



บทที่ 1

บทนำ

## ความเป็นมาและความสำคัญของปัญหา

แบบสอบเป็นเครื่องมือวัดชนิดหนึ่งที่ประกอบด้วยชุดของข้อสอบที่สร้างขึ้นมา เพื่อใช้วัดคุณสมบัตินี้ (Attributes) อย่างใดอย่างหนึ่งของพฤติกรรมของแต่ละบุคคล ได้มีการใช้แบบสอบอย่างแพร่หลายกว้างขวางมากในปัจจุบัน ตามจุดมุ่งหมายที่แตกต่างกันไป นอกเหนือจากเป็นเครื่องมือในการปรับปรุงการเรียนการสอนแล้ว ยังใช้ในการสอบแข่งขันคัดเลือกบุคคลเข้าเรียนเข้าทำงาน จัดตำแหน่ง เลื่อนระดับ ซึ่งการสอบต่าง ๆ เหล่านี้มีการสอบประจำทุกปี ทำให้มีการขนควายหาข้อสอบเก่ามาศึกษา และชักซ้อมความชำนาญเพื่อให้ได้คะแนนเพิ่มขึ้น ดังนั้นเพื่อให้เกิดความยุติธรรมแก่ผู้เข้าสอบจึงต้องมีการสร้างข้อสอบใหม่ขึ้นทุกครั้งที่มีการสอบ โดยสร้างให้มีลักษณะเหมือน ๆ กัน เรียกว่า แบบสอบคู่ขนาน การสร้างแบบสอบคู่ขนานเพื่อที่จะให้ได้ลักษณะของข้อสอบ ที่มีความใกล้เคียงกันด้านเนื้อหา ค่าสถิติของข้อสอบ และแบบสอบ (Guilliksen 1950 cited by Angoff 1984: 86) ซึ่งในความเป็นจริงทำได้ยากมาก และไม่คุ้มกับการลงทุน เพราะสร้างแล้วมักจะได้แต่แบบสอบที่เทียบเคียงกันได้เท่านั้น จึงมีผู้คิดค้นวิธีการที่จะทำให้คะแนนที่วัดจากแบบสอบที่มีเนื้อหาเดียวกัน แต่ต่างฉบับกันสามารถนำมาเทียบกันได้ เป็นกระบวนการที่ใช้วิธีการทางสถิติเพื่อปรับคะแนนที่ได้ให้อยู่ในมาตราเดียวกัน (ชูศักดิ์ ชัมภลิจิต 2529: 144) เรียกว่า การเทียบมาตราคะแนนของแบบสอบ (Test Equating) ซึ่งจำแนกได้เป็นสองลักษณะใหญ่ คือ (สงบ ลักษณะ 2525: 23)

1. การเทียบมาตราในแนวราบหรือแนวนอน (Horizontal equating) เป็นการเทียบคะแนนระหว่างแบบสอบสองชุดที่มีความยากพอ ๆ กัน สอบกับบุคคลกลุ่มเดียวกันเพื่อเทียบคะแนนจากแบบสอบฉบับหนึ่งว่าจะเท่ากับกี่หน่วยคะแนนของแบบสอบอีกฉบับหนึ่ง

2. การเทียบมาตราในแนวตั้งหรือแนวตั้ง (Vertical equating) เป็นการเทียบคะแนนระหว่างแบบสอบสองชุดที่มีความยากต่างกัน สอบกับบุคคลที่มีความสามารถต่างกัน การเทียบมาตราคะแนนของแบบสอบจึงเป็นวิธีการศึกษาเชิงประจักษ์ (Empirical Study) ( Petersen, Marco and Stewart 1982: 73; Petersen, Kolen and Hoover 1989: 242) ที่ต้องการการออกแบบเพื่อรวบรวมข้อมูล และจัดกระทำข้อมูลภายใต้สภาพของความจำกัดที่ไม่สามารถวางเงื่อนไขให้มีความพอเพียงตามนิยามได้ทุกประการ จึงมี

องค์ประกอบหลายประการที่เข้ามาเกี่ยวข้องกับการเทียบมาตรา อันได้แก่ คุณลักษณะของกลุ่มผู้สอบที่ได้มาจากการสุ่มหรือไม่ได้สุ่ม ระดับความสามารถของกลุ่มผู้สอบแตกต่างกันหรือใกล้เคียงกัน ลักษณะการกระจายของคะแนน เป็นต้น องค์ประกอบต่าง ๆ เหล่านี้ ทำให้ไม่สามารถที่จะใช้รูปแบบเดียวกันในการเทียบมาตราได้ รูปแบบการเทียบมาตราจึงเป็นองค์ประกอบหนึ่งที่สำคัญ เพราะเป็นการกำหนดเงื่อนไข และสร้างกฎการแปลงคะแนนในแต่ละครั้ง รูปแบบการเทียบมาตราที่ใช้กันส่วนใหญ่ คือ

1. รูปแบบอีควิเปอร์เซ็นไทล์ (Equipercetile Model) เป็นการเทียบคะแนนดิบระหว่างแบบสอบต่างชุดที่ตำแหน่งเปอร์เซ็นไทล์เดียวกัน (Angoff 1971: 563-568; Petersen, Marco and Stewart 1982: 73)

2. รูปแบบเชิงเส้นตรง (Linear Model) เป็นการเทียบคะแนนดิบระหว่างแบบสอบต่างชุดที่ตรงกับคะแนนมาตรฐาน (Z-score) ค่าเดียวกัน (Angoff 1971: 564; Petersen, Macro and Stewart 1982: 73)

3. รูปแบบอิงทฤษฎีการตอบข้อสอบ (Item Response Theory Model) เป็นการเทียบคะแนนโดยใช้คะแนนจริง (true score equating) ที่ระดับค่าประมาณความสามารถเท่าเดียวกันระหว่างแบบสอบต่างชุด (Petersen, Macro and Stewart 1982: 73)

4. รูปแบบการใช้เทคนิคการวิเคราะห์องค์ประกอบ (Equating Using the Confirmatory factor Analysis Model) ซึ่งเรียกเป็นผู้เสนอไว้ (Rock 1982: 247-258)

จากการศึกษาในเชิงเปรียบเทียบระหว่างรูปแบบการเทียบมาตรา ยังไม่อาจสรุปได้ว่า รูปแบบการเทียบมาตรารูปแบบใดมีประสิทธิภาพมากที่สุด เนื่องจากการเทียบมาตรารูปแบบหนึ่งอาจให้ผลดี ในสถานการณ์หนึ่ง แต่อาจให้ผลด้อยในอีกสถานการณ์หนึ่ง (Kolen and Whitney 1982: 791; Petersen, Cook and Stocking 1983: 155) เพราะผลการเทียบมาตราขึ้นอยู่กับองค์ประกอบหลายประการดังกล่าวแล้ว จึงควรเลือกใช้รูปแบบการเทียบมาตราที่เหมาะสม เพื่อให้ได้ผลการเทียบมาตราที่ถูกต้องหรือมีคุณภาพ เป็นการประเมินก่อนการเทียบมาตราจากการรวบรวมแนวคิดเชิงทฤษฎี นิยาม และงานวิจัยต่าง ๆ พอสรุปได้ว่า

1. การเทียบมาตราเชิงเส้นตรง ซึ่งเป็นการเทียบมาตราในรูปแบบของทฤษฎีดั้งเดิม (Classical Theory) ที่พัฒนามาพร้อม ๆ กับรูปแบบอีควิเปอร์เซ็นไทล์นั้น เป็นรูปแบบที่ง่าย และสะดวกในการแปลงคะแนน (Kolen and Whitney 1982: 291; Hills, Subhiyah and Hirsch 1988: 229) มีความเหมาะสมในการแปลงคะแนนอย่างมาก เมื่อการแจกแจงคะแนนของแบบสอบสองชุดที่นำมาเทียบมาตรานั้นมีรูปร่าง (Shape) เหมือนกัน (Angoff 1984: 87) แต่ถ้ารูปร่างของการแจกแจงคะแนนของแบบสอบต่างชุดนั้นแตกต่างกันแล้ว การเทียบมาตราแบบอีควิเปอร์เซ็นไทล์ จะให้ผลดีกว่า การเทียบมาตราเชิงเส้นตรง

(Angoff 1984: 88) อย่างไรก็ตามหากรูปร่างของการแจกแจงคะแนนของแบบสอบคล้ายคลึงกันมาก ควรใช้การเทียบมาตราเชิงเส้นตรง ซึ่งจะให้ผลดีกว่าการเทียบแบบอิกวิเปอร์เซ็นไต์ล เพราะเป็นปรนัยมากกว่า และไม่มี ความคลาดเคลื่อนในการปรับหรือเกลามาตรภาพให้เรียบตามวิธีการเทียบมาตราแบบอิกวิเปอร์เซ็นไต์ล ซึ่งจะมีความคลาดเคลื่อนมากในช่วงที่มีข้อมูลน้อย (Angoff 1971: 564; Potthoff 1982: 209)

2. สำหรับการเทียบมาตราแบบอิงทฤษฎีการตอบข้อสอบ วิธีนี้มีความสมบัติของความไม่แปรผันไปตามกลุ่มตัวอย่างที่นำมาใช้เทียบมาตรา (Invariance across groups) (Lord 1980: 199) วิธีนี้จึงให้ผลดี แม้ว่าในกรณีที่มีการแจกแจงคะแนนจากแบบสอบต่างชุดแตกต่างกัน (Kolen and Whitney 1982: 279-294)

นอกจากรูปแบบการเทียบมาตราจะเป็นองค์ประกอบสำคัญในการเทียบมาตราแล้วการเทียบมาตรายังขึ้นอยู่กับวิธีการรวบรวมข้อมูลในการนำมาใช้ จากที่แองกอฟฟ์ (Angoff 1984: 93-94) ได้รวบรวมไว้ สามารถจำแนกได้เป็น 2 กลุ่มใหญ่ ๆ คือ กลุ่มเทียบมาตราโดยใช้แบบสอบร่วม (Anchor or common Test) และกลุ่มเทียบมาตราโดยไม่ใช้แบบสอบร่วม ซึ่งเป็นรูปแบบที่มีเงื่อนไขเข้มงวดหลายประการ คือ ความเป็นตัวอย่างสุ่มสมมูลของผู้รับการทดสอบ ความเป็นคู่ขนานของแบบสอบโดยเน้นความสมมูลของค่าความเที่ยงของแบบสอบทั้งสอง เมื่อเงื่อนไขเข้มงวดมาก การปฏิบัติมักมีข้อขัดแย้ง เพราะในสภาพการณ์จริงโอกาสในการสุ่มและการตรวจสอบความสมมูลของความเที่ยงเป็นไปได้ยาก ดังนั้น การเทียบมาตราโดยใช้แบบสอบร่วมจึงเป็นทางออกที่เหมาะสมกว่า

แบบสอบร่วมที่ใช้ในการเทียบมาตรานั้น เป็นกลุ่มข้อสอบที่กำหนดให้ผู้สอบสองกลุ่มที่ทำแบบสอบคนละชุด (X หรือ Y) ได้ทำเหมือนกัน แบบสอบร่วมจะทำหน้าที่ลดความแตกต่างของกลุ่มตัวอย่างที่เนื่องมาจากความแปรปรวนของการสุ่ม (Angoff 1984: 105) โดยมีจุดมุ่งหมายเพื่อเป็นตัวเชื่อมข้อมูลเข้าด้วยกัน ทำให้ค่าพารามิเตอร์ที่วิเคราะห์ออกมาอยู่ในระดับเดียวกัน ผลการเทียบมาตราจึงมีความแม่นยำ (precision) มากขึ้น (Angoff 1984: 106) ดังนั้นแบบสอบร่วมจึงควรมีเนื้อหาเดียวกับแบบสอบเทียบมาตรา โดยดำเนินการสร้างและจัดชุดแบบสอบร่วมอย่างพิถีพิถันเพื่อให้มีลักษณะเทียบเท่ากับแบบสอบเทียบมาตรา (Petersen, Marco and Stewart 1982: 134)

แบบสอบร่วมสามารถจะแยก หรือรวมอยู่ในแบบสอบที่ต้องการเทียบมาตราก็ได้ (Angoff 1984: 107) แบบสอบร่วมที่แยกจากแบบสอบที่ต้องการเทียบมาตรา โดยไม่รวมคะแนนส่วนนี้เข้าในผลการสอบรายบุคคลเรียกว่า แบบสอบร่วมภายนอก (External or exclusive Anchor test) ถ้ามีจำนวนข้อมากจะต้องใช้เวลาทำมากขึ้น สำหรับแบบสอบร่วมที่ผนวกเข้าไปในแบบสอบที่ต้องการเทียบมาตราทั้งสองชุด เรียกว่าแบบสอบร่วมภายใน (Internal or inclusive anchor test) ถ้าแบบสอบร่วมประเภทนี้มีควมยาวมาก

เท่าไร ผู้สอบที่เคยรับการทดสอบมาก่อนย่อมได้เปรียบ เพราะเกิดการเรียนรู้หรือจำข้อสอบใน ส่วนของแบบสอบร่วมได้ อย่างไรก็ตามแบบสอบร่วมจะต้องมีความยาวเพียงพอ ที่จะไปปรับความ แตกต่างระหว่างกลุ่มตัวอย่างหรือแบบสอบที่ต่างชุดกัน

แองกอฟฟ์ (Angoff 1984: 107) ได้กล่าวเกี่ยวกับความยาวของแบบสอบร่วมไว้ ว่าควรมีจำนวนไม่น้อยกว่า 20 ข้อ หรือน้อยกว่าร้อยละ 20 ของจำนวนข้อสอบในแบบสอบ เทียบมาตรา แล้วแต่จำนวนไหนมากกว่าให้ใช้จำนวนนั้น สำหรับไรท์และสโตน (Wright and Stone 1979: 98) ได้กล่าวว่าแบบสอบร่วมควรเป็นข้อสอบที่วัดเรื่องเดียวกันกับแบบสอบ ที่ใช้เทียบมาตราทั้งสองชุด และมีจำนวนเพียง 10 ข้อก็เพียงพอ ประเด็นความยาวเกี่ยวกับแบบ สอบร่วมหรือจำนวนข้อในแบบสอบร่วมจึงเป็นองค์ประกอบที่สำคัญในการเทียบมาตรา จำนวนข้อ ของแบบสอบร่วมที่เหมาะสมจะช่วยให้ผลการเทียบมาตรามีความถูกต้องมากขึ้น ทำให้ต้นทุน ค่าใช้จ่ายและเวลาดำเนินการ

ในต่างประเทศได้มีการศึกษาเกี่ยวกับจำนวนข้อ และลักษณะของแบบสอบร่วมที่ใช้ใน การเทียบมาตราอย่างกว้างขวาง ปีเตอร์เซนและคูก (Petersen and Cook 1978: 238) ได้ชี้ให้เห็นว่าในการเทียบมาตรารูปแบบดั้งเดิม จำนวนข้อและความเป็นตัวแทนเนื้อหาของ แบบสอบร่วมเป็นตัวแปรที่มีความสำคัญ ในการศึกษาความยาวของแบบสอบร่วมนั้น แมคคินเลย์ และเล็กเคส (Mckinley and Reckase 1981 cited by Cook and Petersen 1987) ได้ศึกษาความยาวของแบบสอบร่วม 3 ขนาด คือ 5 15 และ 25 จากข้อมูลจริงของแบบสอบ วัดผลสัมฤทธิ์ พบว่าจำนวนข้อของแบบสอบร่วมตั้งแต่ 15 ข้อ เป็นจำนวนที่เพียงพอในการ เป็นตัวเชื่อมสำหรับการเทียบมาตรารูปแบบ IRT ชนิด 3 พารามิเตอร์ วิธี Concurrent Calibration ถึงแม้ว่าแบบสอบร่วม 25 ข้อ จะให้ผลการเทียบมาตราที่ดีกว่า 15 ข้อ แต่ แบบสอบร่วมขนาด 5 ข้อ เป็นจำนวนที่ไม่เพียงพอ สอดคล้องกับฮิลล์และคณะ (Hills et al 1988) ที่ศึกษาความยาวของแบบสอบร่วม 6 ขนาด คือ 30 25 20 15 10 และ 5 ข้อ จาก ข้อมูลจริงของแบบสอบวัดผลสัมฤทธิ์แบบ Minimum-Competency ที่สร้างในแต่ละปีการศึกษามี ความยาวฉบับละ 45 ข้อ เทียบมาตรารูปแบบเดียวกับแมคคินเลย์และเล็กเคส และพบว่าข้อสอบ ร่วมจำนวน 5 ข้อ (คิดเป็นร้อยละ 11 ของแบบสอบเทียบมาตรา) ยังไม่สามารถใช้ได้ แบบ สอบร่วมขนาด 10 ข้อ (คิดเป็นร้อยละ 22 ของแบบสอบเทียบมาตรา) ขึ้นไป เป็นจำนวนที่ เพียงพอในการเทียบมาตรา และให้ประสิทธิภาพเท่ากับแบบสอบร่วมขนาด 30 ข้อ ในขณะที่ราจู และคณะ (Raju et al 1986 cited by Hills et al 1988) ซึ่งเทียบมาตรารูปแบบ IRT ชนิด 3 พารามิเตอร์ กลับพบว่าข้อสอบร่วมจำนวน 5-6 ข้อ สามารถทำให้การเทียบ มาตรามีความแม่นยำได้ สำหรับในประเทศไทย ภาวิณี ศรีสุขวัฒนานันท์ (2529) ได้ศึกษา เกี่ยวกับความยาวของแบบสอบร่วม 3 ขนาด คือ 7 14 และ 21 ข้อ ซึ่งคิดเป็นร้อยละ 20 40 และ 60 ของแบบสอบเทียบมาตรา ดำเนินการเทียบมาตรา 3 รูปแบบ รวบรวมข้อมูลโดยการ จัดกลุ่มข้อสอบจากแบบสอบฉบับเดียวกัน แบ่งเป็นแบบสอบเทียบมาตราสองฉบับ ๆ ละ 35 ข้อ

และแบบสอบร่วมอีก 1 กลุ่ม แล้วนำข้อมูลผลการสอบ ซึ่งได้ดำเนินการโดยแบบสอบฉบับนี้มาแล้ว มาจัดกระทำตามรูปแบบที่ออกแบบไว้ พบว่าแบบสอบร่วมขนาด 7 ข้อ ซึ่งคิดเป็นร้อยละ 20 ของแบบสอบเทียบมาตรฐาน สามารถทำให้ผลการเทียบมาตรฐานมีคุณภาพอยู่ในระดับที่น่าพอใจ

จากข้อมูลเชิงประจักษ์ที่ได้จากการวิจัย พบว่าจำนวนข้อของแบบสอบร่วมที่ใช้ในการเทียบมาตรฐาน ในแต่ละรูปแบบ แต่ละสถานการณ์นั้น แตกต่างกันไป แม้แต่การเทียบมาตรฐานในรูปแบบเดียวกัน แต่ต่างสถานการณ์และแบบสอบ ยังให้ผลที่แตกต่างกัน ผู้วิจัยจึงสนใจที่จะเปรียบเทียบคุณภาพของการเทียบมาตรฐาน จากการใช้แบบสอบร่วมขนาดความยาวต่าง ๆ กัน เมื่อนำมาใช้กับแบบสอบเทียบมาตรฐาน ซึ่งครูเป็นผู้สร้างสำหรับวัดผลสัมฤทธิ์ในแต่ละรายวิชา แต่ละปีการศึกษาที่มีความยาวไม่เกินฉบับละ 60 ข้อ โดยนำแบบสอบร่วมวัดผลสัมฤทธิ์ที่ผู้วิจัยสร้างขึ้น เข้าไปร่วมอยู่ในแบบสอบในแต่ละปีการศึกษา ให้สามารถเทียบมาตรฐานได้ตามสถานการณ์ปกติ โดยผู้วิจัยสนใจที่จะศึกษาความยาวของข้อสอบร่วมขนาด 15 20 และ 25 ข้อ ซึ่งคิดเป็นร้อยละ 25 33 และ 41 ของแบบสอบเทียบมาตรฐาน เป็นช่วงความยาวที่ไม่น้อยกว่าร้อยละ 20 ของแบบสอบเทียบมาตรฐาน ตามข้อเสนอแนะของแองกอฟฟ์ที่ให้ไว้ และสอดคล้องกับข้อมูลเชิงประจักษ์จากงานวิจัยของฮิลล์และคณะ และของภาวิณี ศรีสุขวัฒนานันท์ ที่พบว่าแบบสอบร่วมที่มีความยาวไม่น้อยกว่าร้อยละ 20 เช่นกัน ให้ผลการเทียบมาตรฐานอยู่ระดับที่น่าพอใจ นอกจากนี้ยังได้ศึกษาความยาวของแบบสอบร่วมขนาด 10 ข้อ ซึ่งคิดเป็นร้อยละ 16 ของแบบสอบเทียบมาตรฐาน แม้ว่าจะต่ำกว่าร้อยละ 20 ก็ตาม แต่จากข้อเสนอของไรท์และสโตนที่ให้ไว้ และเมื่อพิจารณาความยาวแบบสอบวัดผลสัมฤทธิ์ ส่วนใหญ่ที่ใช้ในการเรียนการสอนปกติ มีความยาวประมาณ 45-60 ข้อแล้ว ผู้วิจัยคิดว่าแบบสอบร่วมขนาด 10 ข้อ เป็นความยาวที่น่าจะเหมาะสม ที่จะนำมาศึกษาเพิ่มเติมว่ายังคงให้คุณภาพของการเทียบมาตรฐานอยู่ในระดับที่ยอมรับได้หรือไม่ จึงกำหนดความยาวของแบบสอบร่วมที่ใช้ในการศึกษาครั้งนี้เป็น 4 ขนาด คือ 10 15 20 และ 25 ข้อ

สำหรับการตรวจสอบคุณภาพของการเทียบมาตรฐาน ผู้วิจัยจะตรวจสอบจากค่าความคลาดเคลื่อนมาตรฐานของการเทียบมาตรฐาน ซึ่งเป็นเกณฑ์ภายใน และค่าดัชนีความแตกต่างซึ่งเป็นค่าความคลาดเคลื่อนรวมที่ได้จากเกณฑ์ภายนอก โดยให้แบบสอบร่วมที่มีขนาดสั้นกว่าเป็นฉบับเซทของแบบสอบร่วมที่มีขนาดยาวกว่า ตามวิธีการที่โคเลนและวิทนีย์ (Kolen and Whitney 1982) และฮิลล์และคณะ (Hills et al 1988) ได้ดำเนินการไว้ นอกจากนี้ผู้วิจัยได้รูปแบบการเทียบมาตรฐานเชิงเส้นตรงมาใช้ศึกษาครั้งนี้โดยมีเหตุผล คือ

1. การเทียบมาตรฐานเชิงเส้นตรง เป็นวิธีการแปลงคะแนนที่ง่าย และสะดวกที่สุดที่จะใช้ (Kolen and Whitney 1982: 291; Petersen, Cook and Stocking 1983; Hills, Subhiyah and Hirsch 1988: 229)

2. การเทียบมาตรฐานเชิงเส้นตรง ให้ผลดีกว่ารูปแบบอื่นในสถานการณ์สอบวัดผลสัมฤทธิ์ (Petersen, Cook and Stocking 1983: 155; ภาวิณี ศรีสุขวัฒนานนท์ 2529: 155-170) เนื่องจากในการทดสอบเพื่อวัดผลสัมฤทธิ์ทางการเรียนตามหลักสูตรในชั้นเรียนมีขอบเขตเนื้อหาที่จำกัดในหลักสูตร แบบสอบชุดใหม่ที่สร้างขึ้นจะประกอบด้วยข้อสอบที่มีลักษณะต่าง ๆ ใกล้เคียงกับแบบสอบที่สร้างขึ้นชุดก่อน ๆ เมื่อแบบสอบวัดเนื้อหาเดียวกันวิธีการเทียบมาตรฐานเชิงเส้นตรงจึงให้ผลดีกว่าวิธีอื่น ซึ่งกลุ่มตัวอย่างเป็นกลุ่มสุ่มสมมูล 2 กลุ่ม (Kolen and Whitney 1982: 291)

3. การเทียบมาตรฐานเชิงเส้นตรงยังคงใช้ได้กับกลุ่มตัวอย่างที่มีขนาดเล็ก โดยเฉพาะการเทียบมาตรฐานตามแนวระนาบ (Kolen and Whitney 1982: 292) ครู-อาจารย์จึงสามารถนำไปปฏิบัติได้จริงในสภาพการเรียนการสอนปัจจุบัน ในชั้นเรียนซึ่งมีจำนวนนักเรียนไม่มาก

ดังนั้นในการวิจัยครั้งนี้ ผู้วิจัยจึงมีความสนใจที่จะศึกษาในเชิงเปรียบเทียบเกี่ยวกับคุณภาพของการเทียบมาตรฐานเชิงเส้นตรงเมื่อใช้แบบสอบร่วมภายใน 4 ขนาด คือ 10 15 20 และ 25 ข้อ ซึ่งคิดเป็นร้อยละ 16 25 33 และ 41 ของแบบสอบเทียบมาตรฐาน ในสถานการณ์สอบวัดผลสัมฤทธิ์ วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่สอง เพื่อให้ได้สารสนเทศเกี่ยวกับการใช้แบบสอบร่วมมากขึ้น และสร้างความสมบูรณ์ให้เกิดขึ้นกับระบบการสอบต่อไป

### วัตถุประสงค์ของการวิจัย

เพื่อตรวจสอบคุณภาพของการเทียบมาตรฐานเชิงเส้นตรง เมื่อใช้แบบสอบร่วมภายในที่มีช่วงความยาวต่างกัน 4 ขนาด คือ 10 15 20 และ 25 ข้อ โดยการเปรียบเทียบ

1. ความคลาดเคลื่อนมาตรฐานของการเทียบมาตรฐานเชิงเส้นตรง (Standard Error of Linear Equating: SEE)
2. ดัชนีความแตกต่าง (Discrepancy index: C) ซึ่งเป็นความคลาดเคลื่อนรวมที่ได้จากการวิเคราะห์กลุ่มสอบทานผล (Cross Validation groups)



## สมมติฐานของการวิจัย

การเทียบมาตรฐานโดยใช้แบบสอบร่วมกัน แบบสอบร่วมจะทำหน้าที่เชื่อมและปรับความแตกต่างอันเนื่องมาจากกลุ่มตัวอย่างหรือแบบสอบที่ต่างชุดกัน แบบสอบร่วมที่ยาวกว่าจึงมีส่วนทำให้การประมาณค่าเฉลี่ย และความแปรปรวนในแบบสอบเทียบมาตรฐานได้แม่นยำกว่า ประกอบกับการศึกษางานวิจัยเชิงประจักษ์ หลักและข้อเสนอแนะของนักวัดผลที่เกี่ยวกับความยาวของแบบสอบร่วม ตามที่ได้กล่าวมา ในส่วนของความเป็นมาและความสำคัญของปัญหาแล้วนั้น

ผู้วิจัยจึงคิดว่าแบบสอบร่วมขนาดตั้งแต่ 10 ข้อ เป็นจำนวนที่เพียงพอ และน่าจะอยู่ในเกณฑ์ที่ทำให้ผลการเทียบมาตรฐานมีคุณภาพอยู่ในระดับน่าพอใจ จึงกำหนดสมมติฐานในการวิจัยดังนี้

1. ความคลาดเคลื่อนมาตรฐานของการเทียบมาตรฐาน เมื่อใช้แบบสอบร่วมที่ยาวกว่าให้ค่าน้อยกว่าเมื่อใช้แบบสอบร่วมที่สั้นกว่า
2. ดัชนีความแตกต่างในการเทียบมาตรฐานเชิงเส้นตรงของแบบสอบร่วมตั้งแต่ 10 ข้อขึ้นไปนั้น น่าจะแตกต่างอย่างไม่มีนัยสำคัญทางสถิติ

## ขอบเขตของการวิจัย

1. การวิจัยครั้งนี้ เป็นการเทียบมาตรฐานโดยใช้แบบสอบร่วมที่มีความยาวต่างกัน 4 ขนาด กับแบบสอบเทียบมาตรฐานวัดผลสัมฤทธิ์ปลายภาควิชาคณิตศาสตร์ (ค203) ของกลุ่มโรงเรียนมัธยมศึกษาตอนต้น จังหวัดขอนแก่น กลุ่มที่ 2 ฉบับปีการศึกษา 2532 และ 2533 และเป็นแบบสอบเลือกตอบชนิด 4 ตัวเลือก
2. แบบสอบร่วมที่ใช้ในการเทียบมาตรฐาน เป็นแบบสอบวัดผลสัมฤทธิ์ที่มีความยาวต่างกัน 4 ขนาด คือ 10 15 20 และ 25 ข้อ และเป็นแบบสอบเลือกตอบชนิด 4 ตัวเลือก ที่ผู้วิจัยสร้างขึ้น
3. ประชากรที่ใช้ในการวิจัยเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 2 ปีการศึกษา 2534 สังกัดกรมสามัญศึกษา กลุ่มโรงเรียนมัธยมศึกษาตอนต้น จังหวัดขอนแก่น กลุ่มที่ 2 จำนวนทั้งสิ้น 3,018 คน
4. รูปแบบการเทียบมาตรฐานเป็นรูปแบบเชิงเส้นตรง (Linear Equating) และเทียบมาตรฐานตามแนวระนาบ (Horizontal Equating) จากแบบสอบเทียบมาตรฐานฉบับปีการศึกษา 2533 (ฉบับ Y) ไปยังแบบสอบเทียบมาตรฐานฉบับปีการศึกษา 2532 (ฉบับ X) โดยแบบสอบทั้งสองฉบับ มีการแจกแจงคะแนนที่คล้ายคลึงกัน

## 5. ตัวแปรที่ศึกษา

5.1 ตัวแปรต้น (Independent Variables) คือ ความยาวของแบบสอบรวมภายใน (Internal or Inclusive anchor test) มีสี่ขนาด คือ 10 15 20 และ 25 ข้อ

5.2 ตัวแปรตาม (Dependent Variables) คือ ค่าความคลาดเคลื่อนมาตรฐานของการเทียบมาตราและค่าดัชนีความแตกต่างในแต่ละความยาวของแบบสอบรวม

### ข้อจำกัดของการวิจัย

เนื่องจากการวิจัยครั้งนี้ อาจจะมีผลกระทบ จากการที่ต้องตัดข้อสอบรวมบางส่วน ออกจากแบบสอบฉบับ X หรือฉบับ Y ที่ผู้สอบแต่ละคนได้สอบเต็มฉบับ 85 ข้อ ทั้งนี้เพื่อให้ได้แบบสอบรวม 4 ขนาด ตามความยาวที่ต้องการศึกษา คือ 10 15 20 และ 25 ข้อ ดังนั้นผลการเทียบมาตราครั้งนี้ จึงอาจยังมีความคลาดเคลื่อนอยู่บ้าง

### ข้อตกลงเบื้องต้น

คะแนนที่ได้จากการเทียบมาตราเป็นคะแนนทดสอบที่ได้จากนักเรียนที่เข้าสอบทุกคน ตั้งใจทำข้อสอบอย่างเต็มความสามารถ

### คำจำกัดความที่ใช้ในการวิจัย

1. คุณภาพของวิธีการเทียบมาตรา หมายถึง ผลการเทียบมาตราที่พิจารณาจากค่าความคลาดเคลื่อนมาตรฐานในการเทียบมาตรา และค่าดัชนีความแตกต่างจากการวิเคราะห์กลุ่มสอบทานผล

2. การเทียบมาตราเชิงเส้นตรง (Linear equating) หมายถึง กระบวนการที่ใช้วิธีการทางสถิติเพื่อปรับคะแนนจากแบบสอบต่างชุด ที่มีโครงสร้างเนื้อหาเดียวกัน ให้สามารถเปรียบเทียบกันได้ โดยยึดหลักการแปลงให้เป็นคะแนนมาตรฐาน และใช้รูปแบบแบบสอบรวมเป็นตัวปรับความแตกต่างของกลุ่มตัวอย่างหรือแบบสอบ



3. แบบสอบเทียบมาตรา หมายถึง แบบสอบที่ใช้ทดสอบผลสัมฤทธิ์ปลายภาคของกลุ่มโรงเรียนมัธยมศึกษาตอนต้น จังหวัดขอนแก่น กลุ่มที่ 2 จำนวน 2 ฉบับ สร้างขึ้นตามหลักสูตรกระทรวงศึกษาธิการ พุทธศักราช 2521 ในวิชาคณิตศาสตร์ (ค203) เมื่อปีการศึกษา 2532 (ฉบับ X) และ 2533 (ฉบับ Y) ซึ่งมีโครงสร้างเนื้อหาเดียวกัน แต่ละฉบับประกอบด้วยข้อสอบจำนวน 60 ข้อ

4. แบบสอบร่วมภายใน (Internal or Inclusive anchor test) หมายถึง กลุ่มข้อสอบที่รวมอยู่ในแบบสอบเทียบมาตราทั้งสองฉบับ ซึ่งผู้วิจัยเป็นผู้สร้างตามหลักสูตรกระทรวงศึกษาธิการ พุทธศักราช 2521 วิชาคณิตศาสตร์ (ค203) ที่มีโครงสร้างเนื้อหาเป็นสัดส่วนกับแบบสอบเทียบมาตราทั้งสองชุด ประกอบด้วยความยาวสี่ขนาด คือ 10 15 20 และ 25 ข้อ

5. ความคลาดเคลื่อนมาตรฐานของการเทียบมาตรา (Standard error of equating: SEE) หมายถึง ค่าเบี่ยงเบนมาตรฐานของคะแนนแปลงในการเทียบมาตราเชิงเส้นตรง โดยคำนวณจากสูตรของลอร์ดซึ่งแองกอฟรวบรวมไว้ในรูปแบบที่ 3 (Angoff 1984: 106)

6. คะแนนสมมูล (equivalent scores) หมายถึงคะแนนระหว่างแบบสอบต่างชุดที่เทียบคะแนนกันได้ว่าเป็นคะแนนที่เท่าเทียมกัน

7. คะแนนเกณฑ์ หมายถึง คะแนนที่ได้จากแบบสอบฉบับแรก (ฉบับ X) ของกลุ่มสอบทานผล

8. กลุ่มตัวอย่างเทียบมาตรา (Equating Samples) หมายถึง กลุ่มตัวอย่างที่ได้รับการทดสอบจากแบบสอบเทียบมาตราฉบับใดฉบับหนึ่งและแบบสอบร่วม เพื่อนำผลการสอบไปใช้ในการแปลงคะแนนสำหรับสร้างคะแนนสมมูลในการเทียบมาตราเชิงเส้นตรง

9. กลุ่มสอบทานผล (Cross Validation Group) หมายถึง กลุ่มตัวอย่างที่สุ่มมาจากประชากรเดียวกันกับกลุ่มตัวอย่างเทียบมาตรา โดยไม่มีหน่วยตัวอย่างซ้ำกันเลย ได้รับการสอบจากแบบสอบเทียบมาตราทั้ง 2 ฉบับ และแบบสอบร่วม เพื่อนำผลการสอบไปใช้ในการตรวจสอบหาค่าดัชนีความแตกต่าง

10. ตารางเทียบคะแนน หมายถึง ตารางแสดงตัวเลขคะแนนที่สมมูลกัน (Equivalent scores) ระหว่างแบบสอบต่างชุด

11. ประสิทธิภาพสัมพัทธ์ (Relative efficiency: RE) หมายถึง อัตราส่วนร้อยละของค่าความคลาดเคลื่อนมาตรฐานของการเทียบมาตรา ที่เกิดจากการใช้แบบสอบร่วมขนาด 10 15 หรือ 20 ข้อ ต่อความคลาดเคลื่อนมาตรฐานของการเทียบมาตราที่ใช้แบบสอบร่วม 25 ข้อ

12. ดัชนีความแตกต่าง (discrepancy index: C) หมายถึง ค่าความคลาดเคลื่อนรวม (Total error) ในการเทียบมาตรฐานที่ได้จากกลุ่มสอบทานผล ซึ่งคำนวณจากค่าเฉลี่ยกำลังสองของความแตกต่างระหว่างคะแนนสมมูล ซึ่งได้จากการนำคะแนนของแบบสอบฉบับ Y ไปแปลงคะแนนจากตารางเทียบมาตรฐานกับคะแนนเกณฑ์ แล้วถ่วงน้ำหนักด้วยความแปรปรวนของคะแนนเกณฑ์ จึงเป็นค่ามาตรฐานที่สามารถเปรียบเทียบกันได้โดยตรง เพื่อประเมินคุณภาพในการเทียบมาตรฐาน

13. ความเพียงพอในการเทียบมาตรฐาน (adequacy of Equating) หมายถึง ระดับคุณภาพของวิธีการเทียบมาตรฐานเชิงเส้นตรง ซึ่งประเมินจากค่าดัชนีความแตกต่าง โดยใช้เกณฑ์ของปีเตอร์เซนและคณะ (Petersen et al 1982: 93-94)

### ประโยชน์ที่คาดว่าจะได้รับ

1. ได้ทราบถึงเทคนิควิธีการเทียบมาตรฐาน เพื่อนำไปใช้ปรับคะแนนของนักเรียนที่สอบแบบสอบเนื้อหาเดียวกันคนละฉบับในแต่ละโรงเรียน กลุ่มโรงเรียนหรือเขตการศึกษา ได้อย่างมีประสิทธิภาพและความหมาย และนำไปใช้ในงานวัดและประเมินผลของโรงเรียน
2. ได้สารสนเทศเกี่ยวกับจำนวนข้อสอบร่วมที่ทำให้การเทียบมาตรฐานมีคุณภาพ เพื่อนำไปสู่การสร้างข้อสอบชุดใด ๆ ให้มีสัดส่วนของแบบสอบร่วมเท่าที่จำเป็นเพื่อการเปรียบเทียบผลของนักเรียนได้แน่ชัดขึ้น