



บทที่ 5

สรุปผลการวิจัยและข้อเสนอแนะ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อศึกษาคะแนนจุดตัดที่คำนวณจากทฤษฎีการตอบสนองข้อกระทง (IRT) และทฤษฎีตามแนวคลาสสิก โดยกำหนดคะแนนโดเมนต่างกันและความยาวแบบสอบต่างกันเพื่อให้ผลต่างกันเพียงใด และจะมีผลต่อค่าความน่าจะเป็นในการจำแนกผิดพลาดในลักษณะใด เพื่อจะใช้จำแนกกลุ่มผู้รอบรู้และไม่รอบรู้ และศึกษาว่าผลของคะแนนจุดตัดที่มีต่อความเที่ยงในการตัดสินใจที่คำนวณตามแนวคิดทฤษฎีต่างกันแนวโน้มนต่างกันอย่างไร ๕ จุดตัดที่ใช้ในการจำแนก เมื่อใช้วิธีการคำนวณค่าความเที่ยงตามวิธีประยุกต์ของวอร์ม (Warm, 1980) และวิธีหาค่าความเที่ยงตามแนวคลาสสิกอิน โดยใช้สัมประสิทธิ์แคปปา

กลุ่มตัวอย่างที่ใช้ในการวิจัยมี 2 กลุ่ม คือ กลุ่มแรกเป็นกลุ่มตัวอย่างที่ใช้สร้างเครื่องมือ ซึ่งเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 4 แผนการเรียนวิทย์-คณิต และศิลป์คำนวณ โดยทำการสอบกับกลุ่มตัวอย่างนี้ในสัปดาห์แรกของการเปิดภาคเรียนที่ 1 จำนวนนักเรียน 745 คน จากโรงเรียน 8 โรงเรียน จำนวน 21 ห้องเรียน ซึ่งผู้วิจัยถือว่านักเรียนกลุ่มนี้เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 3 กลุ่มที่สองเป็นกลุ่มตัวอย่างที่ใช้ทำการวิจัย ซึ่งเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 3 สังกัดกรมสามัญศึกษา กระทรวงศึกษาธิการ เขตกรุงเทพมหานคร จำนวนนักเรียน 1,005 คน จากโรงเรียน 11 โรงเรียน จำนวน 22 ห้องเรียน ให้ครูในหมวดวิชาคณิตศาสตร์จำนวน 5 ท่าน เป็นผู้เชี่ยวชาญในการประเมินจุดประสงค์ให้สอดคล้องกับโดเมนที่กำหนด กำหนดน้ำหนักความสำคัญของแต่ละจุดประสงค์ เพื่อให้เป็นเกณฑ์สร้างคลังข้อสอบของแต่ละจุดประสงค์ และให้ประเมินข้อกระทงต่าง ๆ ถึงความสอดคล้องกับโดเมนและรูปแบบข้อกระทงที่กำหนด

เครื่องมือที่ใช้ในการวิจัย เป็นแบบสอบอิงเกณฑ์ลักษณะอิงโดเมนแบบเลือกตอบ 5 ตัวเลือก จำนวน 120 ข้อ ผู้วิจัยสร้างขึ้นตามโดเมนที่กำหนด ซึ่งแต่ละข้อวิเคราะห์คุณภาพในด้านความตรงเชิงเนื้อหา (Content Validity) ซึ่งให้ผลการวิเคราะห์ดังนี้

ความตรงเชิงเนื้อหา ได้จากการตัดสินความสอดคล้องระหว่างข้อกระทงกับ จุดประสงค์ของผู้เชี่ยวชาญเนื้อหาวิชา จำนวน 5 ท่าน และใช้สูตรการคำนวณดัชนีความ สอดคล้อง (IOC) ของโรวิเนลลี และแอมเบิลตัน (1977) ปรากฏว่า ข้อกระทงมี ความตรงเชิงเนื้อหาเป็น 1.00 แสดงว่าทุกข้อสามารถวัดได้สอดคล้องกับพฤติกรรมที่ กำหนดในจุดประสงค์ อันเป็นคุณภาพที่สำคัญมากสำหรับแบบสอบอิงเกณฑ์

การคัดเลือกข้อกระทง ใช้การคัดเลือกตามแนวทฤษฎีการตอบสนองข้อกระทง ดังนั้น ต้องทำการตรวจสอบว่า ข้อมูลที่ได้สอดคล้องกับโมเดล IRT ในข้อที่ว่าข้อมูลที่วัดเป็นมิติเดียว (Unidimension) และข้อมูลแต่ละข้อเป็นอิสระต่อกัน จึงต้องดำเนินการตรวจสอบความเป็น มิติเดียวของแบบสอบที่วัด โดยการวิเคราะห์ตัวประกอบ (Factor Analysis) ด้วย โมเดลอัลฟา (Model Alpha) เพราะเป็นแบบสอบผลสัมฤทธิ์เพื่อการอ้างอิงไปยังประชากร ข้อสอบ และทำการสกัดตัวประกอบด้วยวิธีการหมุนแกนแบบแวนแมกซ์ (Varimax) เพราะ เป็นวิธีการที่ตัวประกอบที่สกัดได้แต่ละตัวอยู่ในลักษณะออร์โทกอนอล (Orthogonal) กล่าวคือ ตัวประกอบที่สกัดได้ตั้งฉากกัน มีความสัมพันธ์กันน้อย แสดงให้เห็นว่าตัวประกอบแต่ละตัวมี ความสัมพันธ์กันน้อยมาก หรือไม่มีความสัมพันธ์กันเลย จากนั้นใช้วิธีการวิเคราะห์ข้อกระทง ด้วยคอมพิวเตอร์ โปรแกรมโลจิสติก 5 3 พารามิเตอร์ นิยามข้อที่ให้ค่า $a > .3$, $-2 < b < 2$ และ $c > .3$ สำหรับคัดเลือกเป็นคลังข้อสอบ และนำมาเลือกให้เป็นแบบ สอบความยาว 20 ข้อ และ 30 ข้อ โดยนิยามข้อที่มีความสามารถให้โค้งลักษณะข้อกระทง (ICC) สูง ๆ ณ ความสามารถที่สัมพันธ์กับคะแนนโดเมนที่กำหนด ซึ่งหาได้จากการสุ่มข้อกระทง เมื่อกำหนดคะแนนโดเมน .5, .6, .7 และ .8 ตามลำดับ สำหรับความยาวแบบสอบ 20 และ 30 ข้อตามลำดับ

การเก็บรวบรวมข้อมูลและการวิเคราะห์ข้อมูล

การเก็บรวบรวมข้อมูล ผู้วิจัยนำแบบสอบถามเชิงเทคนิคเลือกตอบ 5 ตัวเลือก ซึ่งสร้างตามโดเมนเนื้อหาที่กำหนดขึ้นและตามรูปแบบข้อกระทงตามวิธีของสงบ ลักษณะ (2527) ไปทดลองสอบกับนักเรียนชั้นมัธยมศึกษาปีที่ 4 ในสัปดาห์แรกของการเรียนการสอนในภาคเรียนที่ 1 จำนวนนักเรียน 745 คน จากโรงเรียน 8 โรงเรียน จำนวน 21 ห้องเรียน ในการสร้างเครื่องมือเพื่อปรับปรุงมาใช้ทดสอบจริงต่อไป หลังจากปรับปรุงแก้ไขข้อสอบแล้ว ทำการทดสอบจริงกับกลุ่มตัวอย่างที่ใช้ทำการวิจัยซึ่งเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 3 จำนวนนักเรียน 1,005 คน จากโรงเรียน 11 โรงเรียน จำนวนห้องเรียน 22 ห้องเรียน แยกแบบสอบเป็น 4 ฉบับ ๆ ละ 30 ข้อ วิชาคณิตศาสตร์ เนื้อหาการคูณหารพหุนาม ค 311 โดยขอความร่วมมือครูที่สอนประจำวิชาคณิตศาสตร์ที่ใช้เป็นกลุ่มตัวอย่างดำเนินการเรียนการสอนเรื่องการคูณหารพหุนามให้ครบถ้วนตามเนื้อหาของหลักสูตร แล้วทำการสอบหลังจากการเรียนการสอน โดยทำการสอบ 4 ครั้ง ในแต่ละห้องที่เป็นกลุ่มตัวอย่าง ทั้งนี้ได้ให้ครูที่สอนวิชาคณิตศาสตร์ ค 311 เนื้อหาการคูณหารพหุนามทำการประเมินจุดประสงค์ในการวัด และพิจารณาว่าข้อกระทงแต่ละข้อสอดคล้องกับจุดประสงค์ และการสร้างตามรูปแบบข้อกระทงด้วย จำนวน 5 ท่าน

การวิเคราะห์ข้อมูล เมื่อเก็บรวบรวมข้อมูลได้นำมาทำการตรวจให้คะแนน 0 เมื่อตอบผิด และคะแนน 1 เมื่อตอบถูก จากนั้นบันทึกข้อมูลของนักเรียนที่เป็นกลุ่มตัวอย่างลงแผ่นแม่เหล็ก (Disket) แล้วถ่ายโอน (Transfer) ข้อมูลลงเทปแม่เหล็ก (Magnetic Tape) ทำการวิเคราะห์ข้อมูลเรื่องการคูณหารพหุนามว่าเป็นไปตามข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อกระทง (IRT) ในการเลือกข้อกระทงมาใช้ คือข้อสอบที่ใช้วัดนั้นเป็นมิติเดียว (Unidimension) โดยการวิเคราะห์ตัวประกอบ (Factor Analysis) ด้วยคอมพิวเตอร์ โปรแกรมสำเร็จรูป SPSSX ที่สถาบันบริการคอมพิวเตอร์ จุฬาลงกรณ์มหาวิทยาลัย ใช้การวิเคราะห์ด้วยโมเดลอัลฟา สกัดตัวประกอบด้วยวิธีเวรีแมกซ์ จากนั้นมาวิเคราะห์ข้อมูลด้วยโปรแกรมสำเร็จรูปโลจิสติก 5 3 พารามิเตอร์ เพื่อมาสร้างคลังข้อสอบ พิจารณาข้อกระทงที่มีค่าอำนาจจำแนก (α) มากกว่า .3 ค่าความยากอยู่ระหว่าง -2 ถึง

$-2 < b < 2$) และค่าการเดา (c) น้อยกว่า .3 มาทำการสุ่มข้อกระทงให้ได้จำนวนข้อกระทง 20 ข้อ และ 30 ข้อตามลำดับ ค่าแนวหาค่า (θ) ค่าความน่าจะเป็นในการตอบถูกของแต่ละข้อในแต่ละความสามารถ (θ) ซึ่งในทางปฏิบัติจะใช้ θ อยู่ระหว่าง -3 ถึง 3 และคะแนนโดเมนโดยใช้คอมพิวเตอร์คำนวณค่า $P_u(\theta)$ ด้วยภาษาฟอร์แทรน กำหนดคะแนนโดเมนที่ต้องการวัด .5, .6, .7 และ .8 ว่าสัมพันธ์กับความสามารถระดับใด เพื่อใช้เป็นแนวทางเลือกข้อกระทงที่ให้ $I \in \{0, u, 3\}$ สูง ใกล้เคียงความสามารถที่คำนวณหาได้ ซึ่งจะทำให้ได้แบบสอบฉบับ 20 ข้อ 4 ฉบับ และฉบับความยาว 30 ข้อ 4 ฉบับเช่นเดียวกัน ลดหลั่นตามค่าความยากของแบบสอบ นำมาหาคะแนนจุดตัดตามแนวคลาสสิกประยุกต์ตามวิธีของแอมเบิลตันและกรูจเตอร์ (1983) คะแนนจุดตัดตามแนว IRT ประยุกต์ จากวิธีของ วิลคอกซ์ (1976) ค่าแนวหาค่าความน่าจะเป็นในการจำแนกความรู้โดยวิธีของไบโนเมียล โดเมล และหาค่าความเที่ยงโดยแนวคลาสสิกของฮวิน (Huynh) กับค่าความเที่ยงตามแนว IRT ประยุกต์ใช้จากวิธีของวอร์ม (Warm)

ผลการวิจัย

ผลการวิจัย สรุปได้ดังนี้

1. การวิเคราะห์ตัวประกอบเพื่อทดสอบความเป็นมิติเดียวของแบบสอบ ปรากฏว่าได้ตัวประกอบ 33 ตัว แต่ตัวประกอบที่ 1 ให้ค่าน้ำหนักตัวประกอบมากที่สุด คือ 20.54681 ส่วนตัวประกอบรองลงมาให้ค่าน้ำหนักตัวประกอบ 1.93314 ซึ่งเป็นค่าที่แตกต่างกันมากกับตัวประกอบที่สกัดด้วยวิธีแวนแม็กซ์ตัวที่ 1 ส่วนตัวประกอบที่ 3 และตัวถัดไปจะไม่ห่างจากกันมากนัก แสดงว่าแบบสอบที่สร้างนี้วัดมิติเดียว อีกทั้งในการสร้างแบบสอบก็ได้กำหนดโดเมนที่จะวัดและสร้างตามรูปแบบข้อกระทง พร้อมค่าสถิติที่ได้จากการวิเคราะห์ตัวประกอบดังกล่าวก็น่าจะยืนยันได้ว่าแบบสอบที่สร้างขึ้นวัดมิติเดียว สามารถวัดเนื้อหาดังกล่าวได้จริง

2. การวิเคราะห์ข้อสอบด้วยคอมพิวเตอร์ โปรแกรมโลจิสติก 5 ที่ใช้นารามิเตอร์ 3 ตัว นิยามค่าอำนาจจำแนก (a) มากกว่า .3 ค่าความยากอยู่ระหว่าง -2 ถึง 2 และค่าการเดาน้อยกว่า .3 ปรากฏว่าข้อสอบที่สร้างขึ้นมา 120 ข้อ ได้คลังข้อสอบที่สอดคล้องกับเกณฑ์ที่กำหนดดังกล่าว 104 ข้อ

3. การคัดเลือกข้อกระทงที่มีค่าความสามารถที่ให้โค้งสารสนเทศข้อกระทง (Item Information curve) สูง ๆ ที่สอดคล้องกับคะแนนโดเมนที่กำหนด .5, .6, .7 และ .8 ตามลำดับ สำหรับความยาวแบบสอบที่ได้จากการสุ่ม 20 ข้อ คือ -.5, -.1, .3 และ .7 ตามลำดับ และที่ระดับความสามารถ -.6, -.2, .2 และ .6 ตามลำดับ สำหรับความยาวแบบสอบที่ได้จากการสุ่ม 30 ข้อ จึงได้แบบสอบ 8 ฉบับเป็นแบบสอบที่มีความยาว 20 ข้อ 4 ฉบับ วัดที่ความสามารถ -.5, -.1, .3 และ .7 กับแบบสอบที่มีความยาว 30 ข้อ 4 ฉบับ วัดที่ระดับความสามารถ -.6, -.2, .2 และ .6 ตามลำดับ

4. แบบสอบฉบับความยาว 20 และ 30 ข้อ เมื่อนำมาวิเคราะห์หาค่าเฉลี่ย (X) ความแปรปรวน (S^2) ค่าความเที่ยง (α_{21}) ของแต่ละฉบับจำนวน 8 ฉบับ ด้วยโปรแกรมสำเร็จรูป SPSSX ปรากฏว่าฉบับความยาว 20 และ 30 ข้อ ค่าเฉลี่ยมีแนวโน้มลดลง ตามคะแนนโดเมนที่กำหนดเพิ่มขึ้น สำหรับฉบับที่มีความยาว 20 ข้อ ความแปรปรวนอยู่ระหว่าง 20 - 23 ค่าความเที่ยง (α_{21}) อยู่ระหว่าง .80 - .84 ส่วนความยาวแบบสอบ 30 ข้อ ความแปรปรวนอยู่ระหว่าง 41-46 ค่าความเที่ยง (α_{21}) อยู่ระหว่าง .85 - .88 ซึ่งค่าที่คำนวณได้เหล่านี้เพื่อนำมาหาค่าความเที่ยงในการจำแนกความรู้ตามแนวทฤษฎีคลาสสิก โดยวิธีของฮวิน จะต้องสอดคล้องกับคะแนนโดเมนที่กำหนด โดยที่คะแนนโดเมนตามแนวทฤษฎีคลาสสิกนี้ประยุกต์จากวิธีของแฮมเบิลตันและกรูจเตอร์ (Hambleton and De Gruijter 1983: 357) โดยใช้วิธีเรียงค่าแนจมากไปน้อย เรียงจากข้อง่ายไปหาข้อยาก ซึ่งปรากฏว่าคะแนนโดเมนตามแนวทฤษฎีคลาสสิกนี้จะเท่ากับคะแนนโดเมนที่กำหนด คือ .5, .6, .7 และ .8 ทั้งความยาวแบบสอบ 20 ข้อ และ 30 ข้อ ซึ่งมีผลทำให้คะแนนจุดตัดที่หาโดยวิธีการดังกล่าวได้ค่าคงที่ คือ คะแนนโดเมน .5, .6, .7 และ .8 มีคะแนนจุดตัดคือ 10, 12, 14 และ 16 ตามลำดับ สำหรับความยาวแบบสอบ 20 ข้อ และคะแนนจุดตัด 15, 18, 21 และ 24 คะแนนตามลำดับ สำหรับความยาวแบบสอบ

30 ข้อ แต่เมื่อนำแบบสอบดังกล่าวมาหาคะแนนจุดตัดตามแนว IRT ที่สอดคล้องกับคะแนนโดเมนที่กำหนด คะแนนจุดตัดกับความยาวแบบสอบ 20 ข้อ คือ 12, 13, 12 และ 13 คะแนน ประมาณ 60% ของความยาวแบบสอบ ส่วนความยาวแบบสอบ 30 ข้อ คะแนนจุดตัดคือ 18, 16, 18 และ 19 คะแนน ประมาณ 60 % ของความยาวแบบสอบเช่นเดียวกัน

5. การหาค่าความน่าจะเป็นในการจำแนกความรอบรู้ผิดพลาดกับความยาว 20 ข้อ และ 30 ข้อที่คิดคำนวณคะแนนจุดตัดตามแนวคลาสสิกกับไบโนเมียลโมเดล จะให้ค่าที่ดังตารางที่ 17, 18 ส่วนที่คิดคำนวณคะแนนจุดตัดตามแนวทฤษฎี IRT ความน่าจะเป็นในการจำแนกความรอบรู้ผิดพลาดจะแปรเปลี่ยนไปตามคะแนนโดเมนและคะแนนจุดตัด โดยที่ความยาวแบบสอบ 20 ข้อ และ 30 ข้อ จะให้ค่าความน่าจะเป็นในการจำแนกความรอบรู้ผิดพลาดประมาณ 18% และ 15% ตามลำดับ ซึ่งให้ผลที่สอดคล้องกับการกำหนดค่าความน่าจะเป็นในการจำแนกความรอบรู้ผิดพลาดไม่ควรเกิน 20% ตามที่วิลคอกซ์กำหนด (Wilcox 1978; citing Hambleton and De Gruijter 1983: 361)

6. ค่าความเที่ยงในการจำแนกความรอบรู้ ถ้าใช้วิธีตามแนวทฤษฎีคลาสสิกที่สอดคล้องกับคะแนนโดเมน ความยาวแบบสอบที่กำหนด ค่าที่ได้จะประมาณ .5 - .6 ส่วนค่าความเที่ยงที่ใช้ตามแนวทฤษฎี IRT ค่าความเที่ยงในการจำแนกความรอบรู้ที่ได้จะให้ค่าสูงมาก คือ ประมาณ .94 - .95

อภิปรายผล

จากผลการวิจัย ผู้วิจัยได้เสนอการอภิปรายผลแยกเป็นประเด็น ดังนี้

1. การทดลองสร้างเครื่องแบบสอบอิงเกณฑ์ที่วัดผลสัมฤทธิ์ในลักษณะอิงโดเมน โดยกำหนดขอบเขตโดเมนของเนื้อหา และนำมาทำการวิจัยเพื่อสร้างคลังข้อสอบ 120 ข้อ โดยวิเคราะห์ด้วยโปรแกรมสำเร็จรูป SPSSX ซึ่งเป็นการวิเคราะห์ตามแนวทฤษฎีคลาสสิก ให้ค่าความเที่ยง ($\alpha_{2,1}$) ทั้งฉบับ .9605 ความแปรปรวน 528.4035 ส่วนเบี่ยงเบนมาตรฐาน 22.9870 ค่าเฉลี่ย 78.2358 ซึ่งเป็นค่าสถิติที่น่าจะสามารถยืนยันได้ว่าแบบสอบทั้ง 120 ข้อนี้สามารถนำไปใช้วัดความรอบรู้ได้คงเส้นคงวา ทั้งนี้เพราะค่าความเที่ยงทั้งฉบับ

ให้ค่าที่สูงมาก เนื่องจากเป็นการวัดจากคนจำนวนมากคือ 1,005 คน แบบสอบมีความยาว และค่าความแปรปรวนสูง ซึ่งเป็นไปตามหลักการที่ว่า แบบสอบที่มีจำนวนข้อมากจะมีค่าความเที่ยงสูงกว่าแบบสอบที่มีจำนวนข้อน้อย นอกจากนี้ในการวิจัยครั้งนี้ได้ทดสอบกับคนจำนวนมากที่มีความสามารถแตกต่างกันมาก จึงมีผลทำให้ค่าความเที่ยงมีค่าสูงมาก ส่วนผลการวิจัยค่าเฉลี่ยของแบบสอบปรากฏว่า ค่าเฉลี่ยของแบบสอบสูงกว่า 50 % คือ 65% แสดงว่าแบบสอบอยู่ในเกณฑ์ค่อนข้างง่าย

2. การวิเคราะห์ตัวประกอบ (Factor Analysis) ของแบบสอบทั้งฉบับ เพื่อพิจารณาว่า แบบสอบนั้นวัดมิติเดียว ให้ผลอยู่ในขั้นที่สามารถแสดงว่า แบบสอบวัดมิติเดียว เนื่องจากน้ำหนักตัวประกอบตัวที่หนึ่งและตัวประกอบอื่น ๆ ถัดมาแตกต่างกันมาก ซึ่งเป็นไปตามหลักการตรวจสอบวัดความเป็นมิติเดียวของแบบสอบที่ลอร์ดได้เสนอไว้ (Lord 1968: 25) นอกจากนี้ลอร์ดให้การสนับสนุนว่า แบบสอบที่สามารถวัดความเป็นมิติเดียว ได้แก่ แบบสอบวัดเหตุผลเชิงคณิตศาสตร์ ขณะเดียวกันแบบสอบนั้นจะมีความเป็นอิสระจากการตอบด้วย (Warm 1978: 107; Lord 1980: 19) ซึ่งเป็นคุณสมบัติที่สอดคล้องกับข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อกระทง จึงสามารถนำข้อกระทงไปวิเคราะห์ด้วยโปรแกรมโลจิสติก 5 3 พารามิเตอร์ ให้ผลที่เหมาะสมกับโมเดล ประกอบกับการสร้างข้อกระทงนั้นได้กำหนดโดเมนเนื้อหาไว้ และสร้างตามรูปแบบข้อกระทง ก็น่าจะยืนยันได้ว่า แบบสอบนี้สามารถวัดเนื้อหาที่ต้องการวัดได้จริง เพราะวูดและคณะ (Wood and others 1976: 5; citing Warm 1978: 90) กล่าวไว้ว่า การวัดให้ได้ผลการคำนวณจากรูปแบบโลจิสติก 3 พารามิเตอร์ ควรใช้กับกลุ่มตัวอย่างขนาดใหญ่ เช่น ประมาณ 1,000 คนขึ้นไป และข้อสอบควรมีจำนวนมาก เช่น ประมาณ 40 ข้อ ซึ่งในงานวิจัยครั้งนี้ได้ใช้แบบสอบที่สอดคล้องกับหลักเกณฑ์ข้อค้นพบของวูด และคณะ ไรท์และปัญจปากีสาน (Wright and Panchapakesan 1969: 25) ได้กล่าวไว้ว่าค่าอำนาจจำแนกไม่จำเป็นต้องเท่ากัน นอกจากนี้การวิเคราะห์ตามโมเดลโลจิสติก 3 พารามิเตอร์นี้ให้ค่าต่าง ๆ ไม่เปลี่ยนแปลงตามกลุ่มตัวอย่างและไม่เปลี่ยนแปลงไปตามตำแหน่งของข้อสอบ อันเป็นคุณสมบัติของความเป็นปรนัย (Objective of Measurement) ที่มีความสำคัญมากในทางวัดผล

3. การหาคะแนนจุดตัดเมื่อแยกฉบับความยาว 20 ข้อ และ 30 ข้อ คะแนนโดเมน .5, .6, .7 และ .8 ตามลำดับ ผลปรากฏว่าคะแนนจุดตัดตามแนวทฤษฎีประมาณ 60 x ทุกฉบับ นอกจากฉบับความยาว 30 ข้อ คะแนนโดเมน .7 ที่คะแนนจุดตัดตามแนวทฤษฎี IRT คือ 16 คะแนนคิดเป็น 50 x และฉบับคะแนนโดเมน .8 คะแนนจุดตัดตามแนวทฤษฎี IRT คือ 20 คะแนนคิดเป็น 65x ซึ่งอาจเกิดจากคลังข้อสอบมีไม่เพียงพอสำหรับความสามารถระดับคะแนนโดเมนดังกล่าว ทำให้มีบางข้อที่จะต้องเลือกข้อที่มีค่าความยากและค่าความสามารถห่างจากระดับความสามารถที่ต้องการวัด คือ .2 สำหรับแบบสอบฉบับคะแนนโดเมน .7 ต่ำไป จึงทำให้คะแนนจุดตัดต่ำกว่า 60 x และแบบสอบฉบับความยาว 30 ข้อคะแนนโดเมน .8 ที่ให้คะแนนจุดตัดเป็น 20 คะแนนคิดเป็น 65x นั้น แม้ว่าจะมีจำนวนข้อซ้ำกันถึง 20 ข้อ น่าจะได้คะแนนจุดตัดใกล้เคียงกัน หรือปาน ๆ กัน แต่กลับได้คะแนนจุดตัดสูงกว่า เพราะว่าความยาวมีค่าใกล้เคียงกับฉบับ .7 แต่ใช้ความสามารถสูงกว่าคือ .8 ย่อมเป็นไปได้ว่าเมื่อต้องการความสามารถสูงขึ้นก็ย่อมจะมีผลทำให้คะแนนจุดตัดเพิ่มขึ้นด้วยและค่อนข้างสูงกว่าฉบับอื่น ๆ แต่ถ้ามีคลังข้อสอบที่เพียงพอกับระดับความสามารถต่าง ๆ ที่ต้องการวัด ผู้วิจัยคิดว่าน่าจะให้ผลคะแนนจุดตัดที่ใกล้เคียงกัน คือ 60x ของความยาวแบบสอบ ดังเช่น ความยาวแบบสอบ 20 ข้อ ที่ให้ผลประมาณ 60x ทั้ง 4 ฉบับ ถ้าใช้วิธีตามแนวคิดทฤษฎี IRT 3 พารามิเตอร์ ซึ่งจากการวิจัยหาคะแนนจุดตัดของชวลิต โพรธีนคร (ชวลิต โพรธีนคร, 2528) โดยวิธีประยุกต์ของราสส์โมเดล (Rasch Model) พบว่าคะแนนจุดตัดเป็น 50x ของความยาวแบบสอบจากทุกกลุ่มที่ทำการวัดในกลุ่มสูง ปานกลางและกลุ่มต่ำ ซึ่งให้ผลที่ใกล้เคียงกับงานวิจัยครั้งนี้ แต่ที่คะแนนจุดตัดของงานวิจัยครั้งนี้สูงกว่าน่าจะเป็นเพราะ ใช้วิธีการคิดคำนวณโดย 3 พารามิเตอร์ ซึ่งผู้วิจัยคิดว่าแบบสอบเลือกตอบค่าอำนาจจำแนกและค่าการเดา น่าจะมีผลต่อคะแนนจุดตัด การที่จะให้ค่าอำนาจจำแนกและค่าการเดามีค่าเท่ากันทุกข้อนั้นน่าจะขาดหลักการและเหตุผลที่เป็นจริง โดยเฉพาะอย่างยิ่งเป็นแบบสอบชนิดเลือกตอบ เพราะจากการวิเคราะห์ ผลออกมานั้น ค่าพารามิเตอร์ทั้ง 3 คือ ค่าอำนาจจำแนก ค่าความยาก และค่าการเดาไม่เท่ากัน แต่ใกล้เคียงกันและเท่ากันสำหรับค่าการเดาเท่านั้นในบางข้อ และเออร์รี่ (Urry 1970; citing Warm 1978: 20) ได้ให้การสนับสนุนว่า แบบสอบอิงเกณฑ์ชนิดเลือกตอบ โมเดลที่เหมาะสมสำหรับแนวคิดทฤษฎี IRT

คือโมเดลที่ใช้ 3 พารามิเตอร์ ส่วนคะแนนจุดตัดตามแนวคลาสสิก พิจารณาจากคะแนนโดเมนจะให้คะแนนที่คงที่และคะแนนจุดตัดจะเพิ่มขึ้นตามคะแนนโดเมนที่กำหนดสูงขึ้น

4. การหาค่าความน่าจะเป็นในการจำแนกผิดพลาดเพื่อให้ได้กลุ่มผู้รอบรู้และไม่รอบรู้ ผู้วิจัยได้เสนอการอภิปรายผลการวิจัยตามข้อสมมติฐาน แยกตามความยาวแบบสอบดังนี้

4.1 สำหรับความยาวแบบสอบ 20 ข้อ

4.1.1 ผลการวิจัยเกี่ยวกับค่าความน่าจะเป็นในการจำแนกผิดพลาดโดยใช้รูปแบบไบนอมิเยลโมเดลในครั้งนี้ให้ผลขัดแย้งกับสมมติฐานในข้อ 1.1 กล่าวคือ เมื่อกำหนดคะแนนโดเมน .5 ความน่าจะเป็นในการจำแนกผิดพลาดของแบบสอบ ๓ คะแนนจุดตัดเพื่อให้ได้กลุ่มผู้รอบรู้และไม่รอบรู้ตามแนวคิดทฤษฎี IRT น่าจะต่ำกว่าความน่าจะเป็นในการจำแนกผิดพลาดตามแนวคิดทฤษฎี CCT แต่ผลที่ได้ขัดแย้งกับสมมติฐาน การที่ผลเกิดขึ้นดังกล่าวข้างต้นมีสาเหตุมาจากคะแนนจุดตัดที่ได้ตามแนวคิดทฤษฎี IRT และคะแนนโดเมนที่จุดตัดตามทฤษฎี IRT เท่ากับ .6022 ซึ่งสูงกว่าคะแนนโดเมนที่จุดตัดตามทฤษฎี CCT ซึ่งเท่ากับ .5 จึงมีผลให้ค่าความน่าจะเป็นในการจำแนกผิดพลาดของทฤษฎี IRT เท่ากับ .1797 สูงกว่าค่าดังกล่าวของทฤษฎี CCT ซึ่งเท่ากับ .1762 ประมาณ .0035

4.1.2 ผลการวิจัยเกี่ยวกับค่าความน่าจะเป็นในการจำแนกผิดพลาดโดยใช้รูปแบบไบนอมิเยลโมเดลให้ผลขัดแย้งกับสมมติฐานข้อ 1.2 กล่าวคือ เมื่อกำหนดคะแนนโดเมน .6 ความน่าจะเป็นในการจำแนกผิดพลาดของแบบสอบ ๓ คะแนนจุดตัดเพื่อให้ได้กลุ่มผู้รอบรู้และไม่รอบรู้ตามแนวคิดทฤษฎี IRT น่าจะต่ำกว่าความน่าจะเป็นในการจำแนกผิดพลาดตามแนวคิดทฤษฎี CCT แต่ผลที่ได้ขัดแย้งกับสมมติฐานการที่ผลเกิดขึ้นดังกล่าวข้างต้นมีสาเหตุมาจากคะแนนจุดตัดที่ได้ตามแนวคิดทฤษฎี IRT สูงกว่าคะแนนจุดตัดที่ได้ตามแนวคิดทฤษฎี CCT และคะแนนโดเมนที่จุดตัดตามทฤษฎี IRT เท่ากับ .6305 ซึ่งสูงกว่าคะแนนโดเมนที่จุดตัดตามทฤษฎี CCT ซึ่งเท่ากับ .6 จึงมีผลให้ค่าความน่าจะเป็นในการจำแนกผิดพลาดของทฤษฎี IRT เท่ากับ .1814 สูงกว่าค่าดังกล่าวของทฤษฎี CCT ซึ่งเท่ากับ

.1797 ประมาณ .0017 แต่ค่าความน่าจะเป็นดังกล่าวของทฤษฎี IRT สูงกว่าค่าความน่าจะเป็นในการจำแนกผิดพลาดของทฤษฎี CCT ในปริมาณที่น้อยมาก

4.1.3 ผลการวิจัยเกี่ยวกับค่าความน่าจะเป็นในการจำแนกผิดพลาด โดยใช้รูปแบบไบโนเมียลโมเดลให้ผลสอดคล้องกับสมมติฐานในข้อ 1.3 กล่าวคือ เมื่อกำหนดคะแนนโดเมน .7 ความน่าจะเป็นในการจำแนกผิดพลาดของแบบสอบ ๗ คะแนนจุดตัดเพื่อให้ได้กลุ่มผู้รอบรู้และไม่รอบรู้ตามแนวคิดทฤษฎี IRT ต่ำกว่าค่าความน่าจะเป็นไปในการจำแนกผิดพลาดตามแนวคิดทฤษฎี CCT การที่ผลเกิดขึ้นดังกล่าวเป็นผลมาจากคะแนนจุดตัดที่ได้ตามแนวคิดทฤษฎี IRT ต่ำกว่าคะแนนจุดตัดที่ได้ตามแนวคิดทฤษฎี CCT และคะแนนโดเมนที่จุดตัดที่ตามทฤษฎี IRT เท่ากับ .5955 ซึ่งต่ำกว่าคะแนนโดเมนที่จุดตัดตามแนวคิดทฤษฎี CCT ซึ่งเท่ากับ .7 จึงมีผลให้ค่าความน่าจะเป็นในการจำแนกผิดพลาดตามแนวคิดทฤษฎี IRT เท่ากับ .1796 ต่ำกว่าค่าดังกล่าวของทฤษฎี CCT ซึ่งเท่ากับ .1916 ประมาณ .0120

4.1.4 ผลการวิจัยเกี่ยวกับค่าความน่าจะเป็นในการจำแนกผิดพลาด โดยใช้รูปแบบไบโนเมียลโมเดลให้ผลสอดคล้องกับสมมติฐานในข้อ 1.4 กล่าวคือ เมื่อกำหนดคะแนนโดเมน .8 ค่าความน่าจะเป็นในการจำแนกผิดพลาดของแบบสอบ ๗ คะแนนจุดตัดเพื่อให้ได้กลุ่มผู้รอบรู้และไม่รอบรู้ตามแนวคิดทฤษฎี IRT ต่ำกว่าค่าความน่าจะเป็นไปในการจำแนกผิดพลาดตามแนวคิดทฤษฎี CCT การที่ผลเกิดขึ้นดังกล่าวเป็นผลมาจากคะแนนจุดตัดที่ได้ตามแนวคิดทฤษฎี IRT ต่ำกว่าคะแนนจุดตัดที่ได้ตามแนวคิดทฤษฎี CCT และคะแนนโดเมนที่จุดตัดตามแนวคิดทฤษฎี IRT เท่ากับ .6297 ซึ่งต่ำกว่าคะแนนโดเมนที่จุดตัดตามแนวคิดทฤษฎี CCT ซึ่งเท่ากับ .8 จึงมีผลให้ค่าความน่าจะเป็นในการจำแนกผิดพลาดตามแนวคิดทฤษฎี IRT เท่ากับ .1811 ต่ำกว่าค่าดังกล่าวของทฤษฎี CCT ซึ่งเท่ากับ .2182 ประมาณ .0371

4.2 สำหรับความยาวแบบสอบ 30 ข้อ

4.2.1 ผลการวิจัยเกี่ยวกับค่าความน่าจะเป็นในการจำแนกผิดพลาด โดยใช้รูปแบบไบโนเมียลโมเดลในครั้งนี้ให้ผลขัดแย้งกับสมมติฐานในข้อ 2.1 กล่าวคือ เมื่อกำหนดคะแนนโดเมน .5 ความน่าจะเป็นในการจำแนกผิดพลาดของแบบสอบ ๗ คะแนน

จุดตัดเพื่อให้ได้กลุ่มผู้รอบรู้และไม่รอบรู้ตามแนวคิดทฤษฎี IRT น่าจะต่ำกว่าความน่าจะเป็นในการจำแนกผิดพลาดตามแนวคิดทฤษฎี CCT แต่ผลที่ได้ขัดแย้งกับสมมติฐานเป็นเพราะคะแนนจุดตัดที่ได้ตามแนวคิดทฤษฎี IRT สูงกว่าคะแนนจุดตัดที่ได้ตามแนวคิดทฤษฎี CCT และคะแนนโดเมนที่จุดตัดตามทฤษฎี IRT เท่ากับ .6043 ซึ่งสูงกว่าคะแนนโดเมนที่จุดตัดตามทฤษฎี CCT เท่ากับ .1472 สูงกว่าค่าดังกล่าวของทฤษฎี CCT ซึ่งเท่ากับ .1444 ประมาณ .0028

4.2.2 ผลการวิจัยเกี่ยวกับค่าความน่าจะเป็นในการจำแนกผิดพลาด โดยใช้รูปแบบไบโนเมียลโมเดลให้ผลขัดแย้งกับสมมติฐานในข้อ 2.2 กล่าวคือ เมื่อกำหนดคะแนนโดเมน .6 ความน่าจะเป็นในการจำแนกผิดพลาดของแบบสอบ ณ คะแนนจุดตัดเพื่อให้ได้กลุ่มผู้รอบรู้และไม่รอบรู้ตามแนวคิดทฤษฎี IRT น่าจะต่ำกว่าค่าความน่าจะเป็นในการจำแนกผิดพลาดตามแนวคิดทฤษฎี CCT แต่ผลที่ได้ขัดแย้งกับสมมติฐานเป็นเพราะคะแนนจุดตัดที่ได้ตามแนวคิดทฤษฎี IRT สูงกว่าคะแนนจุดตัดที่ได้ตามแนวคิดทฤษฎี CCT และคะแนนโดเมนที่จุดตัดตามทฤษฎี IRT เท่ากับ .6214 สูงกว่าคะแนนโดเมนที่จุดตัดตามทฤษฎี CCT ซึ่งเท่ากับ .6 จึงมีผลทำให้ค่าความน่าจะเป็นในการจำแนกผิดพลาดของทฤษฎี IRT เท่ากับ .1484 สูงกว่าค่าดังกล่าวของทฤษฎี CCT ซึ่งเท่ากับ .1474 ประมาณ .001 แต่ที่สูงกว่าในปริมาณที่น้อยมาก

4.2.3 ผลการวิจัยเกี่ยวกับค่าความน่าจะเป็นในการจำแนกผิดพลาด โดยใช้รูปแบบไบโนเมียลโมเดลให้ผลสอดคล้องกับสมมติฐานในข้อ 2.3 กล่าวคือ เมื่อกำหนดคะแนนโดเมน .7 ค่าความน่าจะเป็นในการจำแนกผิดพลาดของแบบสอบ ณ คะแนนจุดตัดเพื่อให้ได้กลุ่มผู้รอบรู้และไม่รอบรู้ตามแนวคิดทฤษฎี IRT ต่ำกว่าค่าความน่าจะเป็นในการจำแนกผิดพลาดตามแนวคิดทฤษฎี CCT การที่ผลเกิดขึ้นดังกล่าวเป็นผลมาจากคะแนนจุดตัดที่ได้ตามแนวคิดทฤษฎี IRT ต่ำกว่าคะแนนจุดตัดที่ได้ตามแนวคิดทฤษฎี CCT และคะแนนโดเมนที่จุดตัดตามแนวคิดทฤษฎี IRT เท่ากับ .5421 ต่ำกว่าคะแนนโดเมนที่จุดตัดตามแนวคิดทฤษฎี CCT ซึ่งเท่ากับ .7 จึงมีผลให้ค่าความน่าจะเป็นในการจำแนกผิดพลาดตามแนวคิดทฤษฎี IRT เท่ากับ .1444 ต่ำกว่าค่าดังกล่าวตามแนวคิดทฤษฎี CCT ซึ่งเท่ากับ .1573 ประมาณ .0129

4.2.4 ผลการวิจัยเกี่ยวกับค่าความน่าจะเป็นในการจำแนกผิดพลาด โดยใช้รูปแบบไบโนเมียลโมเดลให้ผลสอดคล้องกับสมมติฐานในข้อ 2.4 กล่าวคือ เมื่อ กำหนดคะแนนโดเมน .8 ค่าความน่าจะเป็นในการจำแนกผิดพลาดของแบบสอบ ๗ คะแนน จุดตัดเพื่อได้กลุ่มผู้รอบรู้และไม่รอบรู้ตามแนวคิดทฤษฎี IRT ต่ำกว่าค่าความน่าจะเป็นในการ จำแนกผิดพลาดตามแนวคิดทฤษฎี CCT การที่ผลเกิดขึ้นดังกล่าวเป็นผลมาจากคะแนนจุดตัดที่ได้ ตามแนวคิดทฤษฎี IRT ต่ำกว่าคะแนนจุดตัดที่ได้ตามแนวคิดทฤษฎี CCT และคะแนนโดเมนที่จุดตัด ตามแนวคิดทฤษฎี IRT เท่ากับ .6368 ต่ำกว่าคะแนนโดเมนที่จุดตัดตามแนวคิดทฤษฎี CCT ซึ่งเท่ากับ .8 จึงมีผลให้ค่าความน่าจะเป็นในการจำแนกผิดพลาดตามแนวคิดทฤษฎี IRT เท่ากับ .1496 ต่ำกว่าค่าดังกล่าวตามแนวคิดทฤษฎี CCT ซึ่งเท่ากับ .1795 ประมาณ .0299

เป็นที่น่าสังเกตว่าค่าความน่าจะเป็นในการจำแนกผิดพลาดสำหรับการ คิดคำนวณที่ใช้คะแนนโดเมนและคะแนนจุดตัดที่คำนวณตามแนวคิดทฤษฎี CCT โดยใช้รูปแบบ ไบโนเมียลโมเดลจะให้ค่าคงที่ ทั้งนี้เนื่องจากคะแนนโดเมนและคะแนนจุดตัดที่ได้ตามแนวคิด ทฤษฎี CCT ให้ค่าคงที่เป็นสัดส่วนโดยตรงกัน จึงมีผลทำให้ค่าความน่าจะเป็นในการจำแนก ความรอบรู้ผิดพลาดมีค่าคงที่ ในขณะที่คะแนนโดเมนและคะแนนจุดตัดตามแนวคิดทฤษฎี IRT มีค่าแปรเปลี่ยนไปตามคะแนนโดเมนที่กำหนด จึงทำให้ค่าความน่าจะเป็นในการจำแนก ผิดพลาดแปรเปลี่ยนไปด้วย สำหรับแบบสอบความยาว 20 ข้อ เมื่อกำหนดคะแนนโดเมน .5, .6, .7 และ .8 จะให้ค่าความน่าจะเป็นในการจำแนกความรอบรู้ผิดพลาดตามทฤษฎี IRT ประมาณ 18% และความยาวแบบสอบ 30 ข้อ ให้ค่าความน่าจะเป็นในการจำแนกความ รอบรู้ผิดพลาดประมาณ 15% ซึ่งตามวิธีของวิลคอกซ์ (Wilcox 1976; cited by Hambleton and De Gruijter 1983: 361) กำหนดให้ค่าดังกล่าวน้อยกว่า .2 (20%) ซึ่งผลการวิจัยค่าความน่าจะเป็นในการจำแนกความรอบรู้ผิดพลาดตามแนวคิดทฤษฎี IRT ให้ค่าน้อยกว่าเกณฑ์ที่กำหนด คือ 20 % แสดงว่าผลการวิจัยค่าดังกล่าวตามแนวคิดทฤษฎี IRT น่าจะสมเหตุสมผลมากกว่าวิธีตามแนวคิดทฤษฎี CCT ที่ให้ค่าความน่าจะเป็นในการจำแนก ผิดพลาดสูงขึ้นตามการกำหนดคะแนนโดเมนที่เพิ่มขึ้นและคะแนนจุดตัดที่สูงเพิ่มขึ้นตามไปด้วย โดยไม่คำนึงถึงว่าแบบสอบนั้นจะยากหรือง่าย และจากผลการวิจัยจะเห็นว่าคะแนนจุดตัดที่ได้ ตามแนวคิดทฤษฎี IRT ค่อนข้างเบี่ยงเบนมาที่คะแนน 60 % ของความยาวแบบสอบ

5. การหาค่าความเที่ยงในการตัดสินใจในการจำแนกความรอบรู้ ผู้วิจัยเสนอการอภิปรายผลการวิจัยตามข้อสมมติฐานแยกตามความยาวแบบสอบ ดังนี้

5.1 สำหรับความยาวแบบสอบ 20 ข้อ

ผลการวิจัยเกี่ยวกับค่าความเที่ยงในการตัดสินใจที่คะแนนจุดตัดของแบบสอบฉบับวัดคะแนนโดเมน .5, .6, .7 และ .8 ตามลำดับให้ผลสอดคล้องกับสมมติฐานข้อที่ 3.1, 3.2, 3.3 และ 3.4 ตามลำดับ กล่าวคือ การหาค่าความเที่ยงในการตัดสินใจที่คะแนนจุดตัดตามแนวทฤษฎี IRT สูงกว่าการหาค่าความเที่ยงในการตัดสินใจที่คะแนนจุดตัดตามแนวทฤษฎี CCT ทั้งนี้เพราะว่าวิธีการหาค่าความเที่ยงตามแนวทฤษฎี IRT นิยามจากฟังก์ชันสารสนเทศของแบบสอบ $I(\theta_0)$ ซึ่งเป็นดัชนีชี้บอกคุณภาพของข้อสอบ ถ้า $I(\theta_0)$ มีค่าสูงมากเพียงใดก็แสดงว่า แบบสอบชุดนั้นมีคุณภาพชี้บอกผู้รอบรู้และไม่รอบรู้ได้ดีมากที่ $\theta = \theta_0$ (Van der Linden 1981: 393) จึงมีผลทำให้ค่าความเที่ยงประยุกต์ของวอร์มมีค่าสูงถึง .9295, .9377, .9410 และ .9435 สำหรับแบบสอบฉบับวัดคะแนนโดเมน .5, .6, .7 และ .8 ตามลำดับ ในขณะที่วิธีการหาค่าความเที่ยงของฮวินตามทฤษฎี CCT ใช้สัมประสิทธิ์แคปป่าเป็นดัชนีประมาณค่าความเที่ยงในการจำแนกความรอบรู้ ซึ่งได้จากสัดส่วนความน่าจะเป็นในการจำแนกความรอบรู้ โดยให้ผลค่าความเที่ยงเป็น .5509, .6389, .5326 และ .5816 สำหรับแบบสอบฉบับวัดคะแนนโดเมน .5, .6, .7 และ .8 ตามลำดับ ซึ่งวิธีประมาณค่าความเที่ยงของฮวินนี้จะให้ค่าความเที่ยงประมาณ 53-64% สำหรับความยาวแบบสอบ 20 ข้อ และค่าความเที่ยงวิธีของฮวินจะมีค่าสูง ฉบับที่วัดคะแนนโดเมน .6 ทั้งนี้เพราะแบบสอบฉบับวัดคะแนนโดเมน .6 นี้มีคะแนนจุดตัดคือ 12 ใกล้ค่าเฉลี่ยของแบบสอบซึ่งเท่ากับ 13 ซึ่งค่าความเที่ยงของแบบสอบอิงเกณฑ์ที่ใช้สูตรของฮวินจะมีค่าสูงสุดที่คะแนนจุดตัดอยู่ใกล้คะแนนเฉลี่ย และค่าความเที่ยงจะมีค่าต่ำเมื่อคะแนนจุดตัดอยู่ใกล้คะแนนศูนย์หรือคะแนนเต็ม นอกจากนี้ค่าความเที่ยงของแบบสอบจะมีค่าต่ำเป็นเพราะจำนวนข้อสอบมีน้อยข้อด้วย (Huynh 1976: 260)

5.2 สำหรับความยาวแบบสอบ 30 ข้อ

ผลการวิจัยเกี่ยวกับค่าความเที่ยงในการตัดสินใจที่คะแนนจุดตัดของแบบสอบฉบับวัดคะแนนโดเมน .5, .6, .7 และ .8 ให้ผลสอดคล้องกับสมมติฐานข้อที่ 4.1, 4.2, 4.3 และ 4.4 ตามลำดับ กล่าวคือ การหาค่าความเที่ยงในการตัดสินใจที่คะแนนจุดตัดตามแนวทางทฤษฎี IRT สูงกว่าการหาค่าความเที่ยงในการตัดสินใจที่คะแนนจุดตัดตามแนวทางทฤษฎี CCT ทั้งนี้เพราะว่าวิธีการหาค่าความเที่ยงตามแนวทางทฤษฎี IRT นิยามมาจากฟังก์ชันสารสนเทศของแบบสอบ $[I(\theta_0)]$ ซึ่งเป็นดัชนีชี้บอกคุณภาพของข้อสอบ ถ้า $I(\theta_0)$ มีค่าสูงมากเพียงใดก็แสดงว่า แบบสอบชุดนั้นมีคุณภาพที่บอกผู้รอบรู้และไม่รอบรู้ได้ดีมากที่ $\theta = \theta_0$ (Van der Linden 1981: 393) จึงมีผลทำให้ค่าความเที่ยงประยุกต์ของวอร์มมีค่าสูงถึง .9501, .9554, .9570 และ .9607 สำหรับแบบสอบฉบับวัดคะแนนโดเมน .5, .6, .7 และ .8 ตามลำดับ ในขณะที่วิธีการหาค่าความเที่ยงของฮวินตามทฤษฎี CCT ใช้สัมประสิทธิ์แคปป่าเป็นดัชนีประมาณค่าความเที่ยงในการจำแนกความรอบรู้ ซึ่งได้จากสัดส่วนความน่าจะเป็นในการจำแนกความรอบรู้ โดยให้ผลค่าความเที่ยงเป็น .6372, .6521, .5735 และ .5495 สำหรับแบบสอบฉบับวัดคะแนนโดเมน .5, .6, .7 และ .8 ตามลำดับ ซึ่งวิธีประมาณค่าความเที่ยงของฮวินสำหรับความยาวแบบสอบ 30 ข้อนี้จะให้ค่าความเที่ยงประมาณ 55-65% และค่าความเที่ยงวิธีของฮวินจะมีค่าสูงฉบับที่วัดคะแนนโดเมน .6 ทั้งนี้เพราะค่าความเที่ยงของแบบสอบอิงเกณฑ์ที่ใช้วิธีของฮวินจะมีค่าสูงสุดที่คะแนนจุดตัดอยู่ใกล้ค่าเฉลี่ย (Huynh 1976: 260) ซึ่งแบบสอบวัดคะแนนโดเมน .6 ตามแนวทางทฤษฎี CCT นี้มีคะแนนจุดตัดเท่ากับ 18 คะแนนซึ่งมีค่าใกล้คะแนนเฉลี่ยซึ่งเท่ากับ 20 จึงทำให้ค่าความเที่ยงของแบบสอบฉบับวัดคะแนนโดเมน .6 มีค่าสูงกว่าแบบสอบฉบับอื่น ๆ

อย่างไรก็ดีในการประมาณค่าความเที่ยงตามแนวทางทฤษฎี CCT เป็นการประมาณค่าความเที่ยงซึ่งเป็นความคลาดเคลื่อนเฉลี่ยของคนในกลุ่มที่วัด แต่ถือว่าเป็นความคลาดเคลื่อนของการวัดของแต่ละคน โดยถือเสมือนว่าเท่ากับที่สัมพันธ์กับความคลาดเคลื่อนมาตรฐานของการวัด (Standard error of Measurement SEM) ส่วนการประมาณค่าความเที่ยงตามแนวทางทฤษฎี IRT ประยุกต์ของวอร์ม เป็นการประมาณค่าความเที่ยงซึ่งเป็นความคลาดเคลื่อน

ของคนในกลุ่มที่วัดซึ่งถือว่าผู้มีความสามารถตรงที่วัดจะมีความคลาดเคลื่อนที่สัมพันธ์กับความคลาดเคลื่อนมาตรฐานของการประมาณ (Standard error of Estimate SEE) ดังนั้นการเปรียบเทียบความเที่ยงจึงไม่สามารถทดสอบความแตกต่างได้โดยตรง

จากผลการวิจัย เมื่อคะแนนโดเมนเท่ากันผลการหาคะแนนจุดตัดตามแนวทฤษฎีการตอบสนองข้อกระทงและแนวทฤษฎีคลาสสิกจะมีจำนวนข้อต่างกัน และเมื่อกำหนดจุดตัดตามคะแนนโดเมนต่าง ๆ ที่ต้องการวัดวิธีการตามทฤษฎีการตอบสนองข้อกระทงมีคะแนนที่ถือว่าเป็นผู้รอบรู้ต่ำกว่าของวิธีตามทฤษฎีคลาสสิก ซึ่งต้องใช้จำนวนข้อผ่านมากกว่า จึงทำให้เห็นว่าการใช้ทฤษฎีการตอบสนองข้อกระทงในการหาคะแนนจุดตัดน่าจะให้ความแม่นยำ ถูกต้อง เป็นธรรมและมีความเหมาะสมในการปฏิบัติมากกว่าจะใช้วิธีการหาคะแนนจุดตัดตามแนวทฤษฎีคลาสสิก ซึ่งจำเป็นต้องใช้จำนวนข้อรอบรู้มากกว่า เมื่อคะแนนโดเมนสูงขึ้นซึ่งทำให้จำแนกผู้รอบรู้ผิดในทางบวกที่เรียก Type I Error

การวิจัยครั้งนี้พบว่าการสร้างข้อกระทงในวิชาคณิตศาสตร์ซึ่งเป็นแบบสอบวัดความสามารถ (Ability Test) ให้กลุ่มโดเมนเนื้อหาที่กำหนดตามรูปแบบการสร้างรูปกระทงให้ได้จำนวนมาก สามารถวิเคราะห์ข้อกระทงด้วยโปรแกรมโลจิสติก 5 ได้ และถ้ามีข้อกระทงที่วัดแต่ละระดับความสามารถต่าง ๆ ในแต่ละจุดประสงค์เพียงพอกับความยาวแบบสอบ 20 และ 30 ข้อ ตามที่กำหนดในงานวิจัย กรณีการเลือกข้อกระทงตามทฤษฎี IRT ที่สัมพันธ์กับค่าความสามารถ (θ_0) ที่ต้องการวัดและสอดคล้องกับการกำหนดคะแนนโดเมน จะเป็นวิธีการเลือกข้อกระทงที่เหมาะสมในการนำมาหาคะแนนจุดตัดตามแนวทฤษฎี IRT เพราะหาผู้ที่ได้คะแนนสอบมากกว่าคะแนนจุดตัดที่หาได้ตามแนวทฤษฎี IRT สามารถประมาณการอ้างอิงว่าผู้สอบมีความรู้ในเนื้อหาที่ต้องการวัดในระดับที่ต้องการได้ดี และใช้จำนวนข้อผ่านความรอบรู้ที่สมเหตุสมผลกว่าวิธีการหาคะแนนจุดตัดตามแนวทฤษฎีคลาสสิกที่ใช้จำนวนข้อผ่านความรอบรู้มากขึ้นเป็นสัดส่วนตามที่กำหนดคะแนนโดเมน โดยไม่คำนึงว่าแบบสอบจะยากง่ายเพียงใด ซึ่งในความเป็นจริงถ้าแบบสอบง่ายจำนวนข้อผ่านความรอบรู้ควรใช้เกณฑ์ผ่านสูง และถ้าแบบสอบยากจำนวนข้อผ่านควรกำหนดให้เกณฑ์ต่ำ

ข้อเสนอแนะในการวิจัย

จากผลการวิจัย และข้อค้นพบ ผู้วิจัยจึงขอเสนอแนะการวิจัย ดังนี้

1. ควรจะสร้างคลังข้อสอบให้มากกว่างานวิจัยในครั้งนี เพื่อจะได้มีข้อกระทบ สำหรับได้เลือกมากพอที่จะให้โค้งลักษณะข้อสอบ (Test Characteristic Curve) ที่พอเพียง และใกล้เคียงกับค่าความสามารถ (θ_0) ที่ต้องการวัด
2. ในการคัดเลือกข้อกระทบควรใช้วิธีการเลือกโดยใช้ทฤษฎี IRT 3 พารามิเตอร์ วิเคราะห์ เพราะจะให้ Test Information curve สูง ณ ระดับความสามารถที่ต้องการวัด ซึ่งจะทำให้จำแนกผู้สอบเป็นผู้รอบรู้และไม่รอบรู้ที่ระดับความสามารถนั้น
3. เพื่อใช้เป็นแนวทางให้หน่วยงานทางการศึกษานำเอาวิชาและเรื่องอื่น ๆ ที่ เกี่ยวข้องกับการวิจัยทำเป็นธนาคารข้อสอบที่มีประสิทธิภาพ ในการวัดระดับความสามารถต่าง ๆ

งานวิจัยครั้งต่อไป

1. น่าจะมีการหาคะแนนจุดตัดด้วยวิธีการอื่นมาประยุกต์ใช้ แต่ยังใช้การคัดเลือก ข้อกระทบตามทฤษฎี IRT เพราะข้อกระทบที่ให้ Test Information curve สูงสุดที่ระดับ ความสามารถที่ต้องการวัดจะสามารถจำแนกระดับความสามารถของผู้สอบได้ดี ณ ระดับ ความสามารถนั้น (Hambleton 1977: 66)
2. ให้ใช้การวิเคราะห์ข้อกระทบด้วยโมเดลโลจิสติก 1 พารามิเตอร์ 2 พารา 2 พารามิเตอร์ และ 3 พารามิเตอร์ ในการคัดเลือกข้อกระทบ แล้วมาคำนวณหาคะแนนจุด ตัดตามแนวคิดวิธีการประยุกต์ของวอร์ม เพื่อจะได้ดูว่าให้ผลต่างกันมากเพียงใดในการศึกษา ทางทฤษฎี
3. ในการสร้างคลังข้อสอบให้ใช้แนวคิดการวิเคราะห์เช่นเดียวกับงานวิจัยครั้งนี้ แต่น่าจะใช้วิธีการเลือกข้อกระทบอย่างสุ่มและเลือกข้อกระทบเช่นเดียวกับงานวิจัย แล้วใช้ วิธีการคิดคำนวณคะแนนจุดตัดด้วยวิธีการอื่น ๆ เพื่อจะดูว่าคะแนนจุดตัดที่ได้แตกต่างจากงาน วิจัยอย่างไร

4. น่าจะทำการวิจัยซ้ำอีก แต่ให้เพิ่มคลังข้อสอบให้มากกว่างานวิจัยครั้งนี้ เพื่อจะได้ข้อกระทบที่ให้ Item Information Curve สูง ๆ ใกล้เคียงกับความสามารถที่ต้องการวัดได้เพียงพอ จะได้พิจารณาว่าผลที่ได้แตกต่างจากงานวิจัยครั้งนี้หรือไม่
5. น่าจะนำผลการวิจัยครั้งนี้เทียบกับผลการสอบวิชาคณิตศาสตร์ของนักเรียนในโรงเรียน เพื่อพิจารณาว่าจะมีความสอดคล้องกันหรือไม่ หรือเลือกใช้แบบสอบฉบับวัดคะแนนโดเมนที่ต้องการวัด เพื่อใช้เป็นแนวทางในการทำนายผลการเรียนของนักเรียนและการวินิจฉัยนักเรียนต่อไป
6. ควรทำการวิจัยในลักษณะเช่นนี้กับวิชาอื่น โดยกำหนดความยาวแบบสอบให้สอดคล้องกับเนื้อหาที่ต้องการวัด ซึ่งอาจจะเป็นความยาว 40, 50 หรือ 60 ข้อ เป็นต้น แล้วแต่ความเหมาะสม