

## บทที่ 5

### สรุปผลการวิจัย ทบทวนรายผล และข้อเสนอนะ

การวิจัยครั้งนี้มีจุดมุ่งหมายเพื่อสร้างแบบสอบอิงโดเมนวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 เรื่องพหุนาม และใช้โมเดลโลจิสติก 3 พารามิเตอร์ในการวิเคราะห์แบบสอบที่สร้างขึ้น เพื่อหาค่าพารามิเตอร์ของข้อสอบ ได้แก่ ค่าอำนาจจำแนก ค่าความยาก และค่าการเดา และการคัดเลือกข้อสอบตามเกณฑ์ที่เหมาะสมของค่าพารามิเตอร์ พร้อมทั้งตรวจสอบคุณภาพของแบบสอบในด้านความตรงและความเที่ยงตามแนวของแบบสอบอิงโดเมน และ IRT นอกจากนี้ยังมุ่งใช้โมเดลโลจิสติก 3 พารามิเตอร์ เพื่อการประมาณค่าความสามารถและคะแนนโดเมนของผู้สอบ ตลอดจนเพื่อการกำหนดคะแนนเกณฑ์ของแบบสอบอิงโดเมนที่สร้างขึ้น

กลุ่มตัวอย่างที่ใช้ในการศึกษาค้นคว้าครั้งนี้ ได้แก่ นักเรียนชั้นมัธยมศึกษาปีที่ 3 ภาคเรียนที่ 1 ปีการศึกษา 2531 ของโรงเรียนสังกัดกรมสามัญศึกษา กระทรวงศึกษาธิการ ในจังหวัดศรีสะเกษ จำนวน 2,145 คน ซึ่งได้จากการสุ่มแบบสองขั้นตอน และในแต่ละชั้นใช้วิธีการสุ่มตัวอย่างง่ายจากประชากรนักเรียนจำนวน 4,531 คน

เครื่องมือที่ใช้ในการวิจัย ได้แก่ แบบสอบอิงโดเมนวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 เรื่อง พหุนาม ซึ่งผู้วิจัยได้สร้างขึ้น โดยมุ่งวัดตามกลุ่มพฤติกรรมใหญ่ (Domain of Behaviour) 4 กลุ่ม และกลุ่มพฤติกรรมย่อย (Sub Domain) 22 กลุ่ม ซึ่งได้จากการวิเคราะห์เนื้อหาเกี่ยวกับจุดประสงค์เฉพาะวิชา และพฤติกรรมด้านพุทธิพิสัยวิชาคณิตศาสตร์ ในขั้นการคิดคำนวณของเจมส์ ดับบลิว วิตสัน ข้อสอบแต่ละข้อสร้างจากลักษณะเฉพาะของข้อสอบ (Item Specification) ซึ่งกำหนดขึ้นตามสลับโดเมนของแต่ละโดเมน แบบสอบเป็นชนิดเลือกตอบ 4 ตัวเลือก จำนวน 2 ฉบับ ที่วัดกลุ่มพฤติกรรมเดียวกัน ฉบับละ 50 ข้อ

ในการเก็บรวบรวมข้อมูลและการวิเคราะห์นั้น ผู้วิจัยนำแบบสอบอิงโดเมน 2 ฉบับที่สร้างขึ้น ทำการทดลองสอบ 2 ครั้ง กับกลุ่มตัวอย่างทั้งสิ้น 525 คน เพื่อการปรับปรุงแก้ไขข้อสอบที่บกพร่องและเพื่อศึกษาปัญหาต่าง ๆ ในการสอบ ตลอดจนเพื่อการกำหนดเวลาที่เหมาะสม

จากนั้นนำแบบสอบทั้ง 2 ฉบับ ที่ได้ปรับปรุงแล้วไปทดสอบกับกลุ่มตัวอย่างในการวิจัยจำนวน 2,025 คน นำข้อมูลที่ได้มาวิเคราะห์ด้วยประกอบเพื่อ ตรวจสอบคุณสมบัติ การวัดเพียงมิติเดียว และความเป็นอิสระในการตอบข้อสอบ วิเคราะห์ข้อสอบด้วยโปรแกรม LOGIST 5 Version 2.5 เพื่อหาค่าพารามิเตอร์ของข้อสอบและคัดเลือกข้อสอบไว้เป็นกลุ่มข้อสอบ (Item Pool) ตามเกณฑ์ที่เหมาะสมของค่าพารามิเตอร์ จากนั้นทำการสุ่มข้อสอบจากกลุ่มข้อสอบแต่ละโดเมน โดยวิธีสุ่มอย่างง่าย ได้แบบสอบอิงโดเมนฉบับสมบูรณ์ 1 ฉบับ จำนวน 50 ข้อ นำแบบสอบอิงโดเมนฉบับสมบูรณ์ไปทดสอบกับกลุ่มตัวอย่างในการวิจัย จำนวน 120 คน ตรวจสอบคุณภาพแบบสอบทางด้านความเที่ยงในรูปอิงโดเมนโดยการคำนวณค่าดัชนีความเชื่อถือได้ตามสูตรของ เบรนนัน และ เคน และหาค่าความเที่ยงในรูปแบบของ IRT จากการคำนวณค่าฟังก์ชันสารสนเทศ แบบสอบ การตรวจสอบคุณภาพทางด้านความตรงเชิงเนื้อหาใช้วิธีของโรวิเนลลี และ แสมเบิลตัน ความตรงเชิงโครงสร้างได้จากการวิเคราะห์ตัวประกอบดูความสอดคล้องของข้อสอบที่ถูกจัดเข้าไว้ในตัวประกอบเดียวกันกับข้อสอบที่อยู่ในโดเมนเดียวกัน ส่วนความตรงตามสภาพได้จากการคำนวณค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนของแบบสอบกับ เกรควิชาคณิตศาสตร์ (ก 311) ในภาคเรียนที่ 1 ของนักเรียนแต่ละคนและใช้โมเดลโลจิสติก 3 พารามิเตอร์ กำหนดคะแนนเกณฑ์ของแบบสอบแต่ละโดเมนและทั้งฉบับที่ได้จากการพิจารณาตัดสินของผู้เชี่ยวชาญตามวิธีของ แองกอฟให้อยู่บนสเกลความสามารถ และประมาณค่าความสามารถของผู้สอบ ตลอดจนคะแนนโดเมนในระดับความสามารถต่าง ๆ ในช่วง -5 ถึง +5

### สรุปผลการวิจัย

การวิจัยครั้งนี้มีจุดมุ่งหมายเพื่อให้ได้แบบสอบอิงโดเมนวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 เรื่องพหุนามที่มีคุณภาพจำนวน 1 ฉบับ โดยให้มีคะแนนเกณฑ์เพื่อการตัดสินระดับความรอบรู้และมีแนวทางในการประมาณค่าความสามารถ ตลอดจนคะแนนโดเมนของผู้สอบตามแนวของโมเดลโลจิสติก 3 พารามิเตอร์ ซึ่งสามารถสรุปผลการวิจัยได้ดังนี้

1. ผลการใช้โมเดลโลจิสติก 3 พารามิเตอร์ ในการตรวจสอบคุณสมบัติและการวิเคราะห์แบบสอบอิงโดเมน

### 1.1 ผลการตรวจสอบคุณสมบัติการวัดเพียงลักษณะเดียวและความเป็นอิสระ

#### ในการตอบ

จากการนำผลการสอบแบบสอบอิงโดเมนฉบับที่ 1 และ 2 มาทำการวิเคราะห์ที่ตัวประกอบ เพื่อตรวจสอบคุณสมบัติการวัดเพียงคุณลักษณะเดียว และความเป็นอิสระในการตอบข้อสอบ พบว่า แบบสอบฉบับที่ 1 และ 2 มีตัวประกอบที่มีค่าไอเกนมากกว่า 1 อยู่ 12 และ 13 ตัวประกอบตามลำดับ โดยที่ค่าไอเกนของตัวประกอบที่ 1 มายังตัวประกอบที่ 2 มีค่าลดลงอย่างรวดเร็ว ในขณะที่ค่าไอเกนของตัวประกอบที่ 2 มายังตัวประกอบลำดับถัด ๆ ไป จนถึงตัวประกอบตัวสุดท้ายจะลดลงที่ละน้อย ซึ่งจากลักษณะดังกล่าวจึงสามารถเชื่อถือได้ว่าแบบสอบทั้งสองฉบับมีคุณสมบัติดังกล่าว ดังนั้นจึงสามารถนำมาวิเคราะห์ข้อสอบโดยโปรแกรม LOGIST5 version 2.5 ตามข้อตกลงของโมเดลโลจิสติกได้ และจะถือว่าแบบสอบอิงโดเมนฉบับสมบูรณ์มีคุณสมบัติดังกล่าวนี้ด้วย เนื่องจากข้อสอบได้มาโดยการสุ่มจากกลุ่มข้อสอบของแบบสอบฉบับที่ 1 และ 2

### 1.2 ผลการวิเคราะห์และคัดเลือกข้อสอบตามเกณฑ์ที่เหมาะสมของค่าพารามิเตอร์

จากการวิเคราะห์ข้อสอบของแบบสอบอิงโดเมน ฉบับที่ 1 และ 2 ด้วยโปรแกรม LOGIST 5 พบว่า จำนวนข้อสอบที่สามารถนำมาวิเคราะห์ตามรูปแบบของโมเดลนี้มีทั้งสิ้น 97 ข้อ อีก 3 ข้อ ตัดทิ้งไป เนื่องจากรูปแบบการตอบไม่เหมาะสมกับโปรแกรม และเมื่อทำการคัดเลือกข้อสอบเพื่อใช้เป็นกลุ่มข้อสอบ โดยถือเกณฑ์ที่เหมาะสม คือ  $a > 0.8$   $-2.5 < b < 2.5$  และ  $c < 0.3$  ทั้งนี้โดยพิจารณาถึงความครอบคลุมเนื้อหาตามที่ได้กำหนดน้ำหนักความสำคัญไว้ด้วย จะได้ข้อสอบที่ดีตามเกณฑ์ จำนวน 9 ข้อ และทำการสุ่มข้อสอบในแต่ละโดเมนให้ได้แบบสอบอิงโดเมนฉบับสมบูรณ์ จำนวน 50 ข้อ ซึ่งได้ค่าอำนาจจำแนกอยู่ในช่วงตั้งแต่ 0.82293 ถึง 2.00000 มีค่าเฉลี่ยเป็น 1.27235 ค่าความยากอยู่ในช่วงตั้งแต่ -0.79607 ถึง 2.32 มีค่าเฉลี่ยเป็น 0.79895 และค่าการเดาอยู่ในช่วงตั้งแต่ 0.03223 ถึง 0.29393 โดยมีค่าเฉลี่ยเป็น 0.18271

### 1.3 ผลการตรวจสอบความสอดคล้องในการประมาณค่าความสามารถของ

#### แบบสอบ

จากการคำนวณค่าฟังก์ชันสารสนเทศแบบสอบของแบบอิงโดเมนฉบับที่ 1, 2 และฉบับสมบูรณ์ที่ระดับความสามารถต่าง ๆ ในช่วง -3 ถึง +3 พบว่า ค่าสารสนเทศ

แบบสอบของแบบสอบอิงโดเมนฉบับที่ 1, 2 และ ฉบับสมบูรณ์ มีค่าสูงสุดใกล้เคียงกัน คือเท่ากับ 27.86401, 28.79131 และ 30.55542 ตามลำดับ ที่ระดับความสามารถ 0.8 เท่ากัน และที่ระดับความสามารถดังกล่าว พบว่ามีค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถเท่ากับ 0.18944, 0.18637 และ 0.18091 ตามลำดับ ซึ่งสามารถอธิบายตามแนวของ IRT ได้ว่า แบบสอบทั้ง 3 ฉบับนี้ มีความเหมาะสมหรือมีประสิทธิภาพสูงสุดในการวัดผู้สอบที่มีระดับความสามารถปานกลางและค่อนข้างสูงเหมือนกัน ทั้งนี้อาจจะเป็น เพราะแบบสอบฉบับที่ 1 และ 2 สร้างมาจากลักษณะเฉพาะข้อสอบในโดเมนเดียวกัน และแบบสอบฉบับสมบูรณ์ก็ประกอบด้วยข้อสอบที่ดี ซึ่งสุ่มมาจากกลุ่มข้อสอบของแบบสอบฉบับที่ 1 และ 2 ที่ได้คัดเลือกไว้ ตามเกณฑ์ที่เหมาะสมของค่าพารามิเตอร์ในแต่ละโดเมน ซึ่งในการสุ่มได้พิจารณาถึงความครอบคลุมเนื้อหาตามที่ได้กำหนดน้ำหนักความสำคัญไว้จากข้อ 2.3 ด้วย ดังนั้นแบบสอบอิงโดเมนฉบับสมบูรณ์ที่ได้จึงมีความสอดคล้องกับแบบสอบอิงโดเมนฉบับที่ 1 และ 2 ที่ได้สร้างไว้ในครั้งแรกทั้งในแง่ของความครอบคลุมสัปดาห์โดเมน น้ำหนักความสำคัญของสัปดาห์โดเมน ซึ่งส่งผลให้แบบสอบทั้งสามฉบับมีความสอดคล้องกันในการประมาณค่าความสามารถและยังมีผลให้แบบสอบฉบับสมบูรณ์มีค่าฟังก์ชันสารสนเทศแบบสอบสูงสุด กล่าวคือ มีประสิทธิภาพสูงสุดในการวัดผู้สอบที่มีระดับความสามารถดังกล่าวด้วย

## 2. ผลการตรวจสอบคุณภาพของแบบสอบอิงโดเมน

### 2.1 ผลการตรวจสอบความตรงของแบบสอบ

#### 2.1.1 ความตรงเชิงเนื้อหา จากการพิจารณาตัดสินโดยผู้เชี่ยวชาญ

ตามวิธีการของโรวีเนลลี และ แสมเบิลตัน พบว่า ค่าเฉลี่ยของคะแนนการตัดสินความครอบคลุมของสัปดาห์โดเมนในแต่ละโดเมน และค่าเฉลี่ยของคะแนนการตัดสินความสอดคล้องระหว่างลักษณะเฉพาะของข้อสอบกับสัปดาห์โดเมน ตลอดจนค่าเฉลี่ยของคะแนนการตัดสินความสอดคล้องระหว่างข้อสอบกับลักษณะเฉพาะของข้อสอบในแต่ละสัปดาห์โดเมน มีค่าสูงเกินเกณฑ์ 0.5 ดังนั้นจากผลการตัดสินทั้งสามขั้นตอนนี้จึงสามารถสรุปได้ว่า แบบสอบมีความตรงเชิงเนื้อหา

#### 2.1.2 ความตรงเชิงโครงสร้าง จากการวิเคราะห์ตัวประกอบพบว่า

ข้อสอบทั้ง 50 ข้อ สามารถจัดเข้าเป็นตัวประกอบที่มีค่าไอเกน ตั้งแต่ 1.00 ขึ้นไปได้ 13 ตัวประกอบ และเมื่อเลือกเฉพาะตัวประกอบที่ประกอบด้วยข้อสอบอย่างน้อย 3 ข้อ และแต่ละข้อมีค่าน้ำหนักตัวประกอบตั้งแต่ 0.5 ขึ้นไป จะได้ตัวประกอบ 5 ตัวประกอบ โดยที่ข้อสอบในโดเมนที่ 1 ถูกจัดไว้ในตัวประกอบที่ 1 ข้อสอบในโดเมนที่ 2 ถูกจัดไว้ในตัวประกอบที่ 2 และ 5 ข้อสอบในโดเมนที่ 3 ถูกจัดไว้ในตัวประกอบที่ 3 และข้อสอบในโดเมนที่ 4 ถูกจัดไว้ในตัวประกอบที่ 4 จะเห็นว่าข้อสอบในโดเมนเดียวกันจะถูกจัดไว้ในตัวประกอบเดียวกัน เป็นส่วนใหญ่ ดังนั้นจึงสรุปได้ว่าแบบสอบมีความตรงเชิงโครงสร้าง

2.1.3 ความตรงร่วมสมัย จากการคำนวณค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สันระหว่างคะแนนที่ได้จากการสอบแบบสอบอิงโดเมน ในช่วงปลายภาคเรียนที่ 1 กับเกรดวิชาคณิตศาสตร์ (ค 311) ของภาคเรียนที่ 1 ปีการศึกษา 2531 ของนักเรียนผู้สอบได้ค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ 0.643 และมีความสัมพันธ์กันอย่างมีนัยสำคัญที่ระดับ .001 ดังนั้นจึงสรุปได้ว่า แบบสอบมีความตรงร่วมสมัย

## 2.2 ผลการตรวจสอบความเที่ยงของแบบสอบ

2.2.1 ความเที่ยงในรูปแบบอิงโดเมน โดยการคำนวณค่าดัชนีความเชื่อถือได้ ( $M(c)$ ) จากสูตรของเบรนนัน และ เคน พบว่าค่า  $M(c)$  ของแบบสอบในโดเมนที่ 1-4 และทั้งฉบับมีค่า 0.8884, 0.8485, 0.8443, 0.7700 และ 0.9428 ที่คะแนนเกณฑ์ ( $\pi_0$ ) เป็น 0.62, 0.63, 0.58, 0.54 และ 0.60 ตามลำดับ จะเห็นว่าค่าดัชนีความเชื่อถือได้ของแบบสอบอิงโดเมนฉบับนี้ มีค่าอยู่ในระดับสูง ดังนั้นจึงสรุปได้ว่า แบบสอบฉบับนี้มีคุณสมบัติด้านความเที่ยงตามรูปแบบอิงโดเมน

2.2.2 ความเที่ยงในรูปแบบของ IRT โดยการคำนวณค่าฟังก์ชันสารสนเทศแบบสอบ ณ ระดับความสามารถในช่วง -3 ถึง +3 พบว่ามีค่าสูงมากในช่วงที่ 0 มีค่าระหว่าง 0.0 ถึง 1.8 โดยมีค่าสูงสุดที่ระดับความสามารถ 0.8 คือเท่ากับ 30.55542 แสดงว่า แบบสอบฉบับนี้มีความเที่ยงสูงหรือจะมีประสิทธิภาพในการวัดสูงสุด เมื่อใช้กับผู้สอบที่มีระดับความสามารถปานกลางค่อนข้างสูง

3. ผลการกำหนดคะแนนเกณฑ์ของแบบสอบ จากการนำแบบสอบไปให้ผู้เชี่ยวชาญการสอนคณิตศาสตร์ ชั้น ม.3 พิจารณาเนื้อหาและความยากของข้อสอบ แล้วตัดสินเกณฑ์ผ่านขั้นต่ำของข้อสอบแต่ละข้อตามวิธีของแองกอฟ ผู้วิจัยนำผลการตัดสินมาเฉลี่ยเป็นรายข้อและรายโดเมนตามลำดับ ได้คะแนนเกณฑ์ของแบบสอบในโดเมนที่ 1-4 และทั้งฉบับ ซึ่งอยู่บนสเกลของคะแนนโดเมน ( $\pi_0$ ) เป็น 0.62, 0.63, 0.58, 0.54 และ 0.60 ตามลำดับ และเมื่อแปลงคะแนนโดเมนที่เป็นคะแนนเกณฑ์ ( $\pi_0$ ) ให้อยู่บนสเกลความสามารถ โดยอาศัยโค้งลักษณะแบบสอบ จะได้คะแนนเกณฑ์ตัดสินความรอบรู้ของแบบสอบในโดเมนที่ 1-4 และทั้งฉบับ ซึ่งอยู่บนสเกลความสามารถเป็น 1.48, 0.59, 0.37 1.04 และ 0.93 ตามลำดับ

#### 4. ผลการประมาณค่าความสามารถ และคะแนนโดเมน

4.1 ผลการประมาณค่าความสามารถของผู้สอบ จากการนำคะแนนการตอบแบบสอบถามโดเมนของผู้สอบไปประมาณค่าความสามารถในแต่ละโดเมนและทั้งฉบับ พบว่าในโดเมนที่ 1 ซึ่งมีข้อสอบ 11 ข้อ คะแนนที่ผู้สอบทำได้อยู่ในช่วง 0 ถึง 11 คะแนน โดยที่ค่าความสามารถเฉลี่ยของคะแนนแต่ละคะแนนจาก 1 ถึง 10 เป็น  $-3.33, -1.54, -0.30, 0.05, 0.24, 0.53, 0.83, 0.97, 1.30$  และ  $1.51$  ตามลำดับ

ในโดเมนที่ 2 ซึ่งมีข้อสอบ 17 ข้อ คะแนนที่ผู้สอบทำได้ในช่วง 1 ถึง 17 คะแนน โดยที่ค่าความสามารถเฉลี่ยของคะแนนแต่ละคะแนนจาก 1 ถึง 10 เป็น  $-2.90, -2.43, -2.53, -1.29, -1.41, -0.89, -0.37, -0.35, 0.00, 0.13, 0.37, 0.56, 0.76, 0.89, 1.13$  และ  $1.40$  ตามลำดับ

ในโดเมนที่ 3 ซึ่งมีข้อสอบ 10 ข้อ คะแนนที่ผู้สอบทำได้ อยู่ในช่วง 0 ถึง 10 คะแนน โดยที่ค่าความสามารถเฉลี่ยของคะแนนแต่ละคะแนนจาก 1 ถึง 9 เป็น  $-3.44, -2.61, -1.78, -1.14, -0.65, 0.22, 0.50, 0.87$  และ  $1.45$  ตามลำดับ

ในโดเมนที่ 4 ซึ่งมีข้อสอบ 12 ข้อ คะแนนที่ผู้สอบทำได้ในช่วง 0 ถึง 11 คะแนน โดยที่ค่าความสามารถเฉลี่ยของคะแนนแต่ละคะแนนจาก 1 ถึง 11 เป็น  $-4.87, -2.30, -1.95, -0.84, -0.31, -0.31, 0.54, 0.77, 1.03, 1.22$  และ  $1.94$  ตามลำดับ

เมื่อประมาณค่าความสามารถโดยคิดจากคะแนนการตอบแบบสอบถามทั้งฉบับจำนวน 50 ข้อ พบว่าคะแนนที่ผู้สอบทำได้ในช่วง 9 ถึง 49 คะแนน โดยมีค่าความสามารถเฉลี่ยของคะแนนแต่ละคะแนนเป็น  $-1.23, -1.77, -1.00, -1.04, -1.10, -0.87, -0.84, -0.62, -0.43, -0.35, -0.40, -0.20, -0.05, -0.09, -0.04, -0.01, 0.27, 0.26, 0.37, 0.34, 0.53, 0.52, 0.65, 0.66, 0.79, 0.96, 0.99, 1.07, 0.99, 1.16, 1.28, 1.31, 1.47, 1.63, 1.58, 1.72, 2.01$  และ  $3.06$  ตามลำดับ

4.2 ผลการประมาณคะแนนโดเมน จากการนำค่าพารามิเตอร์ของข้อสอบมาคำนวณค่าคะแนนโดเมนของแบบสอบถามแต่ละโดเมนและทั้งฉบับที่ระดับความสามารถในช่วง  $-5$  ถึง  $+5$  โดยห่างช่วงละ  $0.01$  พบว่า ในโดเมนที่ 1 ที่ระดับความสามารถดังกล่าว คะแนน

โดเมนมีค่าระหว่าง 0.13 ถึง 0.99 ในโดเมนที่ 2 คะแนนโดเมนมีค่าระหว่าง 0.20 ถึง 0.99 ในโดเมนที่ 3 คะแนนโดเมนมีค่าระหว่าง 0.17 ถึง 0.99 ในโดเมนที่ 4 คะแนนโดเมนมีค่าระหว่าง 0.20 ถึง 0.99 และเมื่อคิดจากแบบสอบทั้งสองฉบับ คะแนนโดเมนที่คำนวณได้ในช่วงความสามารถดังกล่าว มีค่าระหว่าง 0.18 ถึง 0.99

### อภิปรายผลการวิจัย

1. ในการสร้างแบบสอบอิงโดเมนฉบับนี้ ผู้วิจัยได้ดำเนินการอย่างละเอียดหลายขั้นตอน โดยได้ทำการวิเคราะห์เนื้อหาและจุดประสงค์เพื่อกำหนดเป็นกลุ่มพฤติกรรมหลักภายในขอบเขตเนื้อหาที่ต้องการวัดหรือโดเมนได้เป็น 4 โดเมน และวิเคราะห์เป็นพฤติกรรมย่อยหรือสับโดเมนได้เป็น 22 สับโดเมน ในการสร้างข้อสอบได้ใช้เทคนิคการกำหนดลักษณะเฉพาะของข้อสอบ ซึ่งเป็นวิธีการแปลงพฤติกรรมย่อยให้มีความเฉพาะเจาะจงและช่วยให้ได้ข้อสอบที่มีความเป็นตัวแทนของกลุ่มพฤติกรรมภายในขอบเขตเนื้อหาที่กำหนดขึ้น และในการกำหนดลักษณะของตัวเลือกและตัวลวงในการเขียนลักษณะเฉพาะของข้อสอบนั้น ผู้วิจัยได้รวบรวมมาจากข้อบกพร่องของนักเรียนจากการทำแบบฝึกหัดเรื่องพหุนาม จากข้อเสนอแนะของผู้เชี่ยวชาญการสอนและจากประสบการณ์การสอนในระดับชั้นมัธยมศึกษาปีที่ 1-3 ของผู้วิจัย เป็นต้น นอกจากนี้แบบสอบที่สร้างขึ้น ก็ได้ผ่านการตรวจสอบคุณภาพทั้งโดยผู้เชี่ยวชาญและจากการวิเคราะห์ทางสถิติทั้งในแนวอิงโดเมน และแนวของ IRT ดังนั้นจึงน่าจะเชื่อถือได้ว่า นอกจากข้อสอบจะมีความเป็นตัวแทนขอบเขตเนื้อหาที่ต้องการวัด ซึ่งช่วยให้การแปลผลคะแนนที่ได้มีความหมายพาดพิงไปสู่โดเมนนั้น ๆ แล้ว ยังเป็นข้อสอบที่มีคุณภาพที่จะช่วยระบุความรู้รอบและข้อบกพร่องในการเรียนเนื้อหานั้น ๆ ของผู้สอบอีกด้วย และนอกจากนี้ การจำแนกแบบสอบอิงโดเมนฉบับนี้ออกเป็น 4 โดเมน น่าที่จะให้ผลดังกล่าวได้ดีกว่าทั้งฉบับ อีกทั้งแบบสอบยังสามารถแยกย่อยไปสอบเป็นรายโดเมนได้ หลังจากที่ผู้สอนสอนจบเนื้อหาในโดเมนนั้น ๆ แล้ว ผู้สอนสามารถใช้แบบสอบเฉพาะโดเมนนั้น เพื่อช่วยในการระบุถึงความรู้รอบและข้อบกพร่องของผู้สอบ ซึ่งจะเป็นผลให้สามารถแก้ไขจุดบกพร่องของผู้เรียนได้ก่อนที่จะเรียนเนื้อหาในโดเมนต่อไป

2. ในขั้นการใช้โมเดลโลจิสติก 3 พารามิเตอร์ ในการตรวจสอบคุณสมบัติและการวิเคราะห์แบบสอบอิงโดเมนนั้น พบว่าแบบสอบมีคุณสมบัติการวัดเพียงคุณลักษณะเดียว และมีความเป็นอิสระของตำแหน่งข้อสอบจริง ซึ่งแสดงว่าแบบสอบมุ่งวัดกลุ่มพฤติกรรมในขอบเขต

เนื้อหาเรื่องพหุนามเพียงเรื่องเดียว ทั้งนี้อธิบายได้ว่าแม่แบบสอบจะแบ่งย่อยเป็น 4 โดเมน และแต่ละโดเมนก็มุ่งวัดผลพฤติกรรมในขอบเขตเนื้อหาเฉพาะเรื่อง แต่ก็ยังคงเป็นความรู้ที่เชื่อมโยงกันในขอบข่ายของเนื้อหาเรื่องพหุนามทั้งหมด ดังนั้นจึงสามารถนำแบบสอบอิงโดเมนนี้ไปทำการวิเคราะห์ด้วยโมเดลโลจิสติก 3 พารามิเตอร์ได้ ทั้งนี้เพราะเป็นไปตามข้อตกลงเบื้องต้นของ IRT ที่ว่า แบบสอบที่จะใช้วิเคราะห์ด้วยโมเดลต่าง ๆ ในกรอบทฤษฎีนี้จะต้องมีคุณสมบัติการวัดเพียงคุณลักษณะเดียว และความเป็นอิสระในตำแหน่งของข้อสอบ และ ฮัทเทน (Hutten 1982 : 4799-A) ได้ตรวจสอบพบว่า ความเหมาะสม (Fit) ของแบบสอบกับโมเดลในกรอบทฤษฎี IRT มีความสัมพันธ์กับคุณสมบัติการวัดเพียงคุณลักษณะ เดียวอย่างมีนัยสำคัญ ดังนั้นจึงกล่าวได้ว่าแบบสอบอิงโดเมนที่สร้างขึ้นมีความเหมาะสม (Fit) กับโมเดลโลจิสติก 3 พารามิเตอร์ และในการนำโมเดลโลจิสติก 3 พารามิเตอร์มาวิเคราะห์แบบสอบ จึงสามารถเชื่อถือได้ตามทฤษฎี IRT ที่ว่าค่าพารามิเตอร์ที่ได้ไม่แปรเปลี่ยนไปตามกลุ่มตัวอย่าง และจะให้สารสนเทศในการวัดได้อย่างมีประสิทธิภาพ

ผลจากการวิเคราะห์ข้อสอบรายข้อด้วยโมเดลโลจิสติก 3 พารามิเตอร์ และการคัดเลือกข้อสอบที่มีค่าพารามิเตอร์ตามเกณฑ์ที่เหมาะสม ทำให้ได้กลุ่มข้อสอบ (Item Pool) ที่ดีจำนวน 90 ข้อ และผู้วิจัยได้ทำการสุ่มข้อสอบในแต่ละโดเมน เพื่อจัดเป็นแบบสอบอิงโดเมน 1 ฉบับ จำนวน 50 ข้อ ซึ่งมีค่าอำนาจจำแนก ค่าความยาก และค่าการเดา ดังตารางที่ 9 ซึ่งสามารถอธิบายได้ว่า สำหรับค่าอำนาจจำแนกของแบบสอบ มีค่าตั้งแต่ 0.82293 ถึง 2.00000 มีค่าเฉลี่ยเป็น 1.27235 ซึ่งตามที่วอร์ม (Warm 1978 : 52) ได้กล่าวว่า ตามทฤษฎี IRT ช่วงของค่าอำนาจจำแนกจะอยู่ระหว่าง  $-\infty$  ถึง  $\infty$  แต่โดยทั่วไปค่าอำนาจจำแนกของแบบสอบจะมีค่าตั้งแต่ 0.5 ถึง 2.5 และค่าอำนาจจำแนกที่มากกว่า 0.8 นับว่าเป็นค่าที่ดีมาก ดังนั้นจึงถือได้ว่าข้อสอบทุกข้อในแบบสอบอิงโดเมนฉบับนี้ สามารถจำแนกผู้สอบได้ดีมาก ส่วนค่าความยากของแบบสอบซึ่งมีค่าตั้งแต่  $-0.79607$  ถึง  $2.32667$  มีค่าเฉลี่ยเท่ากับ  $0.79895$  ซึ่งวอร์มได้กล่าวว่า ค่าความยากตามทฤษฎี IRT จะมีค่าตั้งแต่  $-\infty$  ถึง  $\infty$  แต่ส่วนใหญ่แบบสอบวัดผลสัมฤทธิ์โดยทั่วไป จะมีค่าตั้งแต่  $-2.5$  ถึง  $2.5$  โดยถือว่าคุณค่าความยากที่น้อยกว่า  $-2.5$  เป็นข้อที่ง่ายมาก ค่าความยากที่มากกว่า  $2.5$  เป็นข้อที่ยากมาก และข้อสอบที่มีค่าความยากเท่ากับ  $0.0$  ถือว่ายากง่ายปานกลาง สำหรับแบบสอบอิงโดเมนฉบับนี้ จะเห็นว่าข้อสอบทุกข้อมีค่าความยากอยู่ในเกณฑ์ที่เหมาะสม และส่วนใหญ่เป็นข้อสอบที่มีความยากอยู่ในระดับปานกลาง



ค่อนข้างสูง ข้อสอบที่จัดว่าง่ายมีความยากเป็นลบได้แก่ ข้อ 16, 20, 29, 30 และ 31 ซึ่งเป็นการวัดพฤติกรรมกรรมการบวกและการคูณเนื้อหาเอกลนามอย่างง่าย ๆ และสำหรับค่าการเดาของแบบสอบพบว่ามีค่าตั้งแต่ 0.03223 ถึง 0.29393 โดยมีค่าเฉลี่ยเท่ากับ 0.18271 ตามทฤษฎีของ IRT ค่าการเดามีได้ตั้งแต่ 0 ถึง 1 แต่โดยทั่วไปค่าการเดาของแบบสอบจะมีค่าอยู่ระหว่าง 0.0 ถึง 0.5 (Wingersky, and Others 1982 : 2) โดยที่ถ้ามีค่าการเดามากกว่า 0.3 ถือว่าเป็นข้อสอบที่ไม่ดี โอกาสที่ผู้สอบที่มีความสามารถต่ำ จะสามารถตอบข้อสอบข้อนั้นได้ถูกมีมาก ค่าการเดายิ่งน้อยข้อสอบยิ่งดี และถ้าค่าการเดาเป็นศูนย์ จะถือว่าดีที่สุด (Warm 1978 : 53) สำหรับแบบสอบอิงโดเมนฉบับนี้ถือได้ว่ามีค่าการเดาอยู่ในระดับที่ยอมรับได้ทุกข้อ

### 3. การตรวจสอบคุณภาพแบบสอบในด้านความตรง สามารถอภิปรายผลได้ดังนี้

3.1 ความตรงเชิงเนื้อหา ผลจากการตรวจสอบผู้เชี่ยวชาญพบว่า ข้อสอบแต่ละข้อมีความสอดคล้องกับลักษณะเฉพาะของข้อสอบ ลักษณะเฉพาะของข้อสอบมีความสอดคล้องกับสับโดเมน และสับโดเมนของแต่ละโดเมนมีความครอบคลุมโดเมนนั้น ๆ ดังนั้นจึงสรุปได้ว่าแบบสอบมีความตรงเชิงเนื้อหา ทั้งนี้อธิบายได้ว่าการสร้างแบบสอบ ผู้วิจัยได้ทำการวิเคราะห์พฤติกรรมและเนื้อหาเรื่องพหุนามอย่างละเอียด โดยได้ปรึกษาผู้เชี่ยวชาญการสอนในทุกชั้นตอน ซึ่งสงบ ลักษณะ (2523 : 37-39) กล่าวไว้ว่า วิธีตรวจสอบความตรงเชิงเนื้อหาที่นิยมใช้มากที่สุดคือ การให้ผู้เชี่ยวชาญทางเนื้อหาวิชาเป็นผู้พิจารณา และ ประคอง กรรณสูต (2524 : 49) กล่าวว่า การใช้ผู้เชี่ยวชาญเป็นครูผู้สอน นับว่ามีความเหมาะสมที่สุด เพราะครูผู้สอนเป็นผู้ที่ใกล้ชิดและทราบถึงปัญหาในการเรียนของผู้สอบมากที่สุด และ มิลแมน (Millman 1974 : 307-397) กล่าวว่า ถ้าสามารถนิยามประชากรของข้อสอบได้อย่างชัดเจนเพียงพอแล้ว ข้อสอบจะมีความตรงเชิงเนื้อหาทันที ซึ่งวิธีการที่จะนิยามประชากรของข้อสอบได้อย่างชัดเจนที่สุด ก็คือการใช้เทคโนโลยีในการเขียนข้อสอบ เช่น ลักษณะเฉพาะของข้อสอบ เป็นต้น ดังนั้นจากที่กล่าวมาแล้ว จึงพอเชื่อถือได้ว่า แบบสอบฉบับนี้วัดได้ครอบคลุมเนื้อหาที่ต้องการวัดตาม โนทัศน์ของแบบสอบอิงโดเมน

3.2 ความตรงเชิงโครงสร้าง ผลจากการวิเคราะห์ตัวประกอบของแบบสอบอิงโดเมน จำนวน 50 ข้อ พบว่าข้อสอบในโดเมนเดียวกันจะถูกจัดเข้าไว้ในตัวประกอบเดียวกันเป็นส่วนใหญ่ ได้แก่ ในโดเมนที่ 1, 3 และ 4 ซึ่งในโดเมนเหล่านี้มุ่งวัดกลุ่มพฤติกรรมในขอบเขตเนื้อหาเฉพาะ ในการสร้างข้อสอบแต่ละโดเมนที่วัดเฉพาะ เรื่องจะมีรูปแบบคำถาม

คล้ายคลึงกัน มีวิธีการคิดหาคำตอบที่เหมือนกันเป็นส่วนใหญ่ เช่น โดเมนที่ 3 มุ่งวัดพฤติกรรม การคิดคำนวณเรื่องการคูณพหุนาม เป็นต้น ดังนั้นข้อสอบที่อยู่ในโดเมนเดียวกันจึงมีคุณสมบัติวัด ในสิ่งเดียวกันเป็นส่วนใหญ่ ส่วนในโดเมนที่ 2 ซึ่งมีกลุ่มข้อสอบที่จัดไว้ในตัวประกอบ 2 ตัว คือ ข้อสอบข้อที่ 15, 16, 18, 20, 21, 23 และ 24 ถูกจัดเข้าไว้ในตัวประกอบที่ 2 ส่วนข้อที่ 25, 26 และ 28 ถูกจัดไว้ในตัวประกอบที่ 5 ซึ่งอธิบายได้ว่าในโดเมนที่ 2 วัดพฤติกรรม การคิดคำนวณด้านการบวกและการลบพหุนาม ถึงแม้ข้อสอบจะมีรูปแบบคำถามคล้ายกัน มีวิธีการคิดใน ท่านองเดียวกัน แต่ต้องใช้ความรู้ที่ต่างกัน เช่น ในเรื่องของการเปลี่ยนเครื่องหมายในการลบ เป็นต้น จึงทำให้โดเมนนี้ดูเหมือนว่าจะวัดในหลายสิ่ง

3.3 ความตรงร่วมสมัย ผลการคำนวณค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สันระหว่าง คะแนนที่ได้จากการสอบแบบสอบอิงโดเมน กับเกรดวิชาคณิตศาสตร์ (ค 311) ภาคเรียนที่ 1 ปีการศึกษา 2531 มีค่าเท่ากับ 0.643 ที่ระดับ .001 ดังนั้นจึงถือได้ว่าแบบสอบมีคุณสมบัติของ ความตรงร่วมสมัย ถึงแม้ว่าค่าความสัมพันธ์ที่ได้จะไม่สูงนัก ซึ่งอาจอธิบายได้ว่า เนื้อหาเรื่อง พหุนาม เป็นเพียงเนื้อหาย่อยของเนื้อหาวิชาคณิตศาสตร์ ชั้น ม.3 (ค 311) นักเรียนที่ทำคะแนน จากการสอบแบบสอบอิงโดเมน เรื่องพหุนามได้ดี อาจทำไม่ได้ดีนักเมื่อสอบ เนื้อหารวมทั้งหมด นอกจากนี้ อาจขึ้นอยู่กับสภาพความพร้อมในการสอบของนักเรียนแต่ละคนในสภาพการณ์ที่แตกต่าง ออกไป เป็นต้น

#### 4. การตรวจสอบคุณภาพแบบสอบในด้านความเที่ยง

4.1 ความเที่ยงในรูปแบบอิงโดเมน โดยการประมาณค่าดัชนีความ เชื่อถือได้ :  $M(c)$  ตามสูตรของเบรนนันและเคน ซึ่งพัฒนาจากแนวคิดของทฤษฎีการอ้างอิงสรุปส่วนหนึ่งที่ มุ่งหาความเที่ยงจากการสอบกลุ่มตัวอย่างแล้วอ้างอิงสรุปไปยังมวลของสิ่งที่วัด และคำนึงถึง ความคลาดเคลื่อนในการวัด ซึ่งเกิดจากความแตกต่างระหว่างคะแนนมวลของสิ่งที่วัดกับคะแนน ปรากฏซึ่งคาดหวังไว้ และแบบสอบอิงโดเมนนั้น ข้อสอบที่สุ่มมาใช้ทดสอบ เป็นเพียงส่วนย่อยของ ประชากรข้อสอบ ซึ่งจำเป็นต้องคำนึงถึงความแปรปรวนคลาดเคลื่อนอันเนื่องมาจากการสุ่มข้อสอบ ในการพิจารณาประกอบการคำนวณค่าความเที่ยงด้วย และในการคำนวณความแปรปรวนของคะแนน ในแบบสอบอิงโดเมน ควรคิดจากผลต่างของค่าคะแนนที่สอบได้กับ เกณฑ์มาตรฐานยกกำลังสอง นอกจากนี้ในการคำนวณค่า  $M(c)$  ยังมีข้อดกลง เบื้องต้นที่สอดคล้องกับลักษณะของแบบสอบอิง โดเมนที่สร้างขึ้น กล่าวคือ ข้อสอบต้องมาจากขอบข่าย เนื้อหาที่ได้นิยามไว้อย่างดีแล้วและได้

จากการสุ่มอย่างง่าย หรือการสุ่มแบ่งชั้นจากมวลของสิ่งที่วัดจำนวนมาก พอร์มข้อสอบต้องเป็นพอร์ม  
 คู่ขนานแบบสุ่มซึ่งหมายถึงข้อสอบสร้างขึ้นตามเทคโนโลยีการเขียนข้อสอบอย่างใดอย่างหนึ่ง มี  
 จำนวนครั้งในการสอบเพียง 1 ครั้ง ให้คะแนนการสอบโดยวิธีประเพณีนิยม นอกจากนี้ยังไม่  
 จำกักรูปแบบการกระจายของคะแนนด้วย (Brennan 1980 : 223, Berk 1980 : 336,  
 Brennan and Kane 1977 : 277) ดังนั้นจะเห็นว่าการหาค่าความเที่ยงโดยการคำนวณค่า  
 $M(c)$  จึงมีความเหมาะสมกับแบบสอบอิงโดเมนที่สร้างขึ้น

สำหรับค่าความเที่ยงที่คำนวณได้พบว่า เมื่อคิดจากแบบสอบทั้งฉบับจะได้ความ  
 เที่ยงสูงสุด คือ 0.9428 และเมื่อคิดรายโดเมน พบว่า โดเมนที่ 1 มีความเที่ยงสูงสุด คือ  
 0.8884 และโดเมนที่ 4 มีความเที่ยงต่ำสุด คือ 0.7700 เมื่อพิจารณาถึงสิ่งที่มีอิทธิพลทำให้  
 ค่าความเที่ยงของแบบสอบในแต่ละโดเมนหรือทั้งฉบับ มีค่าสูงหรือต่ำ พบว่าน่าจะเกิดจากองค์  
 ประกอบที่สำคัญ คือ จำนวนข้อสอบ กับระยะห่างระหว่างค่าเฉลี่ยของคะแนนกับคะแนน เกณฑ์  
 กล่าวคือ ถ้ามีจำนวนข้อสอบจะทำให้แบบสอบฉบับนั้น หรือในโดเมนนั้นมีความเที่ยงสูง ดังจะ  
 เห็นได้จากแบบสอบทั้งฉบับ ซึ่งมีข้อสอบจำนวน 50 ข้อ จะมีค่าความเที่ยงสูงกว่าแบบสอบใน  
 แต่ละโดเมน ที่มีข้อสอบอย่างมากเพียง 17 ข้อ และแบบสอบในโดเมนที่ 2 ซึ่งมีข้อสอบจำนวน  
 17 ข้อ ก็จะมีค่าความเที่ยงสูงกว่าแบบสอบในแต่ละโดเมนที่มีข้อสอบอย่างมากเพียง 17 ข้อ และ  
 แบบสอบในโดเมนที่ 2 ซึ่งมีข้อสอบจำนวน 17 ข้อ ก็จะมีค่าความเที่ยงสูงกว่าแบบสอบในโดเมน  
 ที่ 3 และ 4 ซึ่งมีข้อสอบจำนวน 10 และ 12 ข้อ ตามลำดับ และแบบสอบในโดเมนที่ 4 ก็จะมี  
 ค่าความเที่ยงสูงกว่าโดเมนที่ 3 ซึ่งมีจำนวนข้อสอบน้อยกว่า แต่แบบสอบในโดเมนที่ 1 แม้จะมี  
 จำนวนข้อสอบเพียง 11 ข้อ ซึ่งน้อยกว่าในโดเมนที่ 2 และโดเมนที่ 4 แต่ก็มีค่าความเที่ยงสูงกว่า  
 โดเมนทั้งสอง ที่เป็นเช่นนี้อาจเป็นเพราะว่า ค่าเฉลี่ยของคะแนนในโดเมนที่ 1 ซึ่งเท่ากับ 0.27  
 ต่างจากคะแนนเกณฑ์ 0.62 มาก คิดเป็น 0.35 ในขณะที่ในโดเมนที่ 2 และ 4 มีค่าเฉลี่ย  
 ของคะแนนเท่ากับ 0.49 และ 0.43 ตามลำดับ ซึ่งต่างจากคะแนนเกณฑ์ 0.63 และ 0.54  
 เป็น 0.14 และ 0.11 ตามลำดับ ซึ่งสอดคล้องกับที่เบรนนัน (Brennan 1980 : 203) ได้  
 กล่าวไว้ว่า ยิ่งระยะห่างระหว่างค่าเฉลี่ยของคะแนนกับคะแนนเกณฑ์มากขึ้นเท่าใด ค่าความ  
 เที่ยงหรือค่าดัชนีความเชื่อถือได้ :  $M(c)$  ก็จะมากขึ้นเท่านั้น ซึ่งแสดงว่าผู้สอบยังมีคะแนน  
 ต่างจากคะแนน เกณฑ์มาก ก็ยังสามารถถูกแบ่งให้อยู่ในกลุ่มผู้รอบรู้หรือกลุ่มผู้ยังไม่รู้ได้ถูกต้องขึ้น

นอกจากนี้การกำหนดระดับคะแนนเกณฑ์ก็มีผลต่อค่าดัชนีความสอดคล้อง จะเห็นจากผลการวิจัยของ สุวัฒน์ สุขมลสันต์ (2528 : 101-103) พบว่า ค่า  $M(c)$  จะมีค่าสูงสุดเมื่อคะแนนเกณฑ์เท่ากับร้อยละ 0 ค่านี้จะค่อย ๆ ลดลงเมื่อขนาดของคะแนนเกณฑ์เพิ่มขึ้น จนกระทั่งคะแนนเกณฑ์อยู่ระหว่างร้อยละ 50 ถึง 60 ค่า  $M(c)$  จะน้อยที่สุด และจะมีค่าเพิ่มขึ้นเรื่อย ๆ เมื่อคะแนนเกณฑ์มีขนาดเพิ่มขึ้น แล้วจะมีค่าสูงสุดอีกครั้ง เมื่อคะแนนเกณฑ์เท่ากับร้อยละ 100 โดยเมื่อคะแนนเกณฑ์เท่ากับร้อยละ 50 และ 60 จะได้ค่า  $M(c)$  เป็น 0.862 และ 0.821 ตามลำดับ

4.2 ความเที่ยงในรูปแบบ IRT จากการศึกษาค่าฟังก์ชันสารสนเทศแบบสอบและค่าความคลาดเคลื่อนมาตรฐานในช่วงความสามารถ -3 ถึง +3 พบว่า ที่ระดับความสามารถต่ำและสูงมาก ค่าฟังก์ชันสารสนเทศที่ได้จะต่ำ ส่วนค่าความคลาดเคลื่อนมาตรฐานจะสูง และที่ระดับความสามารถปานกลาง และค่อนข้างสูงนั้น ค่าฟังก์ชันสารสนเทศจะสูง ส่วนค่าความคลาดเคลื่อนมาตรฐานจะต่ำ ทั้งนี้เนื่องมาจากค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถ เป็นสัดส่วนกลับของรากที่สองของค่าฟังก์ชันสารสนเทศแบบสอบ ดังนั้นถ้าค่าฟังก์ชันสารสนเทศยิ่งมากค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถก็จะยิ่งน้อยลง นั่นคือถ้าค่าฟังก์ชันสารสนเทศแบบสอบมีค่าสูงสุด ณ ระดับความสามารถใด แสดงว่า แบบสอบฉบับนั้นมีความเหมาะสมหรือมีประสิทธิภาพในการวัดสูงสุดที่ระดับความสามารถนั้น ๆ ก็จะทำให้ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่า ณ ระดับความสามารถนั้นลดลง ทำให้มีความแม่นยำ (Precision) ในการประมาณค่า ณ ระดับความสามารถนั้นสูง ในทำนองเดียวกัน ถ้าค่าฟังก์ชันสารสนเทศของแบบสอบมีค่าต่ำ ณ ระดับความสามารถใดก็จะทำให้ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถเพิ่มขึ้น แสดงว่าความถูกต้องแม่นยำในการประมาณค่าความสามารถ ณ ระดับนั้นก็น้อยลง ดังนั้นจากผลการวิจัยจึงสามารถสรุปได้ว่า แบบสอบอิงโดเมนฉบับนี้มีประสิทธิภาพสูงหรือมีความแม่นยำในการวัดครอบคลุมที่ระดับความสามารถปานกลางและค่อนข้างสูง ซึ่งสอดคล้องกับที่กรีนและคณะ (Green, et al. 1984 : 352) กล่าวว่า "... การทดสอบตามรูปแบบทั่วไปใช้แบบสอบที่สร้างขึ้น เพื่อจำแนกผู้สอบในช่วงความสามารถปานกลาง ดังนั้นจะมีความแปรปรวนคลาดเคลื่อนค่อนข้างต่ำในช่วงความสามารถปานกลาง และจะมีความแปรปรวนคลาดเคลื่อนค่อนข้างสูง เมื่อผู้สอบมีความสามารถสูงมากหรือต่ำมาก" ซึ่งน่าจะเหมาะสมกับแบบสอบอิงโดเมน ซึ่งมีจุดมุ่งหมายเพื่อการประมาณคะแนนโดเมน และหาสถานะ

การรื้อรอบของผู้สอบแต่ละคน เพื่อการปรับปรุงการเรียนการสอนให้ดีขึ้น โดยสามารถนำไปใช้กับผู้สอบได้อย่างมีประสิทธิภาพในวงกว้าง สำหรับผู้สอบที่มีระดับความสามารถต่ำมากย่อมมีข้อบกพร่องในการเรียนรู้หลายประเด็น ซึ่งอาจต้องใช้วิธีการสอนซ้ำทั้งหมด และสำหรับผู้สอบที่มีระดับความสามารถสูง ๆ มักจะเรียนรู้ได้ดี มีข้อบกพร่องน้อยหรือไม่มีเลย แต่ในผู้สอบที่มีระดับความสามารถปานกลางหรือค่อนข้างสูง มักจะบกพร่องในบางประเด็น ซึ่งการได้แบบสอบอิงโคเมนที่สร้างขึ้นอย่างครอบคลุม เนื้อหาที่มุ่งวัดและสามารถชี้ประเด็นความบกพร่องของผู้เรียนได้ และมีประสิทธิภาพสูงสุดในการวัดกับผู้สอบที่มีระดับความสามารถปานกลางและค่อนข้างสูง จึงนับได้ว่าเป็นแบบสอบที่มีความเหมาะสมสำหรับการวัดเพื่อประโยชน์ในการประเมินและปรับปรุงการเรียนการสอนอย่างแท้จริง

สำหรับ เหตุผลที่แบบสอบฉบับนี้มีประสิทธิภาพสูงสุดในการวัดผู้สอบที่มีระดับความสามารถปานกลาง และค่อนข้างสูง ทั้งนี้อาจเป็นเพราะแบบสอบฉบับนี้ประกอบด้วยข้อสอบที่มีค่าความยากอยู่ในระดับปานกลางและค่อนข้างสูง เป็นส่วนใหญ่ แล้วเป็นผลให้ฟังก์ชันสารสนเทศของข้อสอบแต่ละข้อส่วนใหญ่มีค่าสูงที่ระดับความสามารถปานกลางและค่อนข้างสูง ดังนั้นค่าฟังก์ชันสารสนเทศแบบสอบ ซึ่งเป็นผลรวมของฟังก์ชันสารสนเทศข้อสอบทั้งฉบับจึงมีประสิทธิภาพสูงสุด ณ ระดับความสามารถดังกล่าว

5. ผลการกำหนดคะแนนเกณฑ์ของแบบสอบ จากการพิจารณาตัดสินของผู้เชี่ยวชาญ การสอนได้คะแนนเกณฑ์ของแบบสอบในโคเมนที่ 1-4 และทั้งฉบับซึ่งอยู่บนสเกลของคะแนนโคเมน เป็น 0.62, 0.63, 0.58, 0.54 และ 0.60 ตามลำดับ ซึ่งเมื่อแปลงให้อยู่บนสเกลความสามารถจะได้คะแนนเกณฑ์เป็น 1.48, 0.59, 0.37, 1.04 และ 0.93 ตามลำดับ ค่าความสามารถที่เป็นคะแนนเกณฑ์ของแบบสอบดังกล่าวนี้ จะใช้เป็นตัวตัดสินถึงความรอบรู้หรือไม่ของผู้สอบ โดยที่ถ้าวัดความสามารถของผู้สอบมีน้อยกว่าค่าความสามารถที่เป็นคะแนนเกณฑ์ จะได้รับการตัดสินว่าไม่รอบรู้ในโคเมนนั้น จะเห็นว่าคะแนนเกณฑ์ของแต่ละโคเมนมีค่าไม่เท่ากัน ทั้งนี้ อาจเป็นเพราะผู้เชี่ยวชาญการสอนเห็นว่าความยากของข้อสอบในแต่ละโคเมนหรือความสำคัญของแต่ละโคเมนแตกต่างกัน จึงทำให้คะแนนเกณฑ์ที่จะตัดสินความรอบรู้ของผู้สอบไม่เท่ากันไปด้วย ซึ่งผลของคะแนนเกณฑ์ที่ได้จากการตัดสินของผู้เชี่ยวชาญการสอนนี้มีความสอดคล้องกับผลการวิจัยของอีกหลายท่านที่ได้คะแนนเกณฑ์ที่เหมาะสมของแบบสอบอิงโคเมนแยกเป็นรายฉบับที่แตกต่างกัน เช่น สวามินาธาน แสมเบิลตัน และ อัลจินา ที่ใช้จุดตัดเป็น 0.60 และ 0.80

ชมพู จันทรอมพร ใช้จุดตัดเป็น 0.50 และ 0.60 บุญเลิศ คำหอม ใช้จุดตัดเป็น 0.55 และ 0.60 สุโชติ สันตติวงศ์ไชย ได้จุดตัดคิดเป็นร้อยละ 50-70 ของคะแนนเต็ม และ สมหวัง พิธิยานุวัฒน์ ได้หาจุดตัดที่พอดีได้จุดตัดขั้นต่ำอย่างสูงเป็นร้อยละ 53.57 และได้จุดตัดขั้นสูงอย่างต่ำเท่ากับร้อยละ 70 เป็นต้น ซึ่งจะเห็นว่าคะแนนเกณฑ์หรือจุดตัดของวิชาคณิตศาสตร์ส่วนใหญ่จะอยู่ในช่วงตั้งแต่ 0.50 ถึง 0.80 ทั้งนี้ย่อมขึ้นอยู่กับความยากง่ายของเนื้อหาและข้อสอบด้วย

6. ในการประมาณค่าความสามารถของผู้สอบ ผู้วิจัยได้ทำการประมาณค่าในแต่ละโดเมน ทั้งนี้เพื่อให้การตัดสินใจความสามารถในการเรียนรู้เป็นไปอย่างชัดเจนยิ่งขึ้น และตามทฤษฎี IRT นั้น เมื่อทราบลักษณะการตอบข้อสอบแต่ละข้อของผู้สอบและค่าพารามิเตอร์ของข้อสอบข้อนั้น ๆ เราสามารถประมาณค่าความสามารถของผู้สอบให้อยู่บนมาตราความสามารถ (ability scale) อันเดียวกัน ได้จากข้อสอบชุดย่อยชุดใดชุดหนึ่งจากข้อสอบในโดเมน ซึ่งเราทราบค่าพารามิเตอร์นั้น (ผจงจิต อินทสุวรรณ 2525 : 66) และจากผลการประมาณค่าความสามารถของผู้สอบในแต่ละโดเมน พบว่า เมื่อผู้สอบมีคะแนนการสอบสูงขึ้น ก็จะมีค่าความสามารถสูงตามไปด้วย แต่ในผู้สอบบางคนอาจมีค่าความสามารถต่ำกว่าค่าความสามารถของผู้สอบที่ได้คะแนนการสอบต่ำกว่า ทั้งนี้อาจเป็นเพราะมีความผิดพลาดในรูปแบบการตอบข้อสอบของผู้สอบผู้นั้น เช่น อาจตอบข้อสอบข้อง่าย ๆ ผิด แต่ตอบข้อสอบที่ยากได้ถูกต้อง เป็นต้น ซึ่งทำให้มีผลต่อการประมาณค่าความสามารถของผู้สอบ ซึ่งโดยหลักการของ IRT แล้ว ผู้สอบที่มีความสามารถระดับสูง ๆ ไม่ควรอย่างยิ่งที่จะตอบข้อง่าย ๆ ผิด ซึ่งการตอบผิดโดยประมาทนี้มีผลต่อโมเดลโลจิสติก (Lord 1980 อ้างถึงใน ผจงจิต อินทสุวรรณ 2528 : 17) ในการพิจารณาค่าความสามารถของนักเรียน อาจยึดเกณฑ์ดังนี้ -3 ความสามารถต่ำมาก -2 ความสามารถต่ำ -1 ความสามารถค่อนข้างต่ำ 0 ความสามารถปานกลาง +1 ความสามารถค่อนข้างสูง +2 ความสามารถสูง และ +3 ความสามารถสูงมาก

7. จากผลการประมาณคะแนนโดเมนที่ระดับความสามารถต่าง ๆ ในช่วง -5 ถึง +5 พบว่า เมื่อระดับความสามารถสูงขึ้น ค่าคะแนนโดเมนจะสูงขึ้นด้วย โดยจะมีค่าระหว่าง 0 ถึง 1 โดย ๓ ระดับความสามารถต่ำและสูงมาก ช่วงคะแนนโดเมนจะแคบ กล่าวคือ คะแนนไม่แปรเปลี่ยนมาก ส่วนที่ระดับความสามารถปานกลางและค่อนข้างสูง ช่วงคะแนนโดเมนจะกว้าง ซึ่งเมื่อนำคะแนนโดเมนที่ระดับความสามารถต่าง ๆ ไปเขียนกราฟ จะได้กราฟของโค้งลักษณะ

แบบสอบที่มีลักษณะเพิ่มขึ้นทางเดียว (monotonic increasing) ตามระดับความสามารถที่เพิ่มขึ้น ซึ่งเป็นไปตามแนวคิดของ IRT ที่ว่าสำหรับผู้สอบที่มีความสามารถสูงกว่าควรจะมีโอกาสที่จะตอบข้อสอบข้อนั้นได้ถูกต้องหรือมีคะแนนโดเมนมากกว่าผู้ที่มีระดับความสามารถต่ำกว่า และ การที่โค้งลักษณะแบบสอบมีความชันในช่วงความสามารถระดับกลาง ๆ และค่อนข้างสูงนี้ และ ปลายโค้งด้านซ้ายจะมีค่าคะแนนโดเมนในช่วง 0.1 ถึง 0.2 สามารถอธิบายได้ว่า คะแนนโดเมนนั้นคำนวณได้จากฟังก์ชันคุณลักษณะข้อสอบและฟังก์ชันคุณลักษณะแบบสอบ ซึ่งขึ้นอยู่กับค่าพารามิเตอร์ข้อสอบ ได้แก่ ค่าความยาก (b) ค่าอำนาจจำแนก (a) และค่าการเดา (c) ซึ่งค่าพารามิเตอร์ของแบบสอบอิงโดเมนฉบับนี้ มีลักษณะดังนี้ คือ มีค่า a ค่อนข้างสูง โดยที่ค่าเฉลี่ยของค่า a เท่ากับ 1.27235 ค่า b อยู่ในระดับความยากปานกลางค่อนข้างสูง โดยที่ค่าเฉลี่ยของค่า b เท่ากับ 0.79895 และมีค่าการเดา โดยที่ค่าเฉลี่ยของค่า c เท่ากับ 0.18271 ซึ่งสอดคล้องกับที่ลอร์ด (Lord 1980 อ้างถึงใน พจนานุกรม 2528 : 73) ได้กล่าวว่า ถ้าแบบสอบมีค่าอำนาจจำแนกสูง โค้งลักษณะแบบสอบนี้จะชันตรงกลาง ขอบเขตของปลายด้านขวาจะมีค่าคะแนนโดเมนเท่ากับ 1 และที่ปลายด้านซ้ายเท่ากับค่าเฉลี่ยของค่า c ซึ่งผลการวัดจะดีมากในช่วงความสามารถระดับกลาง ๆ ซึ่งเหมาะสมกับคุณภาพของแบบสอบที่พบว่าแบบสอบฉบับนี้มีประสิทธิภาพสูงสุดในการวัด ผู้สอบที่มีระดับความสามารถปานกลางและค่อนข้างสูง ดังนั้นจึงน่าเชื่อถือได้ว่า เมื่อทราบคะแนนความสามารถของผู้สอบแล้ว จะสามารถเทียบเป็นคะแนนโดเมนจากคะแนนโดเมนที่คำนวณไว้ได้อย่างถูกต้องเหมาะสม

#### ข้อเสนอแนะ

ผู้วิจัยมีข้อเสนอแนะดังต่อไปนี้

##### 1. ข้อเสนอแนะสำหรับการวิจัย

1.1 น่าจะมีการศึกษาวิจัยเกี่ยวกับการนำโมเดลโลจิสติก 3 พารามิเตอร์ในการใช้พัฒนาแบบสอบอิงโดเมนในเนื้อหาวิชาอื่น และในระดับอื่น ๆ อีก เพื่อให้ได้ตัวอย่างแบบสอบที่มีคุณภาพสำหรับประโยชน์ในการนำไปใช้และเป็นแนวทางของการประยุกต์ใช้เพื่อพัฒนาแบบสอบต่อไป หรือน่าจะมีการสร้างแบบสอบอิงโดเมนที่มุ่งวัดโดเมนอื่น นอกเหนือจากโดเมนความรู้ (Cognitive Domain) เช่น โดเมนด้านจิตพิสัย (Affective Domain) เป็นต้น

1.2 ในการเขียนข้อสอบเพื่อให้มีความเป็นตัวแทนของโดเมนที่มุ่งวัด อาจใช้เทคโนโลยีในการเขียนข้อสอบ วิธีอื่นนอกจากวิธีการกำหนดลักษณะเฉพาะของข้อสอบ เช่น การใช้ฟอร์มข้อสอบ (Item Form) จุดประสงค์ขยายความ (Amplified Objectives) หรือการออกแบบแง่มุมคำถาม (The Facet Design Approach) เป็นต้น โดยเลือกใช้ให้เหมาะกับลักษณะเนื้อหาวิชา

1.3 ในการวิเคราะห์และตรวจสอบคุณภาพของแบบสอบอิงโดเมนนี้ นอกจากการใช้วิธีของ IRT แล้ว ยังสามารถใช้วิธีแบบอิงโดเมนหรืออิงเกณฑ์วิธีอื่น ๆ นอกเหนือจากในงานวิจัยฉบับนี้ เช่น การหาค่าความยาก ค่าอำนาจจำแนก (sensitivity Index : S) ตามวิธีของคริสปิน และเฟลด์ฮูเซน (Kryspin and Feldhusen 1974) หาค่าความเที่ยงโดยใช้สูตรของคาร์เวอร์ (Carver 1970) หรือวิธีของฮวิน (Huynh 1976) หาคะแนนเกณฑ์ โดยวิธี Pmax ของฮวิน (Huynh 1980) เป็นต้น ซึ่งอาจศึกษาต่อไปว่าวิธีใดจะเหมาะสมกับการนำมาใช้กับแบบสอบอิงโดเมนมากที่สุด ในแง่ของการคัดเลือกข้อสอบ และสารสนเทศที่ได้

1.4 ในด้านการประมาณค่าความสามารถของผู้สอบโดยวิธีความเป็นไปได้สูงสุด (maximum likelihood estimation) นั้น มีข้อจำกัดในแง่ที่ไม่สามารถประมาณค่าความสามารถของผู้สอบที่ได้คะแนนเต็มหรือศูนย์ ดังนั้นอาจศึกษาเกี่ยวกับการประมาณค่าความสามารถวิธีอื่น เช่น วิธีการประมาณค่าแบบเบย์ (Bayesian Estimation) วิธีการประมาณค่าแบบฮิวริสติก (Heuristic Estimation) เป็นต้น

## 2. ข้อเสนอแนะในการนำแบบสอบไปใช้

2.1 ในการนำแบบสอบฉบับนี้ไปใช้ควรดำเนินการตามคู่มือการใช้แบบสอบโดยเคร่งครัด

2.2 ควรนำแบบสอบนี้ใช้ทดสอบกับนักเรียนทันทีภายหลังจากที่สอนเนื้อหาในแต่ละบทเรียนย่อย ๆ หรือโดเมนหนึ่ง ๆ จบลง ทั้งนี้ควรดำเนินการสอบและตัดสินผลการสอบรู้โดยแยกเป็นรายโดเมน เพื่อให้ทราบข้อบกพร่องและระดับความรู้ของผู้เรียนในโดเมนนั้น ๆ หรืออาจดำเนินการสอบทั้งฉบับก็ได้ และควรแจ้งผลการสอบให้นักเรียนทราบอย่างรวดเร็ว ตลอดจนจัดการสอนซ่อมเสริมให้ทันที่ก่อนขึ้นบทเรียนใหม่ หรือโดเมนใหม่



2.3 ในการประมาณค่าความสามารถของผู้เรียน นอกจากจะประมาณโดยใช้โปรแกรม LOGIST 5 แล้ว ถ้ามีเครื่องไมโครคอมพิวเตอร์ อาจใช้โปรแกรมการประมาณค่าความสามารถ ซึ่งผู้วิจัยได้รับความอนุเคราะห์จาก คุณนิรันดร์ สาโรวาท นิสิตปริญญาโท-บัณฑิต ภาควิชาโสตทัศนศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ในการเขียนโปรแกรม หรืออาจจะใช้วิธีเฉลี่ยค่าความยากของข้อที่ตอบถูก จากจำนวนข้อสอบที่นักเรียนทำทั้งหมด ซึ่งอาจจะทำให้ค่าความสามารถที่ประมาณได้ใกล้เคียงกับค่าความสามารถที่คำนวณจากโปรแกรม LOGIST 5 และไวส์ (Weiss, D.J. 1974 : 4) ได้กล่าวว่ววิธีนี้น่าจะเหมาะในการประมาณค่าความสามารถของนักเรียนอย่างหายาบ ๆ และรวดเร็วไม่ต้องเสียเวลาไปใช้เครื่องคอมพิวเตอร์ในการคำนวณ

2.4 เนื่องจากการประมาณค่าความสามารถของผู้สอบโดยวิธีความเป็นไปได้สูงสุดของโมเดลโลจิสติก 3 พารามิเตอร์นี้ มีความซับซ้อนเนื่องจากใช้ฟังก์ชันคณิตศาสตร์ขั้นสูงในการคำนวณ ดังนั้นในการนำแบบสอบไปใช้จึงเหมาะกับการดำเนินการสอบแบบ ซี เอ ที (Computerized Adaptive Testing : CAT) ซึ่งสามารถลุ่มข้อสอบที่ค่าพารามิเตอร์ข้อสอบของแต่ละโดเมน ถูกบันทึกไว้ในคอมพิวเตอร์ ซึ่งเมื่อผู้สอบตอบข้อสอบเครื่องคอมพิวเตอร์ จะตรวจให้คะแนนและประมาณค่าความสามารถของผู้สอบได้ทันที ซึ่งจะทำให้มีความสะดวกและค่าความสามารถที่ได้ก็จะมีค่าถูกต้องยิ่งขึ้น และถ้าจะนำมาใช้ทดสอบโดยการเขียนตอบ ก็อาจจะพัฒนาแบบสอบโดยใช้โมเดลโลจิสติกชนิด 1 พารามิเตอร์ ก็ได้ เนื่องจากการคำนวณค่าความสามารถมีความซับซ้อนน้อยกว่าชนิด 3 พารามิเตอร์ เพราะพิจารณาเฉพาะค่าความยาก ซึ่งจะสามารถจัดทำเป็นตารางเทียบค่าความสามารถได้ง่ายและสะดวกขึ้น

2.5 ในการประยุกต์ใช้โมเดลโลจิสติกในการพัฒนาแบบสอบอิงโดเมนนี้ ควรจะเริ่มต้นอย่างน้อยในระดับของกลุ่มโรงเรียนก่อน เพราะจะได้เป็นการระดมพลังสมองในการสร้าง ซึ่งจะช่วยให้แบบสอบมีคุณภาพดียิ่งขึ้น และจะมีความพร้อมในด้านการใช้เครื่องคอมพิวเตอร์มากกว่า