรียน 6 เทอม ทั้งที่ความเป็นจริงไม่ควรเกิดขึ้นเพราะคะแนนมาตรฐานของผลการเรียน 6 เทอม ต้องรวมรายวิชาที่นักเรียนถนัดและทำเกรดได้ดี ควรให้ค่าคะแนนมาตรฐานที่สูงกว่าเสมอ แต่ในกรณีนี้กลับน้อยกว่า อาจจะเป็นการบ่งบอกถึงว่านักเรียนอาจจะไม่ได้มีความถนัดหรือความชื่นชอบในสาย/แผนการเรียนในปัจจุบันก็ได้ จึงส่งผลให้คะแนนมาตรฐานของผลการเรียน 6 เทอม ต่ำกว่า คะแนนมาตรฐานของความสามารถที่แท้จริง (θ) ซึ่งมีเฉพาะรายวิชาพื้นฐานเท่านั้น

ข้อเสนอแนะ

1. การศึกษาครั้งนี้เป็นการศึกษาโดยใช้ข้อมูลที่เป็นกลุ่มตัวอย่างเท่านั้น เพื่อให้ได้สารสนเทศที่ครบถ้วนมากยิ่งขึ้นควรใช้ข้อมูลทั้งหมด เพื่อให้ถูกต้องและครบถ้วนทั้งหมด

2. หลักฐานความตรงที่สัมพันธ์กับเกณฑ์อื่น งานวิจัยครั้งนี้ได้ใช้ผลการเรียนเฉลี่ย 6 เทอม เพื่อให้ได้สารสนเทศหรือความตรงที่น่าเชื่อมั่นควรเทียบกับเกณฑ์ภายนอกอื่น เช่น ผลการทดสอบทางการศึกษาแห่งชาติ (O-NET) หรือผลการสอบเข้าในระดับมหาวิทยาลัย

บรรณานุกรม

เจนรบ โกรธา และพัชรี จันทร์เพ็ง. (2561). การประเมินระดับความสามารถทางคณิตศาสตร์ของผู้เรียนแบบพหุมิติ โดยใช้เทคโนโลยีสารสนเทศในการรายงานผลการประเมินเพื่อการเรียนรู้. ใน พัชรี จันทร์เพ็ง (บรรณาธิการ). **เอกสารประกอบการประชุมสัมมนาวิชาการการวัด ประเมินผลและวิจัยสัมพันธ์แห่งประเทศไทย ครั้งที่ 26.** (หน้า 181-198). ขอนแก่น. โรงพิมพ์มหาวิทยาลัยขอนแก่น.

ชัยวิชิต เชียรชนะ. (2552). **การวิเคราะห์พหุมิติ (Multidimensional Analysis)**. ขอนแก่น: คณะศึกษาศาสตร์ มหาวิทยาลัยขอนแก่น.

พัชรี จันทร์เพ็ง. (2561). **การประยุกต์ทฤษฎีการตอบสนองข้อสอบแบบพหุมิติเพื่อการวิจัย.** ขอนแก่น: โรงพิมพ์มหาวิทยาลัยขอนแก่น.

ศิริชัย กาญจนวาสี. (2550). **ทฤษฎีการทดสอบแนวใหม่.** พิมพ์ครั้งที่ 3. กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

_____. (2550). **ทฤษฎีการทดสอบแบบดั้งเดิม.** พิมพ์ครั้งที่ 6. กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

ศิริเดช สุชีวะ. (2550). การวิเคราะห์จุดอ่อนและจุดแข็งของผู้เรียน. ใน สุวิมล ว่องวานิช (บรรณธิการ). **หนังสือชุดปฏิรูปการศึกษา “การประเมินผลการเรียนรู้แนวใหม่**”. (หน้า 154-176). พิมพ์ครั้งที่ 2. กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

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

Chappuis, J. (2009). **Seven strategies of assessment for learning**. Upper Saddle River, NJ: Pearson Education.

Hattie, J., & Timperley, H. (2007). The power of feedback. **Review of Educational Research**, 77, 81-112.

Heritage, M. (2008). **Learning progressions: Supporting instruction and formative assessment**. Washington, D.C.: Council of Chief State School Officers.

Chappuis, J. (2012). How am I doing?. **Education Leadership**, 70(1).

Wang, W., Wilson, M., & Adams, R. J. (1997). Rasch models for multidimensionality between items and within items. In M. Wilson & G. Engelhard (Eds.). **Objective measurement: Theory into practice**. (vol. 4, pp. 139-155). [n.p.].

Wilson, M. (2005). **Constructing Measures: An Item Response Modeling Approach (Har/Cdr edition). Mahwah**, N.J: Routledge.

Wilson, M., & Hoskens, M. (2005). Multidimensional item response: Multimethod/Multitrait perspective. In S. Alagumalai, D.D. Curtis & N. Hungi (Eds.). **Applied rasch measurement: A book of exemplars papers in honour of John P.Keeves**. (pp.287-307). Netherlands: Springer.

Wu, M. L., Adams, R. J., & Wilson, M. R. (2007). **ACER ConQuest: Generalised item response modelling software**. Australia: ACER Press

ภาคผนวก

คำสั่งในการประมวลผล

```
Title Partial Credit Model: What happened last night;
data data.dat;
format name 1-3 responses 5-21;
labels << mps.lab;
codes 0,1,2,3,4,5,6,7;
recode (1,2,3,4,5,6,7,8) (0,1,2,3,4,5,6,7)              !items(1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17);
model item + item*step;
estimate;
show !estimates=latent >> data.shw;
itanal >> data.itn;
plot expected! gins=2;
plot icc! gins=2;
plot ccc! gins=2;
show cases !estimates=eap >> data.eap;
reset;
Title Partial Credit Model: What happened last night;
data data.dat;
format name 1-3 responses 5-21;
labels << mps.lab;
codes 0,1,2,3,4,5,6,7;
recode (1,2,3,4,5,6,7,8) (0,1,2,3,4,5,6,7)              !items(1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17);
model item + item*step;
estimate;
show !estimates=latent >> data.shw;
itanal >> data.itn;
plot expected! gins=2;
plot icc! gins=2;
plot ccc! gins=2;
show cases !estimates=eap >> data.eap;
```

ไฟล์การประมาณค่าความสามารถนักเรียนรายบุคคล (Estimate)

```
   2     2.72572    0.36887    0.92472
   3     0.55735    0.23659    0.96903
   4     2.08262    0.29563    0.95165
   5     2.72572    0.36887    0.92472
   6     1.24786    0.25001    0.96542
   7     3.91282    0.57211    0.81891
   8     3.91282    0.57211    0.81891
   9     3.02615    0.40924    0.90734
  10     1.37564    0.25659    0.96357
  11     0.24859    0.21943    0.97336
  12     0.61434    0.24093    0.96788
  13     1.82584    0.28942    0.95366
  14     2.36488    0.32383    0.94198
  15     2.59506    0.35417    0.93060
  16     2.36488    0.32383    0.94198
  17     4.28805    0.64330    0.77104
  18    -0.51844    0.21725    0.97389
  19     1.24786    0.25001    0.96542
  20     1.31091    0.25248    0.96473
  21    -0.60809    0.20587    0.97655
  22     0.86351    0.25614    0.96370
  23     2.36488    0.32383    0.94198
  24     2.17136    0.30080    0.94994
  25    -0.12408    0.21632    0.97411
  26     1.99598    0.29334    0.95239
  27     1.91037    0.29183    0.95288
  28     1.18557    0.24950    0.96556
  29     0.44899    0.22922    0.97093
  30     0.55735    0.23659    0.96903
  31     1.82584    0.28942    0.95366
  32     0.73492    0.25003    0.96541
  33     1.91037    0.29183    0.95288
  34     1.66312    0.28912    0.95659
```

ไฟล์ผลการวิเคราะห์รายวิชา

```
================================================================================
Partial Credit Model: What happened last night             Thu Jan 02 19:22 2025
SUMMARY OF THE ESTIMATION
================================================================================

Estimation method was: Gauss-Hermite Quadrature with 15 nodes
Assumed population distribution was: Gaussian
Constraint was: DEFAULT
The Data File: data.dat
The format:  name 1-3 responses 5-21
The regression model:
Grouping Variables:
The item model: item+item*step
Sample size: 160
Final Deviance:     6407.62626
Total number of estimated parameters: 103
The number of iterations: 810
Termination criteria:  Max iterations=1000, Parameter Change= 0.00010
                       Deviance Change= 0.00010
Iterations terminated because the deviance convergence criteria was reached
Random number generation seed:    1.00000
Number of nodes used when drawing PVs: 2000
Number of nodes used when computing fit: 200
Number of plausible values to draw: 5
Maximum number of iterations without a deviance improvement: 100
Maximum number of Newton steps in M-step: 10
Value for obtaining finite MLEs for zero/perfects:    0.30000



================================================================================

================================================================================
Partial Credit Model: What happened last night             Thu Jan 02 19:22 2025
TABLES OF RESPONSE MODEL PARAMETER ESTIMATES
================================================================================
TERM 1: item
--------------------------------------------------------------------------------------------------
   VARIABLES                              UNWEIGHTED FIT             WEIGHTED FIT
---------------                       -----------------------   -----------------------
     item           ESTIMATE  ERROR^   MNSQ       CI        T    MNSQ       CI        T
--------------------------------------------------------------------------------------------------
 1   THAI1            0.406   0.053    0.76 ( 0.78, 1.22) -2.3   0.77 ( 0.78, 1.22) -2.2
 2   MATH1            0.863   0.055    1.05 ( 0.78, 1.22)  0.4   1.07 ( 0.78, 1.22)  0.6
 3   SCI1            -0.259   0.056    0.69 ( 0.78, 1.22) -3.0   0.75 ( 0.72, 1.28) -1.9
 4   COM1             0.081   0.057    1.17 ( 0.78, 1.22)  1.5   1.23 ( 0.78, 1.22)  1.9
 5   SOC1            -1.242   0.059    1.35 ( 0.78, 1.22)  2.9   1.15 ( 0.54, 1.46)  0.7
 6   PHE1            -1.621   0.068    9.64 ( 0.78, 1.22) 30.3   1.42 ( 0.15, 1.85)  0.9
 7   ART1             0.960   0.052    1.40 ( 0.78, 1.22)  3.2   1.24 ( 0.76, 1.24)  1.9
 8   WORK1            0.009   0.057    0.78 ( 0.78, 1.22) -2.1   0.81 ( 0.76, 1.24) -1.6
 9   ENG1             0.718   0.056    0.92 ( 0.78, 1.22) -0.7   0.94 ( 0.78, 1.22) -0.5
 10  THAI2            0.193   0.057    0.69 ( 0.78, 1.22) -3.1   0.67 ( 0.78, 1.22) -3.3
 11  MATH2            0.981   0.052    1.09 ( 0.78, 1.22)  0.8   1.16 ( 0.77, 1.23)  1.4
 12  COM2             0.348   0.054    1.18 ( 0.78, 1.22)  1.6   1.24 ( 0.78, 1.22)  2.0
 13  SOC2            -1.027   0.060    1.59 ( 0.78, 1.22)  4.5   1.27 ( 0.68, 1.32)  1.6
 14  PHE2            -1.699   0.070    0.69 ( 0.78, 1.22) -3.1   1.05 ( 0.18, 1.82)  0.2
 15  ART2            -0.157   0.055    1.11 ( 0.78, 1.22)  1.0   1.33 ( 0.73, 1.27)  2.2
 16  WORK2            0.305   0.057    0.91 ( 0.78, 1.22) -0.8   0.95 ( 0.73, 1.27) -0.3
 17  ENG2             1.142*  0.230    0.88 ( 0.78, 1.22) -1.1   0.89 ( 0.78, 1.22) -1.0
--------------------------------------------------------------------------------
An asterisk next to a parameter estimate indicates that it is constrained
Separation Reliability =  0.996
Chi-square test of parameter equality =    3181.58,  df = 16,  Sig Level = 0.000
^ Quick standard errors have been used
--------------------------------------------------------------------------------
```

ไฟล์ผลการวิเคราะห์แต่ละรายวิชาในแต่ละขั้นความสามารถ

```
================================================================================
Partial Credit Model: What happened last night            Thu Jan 02 19:22 2025
GENERALISED ITEM ANALYSIS
================================================================================
Item 1
------
item:1 (THAI1)
Cases for this item    160   Discrimination  0.88
Item Threshold(s):    -1.01 -0.60  0.05  0.75  1.23  1.95   Weighted MNSQ   0.77
Item Delta(s):     -0.09 -1.07  0.03  1.01  1.09  1.47
--------------------------------------------------------------------------------
 Label    Score     Count   % of tot  Pt Bis     t  (p)   PV1Avg:1 PV1 SD:1
--------------------------------------------------------------------------------
   0       0.00       17      10.63   -0.55    -8.27(.000) -0.89     0.63
   1       1.00        9       5.63   -0.35    -4.73(.000) -0.63     0.58
   2       2.00       21      13.13   -0.29    -3.84(.000) -0.13     0.57
   3       3.00       28      17.50   -0.03    -0.32(.752)  0.59     0.57
   4       4.00       22      13.75    0.06     0.71(.477)  0.98     0.60
   5       5.00       25      15.63    0.24     3.15(.002)  1.46     0.65
   6       6.00       38      23.75    0.59     9.18(.000)  2.62     1.16
================================================================================

Item 2
------
item:2 (MATH1)
Cases for this item    160   Discrimination  0.81
Item Threshold(s):    -1.18 -0.37  0.48  1.32  2.16  2.80   Weighted MNSQ   1.07
Item Delta(s):     -0.71 -0.49  0.49  1.28  2.43  2.18
--------------------------------------------------------------------------------
 Label    Score     Count   % of tot  Pt Bis     t  (p)   PV1Avg:1 PV1 SD:1
--------------------------------------------------------------------------------
   0       0.00       16      10.00   -0.50    -7.30(.000) -0.74     0.68
   1       1.00       18      11.25   -0.47    -6.75(.000) -0.63     0.57
   2       2.00       30      18.75   -0.06    -0.78(.438)  0.61     0.78
   3       3.00       34      21.25    0.05     0.58(.560)  0.82     0.71
   4       4.00       29      18.13    0.25     3.29(.001)  1.52     0.90
   5       5.00       14       8.75    0.24     3.15(.002)  1.89     0.96
   6       6.00       19      11.88    0.43     5.98(.000)  3.02     1.36
================================================================================

Item 3
------
item:3 (SCI1)
Cases for this item    160   Discrimination  0.86
Item Threshold(s):    -1.12 -0.96 -0.73 -0.46  0.22  1.33   Weighted MNSQ   0.75
Item Delta(s):      0.41 -1.48 -0.78 -0.93  0.20  1.02
--------------------------------------------------------------------------------
 Label    Score     Count   % of tot  Pt Bis     t  (p)   PV1Avg:1 PV1 SD:1
--------------------------------------------------------------------------------
   0       0.00       13       8.13   -0.57    -8.75(.000) -1.11     0.43
   1       1.00        3       1.88   -0.18    -2.29(.023) -0.74     0.43
   2       2.00        6       3.75   -0.25    -3.31(.001) -0.57     0.25
   3       3.00        9       5.63   -0.24    -3.10(.002) -0.25     0.90
   4       4.00       26      16.25   -0.22    -2.77(.006)  0.14     0.66
   5       5.00       42      26.25    0.09     1.12(.263)  0.90     0.60
   6       6.00       61      38.13    0.67    11.27(.000)  2.16     1.15
================================================================================

Item 4
------
item:4 (COM1)
Cases for this item    160   Discrimination  0.73
Item Threshold(s):    -1.75 -1.42 -0.70  0.21  1.24  2.82   Weighted MNSQ   1.23
Item Delta(s):     -0.66 -2.09 -0.73  0.20  1.12  2.64
--------------------------------------------------------------------------------
 Label    Score     Count   % of tot  Pt Bis     t  (p)   PV1Avg:1 PV1 SD:1
```

ประวัติผู้วิจัย

ชื่อ-นามสกุล	นายเจนรบ โกรธา
ตำแหน่ง	ครู วิทยฐานะ ครูชำนาญการ โรงเรียนบ้านฉางกาญจนกุลวิทยา สำนักงานเขตพื้นที่การศึกษามัธยมศึกษาชลบุรี ระยอง
การศึกษา	ปริญญาตรี: ศึกษาศาสตรบัณฑิต (คอมพิวเตอร์ศึกษา) คณะศึกษาศาสตร์ มหาวิทยาลัยขอนแก่น (เกียรตินิยมอันดับ 1) ปริญญาโท: ศึกษาศาสตรมหาบัณฑิต (การวัดและประเมินผลการศึกษา) คณะศึกษาศาสตร์ มหาวิทยาลัยขอนแก่น
ผลงานวิชาการ	• **การประเมินระดับความสามารถทางคณิตศาสตร์ของผู้เรียนแบบพหุมิติโดยใช้เทคโนโลยีสารสนเทศในการรายงานผลการประเมินเพื่อการเรียนรู้** (https://thekunkrootum.com/aca-view.php?m_id=67) • **Constructing Progress Maps of Digital Technology for Diagnosing Mathematical Proficiency** (https://files.eric.ed.gov/fulltext/EJ1232908.pdf) • **การพัฒนาระบบการประมวลผลระดับความสามารถทางคณิตศาสตร์ของผู้เรียนแบบพหุมิติโดยใช้เทคโนโลยีสารสนเทศในการรายงานผลการประเมินเพื่อการเรียนรู้** (https://opac.kku.ac.th/catalog/BibItem.aspx?BibID=b00437159) • **DIAGNOSTIC OF MATHEMATICAL PROFICIENCY USING DIGITAL TRANSFORMATION** (https://futureofedu.co/future-edu-2019/) • **โครงการการสร้างเครื่องมือวินิจฉัยระดับความสามารถทางคณิตศาสตร์ของผู้เรียนโดยใช้เทคโนโลยีสารสนเทศในการรายงานผลการประเมินเพื่อการเรียนรู้** (https://digital.library.tu.ac.th/tu_dc/frontend/Info/item/dc:176530)