

บทที่ ๑



บทนำ

## ความเป็นมาและความสำคัญของปัญหา

ความนิยมและความต้องการใช้แบบสอบ เน้นวันก็มีแต่จะมากขึ้นตามการเพิ่มของประชากร และความซับซ้อนของสังคม ทั้งนี้เพื่อให้กระบวนการทดสอบมีบทบาทในการสร้างความยุติธรรม และความเสมอภาคในการแข่งขัน อาทิเช่น การสอบคัดเลือกเข้าศึกษาต่อในมหาวิทยาลัย การสอบเข้ารับราชการในตำแหน่งต่าง ๆ เมื่อพิจารณาอัตราส่วนของผู้ผ่านเข้ารับการคัดเลือกแต่ละครั้ง ย่อมเป็นที่ยอมรับให้เห็นถึงความสำคัญของการสอบรายบุคคลอย่างยิ่ง จนเป็นเหตุกระตุ้นให้ผู้สอบมีความกระตือรือร้นแสวงหาทางเพิ่มทุนคะแนนด้วยวิธีต่าง ๆ รวมถึงวิธีการปิดช่องความชำนาญในการทำข้อสอบจากข้อสอบเก่าที่โซมาแล้ว ทั้งนี้ จึงเป็นเหตุผลให้ผู้รับนิชอบเกี่ยวกับโปรแกรมการสอบ ท้องหาแบบสอบชุดใหม่ ๆ ที่ไม่เคยใช้สอบมาก่อนมาใช้ในการสอบครั้งต่อไป เพื่อไม่ให้เกิดความได้เปรียบเสียเปรียบในระหว่างผู้สอบด้วยกัน

การสร้างแบบสอบชุดใหม่ โดยหลักการแล้วเป็นเพียงการเปลี่ยนแปลงคำถามเท่านั้น แต่ยังคงทำหน้าที่วัดความรู้ ความสามารถ หรือทักษะในขอบเขตเดียวกับแบบสอบชุดก่อน หรือกล่าวอีกนัยหนึ่ง คือ พยายามสร้างแบบสอบชุดนั้นเอง เพื่อให้ผลการสอบเป็นคะแนนที่ให้ความหมายเปรียบเทียบโดยตรงจากการสอบแต่ละครั้งด้วยแบบสอบต่างชุดกัน แต่โดยความเป็นจริง การสร้างแบบสอบชุดขึ้นที่มีความหมายสมบูรณ์ จนก่อให้เกิดคะแนนสมมูล (Equivalent scores) มีโอกาสเป็นไปได้น้อยมาก แนวทางปฏิบัติต่าง ๆ ที่นักวิจัยได้แนะนำก็เพียงให้เกิดความใกล้เคียงกันเนื้อหา ค่าสถิติของข้อสอบ และแบบสอบทั้งฉบับเท่านั้น (Gulliksen 1958: 174) ทั้งนี้ การลงทุนเพื่อผลิตแบบสอบชุดใหม่ให้มีคุณสมบัติของความเป็นชุดขึ้นจนสามารถให้คะแนนสมมูลอย่างแท้จริง เป็นเรื่องไม่คุ้มค่าการลงทุน และไม่เหมาะสมอย่างยิ่ง เพราะการสอบครั้งต่อไปอาจมีการเปลี่ยนแปลงจุดเน้นที่ต้องการวัด หรือเพิ่มเพิ่มวัดคุณประสงค่างประการ ซึ่งทำให้แบบสอบมีการเปลี่ยนแปลง

แปลงเนื้อหา หรือระดับความยากไปบ้าง ทางออกที่ดีที่สุด คือ การใช้เทคนิคทางสถิติมาปรับคะแนน ให้เกิดความเท่าเทียมกันได้ กระบวนการเช่นนี้ เรียกว่า การเทียบมาตราคะแนนของแบบสอบ (Test equating)

เรื่องราวเกี่ยวกับการเทียบมาตราในประเทศไทยมีน้อยมาก ไม่มีตำราวิจัยผลทางการศึกษา เล่มใดใดกล่าวถึงเรื่องนี้เลย จึงเป็นเรื่องที่ควรแก่การศึกษา ความจริง เยาวดี ราชชัยกุล (Rangchaikul 1975) เคยศึกษาและชี้ชวนให้เห็นความจำเป็นที่คงต้องใช้กระบวนการเทียบมาตรา กับการสอบไล่ชั้นประโยคมัธยมศึกษาตอนปลาย (มศ. ๕) เพราะปรากฏหลักฐานว่า แบบสอบที่ใช้แต่ละปีมีการเปลี่ยนแปลงต่าง ๆ เช่น กำหนดเวลาในการสอบ ความยาวของแบบสอบ และอื่น ๆ นอกจากนี้ยังพบว่า จำนวนนักเรียนสอบได้และสอบตกมีสภาพผันแปรมากเกินกว่าจะยอมรับว่าเป็นสภาพ ความสามารถที่ต่างกันอย่างแท้จริงของนักเรียน ดังนั้น คะแนนดิบที่ซาคคุณสมบัติของความสมมูล แต่ยังคงยึดการตัดสินด้วยเกณฑ์ผ่านขั้นต่ำ โดยคิดเป็นร้อยละ ๕๐ ของคะแนนดิบตลอดไป เป็นเรื่อง ที่นักวิชาการควรตระหนักและป้องกันความไม่ยุติธรรมที่อาจเกิดขึ้นได้ เยาวดี ราชชัยกุล (๑๙๗๕) ได้ศึกษาถึงการสอบที่กระทรวงศึกษาธิการจัดทำเนมีการให้แก่นักเรียนในระบบโรงเรียน ซึ่งสอบเดือน มีนาคม ๑๙๗๓ และ ๑๙๗๔ และการสอบที่จัดให้บุคคลทั่ว ๆ ไป ซึ่งอยู่นอกโรงเรียน ซึ่งสอบ เดือนสิงหาคม ๑๙๗๓ และ ๑๙๗๔ จากการเทียบมาตราคะแนนของแบบสอบทั้ง ๔ ชุด ก็มาตราคะแนน รวม (Common scale score) เทียบที่ละวิชาด้วยรูปแบบเชิงเส้นตรง ผลการเทียบพบว่า คะแนนดิบจากแบบสอบทั้ง ๔ ชุด มีความแตกต่างกันในความหมายของตัวเลข ภายหลังจากเทียบมาตรา และตรวจสอบสัดส่วนของผู้สอบที่ได้รับการตัดสินได้และตกด้วยเกณฑ์ร้อยละ ๕๐ แล้วนั้น พบว่า สัดส่วน นักเรียนที่น่าจะสอบได้แต่ได้รับการตัดสินให้ตก มีจำนวนน่าสนใจ เช่น ถ้ายึดการสอบเมื่อเดือนมีนาคม ๑๙๗๓ เป็นหลัก เทียบกับการสอบเมื่อ เดือนสิงหาคม ๑๙๗๓ และมีนาคม ๑๙๗๔ แล้ว ร้อยละ ๓๗.๖ ที่ตกเมื่อ เดือนสิงหาคม ๑๙๗๓ และร้อยละ ๕ ที่ตกเมื่อเดือนมีนาคม ๑๙๗๔ ควรเป็นผู้ที่สอบได้เมื่อใช้ คะแนนตัดสินเทียบเท่าการสอบเมื่อเดือนมีนาคม ๑๙๗๓ ที่ร้อยละ ๕๐ ของแบบสอบเช่นเดียวกัน

ความยุติธรรมจากการสอบในทำนองนี้ เมื่омองให้ลึกซึ้งแล้ว น่าจะเป็นสิทธิมนุษยชนที่ควร จะได้รับความคุ้มครอง เพราะผลการสอบมีความสำคัญต่อโอกาสการสอบเข้ามหาวิทยาลัย ท่อการ ทำงานหา และต่อความเชื่อมั่นและเกียรติประวัติของบุคคลนั้น ๆ ในสหรัฐอเมริกาได้เกิดเสียงวิพาก

วิจารณ์ในทำนองนี้มานาน และรุนแรงขึ้นเป็นลำดับ จนได้มีกฎหมายเกิดขึ้นเพื่อคุ้มครองผู้รับการทดสอบที่รัฐนิวยอร์ก ซึ่งมีผลบังคับใช้เมื่อวันที่ ๑ มกราคม ๑๙๘๐ เรียกกฎหมายนี้ว่า Truth-in Testing law (LaValle 1980) หรือ The LaValle Act (Marco 1981) ตามที่ผู้เสนอร่างกฎหมายนี้ และมีแนวโน้มที่จะเป็นกฎหมายใช้ในประเทศอื่น ๆ อีกหลายรัฐ กฎหมายฉบับนี้บังคับใช้กับการทดสอบด้วยแบบสอบมาตรฐานที่ใช้กับระดับสูงกว่ามัธยมศึกษา ให้หน่วยงานผู้ผลิตแบบสอบปฏิบัติ ดังนี้

- (๑) ส่งเอกสารที่ทำการศึกษารายละเอียด ตลอดจนรายงานทางสถิติที่เกี่ยวข้องกับแบบสอบไปยังกรมการการศึกษาของรัฐ
- (๒) ส่งเอกสารต่อไปถึงคณะกรรมการการศึกษาภายใน ๓๐ วัน หลังจากได้นำแบบสอบมาตรฐานออกใช้แล้ว: คำถามข้อสอบ ๑ ชุด ที่เป็นฉบับให้คะแนนดิบ พร้อมทั้งค่าเฉลี่ยกฎเกณฑ์ที่ใช้ในการแปลงคะแนนดิบไปสู่คะแนนรายงานผลการสอบ พร้อมคำอธิบาย สำหรับผู้รับการทดสอบมีสิทธิขอให้หน่วยงานทดสอบส่งคำถามข้อสอบให้หนึ่งชุด กระจายคำตอบของผู้สอบเอง พร้อมทั้งการตรวจให้คะแนน และผลสอบที่เป็นคะแนนดิบที่นำไปแปลงเป็นคะแนนรายงาน ผู้รับการทดสอบขอได้ภายใน ๔๐ วัน นับตั้งแต่วันประกาศผลสอบ และผู้ขอต้องชำระเงินเพื่อขอรับบริการนี้
- (๓) สำนักงานทดสอบต้องจัดส่งเอกสารขอสมัครสอบให้แก่ผู้ขอ พร้อมคำชี้แจงในการสอบ สิ่งที่มีแบบสอบมุ่งวัดการให้คะแนน ผลการสอบที่ส่งถึงผู้ใช้ผลสอบต้องส่งคะแนน พร้อมทั้งสารสนเทศต่าง ๆ ที่เสนอในรายงานผลการสอบ อันได้แก่ ภูมิหลังของผู้สอบ
- (๔) การเปิดเผยผลการสอบเป็นสิทธิของผู้สอบแต่ผู้เดียว

จากที่กล่าวมานี้จะเห็นได้ว่า เรื่องของการเทียบมาตรฐานเป็นกระบวนการที่จำเป็นในกฎหมายฉบับนี้ เคยมีผู้คัดค้านร่างกฎหมายนี้ขณะที่อยู่ในระหว่างการพิจารณาว่า หากร่างกฎหมายฉบับนี้ผ่านรัฐสภาแล้ว จะทำให้ไม่สามารถหาเทคนิคการเทียบมาตรฐานที่เหมาะสมได้ และจะมีผลทำให้ค่าใช้จ่ายในการสอบต้องเพิ่มสูงขึ้น เนื่องจากต้องสร้างข้อสอบใหม่หมดทุกข้อ แต่เมื่อกาลเวลาผ่านไปประเด็นโต้แย้งดังกล่าวก็ไต่หมอกไปด้วย โดยฝ่ายสนับสนุนร่างกฎหมายได้อธิบายว่า กฎหมายยอมให้ใช้วิธีการเทียบมาตรฐานที่เหมาะสม เช่น วิธีการที่ใช้อยู่ตามปกติที่แทรกข้อสอบจำนวนหนึ่งในแบบสอบชุดต่าง ๆ ซึ่งเป็นวิธีที่นักวิชาการสร้างแบบสอบมาตรฐานวัดความถนัดเชิงวิชาการของสหรัฐอเมริกา (SAT) คัดชั้นข้อสอบจำนวนหนึ่งที่แทรกเข้าไปไม่มีผลต่อผลการสอบในส่วนที่เป็นคะแนนดิบ จึงไม่จำเป็นต้องเปิดเผยคะแนนสำนักทดสอบที่ขายแบบสอบก็เพียงแต่เพิ่มส่วนที่เป็นการเทียบมาตรฐานลักษณะนี้ได้ นอกจากนี้กฎหมายก็ไม่ได้เป็นอุปสรรคต่อการพัฒนาเทคนิคทางสถิติอื่น ๆ เลย (LaValle 1980)

ภัยแรงผลักดันจากความต้องการของสังคมที่ปรากฏในช่วง ๑๐ ปีนี้ มีนักวิชาการให้ความสนใจเรื่องราวการเทียบมาตรฐานของแบบสอบ และพัฒนาทฤษฎี และเทคนิคต่าง ๆ อย่างจริงจัง ซึ่งมาร์โก (Marco 1981) เรียกบรรยากาศนี้ว่า ยุคแห่งการเปิดเผยแบบสอบ (An era of test disclosure) ในสหรัฐอเมริกา ที่ ETS (Educational Testing Service) ซึ่งเป็นแหล่งที่ผลิตแบบสอบมาตรฐานชนิดต่าง ๆ ให้บริการทั้งในและนอกประเทศอย่างกว้างขวาง ได้เริ่มพัฒนาวิธีการเทียบมาตรฐานกับแบบสอบ SAT ตั้งแต่ปี ๑๙๖๒ โดยใช้มาตรฐานคะแนนของแบบสอบ SAT ฉบับที่ใช้ทดสอบเมื่อเดือนเมษายน ๑๙๖๐ เป็นหลัก ตัวอย่างแบบการเทียบมาตรฐานเชิงเส้นตรง (Donlon and Angoff 1971: 32) ในปี ๑๙๗๒ ได้เริ่มโครงการวิจัยเพื่อพัฒนาตรวจสอบวิธีการที่ใช้อยู่ และศึกษาหาช่องทางที่เหมาะสมยิ่งขึ้น ในช่วงปี ๑๙๗๕ ถึง ๑๙๘๐ ETS ได้จัดสัมมนาขึ้น ๓ ครั้ง โดยเน้นการพัฒนาเทคนิคการเทียบมาตรฐาน และผลิตเอกสาร บทความ ตลอดจนถึงตำรา (Holland and Rubin 1982: xiii) ในเวลาเดียวกัน นักวิชาการที่ไม่ได้สังกัด ETS ก็ได้ให้ความสนใจมากเช่นกัน ทั้งปรากฏบทความในวารสารที่เกี่ยวกับการศึกษาหลายฉบับอย่างต่อเนื่องสม่ำเสมอ

พัฒนาการเกี่ยวกับการเทียบมาตรฐานนับตั้งแต่อดีตจนถึงปัจจุบัน มีทั้งทางด้านกาหนดค่านิยาม ทฤษฎีว่าด้วยการทดสอบที่เข้ามามีส่วนสนับสนุน เทคนิคเชิงสถิติ และการออกแบบรวบรวมข้อมูลคะแนน ทั้งนี้เป็นพัฒนาการที่สอดคล้องกับความก้าวหน้าของทฤษฎีการทดสอบ และสถานการณ์เพื่อการประยุกต์ใช้ ในค่านิยามของการเทียบมาตรฐาน แรกเริ่มเน้นความจำเป็นของการจับคู่เหมือนของการแจกแจงคะแนนของผู้สอบที่ทำแบบสอบสองชุด (Gulliksen 1950; Lord 1950) ถึงแม้จะไม่ใช่อธิบายค่านิยาม แต่เป็นที่เข้าใจกันว่าแบบสอบที่จะนำมาเทียบกันนั้นต้องตั้งอยู่บนรากฐานของสมมุติฐานของความเป็นคู่ขนานของแบบสอบ (Marco 1980) นิยามในระยะต่อมาได้ให้ความสนใจกระจำจเพิ่มขึ้นว่า คุณลักษณะที่ทำกาการวัดด้วยแบบสอบที่ต้องการเทียบกัน ต้องเป็นคุณลักษณะเดียวกัน คะแนนจะสมมูลกันเมื่อสมมุติที่ระดับความสามารถเดียวกัน (Angoff 1984: 92) และเมื่อเร็ว ๆ นี้ ลอร์ด (Lord 1977, 1980: 197) ได้เน้นความหมายของการเทียบมาตรฐาน โดยการเชื่อมโยงความคิดเข้ากับความเสมอภาค (Equity) ส่วนมาร์โก (Marco 1981) ได้วิเคราะห์ค่านิยามต่าง ๆ และสรุปว่า แต่ละนิยามต่างสัมพันธ์กับวิธีการทางสถิติชุดหนึ่ง ๆ ที่จะทำให้อ

เกิดผลคงที่ตามนิยามนั้น นิยามต่าง ๆ แสดงออกในรูปของการเปรียบเทียบในระหว่างคะแนนด้วยกัน ไม่ใช่ตัวแบบสอบ ทุกนิยามใช้กับประชากรที่ระบุหรือนิยามไว้แล้ว โดยทฤษฎีของการเทียบมาตรา ไม่มีความแปรเปลี่ยนใด ๆ เมื่อข้ามกลุ่มประชากร แต่โดยทางปฏิบัติอย่างน้อยผลลัพธ์จะคงขึ้นอยู่กับลักษณะเฉพาะของรายบุคคลที่ประกอบขึ้นเป็นประชากร หากพิจารณาคำนิยาม ดูเหมือนว่าทุกคนในประชากรได้ทำแบบสอบทั้งสองฉบับ แต่โดยข้อเท็จจริงข้อมูลที่ได้อาจไม่สมบูรณ์เช่นนั้น กุญแจในการเทียบมาตราจึง เป็นวิธีการเชิงประจักษ์ที่ของการการออกแบบเพื่อรวบรวมข้อมูล และจัดกระทำข้อมูลภายใต้สภาพของความจำกัดที่ไม่สามารถวางเงื่อนไขให้ความเพียงพอกตามนิยามได้ทุกประการ แองกอฟ (Angoff 1984: 94-121) ได้บรรยายการออกแบบที่ครอบคลุมทั้งสถานการณ์การทดสอบ และรูปแบบการเทียบมาตราไว้ ๖ แบบ จำแนกได้เป็น ๒ กลุ่มใหญ่ คือ กลุ่มวิธีการเทียบมาตราโดยใช้แบบสอบร่วม (Anchor test หรือ Common test) และกลุ่มวิธีการเทียบโดยไม่ใช้แบบสอบร่วม ซึ่งในกรณีที่ไม่ใช้แบบสอบร่วมนี้เป็นรูปแบบที่ของการเงื่อนไขที่เข้มงวดหลายประการ คือ ความเป็นตัวอย่างสมบูรณ์ของผู้รับการทดสอบ และความเป็นคู่ขนานของแบบสอบ โดยเน้นความสมบูรณ์ของค่าความเที่ยงของแบบสอบทั้งสอง เมื่อเงื่อนไขเข้มงวดมาก การปฏิบัติก็มีข้อจำกัดอย่างมาก เพราะในสภาพการทดสอบจริงโอกาสการสุ่ม และการตรวจความสมบูรณ์ของความเที่ยงย่อมเป็นไปได้ยากมาก โดยเฉพาะการทดสอบที่อยู่ภายใต้การบังคับของกฎหมาย เช่น Truth-in-Testing law ในสหรัฐอเมริกาแบบสอบจะไม่สามารถเก็บเป็นความลับอีกต่อไป ผู้สอบจะเกิดการเรียนรู้ได้จากการทบทวนแบบสอบได้ ซึ่งมีผลต่อคะแนนที่จะได้จากแบบสอบใหม่ การเทียบมาตราด้วยการใช้ข้อมูลเก่าจะทำให้ได้คะแนนแปลงในชุดใหม่ที่สูงเกินไป (Marco 1981) ดังนั้น การออกแบบที่ใช้แบบสอบร่วมจึง เป็นทางออกที่เหมาะสมกว่า

แบบสอบร่วม หมายถึง กลุ่มของข้อสอบที่กำหนดให้ผู้สอบสองกลุ่มที่ทำแบบสอบคนละชุด (X หรือ Y) ได้ทำเหมือนกัน แบบสอบร่วมจะทำหน้าที่ลดความแตกต่างของกลุ่มผู้สอบทั้งสองกลุ่ม อันเนื่องมาจากความแปรปรวนของการสุ่มส่วนหนึ่ง (Angoff 1984: 106) คะแนนจากแบบสอบร่วมนำมาใช้เป็นฐานเพื่อประมาณความแตกต่างของกลุ่มคนสองกลุ่ม และสารสนเทศนี้นำไปใช้ในการประมาณความแตกต่างระหว่างแบบสอบสองชุดที่ผู้สอบ เพื่อหาคะแนนแปลงที่เทียบเท่ากันต่อไป (Klein, Kolen 1985) ปัจจุบันการศึกษาลักษณะต่าง ๆ ของแบบสอบร่วม และผลกระทบที่ส่งถึง

ผลการเทียบมาตราค่าง ๆ ได้ให้ความรู้ที่ชัดเจนบางประการ คือ ข้อสอบในชุดของแบบสอบรวม ควรทำหน้าที่วัดความรู้ ความสามารถเกี่ยวกับแบบสอบที่กองการเทียบมาตรา ค่าสัมประสิทธิ์สหสัมพันธ์ ระหว่างแบบสอบรวมกับแบบสอบที่กองการเทียบมาตราควรมีค่าสูงควย (Angoff 1984: 107) หรือกล่าวอีกนัยหนึ่ง ถ้าสามารถจัดหาแบบสอบรวมที่มีลักษณะเป็นฉบับย่อยส่วนใด จะช่วยให้ผลการเทียบมาตรามีความถูกต้องแม่นยำสูง ความยาวของแบบสอบรวม หรือจำนวนข้อในแบบสอบรวมเป็นอีกองค์ประกอบหนึ่งที่มีความสำคัญยิ่ง เพราะมีประเด็นที่ควรพิจารณาดังนี้ ถ้าใช้แบบสอบรวมเป็นส่วนหนึ่งเพื่อการเทียบมาตราโดยไม่รวมคะแนนส่วนนี้เข้าในผลการสอบรายบุคคล ซึ่งเรียกว่า เป็นแบบสอบรวมประเภทภายนอก (External anchor test) แบบสอบรวมที่ยาวจะทวงใช้เวลามากขึ้น ส่วนอีกประเภทหนึ่ง ใช้ผนวกเข้าไปในแบบสอบที่กองการเทียบมาตราทั้งสองชุด ซึ่งเรียกว่า แบบสอบรวมประเภทภายใน (Internal anchor test) ถ้าแบบสอบรวมประเภทนี้มีความยาวมากเท่าไร ผู้สอบที่เคยรับการทดสอบมาก่อนย่อมได้เปรียบ เพราะเกิดการเรียนรู้หรือจำข้อสอบในส่วน ของแบบสอบรวมได้ ฉะนั้นการระบุจำนวนข้อที่เหมาะสมจะช่วยให้ผลการเทียบมีความถูกต้องมากขึ้น และทำให้การลงทุนทั้งค่าใช้จ่าย และเวลาอยู่ในขอบเขตของความเหมาะสม แองกอฟ (Angoff 1984: 107) ได้แนะนำว่า แบบสอบรวมควรมีความยาวมากพอที่จะก่อให้เกิดสารสนเทศที่จำเป็น ควรมีจำนวนไม่น้อยกว่า ๒๐ ข้อ หรือน้อยกว่าร้อยละ ๒๐ ของจำนวนข้อ แล้วแต่จำนวนโหนดจะมากกว่าในแบบสอบที่กองการเทียบ ไรท์และสโตน (Wright, Stone 1979: 98) แนะนำ ควรเป็นข้อสอบที่วัดเรื่องเดียวกันกับเรื่องที่ทำการวัดในแบบสอบทั้งสองชุด เพียง ๑๐ ข้อก็เพียงพอ การกำหนดคะแนนสมมูลแล้ว จะเห็นว่า ประเด็นความยาวของแบบสอบรวมยังไม่มีข้อแนะนำที่เฉพาะเจาะจง จำเป็นต้องมีการศึกษาก่อนการประยุกต์ คลินและโคเลน (๑๙๘๕) ได้ศึกษาอย่างละเอียดกับกลุ่มตัวอย่างผู้สอบที่ไม่ได้สุ่ม เพื่อตอบคำถามว่า การเทียบมาตรารูปแบบเชิงเส้นตรงควรใช้แบบสอบรวมขนาดกี่ข้อ และได้คำตอบว่า หากกลุ่ม ๒ กลุ่ม มีความสามารถค่างกัน แบบสอบรวมที่ยาวกว่าจะให้ผลต่อความถูกต้องของการแปลงคะแนนได้มากกว่า และจากสถานการณ์ที่ศึกษาพบว่า หากแบบสอบรวมมีขนาดน้อยกว่าร้อยละ ๒๕ ของแบบสอบเทียบมาตราทั้งหมดแล้ว คะแนนแปลงมีความคลาดเคลื่อนสูงเกินกว่าที่จะยอมรับได้ แต่ถ้ามองผู้สอบเปลี่ยนไปเป็นกลุ่มที่มีความสามารถใกล้เคียงกันสองกลุ่ม แบบสอบรวมค่างขนาดไม่ได้ก่อให้เกิดความคลาดเคลื่อนในคะแนนแปลงแตกต่างกัน จากผลการวิจัยของโคเลนนี้ เป็นข้อยืนยันถึงความจำเป็นในการศึกษาถึงประเด็นความยาวของแบบสอบรวมในบริบท

### ของการเทียบมาตราแต่ละกรณี

รูปแบบของการเทียบมาตรา (Equating Models) เป็นอีกองค์ประกอบที่สำคัญมากต่อการเทียบมาตรา เพราะเป็นการกำหนดเงื่อนไข และสร้างกฎการแปลงคะแนนในแต่ละครั้งที่ทำการเทียบมาตรา รูปแบบดั้งเดิมที่ยังมีการใช้อยู่ คือ รูปแบบอีควิเปอร์เซนไทล์ (Equipercentile model) และรูปแบบเชิงเส้นตรง (Linear model) ส่วนรูปแบบอิงทฤษฎีการตอบข้อสอบ (Item response theory equating model: IRT) เป็นรูปแบบใหม่ล่าสุดที่พัฒนาทฤษฎีการวัดความสามารถแฝง (Latent trait theory) รูปแบบอีควิเปอร์เซนไทล์นั้นว่าเป็นรูปแบบเก่าแก่ที่มีวิธีการเทียบที่ให้อัตราส่วนสอดคล้องกับนิยามที่มุ่งกำหนดคะแนนสมมูลจากลำดับคะแนนในแบบสอบสองชุดที่มีคุณสมบัติเป็นแบบสอบสมมูลกัน รูปแบบเชิงเส้นตรงเป็นอีกรูปแบบหนึ่งที่พัฒนามาพร้อม ๆ กับรูปแบบแรก แต่เป็นรูปแบบที่ให้ความสะดวกในขั้นตอนการหาคะแนนแปลง และเป็นวิธีวิเคราะห์ที่ปราศจากความลำเอียงอันจะเกิดจากการปรับเส้นโค้งให้เรียบเช่นวิธีแรก ถ้าข้อมูลคะแนนของแบบสอบสองชุดมีการแจกแจงเหมือนกันแล้ว การใช้รูปแบบเชิงเส้นตรงเป็นวิธีประมาณที่ใกล้เคียงกับรูปแบบอีควิเปอร์เซนไทล์มากที่สุด (Angoff 1984: 87) สำหรับรูปแบบอิงทฤษฎีการตอบข้อสอบเป็นรูปแบบที่มุ่งหาการแปลงที่ระบุคะแนนเทียบเท่าของแบบสอบต่างชุดที่ต่างก็สมนัยกับความสามารถแฝงที่ระดับเดียวกันที่ได้วัดด้วยแบบสอบสองชุดนั้น ๆ (Lord 1977; 1980: 199) ได้มีการศึกษาเปรียบเทียบในเชิงได้เปรียบเสียเปรียบของรูปแบบดั้งเดิม และรูปแบบใหม่กับสถานการณ์ต่าง ๆ เช่น สไลด์และลิน (Slinde and Linn 1977; 1978; 1979) ให้ข้อสังเกตจากการเทียบมาตราระหว่างแบบสอบที่มีความยากแตกต่างกันมาก และระดับความสามารถของกลุ่มตัวอย่างก็ต่างกันด้วยว่า รูปแบบ IRT ชนิดสามพารามิเตอร์ของโลจิสต์ให้ผลการเทียบดีกว่ารูปแบบอื่น บีเทอร์สันและคณะ (Petersen and others 1981) ได้ข้อสรุปจากการศึกษาถึงความคงเส้นคงวาของการเทียบมาตราว่า รูปแบบ IRT ให้ผลดีกว่าในสถานการณ์ของแบบสอบหนึ่ง แต่ให้ผลน้อยกว่าเมื่ออยู่ในอีกสถานการณ์หนึ่ง ทอมป์สันและคณะ (Petersen and others 1982) ได้ชี้ให้เห็นว่านอกจากตัวแปรรูปแบบแล้ว ยังมีตัวแปรเกี่ยวกับความสอดคล้องของเนื้อหา และความยากของแบบสอบที่มีผลกระทบต่อวิธีการเทียบมาตรา จากที่กล่าวมานี้ยังไม่อาจหาข้อสรุปในขณะนี้ว่า รูปแบบการเทียบมาตราใดให้ความถูกต้อง หรือมีประสิทธิภาพเหนือกว่าในภาวะเงื่อนไขต่าง ๆ แต่อย่างไรก็ดี

ผลการศึกษาในอีกทีก็ชี้ให้เห็นถึงข้อได้เปรียบในเชิงทฤษฎีของรูปแบบใหม่ และขณะเดียวกันก็แสดงให้เห็นความเหมาะสมกว่าของรูปแบบดั้งเดิมในภาวะความเป็นจริงบางประการที่กำหนดให้ ลอร์ด (Lord 1977) ก็ชี้ชวนให้เห็นความสำคัญของการศึกษาเชิงประจักษ์ที่รอบคอบ และข้อโต้แย้งของการเทียบมาตรา เพื่อให้ได้ข้อค้นพบที่หนักแน่นพอที่จะสรุปเป็นการทั่วไป และนำไปสู่การประยุกต์ใช้มากยิ่งขึ้น ทั้งนี้ องค์ประกอบด้านสถานการณ์ที่เกี่ยวข้องกับแบบสอบที่กองการเทียบมาตราจำเป็นต้องมีการศึกษาควชูไปค้วย

จากที่ได้อธิบายมานี้ ผู้วิจัยได้เห็นความจำเป็นในการพัฒนาความรู้เกี่ยวกับการเทียบมาตรา เพื่อให้มีส่วนเสริมสร้างความสมบูรณ์ของระบบการทดสอบ ที่จะก่อให้เกิดความถูกต้องในสารสนเทศแก่ผู้ใช้ที่ตัดสินใจ และก่อให้เกิดความยุติธรรม ความเสมอภาคแก่ผู้รับการทดสอบ และยิ่งกว่านั้นผู้วิจัยมีความเห็นว่า น่าจะถึงเวลาแล้วที่นักการศึกษา โดยเฉพาะนักวัดผลทางการศึกษาจะหันมาพิจารณา และตระหนักถึงสิทธิมนุษยชนที่พึงได้จากการถูกทดสอบ ทั้งนี้ การศึกษาเพื่อค้นหาความรู้เรื่องของการเทียบมาตรา เท่ากับเพิ่มทางเลือกสำหรับผู้นำไปประยุกต์ใช้ การวิจัยนี้จึงมีวัตถุประสงค์เพื่อศึกษาในเชิงเปรียบเทียบว่า การเทียบมาตราที่มีรูปแบบ และความยาวแบบสอบรวมทั้งที่ต่างกัน สถานการณ์ของการสอบคัดเลือก และการสอบวัดผลสัมฤทธิ์ซึ่งเป็นสถานการณ์ในบริบทของการเรียนการสอน

วัตถุประสงค์ของการวิจัย

การศึกษาคั้งนี้ มีวัตถุประสงค์เพื่อศึกษาเปรียบเทียบผลจากการเทียบมาตราของแบบสอบสองชุดด้วยรูปแบบการเทียบมาตราที่ต่างกันสามรูปแบบ เมื่อใช้แบบสอบร่วมที่มีความยาวต่างกัน โดยมีวัตถุประสงค์เฉพาะ ดังนี้

- ๑. เพื่อเปรียบเทียบความคลาดเคลื่อนมาตรฐานของการเทียบมาตราเมื่อใช้แบบสอบร่วมที่มีความยาวต่างกันในแต่ละรูปแบบของการเทียบมาตรา
- ๒. เพื่อเปรียบเทียบความเพียงพอของวิธีการเทียบมาตรา อันเกิดจากการใช้แบบสอบร่วมที่มีขนาดต่างกัน และใช้รูปแบบที่ต่างกัน



## สมมติฐานของการวิจัย

จากวัตถุประสงค์ของการวิจัยที่กำหนดไว้แล้ว ท้าแปรที่ตองการศึกษาแบ่ง เป็น ๒ มิติ คือ มิติของความยาวแบบสอบรวม และมิติของรูปแบบการเทียบมาตรา มิติทั้งสองมีทฤษฎี และผลการวิจัย ในอดีตเป็นแนวทางของการกำหนดสมมติฐานของการวิจัยครั้งนี้ ดังต่อไปนี้

มิติของความยาวแบบสอบรวม การเทียบมาตราโดยใช้แบบสอบรวม (Common-item equating) เป็นการออกแบบเพื่อรวบรวมข้อมูลเมื่อกลุ่มตัวอย่างเทียบมาตรา ๒ กลุ่ม ไม่ได้เป็น กลุ่มสมบูรณ์ และแบบสอบเทียบมาตรา ๒ ชุด ไม่ได้เป็นแบบสอบคู่ขนานกัน หน้าที่ของแบบสอบรวม คือ เชื่อมและปรับความแตกต่างอันเนื่องมาจากกลุ่มตัวอย่าง หรือมาจากแบบสอบที่ต่างชุดกัน แบบสอบรวมที่ยาวกว่า มีส่วนทำให้การประมาณค่าความถึในการเทียบรูปแบบอิกวิเปอร์ เช่น โคลโดแมนยักว่า (Braun and Holland 1982: 19-22) มีส่วนทำให้การประมาณค่าเฉลี่ย และความแปรปรวนของประชากรในแบบสอบเทียบมาตราได้แม่นยำกว่า (Braun and Holland 1982: 23-24; Angoff 1984: 97; Budescu 1985) และมีส่วนทำให้การประมาณค่าพารามิเตอร์ของข้อสอบให้อยู่บนมาตราเดียวกัน (Lord 1980: 22) ดังนั้น สมมติฐานของการวิจัย ในมิติของความยาวแบบสอบรวมจึงกำหนดตามเกณฑ์การประสิทธิภาพของการเทียบมาตรา ๒ ประการ ดังนี้

### สมมติฐานข้อที่ ๑

แบบสอบรวมที่มีความยาวมากกว่า มีประสิทธิภาพสัมพัทธ์ของการเทียบมาตราที่สูงกว่า ในทุกรูปแบบของการเทียบมาตราของทั้งสองกรณีของแบบสอบที่ใช้

### สมมติฐานข้อที่ ๒

วิธีการเทียบมาตราที่ใช้แบบสอบรวมที่มีความยาวมากกว่า เป็นวิธีที่มีความเพียงพอกว่า วิธีที่ใช้แบบสอบรวมที่มีความยาวน้อยกว่า

มิติของรูปแบบการเทียบมาตรา รูปแบบการเทียบมาตราเป็นรูปแบบเชิงคณิตศาสตร์ที่เป็น  
 เกณฑ์ของการกำหนดคะแนนสมมูล แต่ละรูปแบบมีนิยามของการกำหนดคะแนนสมมูลที่ต่างกันไป  
 รูปแบบอิกวิเปอร์เซนโตล และรูปแบบเชิงเส้นตรงเป็นรูปแบบที่ให้ความเพียงพอต่อการแปลงคะแนน  
 สมมูลอย่างมาก เมื่อการแจกแจงของคะแนนของแบบสองชุดคล้ายกันมาก (Angoff 1984:  
 98-99; Braun and Holland 1982: 23) ในลักษณะการแจกแจงที่มีพิสัยของคะแนน  
 กว้างมาก รูปแบบอิกวิเปอร์เซนโตลให้ความเพียงพอในการแปลงคะแนนสมมูลโตดีกว่าเมื่อพิสัยแคบ  
 (Angoff 1984: 47) ในลักษณะการแจกแจงที่มีพิสัยแคบ รูปแบบเชิงเส้นตรงเป็นรูปแบบที่  
 ให้คะแนนสมมูลที่มีความเพียงพอมากกว่ารูปแบบอิกวิเปอร์เซนโตล และมีความคงเส้นคงวามากกว่า  
 เพราะเป็นวิธีที่มีลักษณะเป็นปรนัยมากกว่า (Angoff 1984: 97; Potthoff 1982: 209)  
 ส่วนรูปแบบ IRT เป็นวิธีที่อาศัยทฤษฎีการตอบสนองซึ่งมีคุณสมบัติของความไม่แปรเปลี่ยนของค่า  
 ประมาณประชากรตอบสนองเมื่อเปลี่ยนกลุ่มผู้สอบ จึงไม่ได้นิยามที่เกี่ยวข้องกับการแจกแจง แต่จาก  
 การวิจัยเปรียบเทียบความแตกต่างของการใช้รูปแบบ IRT กับรูปแบบอื่น พบว่า รูปแบบ IRT ให้  
 ผลการเทียบมาตราที่มีความเพียงพอมากเมื่อกลุ่มตัวอย่างการเทียบมาตรามีการแจกแจงของคะแนน  
 สองชุดแตกต่างกัน และในกรณีที่รูปแบบอิกวิเปอร์เซนโตล และเชิงเส้นตรงใช้ก่อนนำพอใจแล้ว  
 นำไปใช้กับรูปแบบ IRT ก็พบว่า ให้ผลดีกว่าหรือไม่แตกต่างกัน (Petersen and others  
 1981; Cook, Dunbar and Eignor 1981; Kolen 1981; Kolen and  
 Whitney 1982) จากทฤษฎี และข้อความรู้ที่นำมาบอกถึงที่กล่าวมานี้ เมื่อผู้วิจัยได้เลือก  
 สถานการณ์ของการทดสอบที่ต่างกันสองกรณีศึกษา ซึ่งได้พิจารณาจากสภาพแวดล้อมที่มีส่วนทำให้  
 การแจกแจงของคะแนนมีพิสัยต่างกัน กล่าวคือ กรณีแบบสอบคัดเลือกมีพิสัยคะแนนในประชากรกว้าง  
 กว่ากรณีแบบสอบวัดผลสัมฤทธิ์ จึงได้กำหนดสมมุติฐานเพื่อการวิจัยข้อที่ ๓ และ ๔ ดังนี้

### สมมุติฐานข้อที่ ๓

วิธีการเทียบมาตราที่ใช้รูปแบบอิกวิเปอร์เซนโตล และรูปแบบอิงทฤษฎีการตอบสนอง  
 เป็นวิธีที่มีความเพียงพอมากกว่าวิธีที่ใช้รูปแบบเชิงเส้นตรง ในกรณีแบบสอบคัดเลือก

## สมมุติฐานข้อที่ ๔

วิธีการเทียบมาตราที่ใช้รูปแบบเชิงเส้นตรง และรูปแบบอิงทฤษฎีการตอบข้อสอบ เป็นวิธีที่มีความเที่ยงพอมากกว่าวิธีที่ใช้รูปแบบอิกวีเปอร์เซนไทล์ ในกรณีแบบสอบวัดผลสัมฤทธิ์

### ขอบเขตของการวิจัย

การศึกษากครั้งนี้ ได้จำกัดขอบเขตในค่านของสถานการณ์ของการทดสอบที่เป็นกรอบกำหนดข้อมูลคะแนน และแบบสอบเพื่อการเทียบมาตรา ในค่านความยาวของแบบสอบร่วมที่ใช้ และในค่านรูปแบบการเทียบมาตราไว้ดังนี้

#### ๑. ลักษณะแบบสอบ และสถานการณ์การสอบ

เนื่องจากแบบสอบวัดความรู้ (Cognitive) ที่ใช้อยู่ในปัจจุบัน แบ่งออกได้เป็น ๒ ประเภท ตามโครงสร้าง คือ แบบสอบเชิงความถนัด ซึ่งใช้คัดเลือกบุคคลเข้าทำงาน เป็นแบบสอบที่วัดความรู้ความสามารถทั่วไป ประเภทหลัง คือ แบบสอบวัดผลสัมฤทธิ์ ซึ่งใช้วัดความรู้ในขอบเขตของหลักสูตร และกิจกรรมการเรียนการสอน สำหรับการวิจัยครั้งนี้ ได้ใช้แบบสอบทั้งสองกรณี คือ กรณีแบบสอบคัดเลือก คือ วิชาความรู้ทั่วไป และกรณีแบบสอบวัดผลสัมฤทธิ์ คือ วิชาสถิติวิจัย และวัดผลการศึกษาตามหลักสูตรปริญญาตรีของสถาบันอุดมศึกษาแห่งหนึ่ง

สำหรับสถานการณ์การทดสอบที่ใช้เป็นสถานการณ์สอบจริงที่ผู้สอบทุกคนได้แสดงความสามารถในการตอบคำถามอย่างเต็มที่

#### ๒. ความยาวของแบบสอบรวม

ความยาวของแบบสอบรวม มีสามขนาด โดยคิดเป็นร้อยละของแบบสอบที่นำมาเทียบมาตรา ซึ่งมีชุดละ ๘๕ ข้อ ทั้งนี้ ร้อยละ ๒๐ หรือ ๒๑ ข้อ ร้อยละ ๔๐ หรือ ๑๔ ข้อ และร้อยละ ๒๐ หรือ ๗ ข้อ

### ๓. รูปแบบการเทียบมาตรา

รูปแบบการเทียบมาตรา มีสามรูปแบบ คือ รูปแบบอิกวิเปอร์เซนโตล รูปแบบเชิงเส้นตรง และรูปแบบอิงทฤษฎีการทดสอบแบบโลจิสสามพารามิเตอร์

#### ข้อจำกัดของการวิจัย

##### ๑. ข้อจำกัดในค่านคุณภาพของแบบสอบเทียบมาตรา และแบบสอบรวม

เนื่องจากการวิจัยเชิงประจักษ์โดยใช้สถานการณ์การทดสอบจริง ไม่สามารถควบคุมตัวแปรค่านคุณภาพของแบบสอบ การจัดเตรียมแบบสอบเทียบมาตราสองชุด และแบบสอบรวมจึงดำเนินไปเพื่อให้แบบสอบเทียบมาตรา และแบบสอบรวมมีคุณลักษณะของการวัดความรู้ ความสามารถที่คล้ายคลึงซึ่งกันและกันตามข้อกำหนดของการออกแบบการวิจัยที่อยู่ภายใต้ข้อสอบที่มีอยู่จริง แต่อย่างไรก็ตาม ผลการวิจัยที่สรุปภายใต้เงื่อนไขของบริบทดังกล่าว เป็นสิ่งจำเป็นสำหรับการก้าวไปสู่การสร้างทฤษฎีของการเทียบมาตราที่ชัดเจนขึ้นอีกก้าวหนึ่ง

##### ๒. ข้อจำกัดในค่านความยาวของแบบสอบรวม

สืบเนื่องจากที่กล่าวข้างต้น มีผลทำให้การจัดความยาวของแบบสอบรวมมีจำนวนจำกัด เมื่อแบบสอบที่เลือกมาในแต่ละกรณีมีความยาวสูงสุดฉบับละ ๑๐๐ ข้อ เมื่อจัดให้เป็นแบบสอบเทียบมาตราชุดละ ๑๕ ข้อ ๒ ชุดแล้ว จึงมีจำนวนสำหรับแบบสอบรวมที่ยาวที่สุดไม่เกิน ๓๐ ข้อ แต่เพื่อกำหนดให้เป็นอัตราส่วนเทียบกับร้อยละของแบบสอบรวม จึงกำหนดให้มีสามขนาด คิดเป็นร้อยละ ๒๐ ๔๐ และ ๖๐ ของแบบสอบเทียบมาตรา การใช้แบบสอบรวมที่มีขนาดค่อนข้างสั้นนี้ ข้อสรุปจากผลการวิจัยซึ่งอยู่ในกรอบของบริบท จะเป็นเหตุผลหนึ่งที่ใช้ในการพิจารณาสำหรับในกรณีที่จะมีการใช้แบบสอบรวมที่มีความยาวกว่า ว่า วิธีการเทียบมาตราใดที่ก่อให้เกิดความแม่นยำ และให้ความเพียงพอกับผลของคะแนนแปลงแล้ว วิธีนั้นย่อมให้ผลที่ดีเหนือกว่าเมื่อมีโอกาสใช้แบบสอบรวมที่ยาวกว่า

๓. ข้อจำกัดในการวิเคราะห์ข้อมูลด้วยโปรแกรมคอมพิวเตอร์ LOGIST 5

การวิเคราะห์ข้อมูลคะแนนโดยใช้โปรแกรม LOGIST 5 เพื่อหาค่าประมาณพารามิเตอร์ของข้อสอบ และความสามารถของผู้สอบตามทฤษฎีการตอบสนองรูปแบบโลจิส ได้ดำเนินการภายใต้ข้อจำกัดของระบบคอมพิวเตอร์ที่จัดตั้งโปรแกรมไว้ กล่าวคือ การเรียกใช้คำสั่งอยู่ภายใต้การสั่งครั้งต่อครั้ง ไม่สามารถออกคำสั่งการทำงานต่อเนื่องที่ให้มีการจัดเก็บผลการวิเคราะห์ครั้งก่อนมาเป็นข้อมูลป้อนเข้าเพื่อทำการประมาณค่าในลักษณะวนซ้ำต่อเนื่องได้ ดังนั้น ผลการวิเคราะห์ที่นำมาใช้เมื่อการทำงานตามคำสั่งได้สิ้นสุดลง อาจเป็นค่าประมาณที่ได้รับมาก่อนบรรทัดที่กำหนดในกระบวนการวนซ้ำ ซึ่งมีผลทำให้ค่าประมาณยังมีลักษณะแปรผันอยู่บ้าง

นियามศัพท์เฉพาะ

แบบสอบคัดเลือก หมายถึง แบบสอบวิชาความสามารถทั่วไปที่ใช้ในโปรแกรมการคัดเลือกบุคคลเข้าทำงาน ประกอบด้วยข้อสอบ ๑๐๐ ข้อ

แบบสอบวัดผลสัมฤทธิ์ หมายถึง แบบสอบที่ใช้ในโครงการวัดผลสัมฤทธิ์ทางการเรียนวิชาสถิติวิจัยและวัดผลการศึกษาในระดับปริญญาตรี ประกอบด้วยข้อสอบแบบเลือกตอบ จำนวน ๑๐๐ ข้อ

แบบสอบเทียบมาตรา หมายถึง แบบสอบชุด X และชุด Y มีจำนวนชุดละ ๓๕ ข้อ ที่ได้คัดเลือกแบบสุ่มตามเนื้อหา และค่าสถิติประจำข้อจากจำนวน ๑๐๐ ข้อ แบบสอบชุด X และชุด Y เป็นแบบสอบที่มีความคล้ายคลึงกันในคำเนื้อหา และค่าสถิติประจำข้อเป็นทำนองเดียวกันทั้งกรณีของแบบสอบคัดเลือก และแบบสอบวัดผลสัมฤทธิ์

แบบสอบรวม หมายถึง ข้อสอบจำนวนหนึ่งที่เลือกมาจากข้อสอบ จำนวน ๑๐๐ ข้อ ตามลักษณะทางสถิติ และองค์ประกอบของเนื้อหาเช่นเดียวกับแบบสอบเทียบมาตราโดยไม่มีข้อที่ซ้ำกัน และได้กำหนดไว้สามระดับ คือ แบบสอบรวม V60 ประกอบด้วยข้อสอบ จำนวน ๒๐ ข้อ แบบสอบรวม V40 ประกอบด้วยข้อสอบ จำนวน ๑๕ ข้อ และแบบสอบรวม V20 ประกอบด้วยข้อสอบ จำนวน ๗ ข้อ

กลุ่มตัวอย่างเทียบมาตรา (Equating Samples) หมายถึง กลุ่มตัวอย่างผู้สอบที่ได้รับ การสุ่มอย่างง่ายจากประชากรผู้สอบทั้งหมดของการสอบที่เลือกมาใช้ในการวิจัยครั้งนี้ กรณีแบบสอบ ละสองกลุ่ม เรียกว่า กลุ่ม  $\alpha$  และ  $\beta$  จำนวนกลุ่มละ ๑๕๐๐ คน เพื่อนำผลการสอบมาใช้สร้าง คะแนนแปลงตามรูปแบบการเทียบมาตราต่าง ๆ

กลุ่มตัวอย่างสอบแทนผล (Cross-validation samples) หมายถึง กลุ่มตัวอย่าง ซึ่งได้จากการสุ่มตัวอย่างจากประชากรเกี่ยวกับกลุ่มตัวอย่างเทียบมาตรา โดยไม่มีหน่วยตัวอย่าง ที่ซ้ำกันเลย จำนวน ๑๕๐๐ คน เพื่อนำผลการสอบมาใช้วิเคราะห์สอบแทนผลการเทียบมาตราตาม วิธีต่าง ๆ

วิธีการเทียบมาตรา (Equating methods) หมายถึง วิธีการดำเนินการไปเพื่อให้เกิด การแปลงคะแนนจากมาตรา Y ไปสู่มাত্রา X แต่ละวิธีประกอบด้วยตัวแปรด้านความยาวของแบบ สอบรวม และตัวแปรด้านรูปแบบ วิธีที่ใช้ในการวิจัยนี้มีทั้งหมด ๕ วิธี คือ

- ๑. รูปแบบอิกวิเปอร์เซนไทล์ที่มีแบบสอบรวม V60
- ๒. รูปแบบอิกวิเปอร์เซนไทล์ที่มีแบบสอบรวม V40
- ๓. รูปแบบอิกวิเปอร์เซนไทล์ที่มีแบบสอบรวม V20
- ๔. รูปแบบเชิงเส้นตรงที่มีแบบสอบรวม V60
- ๕. รูปแบบเชิงเส้นตรงที่มีแบบสอบรวม V40
- ๖. รูปแบบเชิงเส้นตรงที่มีแบบสอบรวม V20
- ๗. รูปแบบ IRT ที่มีแบบสอบรวม V60
- ๘. รูปแบบ IRT ที่มีแบบสอบรวม V40
- ๙. รูปแบบ IRT ที่มีแบบสอบรวม V20

ผลการเทียบมาตรฐาน (Equating results) หมายถึง

ก. ความคลาดเคลื่อนมาตรฐานของการเทียบมาตรฐาน (Standard error of equating) คือ รากที่สองของความแปรปรวนแอสิมิโทคของคะแนนแปลงในแต่ละรูปแบบของการเทียบมาตรฐาน ความคลาดเคลื่อนมาตรฐานของการเทียบมาตรฐานนี้ใช้เป็นที่ชี้เปรียบเทียบความแม่นยำของการเทียบมาตรฐานภายในแต่ละรูปแบบของการเทียบมาตรฐาน

ข. กชี้เปรียบเทียบความแตกต่าง (Discrepancy comparison index: C) คือ ค่าเฉลี่ยกำลังสองของความแตกต่างของคะแนนแปลงกับคะแนนเกณฑ์ที่ได้วงน้าหนักภัยส่วนเบี่ยงเบนมาตรฐานของคะแนนเกณฑ์ กชี้นี้ได้จากกรวิเคราะห์กลุ่มสอบทานผล เป็นค่าความคลาดเคลื่อนรวมใช้เปรียบเทียบเพียงพอของวิธีการเทียบมาตรฐาน

ประสิทธิภาพสัมพัทธ์ (Relative efficiency: RE) หมายถึง อัตราส่วนร้อยละของค่าความคลาดเคลื่อนมาตรฐานของการเทียบมาตรฐานที่เกิดจากการใช้แบบสอบรวม V40 หรือ V20 ต่อค่าความคลาดเคลื่อนมาตรฐานของการเทียบมาตรฐานที่ใช้แบบสอบรวม V60 ในรูปแบบการเทียบมาตรฐานเดียวกัน ทั้งสมการ

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

และ

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

เมื่อ RE คือ ประสิทธิภาพสัมพัทธ์  
SEE X\* คือ ความคลาดเคลื่อนมาตรฐานของการเทียบมาตรฐาน

ความเพียงพอทางวิธีการเทียบมาตรา หมายถึง ระบุความถูกต้องของการเทียบคะแนน  
สมมูลจากมาตราของแบบสอบชุด X ไปสู่แบบสอบชุด X ตามกฎการเทียบมาตราที่กำหนด เกณฑ์การ  
ตัดสินความเพียงพอของวิธีการเทียบมาตราที่ใช้ในการวิจัยนี้ คือ ใช้การเปรียบเทียบด้วยค่าดัชนีเปรียบเทียบ  
เทียบความแตกต่าง (C) ถ้าค่า C มาก หมายความว่า การเทียบมาตราด้วยวิธีนั้น ๆ ให้ความ  
เพียงพอ และในทางตรงกันข้าม ค่า C น้อย หมายความว่า วิธีนั้น ๆ ให้ความเพียงพอมาก



ศูนย์วิทยทรัพยากร  
จุฬาลงกรณ์มหาวิทยาลัย