



การดำเนินการวิจัย

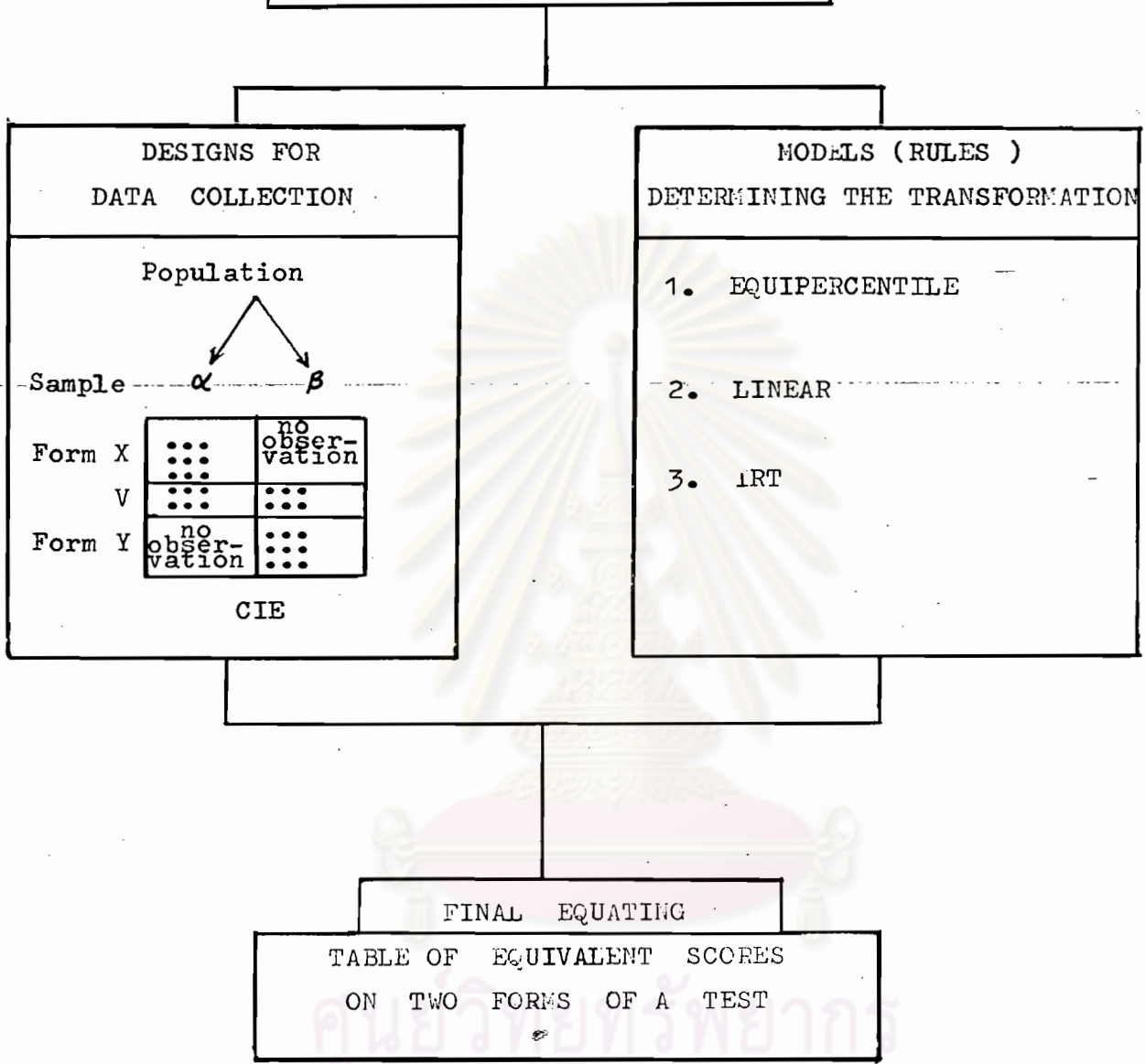
กระบวนการเทียบมาตรฐานคะแนนของแบบสอบ มีองค์ประกอบที่เป็นลักษณะเฉพาะของงาน คือ องค์ประกอบที่ว่าด้วยการออกแบบเพื่อรวบรวมข้อมูลคะแนนสอบที่จะนำมาใช้ระบุคะแนนแปลง และอีกองค์ประกอบหนึ่ง คือ การเลือกสรรรูปแบบการเทียบมาตรฐานเพื่อกำหนดเกณฑ์การแปลงคะแนน จากมาตรฐานของแบบสอบชุดหนึ่ง ไปสู่แบบสอบอีกชุดหนึ่ง สำหรับการศึกษาครั้งนี้ ได้กำหนดตัวแปรในแต่ละองค์ประกอบที่ต่องการศึกษา ดังนี้

๑. การออกแบบเพื่อรวบรวมข้อมูล กำหนดให้ใช้แบบสอบร่วมชนิดแบบสอบร่วมภายใน (Internal anchor test) กับกลุ่มตัวอย่างสองกลุ่ม ความยาวของแบบสอบร่วมแปรผันไป ๓ ขนาด คือ มีจำนวนข้อในแบบสอบร่วมคิดเป็นร้อยละ ๖๐ ๔๐ และ ๒๐ ของแบบสอบเทียบมาตรฐาน และเรียกว่า V60 V40 และ V20 ตามลำดับ

๒. รูปแบบการเทียบมาตรฐาน กำหนดไว้ ๓ รูปแบบ คือ
- ก. รูปแบบการเทียบมาตรฐานอิกวิเปอร์เซนไทล์
 - ข. รูปแบบการเทียบมาตรฐานเชิงเส้นตรง
 - ค. รูปแบบการเทียบมาตรฐานอิงทฤษฎีการทดสอบแบบสามพารามิเตอร์

ผลสุดท้ายของการเทียบมาตรฐานแบบสอบ คือ ตารางคะแนนสมมูลที่เทียบจากมาตรฐานคะแนนของแบบสอบชุด Y ไปยังมาตรฐานคะแนนของแบบสอบชุด X กระบวนการเทียบมาตรฐาน สรุปได้ดังแผนภาพที่ ๑ ส่วนรายละเอียดอีกได้เสนอในลำดับต่อไป

METHODOLOGY OF EQUATING



แผนภาพที่ • แสดงกระบวนการเทียบมาตรฐานคะแนนแบบสอบสองชุด

ศูนย์วิจัยทั้พยาบาล
จุฬาลงกรณ์มหาวิทยาลัย

การออกแบบการเทียบมาตรฐาน

๑. สถานการณ์การทดสอบ

การศึกษาครั้งนี้ได้กำหนดการใช้ข้อมูลคะแนน และแบบสอบจากสถานการณ์การทดสอบจริง ๒ สถานการณ์ คือ การสอบคัดเลือกเพื่อเข้าทำงาน และการสอบวัดผลสัมฤทธิ์ทางการเรียนตามระบบการจัดการเรียนการสอนในระดับปริญญาตรี ในการเลือกกรณีเพื่อทำการศึกษาในครั้งนี้ใช้หลักการพิจารณา ดังนี้

๑.๑ ประเภทของแบบสอบ การสอบคัดเลือกใช้กรณีแบบสอบที่วัดความสามารถทั่วไป ซึ่งเป็นเกณฑ์ในการตัดสินความสามารถของผู้เข้าสอบขั้นต้น ส่วนในสถานการณ์การสอบวัดผลสัมฤทธิ์ทางการเรียน ใช้กรณีการสอบวัดวิชาที่ได้เปิดให้นักศึกษาลงทะเบียนเรียน นักศึกษาได้ผ่านกระบวนการเรียนการสอนตามระบบของสถาบัน และรับการสอบวัดความรู้ตามวัตถุประสงค์ของวิชานั้น ๆ เพื่อรับการพิจารณาให้ผ่านหรือไม่ผ่านในรายวิชานั้น

๑.๒ คุณภาพของแบบสอบ เนื่องจากการเทียบมาตรฐานคะแนน เป็นกระบวนการสุดท้ายในการรายงานผลการสอบวัดที่มุ่งให้ตัวเลขของการรายงานมีคุณสมบัติเปรียบเทียบกันได้โดยตรง คุณภาพของแบบสอบจึงควรมีระดับสูงพอสมควร เพื่อให้ผลการสร้างคะแนนสมมูลจากการใช้วิธีการเทียบมาตรฐานที่ต่อการศึกษาในครั้งนี้อยู่ในระดับที่น่าพอใจ แต่เนื่องจากการเลือกแบบสอบที่นำมาใช้ช่วยการใช้สถานการณ์จริง ผู้วิจัยไม่สามารถทำการศึกษาแบบสอบก่อนตัดสินใจ จึงใช้หลักการพิจารณาจากปัจจัยแวดล้อม ดังนี้

ก. สถาบันที่เป็นเจ้าของแบบสอบมีโปรแกรมการทดสอบในกรณีที่ผู้วิจัยต้องการ และได้มีการดำเนินการต่อเนื่องมานานพอสมควร และยังคงมีการดำเนินการต่อไป

ข. การพัฒนาแบบสอบของสถาบัน เป็นระบบที่มีความพร้อมทั้งด้านบุคลากรและงบประมาณ มีแผนการดำเนินงานที่ชัดเจน

ค. สถาบันนั้น ๆ มีขีดความสามารถในการให้การสนับสนุนความก้าวหน้าในด้านการทดสอบสูง สามารถให้การบริการข้อมูลเพื่อการวิจัยได้

๑.๓ กรณีการทดสอบที่จะนำมาใช้ในการศึกษาค้างนี้ ต้องมีจำนวนผู้สอบไม่น้อยกว่า ๕๐๐๐ คน ทั้งนี้เนื่องจากการประยุกต์ทฤษฎีการวัดความสามารถแฝง ต้องอาศัยข้อมูลจำนวนมาก

๒. ประชากรและกลุ่มตัวอย่าง

กรณีแบบสอบที่เลือกจากสถานการณการทดสอบจริง ๒ กรณี คือ กรณีแบบสอบคัดเลือก และกรณีแบบสอบวัดผลสัมฤทธิ์ทางการ เรียนตามหลักสูตรระดับปริญญาตรี ประชากรในกรณีแบบสอบคัดเลือกประกอบด้วย ผลการสอบรายบุคคลที่ได้รับการทดสอบพร้อมกันเมื่อปี ๑๕๕๔ จำนวน ๑๕๕๓๕ คน ประชากรในกรณีแบบสอบวัดผลสัมฤทธิ์ประกอบด้วย ผลการสอบรายบุคคลที่ไ้ลงทะเบียนเรียนในรายวิชานั้น และได้รับการทดสอบเมื่อสิ้นภาคการศึกษาในปี ๑๕๕๔ เช่นกัน มีจำนวน ๑๒๓๔๓ คน การสุ่มตัวอย่างในแต่ละกรณีใช้ระเบียบวิธีเดียวกัน คือ ทำการสุ่มอย่างง่าย และสุ่มอย่างไม่ใส่คืน (Random without replacement) อาศัยคำสั่ง SAMPLE ของโปรแกรมสำเร็จรูป SPSSX สุ่มโดยใช้รหัสเลขสุ่ม (Set seed) ในแต่ละกรณี สุ่มจำนวน ๕๕๐๐ คน จากนั้นได้ทำการสุ่มอย่างมีระบบจากจำนวน ๕๕๐๐ คน เพื่อกำหนดหน่วยให้อยู่ในกลุ่ม α และกลุ่ม β ซึ่งเป็นกลุ่มตัวอย่างเทียบมาตรา และกลุ่ม γ ซึ่งเป็นกลุ่มตัวอย่างสอบทานผล จำนวนกลุ่มละ ๑๕๐๐ คน

๓. แบบสอบ

แบบสอบที่เลือกมาในแต่ละกรณี เป็นแบบสอบที่ได้รับการสร้าง และพัฒนาขึ้นอย่างมีระบบ และยังคงใช้อยู่ในโครงการต่อเนื่องของสถาบัน แต่ละกรณีมีลักษณะเฉพาะ ดังนี้

๓.๑ กรณีแบบสอบคัดเลือก คือ วิชาความสามารถทั่วไป ประกอบด้วยข้อสอบชนิดสี่ตัวเลือก จำนวน ๑๐๐ ข้อ วัตถุประสงค์ประกอบต่าง ๆ รวม ๕ องค์ประกอบ ใช้เวลาในการทดสอบ ๒ ชั่วโมง

๓.๒ กรณีแบบสอบวัดผลสัมฤทธิ์ แบบสอบที่นำมาศึกษา คือ แบบสอบวิชาสถิติทวิจัย และวัดผลการศึกษา ตามหลักสูตรระดับปริญญาตรี ประกอบด้วยข้อสอบชนิด ๕ ตัวเลือก จำนวน ๑๐๐ ข้อ วัตถุประสงค์ประกอบ ๓ องค์ประกอบ ใช้เวลาในการทดสอบ ๓ ชั่วโมง การทดสอบได้

ดำเนินการตามระเบียบการประเมินผลรายวิชา

๔. การจัดเตรียมแบบสอบเพื่อทำการเทียบมาตรา

ปกติการสร้างแบบสอบชุดใหม่ มักเกิดจากความต้องการรักษาความลับของแบบสอบ เพื่อให้เกิดความยุติธรรม หลักสูตรที่ใช้กำหนดข้อสอบหรือตารางกำหนดลักษณะเฉพาะของข้อสอบ มักไม่มีการเปลี่ยนแปลงมากนัก หรืออาจไม่มีการเปลี่ยนแปลงเลย ผลจากการสอบครั้งหลังจึงนำไปเทียบกับมาตราแบบสอบชุดก่อน เพื่อให้สามารถทำการเปรียบเทียบในเชิงตัวเลขได้ ในการศึกษาครั้งนี้ ใ้จัดให้แบบสอบแต่ละกรณีสองชุด สมมุติให้ชุด X เป็นชุดเก่า และชุด Y เป็นชุดใหม่ การเทียบมาตราจึงเทียบจากแบบสอบ Y ไปหาแบบสอบ X แบบสอบชุด X และ Y ของแต่ละกรณีได้จากการจัดสรรข้อสอบที่มีอยู่ทั้งหมด ๑๐๐ ข้อ ให้เกิดเป็นแบบสอบชุด X ๕๕ ข้อ ชุด Y ๔๕ ข้อ โดยให้ครอบคลุมเนื้อหาที่วัดคล้ายกันมีความยากโดยเฉลี่ยใกล้เคียงกันโดยประมาณ สำหรับแบบสอบรวมของแต่ละกรณีมีลักษณะคล้ายกับชุด X และ Y โดยเลือกจาก ๓๐ ข้อ ที่เหลือ การจัดเตรียมแบบสอบเพื่อการวิจัยในทำนองนี้ได้พบในงานวิจัยหลายชิ้น (เช่น Levine 1955 cited by Klein and Kolen 1985; Petersen Marco and Stewart 1982) ขั้นตอนการดำเนินการจัดชุดแบบสอบ X และ Y และแบบสอบรวมในแต่ละกรณีของแบบสอบได้ดำเนินการ ดังนี้

๔.๑ วิเคราะห์ข้อสอบของแบบสอบแต่ละกรณีด้วยการใช้คำสั่ง RELIABILITY

ของโปรแกรมสำเร็จรูป SPSSX แล้วนำค่าสถิติของข้อสอบรายข้อแจกแจงลงในตาราง ๒ ทาง คือ ค่าเฉลี่ยรายข้อทางหนึ่ง ส่วนอีกทางหนึ่งเป็นค่าสหสัมพันธ์รายข้อกับแบบสอบทั้งฉบับ โดยแบ่งการแจกแจงตามองค์ประกอบของเนื้อหาของแบบสอบ

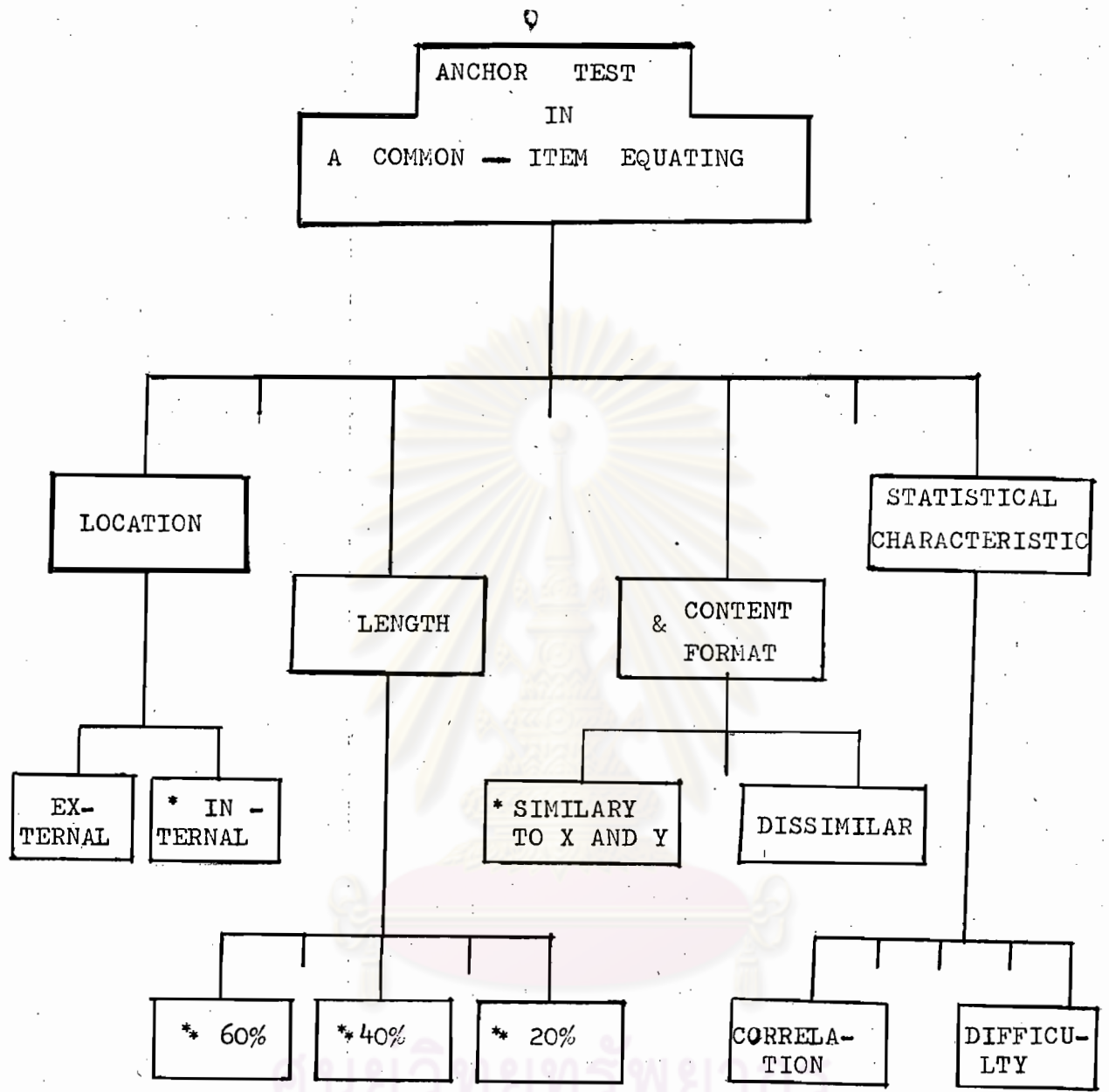
๔.๒ ในแต่ละองค์ประกอบ จัดกลุ่มข้อสอบให้ข้อที่มีค่าเฉลี่ย และค่าสหสัมพันธ์

รายข้อกับแบบสอบทั้งฉบับที่มีค่าใกล้เคียงกันอยู่ในกลุ่มเดียวกัน จากนั้นสุ่มข้อสอบในแต่ละกลุ่มให้เป็นข้อสอบของแบบสอบชุด X และชุด Y และส่วนที่เหลือ ทำเช่นนี้กับทุกกลุ่มของข้อสอบ ผลสุดท้ายของการจัดข้อสอบรายข้อเข้าชุดนั้น คือ ใ้แบบสอบชุด X และชุด Y ชุดละ ๕๕ ข้อ และส่วนที่เหลือ ๓๐ ข้อ ทั้ง ๓ ส่วนนี้มีคุณสมบัติเป็นแบบสอบที่ทำหน้าทีวัดความสามารถที่คล้ายคลึงกัน

๔.๓ การนำแบบสอบรวมเข้ามศึกษาตามแผนของการออกแบบการเทียบมาตรา
 ค่ายการใช้แบบสอบรวม (a common - item equating: CIE) ในครั้งนี้มีตัวแปรที่การ
 วิจัยครั้งนี้กำหนด และตัวแปรที่ไม่ได้มีการควบคุมตั้งเสนอในแผนภาพที่ ๒ ตัวแปรด้านตำแหน่งของ
 แบบสอบรวม (Location) ให้เป็นชนิดภายใน (Internal anchor test) ตัวแปร
 ด้านเนื้อหา และรูปแบบคำถามกำหนดให้มีลักษณะคล้ายกันกับแบบสอบเทียบมาตรา X และ Y ซึ่งได้
 มีขั้นตอนการจัดเตรียมดังกล่าวในลำดับต่อไป สำหรับตัวแปรเกี่ยวกับคุณลักษณะทางสถิติของข้อสอบ
 และแบบสอบไม่ได้อยู่ในขอบเขตของการศึกษา แต่ใช้วิธีการสุ่มข้อสอบเพื่อให้มีความคล้ายคลึงกัน
 ส่วนตัวแปรที่ว่าด้วยความยาวของแบบสอบรวม เป็นตัวแปรที่ของการศึกษาในครั้งนี้ ได้กำหนดให้
 เป็นร้อยละ ๒๐, ๔๐ และ ๖๐ ของแบบสอบเทียบมาตราซึ่งมีชุดละ ๕๕ ข้อ การจัดเตรียมแบบสอบ
 รวมมีขั้นตอน ดังนี้ จากกลุ่มข้อสอบที่เหลือ ๓๓ ข้อ ซึ่งมีองค์ประกอบของเนื้อหาที่ทำการวัด เช่น
 เกี่ยวกับฉบับเดิม นั้น นำมาสุ่มตามสัดส่วนของเนื้อหาทุกองค์ประกอบ จำนวนที่สุ่มมี ๗ ข้อ ซึ่งคิด
 เป็นร้อยละ ๒๐ ของแบบสอบเทียบมาตรา กำหนดให้เป็นแบบสอบรวม V20 ต่อไปสุ่มอีก ๗ ข้อ
 จากส่วนที่เหลือในทำนองเกี่ยวกับการสุ่ม V20 ข้อสอบ ๗ ข้อ ที่ได้นำมารวมกับ V20 รวมเป็น
 ข้อสอบ ๑๔ ข้อ ซึ่งคิดเป็นร้อยละ ๔๐ ของแบบสอบเทียบมาตรา กำหนดให้เป็นแบบสอบรวม V40
 จากนั้นสุ่มอีก ๗ ข้อ รวมเข้ากับ ๑๔ ข้อแรก รวมเป็น ๒๑ ข้อ ซึ่งคิดเป็นร้อยละ ๖๐ ของแบบสอบ
 เทียบมาตรา กำหนดให้เป็นแบบสอบรวม V60 ด้วยวิธีดังกล่าวแบบสอบรวมที่มีขนาดสั้นกว่าเป็น
 ขั้วชี้ของแบบสอบรวมที่มีขนาดยาวกว่า โดยที่ทุกขนาดมีองค์ประกอบของเนื้อหาที่ทำการวัดเหมือนกัน

๕. การจัดชุดแบบสอบรวมเข้ารวมกับแบบสอบเทียบมาตรา

การจัดชุดแบบสอบรวมเข้ารวมกับแบบสอบเทียบมาตราชุด X และชุด Y ตามการออกแบบ
 แผนการใช้แบบสอบรวมชนิดแบบสอบรวมภายใน (Internal anchor test) ได้แบบสอบที่
 นำมาศึกษากรณีละ ๓ ชุด ให้สัญลักษณ์แบบสอบ ดังนี้ XV20-YV20 XV40-YV40 และ
 XV60-YV60



* ตัวแปรที่ไคควบคุม

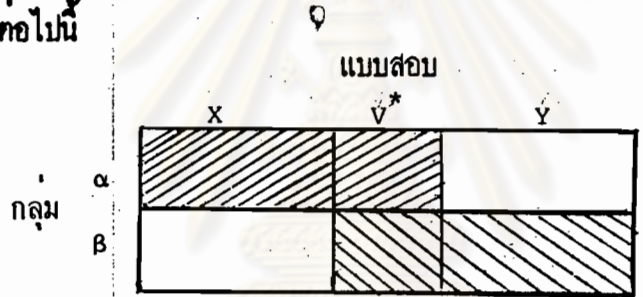
* ตัวแปรที่ทำการศึกษา

แผนภาพที่ ๒ ตัวแปรที่เกี่ยวข้องกับแบบสอบรวม

๖. การรวบรวมข้อมูล

ข้อมูลเป็นประเภทคุณภาพ หลังจากศึกษาขออนุญาตจากหน่วยงานเจ้าของข้อมูลแล้ว ผู้วิจัยได้รับข้อมูลที่บันทึกลงในเทปในลักษณะผลการตอบรายชื่อของผู้สอบทั้งประชากร จากนั้นได้นำมาสุ่มตัวอย่าง และประมวลให้เป็นผลการตอบของรายคนต่อแบบสอบถามแต่ละชุดตามที่ได้ออกแบบไว้ในข้อ ๓ ในกลุ่มตัวอย่าง α และ β ดังนี้

การศึกษารูปแบบการเทียบมาตรการด้วยการออกแบบการใช้แบบสอบถาม กำหนดให้กลุ่มตัวอย่าง α เป็นกลุ่มที่ได้รับการทดสอบด้วยชุด X และกลุ่ม β เป็นกลุ่มที่ได้รับการทดสอบด้วยชุด Y และกลุ่มตัวอย่างทั้งสองจะได้รับการทดสอบด้วยแบบสอบถามรวม V เหมือนกัน จากการจำลองสถานการณ์ด้วยการจัดกระทำให้เกิดแบบสอบถาม X Y และ V ดังได้กล่าวมาแล้วนั้น การรวบรวมข้อมูลจะนำเสนอในตารางต่อไปนี้



สัญลักษณ์ที่ใช้

- V^* คือ แบบสอบถามที่มีขนาดผันแปรตามที่กำหนด
- n_α , n_β คือ จำนวนข้อมูลคะแนนตามกลุ่มตัวอย่าง α และ β ตามลำดับ ซึ่งมีจำนวนกลุ่มละ ๑,๕๐๐ ราย
- n_t คือ กลุ่มตัวอย่างรวมเท่ากับ $n_\alpha + n_\beta$

ระเบียบวิธีการเทียบมาตรา

ระเบียบวิธีการเทียบมาตราที่ใช้ในการศึกษาค้างนี้ เป็นระเบียบวิธีเดียวกันที่ใช้ศึกษากับกรณีแบบสองทั้งสองกรณี คือ กรณีการสอบคัดเลือก และกรณีการวัดผลสัมฤทธิ์ ทั้งนี้รายละเอียดของแต่ละรูปแบบการเทียบมาตราต่อไปนี้

๑. รูปแบบการเทียบมาตราอิกิวเปอร์เซนไทล์

ภายหลังการประมวลข้อมูลคะแนนตามการออกแบบไปแล้ว นำมาเทียบมาตราตามขั้นตอน ดังนี้

๑) ทำตารางประมาณการแจกแจงความถี่ของกลุ่มตัวอย่างรวม t โดย

ก. ทำตารางแจกแจงคะแนนแบบสองรวม v ของกลุ่ม t โดยมีจำนวนคนในการแจกแจง (n_t) เท่ากับ ๓,๐๐๐ คน การแจกแจงความถี่อาศัยโปรแกรมสำเร็จรูป SPSSX คำสั่ง CROSSTAB

ข. หาสัดส่วนความถี่ในแต่ละชั้นคะแนนของ v ในกลุ่มตัวอย่างรวม t เทียบกับความถี่ของชั้นคะแนนเดียวกันของ v ในกลุ่มตัวอย่าง α และ β ตามลำดับ ($f_{it}/f_{i\alpha}$ และ $f_{it}/f_{i\beta}$)

ค. ทำตารางแจกแจงความถี่ 2 ทางของกลุ่มตัวอย่าง ที่ทำแบบสอบชุด x ทางหนึ่งเป็นการแจกแจงคะแนน x ส่วนอีกทางหนึ่งเป็นการแจกแจงตามคะแนน v และทำการวางในทำนองเดียวกันกับกลุ่ม β

ง. ประมาณความถี่ของกลุ่มตัวอย่างรวม t ในแบบสอบชุด x และ y โดยการปรับขยายความถี่แต่ละชั้นของคะแนน x ในกลุ่ม α และ y ในกลุ่ม β ด้วยสัดส่วนของ $f_{it}/f_{i\alpha}$ และ $f_{it}/f_{i\beta}$ กับแต่ละคะแนน v ตามลำดับ ผลรวมของผลคูณความถี่แต่ละ v กับสัดส่วน เป็นความถี่ประมาณของรายคะแนน x และ y ของกลุ่มตัวอย่างรวม t

๒) จากตารางความถี่ประมาณของกลุ่มตัวอย่างรวม t ในแบบสอบชุด x และชุด y นำมาหาตำแหน่ง เปอร์เซนไทล์ของแต่ละชั้นคะแนนของแต่ละชุด อาศัยการคำนวณด้วยคอมพิวเตอร์

ใช้โปรแกรมฟอร์มแทนที่ผู้วิจัยเขียนขึ้น (ในภาคผนวก ค.)

๓) เขียนกราฟแสดงความสัมพันธ์ของคะแนน x กับตำแหน่งเปอร์เซ็นต์ไทล์ โดยใช้กระดาษกราฟชนิดมีมาตราแกนทั้งเป็นค่าความน่าจะเป็น (Arithmetic probability graph) และเขียนกราฟของคะแนน y ลงบนกระดาษเดียวกัน เพื่อแสดงคะแนนสมมูลของ x กับ y ด้วยการอ่านค่าที่เปอร์เซ็นต์ไทล์เดียวกัน การลากเส้นกราฟให้อาศัยการเกลาด้วยมือ โดยใช้กระดูกงูช่วย ถึอหลักการลากเส้นให้ผ่านจุดมากที่สุด ในช่วงที่จุดต่าง ๆ ไม่ได้อยู่ในแนวเดียวกัน ให้นำเส้นลากให้ผ่านจุดต่าง ๆ ให้มีความใกล้เคียงกันทั้งสองข้าง

๔) ทำตารางคะแนนสมมูลของ x และ y โดยกำหนดให้คะแนน y เป็นหลักแล้วเทียบไปหาคะแนน x โดยอ่านจากกราฟซึ่งเป็นวิธีการเทียบเชิงเส้นตรง (lineary interpolated relative cummulative frequency distribution) (Angoff 1984: 11-12; Kolen 1984)

๒. รูปแบบการเทียบมาตราเชิงเส้นตรง

ตามที่ได้ออกแบบให้กลุ่มตัวอย่างผู้สอบมาจากการสุ่มอย่างง่าย และให้กลุ่ม α ทำชุด x และ V ส่วนกลุ่ม β ทำฟอร์ม y และ V การประมาณค่าประชากรต่าง ๆ เพื่อนำมาแทนลงในสมการต้นแบบตามนियามการเทียบมาตราเชิงเส้นตรง

$$x^* = e_x(y) = AY + B \text{ ----- (๑)}$$

กระทำด้วยวิธีการประมาณด้วยความเป็นไปได้สูงสุด (Maximum likelihood method) (Lord 1955; Angoff 1984: 105) ชั้นตอนต่าง ๆ มีดังนี้

๑) สัญลักษณ์ที่เกี่ยวข้องมีดังนี้

$$\hat{\mu}_x \quad \hat{\sigma}_x^2$$

คือ ค่าประมาณคะแนนเฉลี่ยและความแปรปรวนของกลุ่มตัวอย่างรวมที่ทำแบบสอบชุด x

$\hat{\mu}_y$	$\hat{\sigma}_y^2$	คือ ค่าประมาณคะแนนเฉลี่ย และความแปรปรวนของ <u>กลุ่มตัวอย่างรวม</u> ที่ทำแบบสอบถาม <u>Y</u>
$M_{y\alpha}$	$S_{y\alpha}^2$	คือ ค่าเฉลี่ยและความแปรปรวนของคะแนนชุด <u>X</u> จาก <u>กลุ่ม α</u>
$M_{y\beta}$	$S_{y\beta}^2$	คือ ค่าเฉลี่ยและความแปรปรวนของคะแนนชุด <u>Y</u> จาก <u>กลุ่ม β</u>
$M_{v\alpha}$	$S_{v\alpha}^2$	คือ ค่าเฉลี่ยและความแปรปรวนของคะแนนแบบสอบถาม <u>V</u> จาก <u>กลุ่ม α</u>
$M_{v\beta}$	$S_{v\beta}^2$	คือ ค่าเฉลี่ยและความแปรปรวนของคะแนนแบบสอบถาม <u>V</u> จาก <u>กลุ่ม β</u>
M_{vt}	S_{vt}^2	คือ ค่าเฉลี่ยและความแปรปรวนของแบบสอบถาม <u>V</u> จาก <u>กลุ่ม t</u>
$b_{xv\alpha}$		คือ สัมประสิทธิ์ถดถอย <u>X</u> จาก <u>V</u> ใน <u>กลุ่มตัวอย่าง α</u>
$b_{yv\beta}$		คือ สัมประสิทธิ์ถดถอย <u>Y</u> จาก <u>V</u> ใน <u>กลุ่มตัวอย่าง β</u>

๒) หากค่าประมาณต่าง ๆ ตามสูตรต่อไปนี้

$$\hat{\mu}_x = M_{x\alpha} + b_{xv\alpha} (\hat{\mu}_v - M_{v\alpha}) \quad \text{----- (๒)}$$

$$\hat{\sigma}_x^2 = S_{x\alpha}^2 + b_{xv\alpha}^2 (\hat{\sigma}_v^2 - S_{v\alpha}^2) \quad \text{----- (๓)}$$

$$\hat{\mu}_y = M_{y\beta} + b_{yv\beta} (\hat{\mu}_v - M_{v\beta}) \quad \text{----- (๔)}$$

$$\hat{\sigma}_y^2 = S_{y\beta}^2 + b_{yv\beta}^2 (\hat{\sigma}_v^2 - S_{v\beta}^2) \quad \text{----- (๕)}$$

๓) หากค่าประมาณความชัน A และค่าคงที่ B จากสูตรต่อไปนี้ แล้วแทนค่าประมาณประชากรต่าง ๆ ลงในสมการเส้นตรง

$$A_{xy} = \frac{b_{yv\beta} [S_{x\alpha}^2 + b_{xv\alpha}^2 (\sigma_v^2 - S_{v\alpha}^2)]}{b_{xv\alpha} [S_{y\beta}^2 + b_{yv\beta}^2 (\sigma_v^2 - S_{\beta v}^2)]}$$

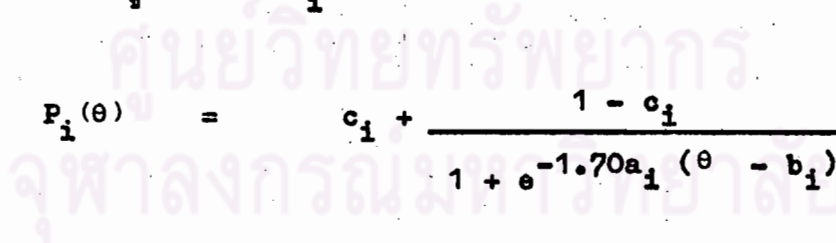
$$B_{xy} = \hat{\mu}_x - A_{xy} \hat{\mu}_y \text{ ----- (b)}$$

๔) ใช้สมการสร้างตารางเทียบคะแนนจากแบบสอบ y ไปยังคะแนนตามมาตราของแบบสอบ x ซึ่งจะได้ตารางแปลงทั้งหมด b ตาราง

๓. รูปแบบการเทียบมาตราอิงทฤษฎีการตอบสนองแบบสามพารามิเตอร์

ตามข้อตกลงของทฤษฎีการตอบสนองแบบสามพารามิเตอร์ที่ว่า มีฟังก์ชันทางคณิตศาสตร์ซึ่งสัมพันธ์กับความน่าจะเป็นของผู้สอบในอันที่จะตอบข้อสอบข้อหนึ่ง ๆ ใดถูกต้องนั้น รูปแบบทางคณิตศาสตร์ที่เลือกมาศึกษาในครั้งนี้ คือ โลจิสติกโมเดลแบบสามพารามิเตอร์ (three - parameter logistic model) ในรูปแบบนี้ θ หมายถึง รัศมีความสามารถของผู้สอบความน่าจะเป็นของการตอบข้อสอบข้อที่ i ถูกต้อง คือ $P_i(\theta)$ โดยที่

$$P_i(\theta) = c_i + \frac{1 - c_i}{1 + e^{-1.70a_i(\theta - b_i)}}$$



เมื่อ	a_i	b_i	และ	c_i	เป็นพารามิเตอร์ที่บรรยายข้อสอบข้อที่ i
	a_i				เป็นสัดส่วนของความชันของ $P_i(\theta)$ ที่จุดเปลี่ยน โค้ง ให้ความหมายว่า เป็นอำนาจจำแนกของข้อสอบ
		b_i			เป็นจุดแกมมาครา ทรงจุดเปลี่ยนโค้งของ $P_i(\theta)$ มีความหมายว่า เป็นความยากของข้อสอบ
	c_i				เป็นแอสิมโทททางต่ำ (lower asymptote) ของ $P_i(\theta)$ ให้ความหมายว่า เป็นค่าการคาด สมมุติ (pseudo - guessing parameter)

การดำเนินการเทียบมาตราความรูปร่างนี้ ประกอบด้วยสองขั้นตอนใหญ่ ๆ คือ
ขั้นการประมาณค่าคะแนนจริง และขั้นการเทียบมาตรา ซึ่งจะกล่าวต่อไปดังนี้

๑) ขั้นการประมาณคะแนนจริง

ก. การประมาณค่าพารามิเตอร์ของข้อสอบ และความสามารถของข้อสอบ

กระทำโดยใช้โปรแกรมสำเร็จรูป LOGIST 5 (Wingersky, Barton and Lord 1982)
ค่าประมาณได้มาโดยกระบวนการประมาณตามวิธีความเป็นไปได้สูงสุด (maximum likelihood
procedure) โดยได้ดำเนินการตามขั้นตอนที่กล่าวไว้ในคู่มือ ดังนี้

LOGIST 5 กำหนดให้ป้อนข้อมูลที่ เป็นผลการตอบรายข้อของกลุ่มตัวอย่าง
ผู้สอบ โดยแปลงเป็นรหัสของการตอบถูกต้อง (๑) ไม่ถูกต้อง (๐) เว้นไม่ตอบ (๒) และ
ทำไม่ถึง (๓) ขั้นตอนการประมาณแบ่งเป็น ๔ ขั้นตอน (steps) การทำงานในแต่ละขั้นเป็นดังนี้

ขั้นที่ ๑ ประมาณความสามารถ θ และค่า b โดยให้ a และ c คงที่

ขั้นที่ ๒ ประมาณค่า a b และ c โดยให้ θ คงที่

ขั้นที่ ๓ ประมาณค่า θ และค่า b โดยให้ a และ c คงที่

ขั้นที่ ๔ ประมาณค่า a b และ c โดยให้ θ คงที่

ในแต่ละขั้นได้แตกเป็นชั้นย่อยหลายชั้น (stages) การทำงานจะจบ
แต่ละขั้น (step) เมื่อการเพิ่มขึ้นของค่าฟังก์ชันเกณฑ์ (criterion function) ระหว่าง

๒. ขั้นตอนน้อยกว่าอัตราส่วนร้อยละที่กำหนดไว้สำหรับแต่ละชั้น การสิ้นสุดการทำงานเป็นไปตามเงื่อนไข ๓ ประการ คือ เมื่อการประมาณค่าได้ครบตามขั้นตอนทั้ง ๔ (steps) หรือเมื่อการประมาณค่าได้ดำเนินไปจนครบ ๒๐ ขั้นตอน (stages) หรือการประมาณค่าล้มเหลวด้วยสาเหตุของลักษณะข้อมูลคะแนนที่ทำให้เกิดการติดลบในการคำนวณค่าอนุพันธ์บางขั้นตอน

จากการใช้ LOGIST 5 ประมาณค่าความสามารถของผู้สอบในกลุ่มตัวอย่างของแต่ละกรณีพร้อมกัน ๓,๐๐๐ คน และประมาณค่าพารามิเตอร์ของข้อสอบทั้งหมด ๕๐ ข้อ ที่ได้เลือกสรรแล้ว การประมาณค่าทั้งหมดทำพร้อมกันครั้งเดียว ผลของค่าประมาณจึงเป็นมาคร่าเดียวกัน

ข. การหาค่าประมาณคะแนนจริงของผู้สอบ ค่าประมาณที่ใช้ คือ คะแนนจริงตามจำนวนข้อที่ตอบถูกต้องของผู้สอบ ซึ่งมีสูตรดังนี้

$$\xi = \frac{\sum_{i=1}^{n_x} x_i}{n_x} P_i(\theta) \quad \text{----- (๓)}$$

$$\eta = \frac{\sum_{j=1}^{n_y} y_j}{n_y} P_j(\theta) \quad \text{----- (๔)}$$

เมื่อ ξ และ η คือ คะแนนจริงของผู้สอบที่มีระดับความสามารถ θ เกี่ยวกันที่ทำแบบสอบชุด X และ Y ตามลำดับ และ $P_i(\theta)$ หรือ $P_j(\theta)$ เป็นฟังก์ชันของการตอบข้อสอบข้อที่ i ในแบบสอบชุด X และ j ในแบบสอบชุด Y ตามลำดับ

ในการคำนวณค่า $P_i(\theta)$, ξ และ η ได้อาศัยโปรแกรมคอมพิวเตอร์ (ภูภาคผนวก ค.) ให้ ξ เป็นคะแนนจริงของผู้สอบที่ทำแบบสอบ X ส่วน η เป็นคะแนนจริงของผู้สอบที่ทำแบบสอบ Y ในการคำนวณได้แยกคำนวณแต่ละกรณีของแบบสอบแต่ละชุด และใช้วิธีการหาค่า θ ที่ครอบคลุมพิสัยของกลุ่มตัวอย่าง

๒) ขั้นตอนการเทียบมาตรา

ก. จากสมการที่ ๗ และ ๘ ได้กำหนดความสัมพันธ์เชิงฟังก์ชันระหว่าง ξ และ η ดังนั้น ค่าที่สมมูลกันของ ξ และ η มีความหมายเป็นอย่างเดียวกัน จึงเป็นการเทียบมาตราของแบบสอบ ๒ ชุด ณ ที่จุดแต่ละจุดของ θ ที่กำหนดขึ้น

ข. เนื่องจากการเทียบมาตราด้วยการใช้คะแนนจริงมีข้อจำกัด คือ พิสัยของคะแนนจริงไม่คลุมตลอดช่วงความยาวของข้อสอบ ทั้งนี้เพราะ ξ จะมีค่าไม่น้อยกว่า $\sum_i c_i$ ขณะที่ x อาจมีค่าต่ำลงไปถึง ๐ และเป็นห่านองเดียวกับ η ด้วย การแก้ไขจุดนี้ได้ใช้วิธีขยายพิสัยด้วยวิธีแปลงเชิงเส้นตรง (linear interpolation) ซึ่งจะทำให้การเทียบมาตราคะแนนที่อยู่ต่ำกว่าระดับการเดามีความหมายขึ้น (Lord 1982: 211) สูตรในการแปลง คือ

$$\frac{x(y) - \sum_i c_i}{\sum_i c_i (1 - c_i)} = \frac{y - \sum_j c_j}{\sum_j c_j (1 - c_j)}$$

จัดรูปสมการใหม่ จะได้

$$x^* = x(y) = \frac{\sum_i c_i (1 - c_i)}{\sum_j c_j (1 - c_j)} (y - \sum_j c_j) + \sum_j c_j \quad (๙)$$

ค. จากค่าคะแนนจริง ξ และ η ของแต่ละ θ ในแบบสอบแต่ละคู่มือที่ทำการเทียบมาตรา นำมาสร้างเป็นตารางเพื่อเทียบคะแนนจากชุด y ไปสู่ชุด x

การวิเคราะห์ความคลาดเคลื่อนมาตรฐานของการเทียบมาตรา

ความคลาดเคลื่อนมาตรฐานของการเทียบมาตราที่เป็นผลจากการใช้รูปแบบการเทียบมาตราต่างกัน และความยาวของแบบสอบรวมต่างกัน เกิดจากการผันแปรของการสุ่ม (sampling fluctuation) ของค่าประมาณที่ใช้ในรูปแบบของการเทียบมาตรา ในการวิเคราะห์ความ



คลาดเคลื่อนในแต่ละรูปแบบแตกต่างกัน ดังนี้

๑. ความคลาดเคลื่อนมาตรฐานของรูปแบบอควิเปอร์เซนไทล์

(๑) ประมาณความคลาดเคลื่อนมาตรฐานของการเทียบมาตรา จากสูตรความแปรปรวน (Lord 1982) ดังนี้

$$\text{Var } X^* = \frac{PQ}{g_o^2} \left(\frac{1}{n_\alpha} + \frac{1}{n_\beta} \right) \text{----- (๑๐)}$$

เมื่อ $SEE = \sqrt{\text{Var } X^*}$

- Var X^* คือ ความแปรปรวนของคะแนนแปลง
- Q คือ สัดส่วนประชากรที่อยู่ต่ำกว่าคะแนนที่กำหนด สำหรับการประมาณค่า y_q นี้ แทนด้วยค่าของตัวอย่าง
- P คือ $1 - Q$ แทนด้วยค่าของตัวอย่าง
- g_o คือ ความน่าจะเป็นของการแจกแจงของคะแนน X เมื่อกำหนดคะแนน y_o หากได้จากข้อมูลเชิงประจักษ์ด้วยสูตร
 $g_q = \text{Ordinate}/SD_x$
 เมื่อ Ordinate คือ ค่าที่อ่านจากตารางแจกแจงโค้งปกติ ตามค่า p ที่กำหนด และ
 SD_x คือ ส่วนเบี่ยงเบนมาตรฐานของคะแนนใน X ที่ศึกษา

จากสูตรข้างต้นนี้ นำมาใช้กับการเทียบมาตรารูปแบบอควิเปอร์เซนไทล์ที่กระทำกับแบบสอบกรณีคัดเลือก และกรณีแบบสอบวัดผลสัมฤทธิ์ ซึ่งจะได้อาคความคลาดเคลื่อนมาตรฐานกรณีละ ๓ ค่า

การกำหนดค่า Y_0 เพื่อหาค่า SEE ได้ใช้คะแนนมาตรฐานของ Y เป็นหลัก คือ Y_0 ที่มีค่าคะแนนมาตรฐานดังต่อไปนี้ ๐ .๕ ๑.๐ ๑.๕ ๒.๐ ๒.๕ และ ๓.๐

(๒) หาค่าประสิทธิภาพสัมพัทธ์ (Relative efficiency) ของความยาวของแบบสอบถามเพื่อเปรียบเทียบขนาดของความคลาดเคลื่อนมาตรฐานของการเทียบมาตรา เมื่อใช้แบบสอบถามที่มีขนาด $V60$ $V40$ และ $V20$ เทียบกับ $V60$ ซึ่งเป็นวิธีที่ให้ค่า SEE น้อยที่สุดตามสมมุติฐาน

$$RE_{V20.V60} = \frac{SEE X^* (V60)}{SEE X^* (V20)} \times 100\% \quad \text{--- (๑๑)}$$

และ

$$RE_{V40.V60} = \frac{SEE X^* (V60)}{SEE X^* (V40)} \times 100\% \quad \text{--- (๑๒)}$$

ค่าประสิทธิภาพสัมพัทธ์ เป็นค่าที่แสดงอัตราร้อยละของความแม่นยำของการเทียบมาตราที่ใช้แบบสอบถาม $V40$ และ $V20$ เมื่อเทียบกับผลการเทียบของ $V60$

๒. ความคลาดเคลื่อนมาตรฐานของการเทียบมาตราแบบเชิงเส้นตรง

๑. เปรียบเทียบผลการเทียบมาตราด้วยการใช้แบบสอบถามที่มีความยาวต่างกันด้วยค่าความคลาดเคลื่อนมาตรฐานของการเทียบมาตรา โดยใช้สูตรการประมาณความคลาดเคลื่อนของการเทียบมาตราที่เสนอโดยลอร์ด (Lord 1955 cited by Angoff 1984: 104)

ดังนี้

$$SEE^2 = 2 \frac{\sigma_x^2}{x} (1 - \hat{r}^2) \frac{(1 + \hat{r}^2) Z_y^2}{n_t} + 2 \quad \text{--- (๑๓)}$$

เมื่อ SEE คือ ความคลาดเคลื่อนมาตรฐานของการเทียบมาตรา
 $\hat{\sigma}_x^2$ คือ ค่าประมาณความแปรปรวนของคะแนน X
 \hat{r} คือ ค่าประมาณเฉลี่ยของความสัมพันธ์ระหว่างแบบสอบ X กับ V และ Y กับ V โดยมีข้อตกลงว่า

$$\hat{r} = \frac{b_{yV\alpha} \hat{\sigma}_v}{\hat{\sigma}_x} = \frac{b_{yV\beta} \hat{\sigma}_v}{\hat{\sigma}_y}$$

ในการแทนค่าเพื่อคำนวณ ค่า r ซึ่งเฉลี่ยจาก

$$\frac{b_{xV\alpha} \hat{\sigma}_v}{\hat{\sigma}_x} \quad \text{และ} \quad \frac{b_{yV\beta} \hat{\sigma}_v}{\hat{\sigma}_y}$$

เนื่องจากแบบสอบแต่ละคู่ที่ทำการเทียบมาตราจะมีความเฉพาะแปรผันตามขนาดของความยาวแบบสอบรวมที่ใช้ ดังนั้น การคำนวณค่าความคลาดเคลื่อนที่จะให้นำมาเปรียบเทียบกันนั้น จะใช้มาตราคะแนน (scaled score) ที่มีค่าเฉลี่ยเท่ากับ ๕๐ และค่าเบี่ยงเบนมาตรฐานเท่ากับ ๑๐ หมก ทุก ๆ การเทียบมาตรา

การคำนวณค่าความคลาดเคลื่อนของการเทียบมาตรา ได้คำนวณทั้งหมด ๖ ชุด ตามจำนวนการแปลงที่กระทำในขั้นก่อน เนื่องจากค่าความคลาดเคลื่อนจะแตกต่างกันตามระดับของคะแนน และมีลักษณะสมมาตรกันเมื่อคะแนนอยู่เหนือ และต่ำกว่าค่าเฉลี่ย จึงกำหนดการเสนอค่าความคลาดเคลื่อนที่ค่าคะแนนมาตรฐาน Z_y เท่ากับ ๐.๐ ๐.๕ ๑.๐ ๑.๕ ๒.๐ ๒.๕ และ ๓.๐

๒. หากค่าประสิทธิภาพสัมพัทธ์ของความยาวแบบสอบรวมทวิสูตร (๑๑) และ (๑๒)

๓. ความคลาดเคลื่อนมาตรฐานของการเทียบมาตราแบบ IRT

๑) ความคลาดเคลื่อนเชิงสุ่มของการเทียบมาตราแบบ IRT เกิดขึ้นจากการประมาณความสัมพันธ์ของการเทียบ ϵ ไปยัง ϵ ทั้งนี้เนื่องจากการประมาณคะแนนจริงของผู้สอบได้มาจากค่าประมาณประชากรข้อสอบรายข้อ ค่าเหล่านี้ คือ ที่มาของความคลาดเคลื่อนเชิงสุ่มของการเทียบมาตรา (Lord 1982: 2)

สูตรสำหรับคำนวณค่าความคลาดเคลื่อนได้ก็เปลี่ยนแปลงตามความเหมาะสม จากสูตรที่ลอว์ได้สังเคราะห์ขึ้น (๑๔๔๑) โดยที่เปลี่ยนแปลงในส่วนของการจัดเฉพาะกรณีการเทียบมาตราที่ใช้แบบสอบรวมภายใน และการหาค่าประมาณประชากรข้อสอบรายชื่อจากการวิเคราะห์พร้อมกันทุกข้อ ทั้งสองชุดสมการเพื่อคำนวณความแปรปรวนแอสซิมโทติก มีดังนี้

$$\text{Var } \hat{\eta} = \sum_{p=1}^2 \sum_{g=1}^n \sum_{r=1}^3 \sum_{s=1}^3 \eta'_{rgp} \eta'_{sgp} \text{Cov}(t_{rgp}, t_{sgp}) \text{----- (๑๕)}$$

เมื่อ η' หมายถึง $\frac{\partial \eta}{\partial t_{rgp}}$ คืออนุพันธ์ของประชากรข้อสอบ โดยที่ t_{rgp} คือสัญลักษณ์แทนค่าประชากร a_g, b_g และ c_g ของกลุ่มตัวอย่าง p เมื่อ $r = 1, 2$ และ 3 แทน a, b และ c ตามลำดับ สูตรการคำนวณค่าอนุพันธ์มีดังนี้ (Lord 1980: 180)

เมื่อ $P = 1$

$$\frac{\partial \eta}{\partial t_{rg1}} = -p_{g1}^{(r)}(\theta) \frac{\sum p_{g2}'(\theta)}{\sum p_{g1}'(\theta)}$$

เมื่อ $P = 2$

$$\frac{\partial \eta}{\partial t_{rg}} = p_{g2}^{(r)}(\theta)$$

โดยที่ $p^{(r)}$ เป็นสัญลักษณ์แทนอนุพันธ์ของ p_{gp} ที่ $r = 1$ หรือ 2 หรือ 3 $\text{Cov}(t_{rgp}, t_{sgp})$ คือ ความแปรปรวนรวมรวมของตัวประมาณที่เป็นไปได้สูงสุดของค่าประชากรข้อสอบ (Maximum likelihood estimators of item parameters) เมื่อค่าความสามารถคงที่ (Lord 1980: 191)

ความคลาดเคลื่อนมาตรฐานของการเทียบมาตรฐานรูปแบบ IRT นี้ เท่ากับกรณีที่สองของความแปรปรวนเฉลี่ยกำลังสุทธ

$$SEE = \sqrt{\text{Var} \hat{\eta} / (n_g) (n_\alpha)} \quad \text{-----} \quad (๑๕)$$

เมื่อ n_g และ n_α คือ จำนวนข้อสอบในแต่ละชุดของแบบสอบ และจำนวนคนในกลุ่มตัวอย่าง การวิเคราะห์ในส่วนนี้ได้ดำเนินการ ดังนี้

ก. การกำหนดจุดประเมินความคลาดเคลื่อนเพื่อเสนอในรายงาน ใช้การระบุค่า ตามคะแนนมาตรฐานของคะแนนจริงตามการแจกแจงในแบบสอบแต่ละชุดในช่วง ± 3.0 ซึ่งได้วิเคราะห์จากการแจกแจงความถี่ของคะแนนจริงของกลุ่มตัวอย่าง β ที่ทำแบบสอบ γ โดยแต่ละชุดของแบบสอบจะเริ่มรายงานที่คะแนนจริงต่ำสุดที่เป็นไปได้ (Minimum possible scores) สำหรับค่าที่ต่ำกว่า เป็นค่าที่ไม่มีคามหมายเพราะอยู่ในระดับของการเดา

ข. ประเมินค่า SEE ตามค่าคะแนนมาตรฐานที่สนใจตามสูตรที่ ๑๓ และ ๑๔ โดยอาศัยโปรแกรมคอมพิวเตอร์ที่ผู้วิจัยเขียนเอง (ภาคผนวก ค.) แล้วนำไปเปรียบเทียบระหว่างวิธีที่ใช้แบบสอบรวมต่างกันในแต่ละกรณีของแบบสอบ

๒) คำนวณค่าดัชนีประสิทธิภาพสัมพัทธ์ของแบบสอบรวมในทำนองเดียวกับที่วิเคราะห์ในรูปแบบอิกวิเปอร์เซนไทล์ และเชิงเส้นตรง (สูตร ๑๑ และ ๑๒) แล้วนำไปเปรียบเทียบประสิทธิภาพของแบบสอบรวมร้อยละ ๘๐ และ ๒๐ เมื่อเทียบกับแบบสอบรวมร้อยละ ๒๐ ท่อไป

การวิเคราะห์กลุ่มสอบทานผล

วิธีวิเคราะห์กลุ่มสอบทานผล ได้นำมาใช้ประเมินความเพียงพอของการเทียบมาตรฐานเชิงประจักษ์ กลุ่มตัวอย่างสอบทานผลเป็นกลุ่มตัวอย่างที่ได้จากการสุ่มตัวอย่างง่ายจากประชากร เกี่ยวพันกับกลุ่มตัวอย่างเทียบมาตรฐาน ซึ่งเป็นกลุ่มตัวอย่างที่พัฒนาคะแนนสมมูลในแต่ละวิธีของการเทียบมาตรฐาน การวิเคราะห์กลุ่มสอบทานผลได้เริ่มด้วยการแปลงคะแนน γ ในหน่วยตัวอย่างให้เป็น x'

ตามตารางคะแนนสมมูลที่ได้พัฒนาขึ้นจากรูปแบบที่ทองการ แล้วตรวจสอบความถูกต้องโดยวัดความแตกต่างของคะแนน x^* กับคะแนน x ที่มีอยู่ในหน่วยตัวอย่าง ถ้าความแตกต่างเกิดขึ้นโดยเฉลี่ยทั้งกลุ่มไม่มาก ย่อมหมายถึงวิธีการเทียบมาตราที่เป็นตัวกำหนดตารางคะแนนเทียบนั้นมีความเพียงพอที่ทำให้เกิดการแปลงคะแนนที่สมมูลได้ การดำเนินการวิเคราะห์กลุ่มสอบทานผลได้ดำเนิน ดังนี้

๑. กลุ่มตัวอย่างสอบทานผล

กลุ่มตัวอย่างสอบทานผล หรือกลุ่ม r ได้จากการสุ่มตัวอย่างอย่างง่ายพร้อม ๆ กับการสุ่มตัวอย่างกลุ่ม α และ β ดังที่กล่าวมาแล้วในหัวข้อการออกแบบเทียบมาตรา กลุ่ม r มีขนาด ๑๕๐๐ คน แต่ละหน่วยมีข้อมูล ที่เป็นผลการสอบรายข้อ ๑๐๐ ข้อ เมื่อนำมาประมวลผลคะแนนตามที่วางแผนในการเตรียมชุดแบบสอบเพื่อเทียบมาตราแล้ว ทุกหน่วยข้อมูลจึงมีข้อมูลในลักษณะเป็นชุดของแบบสอบตามที่ทองการ คือ XV20 YV20 XV40 YV40 XV60 YV60

๒. การหาดัชนีเปรียบเทียบความแตกต่าง (Index C)

ดัชนีเปรียบเทียบความแตกต่างที่ใช้กับกลุ่มสอบทานผลที่นำมาเป็นมาตรการประเมินความเพียงพอของวิธีการเทียบมาตรานี้ ได้ประยุกต์จากดัชนีเปรียบเทียบเปอร์เซนไทล์ (The percentile comparison index) ของโคเลน (Kolen 1982) และดัชนีเบี่ยงเบน (discrepancy index) ของปีเตอร์สัน และคณะ (Petersen etc 1982: 91) และคำแนะนำของสตรูด (Stroud 1982: 138)

ดัชนี C เป็นตัววัดความแตกต่างระหว่างการแจกแจงของคะแนนชุด x ซึ่งใช้เป็นคะแนนเกณฑ์ กับคะแนนแปลงที่อาศัยตารางการเทียบคะแนนสมมูลที่ได้จากกระบวนการเทียบมาตราแต่ละวิธี โดยได้ทำการแปลงคะแนน y ของผู้สอบแต่ละคน ให้เป็นคะแนน x^* ที่สมมูลกัน ความแตกต่างของ x^* กับ x ของผู้สอบคน ๆ หนึ่ง เกี่ยวกัน ถือว่าเป็นความคลาดเคลื่อนที่เกิดขึ้นกับวิธีการเทียบมาตราที่เป็นตัวกำหนดคะแนนสมมูลนั้น ๆ

การคำนวณค่าดัชนี c ใช้สูตรดังต่อไปนี้

$$c = \frac{\sum (x - x_i^*)^2 / n s_x^2}{\dots} \quad (๑๖)$$

เมื่อ x_i	คือ	คะแนน x ของผู้สอบคนที่ i จากแบบสอบชุด x
x_i^*	คือ	คะแนน x ของผู้สอบคนที่ i จากการแปลงคะแนน y ที่สมมูลกัน
n	คือ	จำนวนคนในกลุ่มตัวอย่างสอบทานผล
s_x^2	คือ	ค่าความแปรปรวนของการแจกแจงคะแนนจากแบบสอบชุด x

ค่า c เป็นค่าที่ได้ดวงน้ำหนักให้เป็นมาตรฐานของค่าเฉลี่ยของความแตกต่างกำลังสอง ซึ่งเป็นค่าที่ให้น้ำหนักมากที่สุดกับคะแนน x ที่ปรากฏอยู่มากที่สุด และมีความแน่นอนในความหมายของความคลาดเคลื่อนรวมที่ใช้ในวรรณคดีทางสถิติ (Petersen etc. 1982: 91) ค่าดัชนี c ซึ่งได้ทำการดวงน้ำหนักแล้วนี้ สามารถนำไปเปรียบเทียบกับค่า c ที่ได้มาจากสถานการณ์การเทียบมาตราที่ต่าง ๆ กันได้โดยตรง

๓. ประเมินค่าดัชนี c

ค่า c ที่คำนวณได้จากกลุ่มสอบทานผลของแต่ละกรณีของแบบสอบ ซึ่งเป็นค่าเชิงปริมาณที่ยังไม่สามารถระบุคุณภาพได้ ทั้งนี้ จึงได้อาศัยหลักการสร้างเกณฑ์การประเมินของปีเตอร์สัน และคณะ ได้ใช้ในการวิจัยเพื่อศึกษาเปรียบเทียบผลการเทียบมาตราทางรูปแบบ (Petersen and others 1982: 93-94) หลักการดังกล่าวมีดังนี้ ค่า c ที่คำนวณได้เป็นความคลาดเคลื่อนรวม ซึ่งเป็นผลบวกของกำลังสองของความล่าเอียงกับความแปรปรวนของความแตกต่าง ถ้าความคลาดเคลื่อนรวมเกิดขึ้นเนื่องจากความล่าเอียงเพียงอย่างเดียวตามความเห็นของนักจิตมิติ วิธีการเทียบมาตราดังกล่าวจึงกล่าวได้ว่า เป็นวิธีที่ให้ผลที่น่าพอใจอย่างมาก เมื่อค่าความคลาดเคลื่อนรวมนั้นมีค่าน้อยกว่าหรือเท่ากับกำลังสองของร้อยละ ๕ ของส่วนเบี่ยงเบนมาตรฐานของการแจกแจงของคะแนนเกณฑ์ แต่ถ้ามากกว่าร้อยละ ๒๐ ผลการเทียบมาตรานั้นถือว่า ไม่น่าพอใจอย่างยิ่ง จากหลักการดังกล่าวได้กำหนดระดับการยอมรับ ดังนี้

ระดับน่าพอใจอย่างมาก	หมายถึง	$c \leq (.05 SD_X)^2$
ระดับน่าพอใจ	หมายถึง	$(.05 SD_X)^2 < c \leq (.10 SD_X)^2$
ระดับปานกลาง	หมายถึง	$(.10 SD_X)^2 < c \leq (.15 SD_X)^2$
ระดับไม่น่าพอใจ	หมายถึง	$(.15 SD_X)^2 < c \leq (.20 SD_X)^2$
ระดับไม่น่าพอใจอย่างยิ่ง	หมายถึง	$(.20 SD_X)^2 < c$

๔. การทดสอบความมีนัยสำคัญของความแตกต่างของค่า c

สถิติการทดสอบที่เลือกมา คือ Friedman test ซึ่งเป็นการทดสอบนัยพารา-เมตริก โดยการวิเคราะห์ความแปรปรวนสองทางที่ใช้ข้อมูลจำลาคับที่ทำการวัดซ้ำ และไม่ต้องมีข้อตกลงเบื้องต้นของการแจกแจงของค่า c การทดสอบได้กำหนดให้ทำการทดสอบความแตกต่างรายคู่ที่สนใจ (Marascuilo and McSweeney 1977: 376-379, 384-385) มีดังนี้

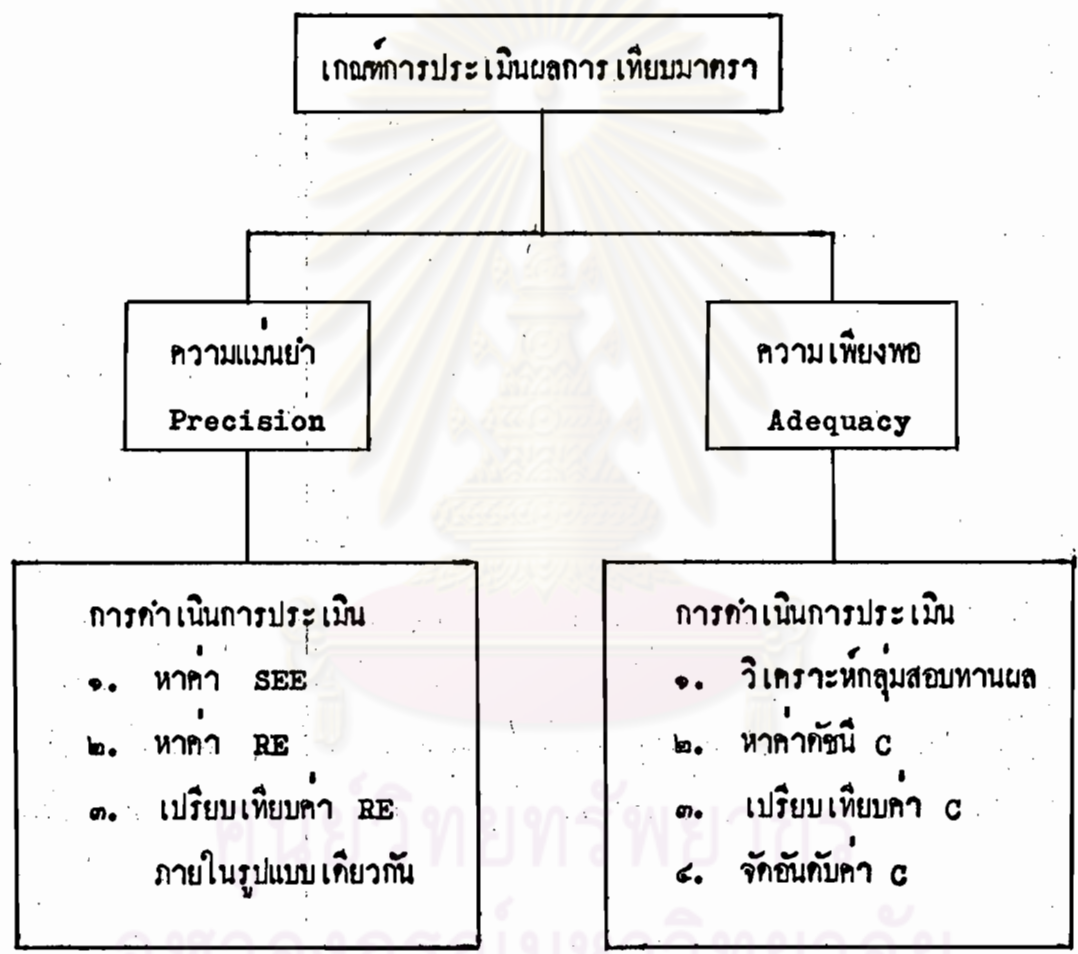
ก. ในประเด็นของแบบสอบถาม เพื่อทดสอบว่าแบบสอบถาม V60 V40 และ V20 มีความแตกต่างกันเป็นคู่ ๆ โดยกำหนดให้ตัวแปรทางรูปแบบของการเทียบมาตราเป็นตัวแปรตามบล็อก

ข. ประเด็นของรูปแบบการเทียบมาตรา เพื่อทดสอบว่ารูปแบบอิคิวเปอร์เซนไทล์เชิงเส้นตรง และ IRT มีความแตกต่างกันเป็นคู่ ๆ โดยกำหนดให้ตัวแปรตามความยาวแบบสอบถามเป็นตัวแปรบล็อก

สรุปการดำเนินการวิจัย

จากวัตถุประสงค์ของการวิจัยที่ว่าเพื่อศึกษาผลของการเทียบมาตราสามรูปแบบ ที่ใช้แบบสอบถามในการเทียบมาตราต่าง ๆ กันสามขนาดนั้น ได้กำหนดวัตถุประสงค์ตามนิยามของคำว่า ผลของการเทียบมาตราว่า เป็นการศึกษาถึงความคลาดเคลื่อนมาตรฐานของการเทียบมาตรา และความเพียงพอของการเทียบมาตราต่าง ๆ เหล่านั้น การดำเนินการวิจัยได้เริ่มด้วยการออกแบบการเทียบมาตราด้วยการกำหนดแบบสอบถาม และประชากรผู้สอบ ตลอดจนการจัดกระทำข้อมูลคะแนนให้พร้อมที่จะใช้เพื่อการดำเนินการเทียบมาตรา ซึ่งได้ดำเนินการตามระเบียบวิธีการเทียบมาตราทั้งสาม

รูปแบบ จนได้ตารางเทียบคะแนนสมมูลของแบบสอบแต่ละคู่ จากนั้นจึงเป็นขั้นตอนของการวิเคราะห์ ผลการเทียบมาตรา ซึ่งได้แบ่งเป็นสองส่วนตามที่กำหนดในวัตถุประสงค์ คือ ส่วนของการวิเคราะห์ ความคลาดเคลื่อนมาตรฐาน และส่วนของการวิเคราะห์กลุ่มสอบทานผล ประเมินความแม่นยำ และ ความเพียงพอของวิธีการเทียบมาตรา ดังสรุปในแผนภาพที่ ๓ ผลการวิเคราะห์ได้เสนอในบทต่อไป



แผนภาพที่ ๓ สรุปเกณฑ์การประเมินผลการเทียบมาตรา