



บทที่ 1

บทนำ

### ความสำคัญของปัญหา

การสอบมีความจำเป็นในหลายวงการไม่ว่าจะเป็นวงการทหาร วงการอุตสาหกรรม วงการธุรกิจ โดยเฉพาะอย่างยิ่งในวงการศึกษ การสอบในแต่ละครั้งมีเป้าหมายแตกต่างกันออกไป เช่น เพื่อคัดเลือกคนเข้าทำงาน เพื่อเลื่อนชั้น เพื่อเข้าศึกษาต่อ เพื่อให้ระดับคะแนนแก่นักเรียน และอื่น ๆ อีกมากมาย ปัญหาที่พบในการสอบที่มีผู้สอบจำนวนมาก ๆ ได้แก่ การจัดเตรียมการสอบ และการดำเนินการสอบ และต้องมีการจัดการสอบทั่วประเทศพร้อมกันโดยใช้สถานที่สอบเพียงแห่งใดแห่งหนึ่ง ปัญหาที่ตามมาได้แก่ การจราจรติดขัดเพราะทุกคนต้องเดินทางเพื่อไปยังสนามสอบ บางครั้งไม่ทันเข้าสอบในวิชาแรก สถานที่สอบไม่เพียงพอ กรรมการคุมสอบขาดแคลน และสิ้นเปลืองค่าใช้จ่าย สาเหตุสำคัญที่อยู่เบื้องหลังของการที่ต้องจัดสอบพร้อมกันคือ การรั่วไหลของข้อสอบ เพราะถ้าไม่จัดสอบให้พร้อมกันแล้วถ้าหากมีการใช้ข้อสอบฉบับเดียวกันคนสอบที่หลังย่อมได้เปรียบคนที่สอบก่อนเนื่องจากมีโอกาสที่จะได้รับรู้ข้อสอบมากกว่า ผลการสอบที่ออกมาจึงไม่มีความยุติธรรม(Linn, 1989)และถ้าจะให้เกิดความยุติธรรมแก่ทุกฝ่ายก็ต้องจัดการสอบใหม่เพื่อหาความสามารถที่แท้จริง

แนวทางการแก้ปัญหาหนึ่งประการหนึ่งคือการใช้ข้อสอบคู่ขนานที่วัดในเนื้อหาเดียวกัน และมีความยากเท่าเทียมกันแต่การสร้างข้อสอบคู่ขนานทำได้ยากในทางปฏิบัติ ซึ่งลินน์(Linn, 1986) ได้กล่าวว่า แม้ว่าแบบสอบหลายฉบับที่สร้างขึ้นมาจากมีเนื้อหาเดียวกันเป็นข้อสอบประเภทเดียวกัน และพิสัยความยากอยู่ในระดับเดียวกัน แต่ตัวคำถามที่ใช้ในแต่ละฉบับย่อมมีความแตกต่างกัน ดังนั้นแบบสอบสองฉบับใด ๆ จึงไม่มีโอกาสที่จะมีความเท่าเทียมกันในการวัดและระดับความยากผลที่ตามมาก็คือ คะแนนดิบที่ได้จากการสอบไม่สามารถที่จะนำมาเปรียบเทียบกันได้ หรือเมื่อเปรียบเทียบกันก็ไม่มี ความยุติธรรมต่อผู้สอบที่ทำแบบสอบฉบับที่มีความยากสูงกว่า ทางออกที่ดีที่สุดในการแก้ปัญหานี้คือ การใช้เทคนิคทางสถิติมาปรับคะแนนให้เกิดความเท่าเทียมกันกระบวนการเช่นนี้เรียกว่าการปรับเทียบคะแนน (Test Equating)



เบรนนอน (Brennan, 1987) ได้กล่าวถึงความจำเป็นในการปรับเทียบคะแนนประการหนึ่งว่า ปัจจุบันนี้ได้มีการสอบจำนวนมากที่ต้องใช้ข้อสอบหลาย ๆ ฉบับ ผู้ที่รับผิดชอบในการสอบต้องตระหนักว่า คะแนนที่ได้จากแบบสอบเหล่านั้นต้องมีการปรับเทียบคะแนน เพื่อให้ 2 คนที่ได้ทำข้อสอบฉบับยากกว่าไม่เกิดการเสียเปรียบ

แองกอฟ (Angoff, 1984) ได้ให้ความหมายของการปรับเทียบคะแนนระหว่างแบบสอบว่าเป็นกระบวนการปรับระบบคะแนนของแบบสอบหนึ่ง ให้อยู่บนสเกลของแบบสอบฉบับอื่นที่มุ่งวัดในสิ่งเดียวกัน เพื่อให้คะแนนสามารถเปรียบเทียบกันได้ สกักส์ (Skaggs, 1986) ได้ชี้ว่า การที่จะนำคะแนนจากแบบสอบทั้งสองมาปรับเทียบกันนั้นแบบสอบทั้งสอง จะต้องวัดในสิ่งเดียวกันต้องมีความเที่ยงเท่ากันและคู่ขนานกัน การปรับเทียบคะแนนระหว่างแบบสอบได้มีผู้ศึกษาและพัฒนาขึ้นมาเรื่อย ๆ จึงเกิดวิธีการปรับเทียบแบบต่าง ๆ ได้แก่ การปรับเทียบคะแนนตามแนวตั้ง (Vertical Equating) และการปรับเทียบคะแนนตามแนวระดับ (Horizontal Equating) แฮมเบิลตันและ สวามินาธาน (Hambleton และ Swaminathan, 1985) ได้อธิบายถึงการปรับเทียบคะแนนตามแนวตั้งว่าเป็นการปรับเทียบ คะแนนระหว่างแบบสอบที่มีความยากแตกต่างกัน และผู้สอบมีความสามารถแตกต่างกัน ส่วนการปรับเทียบคะแนนตามแนวระดับเป็นการปรับเทียบคะแนนระหว่างแบบสอบสองฉบับที่มีความยาก ใกล้เคียงกันและกลุ่มผู้สอบมีความสามารถคล้ายกัน

การปรับเทียบคะแนนนอกจากจะจำแนกตามความแตกต่างของความยากและความสามารถของผู้สอบคือการปรับเทียบตามแนวตั้งและแนวระดับแล้ว ยังสามารถจำแนกตามทฤษฎีการวัดที่ใช้ได้อีกคือการปรับเทียบตามแนวทฤษฎีการวัดแบบดั้งเดิม (Classical Test Theory) และการปรับเทียบตามแนวทฤษฎีการตอบสนองรายข้อ (Item Response Theory หรือ IRT) (Skaggs, 1986) การปรับเทียบตามแนวทฤษฎีการวัดแบบดั้งเดิมได้แก่ วิธีการปรับเทียบแบบอีควิเปอร์เซ็นต์ (Equipercentile Equating) และการปรับเทียบเชิงเส้นตรง (Linear Equating) ส่วนการปรับเทียบตามแนว IRT ได้จำแนกออกเป็นการปรับเทียบตามโมเดล 1 พารามิเตอร์หรือบางครั้งนิยมเรียกว่าราชัจโมเดล (Rasch Model) การปรับเทียบตามโมเดล 2 พารามิเตอร์ (Two Parameter Model) และการปรับเทียบคะแนนตามโมเดล 3 พารามิเตอร์ (Three Parameter Model)

วิธีการปรับเทียบคะแนนเป็นองค์ประกอบที่สำคัญประการหนึ่งต่อการปรับเทียบคะแนน เพราะเป็นการกำหนดเงื่อนไขบางประการในการปรับเทียบคะแนน การศึกษาในเรื่องนี้ได้แก่



งานวิจัยของ สไลด์ และลินน์ (Slind และ Linn, 1977, 1979) ซึ่งให้ข้อสังเกตจากการเปรียบเทียบคะแนนระหว่างแบบสอบที่มีความยากแตกต่างกัน และระดับความสามารถของผู้สอบแตกต่างกันว่าโมเดล IRT ชนิด 3 พารามิเตอร์ ให้ผลการปรับเทียบน่าเชื่อถือได้ สแกกส์ (Skaggs, 1986) ให้ความเห็นว่าในการปรับเทียบคะแนนไม่มีวิธีการใดที่ดีที่สุด ทั้งนี้ขึ้นอยู่กับสถานการณ์ของการปรับเทียบ อย่างไรก็ตามการปรับเทียบตามแนวทฤษฎีการตอบสนองรายข้อดูเหมือนจะให้ผลน่าเชื่อถือได้ และเนื่องจากเงื่อนไขการปรับเทียบคะแนนจะต้องมีความเสมอภาค (equity) มีความสมมาตร (symmetry) และมีความไม่แปรเปลี่ยน (invariance) ดังนั้นถ้าใช้วิธีการปรับเทียบตามแนวทฤษฎีแบบดั้งเดิมแล้วไม่สามารถเป็นไปตามเงื่อนไขได้ การปรับเทียบตามแนวทฤษฎี IRT เท่านั้นที่เป็นไปตามเงื่อนไขได้ (Hambleton และ Swaminathan, 1985) อันเป็นเหตุผลหนึ่งที่ทำให้การวิจัยครั้งนี้เลือกศึกษาเฉพาะการปรับเทียบตามแนวทฤษฎีการตอบสนองรายข้อ

นอกจากนี้การปรับเทียบคะแนนยังมีความแตกต่างกันในเรื่องแบบแผนการทดลองหรือแบบแผนการเก็บรวบรวมข้อมูล (Experimental Design) เรื่องนี้แองกอฟ (Angoff, 1982, 1984) ได้แบ่งแบบแผนการเก็บรวบรวมข้อมูลไว้ 3 แบบด้วยกันคือ แบบแผนกลุ่มสมมูล (Equivalent Group Design) ตามแบบแผนนี้ประกอบด้วยกลุ่มตัวอย่างสองกลุ่ม แต่ละกลุ่มทำแบบสอบคนละฉบับที่ต้องการนำมาปรับเทียบกัน เนื่องจากกลุ่มตัวอย่างนี้ได้มาโดยการสุ่มจึงคาดหมายว่ากลุ่มตัวอย่างทั้งสองนั้นมีความสามารถใกล้เคียงกัน ในแบบแผนที่สองคือแบบแผนกลุ่มเดี่ยว (Single Group Design) ตามแบบแผนนี้แบบสอบทั้งสองที่จะนำมาปรับเทียบจะถูกนำไปสอบกับกลุ่มตัวอย่างเดี่ยวกันเพียงกลุ่มเดียว ส่วนในแบบแผนที่สามคือแบบแผนข้อสอบร่วม (Anchor Test Design) ตามแบบแผนนี้ในแบบสอบแต่ละฉบับที่จะนำไปสอบกับกลุ่มตัวอย่าง 2 กลุ่มจะมีข้อสอบร่วมอยู่ด้วยข้อสอบร่วมนี้อาจปะปนอยู่ในแบบสอบซึ่งเรียกว่าข้อสอบร่วมภายใน (Internal Anchor Test) หรืออาจจัดเป็นฉบับแยกต่างหากซึ่งเรียกว่าข้อสอบร่วมภายนอก (External Anchor Test) ก็ได้ การที่มีข้อสอบร่วมนี้จะช่วยในการปรับความแตกต่างระหว่างกลุ่มผู้สอบได้

เกี่ยวกับประเภทของข้อสอบร่วม ลอยด์ (Loyd, 1983 อ้างถึงใน Skaggs, 1986) ได้ตรวจสอบการจัดข้อสอบร่วมทั้งสองแบบและใช้การปรับเทียบคะแนนตามโมเดล IRT ชนิด 1, 2 และ 3 พารามิเตอร์ ผลการวิจัยพบว่าแบบสอบร่วมชนิดภายนอกให้ผลการปรับเทียบที่พึงพอใจ



มากกว่าแบบสอบร่วมชนิดภายใน ส่วนการวิจัยเกี่ยวกับจำนวนข้อสอบร่วม ได้มีผู้ ศึกษาไว้หลายคน เช่น แองกอฟ (Angoff, 1984) เสนอแนะว่าแบบสอบร่วมควรมีความยาวมากพอที่จะก่อให้เกิด สารสนเทศ และความไม่น้อยกว่า 20 ข้อ หรือไม่น้อยกว่าร้อยละ 20 ของจำนวนข้อสอบที่นำมา ปรับเทียบโดยเลือกจำนวนที่มีค่ามากอย่างใดอย่างหนึ่ง ไรท์และ สโตน (Wright และ Stone, 1979) แนะนำว่าข้อสอบร่วมควรเป็นข้อสอบที่วัดเนื้อหาเดียวกันกับข้อสอบทั้งฉบับ และจำนวน เพียง 10 ข้อก็เพียงพอต่อการกำหนดคะแนนสมมูล ส่วนเวลและคณะ (Vale และคณะ, 1981) ได้ศึกษาโดยจำลองข้อมูลตามโมเดล IRT โดยใช้ข้อสอบร่วมจำนวน 5, 15 และ 25 ข้อ พบว่า ข้อสอบร่วม 15 ถึง 20 ข้อมีความจำเป็น แม็คคินเลย์และเร็คเคส (Mckinley และ Reckase, 1981) ได้ศึกษาจำนวนข้อสอบของแบบสอบร่วมโดยใช้ข้อมูลจริงจากการสอบผลสัมฤทธิ์ทางการเรียน ในหลาย ๆ วิชา และได้ข้อสรุปว่าข้อสอบร่วมเพียง 5 ข้อไม่มีความเพียงพอในการปรับเทียบ คะแนน แต่ถ้าใช้ข้อสอบร่วม 25 ข้อจะได้ผลดีกว่า 15 ข้อ คลีนและโคเลน (Klein และ Kolen, 1985) ได้ศึกษากับกลุ่มตัวอย่างของผู้สอบที่ไม่มี การสุ่มโดยใช้การปรับเทียบคะแนน เชิงเส้นตรง พบว่าถ้าสองกลุ่มมีความสามารถแตกต่างกันแบบสอบร่วมที่ยาวกว่าให้ผลในการแปลง คะแนนได้ถูกต้องมากกว่าและถ้าหากแบบสอบมีขนาดร้อยละ 24 ของแบบสอบที่ใช้ในการปรับเทียบ คะแนนแล้วคะแนนแปลงเกิดความคลาดเคลื่อนสูงกว่าที่จะยอมรับได้ ซึ่งผลนี้สอดคล้องกับทฤษฎี (2528) ที่ศึกษาโดยใช้ข้อสอบร่วมในการสอบคัดเลือกคนเข้าทำงานและในการสอบวัดผลสัมฤทธิ์ ทางการเรียน โดยใช้ข้อสอบร่วมที่มีความยาวเป็นร้อยละ 20, 40 และ 60 ของแบบสอบที่นำมา ปรับเทียบคะแนน พบว่าเมื่อความยาวของแบบสอบร่วมเพิ่มขึ้นความคลาดเคลื่อนมาตรฐานของ การปรับเทียบจะน้อยลง จากลักษณะและจำนวนข้อสอบร่วมที่ได้อภิปรายมาพอสรุปได้ว่า แบบสอบร่วมแบบภายนอกได้ผลดีกว่าแบบสอบร่วมแบบภายใน ส่วนในเรื่องจำนวนข้อยังไม่ได้ ข้อสรุปที่ชัดเจนว่าจำนวนข้อสอบร่วมเท่าใดจึงจะพอเหมาะ แต่ที่สรุปได้ก็คือจำนวนข้อสอบร่วม ยิ่งมากความคลาดเคลื่อนในการปรับเทียบยิ่งน้อยลง

ในการปรับเทียบคะแนน เพเชียน (Patience, 1981) ได้แนะนำว่าองค์ประกอบที่ทำให้ การปรับเทียบคะแนนโดยใช้โมเดล IRT 3 พารามิเตอร์ได้ผลไม่น่าพอใจคือ ขนาดกลุ่มตัวอย่าง ที่เล็กเกินไป โควเวลล์ (Cowell, 1981) ได้ศึกษาในเรื่องนี้โดยใช้ขนาดของกลุ่มตัวอย่างเป็น ตัวแปรอิสระ ข้อมูลที่ใช้คือคะแนนจากแบบสอบ TOEFL โดยใช้การปรับเทียบตามแนว IRT ขนาดกลุ่มตัวอย่างที่ใช้มีขนาดเล็กตั้งแต่ 300 จนถึงขนาดใหญ่ 2,000-3,000 ผลการวิจัยพบว่า



กลุ่มตัวอย่างขนาดเล็กสามารถทำให้ผลการเปรียบเทียบมีความคงตัวได้ โคลเลนและวิทเนย์ (Kolen และ Whitney, 1982) ใช้ข้อมูลจากแบบสอบพัฒนาการศึกษาทั่วไป (GED) พบว่ากลุ่มตัวอย่างขนาดเล็ก(170-190)ในโมเดล IRT 3 พารามิเตอร์จะให้พารามิเตอร์ข้อสอบในลักษณะสุดโต่ง อเยอร์เว (Ayerve, 1992) ได้ศึกษาการเปรียบเทียบคะแนนตามแนวคิดโดยใช้วิธีอควิเปอร์เซ็นไต์ล และโมเดล IRT 3 พารามิเตอร์ และใช้ความยาวแบบสอบและขนาดกลุ่มตัวอย่างเป็นตัวแปร ขนาดของกลุ่มตัวอย่างที่ใช้คือ 200, 500 และ 1,000 และความยาวของแบบสอบคือ 30 และ 60 ข้อ พบว่าสำหรับโมเดล IRT 3 พารามิเตอร์ ขนาดของกลุ่มตัวอย่างเป็นปัจจัยสำคัญโดย กลุ่มตัวอย่างขนาด 200 ให้ผลการเปรียบเทียบที่ไม่เที่ยงตรง ขณะที่กลุ่มตัวอย่างขนาด 500 และ 1000 ให้ผลที่เที่ยงตรงกว่า สำหรับแบบอควิเปอร์เซ็นไต์ลความยาวของแบบสอบเป็นองค์ประกอบที่สำคัญ สแกกส์ และ ลิสสิท (Skaggs และ Lissitz, 1986) แนะนำว่า ขนาดกลุ่มตัวอย่าง จะมากหรือน้อย ขึ้นอยู่กับลักษณะของแบบแผนการเปรียบเทียบคะแนน ฮิวลิน ลิสส์ค และดราสโกว์ (Hulin, Lissak และ Drasgow, 1982 อ้างถึงใน Skaggs และ Lissitz, 1986) ศึกษา ขนาดกลุ่มตัวอย่าง ความยาวแบบสอบ และการประมาณค่าพารามิเตอร์ด้วยโปรแกรม LOGIST โดยใช้การจำลองข้อมูลตามโมเดล IRT 3 พารามิเตอร์ และพารามิเตอร์ความสามารถแจกแจง แบบโค้งปกติพบว่าขนาดของกลุ่มตัวอย่าง และความยาวของแบบสอบมีผลต่อการประมาณค่า พารามิเตอร์ โดยเมื่อประมาณค่าใช้ข้อสอบ 60 ข้อ กับกลุ่มผู้สอบ 1,000 คน พบว่าให้ผลออกมา ในลักษณะคงที่ ถ้าลดข้อสอบลงครึ่งหนึ่งคือเหลือ 30 ข้อ และเพิ่มขนาดกลุ่มตัวอย่างเป็น 2 เท่า ให้ผลที่ใกล้เคียงกัน และพบว่าโปรแกรม LOGIST ใช้กับกลุ่มตัวอย่าง 15 คนไม่ได้

จะเห็นได้ว่าในเรื่องขนาดกลุ่มตัวอย่างนั้นมีการศึกษา 2 แนว คือกลุ่มตัวอย่างสำหรับการประมาณค่าพารามิเตอร์ และกลุ่มตัวอย่างสำหรับการเปรียบเทียบคะแนนระหว่างแบบสอบ โดยเฉพาะในเรื่องการเปรียบเทียบคะแนนยังไม่ได้ข้อสรุปที่แน่นอนแต่ที่ได้ข้อสรุปคือขนาดกลุ่มตัวอย่าง ที่มากจะให้ความคลาดเคลื่อนน้อยลง ในการวิจัยครั้งนี้จึงได้ศึกษาขนาดกลุ่มตัวอย่างขั้นต่ำในการ เปรียบเทียบคะแนนให้เกิดประสิทธิภาพหรือให้ผลไม่แตกต่างจากกลุ่มตัวอย่างขนาดใหญ่

ที่กล่าวมาทำให้เห็นว่าความยาวแบบสอบเป็นองค์ประกอบที่มีผลต่อการเปรียบเทียบคะแนน ผู้วิจัยจึงกำหนดความยาวแบบสอบเป็นตัวแปร โดยกำหนดความยาวแบบสอบตามที่ได้พบเห็นในการ สอบต่างๆไป ซึ่งเริ่มจาก 30 ข้อ และเพิ่มขึ้นครั้งละ 20 ข้อคือเป็น 50 ข้อ 70 ข้อ 90 ข้อ 110 ข้อ 130 ข้อ และ 150 ข้อ เป็นความยาวแบบสอบที่ศึกษาทั้งหมด 7 ค่า



ในเรื่องแบบแผนการปรับเทียบคะแนน ถึงแม้จะมี 3 แบบแผน คือ แบบแผนกลุ่มเดี่ยว แบบแผนกลุ่มสมมูล และแบบแผนข้อสอบร่วม แต่การวิจัยเท่าที่ผ่านมามักจะไม่ค่อยพบเห็นการศึกษาตามแบบแผนกลุ่มเดี่ยว เพราะในทางปฏิบัติแบบแผนการปรับเทียบกลุ่มเดี่ยวนี้นำมาใช้น้อยมาก หรือแทบไม่ใช้เลย ดังนั้นการวิจัยครั้งนี้จึงเลือกศึกษาแบบแผนการปรับเทียบแบบกลุ่มสมมูลและแบบแผนข้อสอบร่วม

ทฤษฎีที่ใช้ในการปรับเทียบที่ศึกษา มีทั้งตามแนวทฤษฎีการวัดแบบดั้งเดิม ได้แก่ แบบอควิเปอร์เซ็นไต์ลด์ แบบเส้นตรง และศึกษาตามแนวทฤษฎีการตอบสนองรายข้อ แต่เนื่องจากทฤษฎีการวัดแบบดั้งเดิมนั้นไม่สามารถแก้ปัญหาบางจุดได้คือ ความเสมอภาค ความสมมาตร และความไม่แปรเปลี่ยน ด้วยเหตุนี้ผู้วิจัยจึงเลือกศึกษาตามแนวทฤษฎีการตอบสนองรายข้อซึ่งสามารถแก้ปัญหาเหล่านี้ได้ และโมเดลที่นิยมใช้กันในทฤษฎีการตอบสนองรายข้อ คือ 1 และ 3 พารามิเตอร์ การศึกษาค้นคว้าครั้งนี้จึงเลือกศึกษาตามโมเดล 1 และ 3 พารามิเตอร์นี้

ในแบบแผนการปรับเทียบคะแนนแบบข้อสอบร่วม ที่กล่าวมาข้างต้นมักศึกษาหาความยาวแบบข้อสอบร่วมที่ทำให้การปรับเทียบคะแนนเกิดประสิทธิภาพ ดังนั้นในแบบแผนข้อสอบร่วม (CIE) ที่ใช้ในการวิจัยครั้งนี้ผู้วิจัยจึงไม่ศึกษาความยาวของแบบข้อสอบร่วมในลักษณะที่เป็นสัดส่วนกับฉบับที่จะนำมาปรับเทียบกัน แต่ศึกษาว่าขนาดกลุ่มตัวอย่างขั้นต่ำควรเป็นเท่าใดเมื่อความยาวแบบข้อสอบที่แตกต่างกันโดยมีข้อสอบร่วมเป็นแบบภาษาในที่จำนวนข้อละ 20 ตามข้อเสนอของแองกอฟ (Angoff, 1984)

จากการศึกษางานวิจัยที่เกี่ยวข้องกับการปรับเทียบคะแนนยังพบว่าได้มีการศึกษาในลักษณะการเก็บข้อมูลจริงจากการสอบของกลุ่มตัวอย่างและอีกลักษณะหนึ่งคือ การจำลองข้อมูลโดยใช้คอมพิวเตอร์แล้วนำข้อมูลนั้นมาประกอบการศึกษาค้นคว้า ซึ่งเรียกว่า Simulation หรือ Monte Carlo เมื่อเปรียบเทียบข้อดีและข้อเสียของการใช้ข้อมูลสองวิธีนี้พบว่า วิธีการเก็บข้อมูลจริงมักมีความคลาดเคลื่อนเกิดขึ้นเสมอ เนื่องจากมีตัวแปรแทรกซ้อนมากผู้เก็บข้อมูลไม่สามารถควบคุมตัวแปรเหล่านั้นได้ (ภาวิณี, 2528) ศึกษาตัวแปรได้น้อย และยังมีสิ่งเปลี่ยนแปลงทรัพยากร (Cook และ Petersen, 1987) แนะนำว่าควรได้มีการศึกษาตัวแปรที่เกี่ยวข้องกับการปรับเทียบคะแนนโดยใช้วิธีการจำลองข้อมูลเพราะจะทำให้ได้ผลที่ชัดเจนกว่า



จากที่กล่าวมาตัวแปรที่เกี่ยวข้องกับการเปรียบเทียบคะแนนมีมากมาย ได้แก่ ความยาวของแบบสอบถาม ชนิดของแบบสอบถาม ความยาวของแบบสอบ วิธีการเก็บรวบรวมข้อมูล แบบแผนการเปรียบเทียบ ขนาดกลุ่มตัวอย่าง วิธีการประเมินความคลาดเคลื่อน แต่ที่ยังเป็นปัญหาและยังหาข้อสรุปไม่ได้ที่สำคัญคือ ขนาดของกลุ่มตัวอย่าง สำหรับในเรื่องการประมาณค่าพารามิเตอร์แล้ว กลุ่มตัวอย่างที่ใหญ่ให้การประมาณค่าที่เที่ยงตรงกว่า และในเรื่องการเปรียบเทียบคะแนนตามทฤษฎีการตอบสนองรายชื่อก่อนที่จะนำคะแนนมาเปรียบเทียบกัน ก็ต้องผ่านกระบวนการประมาณค่าพารามิเตอร์ด้วย ดังนั้นขนาดกลุ่มตัวอย่างจึงมีผลต่อการเปรียบเทียบคะแนนและตามหลักนี้กลุ่มตัวอย่างที่ใหญ่คือประมาณ 3,000 ให้ผลการเปรียบเทียบคะแนนที่มีประสิทธิภาพซึ่งในทางปฏิบัติสำหรับการสอบทั่วไปมักหากกลุ่มตัวอย่างขนาดใหญ่เช่นนี้ได้ลำบาก เมื่อผู้วิจัยสนใจหาขนาดกลุ่มตัวอย่างขั้นต่ำในการเปรียบเทียบคะแนนที่ให้ผลไม่แตกต่างจากกลุ่มตัวอย่างขนาดใหญ่ซึ่งในการศึกษาลักษณะนี้ถ้าใช้ข้อมูลจริงมีขีดจำกัดต่อการศึกษาเรื่องนี้มาก เพราะต้องมีการปรับเปลี่ยนกลุ่มตัวอย่างขนาดต่างๆ โดยเริ่มจากขนาดใหญ่ เช่น เริ่มจาก 3,000 ลดลงมาเป็นกลุ่มตัวอย่างขนาดเล็กกว่าจะได้คำตอบต้องใช้กลุ่มตัวอย่างเป็นหลายหมื่นคน จึงทำให้ต้องสิ้นเปลืองทรัพยากรอย่างมหาศาล และถ้ายิ่งกลุ่มตัวอย่างที่เลือกมาเป็นกลุ่มเดิมหรือคนเดิม ผู้สอบจะเกิดความเมื่อยล้าและเกิดการเรื้อรังจากการสอบตอนแรกจะส่งผลไปยังการสอบครั้งต่อไป (Linn, 1989) การจำลองข้อมูลจึงสามารถแก้ปัญหาที่กล่าวมาได้ สำหรับการจำลองข้อมูลในการเปรียบเทียบคะแนนระหว่างแบบสอบในแนวดิ่ง อเยอร์เว (Ayerve, 1992) ได้ทำการศึกษามาแล้ว ในการวิจัยครั้งนี้จึงสนใจศึกษาการเปรียบเทียบคะแนนตามแนวระดับ

### วัตถุประสงค์ของการวิจัย

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อหาขนาดของกลุ่มตัวอย่างขั้นต่ำ ที่ทำให้การเปรียบเทียบคะแนนในแนวระดับ ตามทฤษฎีการตอบสนองรายชื่อ โมเดล หนึ่ง และสามพารามิเตอร์ ให้เกิดประสิทธิภาพ โดยมีวัตถุประสงค์เฉพาะดังนี้

1. เพื่อหาขนาดของกลุ่มตัวอย่างขั้นต่ำในการเปรียบเทียบคะแนนในแนวระดับตามแนวทฤษฎีการตอบสนองรายชื่อโมเดล หนึ่ง และ สามพารามิเตอร์ ในแบบแผนกลุ่มสมมูลเมื่อความยาวของแบบสอบมีขนาดต่าง ๆ กัน



2. เพื่อหาขนาดของกลุ่มตัวอย่างขั้นต่ำในการปรับเทียบคะแนนในแนวระดับตามแนว ทฤษฎีการตอบสนองรายข้อโมเดล หนึ่ง และ สามพารามิเตอร์ ในแบบแผนข้อสอบร่วมเมื่อ ความยาวของแบบสอบร่วมมีขนาดต่าง ๆ กัน

3. เพื่อเปรียบเทียบขนาดของกลุ่มตัวอย่างขั้นต่ำในการปรับเทียบคะแนนในแนวระดับ ตามแนวทฤษฎีการตอบสนองรายข้อโมเดล หนึ่ง และ สามพารามิเตอร์ ในแบบแผนกลุ่มสมมูล เมื่อความยาวของแบบสอบมีขนาดต่าง ๆ กัน

4. เพื่อเปรียบเทียบขนาดของกลุ่มตัวอย่างขั้นต่ำในการปรับเทียบคะแนนในแนวระดับ ตามแนว ทฤษฎีการตอบสนองรายข้อโมเดล หนึ่ง และ สามพารามิเตอร์ ในแบบแผนข้อสอบ ร่วมเมื่อความยาวของแบบสอบมีขนาดต่าง ๆ กัน

5. เพื่อเปรียบเทียบขนาดกลุ่มตัวอย่างขั้นต่ำในการปรับเทียบคะแนนในแนวระดับตาม แบบแผนกลุ่มสมมูลและแบบแผนข้อสอบร่วม ตามโมเดล 1 และ 3 พารามิเตอร์เมื่อความยาว แบบสอบมีขนาดต่าง ๆ กัน

#### สมมติฐานการวิจัย

จากหลักการเชิงทฤษฎีและผลการวิจัยของ เรคเคส (Reckase, 1979) โคเวลล์ (Cowell, 1981) ฮิวลิน ลิสซัก และ ดราสโกว์ (Hulin, Lissak, และ Drasgow, 1982) คาลด์เวลล์ (Caldwell, 1984) สแกกส์ (Skaggs, 1984) และคนอื่น ๆ ที่ได้ศึกษามา (รายละเอียดแสดงในบทที่ 2) สามารถนำมาเป็นแนวทางในการกำหนดสมมติฐานการวิจัยครั้งนี้ ซึ่งผู้วิจัยกำหนดสมมติฐานการวิจัยไว้ว่า

1. ขนาดของกลุ่มตัวอย่างขั้นต่ำที่ใช้วิธีการปรับเทียบคะแนนในแนวระดับตามแบบแผน กลุ่มสมมูล โมเดล 3-พารามิเตอร์ น่าจะมีขนาดใหญ่กว่าขนาดของกลุ่มตัวอย่างขั้นต่ำของโมเดล 1-พารามิเตอร์

2. ขนาดของกลุ่มตัวอย่างขั้นต่ำที่ใช้วิธีการปรับเทียบคะแนนในแนวระดับตามแบบแผน ข้อสอบร่วม โมเดล 3-พารามิเตอร์ น่าจะมีขนาดใหญ่กว่าขนาดของกลุ่มตัวอย่างขั้นต่ำของโมเดล 1-พารามิเตอร์



3. ความยาวของแบบสอบขนาด 150 ข้อในแบบแผนกลุ่มสมมูลและแบบแผนข้อสอบรวมทั้งโมเดล 1 พารามิเตอร์และ 3 พารามิเตอร์ น่าจะทำให้ขนาดของกลุ่มตัวอย่างขั้นต่ำที่น้อยกว่าความยาวแบบสอบขนาด 130 ข้อ 110 ข้อ, 90 ข้อ, 70 ข้อ, 50 ข้อ และ 30 ข้อในลักษณะที่รองลงมาตามลำดับ

4. ขนาดของกลุ่มตัวอย่างขั้นต่ำในแบบแผนกลุ่มสมมูลและแบบแผนข้อสอบรวมทั้งโมเดล 1 และ 3 พารามิเตอร์ ในแต่ละความยาวแบบสอบน่าจะมีขนาดแตกต่างกัน

### ขอบเขตของการวิจัย

1) การวิจัยครั้งนี้มีขอบเขตการศึกษาอยู่ที่การปรับเทียบคะแนนในแนวระดับตามแนวทฤษฎีการตอบสนองรายข้อ(IRT) ตามโมเดล 1 และ 3 พารามิเตอร์ โดยการแจกแจงของพารามิเตอร์ b เป็นโค้งปกติส่วนพารามิเตอร์ a และ c มีการแจกแจงแบบสม่ำเสมอ(uniform) และใช้แบบแผนการปรับเทียบเป็นแบบกลุ่มสมมูล และแบบข้อสอบร่วมชนิดภายใน

2) ตัวแปรในการวิจัยประกอบด้วย

ตัวแปรอิสระ ได้แก่ โมเดลของการตอบสนองรายข้อ แบบแผนการปรับเทียบคะแนน และความยาวของแบบสอบ

ตัวแปรตาม ได้แก่ ขนาดของกลุ่มตัวอย่างขั้นต่ำ

### นิยามคำศัพท์

1) การปรับเทียบคะแนน(test equating) หมายถึงกระบวนการนำคะแนนที่ได้จากแบบสอบ 2 ฉบับมาปรับให้อยู่บนสเกลเดียวกันเพื่อให้คะแนนจากแบบสอบทั้งสองฉบับนั้นสามารถเปรียบเทียบกันได้ สำหรับในการวิจัยครั้งนี้หมายถึงการนำคะแนนที่จำลองได้จากโปรแกรม IRTDATA โดยการจำลองเหตุการณ์ที่มีการใช้แบบสอบไปสอบกับกลุ่มตัวอย่างซึ่งกำหนดเงื่อนไขการจำลอง ให้เป็นไปตามแบบแผนการเก็บรวบรวมข้อมูลที่กำหนดคือ แบบกลุ่มสมมูล และแบบข้อสอบร่วม แล้วปรับเทียบคะแนนในแนวระดับให้อยู่บนสเกลเดียวกันตามแนวทฤษฎีการตอบสนองรายข้อ (IRT)



2) การปรับเทียบคะแนนในแนวระดับ (horizontal test equating) หมายถึง การปรับเทียบคะแนนที่ได้จากกลุ่มผู้สอบสองกลุ่มที่มีความสามารถในระดับเดียวกัน และแบบสอบที่ใช้มีความยากเท่ากันหรือใกล้เคียงกัน สำหรับการวิจัยครั้งนี้หมายถึงการปรับเทียบคะแนนที่ได้จากการจำลองข้อมูลจากกลุ่มผู้สอบสองกลุ่มที่มีความสามารถในระดับเดียวกัน คือกำหนดค่าเฉลี่ยของพารามิเตอร์ความสามารถ (a) ให้มีค่าเท่ากับ 0 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1 และแบบสอบที่ใช้ทั้งสองฉบับมีความยากเท่ากันคือกำหนดค่าเฉลี่ยของพารามิเตอร์ความยาก (b) ให้เท่ากับ 0 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1

3) การปรับเทียบคะแนนตามแนวทฤษฎีการตอบสนองรายข้อ หมายถึงการนำหลักการของ ทฤษฎีการตอบสนองรายข้อ (Item Response Theory) อันได้แก่การประมาณพารามิเตอร์ ความสามารถ และพารามิเตอร์ข้อสอบ มาใช้ในการปรับเทียบคะแนน โดยยึดหลักการที่ว่าถ้า พารามิเตอร์ความสามารถอยู่ในสเกลเดียวกันแล้วสามารถนำคะแนนจากแบบสอบสองฉบับมาปรับเทียบ กันได้ สำหรับการวิจัยครั้งนี้หมายถึงการนำคะแนนจากแบบสอบฉบับหนึ่ง (X) ไปเทียบกับคะแนนของ แบบสอบอีกฉบับหนึ่ง (Y) ที่ระดับความสามารถ (a) เดียวกัน

4) แบบแผนในการปรับเทียบคะแนน (equating design) หมายถึงวิธีการต่างๆ ที่ใช้ในการเลือกกลุ่มตัวอย่างในการสอบ และเลือกหรือจัดแบบสอบเพื่อนำไปสอบกับกลุ่มตัวอย่าง ที่กำหนดไว้มี 3 วิธี คือ แบบแผนกลุ่มเดียว (single group design) แบบแผนกลุ่มสมมูล (equivalent group design) และแบบแผนข้อสอบร่วม (common item equating) สำหรับการวิจัยครั้งนี้ใช้แบบแผนกลุ่มสมมูลและแบบแผนข้อสอบร่วม

5) แบบแผนกลุ่มสมมูล หมายถึงแบบแผนที่เลือกกลุ่มตัวอย่างมา 2 กลุ่มโดยการสุ่ม และกลุ่มตัวอย่างแต่ละกลุ่มทำแบบสอบแต่ละฉบับเพียงฉบับใดฉบับหนึ่งสำหรับการวิจัยครั้งนี้หมายถึง การเลือกกลุ่มตัวอย่างจากการจำลองข้อมูลสองกลุ่มที่มีการกำหนดเลขสุ่มของกลุ่มผู้สอบและเลขสุ่ม ของแบบสอบ

6) แบบแผนข้อสอบร่วม เป็นแบบแผนที่กลุ่มผู้สอบสองกลุ่มทำแบบสอบฉบับเดียวกันซึ่ง เรียกว่าแบบสอบร่วมและนอกจากนี้กลุ่มตัวอย่างทั้งสองต้องทำแบบสอบ 2 ฉบับที่ต้องการนำมา ปรับเทียบอีกกลุ่มละฉบับที่แตกต่างกัน สำหรับการวิจัยครั้งนี้เป็นการใช้กลุ่มตัวอย่างจากการจำลอง ข้อมูลสองกลุ่มที่มีการกำหนดเลขสุ่มของกลุ่มตัวอย่างและเลขสุ่มแบบสอบ โดยกำหนดข้อสอบร่วม เป็นแบบแผนภายในที่มีจำนวนร้อยละ 20 ของความยาวแบบสอบ





7) ความยาวของแบบสอบ(test length) หมายถึงจำนวนข้อสอบที่จัดไว้ในแบบสอบแต่ละฉบับ สำหรับการวิจัยครั้งนี้กำหนดไว้ดังนี้

ทั้งแบบแผนกลุ่มสมบูรณ์และข้อสอบร่วมความยาวแบบสอบที่ใช้คือ 30 ข้อ 50 ข้อ 70 ข้อ 90 ข้อ 110 ข้อ 130 ข้อ และ 150 ข้อ สำหรับแบบแผนข้อสอบร่วมมีความยาวแบบสอบร่วมเป็น 5 ข้อ 8 ข้อ 12 ข้อ 15 ข้อ 18 ข้อ 22 ข้อ และ 25 ข้อ

8) ความคลาดเคลื่อนของการปรับเทียบคะแนน (equating error) หมายถึงความผิดพลาดอันเกิดขึ้นเนื่องจากการปรับเทียบคะแนน สำหรับการวิจัยครั้งนี้พิจารณาจากดัชนีความแตกต่างของการปรับเทียบคะแนน (c)

9) ประสิทธิภาพของการปรับเทียบคะแนน (efficiency) หมายถึงการปรับเทียบคะแนนที่เกิดความคลาดเคลื่อนอยู่ในระดับที่พอใจมากที่สุด ในการวิจัยครั้งนี้หมายถึงการปรับเทียบคะแนนที่มีดัชนีความแตกต่างอยู่ในระดับพอใจมากที่สุด คือ มีค่าน้อยกว่าหรือเท่ากับ  $(.05S_x)^2$  เมื่อ  $S_x^2$  คือค่าความแปรปรวนของคะแนน X

10) ขนาดของกลุ่มตัวอย่างขั้นต่ำ (Minimum sample size) หมายถึงจำนวนผู้สอบที่น้อยที่สุดที่ทำให้การปรับเทียบคะแนนในแต่ละวิธีให้ผลไม่แตกต่างจากกลุ่มตัวอย่างขนาดใหญ่ (3,000)

### ข้อจำกัดของการวิจัย

ข้อจำกัดของการวิจัยครั้งนี้ได้แก่

1. การหาความคลาดเคลื่อนในการปรับเทียบคะแนนไม่ได้หามาจากค่าความคลาดเคลื่อนมาตรฐานของการปรับเทียบคะแนน (Standard error of equating) แต่หาจากดัชนีความแตกต่าง
2. การปรับเทียบคะแนนครั้งนี้เป็นการปรับเทียบคะแนนจากแบบสอบฉบับ Y ไปยังฉบับ X เพียงทางเดียวเท่านั้น





### ข้อตกลงเบื้องต้นของการวิจัย

การวิจัยครั้งนี้มีข้อตกลงเบื้องต้นดังนี้

1. การวิจัยครั้งนี้เป็นการหาขนาดกลุ่มตัวอย่างขั้นต่ำ ที่คำนวณมาจากการกำหนดขนาดกลุ่มตัวอย่างแต่ละช่วงที่มีระยะห่างกันช่วงละ 50 เพื่อให้เกิดประสิทธิภาพในการใช้ทรัพยากรการวิจัย
2. การวิจัยครั้งนี้ข้อมูลที่ใช้เป็นไปตามข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองรายข้อ คือ ความเป็นเอกมิติ (unidimensionality) ความเป็นอิสระ (local independence) มีความสอดคล้องกับโค้งคุณลักษณะข้อสอบ (item characteristic curve) และความเร็วในการสอบไม่มีผลต่อความสามารถของผู้สอบ (speedness)

### ประโยชน์ที่คาดว่าจะได้รับจากการวิจัย

การวิจัยคาดหวังว่าจะได้รับประโยชน์ดังนี้

1. ทำให้ทราบขนาดกลุ่มตัวอย่างขั้นต่ำในการปรับเทียบคะแนนตามแนวระดับ ในแนวทฤษฎีการตอบสนองรายข้อ โมเดล 1 และ 3 พารามิเตอร์ ที่มีแบบแผนการปรับเทียบและความยาวแบบสอบแตกต่างกัน
2. ผลจากการวิจัยสามารถนำไปใช้ประกอบการตัดสินใจเกี่ยวกับขนาดกลุ่มตัวอย่างและความยาวของแบบสอบในการปรับเทียบคะแนนในแนวระดับตามทฤษฎีการตอบสนองรายข้อได้
3. เป็นพื้นฐานในการวิจัยเกี่ยวกับการวัดและประเมินผลโดยเฉพาะอย่างยิ่งในการปรับเทียบคะแนนตามแบบแผนกลุ่มสมมูลและแบบแผนข้อสอบร่วม ในแนวทฤษฎีการตอบสนองรายข้อโมเดล 1 และ 3 พารามิเตอร์ ต่อไป

ศูนย์วิจัยทรัพยากร  
จุฬาลงกรณ์มหาวิทยาลัย