

บทที่ ๕

บทสรุป

วัตถุประสงค์ของการวิจัย

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อศึกษาเปรียบเทียบผลของการใช้รูปแบบการเทียบมาตราที่ต่างกัน ๓ รูปแบบ คือ ๑) รูปแบบอิกวิเปอร์เซนไทล์ ๒) รูปแบบเชิงเส้นตรง และ ๓) รูปแบบอิงทฤษฎีการทอนข้อสอบแบบสามพารามิเตอร์ และที่ใช้การออกแบบเทียบมาตราโดยมีแบบสอบรวมชนิดภายในต่างกันสามขนาด ซึ่งคิดเป็นอัตราส่วนร้อยละของแบบสอบที่ทำการเทียบมาตรา ดังนี้

๑) ขนาดร้อยละ ๒๐ (๒) ขนาดร้อยละ ๔๐ และ ๓) ขนาดร้อยละ ๖๐ ผลของการเทียบมาตราที่นำมาใช้ประเมิน ๒ ลักษณะ คือ ๑) ค่าความคลาดเคลื่อนมาตรฐานของการเทียบมาตรา (SEE) ซึ่งเป็นการวิเคราะห์ที่ในเชิงทฤษฎีของรูปแบบ และ ๒) กซ์นิการเปรียบเทียบความแตกต่าง (Index C) ที่ได้จากการวิเคราะห์กลุ่มตัวอย่างสอบทานผล ซึ่งเป็นการวิเคราะห์ความเพียงพอของวิธีการเทียบมาตราโดยตรงกับผลการเทียบสุดท้าย

สมมุติฐานของการวิจัย

๑. แบบสอบรวมที่มีความยาวมากกว่ามีประสิทธิภาพสัมพัทธ์ของการเทียบมาตราที่สูงกว่าในทุกรูปแบบของการเทียบมาตราของทั้งสองกรณีของแบบสอบที่ใช้
๒. วิธีการเทียบมาตราที่ใช้แบบสอบรวมที่มีความยาวมากกว่า เป็นวิธีที่มีความเพียงพอกว่าวิธีที่ใช้แบบสอบรวมที่มีความยาวน้อยกว่า
๓. วิธีการเทียบมาตราที่ใช้รูปแบบอิกวิเปอร์เซนไทล์ และรูปแบบอิงทฤษฎีการทอนข้อสอบ เป็นวิธีที่มีความเพียงพอมากกว่าวิธีที่ใช้รูปแบบเชิงเส้นตรง ในกรณีแบบสอบคัดเลือก

๔. วิธีการเทียบมาตรฐานที่ใช้รูปแบบเชิงเส้นตรง และรูปแบบอิงทฤษฎีการตอบข้อสอบ เป็นวิธีที่มีความเพียงพอมากกว่าวิธีที่ใช้รูปแบบอควิปเปอร์ เช่น โกล์ ในกรณีแบบสอบวัดผลสัมฤทธิ์

วิธีวิจัย

๑. ประชากรและกลุ่มตัวอย่าง

การวิจัยครั้งนี้ใช้สถานการณการทดสอบที่โหระมุลคะแนน ๒ กรณี คือ ๑) กรณีแบบสอบคัดเลือก ประชากร คือ ผลการสอบรายบุคคลซึ่งมีคุณสมบัติขึ้นทำคามกำหนดของระเบียบการสอบคัดเลือกที่กำหนดไว้ มีจำนวน ๑๕๕๕ คน สุ่มตัวอย่างสำหรับการเทียบมาตรา ๒ กลุ่ม กลุ่มละ ๑๕๐๐ คน และสำหรับการวิเคราะห์กลุ่มสอบทานผลอีก ๑ กลุ่ม จำนวน ๑๕๐๐ คน ๒) กรณีแบบสอบวัดผลสัมฤทธิ์ ประชากร คือ ผลการสอบรายบุคคลในวิชาที่กำหนดในระดัับอุดมศึกษา จำนวน ๑๒๓๘๐ คน สุ่มตัวอย่างห่านองเกี่ยวกับกรณีแรก

๒. แบบสอบ

ในแต่ละกรณีของแบบสอบที่นำมาศึกษา ประกอบด้วยข้อสอบ ๑๐๐ ข้อ ใช้วัดความรู้หนึ่งวิชา ผู้วิจัยได้นำมาวิเคราะห์รายข้อ แล้วนำมาเลือก และจัดเป็นแบบสอบที่วัดในเรื่องเดียวกัน มีความยากง่ายใกล้เคียงกัน ๒ ชุด ชุดละ ๓๕ ข้อ เพื่อใช้เป็นแบบสอบสำหรับทำการเทียบมาตรา และทำการคัดเลือกจัดชุดแบบสอบรวมขนาด ๓ ข้อ ๑๔ ข้อ และ ๒๐ ข้อ จากข้อสอบที่เหลือ โดยให้แบบสอบรวมมีลักษณะเป็นฉบับย่อยส่วนองแบบสอบเทียบมาตราที่จัดไว้แล้ว จากการจัดเตรียมแบบสอบของกรณีสอบคัดเลือก ได้แบบสอบเทียบมาตราซึ่งได้ผนวกเอาแบบสอบรวมแล้ว ๓ ชุด คือ XV60 กับ YV60 XV40 กับ YV40 และ XV20 กับ YV20 ความเที่ยงของแบบสอบแต่ละคู่มักค่าต่างกันเล็กน้อย คือ ต่างกัน .๐๑๒ .๐๒๕ และ .๐๒๓ ตามลำดับ กรณีแบบสอบวัดผลสัมฤทธิ์ได้แบบสอบเทียบมาตรา ๓ ชุด เช่นเดียวกัน มีความเที่ยงต่างกันตามลำดับ ดังนี้ .๐๑๓ .๐๑๕ และ .๐๒๓ ตามลำดับ จากผลการจัดเตรียมแบบสอบทั้งสองกรณี พบว่า แบบสอบกรณีหลังมีความเที่ยงต่างกันในแต่ละคู่อของแบบสอบเทียบมาตรามากกว่ากรณีแรกเล็กน้อย

๓. การรวบรวมข้อมูล

ข้อมูล เป็นประเภททฤษฎี จากการออกแบบการเทียบมาตราให้กลุ่มตัวอย่างหนึ่ง (α) รับการทดสอบด้วยแบบสอบชุด X และอีกกลุ่มหนึ่ง (β) ทำชุด Y และทั้งสองกลุ่มได้รับการทดสอบส่วนที่เป็นแบบสอบร่วมเหมือนกัน การประมวลข้อมูลให้เป็นไปตามสถานการณ์ดังที่กล่าวมาแล้วนี้ ใ้โดยอาศัยการจักรกระทำข้อมูลผลการตอบรายข้อที่บันทึกไว้แล้ว จากการสอบถามสภาพความเป็นจริงของแต่ละกรณี ข้อมูลที่สังเคราะห์ขึ้นเพื่อกำเนิการเทียบมาตราจึงเป็นคะแนนของแบบสอบชุด XV60 XV40 และ XV20 จากกลุ่มตัวอย่าง และ YV60 YV40 และ YV20 จากกลุ่มตัวอย่าง

ผลการรวบรวมข้อมูล ในกรณีของแบบสอบคัดเลือก กลุ่มตัวอย่าง α และ β มีการแจกแจงคล้ายคลึงกันโดยประมาณ คะแนนเฉลี่ยของแบบสอบชุด Y สูงกว่าชุด X ทุกขนาดของแบบสอบร่วมโดยประมาณอยู่ ๐.๕ คะแนน แต่ส่วนเบี่ยงเบนมีค่าต่ำกว่าประมาณ ๐.๖ ในกรณีแบบสอบวัดผลสัมฤทธิ์มีลักษณะกลับกัน คือ คะแนนเฉลี่ยของชุด Y ใ้ต่ำกว่าคะแนนเฉลี่ยของชุด X ประมาณ ๐.๑ คะแนน และส่วนเบี่ยงเบนสูงกว่าประมาณ ๐.๕

๔. การเทียบมาตรา

วิธีการเทียบมาตราในแต่ละกรณีของแบบสอบ ประกอบด้วยรูปแบบที่ใช้กำหนดกฎการแปลงคะแนนสมมูล ซึ่งมีอยู่ ๓ รูปแบบ และการออกแบบที่ใช้แบบสอบร่วมชนิดที่เป็นแบบสอบร่วมภายใน ๓ ขนาด จึงรวมเป็นวิธีการเทียบมาตราทั้งหมด ๔ วิธี ภายหลังจากการประมวลผลคะแนนของกลุ่มตัวอย่างตามแบบสอบชุดต่าง ๆ ที่ต้องการเทียบแล้ว ใ้ดำเนินการเทียบมาตราทั้ง ๔ วิธี แต่ละวิธี ใ้สรูปเป็นการวางคะแนนสมมูล แปลงจากมาตราชุด Y ซึ่งสมมุติเป็นแบบสอบชุดที่สร้างใหม่ไปสู่คะแนนแปลง X^* ที่มีมาตราตามชุด X หรือที่สมมุติว่าเป็นแบบสอบเก่า

การวางคะแนนสมมูลของการเทียบมาตราคะแนนของแบบสอบคู่ใด ๆ ใ้เป็นการวางอ้างอิงของการเทียบมาตราเพื่อจัดทำรายงานผลการสอบตามมาตราที่ต้องการ ลักษณะของคะแนนสมมูลที่ผ่านการเทียบมาตราทั้ง ๔ วิธี มีลักษณะสอดคล้องกันเป็นส่วนใหญ่ในช่วงคะแนนตอนกลาง ๆ และความสอดคล้องก็กล่าวจะน้อยลงในตอนปลาย แบบสอบคู่ที่มีแบบสอบร่วมมากกว่า มีความสอดคล้อง

คล่องของคะแนนแปลงมากกว่า (ทั้งแสดงในแผนภาพที่ ๖ และแผนภาพที่ ๗)

การวิเคราะห์ข้อมูล

๑. วิเคราะห์ความคลาดเคลื่อนมาตรฐานของการเทียบมาตรา (SEE) ทั้ง ๓ รูปแบบ การเปรียบเทียบปริมาณความคลาดเคลื่อนจากวิธีการเทียบมาตราที่ต่างกันในแต่ละกรณี ได้ใช้ค่า SEE ที่ประมาณจากตำแหน่งคะแนนมาตรฐานเดียวกัน โดยพิจารณาในช่วงคะแนนมาตรฐาน ± ๓.๐ ของคะแนน Y ทดสอบความเหมาะสมของวิธีเทียบมาตราใช้พิจารณาจากดัชนีประสิทธิภาพสัมพัทธ์ (RE) ซึ่งเป็นอัตราส่วนร้อยละระหว่างความคลาดเคลื่อนของวิธีเทียบมาตราที่ใช้แบบผสมรวม V40 และ V20 กับเกณฑ์ คือ V60 ภายในแต่ละรูปแบบของการเทียบมาตรา

๒. วิเคราะห์กลุ่มตัวอย่างสองทานผล เพื่อหาดัชนีความแตกต่าง (C) ระหว่างคะแนนแปลงที่อ่านจากการวางเทียบมาตราวิธีต่าง ๆ กับคะแนนเกณฑ์ ซึ่งเป็นคะแนนที่กลุ่มตัวอย่างได้โดยตรง จากการตอบแบบสอบถามนั้น ๆ เป็นความคลาดเคลื่อนรวมที่เกิดจากการใช้ตารางคะแนนสมมูลอันเป็นผลจากการเทียบมาตรา ทั้งนี้เปรียบเทียบความแตกต่าง (C) เป็นค่าที่ได้ถ่วงน้ำหนักให้เป็นมาตรฐานของค่าเฉลี่ยของความแตกต่างกำลังสอง ซึ่งสามารถนำมาเปรียบเทียบในเชิงปริมาณได้โดยตรง ไม่ว่าจะได้จากวิธีการเทียบมาตราที่ต่างรูปแบบ และต่างสถานการณ์ นำค่าดัชนี C มาทำการวิเคราะห์คุณภาพของการเทียบมาตราโดยใช้เกณฑ์ของปีเตอร์สันและคณะก่อน แล้วจึงทดสอบความมีนัยสำคัญของการจัดอันดับของค่า C ด้วยวิธีสถิติแบบพาราเมตริก Friedman model ซึ่งเป็นการวิเคราะห์ความแปรปรวนสองทางที่ใช้กับข้อมูลจัดอันดับที่มีการวัดซ้ำ ทำการวิเคราะห์แบบรายคู่ตามกำหนด

การวิเคราะห์ข้อมูลส่วนใหญ่อาศัยเครื่องคอมพิวเตอร์ของสถาบันบริการคอมพิวเตอร์ของจุฬาลงกรณ์มหาวิทยาลัย โดยใช้โปรแกรมสำเร็จรูป SPSSX LOGIST 5 และโปรแกรมที่ผู้วิจัยสร้างขึ้น

สรุปผลการวิจัย

๑. ผลการเทียบมาตราโดยส่วนรวม

จากการเทียบมาตราแบบสอขแต่ละกรณีด้วยรูปแบบการเทียบมาตรา ๓ รูปแบบ และใช้ความยาวแบบสอขร่วม ๓ ขนาด พบว่า ทั้ง ๔ วิธี ที่ดำเนินไปในแต่ละกรณี ให้ผลการเทียบที่อยู่ในระดับที่ยอมรับได้ทั้งสี่ตามเกณฑ์ของปีเตอร์สันและคณะ กล่าวคือ ในกรณีแบบสอขคัดเลือกวิธีที่ใช้แบบสอขร่วม ๒๑ ข้อ (ร้อยละ ๒๐) ๑๔ ข้อ (ร้อยละ ๘๐) กับรูปแบบอีคิวเปอร์เซนไทล์ และ IRT ให้ผลการเทียบมาตราในระดับที่น่าพอใจอย่างมาก และวิธีที่ใช้แบบสอขร่วม ๒๑ ข้อ (ร้อยละ ๒๐) กับรูปแบบเชิงเส้นตรงให้ผลที่น่าพอใจอย่างมากเช่นกัน ส่วนวิธีอื่นที่เหลือให้ผลในระดับที่ต่ำลงมาเล็กน้อย คือ อยู่ในระดับที่น่าพอใจ สำหรับกรณีแบบสอขวัดผลสัมฤทธิ์ทั้ง ๔ วิธี ให้ผลการเทียบมาตราในระดับที่น่าพอใจทั้งสิ้น

๒. มิตินี้เกี่ยวกับความยาวของแบบสอขร่วม

ในการศึกษาผลการเทียบมาตราที่ใช้แบบสอขขนาดต่างกัน พบว่า ความยาวของแบบสอขร่วมเป็นปัจจัยหนึ่งที่มีผลกระทบต่อความแม่นยำ และความเพียงพอโดยตรง เมื่อแบบสอขร่วมมีความยาวเพิ่มขึ้น ความคลาดเคลื่อนมาตรฐานของการเทียบมาตราก็ลดลง และในทำนองเดียวกัน ความคลาดเคลื่อนรวมหรือค่าคิซีบีเปรียบเทียบความแตกต่าง c ก็ลดลงด้วย แต่มีข้อน่าสังเกตที่ว่า ในกรณีของแบบสอขคัดเลือก วิธีที่ใช้แบบสอขร่วม ๒๑ ข้อ และ ๑๔ ข้อ ค่าความคลาดเคลื่อนมาตรฐานของการเทียบมาตรา (SEE) และความคลาดเคลื่อนรวม (ค่าคิซีบี c) มีค่าต่างกันเพียงเล็กน้อย ซึ่งทำให้เห็นว่าการเพิ่มความยาวของแบบสอขร่วมให้ผลต่อการเพิ่มประสิทธิภาพของการเทียบมาตราที่ไม่สู้จักเจนนัก การเปรียบเทียบผลการเทียบมาตราภายในรูปแบบโดยใช้เกณฑ์การประเมินภายในและภายนอก มีดังนี้

๒.๑ ผลการศึกษาเปรียบเทียบความคลาดเคลื่อนมาตรฐานของการเทียบมาตราที่ใช้แบบสอขร่วมขนาดต่าง ๆ กันในแต่ละรูปแบบของการเทียบมาตรา พบว่า การเทียบด้วยรูปแบบอีคิวเปอร์เซนไทล์ และเชิงเส้นตรง ให้ผลสอดคล้องกับสมมุติฐานของการวิจัยข้อที่ ๑ ทั้งสองกรณี

ของแบบสอบที่นำมาใช้ในการศึกษาคั้งนี้ กล่าวคือ เมื่อใช้แบบสอบรวมที่ยาวที่สุด (๒๐ ข้อ หรือ ร้อยละ ๖๐ ของแบบสอบเทียบมาตรา) ค่าความคลาดเคลื่อนมาตรฐานของการเทียบมাত্রามีค่าน้อยที่สุด เมื่อ เทียบกับวิธีแบบสอบรวมที่ขนาดสั้นลงอีกสองขนาด และเมื่อใช้แบบสอบรวมขนาดกลาง (๑๔ ข้อ หรือร้อยละ ๔๐) ค่าความคลาดเคลื่อนที่คำนวณได้จัดอยู่ในอันดับกลาง และสุดท้าย คือ วิธีที่ใช้แบบสอบรวมขนาดสั้นที่สุด (๗ ข้อ หรือร้อยละ ๒๐) ค่าความคลาดเคลื่อนมีค่ามากที่สุด ส่วนวิธีการเทียบมาตราด้วยรูปแบบ IRR ให้ผลในทำนองเดียวกัน ยกเว้นวิธีที่ใช้แบบสอบรวมขนาด ๑๔ ข้อ ให้ผลไม่แตกต่างจากวิธีที่ใช้ ๒๐ ข้อ จึงสรุปได้ว่า การเทียบมาตราที่ใช้กรณีแบบสอบคัดเลือก และวัดผลสัมฤทธิ์ตามที่นำมาศึกษาคั้งนี้ ด้วยรูปแบบการเทียบมาตราสามรูปแบบ พบว่าการใช้แบบสอบรวมที่ยาวกว่าให้ประสิทธิภาพของการเทียบมาตราที่สูงกว่า

๒.๒ การศึกษาความเพียงพอของวิธีการเทียบมาตราอันเกิดจากการใช้แบบสอบรวมที่มีขนาดต่างกันในแต่ละรูปแบบการเทียบมาตราด้วยเทคนิคการวิเคราะห์กลุ่มสอบทานผล ซึ่งเป็นเกณฑ์ประเมินอีกเกณฑ์หนึ่งที่น่ามาประเมินความเพียงพอของวิธีการเทียบมาตรา ผลการศึกษา พบว่า ในกรณีแบบสอบคัดเลือก วิธีการเทียบมาตราที่ใช้แบบสอบรวมขนาด ๒๐ ข้อ (ร้อยละ ๖๐) และ ๑๔ ข้อ (ร้อยละ ๔๐) มีความเพียงพอในระดับที่น่าพอใจมากกว่าทุกรูปแบบที่ใช้ ยกเว้นรูปแบบเชิงเส้นตรงที่ใช้แบบสอบรวม ๑๔ ข้อ ให้ผลในระดับที่น่าพอใจ ส่วนวิธีที่ใช้แบบสอบรวม ๗ ข้อ (ร้อยละ ๒๐) กับรูปแบบทั้งสามให้ผลการเทียบในระดับที่น่าพอใจ สำหรับกรณีแบบสอบวัดผลสัมฤทธิ์ พบว่า ทั้ง ๔ วิธี ของการเทียบมাত্রามีความเพียงพอในระดับที่น่าพอใจ จากการจัดอันดับตามค่าดัชนี C พบว่า ในมิติของความยาวแบบสอบรวมมีแบบแผนเดียวกันทั้งสองกรณีของแบบสอบ (ตารางที่ ๑๓) คือ วิธีที่ใช้แบบสอบรวมร้อยละ ๖๐ เป็นวิธีที่มีความเพียงพอมากเป็นอันดับ ๑ แบบสอบรวมขนาด ร้อยละ ๔๐ เป็นอันดับ ๒ และแบบสอบรวมขนาดร้อยละ ๒๐ เป็นอันดับ ๓ ข้อค้นพบดังกล่าวสอดคล้องกับสมมุติฐานที่กำหนดไว้ในข้อ ๒

๓. มิติที่เกี่ยวกับรูปแบบการเทียบมาตรา

รูปแบบการเทียบมาตรา ๓ รูปแบบ ที่ใช้ในการเทียบมาตราแบบสอบสองกรณีให้ผล
ต่างกัน ดังนี้

๓.๑ กรณีแบบสอบคัดเลือก ผลการวิเคราะห์กลุ่มสอบทานผลแสดงให้เห็นแบบแผนของค่าดัชนี C ซึ่งเป็นความคลาดเคลื่อนรวม สามารถจัดอันดับของความเพียงพออันเกิดจากรูปแบบการเทียบมาตราในแง่ความยาวแบบสอบรวมที่ใช้เป็นทำนองเดียวกันทั้งหมด แบบแผนของอันดับความเพียงพอ มีดังนี้ วิธีที่ใช้รูปแบบอิกวิเปอร์เซนไทล์เป็นอันดับที่หนึ่ง IRT เป็นอันดับที่สอง และเชิงเส้นตรงเป็นอันดับที่สาม ข้อค้นพบในส่วนนี้มีความสอดคล้องกับสมมุติฐานทางการวิจัยข้อ ๓ ในส่วนที่กล่าวไว้ว่า รูปแบบอิกวิเปอร์เซนไทล์เป็นรูปแบบที่มีความเพียงพอมากกว่ารูปแบบเชิงเส้นตรง ส่วนการเปรียบเทียบรูปแบบคู่อื่นไม่สามารถสรุปตามความแตกต่างอย่างมีนัยสำคัญ

๓.๒ กรณีแบบสอบวัดผลสัมฤทธิ์ ผลการจัดอันดับในทำนองเกี่ยวกับ ๓.๑ คือ จากความเพียงพอมากไปหาน้อยได้ดังนี้ รูปแบบเชิงเส้นตรงเป็นอันดับที่หนึ่ง อิกวิเปอร์เซนไทล์เป็นอันดับที่สอง และ IRT เป็นอันดับที่สาม การทดสอบความแตกต่างของค่าอันดับด้วยสถิติที่ไม่ใช่พาราเมตริกในทำนองเกี่ยวกับกรณีแบบสอบคัดเลือก พบว่า อันดับที่หนึ่ง และที่สาม มีความแตกต่างกันอย่างมีนัยสำคัญที่ .๐๕ ซึ่งหมายถึง วิธีที่ใช้รูปแบบเชิงเส้นตรง เป็นวิธีที่มีความเพียงพอมากกว่าวิธีที่ใช้รูปแบบ IRT ส่วนการเปรียบเทียบคู่อื่นไม่พบความแตกต่างอย่างมีนัยสำคัญ ข้อค้นพบนี้เป็น การปฏิเสธสมมุติฐานของการวิจัยที่กำหนดไว้ในข้อ ๔

การอภิปรายผล

๑. จากการประเมินผลความเพียงพอของวิธีการเทียบมาตราของทั้งสองกรณีของแบบสอบที่นำมาศึกษาด้วยเกณฑ์ของปีเตอร์สันและคณะ พบว่า ทุกวิธีที่ทำการศึกษากลับมาเป็นวิธีที่ให้ผลการแปลงคะแนนสมมูลที่มีความเพียงพออยู่ในระดับน่าพอใจ และน่าพอใจอย่างมาก เมื่อวิเคราะห์ในเชิงเหตุผล พบว่า การเทียบมาตราที่ดำเนินไปนั้น ใช้กลุ่มตัวอย่างสมมูลสองกลุ่มที่ได้จากการสุ่มจากประชากรเดียวกัน การออกแบบการวิจัยลักษณะนี้เป็น การควบคุมให้เกิดความคลาดคลึงของการแจกแจงของคะแนนที่ได้จากการสอบ ส่วนในค่านแบบสอบที่ใช้เทียบมาตรานั้นได้เตรียมจากแบบสอบฉบับเดียวกันที่มีความเที่ยงค่อนข้างสูง โดยทำให้เกิดเป็นแบบสอบเทียบมาตราที่คล้ายคลึงกันสองชุด ยิ่งกว่านั้นการออกแบบด้วยการใช้แบบสอบรวม (CIE) เป็นวิธีที่ช่วยรับความแตกต่างของข้อมูลคะแนนในส่วนที่อาจเกิดขึ้นจากแบบสอบเทียบมาตราที่มีความยากง่ายต่างกัน และความสามารถของผู้สอบสอง

กลุ่มแตกต่างกัน ดังนั้น เมื่อใช้รูปแบบอิคิวเปอร์ เซนโทลจึงให้ผลสอดคล้องกับนิยามที่ว่า การระบุคะแนนสมมูลของแบบสอบต่างชุด เมื่อกลุ่มผู้สอบมีการกระจายของความสามารถอยู่ในลักษณะเดียวกัน แล้วนั้น ให้ระบุจากตำแหน่ง เปอร์ เซนโทลเดียวกัน (Marco 1981) ส่วนผลการใช้รูปแบบเชิงเส้นตรง โดยทฤษฎีแล้วเป็นวิธีการประมาณคะแนนสมมูลตามนิยามของอิคิวเปอร์ เซนโทลที่ดีที่สุด เมื่อการแจกแจงของคะแนนของแบบสอบสองชุดมีลักษณะคล้ายคลึงกัน (Angoff 1984: 93) สำหรับรูปแบบ IRT ซึ่งมีข้อกำหนดที่แตกต่างไปบ้าง คือ แบบสอบที่ไรท์ของมีคุณสมบัติของการวัดความสามารถในมิติเดียวกัน และจำนวนข้อมูลที่ใช้ในการประมาณค่าต้องมีจำนวนไม่น้อยเกินไปนั้น การออกแบบการวิจัยครั้งนี้ไม่ได้มีข้อขัดแย้งกับข้อกำหนด จึงทำให้ได้ผลการเทียบที่น่าพอใจเช่นกัน กับรูปแบบการเทียบมาตราแบบดั้งเดิม

๒. การศึกษาเปรียบเทียบผลการใช้แบบสอบรวมทั้งที่ต่างกันด้วยการประเมินความแม่นยำของการเทียบมาตราจากการวิเคราะห์ความคลาดเคลื่อนมาตรฐานของการเทียบมาตรา (SEE) และคักสินด้วยค่าประสิทธิภาพสัมพัทธ์ (RE) ซึ่งพบว่า ส่วนใหญ่เป็นไปตามสมมุติฐานของการวิจัยที่กล่าวว่า แบบสอบรวมที่ยาวกว่าให้ประสิทธิภาพของการเทียบมาตราที่สูงกว่า ทั้งนี้ เหตุผลสำคัญในการสนับสนุนข้อค้นพบ คือ โขยบทบาทของแบบสอบรวมซึ่งทำหน้าที่ให้สารสนเทศ เพื่อใช้ในการปรับความแตกต่างระหว่างกลุ่มผู้สอบสองกลุ่ม และแบบสอบที่ต่างชุดกันในแต่ละรูปแบบของการเทียบมาตราในรูปแบบอิคิวเปอร์ เซนโทล สักส่วนของความถี่ของคะแนนในแบบสอบรวมที่ได้จากกลุ่มตัวอย่างรวมก็นำไปใช้เพื่อประมาณความถี่ของแต่ละชั้นของคะแนนของแบบสอบเทียบมาตราชุด X และชุด Y สำหรับในรูปแบบเชิงเส้นตรงผลการสอบของกลุ่ม α และ β ในส่วนของแบบสอบรวมก็นำไปใช้ในการประมาณคะแนนของแบบสอบเทียบมาตราชุด X และชุด Y ของประชากรโดยวิธีความเป็นไปได้สูงสุด (Maximum-likelihood) ส่วนรูปแบบ IRT ก็อาศัยแบบสอบรวมเพื่อการประมาณค่าพารามิเตอร์ของข้อสอบ ทำให้ค่าประมาณเป็นค่าที่มีหน่วยบนมาตราเดียวกัน จากผลการวิจัยได้แสดงให้เห็นว่า แบบสอบรวมที่ยาวกว่าทำให้ผลการเทียบมาตราที่มีความแม่นยำมากกว่า ในทำนองเดียวกัน เมื่อประเมินด้วยการใช้เกณฑ์ความคลาดเคลื่อนรวม (คชณิ c) จากเทคนิคการวิเคราะห์กลุ่มสอบตามผลิตภัณฑ์ข้อสรุปเช่นเดียวกัน ซึ่งเป็นการยืนยันสมมุติฐานการวิจัยข้อที่ ๒ ด้วย

แต่อย่างไรก็ตาม จากการวิเคราะห์ความคลาดเคลื่อนมาตรฐานของการเทียบมาตรา (SEE) และหาค่าประสิทธิภาพสัมพัทธ์ของแบบสอบ YV40-XV40 ของกรณีแบบสอบคัดเลือก พบว่า ผลการเทียบมาตราของแบบสอบคู่นี้ให้ประสิทธิภาพสูงใกล้เคียงกับการเทียบมาตราแบบสอบ YV60-XV60 ลักษณะเช่นนี้ไม่ปรากฏกับกรณีของแบบสอบวัดผลสัมฤทธิ์ จากการวิเคราะห์ในเชิงเหตุผล พบว่า เป็นผลกระทบจากคุณลักษณะของข้อสอบที่ประกอบขึ้นเป็นแบบสอบรวม เมื่อแบบสอบรวมขนาด ๑๔ ข้อ มีผลรวมความแปรปรวนรายข้อสูงกว่าผลรวมความแปรปรวนรายข้อของแบบสอบรวมขนาด ๒๐ ข้อ เมื่อเทียบตามสัดส่วนของความแปรปรวนของแต่ละแบบสอบรวม ความแม่นยำของการประมาณโดยอาศัยค่าสถิติจากแบบสอบรวมจึงมีความแม่นยำต่างกัน ซึ่งส่งผลถึงผลสุดท้ายของกระบวนการเทียบมาตราด้วย จากข้อค้นพบนี้เป็นการสนับสนุนผลการวิจัยของคลินและโคเลน (๑๘๘๕) ที่ว่า ความยาวของแบบสอบรวมไม่ใช่ปัจจัยเดียวในองค์ประกอบของแบบสอบรวมที่มีผลกระทบต่อ การเทียบมาตรา และเป็นข้อสันนิษฐานได้ว่า หากไม่มีการวิจัยในเรื่องนี้จึงยังไม่สามารถลงข้อสรุปที่เป็น การทั่วไปได้อย่างชัดเจน ข้อสรุปที่พบจึงต้องมีเงื่อนไขของบริบทประกอบเสมอ เช่น ผลการวิจัยของ ปีเตอร์สันและคณะ (๑๘๘๐, ๑๘๘๑) ของคลินและโคเลน (๑๘๘๕) เป็นต้น

๓. จากข้อค้นพบในมิติของรูปแบบการเทียบมาตราที่ใช้ในกรณีแบบสอบคัดเลือก ซึ่งให้ ผลสอดคล้องกับสมมุติฐานที่กำหนดไว้ในข้อที่ ๓ ที่ว่า รูปแบบอิกวิเปอร์ เซนไทล์ และรูปแบบ IRT เป็นวิธีที่มีความเพียงพอมากกว่ารูปแบบเชิงเส้นตรงนั้น เมื่อย้อนกลับไปพิจารณาสถานการณ์การสอบ คัดเลือกที่ได้เลือกมาใช้ นั้น มีปัจจัยแวดล้อมที่สนับสนุนผลการวิจัยสองประการ คือ ปัจจัยด้านแบบสอบ และปัจจัยด้านประชากรผู้สอบ ในด้านแบบสอบเป็นแบบสอบวัดความรู้ความสามารถทั่วไปในเชิงความ ถนัด มีพิสัยของระดับความยากกว้างพอสมควร (ค่าเฉลี่ยรายข้ออยู่ระหว่าง .๑๓ ถึง .๓๓) แบบสอบที่นำมาใช้มีความเที่ยงค่อนข้างสูง (.๘๘) ภายหลังจากการจัดเตรียมชุดของแบบสอบเทียบ มาตราเป็นคู่ ๆ (X และ Y) แต่ละคู่มิมีความเที่ยงใกล้เคียงกัน ในด้านประชากรผู้สอบเป็นประชากร ที่มีการกระจายของความสามารถกว้าง ข้อมูลกลุ่มตัวอย่างจากการออกแบบมีความดีในแต่ละชั้นของ คะแนนมีพิสัยกว้าง จึงทำให้การประมาณความดีประชากรตามรูปแบบของอิกวิเปอร์ เซนไทล์ได้ผล แม่นยำ ดังนั้น การเทียบมาตราในกรณีแบบสอบคัดเลือกที่อยู่ภายใต้บริบททั้งที่กล่าวมา รูปแบบอิกวิ- เปรอร์ เซนไทล์ให้ความเพียงพอในการแปลงคะแนนสมมูลมากที่สุด เมื่อใช้แบบสอบรวมขนาดเดียวกัน

แต่อย่างไรก็ตาม ถ้าพิจารณาปริมาณค่าดัชนี c แทนการจัดอันดับแล้ว พบว่า รูปแบบ IRT และแม่แบบรูปแบบเชิงเส้นตรงก็สามารถให้ความเพียงพอในการแปลงคะแนนสมมูลในระบับนำพอใจอย่างมากได้เช่นเดียวกับรูปแบบอควิเปอร์เซนไทล์ เมื่อใช้แบบสอบถามร้อยละ ๒๐ หรือ ๒๕ ข้อ ดังนั้น อาจกล่าวได้ว่า เมื่อการเทียบมาตราสามารถออกแบบใช้แบบสอบถามภายในขนาดยาวมากพอ เช่น ร้อยละ ๒๐ ของแบบสอบถามมากรากกับกลุ่มตัวอย่างสมมูลสองกลุ่มแล้ว การใช้รูปแบบใดในการเทียบไม่ใช่สาระสำคัญมากนักในค่านความเพียงพอ แต่อาจไปพิจารณาในเชิงปฏิบัติถึงปัจจัยประกอบที่นอกเหนือจากที่กล่าวถึง เช่น ความพร้อมในการดำเนินการด้านต่าง ๆ มากกว่า

๔. กรณีแบบสอบถามวัดผลสัมฤทธิ์ มีปัจจัยต่าง ๆ แตกต่างจากกรณีแบบสอบถามคัดเลือก ในค่านแบบสอบถามมีขอบเขตของเนื้อหาที่จำกัดภายในกรอบของหลักสูตร และในค่านผู้สอบประชากรมีความสามารถขั้นต่ำค่อนข้างชัดเจน นอกจากนี้ผู้สอบยังได้ผ่านกระบวนการเรียนการสอนในวิชาที่สอบแล้ว จึงทำให้คะแนนที่ได้มีพิสัยที่แคบกว่ากรณีแบบสอบถามคัดเลือก แต่เนื่องจากการออกแบบเทียบมาตราด้วยกลุ่มผู้สอบสมมูลสองกลุ่มที่ได้จากการสุ่ม จึงพบว่า การแจกแจงของคะแนนทั้งสองกลุ่มคล้ายคลึงกัน จากปัจจัยที่กล่าวมานี้เป็นปัจจัยที่ผู้วิจัยได้กำหนดและออกแบบไว้ และจากผลการวิเคราะห์พบความสัมพันธ์กับสมมุติฐานส่วนหนึ่งของข้อที่ ๔ ที่ว่า วิธีการเทียบมาตราที่ใช้รูปแบบเชิงเส้นตรงเป็นวิธีที่มีความเพียงพอมากกว่ารูปแบบอควิเปอร์เซนไทล์

สำหรับสมมุติฐานอีกส่วนหนึ่งของสมมุติฐานข้อที่ ๔ ที่กล่าวว่า รูปแบบ IRT จะให้ความเพียงพอแก่วิธีการเทียบมาตรามากกว่ารูปแบบอควิเปอร์เซนไทล์นั้น ไม่สามารถให้ผลการสนับสนุนจากการวิจัยครั้งนี้ ซึ่งได้อภิปรายในหัวข้อที่ ๕ ต่อไป

๕. การเปรียบเทียบผลการเทียบมาตราต่างรูปแบบ พบว่า วิธีที่ใช้รูปแบบ IRT ไม่ได้เป็นวิธีที่น่าพอใจมากที่สุด ในกรณีแบบสอบถามคัดเลือกได้เป็นอันดับสองรองจากรูปแบบอควิเปอร์เซนไทล์ และในกรณีแบบสอบถามวัดผลสัมฤทธิ์ได้เป็นอันดับสาม ซึ่งเป็นอันดับสุดท้าย ถ้าพิจารณาเทียบกับผลงานวิจัยในอดีตซึ่งพบว่า เมื่อแบบสอบถามทำการเทียบมาตราที่มีความยากต่างกันแล้ว วิธีที่ใช้รูปแบบ IRT จะให้ผลที่น่าพอใจที่สุด (Slinde and Linn 1977; Petersen and others 1981; Kolen 1981; Cook and others 1981) และพบว่า เมื่อการวิเคราะห์กับกลุ่มตัวอย่าง

ขนาดเล็กลงไป มีผลทำให้วิธีการเทียบตามรูปแบบ IRT ให้ผลไม่น่าพอใจทั้ง ๆ ที่แบบสอบมีลักษณะคล้ายคลึงกัน (Kolen and Khitney 1982) สำหรับการวิจัยครั้งนี้ ได้ใช้กลุ่มตัวอย่างขนาดใหญ่ทั้งสองกรณีของแบบสอบ และลักษณะแบบสอบที่เทียบมาครามีความแตกต่างกันบ้างเล็กน้อยเท่านั้น แต่ปรากฏผลดังที่กล่าวสรุปไปนั้น อาจสืบเนื่องจากสาเหตุ ๔ ประการ ดังนี้

ก. ปัญหาการวิเคราะห์ค่าประมาณพารามิเตอร์ข้อสอบที่ใช้โปรแกรมคอมพิวเตอร์ LOGIST ที่สถาบันคอมพิวเตอร์จุฬาลงกรณ์มหาวิทยาลัย ซึ่งมีข้อจำกัดในการใช้โปรแกรม ทั้งนี้เนื่องจากกระบวนการคำนวณค่าเป็นของใช้คำสั่งที่กำหนดไว้ในโปรแกรม (default) เมื่อการวิเคราะห์ประสบปัญหาเนื่องจากการประมาณค่าแต่ละขั้นตอนไม่ไต่บรรลุตามเกณฑ์ จำเป็นต้องใช้วิธีการปรับปรุงค่าภายหลังการทำงานของโปรแกรมแต่ละครั้ง แล้วจึงส่งงานใหม่ การวิเคราะห์จะเริ่มต้นใหม่ทุกครั้ง ทำให้สิ้นเปลืองเวลา และเป็นการกักตุนกำลังของการประมาณ ค่าประมาณที่ได้เมื่อสิ้นสุดการทำงานจึงอาจจะยังมีความแปรปรวนอยู่ ความจริงโดยลักษณะของโปรแกรม LOGIST สามารถรับคำสั่งการทำงานอย่างต่อเนื่องโดยการเก็บผลครั้งก่อนลงในหน่วยความจำที่เตรียมไว้ และใช้เป็นข้อมูลของการทำงานครั้งต่อไป หลาย ๆ ครั้งจนกว่าได้ค่าประมาณที่ไม่แปรเปลี่ยน

ข. การเทียบมาตราตามรูปแบบ IRT สามพารามิเตอร์ วิธีเทียบคะแนนจริงเป็นวิธีเทียบในแนวคิดของแบบสอบมิติเดียวสองชุดที่วัดความสามารถเดียวกัน ซึ่งให้คุณสมบัติของผลการเทียบมาตรา ๓ ประการ คือ ความยุติธรรม ความไม่แปรเปลี่ยนเมื่อเปลี่ยนกลุ่ม และความสมมาตร (Lord 1981: 199) ซึ่งคุณสมบัติทั้งกล่าวนี้จะเกิดได้เฉพาะใช้กับการเทียบคะแนนจริง แต่ในทางปฏิบัติเราไม่สามารถรู้คะแนนจริงของผู้สอบทุกรายบุคคลในการสอบ เมื่อได้ตารางเทียบคะแนนสมมูลจากกระบวนการเทียบมาตราแล้ว จะนำไปใช้กับคะแนนสังเกตของแต่ละบุคคล (Lord 1982: 147) จากที่กล่าวมานี้ อาจเป็นสาเหตุหนึ่งที่ทำให้การวิเคราะห์กลุ่มสอบทานผลได้คะแนนแปลง x^* ที่อ่านค่าจาก y แตกต่างจาก x ที่รายบุคคลได้จริง และสาเหตุนี้อาจได้รับผลกระทบจากการประมาณค่าพารามิเตอร์ทั้งที่อธิบายในข้อ ก. มาก่อนแล้ว เมื่อใช้ดัชนีเปรียบเทียบความแตกต่าง c ทักสินผลสุดท้ายของกระบวนการเทียบมาตราแล้ว จึงให้ข้อสรุปว่า วิธีที่ใช้รูปแบบ IRT กับสถานการณ์แบบสอบที่ศึกษามีความเพียงพอน้อยกว่ารูปแบบอิกวิเปอร์ เชน ไทล์ และ เชิง เส้นตรง

ค. ปัญหาการเทียบมาตรฐานคะแนนส่วนที่อยู่ต่ำกว่าระดับการเคหา เป็นปัญหาในเชิงปฏิบัติ เพื่อให้สามารถเทียบคะแนนของแบบสอบโต้ครบถ้วนตลอดชีพของคะแนน ลอริกโต้แนะนำให้ใช้วิธีปรับขยายเชิงเส้นตรง (Linear interpolation method) (Lord 1981: 210-211) วิธีนี้ไ้มีผู้นำไปใช้มาก เช่น โกลเดน (๑๙๘๑) และคูกู กัมบาร์ และไอเนอร์ (๑๙๘๑) เป็นต้น แต่ยังไม่ไ้มีการศึกษาถึงความเหมาะสม และผลกระทบอย่างจริงจัง และเป็นข้ออภิปรายของนักวิจัยที่ไ้แนวปฏิบัติที่ไ้ตั้งข้อสันนิษฐานว่า ปัญหา IRT ที่มีส่วนหนึ่งอาจเกิดจากสาเหตุนี้ไ้

ข้อเสนอแนะ

๑. ข้อเสนอแนะเกี่ยวกับการประยุกต์

เป็นความจริงที่การวิจัยในอดีตเกือบทุกงานไ้สรุปว่า ไม่สามารถบอกไ้ว่าวิธีการใด หรือรูปแบบใดที่ไ้ทำให้การเทียบมาตรฐานไ้ผลดีที่สุด จำเป็นต้องมีเงื่อนไขซึ่ง เป็นตัวแปรที่อยู่เหนือการควบคุมเป็นตัวประกอบในคำสรุปเสมอ สำหรับผลการวิจัยครั้งนี้ก็เช่นกัน ข้อสรุปที่ไ้ไ้จะตั้งอยู่ภายใต้เงื่อนไข และสถานการณ์ที่มาพร้อมกับข้อมูล ทั้งนี้ การเลือกวิธีเทียบมาตรฐานไปประยุกต์จึงต้องพิจารณาประเด็นต่าง ๆ โดยมีข้อค้นพบจากงานวิจัยต่าง ๆ ช่วยในการตัดสินใจสำหรับข้อเสนอแนะเพื่อนำไปประยุกต์ไ้ เป็นข้อเสนอแนะภายใต้การออกแบบโดยมีแบบสอบร่วม (CIE) เมื่อแบบสอบที่ไ้ต้องการเทียบมาตรฐานไ้ความคล้ายคลึงกันในด้านเนื้อหา และระดับความยาก และกลุ่มตัวอย่างผู้สอบ ๒ กลุ่ม ไ้จากการสุ่มอย่างง่ายจากประชากรเดียวกัน หรือมีการแจกแจงของคะแนนที่คล้ายคลึงกันทั้งสองกลุ่ม

๑.๑ ในสภาพการณ์การสอบคัดเลือกที่มีวัตถุประสงค์เพื่อจำแนกคนตามระดับความสามารถเชิงความถนัดทางวิชาการ เพื่อให้เข้ารับการฝึกงานในอาชีพ หรือเข้าทำงาน โดยมีแบบสอบวัดความถนัด หรือความรู้ความสามารถทั่วไปที่มีความเที่ยงสูงใกล้เคียงกัน มีจำนวนผู้สอบมาก และมีพิสัยของการกระจายคะแนนกว้าง คือ มีความถี่กระจายตั้งแต่คะแนนต่ำสุดถึงสูงสุดแล้วการใช้รูปแบบอิกวิเบอร์ เช่น ไ้ไลในการเทียบมาตรฐานจะให้ความแม่นยำต่อการแปลงคะแนนสมมูลสำหรับคำถามที่เกี่ยวกับความยาวของแบบสอบร่วมที่ไ้ใช้ในสถานการณ์เช่นนี้ แบบสอบร่วมขนาด ๒๐ ข้อ

ที่ต่างกันของแบบสอบใหม่และเก่า เทคนิควิธี เช่นนี้สามารถนำเข้ามาใช้ในโครงการทดสอบที่เป็นโครงการต่อเนื่อง เช่น โครงการทดสอบเพื่อคัดเลือกบุคคลเข้าทำงาน โครงการทดสอบเพื่อวัดผลสัมฤทธิ์ทางการเรียนตามหลักสูตรอย่างมีมาตรฐานของการศึกษา และโครงการทดสอบเพื่อปรับตำแหน่งในการปฏิบัติงาน เป็นต้น ส่วนที่ที่จะเกิดขึ้นจากการนำเทคนิควิธี เทียบมาคราใน แต่ละโครงการ มีดังต่อไปนี้

๒.๑ โครงการทดสอบเพื่อคัดเลือกบุคคลเข้าทำงาน

โดยปกติของโครงการทดสอบคัดเลือกบุคคลเข้าทำงานในส่วนของ การสอบข้อเขียนมักประกอบด้วยหมวดวิชาความรู้ทั่วไป และหมวดวิชาเฉพาะ หมวดวิชาความรู้ทั่วไป ซึ่งหมายถึง ความสามารถในการเชิงความถนัดทางวิชาการทั่ว ๆ ไป ซึ่งเป็นวิชาที่กำหนดให้ผู้สมัครเข้าทำงานในสาขาต่างต้องผ่านการทดสอบ เช่น การสมัครเข้ารับการศึกษาเป็นข้าราชการที่จัดดำเนินการโดยคณะกรรมการข้าราชการพลเรือน (ก.พ.) แต่ละครั้งของการจัดสอบแข่งขัน พบว่า ในระดับวิชาชีพ หรือปริญญาตรี ผู้สมัครจำนวนนับหมื่น การสอบแต่ละครั้งได้นำผลการสอบมาเรียงลำดับ และพิจารณาบุคคลตามลำดับคะแนน ผู้ที่มีลำดับเกินกว่าจำนวนที่รับได้ จะมีชื่อในบัญชีสำรองจำนวนหนึ่ง บัญชีนี้จะถูกยกเลิกเมื่อมีการจัดสอบครั้งใหม่ ปัญหาที่พบ คือ การจัดสอบให้กับคนจำนวนมาก ๆ ในเวลาที่กำหนดทุกครั้ง งานด้านเตรียมสถานที่ ผู้ควบคุมการสอบ ตลอดจนการจัดเตรียมแบบสอบ และอื่น ๆ นับตั้งแต่การรับสมัครจนถึงการประกาศผลการสอบ เป็นงานหนักมาก นอกจากนี้ตามระเบียบที่กำหนดให้มีการยกเลิกผลการสอบเมื่อมีการสอบครั้งใหม่ ทำให้ผู้สมัครที่เคยสอบและที่ต่อการสมัครอีกต้องทำการสอบซ้ำ เป็นการสิ้นเปลืองทั้งเวลา และงบประมาณส่วนตัว และเงินส่วนของรัฐ ทางแก้อย่างหนึ่ง คือ ๑) จัดการสอบให้บ่อยขึ้น และกระจายหลายสนามสอบ เพื่อเฉลี่ยคนสอบในแต่ละครั้งให้น้อยลง ซึ่งให้ความสะดวกในการเตรียมสถานที่สอบ และแบบสอบ ๒) กำหนดให้ใช้ระเบียบวิธีเทียบมาคราสำหรับแปลงคะแนนดิบจากการสอบไปสู่คะแนนมาคราที่สามารถเปรียบเทียบกันได้กับครั้งก่อนหรือครั้งต่อไป โดยการออกแบบการเทียบมาคราอย่างมีแบบสอบร่วม (CIE) ๓) ออกใบรับรองผลการสอบเพื่อให้นำไปใช้เป็นหลักฐานในการสมัครเข้าทำงานตามกำหนดอายุของการใช้ผลการสอบ อาจเป็น ๓ ปี หรือ ๕ ปี เป็นต้น

๒.๒ โครงการทดสอบเพื่อวัดผลสัมฤทธิ์ทางการเรียนตามหลักสูตร

หลักสูตรเป็นโครงการต่อเนื่องระยะยาว โดยมีวัตถุประสงค์ให้ผู้เข้ารับการอบรม หรือผ่านกระบวนการเรียนได้สัมฤทธิ์ผลตามที่ได้กำหนดไว้ในหลักสูตร เช่น หลักสูตรมัธยมศึกษา หลักสูตรระดับอุดมศึกษา เป็นต้น การประเมินผลสัมฤทธิ์ของหลักสูตรที่สำคัญ คือ การสอบวัดผลสัมฤทธิ์ทางการเรียนของนักเรียน แบบสอบที่ใช้กับนักเรียนหรือนักศึกษาแต่ละปีการศึกษา มักมีการเปลี่ยนแปลง การประเมินโดยใช้คะแนนดิบของแบบสอบแต่ละครั้งจึงไม่สามารถนำมาเปรียบเทียบข้ามรุ่นได้อย่างมีความหมายและถูกต้อง ดังนั้น หากนำเทคนิคการเทียบมาตรฐานมีแบบสอบร่วม (CIE) มาใช้กับการสอบประจำปีการศึกษาในตลอดปลายภาคทุกปี ผลการสอบสามารถช่วยให้ผู้ประเมิน และพัฒนาหลักสูตรอย่างมีความมั่นใจในการเปรียบเทียบผลกับคนอื่น ๆ ในหลักสูตรเดียวกันได้ สามารถควบคุมมาตรฐานในการตัดสินได้ - ทดอย่างคง เส้นคงวาลอดโครงการหลักสูตรได้

๒.๓ โครงการทดสอบเพื่อปรับตำแหน่งในระบบการทำงาน

ตามระเบียบการปฏิบัติงานราชการในบางสายบางระดับได้กำหนดให้มีการสอบเพื่อปรับเลื่อนตำแหน่งความรับผิดชอบงานให้สูงขึ้น เมื่อบุคคลนั้นมีคุณสมบัติถึงระดับหนึ่ง เช่น การสอบเพื่อเลื่อนชั้นในสายงานวิชาการจากตำแหน่งที่ ๔ เป็นที่ ๕ เป็นต้น การสอบเป็นเกณฑ์หนึ่งที่นำมาประเมินความเหมาะสมในก้านความรู้ความสามารถของผู้ขอตำแหน่ง ผลการสอบในแต่ละครั้งให้นำมาจัดอันดับ ผู้ที่ได้คะแนนสูงเป็นผู้ได้รับการพิจารณา ก่อน ผลการสอบของผู้ที่ยังไม่ได้รับการปรับตำแหน่งจะถูกยกเลิกโดยอัตโนมัติ เมื่อมีการสอบครั้งใหม่ และในทำนองเดียวกันผู้ที่มิคุณสมบัติแล้วแต่ยังไม่มีการสอบ ผู้นั้นจะต้องรอจนกว่าจะถึงเวลาสอบ วิธีปฏิบัตินี้หากได้นำเทคนิคการเทียบมาตรฐานมีแบบสอบร่วมเขาไปใช้ จะช่วยเสริมค่านคุณค่าที่เป็นมาตรฐานจากการสอบ และให้โอกาสของแต่ละบุคคลเพื่อได้รับการประเมินในส่วนของความรู้ ความสามารถอย่างยุติธรรมมากขึ้น คือ ไม่มีผลกระทบจากข้อสอบที่ง่ายกว่าหรือยากกว่าครั้งก่อน นอกจากนี้การจัดการสอบยังสามารถจัดได้บ่อย ๆ โดยไม่ต้องรอให้เสียโอกาสส่วนหนึ่ง

๓. ข้อเสนอแนะในการวิจัยต่อ

๓.๑ จากตารางออกแบบการศึกษาวิธีการเทียบมาตรฐานที่มีแบบสอบรวม (CIE) ได้ศึกษาเพียงตัวแปรในมิติของรูปแบบสามรูปแบบ และความยาวแบบสอบรวม ๓ ขนาด ในสถานการณ์การทดสอบเพื่อการคัดเลือกบุคคลกับวัตถุประสงค์ทางการเรียน ข้อค้นพบยังมีขีดความสามารถในการอธิบายจำกัดอยู่มาก จึงน่าจะได้มีการทำวิจัยในตัวแปรอื่น ๆ เช่น ศึกษาการเทียบมาตรฐานที่มีแบบสอบรวมชนิดภายนอก (External anchor test) ลักษณะศาสตร์ของแบบสอบรวม ความคล้ายคลึงของแบบสอบเทียบมาตรฐานกับแบบสอบรวม และแม้กระทั่งความยาวของแบบสอบรวมเมื่อแบบสอบเทียบมาตรฐานมีความยาวมากกว่า ๓๕ ข้อ ดังเช่นที่ใช้ในการวิจัยครั้งนี้

๓.๒ ปัจจุบันได้มีรูปแบบการเทียบมาตรฐานเกิดขึ้นอีกหลายรูปแบบ น่าจะได้มีการศึกษาค้นคว้าต่อ เพื่อเป็นสารสนเทศสำหรับการตัดสินใจเลือกใช้ในโครงการทดสอบใดกว้างขวางยิ่งขึ้น เช่น วิธี Pre-equating ซึ่งมีการนำไปใช้ในโครงการทดสอบ TOEFL การศึกษา Pre-equating เป็นวิธีการเทียบมาตรฐานโดยอาศัยทฤษฎีการตอบสนองในการหาค่าประชากรข้อสอบไว้ก่อน เมื่อสร้างชุดแบบสอบแล้วจะคำนวณสารสนเทศต่าง ๆ จากค่าประชากรรายข้อที่หาได้แล้ว สร้างตารางแปลงคะแนนสมมูลจากชุดหนึ่งไปยังอีกชุดหนึ่ง วิธีการนี้ผู้ออกแบบชุดของข้อสอบควรมีคลังข้อสอบที่เคยผ่านการทดลองใช้แล้ว การจัดชุดแบบสอบแต่ละครั้งแตกต่างกันเฉพาะในองค์ประกอบของรายข้อ วิธี Pre-equating นี้ นักวิจัยสามารถสร้างตารางแปลงคะแนนสมมูลไว้ก่อนเสมอ จึงเป็นการสะดวกสำหรับ และสามารถเสนอคะแนนผลการสอบได้อย่างรวดเร็ว

๓.๓ การศึกษาโดยเลือกสถานการณ์ตามความเป็นจริงมาทำการศึกษาสองสถานการณ์ดังที่เสนอมานี้ เป็นเพียงการบรรยายตามสภาพแต่ไม่สามารถควบคุมตัวแปรแทรกซ้อน ดังนั้น น่าจะศึกษาโดยการควบคุมตัวแปรเพื่อคุณผลให้ชัดเจน เช่น ตัวแปรเกี่ยวกับการแจกแจงของคะแนนประชากรที่มีลักษณะต่าง ๆ กัน ความเที่ยงของแบบสอบเทียบมาตรฐาน และแบบสอบรวม เป็นต้น