

การเปรียบเทียบผลการใช้รูปแบบการเทียบ มาตราที่ต่างกัน เมื่อแบบสอบร่วมมีความ ยาวต่างกัน

● ภาวิณี ศรีสุขวัฒน์นันท์

บทนำ

โปรแกรมการสอบที่จัดขึ้นเพื่อการคัดเลือกบุคคลเข้าศึกษาต่อ เข้าทำงาน ตัดสินหรือให้การรับรองความรู้ความสามารถของผู้สอบจำเป็นต้องใช้แบบสอบที่มีคุณภาพ ที่สามารถวัดในสิ่งที่ต้องการอย่างมีมาตรฐาน และให้ความยุติธรรมในเชิงเปรียบเทียบกันได้ หากโปรแกรมนั้น ๆ ยังคงดำเนินอยู่ แบบสอบที่ใช้ในแต่ละครั้งจำเป็นต้องมีคุณสมบัติอย่างเดียวกัน ความเป็นอย่างเดียวกันของแบบสอบ หมายถึง มีจุดประสงค์ในการวัดอย่างเดียวกัน เนื้อหาในแบบสอบ ความยากง่าย และค่าอำนาจจำแนกของข้อสอบ ตลอดจนค่าสถิติอื่น ๆ ต้องเหมือนกันด้วย โดยทฤษฎีแบบสอบสองฉบับใด ๆ ที่มีลักษณะเหมือนกันในทำนองนี้ คือแบบสอบคู่ขนาน (Parallel forms) คะแนนจากแบบสอบคู่ขนานสามารถนำมาเปรียบเทียบกัน ในเชิงปริมาณได้โดยตรง ทั้งนี้ ถือว่าเป็นผลจากการวัดปริมาณจากเครื่องมือชนิดเดียวกัน มีระบบมาตราและหน่วยวัดเหมือนกัน ในทางปฏิบัติ การพัฒนาแบบสอบคู่ขนานเพื่อใช้ในงานสอบต่าง ๆ เป็นสิ่งที่ต้องลงทุนทั้งเวลาและเงินทุนมาก และบางครั้งแบบสอบคู่ขนานที่พัฒนาแล้วอาจไม่เหมาะแก่การใช้ในกรณีที่มีการปรับหรือเปลี่ยนแปลงจุดประสงค์ในการวัดบ้างอันเนื่องจากการปรับปรุงหลักสูตร ดังนั้นในทางปฏิบัติจึงได้มีการพัฒนาเทคนิคทางสถิติเพื่อจัดการกับแบบสอบสองฉบับที่ทำหน้าที่วัดในสิ่งเดียวกัน หรือเรียกว่าเป็นแบบสอบคู่ขนานเชิงหน้าที่ (functional parallel tests) ในส่วนของระบบมาตราคะแนนเพื่อให้สามารถทำการเปรียบเทียบผลการสอบได้อย่างยุติธรรม เทคนิคดังกล่าวเรียกว่า เทคนิคการเทียบมาตราคะแนนของแบบสอบ (Test equating)

พัฒนาการของเทคนิคการเทียบมาตรฐานคะแนนของแบบสอบ เริ่มจากการเสนอแนวคิดเชิงทฤษฎี และนำไปสู่การปฏิบัติในสถานการณ์ทดสอบความถนัดเชิงวิชาการ (SAT) ของสหรัฐอเมริกา ตั้งแต่ปี ค.ศ.1942 ต่อจากนั้น ETS (Educational Testing Services) ได้พัฒนาเทคนิคการเทียบมาตรฐานตลอดมา ในทศวรรษที่ผ่านมา นักวัดผลได้ศึกษาวิจัยกันอย่างจริงจังถึงการใช้และการปรับเทคนิคต่าง ๆ กับสถานการณ์การสอบทั้งด้านวัดผลสัมฤทธิ์ทางการเรียนและสมรรถวิสัยเชิงวิชาการต่าง ๆ ทั้งนี้เป็นผลผลักดันจากความก้าวหน้าของทฤษฎีการวัดความสามารถแฝง (Latent trait test theory) และกฎหมายคุ้มครองผู้รับการทดสอบที่ประกาศใช้เมื่อวันที่ 1 มกราคม 1980 ที่รัฐนิวยอร์ก (Marco 1981) ประเด็นคำถามที่นักวัดผลสนใจ คือ รูปแบบการเทียบมาตรฐานใดที่ให้ผลการแปลงคะแนนที่เกิดความคลาดเคลื่อนน้อยกว่ากัน หรือที่ทำให้การลงทุนด้านดำเนินการน้อยกว่า และ/หรือเสนอผลได้เร็วกว่า

เทคนิคการเทียบมาตรฐาน เป็นองค์ประกอบของมิติการออกแบบรวบรวมข้อมูลกับมิติโมเดลหรือรูปแบบซึ่งเป็นกฎระเบียบวิธีการแปลงคะแนน สำหรับด้านการออกแบบรวบรวมข้อมูลที่ได้รับนิยามคือวิธีที่ใช้แบบสอบร่วม (Anchor Test) ซึ่งเป็นวิธีที่อาศัยข้อสอบจำนวนหนึ่งผนวกเข้าในแบบสอบทั้ง 2 ชุด

(X และ Y) ที่ต้องการเทียบกัน กลุ่มผู้สอบที่ตอบแบบสอบชุดใดชุดหนึ่งจะได้รับแบบสอบร่วมเหมือนกัน ผลการตอบในส่วนของแบบสอบร่วมเป็นส่วนที่นำมาใช้ปรับความแตกต่างในเชิงสถิติด้านความยากง่ายของแบบสอบทั้ง 2 ชุด และความแตกต่างระหว่าง 2 กลุ่มที่ต่างก็ตอบเพียง X หรือ Y เท่านั้น ส่วนมิติโมเดล พบว่าปัจจุบันนักวัดผลให้ความสนใจอยู่ 3 โมเดล คือ Equi-percentile, Linear และ Item response theory ซึ่งเป็นส่วนหนึ่งของทฤษฎีวัดความสามารถแฝง ทั้งสามโมเดลมีความเหมาะสมในการนำไปใช้ต่าง ๆ กัน ยังไม่มีข้อค้นพบที่จะสรุปว่าโมเดลใดดีที่สุด ขณะนี้ยังคงมีงานวิจัยที่นักวัดผลพยายามค้นหาคำตอบนี้อยู่ สำหรับงานวิจัยที่กำลังเสนออยู่นี้ก็เป็นความพยายามอีกครั้งหนึ่งที่ต้องการเพิ่มข้อความรู้ที่เกี่ยวข้องเพื่อพัฒนาเทคนิคการเทียบมาตรฐานต่อไป

วัตถุประสงค์ของการวิจัย

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อศึกษาเปรียบเทียบผลของการใช้รูปแบบการเทียบมาตรฐานที่ต่างกัน 3 รูปแบบ คือ 1) รูปแบบอีควิเปอร์เซนไทล์ 2) รูปแบบเชิงเส้นตรง และ 3) รูปแบบอิงทฤษฎีการตอบข้อสอบแบบสามพารามิเตอร์ และที่ใช้การออกแบบเทียบมาตรฐานโดยมีแบบสอบร่วมชนิดภายในต่างกันสามขนาด ซึ่งคิดเป็นอัตราส่วนร้อยละ

ของแบบสอบที่ทำการเทียบมาตรา ดังนี้
 1) ขนาดร้อยละ 20 2) ขนาดร้อยละ 40 และ
 3) ขนาดร้อยละ 60 ผลของการเทียบมาตรา
 ที่นำมาใช้ประเมินมี 2 ลักษณะ คือ 1) ค่า
 ความคลาดเคลื่อนมาตรฐานของการเทียบ
 มาตรา (SEE) ซึ่งเป็นการวิเคราะห์ในเชิง
 ทฤษฎีของรูปแบบ และ 2) ดัชนีการเปรียบเทียบ
 ความแตกต่าง (Index C) ที่ได้จากการ
 วิเคราะห์กลุ่มตัวอย่างสอบทานผล ซึ่งเป็น
 การวิเคราะห์ความเพียงพอของวิธีการเทียบ
 มาตราโดยตรงกับผลการเทียบสุดท้าย

สมมติฐานของการวิจัย

1. แบบสอบร่วมที่มีความยาวมากกว่า
 มีประสิทธิภาพสัมพัทธ์ของการเทียบมาตรา
 ที่สูงกว่าในทุกรูปแบบของการเทียบมาตรา
 ของทั้งสองกรณีของแบบสอบที่ใช้

2. วิธีการเทียบมาตราที่ใช้แบบสอบร่วม
 ที่มีความยาวมากกว่า เป็นวิธีที่มีความเพียงพอ
 กว่าวิธีที่ใช้แบบสอบร่วมที่มีความยาวน้อย
 กว่า

3. วิธีการเทียบมาตราที่ใช้รูปแบบอีควิ-
 เปรอร์เซนไทล์ และรูปแบบอิงทฤษฎีการตอบ
 ข้อสอบเป็นวิธีที่มีความเพียงพอมากกว่าวิธี
 ที่ใช้รูปแบบเชิงเส้นตรง ในกรณีแบบสอบ
 คัดเลือก

4. วิธีการเทียบมาตราที่ใช้รูปแบบเชิง
 เส้นตรง และรูปแบบอิงทฤษฎีการตอบข้อ
 สอบเป็นวิธีที่มีความเพียงพอมากกว่าวิธีที่
 ใช้รูปแบบอีควิเปอร์เซนไทล์ ในกรณีแบบ
 สอบวัดผลสัมฤทธิ์

วิธีวิจัย

1. ประชากรและกลุ่มตัวอย่าง

การวิจัยครั้งนี้ใช้สถานการณ์ทดสอบที่
 ให้ข้อมูลคะแนน 2 กรณี คือ 1) กรณีแบบ
 สอบคัดเลือก ประชากร คือ ผลการสอบ
 รายบุคคลผู้ซึ่งมีคุณสมบัติขั้นต่ำตามกำหนด
 ของระเบียบการสอบคัดเลือกที่กำหนดไว้ มี
 จำนวน 15,875 คน สุ่มตัวอย่างสำหรับการ
 เทียบมาตรา 2 กลุ่ม กลุ่มละ 1,500 คน และ
 สำหรับการวิเคราะห์กลุ่มสอบทานผลอีก 1
 กลุ่ม จำนวน 1,500 คน 2) กรณีแบบสอบ
 วัดผลสัมฤทธิ์ ประชากร คือ ผลการสอบ
 รายบุคคลในวิชาที่กำหนดในระดับอุดมศึกษา
 จำนวน 12,383 คน สุ่มตัวอย่างทำนองเดียวกับ
 กรณีแรก

2. แบบสอบ

ในแต่ละกรณีของแบบสอบที่นำมาศึกษา
 ประกอบด้วยข้อสอบ 100 ข้อ ใช้วัดความรู้
 หนึ่งวิชา ผู้วิจัยได้นำมาวิเคราะห์รายข้อ แล้ว
 นำมาเลือก และจัดเป็นแบบสอบที่วัดในเรื่อง
 เดียวกันมีความยากง่ายใกล้เคียงกัน 2 ชุด

ชุดละ 35 ข้อ เพื่อใช้เป็นแบบสอบสำหรับการเทียบมาตรฐาน และทำการคัดเลือกจัดชุดแบบสอบร่วมขนาด 7 ข้อ 14 ข้อ และ 21 ข้อ จากข้อสอบที่เหลือ โดยให้แบบสอบร่วมมีลักษณะเป็นฉบับย่อส่วนของแบบสอบเทียบมาตรฐานที่จัดไว้แล้ว จากการจัดเตรียมแบบสอบของกรณีสอบคัดเลือก ได้แบบสอบเทียบมาตรฐานซึ่งได้ผนวกเอาแบบสอบร่วมแล้ว 3 คู่ คือ XV60 กับ YV60 XV40 กับ YV40 และ XV20 กับ YV20 ความเที่ยงของแบบสอบแต่ละคู่มีค่าต่างเล็กน้อย คือ ต่างกัน .032 .025 และ .023 ตามลำดับ กรณีแบบสอบวัดผลสัมฤทธิ์ได้แบบสอบเทียบมาตรฐาน 3 คู่ เช่น เดียวกัน มีความเที่ยงต่างกันตามลำดับ ดังนี้ .047 .039 และ .027 ตามลำดับ จากผลการจัดเตรียมแบบสอบทั้งสองกรณี พบว่า แบบสอบกรณีหลังมีความเที่ยงต่างกันในแต่ละคู่ของแบบสอบเทียบมาตรฐานมากกว่ากรณีแรกเล็กน้อย

3. การรวบรวมข้อมูล

ข้อมูลเป็นประเภททฤษฎีภูมิ จากการออกแบบการเทียบมาตรฐานให้กลุ่มตัวอย่างหนึ่ง (α) รับการทดสอบด้วยแบบสอบชุด X และอีกกลุ่มหนึ่ง (β) ทำชุด Y และทั้งสองกลุ่มได้รับการทดสอบส่วนที่เป็นแบบสอบร่วมเหมือนกัน การประมวลข้อมูลให้เป็นไปตามสถานการณ์ดังที่กล่าวมาแล้วนี้ได้อาศัยการจัดกระทำข้อมูลผลการตอบรายข้อที่บันทึกไว้

แล้ว จากการสอบตามสภาพความเป็นจริงของแต่ละกรณี ข้อมูลที่สังเคราะห์ขึ้นเพื่อดำเนินการเทียบมาตรฐานจึงเป็นคะแนนของแบบสอบชุด XV60 XV40 และ XV20 จากกลุ่มตัวอย่าง และ YV60 YV40 และ YV20 จากกลุ่มตัวอย่าง

ผลการรวบรวมข้อมูล ในกรณีของแบบสอบคัดเลือก กลุ่มตัวอย่าง α และ β มีการแจกแจงคล้ายคลึงกันโดยประมาณ คะแนนเฉลี่ยของแบบสอบชุด Y สูงกว่าชุด X ทุกขนาดของแบบสอบร่วมโดยประมาณอยู่ 0.5 คะแนน แต่ส่วนเบี่ยงเบนมีค่าต่ำกว่าประมาณ 0.6 ในกรณีแบบสอบวัดผลสัมฤทธิ์มีลักษณะกลับกัน คือ คะแนนเฉลี่ยของชุด Y ได้ต่ำกว่าคะแนนเฉลี่ยของชุด X ประมาณ 1.1 คะแนน และส่วนเบี่ยงเบนสูงกว่าประมาณ 0.5

4. การเทียบมาตรฐาน

วิธีการเทียบมาตรฐานในแต่ละกรณีของแบบสอบ ประกอบด้วยรูปแบบที่ใช้กำหนดกฎการแปลงคะแนนสมมูล ซึ่งมีอยู่ 3 รูปแบบ และการออกแบบที่ใช้แบบสอบร่วมชนิดที่เป็นแบบสอบร่วมภายใน 3 ขนาด จึงรวมเป็นวิธีการเทียบมาตรฐานทั้งหมด 9 วิธี หลังจากการประมวลผลคะแนนของกลุ่มตัวอย่างตามแบบสอบชุดต่าง ๆ ที่ต้องการเทียบแล้ว ได้ดำเนินการเทียบมาตรฐานทั้ง 9 วิธี แต่ละวิธีได้สรุปเป็นตารางคะแนนสมมูล แปลง

จากมาตรฐาน Y ซึ่งสมมติเป็นแบบสอบชุดที่สร้างใหม่ไปสู่คะแนนแปลง X^* ที่มีมาตรฐานตามชุด X หรือที่สมมติว่าเป็นแบบสอบเก่า

ตารางคะแนนสมมูลของการเทียบมาตรฐานคะแนนของแบบสอบคู่ใด ๆ ใช้เป็นตารางอ้างอิงของการเทียบมาตรฐานเพื่อจัดทำรายงานผลการสอบตามมาตรฐานที่ต้องการ ลักษณะของคะแนนสมมูลที่ผ่านการเทียบมาตรฐานทั้ง 9 วิธี มีลักษณะสอดคล้องกันเป็นส่วนใหญ่ ในช่วงคะแนนตอนกลาง ๆ และความสอดคล้องดังกล่าวจะน้อยลงในตอนปลาย แบบสอบคู่ที่มีแบบสอบร่วมมากกว่า มีความสอดคล้องของคะแนนแปลงมากกว่า

การวิเคราะห์ข้อมูล

1. วิเคราะห์ความคลาดเคลื่อนมาตรฐานของการเทียบมาตรฐาน (SEE) ทั้ง 3 รูปแบบการเปรียบเทียบปริมาณความคลาดเคลื่อนจากวิธีการเทียบมาตรฐานที่ต่างกันในแต่ละกรณี ได้ใช้ค่า SEE ที่ประมาณจากตำแหน่งคะแนนมาตรฐานเดียวกัน โดยพิจารณาในช่วงคะแนนมาตรฐาน ± 3.0 ของคะแนน Y ตัดสินความเหมาะสมของวิธีเทียบมาตรฐานใช้พิจารณาจากดัชนีประสิทธิภาพสัมพัทธ์ (RE) ซึ่งเป็นอัตราส่วนร้อยละระหว่างความคลาดเคลื่อนของวิธีเทียบมาตรฐานที่ใช้แบบสอบร่วม V40 และ V20 กับเกณฑ์ คือ V60 ภายในแต่ละรูปแบบของการเทียบมาตรฐาน

2. วิเคราะห์กลุ่มตัวอย่างสอบทานผล เพื่อหาดัชนีความแตกต่าง (C) ระหว่างคะแนนแปลงที่อ่านจากตารางเทียบมาตรฐานวิธีต่าง ๆ กับคะแนนเกณฑ์ ซึ่งเป็นคะแนนที่กลุ่มตัวอย่างได้โดยตรง จากการตอบแบบสอบชุดนั้น ๆ เป็นความคลาดเคลื่อนรวมที่เกิดจากการใช้ตารางคะแนนสมมูลอันเป็นผลจากการเทียบมาตรฐาน ดัชนีเปรียบเทียบความแตกต่าง (C) เป็นค่าที่ได้ถ่วงน้ำหนักให้เป็นมาตรฐานของค่าเฉลี่ยของความแตกต่างกำลังสอง ซึ่งสามารถนำมาเปรียบเทียบในเชิงปริมาณได้โดยตรง ไม่ว่าจะได้จากการเทียบมาตรฐานที่ต่างรูปแบบ และต่างสถานการณ์ นำค่าดัชนี C มาทำการวิเคราะห์คุณภาพของการเทียบมาตรฐานโดยใช้เกณฑ์ของปีเตอร์สันและคณะก่อน แล้วจึงทดสอบความมีนัยสำคัญของการจัดอันดับของค่า C ด้วยวิธีสถิติสหพาราเมตริก Friedman model ซึ่งเป็นการวิเคราะห์ความแปรปรวนสองทางที่ใช้กับข้อมูลจัดอันดับที่มีการวัดซ้ำ ทำการวิเคราะห์แบบรายคู่ตามกำหนด

การวิเคราะห์ข้อมูลส่วนใหญ่อาศัยเครื่องคอมพิวเตอร์ของสถาบันบริการคอมพิวเตอร์ของจุฬาลงกรณ์มหาวิทยาลัย โดยใช้โปรแกรมสำเร็จรูป SPSSX LOGIST 5 และโปรแกรมที่ผู้วิจัยสร้างขึ้น

สรุปผลการวิจัย

1. ผลการเทียบมาตราโดยส่วนรวม

จากการเทียบมาตราแบบสอบแต่ละกรณี ด้วยรูปแบบการเทียบมาตรา 3 รูปแบบ และใช้ความยาวแบบสอบร่วม 3 ขนาด พบว่า ทั้ง 9 วิธี ที่ดำเนินไปในแต่ละกรณี ให้ผลการเทียบที่อยู่ในระดับที่ยอมรับได้ทั้งสิ้นตามเกณฑ์ของปีเตอร์สันและคณะ กล่าวคือ ในกรณีแบบสอบคัดเลือกวิธีที่ใช้แบบสอบร่วม 21 ข้อ (ร้อยละ 60) 14 ข้อ (ร้อยละ 40) กับรูปแบบอีควิเปอร์เซนไทล์ และ IRT ให้ผลการเทียบมาตราในระดับน่าพอใจอย่างมาก และวิธีที่ใช้แบบสอบร่วม 21 ข้อ (ร้อยละ 60) กับรูปแบบเชิงเส้นตรงให้ผลที่น่าพอใจอย่างมากเช่นกัน ส่วนวิธีอื่นที่เหลือให้ผลในระดับที่ต่ำลงมาเล็กน้อย คือ อยู่ในระดับน่าพอใจสำหรับกรณีแบบสอบวัดผลสัมฤทธิ์ทั้ง 9 วิธี ให้ผลการเทียบมาตราในระดับน่าพอใจทั้งสิ้น

2. มิติที่เกี่ยวกับความยาวของแบบสอบร่วม

ในการศึกษาผลการเทียบมาตราที่ใช้แบบสอบขนาดต่างกัน พบว่า ความยาวของแบบสอบร่วมเป็นปัจจัยหนึ่งที่มีผลกระทบต่อความแม่นยำ และความเที่ยงพอโดยตรง เมื่อแบบสอบร่วมมีความยาวเพิ่มขึ้น ความคลาดเคลื่อนมาตรฐานของการเทียบมาตราได้ลดลง และในทำนองเดียวกัน ความคลาดเคลื่อน

รวมหรือค่าดัชนีเปรียบเทียบความแตกต่าง C ก็ลดลงด้วย แต่มีข้อน่าสังเกตที่ว่า ในกรณีของแบบสอบคัดเลือก วิธีที่ใช้แบบสอบร่วม 21 ข้อ และ 14 ข้อ ค่าความคลาดเคลื่อนมาตรฐานของการเทียบมาตรา (SEE) และความคลาดเคลื่อนรวม (ค่าดัชนี C) มีค่าต่างกันเพียงเล็กน้อย ซึ่งทำให้เห็นว่าการเพิ่มความยาวของแบบสอบร่วมให้ผลต่อการเพิ่มประสิทธิภาพของการเทียบมาตราที่ไม่สู้ชัดเจนนัก การเปรียบเทียบผลการเทียบมาตราภายในรูปแบบโดยใช้เกณฑ์การประเมินภายในและภายนอก มีดังนี้

2.1 ผลการศึกษาเปรียบเทียบความคลาดเคลื่อนมาตรฐานของการเทียบมาตราที่ใช้แบบสอบร่วมขนาดต่าง ๆ กันในแต่ละรูปแบบของการเทียบมาตรา พบว่า การเทียบด้วยรูปแบบอีควิเปอร์เซนไทล์ และเชิงเส้นตรงให้ผลสอดคล้องกับสมมติฐานของการวิจัยข้อที่ 1 ทั้งสองกรณีของแบบสอบที่นำมาใช้ในการศึกษาคั้งนี้ กล่าวคือ เมื่อใช้แบบสอบร่วมที่ยาวที่สุด (21 ข้อ หรือร้อยละ 60 ของแบบสอบเทียบมาตรา) ค่าความคลาดเคลื่อนมาตรฐานของการเทียบมาตรามีค่าน้อยที่สุดเมื่อเทียบกับวิธีแบบสอบร่วมที่ขนาดสั้นลงอีกสองขนาด และเมื่อใช้แบบสอบร่วมขนาดกลาง (14 ข้อ หรือร้อยละ 40) ค่าความคลาดเคลื่อนที่คำนวณได้จัดอยู่ในอันดับกลางและสุดท้าย คือวิธีที่ใช้แบบสอบร่วมขนาด

สั้นที่สุด (7 ข้อ หรือร้อยละ 20) ค่าความคลาดเคลื่อนมีค่ามากที่สุด ส่วนวิธีการเทียบมาตราด้วยรูปแบบ IRT ให้ผลในทำนองเดียวกันเว้นวิธีที่ใช้แบบสอบร่วมขนาด 14 ข้อ ให้ผลไม่แตกต่างจากวิธีที่ใช้ 21 ข้อ จึงสรุปได้ว่าการเทียบมาตราที่ใช้กรณีแบบสอบคัดเลือก และวัดผลสัมฤทธิ์ตามที่นำมาศึกษาครั้งนี้ ด้วยรูปแบบการเทียบมาตราสามรูปแบบ พบว่าการใช้แบบสอบร่วมที่ยาวกว่าให้ประสิทธิภาพของการเทียบมาตราที่สูงกว่า

2.2 การศึกษาความเพียงพอของวิธีการเทียบมาตราอันเกิดจากการใช้แบบสอบร่วมที่มีขนาดต่างกันในแต่ละรูปแบบการเทียบมาตราด้วยเทคนิคการวิเคราะห์กลุ่มสอบทานผล ซึ่งเป็นเกณฑ์ประเมินอีกเกณฑ์หนึ่งที่น่ามาประเมินความเพียงพอของวิธีการเทียบมาตรา ผลการศึกษา พบว่าในกรณีแบบสอบคัดเลือก วิธีการเทียบมาตราที่ใช้แบบสอบร่วมขนาด 21 ข้อ (ร้อยละ 60) และ 14 ข้อ (ร้อยละ 40) มีความเพียงพอในระดับที่น่าพอใจมากทุกรูปแบบที่ใช้ ยกเว้นรูปแบบเชิงเส้นตรงที่ใช้แบบสอบร่วม 14 ข้อ ให้ผลในระดับน่าพอใจ ส่วนวิธีที่ใช้แบบสอบร่วม 7 ข้อ (ร้อยละ 20) กับรูปแบบทั้งสามให้ผล การเทียบในระดับน่าพอใจ สำหรับกรณีแบบสอบวัดผลสัมฤทธิ์ พบว่าทั้ง 9 วิธี ของการเทียบมาตรามีความเพียงพอในระดับน่า

พอใจ จากการจัดอันดับตามค่าดัชนี C พบว่า ในมิติของความยาวแบบสอบร่วมมีแบบแผนเดียวกันทั้งสองกรณีของแบบสอบ (ตารางที่ 13) คือ วิธีที่ใช้แบบสอบร่วมร้อยละ 60 เป็นวิธีที่มีความเพียงพอมากเป็นอันดับ 1 แบบสอบร่วมขนาดร้อยละ 40 เป็นอันดับ 2 และแบบสอบร่วมขนาดร้อยละ 20 เป็นอันดับ 3 ข้อค้นพบดังกล่าวสอดคล้องกับสมมติฐานที่กำหนดไว้ในข้อ 2

3. มิติที่เกี่ยวกับรูปแบบการเทียบมาตรา

รูปแบบการเทียบมาตรา 3 รูปแบบ ที่ใช้ในการเทียบมาตราแบบสอบสองกรณีให้ผลต่างกัน ดังนี้

3.1 กรณีแบบสอบคัดเลือก ผลการวิเคราะห์กลุ่มสอบทานผลแสดงให้เห็นแบบแผนของค่าดัชนี C ซึ่งเป็นความคลาดเคลื่อนรวม สามารถจัดอันดับของความเพียงพออันเกิดจากรูปแบบการเทียบมาตราในแต่ละความยาวแบบสอบร่วมที่ใช้เป็นทำนองเดียวกันทั้งหมด แบบแผนของอันดับความเพียงพอมีดังนี้ วิธีที่ใช้รูปแบบอิกวิเปอร์เซนไคล์เป็นอันดับที่หนึ่ง IRT เป็นอันดับที่สอง และเชิงเส้นตรงเป็นอันดับที่สาม ข้อค้นพบในส่วนนี้มีความสอดคล้องกับสมมติฐานทางการวิจัยข้อ 3 ในส่วนที่กล่าวว่า รูปแบบอิกวิเปอร์เซนไคล์เป็นรูปแบบที่มีความเพียงพอมากกว่ารูปแบบเชิงเส้นตรง ส่วนการเปรียบเทียบ

รูปแบบคู่อื่นไม่สามารถสรุปตามความแตกต่างอย่างมีนัยสำคัญ

3.2 กรณีแบบสอบวัดผลสัมฤทธิ์ ผลการจัดอันดับในทำนองเดียวกับ 3.1 คือ จากความเพียงพอมากไปหาน้อยได้ดังนี้ รูปแบบเชิงเส้นตรงเป็นอันดับที่หนึ่ง อีควิเปอร์เซนไคล์เป็นอันดับที่สอง และ IRT เป็นอันดับที่สาม การทดสอบความแตกต่างของค่าอันดับด้วยสถิติที่ไม่ใช่พาราเมตริกในทำนองเดียวกับกรณีแบบสอบคัดเลือก พบว่า อันดับที่หนึ่งและที่สาม มีความแตกต่างกันอย่างมีนัยสำคัญที่ .05 ซึ่งหมายถึง วิธีที่ใช้รูปแบบเชิงเส้นตรงเป็นวิธีที่มีความเพียงพอมากกว่าวิธีที่ใช้รูปแบบ IRT ส่วนการเปรียบเทียบคู่อื่นไม่พบความแตกต่างอย่างมีนัยสำคัญ ข้อค้นพบนี้เป็นการปฏิเสธสมมติฐานของการวิจัยที่กำหนดไว้ในข้อ 4

การอภิปรายผล

1. จากการประเมินผลความเพียงพอของวิธีการเทียบมาตรของทั้งสองกรณีของแบบสอบที่นำมาศึกษาด้วยเกณฑ์ของปีเตอร์สันและคณะ พบว่า ทุกวิธีที่ทำการศึกษาคือเป็นวิธีที่ใช้ผลการแปลงคะแนนสมมูลที่มีความเพียงพออยู่ในระดับน่าพอใจ และน่าพอใจอย่างมาก เมื่อวิเคราะห์ในเชิงเหตุผล พบว่าการเทียบมาตรที่ดำเนินไปนั้น ใช้กลุ่มตัวอย่างสมมูลสองกลุ่มที่ได้จากการสุ่มจาก

ประชากรเดียวกัน การออกแบบการวิจัยลักษณะนี้เป็นการควบคุมให้เกิดความคล้อยคลึงของการแจกแจงของคะแนนที่ได้จากการสอบ ส่วนในด้านแบบสอบที่ใช้เทียบมาตรานั้นได้เตรียมจากแบบสอบฉบับเดียวกับที่มีความเที่ยงค่อนข้างสูง โดยทำให้เกิดเป็นแบบสอบเทียบมาตรที่คล้ายคลึงกันสองชุด ยิ่งกว่านี้การออกแบบด้วยการใช้แบบสอบร่วม (CIE) เป็นวิธีที่ใช้ปรับความแตกต่างของข้อมูลคะแนนในส่วนที่อาจเกิดขึ้นจากแบบสอบเทียบมาตรมีความยากง่ายต่างกัน และความสามารถของผู้สอบสองกลุ่มแตกต่างกัน ดังนั้น เมื่อใช้รูปแบบอีควิเปอร์เซนไคล์จึงให้ผลสอดคล้องกับนิยามที่ว่า การระบุคะแนนสมมูลของแบบสอบต่างชุดเมื่อกลุ่มผู้สอบมีการกระจายของความสามารถอยู่ในลักษณะเดียวกันแล้วนั้น ให้ระบุจากตำแหน่งเปอร์เซนไคล์เดียวกัน (Marco 1981) ส่วนผลการใช้รูปแบบเชิงเส้นตรง โดยทฤษฎีแล้วเป็นวิธีการประมาณคะแนนสมมูลตามนิยามของอีควิเปอร์เซนไคล์ที่ดีที่สุด เมื่อการแจกแจงของคะแนนของแบบสอบสองชุดมีลักษณะคล้ายคลึงกัน (Angoff 1984: 93) สำหรับรูปแบบ IRT ซึ่งมีข้อกำหนดที่แตกต่างไปบ้างคือ แบบสอบที่ใช้ต้องมีคุณสมบัติของการวัดความสามารถในมิติเดียวกัน และจำนวนข้อมูลที่ใช้ในการประมาณค่าต้องมีจำนวนไม่น้อยเกินไปนั้น การออกแบบการวิจัยครั้งนี้ไม่ได้มีข้อขัดแย้งกับข้อกำหนด จึงทำได้

ผลการเทียบที่น่าพอใจเช่นกันกับรูปแบบการเทียบมาตรฐานแบบดั้งเดิม

2. การศึกษาเปรียบเทียบผลการใช้แบบสอบร่วมที่ต่างกันด้วยการประเมินความแม่นยำของการเทียบมาตรฐานจากการวิเคราะห์ความคลาดเคลื่อนมาตรฐานของการเทียบมาตรฐาน (SEE) และตัดสินด้วยค่าประสิทธิภาพสัมพัทธ์ (RE) ซึ่งพบว่า ส่วนใหญ่เป็นไปตามสมมติฐานของการวิจัยที่กล่าวว่า แบบสอบร่วมที่ยาวกว่าให้ประสิทธิภาพของการเทียบมาตรฐานที่สูงกว่า ทั้งนี้มีเหตุผลสำคัญในการสนับสนุนข้อค้นพบ คือ โดยบทบาทของแบบสอบร่วมซึ่งทำหน้าที่ให้สารสนเทศ เพื่อใช้ในการปรับความแตกต่างระหว่างกลุ่มผู้สอบสองกลุ่ม และแบบสอบที่ต่างชุดกันในแต่ละรูปแบบของการเทียบมาตรฐานในรูปแบบอีควิเปอร์เซนไตล์ สัดส่วนของความถี่ของคะแนนในแบบสอบร่วมที่ได้จากกลุ่มตัวอย่างรวม

ได้นำไปใช้เพื่อประมาณความถี่ของแต่ละชั้นของคะแนนของแบบสอบเทียบมาตรฐานชุด X และชุด Y สำหรับในรูปแบบเชิงเส้นตรงผลการสอบของกลุ่ม α และ β ในส่วนของแบบสอบร่วมได้นำไปใช้ในการประมาณคะแนนของแบบสอบเทียบมาตรฐานชุด X และชุด Y ของประชากรโดยวิธีความเป็นไปได้สูงสุด (Maximum-likelihood) ส่วนรูปแบบ IRT ได้อาศัยแบบสอบร่วมเพื่อการประมาณ

ค่าพารามิเตอร์ของข้อสอบ ทำให้ค่าประมาณเป็นค่าที่มีหน่วยบนมาตรฐานเดียวกัน จากผลการวิจัยได้แสดงให้เห็นว่า แบบสอบร่วมที่ยาวกว่าทำให้ผลการเทียบมาตรฐานมีความแม่นยำมากกว่า ในทำนองเดียวกันเมื่อประเมินด้วยการใช้เกณฑ์ความคลาดเคลื่อนรวม (ดัชนี C) จากเทคนิคการวิเคราะห์กลุ่มสอบทาน ผลก็พบข้อสรุปเช่นเดียวกัน ซึ่งเป็นการยืนยันสมมติฐานการวิจัยข้อที่ 2 ด้วย

แต่อย่างไรก็ตาม จากการวิเคราะห์ความคลาดเคลื่อนมาตรฐานของการเทียบมาตรฐาน (SEE) และหาค่าประสิทธิภาพสัมพัทธ์ของแบบสอบ YV40-XV40 ของกรณีแบบสอบคัดเลือกพบว่า ผลการเทียบมาตรฐานของแบบสอบคู่นี้ให้ประสิทธิภาพสูงใกล้เคียงกับการเทียบมาตรฐานแบบสอบ YV60-XV60 ลักษณะเช่นนี้ไม่ปรากฏกับกรณีของแบบสอบวัดผลสัมฤทธิ์ จากการวิเคราะห์ในเชิงเหตุผล พบว่าเป็นผลกระทบจากคุณลักษณะของข้อสอบที่ประกอบขึ้นเป็นแบบสอบร่วม เมื่อแบบสอบร่วมขนาด 14 ข้อ มีผลรวมความแปรปรวนรายข้อสูงกว่าผลรวมความแปรปรวนรายข้อของแบบสอบร่วมขนาด 21 ข้อ เมื่อเทียบตามสัดส่วนของความแปรปรวนของแต่ละแบบสอบร่วม ความแม่นยำของการประมาณโดยอาศัยค่าสถิติจากแบบสอบร่วมจึงมีความแม่นยำต่างกัน ซึ่งส่งผลถึงผลสุดท้ายของกระบวนการเทียบมาตรฐานด้วย จากข้อ

ค้นพบนี้เป็นการสนับสนุนผลการวิจัยของ คลีนและโคเลน (1985) ที่ว่า ความยาวของแบบสอบถามไม่ใช่ปัจจัยเดียวในองค์ประกอบของแบบสอบถามที่มีผลกระทบต่อการศึกษาเปรียบเทียบ และเป็นข้อสันนิษฐานได้ว่า ทำไมการวิจัยในเรื่องนี้จึงยังไม่สามารถลงข้อสรุปที่เป็นการทั่วไปได้อย่างชัดเจน ข้อสรุปที่พบจึงต้องมีเงื่อนไขของบริบทประกอบเสมอ เช่น ผลการวิจัยของปีเตอร์สันและคณะ (1981, 1982) ของคลีนและโคเลน (1985) เป็นต้น

3. จากข้อค้นพบในมิติของรูปแบบการเทียบมาตราที่ใช้ในกรณีแบบสอบคัดเลือกซึ่งให้ผลสอดคล้องกับสมมติฐานที่กำหนดไว้ในข้อที่ 3 ที่ว่า รูปแบบอีควิเปอร์เซนไคล์และรูปแบบ IRT เป็นวิธีที่มีความเพียงพอมากกว่ารูปแบบเชิงเส้นตรงนั้น เมื่อย้อนกลับไปพิจารณาสถานการณ์การสอบคัดเลือกที่ได้เลือกมาใช้ นั้น มีปัจจัยแวดล้อมที่สนับสนุนผลการวิจัยสองประการ คือ ปัจจัยด้านแบบสอบและปัจจัยด้านประชากรผู้สอบ ในด้านแบบสอบเป็นแบบสอบวัดความรู้ความสามารถทั่วไปในเชิงความถนัด มีพิสัยของระดับความยากกว้างพอสมควร (ค่าเฉลี่ยรายข้ออยู่ระหว่าง .13 ถึง .73) แบบสอบที่นำมาใช้มีความเที่ยงค่อนข้างสูง (.88) ภายหลังการจัดเตรียมชุดของแบบสอบเทียบมาตราเป็นคู่ ๆ (X และ Y) แต่ละคู่มีความเที่ยงใกล้เคียงกัน ในด้านประชากรผู้สอบเป็นประชากรที่

มีการกระจายของความสามารถกว้าง ข้อมูลกลุ่มตัวอย่างจากการออกแบบมีความถี่ในแต่ละชั้นของคะแนนมีพิสัยกว้าง จึงทำให้การประมาณความถี่ประชากรตามรูปแบบของอีควิเปอร์เซนไคล์ได้ผลแม่นยำ ดังนั้นการเทียบมาตราในกรณีแบบสอบคัดเลือกที่อยู่ภายใต้บริบทดังที่กล่าวมา รูปแบบอีควิเปอร์เซนไคล์ให้ความเพียงพอในการแปลงคะแนนสมมูลมากที่สุดเมื่อใช้แบบสอบร่วมขนาดเดียวกัน

แต่อย่างไรก็ตาม ถ้าพิจารณาปริมาณค่าดัชนี C แทนการจัดอันดับแล้ว พบว่ารูปแบบ IRT และแม้แต่รูปแบบเชิงเส้นตรงก็สามารถให้ความเพียงพอในการแปลงคะแนนสมมูลในระดับน่าพอใจอย่างมากได้เช่นเดียวกับรูปแบบอีควิเปอร์เซนไคล์ เมื่อใช้แบบสอบร่วมร้อยละ 60 หรือ 21 ข้อ ดังนั้น อาจกล่าวได้ว่า เมื่อการเทียบมาตราสามารถออกแบบใช้แบบสอบร่วมภายในขนาดยาวมากพอ เช่น ร้อยละ 60 ของแบบสอบเทียบมาตรากับกลุ่มตัวอย่างสมมูลสองกลุ่มแล้ว การใช้รูปแบบใดในการเทียบไม่ใช่สาระสำคัญมากนักในด้านความเพียงพอ แต่อาจไปพิจารณาในเชิงปฏิบัติถึงปัจจัยประกอบที่นอกเหนือจากที่กล่าวถึง เช่น ความพร้อมในการดำเนินการด้านต่าง ๆ มากกว่า

4. กรณีแบบสอบวัดผลสัมฤทธิ์ มีปัจจัยต่าง ๆ แตกต่างจากกรณีแบบสอบคัดเลือก

ในด้านแบบสอบถามมีขอบเขตของเนื้อหาที่จำกัด ภายในกรอบของหลักสูตร และในด้านผู้สอบประชากรมีความสามารถขั้นต่ำค่อนข้างชัดเจน นอกจากนี้ผู้สอบยังได้ผ่านกระบวนการเรียนการสอนวิชาที่สอบแล้ว จึงทำให้คะแนนที่ได้มีพิสัยที่แคบกว่ากรณีแบบสอบคัดเลือก แต่เนื่องจากการออกแบบเทียบมาตรฐานด้วยกลุ่มผู้สอบสมมูลสองกลุ่มที่ได้จากการสุ่ม จึงพบว่า การแจกแจงของคะแนนทั้งสองกลุ่มคล้ายคลึงกัน จากปัจจัยที่กล่าวมานี้เป็นปัจจัยที่ผู้วิจัยได้กำหนดและออกแบบไว้ และจากผลการวิเคราะห์พบความสอดคล้องกับสมมติฐานส่วนหนึ่งของข้อที่ 4 ที่ว่าวิธีการเทียบมาตรฐานที่ใช้รูปแบบเชิงเส้นตรงเป็นวิธีที่มีความเพียงพอมากกว่ารูปแบบอิกวิเปอร์เซนไคล์

สำหรับสมมติฐานอีกส่วนหนึ่งของสมมติฐานข้อที่ 4 ที่กล่าวว่า รูปแบบ IRT จะให้ความเพียงพอแก่วิธีการเทียบมาตรฐานมากกว่ารูปแบบอิกวิเปอร์เซนไคล์นั้น ไม่สามารถให้ผลการสนับสนุนจากการวิจัยครั้งนี้ ซึ่งได้อภิปรายในหัวข้อที่ 5 ต่อไป

5. การเปรียบเทียบผลการเทียบมาตรฐานต่างรูปแบบ พบว่า วิธีที่ใช้รูปแบบ IRT ไม่ได้เป็นวิธีที่น่าพอใจมากที่สุด ในกรณีแบบสอบคัดเลือกได้เป็นอันดับสองรองจากรูปแบบอิกวิเปอร์เซนไคล์และในกรณีแบบสอบวัดผลสัมฤทธิ์ได้เป็นอันดับสาม ซึ่งเป็น

อันดับสุดท้าย ถ้าพิจารณาเทียบกับผลงานวิจัยในอดีตซึ่งพบว่า เมื่อแบบสอบคู่ที่ทำการเทียบมาตรฐานมีความยากต่างกันแล้ว วิธีที่ใช้รูปแบบ IRT จะให้ผลที่น่าพอใจที่สุด (Slind and Linn 1977; Petersen and others 1981; Kolen 1981; Cook and others 1981) และพบว่า เมื่อการวิเคราะห์กับกลุ่มตัวอย่างขนาดเล็กเกินไป มีผลทำให้วิธีการเทียบตามรูปแบบ IRT ให้ผลไม่น่าพอใจทั้ง ๆ ที่แบบสอบมีลักษณะคล้ายคลึงกัน (Kolen and Khitney 1982) สำหรับการวิจัยครั้งนี้ ได้ใช้กลุ่มตัวอย่างขนาดใหญ่ทั้งสองกรณีของแบบสอบ และลักษณะแบบสอบที่เทียบมาตรฐานมีความแตกต่างกันบ้างเล็กน้อยเท่านั้น แต่ปรากฏผลดังที่กล่าวสรุปไปนั้น อาจสืบเนื่องจากสาเหตุ 4 ประการ ดังนี้

ก. ปัญหาการวิเคราะห์ค่าประมาณพารามิเตอร์ข้อสอบที่ใช้โปรแกรมคอมพิวเตอร์ LOGIST ที่สถาบันคอมพิวเตอร์จุฬาลงกรณ์มหาวิทยาลัย ซึ่งมีขีดจำกัดในการใช้โปรแกรม ทั้งนี้เนื่องจากกระบวนการคำนวณจำเป็นต้องใช้คำสั่งที่กำหนดไว้ในโปรแกรม (default) เมื่อการวิเคราะห์ประสบปัญหาเนื่องจากการประมาณค่าแต่ละขั้นตอนไม่ได้บรรลุตามเกณฑ์ จำเป็นต้องใช้วิธีการปรับปรุงค่าภายหลังการทำงานของโปรแกรมแต่ละครั้ง แล้วจึงส่งงานใหม่ การวิเคราะห์จะเริ่มต้นใหม่ทุกครั้ง ทำให้สิ้น

เปลืองเวลา และเป็นการตัดทอนกำลังของ การประมาณ ค่าประมาณที่ได้เมื่อสิ้นสุด การทำงานจึงอาจจะยังมีความแปรปรวนอยู่ ความจริงโดยลักษณะของโปรแกรม LOGIST สามารถรับคำสั่งการทำงานอย่างต่อเนื่อง โดยการเก็บผลครั้งก่อนลงในหน่วยความจำ ที่เตรียมไว้ และใช้เป็นข้อมูลของการทำงาน ครั้งต่อไป หลาย ๆ ครั้งจนกว่าได้ค่าประมาณ ที่ไม่แปรเปลี่ยน

ข. การเทียบมาตรารูปแบบ IRT สามพารามิเตอร์ วิธีเทียบคะแนนจริงเป็นวิธี เทียบในแนวคิดของแบบสอบมิติเดียวสองชุด ที่วัดความสามารถเดียวกัน ซึ่งให้คุณสมบัติ ของผลการเทียบมาตรา 3 ประการ คือความ ยุติธรรม ความไม่แปรเปลี่ยนเมื่อเปลี่ยนกลุ่ม และความสมมาตร (Lord 1981 : 199) ซึ่ง คุณสมบัติดังกล่าวนี้จะเกิดได้เฉพาะใช้กับการ เทียบคะแนนจริง แต่ในทางปฏิบัติเราไม่ สามารถรู้คะแนนจริงของผู้สอบทุกรายบุคคล ในการสอบ เมื่อได้ตารางเทียบคะแนนสมมูล จากกระบวนการเทียบมาตราแล้ว จะนำไป ใช้กับคะแนนสังเกตของแต่ละบุคคล (Lord 1982 : 147) จากที่กล่าวมานี้ อาจเป็นสาเหตุ หนึ่งที่ทำให้การวิเคราะห์กลุ่มสอบทานผลได้ คะแนนแปลง X^* ที่อ่านค่าจาก y แตกต่าง จาก x ที่รายบุคคลได้จริง และสาเหตุนี้อาจ ได้รับผลกระทบจากการประมาณค่าพารา มิเตอร์ดังที่อภิปรายในข้อ ก. มาก่อนแล้ว เมื่อใช้ดัชนีเปรียบเทียบความแตกต่าง C

ตัดสินผลสุดท้ายของกระบวนการเทียบมาตรา แล้ว จึงให้ข้อสรุปว่า วิธีที่ใช้รูปแบบ IRT กับสถานการณ์แบบสอบที่ศึกษามีความเพียงพอ น้อยกว่ารูปแบบอิกวิเปอร์เซนไคล์ และ เชิงเส้นตรง

ค. ปัญหาการเทียบมาตราคะแนนส่วน ที่อยู่ต่ำกว่าระดับการเดา เป็นปัญหาในเชิง ปฏิบัติ เพื่อให้สามารถเทียบคะแนนของแบบ สอบได้ครบถ้วนตลอดพิสัยของคะแนน ลอร์ด ได้แนะนำให้ใช้วิธีปรับขยายเชิงเส้นตรง (Linear interpolation method) (Lord 1981 : 210-211) วิธีนี้ได้มีผู้นำไปใช้มาก เช่น โคลเลน (1981) และคูก ดันบาร์ และไอเนอร์ (1981) เป็นต้น แต่ยังไม่ได้มีการศึกษาถึงความ เหมาะสม และผลกระทบอย่างจริงจัง และเป็นข้ออภิปรายของนักวิจัยที่ใช้แนวปฏิบัติ นี้ที่ตั้งข้อสันนิษฐานว่า ปัญหา IRT ที่มีส่วน หนึ่งอาจเกิดจากสาเหตุนี้ได้

ข้อเสนอแนะ

1. ข้อเสนอแนะเกี่ยวกับการประยุกต์

เป็นความจริงดังที่การวิจัยในอดีตเกือบ ทุกงานได้สรุปว่า ไม่สามารถบอกได้ว่าวิธี การใด หรือรูปแบบใดที่ทำให้การเทียบมาตรา ได้ผลดีที่สุด จำเป็นต้องมีเงื่อนไขซึ่งเป็นตัว แปรที่อยู่เหนือการควบคุมเป็นตัวประกอบ ในคำสรุปเสมอ สำหรับผลการวิจัยครั้งนี้ก็ เช่นกัน ข้อสรุปที่ให้ไว้จะต้องอยู่ภายใต้เงื่อนไข

และสถานการณ์ที่มาพร้อมกับข้อมูล ดังนั้น การเลือกวิธีเทียบมาตราไปประยุกต์จึงต้องพิจารณาประเด็นต่าง ๆ โดยมีข้อค้นพบจากงานวิจัยต่าง ๆ ช่วยในการตัดสินใจสำหรับข้อเสนอแนะเพื่อนำไปประยุกต์ใช้นี้ เป็นข้อเสนอแนะภายใต้การออกแบบโดยมีแบบสอบร่วม (CIE) เมื่อแบบสอบที่ต้องการเทียบมาตรามีความคล้ายคลึงกันในด้านเนื้อหา และระดับความยาก และกลุ่มตัวอย่างผู้สอบ 2 กลุ่ม ได้จากการสุ่มอย่างง่ายจากประชากรเดียวกัน หรือมีการแจกแจงของคะแนนที่คล้ายคลึงกันทั้งสองกลุ่ม

1.1 ในสภาพการณ์การสอบคัดเลือกที่มีวัตถุประสงค์เพื่อจำแนกคนตามระดับความสามารถเชิงความถนัดทางวิชาการ เพื่อให้เข้ารับการศึกษาในอาชีพ หรือเข้าทำงาน โดยมีแบบสอบวัดความถนัด หรือความรู้ความสามารถทั่วไปที่มีความเที่ยงสูงใกล้เคียงกัน มีจำนวนผู้สอบมาก และมีพิสัยของการกระจายคะแนนกว้าง คือ มีความถี่กระจายตั้งแต่คะแนนต่ำสุดถึงสูงสุดแล้วการใช้รูปแบบอีควิเปอร์เซนไทล์ในการเทียบมาตราจะให้ความแม่นยำต่อการแปลงคะแนนสมมูลสำหรับคำถามที่เกี่ยวกับความยาวของแบบสอบร่วมที่ใช้ในสถานการณ์เช่นนี้ แบบสอบร่วมขนาด 21 ข้อ เมื่อแบบสอบเทียบมาตรามีข้อสอบชุดละ 35 ข้อ ให้ผลการเทียบมาตราที่น่าพอใจอย่างมาก และถ้าได้ทำ

การตรวจสอบค่าความเที่ยงของแบบสอบร่วมพบว่า มีความเที่ยงสูงถึงแม้จำนวนข้อจะลดลง เช่นในกรณีที่พบในการวิจัยครั้งนี้ แบบสอบร่วมนั้น ๆ ก็สามารถทำให้ผลการเทียบมาตราที่น่าพอใจได้เช่นกัน

1.2 ในสภาพการณ์การสอบวัดผลสัมฤทธิ์ที่มีบริบทของกระบวนการเรียนการสอนวัตถุประสงค์ของการสอบ คือ การนำไปสู่การตัดสินใจระดับผลสัมฤทธิ์ในการเรียนวิชานั้น ๆ เมื่อสิ้นภาคการศึกษา การสอบแต่ละครั้งได้ใช้แบบสอบที่คล้ายคลึงกันในเรื่องที่วัด การเทียบมาตราคะแนนของแบบสอบที่ใช้ต่างครั้งกันด้วยการออกแบบโดยใช้แบบสอบร่วมกับรูปแบบเชิงเส้นตรง จะให้ผลต่อการแปลงคะแนนสมมูลได้อย่างเพียงพอ

1.3 ถึงแม้ว่าข้อสรุปโดยส่วนรวมได้กล่าวว่า ความยาวของแบบสอบร่วมที่มากกว่าย่อมให้ผลการเทียบมาตราที่ถูกต้อง และแม่นยำกว่าก็ตาม แต่การตัดสินใจในด้านปฏิบัติย่อมต้องคำนึงถึงตัวแปรอื่น ๆ เช่น

ก. เวลา แบบสอบร่วมที่ยาวกว่าย่อมต้องใช้เวลาสอบมากกว่า

ข. วิธีการใช้แบบสอบร่วม ตามรูปแบบการใช้แบบสอบที่ใช้ในการวิจัยครั้งนี้เป็นแบบสอบร่วมภายใน หมายถึง การนับคะแนนของแบบสอบร่วมรวมเข้าไปในผล

การสอบเพื่อใช้ในการตัดสินใจ จึงเป็นสิ่ง ที่ควรคำนึงเมื่อต้องการใช้แบบสอบรวมขนาด ยาวมาก เพื่อหวังผลความถูกต้องแม่นยำของ คะแนนแปลง เพราะจะทำให้ผู้สอบที่เคยสอบ มาแล้วมีโอกาสรู้ และจดจำข้อสอบในส่วน ของแบบสอบรวมได้ ทำให้เกิดความเสียเปรียบแก่ผู้ที่ไม่เคยสอบโครงการนั้น ๆ มาก่อน

2. ข้อเสนอแนะการใช้เทคนิคการเทียบ มาตราในโครงการทดสอบ

การเทียบมาตราเป็นระเบียบวิธีของการ แปลงคะแนนสอบแต่ละครั้งให้เป็นคะแนนที่ สามารถให้ความหมายในเชิงปริมาณที่เปรียบเทียบกัน ได้โดยตรงระหว่างผู้สอบต่างครั้ง ต่างเวลาคนที่สอบในโครงการทดสอบวิชา เดียวกัน โดยไม่มีใครได้เปรียบหรือเสียเปรียบ อันเนื่องมาจากความยากที่ต่างกันของแบบ สอบใหม่และเก่า เทคนิควิธีเช่นนี้สามารถนำ เข้ามาใช้ในโครงการทดสอบที่เป็นโครงการ ต่อเนื่อง เช่น โครงการทดสอบเพื่อคัดเลือก บุคคลเข้าทำงาน โครงการทดสอบเพื่อวัดผล สัมฤทธิ์ทางการเรียนตามหลักสูตรอย่างมี มาตรฐานของการตัดสินใจ และโครงการทดสอบ เพื่อปรับตำแหน่งในการปฏิบัติงาน เป็นต้น ส่วนดีที่จะเกิดขึ้นจากการนำเทคนิควิธี เทียบมาตราในแต่ละโครงการมีดังต่อไปนี้

2.1 โครงการทดสอบเพื่อคัดเลือกบุคคล เข้าทำงาน

โดยปกติของโครงการทดสอบคัดเลือก บุคคลเข้าทำงานในส่วนของ การสอบข้อเขียน มักประกอบด้วยหมวดวิชาความรู้ทั่วไป และ หมวดวิชาเฉพาะ หมวดวิชาความรู้ทั่วไป ซึ่งหมายถึง ความสามารถในเชิงความถนัด ทางวิชาการทั่ว ๆ ไป ซึ่งเป็นวิชาที่กำหนด ให้ผู้สมัครเข้าทำงานในสาขาต่างต้องผ่าน การทดสอบ เช่น การสมัครเข้ารับ การคัดเลือกบรรจุเป็นข้าราชการที่จัดดำเนินการ โดยคณะกรรมการข้าราชการพลเรือน (ก.พ.) แต่ครั้งของการจัดสอบแข่งขัน พบว่าในระดับวิชาชีพ หรือปริญญาตรี ผู้สมัครจำนวน นับหมื่น การสอบแต่ละครั้งได้นำผลการสอบ มาเรียงลำดับ และพิจารณาบุคคลตามลำดับ คะแนน ผู้ที่มีลำดับเกินกว่าจำนวนที่รับได้ จะมีชื่อในบัญชีสำรองจำนวนหนึ่ง บัญชีนี้จะ ถูกยกเลิกเมื่อมีการจัดสอบครั้งใหม่ ปัญหา ที่พบ คือ การจัดสอบให้กับคนจำนวนมาก ๆ ในเวลาที่กำหนดทุกครั้ง งานด้านเตรียม สถานที่ ผู้ควบคุมการสอบ ตลอดจนการจัด เตรียมแบบสอบ และอื่น ๆ นับตั้งแต่การรับ สมัครจนถึงการประกาศผลการสอบ เป็น งานหนักมาก นอกจากนี้ตามระเบียบที่กำหนด ให้มีการยกเลิกผลการสอบเมื่อมีการสอบ ครั้งใหม่ ทำให้ผู้สมัครที่เคยสอบและที่ต้อง การสมัครอีกต้องทำการสอบซ้ำ เป็นการ สิ้นเปลืองทั้งเวลา และงบประมาณส่วนตัว

และเงินส่วนของรัฐ ทางแก้อ่อนนี้ คือ 1) จัดการสอบให้บ่อยขึ้น และกระจายหลายสนามสอบเพื่อเฉลี่ยคนสอบในแต่ละครั้งให้น้อยลง ซึ่งให้ความสะดวกในการเตรียมสถานที่สอบ และแบบสอบ 2) กำหนดให้ใช้ระเบียบวิธีเทียบมาตรฐานสำหรับแปลงคะแนนดิบจากการสอบไปสู่คะแนนมาตรฐานที่สามารถเปรียบเทียบกันได้อย่างครั้งก่อนหรือครั้งต่อไป โดยการออกแบบการเทียบมาตรฐานอย่างมีแบบสอบถาม (CIE) 3) ออกใบรับรองผลการสอบเพื่อให้นำไปใช้เป็นหลักฐานในการสมัครเข้าทำงานตามกำหนดอายุของการใช้ผลการสอบ อาจเป็น 3 ปี หรือ 5 ปี เป็นต้น

2.2 โครงการทดสอบเพื่อวัดผลสัมฤทธิ์ทางการเรียนตามหลักสูตร

หลักสูตรเป็นโครงการต่อเนื่องระยะยาว โดยมีวัตถุประสงค์ให้ผู้เข้ารับการอบรม หรือผ่านกระบวนการเรียนได้สัมฤทธิ์ผลตามที่ได้กำหนดไว้ในหลักสูตร เช่น หลักสูตรมัธยมศึกษา หลักสูตรระดับอุดมศึกษา เป็นต้น การประเมินผลสัมฤทธิ์ของหลักสูตรที่สำคัญ คือ การสอบวัดผลสัมฤทธิ์ทางการเรียนของนักเรียน แบบสอบที่ใช้กับนักเรียนหรือนักศึกษาแต่ละปีการศึกษา มักมีการเปลี่ยนแปลง การประเมินโดยใช้คะแนนดิบของแบบสอบแต่ละครั้งจึงไม่สามารถนำมาเปรียบเทียบข้ามรุ่นได้อย่างมีความหมายและ

ถูกต้อง ดังนั้น หากนำเทคนิคการเทียบมาตรฐานที่มีแบบสอบถาม (CIE) มาใช้กับการสอบประจำปีการศึกษาในตลอดปลายภาคทุกปี ผลการสอบสามารถช่วยให้ผู้ประเมินและพัฒนาหลักสูตรอย่างมีความมั่นใจในการเปรียบเทียบผลกับคนอื่น ๆ ในหลักสูตรเดียวกันได้ สามารถควบคุมมาตรฐานในการตัดสินได้-ตกอย่างคงเส้นคงวาลดโครงการหลักสูตรได้

2.3 โครงการทดสอบเพื่อปรับตำแหน่งในระบบการทำงาน

ตามระเบียบการปฏิบัติงานราชการในบางสายบางระดับได้กำหนดให้มีการสอบเพื่อปรับเลื่อนตำแหน่งความรับผิดชอบงานให้สูงขึ้น เมื่อบุคคลนั้นมีคุณสมบัติถึงระดับหนึ่ง เช่น การสอบเพื่อเลื่อนชั้นในสายงานวิชาการจากตำแหน่งที่ 4 เป็นที่ 5 เป็นต้น การสอบเป็นเกณฑ์หนึ่งที่น่ามาประเมินความเหมาะสมในด้านความรู้ความสามารถของผู้ขอตำแหน่ง ผลการสอบในแต่ละครั้งได้นำมาจัดอันดับ ผู้ที่ได้คะแนนสูงเป็นผู้ได้รับการพิจารณาก่อน ผลการสอบของผู้ที่ยังไม่ได้รับการปรับตำแหน่งจะถูกยกเลิกโดยอัตโนมัติเมื่อมีการสอบครั้งใหม่ และในทำนองเดียวกันผู้ที่มีคุณสมบัติแล้ว แต่ยังไม่มีการสอบ ผู้นั้นจะต้องรอจนกว่าจะถึงเวลาสอบ วิธีปฏิบัตินี้หากได้นำเทคนิคการเทียบมาตรฐานที่มีแบบสอบถามเข้าไปใช้ จะช่วยเสริมด้านคุณ

ภาพที่เป็นมาตรฐานจากการสอบ และให้โอกาสของแต่ละบุคคลเพื่อได้รับการประเมินในส่วนของคุณวุฒิ ความสามารถอย่างยุติธรรมมากขึ้น คือ ไม่มีผลกระทบจากข้อสอบที่ง่ายกว่าหรือยากกว่าครั้งก่อน นอกจากนี้การจัดการสอบยังสามารถจัดได้บ่อย ๆ โดยไม่ต้องรอให้เสียโอกาสส่วนหนึ่ง

3. ข้อเสนอแนะในการวิจัยต่อ

3.1 จากการออกแบบการศึกษาวิธีการเทียบมาตรฐานที่มีแบบสอบร่วม (CIE) ได้ศึกษาเพียงตัวแปรในมิติของรูปแบบสามรูปแบบ และความยาวแบบสอบร่วม 3 ขนาด ในสถานการณ์การทดสอบเพื่อการคัดเลือกบุคคลกับวัดผลสัมฤทธิ์ทางการเรียน ข้อค้นพบยังมีขีดความสามารถในการอธิบายจำกัดอยู่มาก จึงน่าจะได้มีการทำวิจัยในตัวแปรอื่น ๆ เช่น การศึกษาการเทียบมาตรฐานที่มีแบบสอบร่วมชนิดภายนอก (External anchor test) ลักษณะค่าสถิติของแบบสอบร่วม ความคล้ายคลึงของแบบสอบเทียบมาตรฐานกับแบบสอบร่วม และแม้กระทั่งความยาวของแบบสอบร่วมเมื่อแบบสอบเทียบมาตรฐานมีความยาวมากกว่า 35 ข้อ ดังเช่นที่ใช้ในการวิจัยครั้งนี้

3.2 ปัจจุบันได้มีรูปแบบการเทียบมาตรฐานเกิดขึ้นอีกหลายรูปแบบ น่าจะได้มี

การศึกษาค้นคว้าต่อ เพื่อเป็นสารสนเทศสำหรับการตัดสินใจเลือกใช้ในโครงการทดสอบได้กว้างขวางยิ่งขึ้น เช่น วิธี Pre-equating ซึ่งมีการนำไปใช้ในโครงการทดสอบ TOEFL การศึกษา Pre-equating เป็นวิธีการเทียบมาตรฐานโดยอาศัยทฤษฎีการตอบข้อสอบในการหาค่าประชากรข้อสอบไว้ก่อน เมื่อสร้างชุดแบบสอบแล้วจะคำนวณสารสนเทศต่าง ๆ จากค่าประชากรรายข้อที่หาไว้แล้ว สร้างตารางแปลงคะแนนสมมูลจากชุดหนึ่งไปยังอีกชุดหนึ่ง วิธีการนี้ผู้ออกแบบชุดของข้อสอบควรมีคลังข้อสอบที่เคยผ่านการทดลองใช้แล้ว การจัดชุดแบบสอบแต่ละครั้งแตกต่างกันเฉพาะในองค์ประกอบของรายข้อ วิธี Pre-equating นี้ นักวัดผลสามารถสร้างตารางคะแนนสมมูลไว้ก่อนเสมอ จึงเป็นการสะดวกสำหรับ และสามารถเสนอคะแนนผลการสอบได้อย่างรวดเร็ว

3.3 การศึกษาโดยเลือกสถานการณ์ตามความเป็นจริงมาทำการศึกษาสองสถานการณ์ดังที่เสนอมานี้ เป็นเพียงการบรรยายตามสภาพแต่ไม่สามารถควบคุมตัวแปรแทรกซ้อน ดังนั้น น่าจะศึกษาโดยการควบคุมตัวแปรเพื่อดูผลให้ชัดเจน เช่น ตัวแปรเกี่ยวกับแจกแจงของคะแนนประชากรที่มีลักษณะต่าง ๆ กัน ความเที่ยงของแบบสอบเทียบมาตรฐาน และแบบสอบร่วม เป็นต้น ●

บรรณานุกรม

- Angoff, W.H. *Scales, Norms, and Equivalent Scores*. Educational Testing Service, 1984. Republication of Chapter 15 in R.L. Thorndike (Ed.) *Educational Measurement*. (2nd ed.) Washington, D.C.: American Council on Education, 1971. pp. 508-600.
- Budescu, David. "Efficiency of Linear Equating as a Function of the Length of the Anchor Test." *Journal of Educational Measurement* 22 (1985): 13-20.
- Donlon, T.F. and Angoff, W.H. "The Scholastic Aptitude Test." In W.H. Angoff (Ed), *The College Board Admission Testing Program: A technical report on research and development activities relating to the SAT and AT*. New York: College Entrance Examination Board, 1971. pp.15-48.
- Holland, Paul W. and Rubin Donald B. (Ed) *Test Equating*. Educational Testing Service, Princeton, New Jersey: Academic Press, 1982.
- Jaeger, R.M. "Some Exploratory Indices for Selection of a Test Equating Method." *Journal of Educational Measurement* 18 (1981): 23-38.
- Klein, Lawrence W., Kolen, Michael J. "Effect of Number of Common Items in Common-Item Equating with Nonrandom Groups." Paper presented at Annual Meeting of the American Educational Research Association, Chicago, 1985.
- Kolen, M.J. and Whitney, D.R. "Comparison of Four Procedures for Equating The Tests of General Educational Development." *Journal of Educational Measurement* 19 (1982): 279-294.
- LaValle, K.P. "Truth in Testing." *Today's Education* 69 (1980): 59-64.
- Lord, F.M. *Applications of Item Response Theory to Practical Testing Problems*. Hillsdale, N.J.: Erlbaum, 1980.
- . "Standard Error of an Equating by Item Response Theory." Educational Testing Service, Princeton, New Jersey: ERIC Document Reproduction Service, November, 1981.
- . (a). "The standard error of Equipercentile Equating." *Journal of Educational Statistics* 7 (1982): 165-174.
- Marco, Caryl. "Equating Tests in an Era of Test Disclosure." *New Directions for Testing and Measurement*, (1981, Sep): 105-122.
- Rangchaikul, Yavadee. "Equating and Calibration Technique For Test Program in Thailand." A Dissertation Submitted to the Area of Educational Management Systems, The Florida State University. December, 1975.
- Wingersky M.S.; Barton and Lord, F.M. *Logist 5-A* computer program for estimating examinee ability and item characteristic curve parameter. Princeton NJ: Educational Testing Service, 1982.