

บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

ในบทนี้จะกล่าวถึงเอกสารและงานวิจัยที่เกี่ยวข้องต่างๆ ที่เป็นพื้นฐานของการวิจัยครั้งนี้ โดยผู้วิจัยได้นำเสนอตามหัวข้อดังต่อไปนี้

1. ระบบการรับบุคคลเข้าศึกษา
2. การเชื่อมโยงคะแนนระหว่างแบบสอบ
3. วิธีการเปรียบเทียบผลการเรียน
4. งานวิจัยที่เกี่ยวข้อง

ระบบการรับบุคคลเข้าศึกษา

การคัดเลือกบุคคลเข้าศึกษาต่อในระดับอุดมศึกษาของไทยมีพัฒนาการมายาวนาน เพื่อคัดเลือกนักเรียนที่มีความรู้ความสามารถเหมาะสมกับสาขาวิชาต่างๆ เข้าศึกษาต่อในมหาวิทยาลัยได้ เปิดดำเนินการจัดการเรียนการสอน ซึ่งจะทำให้มหาวิทยาลัยสามารถพัฒนาผู้เรียนในสาขาวิชานั้น ให้เป็นบุคคลที่มีคุณภาพหรือมีสมรรถภาพสูงสุด และช่วยลดปัญหาการออกกลางคันอันเนื่องมาจากการรับบุคคลที่มีคุณสมบัติไม่เหมาะสมเข้าศึกษา ระบบการรับเข้าศึกษาของไทยนั้น มีวิวัฒนาการมาประมาณกึ่งพุทธวรรษ ประกอบกับนโยบายขยายการศึกษาขั้นพื้นฐานของประเทศและการประกาศใช้พระราชบัญญัติการศึกษาแห่งชาติ พ.ศ. 2542 ที่ให้รัฐขยายการศึกษาขั้นพื้นฐานไม่น้อยกว่า 12 ปี อย่างทั่วถึงและมีคุณภาพ โดยไม่เก็บค่าเล่าเรียน ส่งผลให้จำนวนผู้ที่ต้องการศึกษาต่อในระดับ อุดมศึกษาเพิ่มมากขึ้น กระบวนการคัดเลือกบุคคลเข้าศึกษาต่อจึงมีความสำคัญ พัฒนาการของ ระบบการรับเข้าศึกษาพอสรุปได้ดังนี้

1. ความเป็นมาของการสอบคัดเลือกรวม

การรับบุคคลเข้าศึกษาในสถาบันอุดมศึกษา เป็นหน้าที่และความรับผิดชอบโดยตรงของแต่ละสถาบัน โดยสถาบันอุดมศึกษาแต่ละแห่งสามารถที่จะกำหนดหลักเกณฑ์และวิธีการในการคัดเลือกที่เหมาะสมได้อย่างอิสระ แต่การที่แต่ละสถาบันเป็นผู้ดำเนินการเองทั้งหมดนั้น พบว่า มีปัญหาเหมือนกันอย่างน้อย 3 ประการ คือ (CUAS:http://www.cuas.or.th/download_pdf/cuasinfo.pdf)

1. เกิดปัญหาที่ว่าง ที่แต่ละสถาบันไม่สามารถรับนักศึกษาได้เต็มตามเป้าหมายที่กำหนดไว้ เนื่องจากการสละสิทธิ์ของผู้สอบ

2. ต้องประกาศเรียกตัวผู้สอบได้เพื่อทดแทนผู้ที่สละสิทธิ์หลายครั้ง ทำให้การเปิดเรียนของสถาบันต้องล่าช้ากว่ากำหนด

3. ความซ้ำซ้อนในการดำเนินงานของแต่ละสถาบัน ตั้งแต่การรับสมัคร การจัดสอบ เป็นต้น ทำให้สิ้นเปลืองทรัพยากรทั้งของสถาบันอุดมศึกษาและของผู้สมัครสอบเอง

ปัญหาที่เกิดขึ้นดังกล่าว ทำให้แนวคิดในการดำเนินการสอบร่วมกันของสถาบันอุดมศึกษาต่างๆ ได้พัฒนาขึ้นตามลำดับ โดยการสอบคัดเลือกกรมเพื่อรับบุคคลเข้าศึกษาต่อในมหาวิทยาลัยของประเทศไทย ได้เริ่มดำเนินการมาตั้งแต่ปีการศึกษา 2504 ซึ่งมีมหาวิทยาลัยเกษตรศาสตร์ และมหาวิทยาลัยแพทยศาสตร์ (มหาวิทยาลัยมหิดล ในปัจจุบัน) จัดสอบร่วมกันเป็นครั้งแรก โดยมีสำนักงานสภาการศึกษาแห่งชาติเป็นผู้ประสานงาน ในปีการศึกษา 2505 มหาวิทยาลัยทั้ง 5 แห่งที่มีอยู่ในขณะนั้น คือ มหาวิทยาลัยเกษตรศาสตร์ มหาวิทยาลัยแพทยศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย มหาวิทยาลัยธรรมศาสตร์ และมหาวิทยาลัยศิลปากร ได้ดำเนินการจัดสอบร่วมกันต่อมาในปีการศึกษา 2509 คณะรัฐมนตรีได้มีมติตามข้อเสนอของจุฬาลงกรณ์มหาวิทยาลัย ให้มหาวิทยาลัยต่างๆ ดำเนินการสอบคัดเลือกเองโดยแยกกันสอบ และให้สำนักงานสภาการศึกษาแห่งชาติกำหนดวันสอบเพื่อไม่ให้สอบในวันเดียวกัน ผู้สมัครสอบมีสิทธิที่จะสมัครสอบคัดเลือกในสถาบันการศึกษาที่ตนเองต้องการได้หลายสถาบัน ก่อให้เกิดปัญหามากมายกับผู้สมัครเข้าศึกษาและสถาบันการศึกษา เพราะผู้สมัครจำนวนมากสอบได้ในหลายสถาบัน แต่สามารถเลือกลงทะเบียนเข้าเรียนเพียงสถาบันเดียว ทำให้เกิดที่ว่างจากการสละสิทธิ์ และต้องมีการเรียกตัวผู้สมัครสอบเพิ่มหลายครั้ง ทำให้เวลาในการเปิดภาคเรียนล่าช้ากว่ากำหนด ในปีการศึกษา 2510 คณะรัฐมนตรีเห็นชอบตามข้อเสนอของสำนักงานสภาการศึกษาแห่งชาติ ที่เสนอให้ดำเนินการการรับบุคคลเข้าศึกษาต่อในมหาวิทยาลัยโดยการจัดสอบรวมอีกครั้ง โดยให้สำนักงานสภาการศึกษาแห่งชาติเป็นผู้ประสานงานและเป็นผู้ดำเนินการ ทั้งนี้มีการวางโครงสร้างการดำเนินการอย่างเป็นระบบมีการแต่งตั้งอนุกรรมการหลายคณะเพื่อดูแลรับผิดชอบการดำเนินงานในแต่ละด้านเป็นการเฉพาะ

ต่อมาเมื่อมีการจัดตั้งทบวงมหาวิทยาลัยของรัฐขึ้นเมื่อวันที่ 29 กันยายน พ.ศ. 2515 การประสานงานการดำเนินการสอบคัดเลือกเข้าสถาบันอุดมศึกษาจึงอยู่ในความรับผิดชอบของทบวงมหาวิทยาลัย โดยทบวงมหาวิทยาลัยได้ดำเนินการตั้งแต่ปีการศึกษา 2516 เป็นต้นมา โดยกำหนดแนวทางในการรับบุคคลเข้าศึกษาต่อไว้ 3 แนวทางคือ

ก. การรับตรง หรือการรับนักศึกษาตามโควตาของมหาวิทยาลัยส่วนภูมิภาค เป็นการเปิดโอกาสให้นักเรียนในส่วนภูมิภาคเข้าสู่ระบบอุดมศึกษามากขึ้น

ข. การรับนักศึกษาตามโครงการพิเศษต่างๆ ที่มหาวิทยาลัยกำหนดขึ้น เช่น โครงการพิเศษ สำหรับผู้ที่มีความสามารถทางด้านกีฬา ด้านวิชาการ และด้านศิลปะ เป็นต้น

ค. การสอบคัดเลือกรวม ที่ทบวงมหาวิทยาลัยเป็นผู้ดำเนินการ ซึ่งมีรูปแบบการบริหาร และการดำเนินงานพอสรุปได้ดังนี้

1. การบริหารงานการสอบคัดเลือกจัดทำในรูปของคณะกรรมการ โดยมี คณะอนุกรรมการประสานงานการสอบคัดเลือกบุคคลเข้าศึกษาในสถาบันอุดมศึกษาของรัฐ ทำหน้าที่กำหนดนโยบายเกี่ยวกับการสอบคัดเลือกรวม โดยมีผู้แทนมหาวิทยาลัยที่เกี่ยวข้องร่วมเป็น อนุกรรมการ

2. การดำเนินการสอบคัดเลือก คณะอนุกรรมการประสานงานการสอบคัดเลือก จะแต่งตั้งคณะกรรมการการสอบคัดเลือกของทบวงมหาวิทยาลัยขึ้น เพื่อดำเนินการสอบคัดเลือก ในแต่ละปี โดยมีประธานอนุกรรมการฝ่ายต่างๆร่วมเป็นกรรมการ

จากการดำเนินการสอบคัดเลือกรวมระยะหนึ่งพบว่า มีปัญหาที่สำคัญที่จะต้องนำมา พิจารณาแก้ไขดังนี้

1. มีผลกระทบในทางลบต่อการจัดการศึกษาระดับมัธยมศึกษาตอนปลาย
2. มีนักศึกษาที่ผ่านการคัดเลือกจำนวนหนึ่งมีความรู้ความสามารถไม่สอดคล้องกับ สาขาวิชาที่เรียน
3. จำนวนผู้สมัครและจำนวนสถาบันที่ร่วมรับนักศึกษามีแนวโน้มสูงขึ้น
4. สร้างความเครียดแก่ผู้สอบและผู้ปกครอง
5. มีข้อจำกัดด้านเวลา สามารถดำเนินการได้เฉพาะในช่วงปิดภาคการศึกษา

จากปัญหาดังกล่าวทบวงมหาวิทยาลัยจึงได้แต่งตั้งคณะกรรมการพิจารณาปรับปรุงการสอบ คัดเลือกขึ้นเมื่อปี พ.ศ. 2535 เพื่อทำหน้าที่ศึกษาและเสนอแนะแนวทางในการปรับปรุงระบบการ สอบคัดเลือก โดยมีผู้แทนจากกรมวิชาการ ผู้แทนจากกรมสามัญศึกษา กระทรวงศึกษาธิการ ผู้แทน จากสำนักงานคณะกรรมการการศึกษาแห่งชาติ และผู้ทรงคุณวุฒิจากมหาวิทยาลัยต่างๆ ร่วมเป็น อนุกรรมการ

คณะอนุกรรมการพิจารณาปรับปรุงระบบการสอบคัดเลือก ได้เสนอรูปแบบการคัดเลือก บุคคลเข้าศึกษาในสถาบันอุดมศึกษาระบบใหม่ ซึ่งทบวงมหาวิทยาลัยได้ให้ความเห็นชอบและได้ ประกาศให้นำรูปแบบและวิธีการคัดเลือกระบบใหม่มาใช้ตั้งแต่ปีการศึกษา 2542 เป็นต้นไป

2. ระบบการรับเข้าศึกษาระบบใหม่

มหาวิทยาลัยทั้งหลายได้ตระหนักเป็นอย่างดีว่า ระบบการสอบคัดเลือกใหม่ที่ใช้อยู่นี้ แม้จะได้รับการปรับปรุงและมีข้อดีกว่าระบบดั้งเดิมที่เคยใช้อยู่ในอดีตเป็นอย่างมาก แต่ก็ยังมีข้อเสียอยู่ไม่น้อย ที่ประชุมอธิการบดีแห่งประเทศไทย (ทปอ.) มีความประสงค์ที่จะพัฒนากระบวนการคัดเลือกให้ดีขึ้นไปอย่างต่อเนื่องจึงได้จัดให้มี “คณะทำงานว่าด้วยการรับบุคคลเข้าศึกษาและการวัดผล” (Admissions and Assessment Forum) ซึ่งประกอบไปด้วยผู้เชี่ยวชาญด้านการวัดผลจากมหาวิทยาลัยต่างๆ ดำเนินการติดตามการสอบคัดเลือกอย่างต่อเนื่อง เพื่อประเมินผลดีผลเสีย พร้อมทั้งหาแนวทางในการพัฒนาระบบการคัดเลือกให้ดียิ่งขึ้นต่อไป

จากผลของการศึกษาและเฝ้าระวังของคณะทำงานดังกล่าว ที่ประชุมอธิการบดีแห่งประเทศไทยได้เสนอต่อทบวงมหาวิทยาลัยตามหนังสือ ที่ ทปอ. 44/147 ลงวันที่ 19 เมษายน 2544 เกี่ยวกับการปรับปรุงการคัดเลือกผู้เข้าศึกษาต่อในมหาวิทยาลัยในระบบสอบรวม โดยยึดหลักการให้เพิ่มการใช้ผลการเรียนระดับมัธยมศึกษาตอนปลายเป็นส่วนหนึ่งของการคัดเลือกให้มากขึ้น อีกทั้งให้เพิ่มการพิจารณาความสามารถของผู้สมัคร โดยพิจารณาจากองค์ประกอบหลัก ได้แก่ ผลการสอบจากแบบทดสอบวิชาหลัก และ/หรือ แบบทดสอบความสามารถทางการเรียน ซึ่งจะจัดสอบโดยสถาบันทดสอบกลางที่ได้มีการจัดตั้งขึ้น รวมทั้งให้มีการพิจารณาผลการเรียนระดับมัธยมศึกษาตอนปลายในส่วนอื่นๆ ที่มีรายละเอียดและได้รับการตรวจสอบจากสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน ทั้งนี้ มหาวิทยาลัย/สถาบันอาจกำหนดคุณสมบัติ หรืออาจกำหนดให้มีการสอบวิชาเฉพาะหรือวิชาความถนัดเฉพาะด้านเพิ่มเติมด้วยก็ได้ โดยการจัดสอบให้ทำโดยระบบการสอบรวมซึ่งในปัจจุบันนี้นับว่าเป็นระบบที่มีประสิทธิภาพมากอยู่แล้ว ทั้งนี้การปรับปรุงสาระในกระบวนการคัดเลือกดังกล่าว จะเป็นไปโดยสอดคล้องกับการปฏิรูปการศึกษาตามพระราชบัญญัติการศึกษาแห่งชาติ พ.ศ. 2542 และได้เสนอให้เริ่มใช้ระบบที่ปรับปรุงใหม่นี้ตั้งแต่ ปีการศึกษา พ.ศ. 2547 เป็นต้นไป

3. แนวทางและหลักการในระบบการรับเข้าศึกษา

ในการพิจารณาของที่ประชุมอธิการบดีแห่งประเทศไทย (ทปอ.) ร่วมกับสำนักงานคณะกรรมการการอุดมศึกษา (ทบวงมหาวิทยาลัยเดิม) ได้ยึดหลักการแนวทางเพื่อกำหนดเป็นระบบใหม่ในการรับบุคคลเข้าศึกษาในระดับอุดมศึกษาดังต่อไปนี้

3.1 ระบบใหม่จะต้องปรับเปลี่ยนจากระบบสอบแข่งขันเพื่อคัดเลือกบุคคลเข้าศึกษาในมหาวิทยาลัย (Entrance Examination) เป็นระบบการรับเข้า (Admissions) โดยพิจารณาจากผลการเรียนในระดับมัธยมศึกษา และต้องเป็นระบบที่มีความยุติธรรม โปร่งใส และตรวจสอบได้

3.2 การพิจารณาผลการเรียนเพื่อประโยชน์ในการรับเข้าศึกษาในระดับอุดมศึกษา จะพิจารณาจากการวัดผลด้วยวิธีการต่างๆ และตามช่วงเวลาต่างๆ ที่กำหนดในหลักสูตรการศึกษา ขั้นพื้นฐาน ตามกระบวนการปฏิรูปการศึกษา การพิจารณาผลการเรียน เป็นกิจกรรมทั้งในหลักสูตร และกิจกรรมประกอบที่ดำเนินการ โดยสถานศึกษา และสำนักทดสอบกลางแห่งชาติที่จะได้จัดตั้งขึ้น ตามพระราชบัญญัติการศึกษาแห่งชาติ พ.ศ. 2542

3.3 หลักเกณฑ์การสอบเพิ่มเติมโดยตั้งเป้าหมายว่า เมื่อผู้สำเร็จการศึกษาออกจาก สถานศึกษาระดับมัธยมศึกษา ก็จะมีข้อมูลเพียงพอแก่การพิจารณารับเข้าของมหาวิทยาลัย โดยไม่ ต้องมีการจัดสอบคัดเลือกเพิ่มเติมเช่นที่เป็นอยู่ในปัจจุบัน หรือหากแม้มีการสอบเพิ่มเติมกำหนดให้ สอบได้ไม่เกิน 3 รายวิชา

3.4 การดำเนินการคัดเลือกเพื่อเข้าศึกษาจะดำเนินการเป็นระบบกลาง หรือ Central University Admissions System (CUAS) (สำนักทดสอบกลาง สำนักงานคณะกรรมการการอุดมศึกษา. 2549 : 3-5)

ที่ประชุมอธิการบดีแห่งประเทศไทยร่วมกับสำนักงานคณะกรรมการการอุดมศึกษา ได้ บรรลุ ข้อตกลงร่วมกัน ซึ่งกลายมาเป็นองค์ประกอบของระบบใหม่ที่ใช้สำหรับการรับบุคคลเข้า ศึกษาในมหาวิทยาลัย โดยมีข้อสรุปดังตาราง 1 (สุธา เหลือดมัย. 2549 : 6)

ตาราง 1 องค์ประกอบของระบบกลางการรับนิสิตนักศึกษา (Admissions)

องค์ประกอบ	ค่าน้ำหนัก ปีการศึกษา 2549	ค่าน้ำหนัก ปีการศึกษา 2550	ค่าน้ำหนัก ปีการศึกษา 2551
1. ผลการเรียนเฉลี่ยสะสมตลอดหลักสูตรมัธยมศึกษา ตอนปลาย หรือเทียบเท่า (GPAX)	10%	10%	10%
2. ผลการเรียนเฉลี่ยสะสมตลอดหลักสูตรมัธยมศึกษา ตอนปลายตามกลุ่มสาระการเรียนรู้ (GPA กลุ่มสาระ 3-5 กลุ่ม จาก 8 กลุ่ม)	20%	30%	40%
3. ผลการสอบทางศึกษาแห่งชาติขั้นพื้นฐาน (Ordinary National Educational Test :O-NET)	35 – 70%	35 –60%	35 –50%
4. ผลการสอบทางการศึกษาแห่งชาติขั้นสูง (Advanced National Educational Test :A-NET) และ/หรือวิชาเฉพาะไม่เกิน 3 วิชา	0 – 35%	0 – 35%	0 – 35%

การทดสอบทางการศึกษาแห่งชาติขั้นพื้นฐาน (Ordinary National Educational Test) เป็นการวัดผลการจัดการศึกษาขั้นพื้นฐาน ในช่วงชั้นที่ 4 จัดสอบ 5 วิชา ได้แก่ ภาษาไทย คณิตศาสตร์ วิทยาศาสตร์ สังคมศึกษา ภาษาอังกฤษ จัดสอบเพียงครั้งเดียว สำหรับผู้เรียนจบหลักสูตรการศึกษา ขั้นพื้นฐาน จึงนับว่าเป็นส่วนหนึ่งของการศึกษา ซึ่งนักเรียนจะต้องสอบตามระบบการเรียนโดยปกติ อยู่แล้ว ไม่ใช่เป็นการกำหนดให้สอบเพิ่มเติม และสถาบันอุดมศึกษามีความเห็นร่วมกันว่าน่าจะเป็นข้อมูลที่เป็นประโยชน์ต่อการคัดเลือกบุคคลเข้าศึกษาต่อ จึงได้กำหนดให้นำมาใช้เป็นส่วนหนึ่งขององค์ประกอบ

การทดสอบทางการศึกษาแห่งชาติขั้นสูง (Advanced National Educational Test) เป็นแบบทดสอบวัดความรู้ ความคิดวิเคราะห์ สังเคราะห์ โดยเน้นทักษะความคิดมากกว่า O-NET ประกอบด้วยวิชาภาษาไทย คณิตศาสตร์ วิทยาศาสตร์ สังคมศึกษา ภาษาอังกฤษ ในระยะแรกจะจัดสอบเพียงครั้งเดียว โดยจัดสอบต่อจากการสอบ O-NET นักเรียนสามารถสมัครสอบ A-NET ได้มากกว่า 1 ครั้ง โดยใช้ผลคะแนนการสอบครั้งที่ดีที่สุดยื่นสมัครเข้าศึกษาต่อในสถาบันอุดมศึกษา การสอบ A-NET แต่ละครั้งจะมีอายุการใช้ 3 ปี ข้อสอบแต่ละวิชาใช้เวลา 2 ชั่วโมง ยกเว้นวิชา วิทยาศาสตร์ ใช้เวลา 3 ชั่วโมง

คำอธิบายของแต่ละองค์ประกอบ ในระบบการรับเข้าศึกษามีดังนี้

องค์ประกอบที่ 1 ผลการเรียนเฉลี่ยสะสมตลอดหลักสูตร (GPAX) เป็นตัวเลขวัดผลการเรียนที่รวบรวมมาจากการเรียนทุกรายวิชาและทุกชั้นปี จึงเป็นคะแนนที่สะท้อนผลการเรียนจริงของนักเรียน และเป็นหลักประกันว่าผู้สมัครเข้าศึกษาได้ผ่านกระบวนการเรียนในระดับมัธยมศึกษาตอนปลาย มาโดยครบถ้วน เป็นกระบวนการที่พัฒนาบุคคลให้เป็นคนที่สมบูรณ์ตามแผนการศึกษา จึงถือว่าเป็นดัชนีบ่งชี้ที่ดีที่สุดเท่าที่มีในปัจจุบันประกอบกับจากการศึกษาเบื้องต้นและการเฝ้าระวังของหน่วยงานต่างๆ เช่น สำนักงานคณะกรรมการการศึกษาแห่งชาติ และมหาวิทยาลัยได้บ่งชี้ว่า ค่า GPAX จากโรงเรียนมีความน่าเชื่อถือ อย่างไรก็ตาม ในขณะที่ยังไม่มีข้อมูลเชิงละเอียดในทางวิชาการที่มากเพียงพอที่จะยืนยันความเชื่อถือได้ของค่า GPAX โดยสมบูรณ์ ทพบอ. จึงเห็นว่าน้ำหนักของค่าคะแนนดังกล่าวที่จะนำมาใช้ ไม่ควรที่จะมากกว่าที่ใช้กำหนดให้คิดค่าน้ำหนักเพียงร้อยละ 10

องค์ประกอบที่ 2 ผลการเรียนเฉลี่ยสะสมตามกลุ่มสาระการเรียนรู้ (GPA) การนำมาเป็นองค์ประกอบหนึ่งของการคัดเลือกก็เพื่อต้องการคัดนักเรียนได้ตรงตามความสามารถหรือความถนัด โดยแต่ละคณะวิชาได้กำหนดกลุ่มสาระที่จะนำมาพิจารณาสอดคล้องกับสาขาวิชาที่นักเรียนจะศึกษาต่อ อันเป็นการพิจารณาคุณลักษณะของนักเรียนที่ละเอียดกว่าการดูจาก GPAX ซึ่งเป็นการพิจารณาในภาพรวม

องค์ประกอบที่ 3 การทดสอบทางการศึกษาแห่งชาติขั้นพื้นฐาน (O-NET) ในหลักสูตรการศึกษาขั้นพื้นฐาน พ.ศ. 2544 ได้กำหนดให้มีการประเมินคุณภาพระดับชาติ สถานศึกษาต้องจัดให้ผู้เรียนทุกคนที่เรียนในปีสุดท้ายของแต่ละช่วงชั้น ได้แก่ ชั้นประถมศึกษาปีที่ 3 ชั้นประถมศึกษาปีที่ 6 ชั้นมัธยมศึกษาปีที่ 3 และชั้นมัธยมศึกษาปีที่ 6 เข้ารับการประเมินคุณภาพระดับชาติ ในกลุ่มสาระการเรียนรู้ที่สำคัญ สำหรับการประเมินคุณภาพของช่วงชั้นที่ 4 หรือชั้นมัธยมศึกษาปีที่ 6 จะพิจารณาใน 5 กลุ่มสาระการเรียนรู้ คือ ภาษาไทย คณิตศาสตร์ วิทยาศาสตร์ สังคมศึกษา ศาสนาและวัฒนธรรม และภาษาอังกฤษ การสอบตามช่วงชั้นทั้ง 4 ช่วงชั้นจะเป็นการจัดสอบโดยสถาบันทดสอบทางการศึกษาแห่งชาติซึ่งเป็นหน่วยงานกลางที่จัดตั้งขึ้น วัตถุประสงค์เบื้องต้นของการทดสอบเพื่อประโยชน์ในการประกันคุณภาพการศึกษา และเป็นดัชนีบ่งชี้เพื่อใช้ในการปรับปรุงคุณภาพการเรียนการสอนของโรงเรียน โดยหลักการ คือ GPA ที่ได้จากโรงเรียนไม่ควรแตกต่างจากคะแนนสอบ O-NET อย่างมีนัยสำคัญ (สุธา เหลือลมัย, 2549 : 7) หากมีความแตกต่างกันมาก สถานศึกษาก็จะต้องดำเนินการปรับปรุงกระบวนการเรียนการสอนและการวัดผลประเมินผลให้ได้มาตรฐานตามที่กระทรวงศึกษาธิการกำหนด ค่าคะแนน O-NET นี้ นอกจากวัตถุประสงค์ดังกล่าวแล้ว จึงจะนำมาใช้เพื่อประโยชน์ในการรับเข้าศึกษาในมหาวิทยาลัยด้วย

องค์ประกอบที่ 4 การทดสอบทางการศึกษาแห่งชาติขั้นพื้นฐาน (A-NET) การกำหนดสอบ A-NET และ/หรือวิชาเฉพาะตามองค์ประกอบที่ 4 นี้เพื่อให้คณะวิชาต่างๆ ได้มีโอกาสทดสอบเพิ่มเติม หากเห็นว่าคะแนนจาก O-NET และ GPA ไม่เพียงพอสำหรับการพิจารณาความเหมาะสมในการศึกษาต่อ ทั้งนี้ได้กำหนดให้มีรายวิชาที่จำเป็นจะต้องสอบให้น้อยที่สุด และควรทำการทดสอบ ในคราวเดียวกับการสอบ O-NET

โดยสรุปจะเห็นว่ากระบวนการรับบุคคลเข้าศึกษาในมหาวิทยาลัยในระบบ Admissions นี้ จะพิจารณาจากข้อมูลของนักเรียนที่ได้ในระหว่างการศึกษาระดับมัธยมศึกษาตอนปลายแทบทั้งสิ้น โดยหลีกเลี่ยงการสอบเพิ่มเติมหรือหากมีการสอบก็ให้สอบเท่าที่จำเป็นแต่ไม่เกิน 3 รายวิชา (CUAS: http://www.cuas.or.th/download_pdf/cuasinfo.pdf)

การจัดสอบ O-NET และ A-NET สถาบันทดสอบทางการศึกษาแห่งชาติร่วมกับสนามสอบของมหาวิทยาลัยดำเนินการจัดสอบ โดยยึดหลักปฏิบัติตามระบบ Entrance เดิม สนามสอบของมหาวิทยาลัยแบ่งเป็น

- ศูนย์สอบภายในกรุงเทพฯ ได้แก่ มหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตบางเขน สถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ มหาวิทยาลัยศิลปากร วังท่าพระ มหาวิทยาลัยศรีนครินทรวิโรฒ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง จุฬาลงกรณ์มหาวิทยาลัย มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี

- ศูนย์สอบส่วนภูมิภาค ได้แก่ มหาวิทยาลัยธรรมศาสตร์ ศูนย์รังสิต มหาวิทยาลัยขอนแก่น มหาวิทยาลัยเชียงใหม่ มหาวิทยาลัยเทคโนโลยีสุรนารี มหาวิทยาลัยนเรศวร มหาวิทยาลัยบูรพา มหาวิทยาลัยมหาสารคาม มหาวิทยาลัยวลัยลักษณ์ มหาวิทยาลัยศิลปากร วิทยาเขตพระราชวังสนามจันทร์ มหาวิทยาลัยสงขลานครินทร์ วิทยาเขตหาดใหญ่ มหาวิทยาลัยอุบลราชธานี (<http://www.niets.or.th/mynetwork.html#top>)

1.4 การกำหนดกลุ่มสาระการเรียนรู้ รายวิชาเฉพาะ และกลุ่มสาขาวิชาที่เกี่ยวข้อง

การกำหนดกลุ่มสาระการเรียนรู้ รายวิชาเฉพาะ และกลุ่มสาขาวิชาที่เกี่ยวข้องในการรับบุคคลเข้าศึกษาในระบบกลางมีดังนี้

1. กลุ่มสาระการเรียนรู้ ได้แก่ 1. กลุ่มสาระการเรียนรู้ภาษาไทย 2. กลุ่มสาระการเรียนรู้คณิตศาสตร์ 3. กลุ่มสาระการเรียนรู้วิทยาศาสตร์ 4. กลุ่มสาระการเรียนรู้สังคมศึกษา ศาสนาและวัฒนธรรม 5. กลุ่มสาระการเรียนรู้สุขศึกษาและพลศึกษา 6. กลุ่มสาระการเรียนรู้ศิลปะ 7. กลุ่มสาระการเรียนรู้การงานอาชีพและเทคโนโลยี และ 8. กลุ่มสาระการเรียนรู้ภาษาต่างประเทศ

2. วิชาเฉพาะ ได้แก่ วิชาความถนัดต่างๆ เช่น ด้านศิลปะ วัฒนธรรมความเป็นครู ความถนัดทางสถาปัตยกรรม ความถนัดทางด้านวิศวกรรม กำหนดสอบครั้งแรกในเดือนตุลาคม ปีละครั้งเดียว สำหรับวิชาทางด้านภาษาต่างประเทศ ได้แก่ ฝรั่งเศส เยอรมัน จีน ญี่ปุ่น บาลี อาหรับ กำหนดสอบต่อจากการสอบ A – NET

3. กลุ่มสาขาวิชา ในการรับบุคคลเข้าศึกษาในระดับอุดมศึกษาประจำปีการศึกษา 2550 มีกลุ่มสาขาวิชาจำนวน 9 กลุ่มสาขาวิชา ได้แก่

3.1 กลุ่มวิทยาศาสตร์สุขภาพ ได้แก่ คณะแพทยศาสตร์ ทันตแพทยศาสตร์ เทคนิคการแพทย์ เกษศาสตร์ สหเวชศาสตร์ พยาบาลศาสตร์ สาธารณสุขศาสตร์ วิทยาศาสตร์การกีฬา พลศึกษาและสุขศึกษา แยกเป็น 4 กลุ่มย่อยได้แก่

3.1.1 วิทยาศาสตร์สุขภาพ กำหนดใช้ GPA 5 กลุ่มสาระ ได้แก่ ภาษาไทย ภาษาอังกฤษ สังคมศึกษา คณิตศาสตร์ วิทยาศาสตร์ กลุ่มละ 4% เท่าๆกัน รวมน้ำหนัก 20%

สอบ O-NET จำนวน 5 กลุ่มสาระ ได้แก่ ภาษาไทย ภาษาอังกฤษ สังคมศึกษา คณิตศาสตร์ วิทยาศาสตร์ กลุ่มละ 7% รวม 35% สอบ A-NET จำนวน 3 กลุ่มสาระ ได้แก่ ภาษาอังกฤษ 10% คณิตศาสตร์ 10% วิทยาศาสตร์ 15% รวม 35%

3.1.2 สุขศึกษาและพลศึกษา กำหนดใช้ GPA 5 กลุ่มสาระ ได้แก่ ภาษาไทย 2% ภาษาอังกฤษ 2% คณิตศาสตร์ 3% วิทยาศาสตร์ 3% สุขศึกษาและพลศึกษา 10% รวม น้ำหนัก 20% สอบ O-NET จำนวน 5 กลุ่มสาระ ได้แก่ ภาษาไทย ภาษาอังกฤษ สังคมศึกษา คณิตศาสตร์ วิทยาศาสตร์ กลุ่มละ 14% รวม 70% ไม่มีสอบ A-NET

3.1.3 วิทยาศาสตร์การกีฬา (รูปแบบที่ 1) กำหนดใช้ GPA 5 กลุ่มสาระ ได้แก่ ภาษาไทย 2% ภาษาอังกฤษ 2% คณิตศาสตร์ 3% วิทยาศาสตร์ 3% สุขศึกษาและพลศึกษา 10% รวม น้ำหนัก 20% สอบ O-NET จำนวน 5 กลุ่มสาระ ได้แก่ ภาษาไทย ภาษาอังกฤษ สังคมศึกษา คณิตศาสตร์ วิทยาศาสตร์ กลุ่มละ 8% เท่ากัน รวม 40% สอบ A-NET จำนวน 3 กลุ่มสาระ ได้แก่ วิทยาศาสตร์ คณิตศาสตร์ และภาษาอังกฤษ กลุ่มละ 10% รวม 30%

3.1.4 วิทยาศาสตร์การกีฬา (รูปแบบที่ 2) กำหนดใช้ GPA 5 กลุ่มสาระ ได้แก่ ภาษาไทย 2% ภาษาอังกฤษ 2% คณิตศาสตร์ 3% วิทยาศาสตร์ 3% สุขศึกษาและพลศึกษา 10% รวม น้ำหนัก 20% สอบ O-NET จำนวน 5 กลุ่มสาระ ได้แก่ ภาษาไทย ภาษาอังกฤษ สังคมศึกษา คณิตศาสตร์ วิทยาศาสตร์ กลุ่มละ 8% เท่ากัน รวม 40% สอบ A-NET จำนวน 3 กลุ่มสาระ ได้แก่ ภาษาไทย สังคมศึกษาและภาษาอังกฤษ กลุ่มละ 10% รวม 30%

3.2 กลุ่มวิทยาศาสตร์กายภาพ ได้แก่ คณะวิทยาศาสตร์ ทรัพยากรธรรมชาติและ เทคโนโลยีสารสนเทศ เป็นต้น กำหนดใช้ GPA 5 กลุ่มสาระ ได้แก่ ภาษาไทย ภาษาอังกฤษ สังคมศึกษา กลุ่มละ 3% วิทยาศาสตร์ 6% คณิตศาสตร์ 5% รวม 20% สอบ O-NET จำนวน 5 กลุ่มสาระ ได้แก่ ภาษาไทย ภาษาอังกฤษ สังคมศึกษา คณิตศาสตร์ วิทยาศาสตร์ 7% เท่าๆกัน รวม 35% สอบ A-NET จำนวน 3 กลุ่มสาระ ได้แก่ วิทยาศาสตร์ 20% คณิตศาสตร์ 10% ภาษาอังกฤษ 5% รวม 35%

3.3 กลุ่มวิศวกรรมศาสตร์ ได้แก่ คณะวิศวกรรมศาสตร์ กำหนดใช้ GPA 5 กลุ่มสาระ ได้แก่ ภาษาไทย ภาษาอังกฤษ สังคมศึกษา คณิตศาสตร์ วิทยาศาสตร์ กลุ่มละ 4% เท่าๆกัน รวม 20% สอบ O-NET จำนวน 5 กลุ่มสาระ ได้แก่ ภาษาไทย ภาษาอังกฤษ สังคมศึกษา คณิตศาสตร์ วิทยาศาสตร์ กลุ่มละ 8% เท่าๆกัน รวม 40% สอบ A-NET จำนวน 3 กลุ่มสาระ ได้แก่ คณิตศาสตร์ 10% วิทยาศาสตร์ 10% ความถนัดทางวิศวกรรม 10% รวม 30%

3.4 กลุ่มสถาปัตยกรรมศาสตร์ ได้แก่ คณะสถาปัตยกรรมศาสตร์ กำหนดใช้ GPA 5 กลุ่มสาระ ได้แก่ ภาษาไทย สังคมศึกษา กลุ่มละ 2.5% ภาษาอังกฤษ คณิตศาสตร์ วิทยาศาสตร์ กลุ่มละ 5% รวม 20% สอบ O-NET จำนวน 5 กลุ่มสาระ ได้แก่ ภาษาไทย ภาษาอังกฤษ สังคมศึกษา คณิตศาสตร์ วิทยาศาสตร์ กลุ่มละ 8% เท่าๆกัน รวม 40% สอบวิชาเฉพาะ ได้แก่ ความถนัดทางสถาปัตยกรรม 30%

3.5 กลุ่มเกษตรศาสตร์ ได้แก่ คณะเกษตรศาสตร์ อุตสาหกรรมเกษตรวนศาสตร์ เทคโนโลยีการเกษตร เป็นต้น กำหนดใช้ GPA 5 กลุ่มสาระ ได้แก่ คณิตศาสตร์ 4% ภาษาอังกฤษ 4% สังคมศึกษา 3% ภาษาไทย 3% วิทยาศาสตร์ 6% รวม 20% สอบ O-NET จำนวน 5 กลุ่มสาระ

ได้แก่ ภาษาไทย ภาษาอังกฤษ สังคมศึกษา คณิตศาสตร์ วิทยาศาสตร์ ศึกษาศาสตร์ กลุ่มละ 8% เท่าๆกัน รวม 40% สอบ A-NET จำนวน 2 วิชา ได้แก่ วิทยาศาสตร์ 20% คณิตศาสตร์ 10% รวม 30%

3.6 กลุ่มบริหารธุรกิจพาณิชยศาสตร์ การบัญชี การจัดการ การท่องเที่ยวและ เศรษฐศาสตร์ โดยแบ่งเป็น

3.6.1 สาขาวิชาบริหารธุรกิจ พาณิชยศาสตร์ การบัญชี เศรษฐศาสตร์ กำหนดใช้ GPA 5 กลุ่มสาระ ได้แก่ ภาษาไทย ภาษาอังกฤษ สังคมศึกษา คณิตศาสตร์ วิทยาศาสตร์ กลุ่มละ 4% เท่าๆกัน รวม 20% สอบ O-NET จำนวน 5 กลุ่มสาระ ได้แก่ ภาษาไทย ภาษาอังกฤษ สังคมศึกษา คณิตศาสตร์ วิทยาศาสตร์ กลุ่มละ 7% รวม 35% สอบ A-NET จำนวน 2 วิชา ได้แก่ คณิตศาสตร์ 20% ภาษาอังกฤษ 15% รวม 35%

3.6.2 สาขาวิชาการท่องเที่ยวและโรงแรม กำหนดใช้ GPA 3 กลุ่มสาระ ได้แก่ ภาษาไทย 5% คณิตศาสตร์ 5% ภาษาอังกฤษ 10% รวม 20% สอบ O-NET จำนวน 5 กลุ่มสาระ ได้แก่ ภาษาไทย ภาษาอังกฤษ สังคมศึกษา คณิตศาสตร์ วิทยาศาสตร์ กลุ่มละ 14% รวม 70% และไม่มีการสอบ A-NET

3.7 กลุ่มครุศาสตร์/ศึกษาศาสตร์ (การสอน) กำหนดใช้ GPA 5 กลุ่มสาระ ได้แก่ ภาษาไทย ภาษาอังกฤษ สังคมศึกษา คณิตศาสตร์ วิทยาศาสตร์ กลุ่มละ 4% เท่าๆกัน รวม 20% สอบ O-NET จำนวน 5 กลุ่มสาระ ได้แก่ ภาษาไทย ภาษาอังกฤษ สังคมศึกษา คณิตศาสตร์ วิทยาศาสตร์ กลุ่มละ 8% รวม 40% สอบ A-NET เฉพาะวิชาเฉพาะ ได้แก่ ความถนัดทางวิชาชีพครู 10% วิชาเอก/วิชาเฉพาะไม่เกิน 2 วิชาๆ ละ 10% หรือสอบวิชาเดียว 20%

3.8 กลุ่มศิลปกรรม ทัศนศิลป์และประยุกต์ศิลป์ ได้แก่ คณะศิลปกรรมศาสตร์ ทัศนศิลป์ มัณฑนศิลป์ จิตรกรรม ประติมากรรม และภาพพิมพ์ กำหนดใช้ GPA 3 กลุ่มสาระ ได้แก่ ภาษาไทย สังคมศึกษา กลุ่มละ 6% และภาษาอังกฤษ 8% รวม 20 % สอบ O-NET จำนวน 5 กลุ่ม สาระได้แก่ ภาษาไทย ภาษาอังกฤษ สังคมศึกษา คณิตศาสตร์ วิทยาศาสตร์ กลุ่มละ 7% รวม 35% ไม่มีสอบ A-NET แต่จะทดสอบเฉพาะตามกลุ่มสาขา จำนวน 35% แบ่งเป็น 1. กลุ่มที่สอบปฏิบัติ อย่างเดียว ได้แก่ วาดเส้น องค์ประกอบศิลป์ 2. สอบทฤษฎีและปฏิบัติด้านทัศนศิลป์และการ ออกแบบ 3. สอบทฤษฎีและปฏิบัติด้านดนตรี และ 4. สอบทฤษฎีและปฏิบัติด้านนาฏศิลป์และ ศิลปการแสดง

3.9 กลุ่มมนุษยและสังคมศาสตร์ ได้แก่ คณะนิเทศศาสตร์ วารสารศาสตร์ อักษร ศาสตร์ ศิลปศาสตร์ มนุษยศาสตร์ รัฐศาสตร์ นิติศาสตร์ สังคมวิทยา สังคมสงเคราะห์ศาสตร์ เป็นต้น แบ่งออกเป็น 6 กลุ่มย่อย ได้แก่

3.9.1 นิเทศศาสตร์/วารสารศาสตร์ กำหนดใช้ GPA 4 กลุ่มสาระ ได้แก่ ภาษาไทย 5% สังคมศึกษา 5% ภาษาอังกฤษ 5% และคณิตศาสตร์/วิทยาศาสตร์/ภาษาต่างประเทศ/ศิลปะ 5% รวม 20% สอบ O-NET จำนวน 5 กลุ่มสาระ ได้แก่ ภาษาไทย ภาษาอังกฤษ สังคมศึกษา คณิตศาสตร์ วิทยาศาสตร์ กลุ่มละ 14% รวม 70% ไม่มีสอบ A-NET

3.9.2 สังคมวิทยา/มานุษยวิทยา/สังคมศาสตร์ มี 2 รูปแบบคือ

รูปแบบที่ 1 กำหนดใช้ GPA 5 กลุ่มสาระ ได้แก่ ภาษาไทย ภาษาอังกฤษ สังคมศึกษา คณิตศาสตร์ วิทยาศาสตร์ กลุ่มละ 4% เท่ากัน รวม 20% สอบ O-NET จำนวน 5 กลุ่มสาระ ได้แก่ ภาษาไทย ภาษาอังกฤษ สังคมศึกษา คณิตศาสตร์ วิทยาศาสตร์ กลุ่มละ 14% รวม 70% ไม่มีสอบ A-NET

รูปแบบที่ 2 กำหนดใช้ GPA 5 กลุ่มสาระ ได้แก่ ภาษาไทย ภาษาอังกฤษ สังคมศึกษา คณิตศาสตร์ วิทยาศาสตร์ กลุ่มละ 4% เท่ากัน รวม 20% สอบ O-NET จำนวน 3 กลุ่ม สาระ ได้แก่ ภาษาไทย 10% สังคมศึกษา 10% และ ภาษาอังกฤษ 15% รวม 35% สอบ A-NET 2 กลุ่ม สาระ ได้แก่ คณิตศาสตร์ 20% และวิทยาศาสตร์ 15% รวม 35%

3.9.3 รัฐศาสตร์ แยกเป็น 3 สาขา ได้แก่

1) สาขาการปกครอง/บริหารรัฐกิจ กำหนดใช้ GPA 3 กลุ่มสาระ ได้แก่ ภาษาไทย 5% สังคมศึกษา 10% ภาษาต่างประเทศ/คณิตศาสตร์ 5% รวม 20% สอบ O-NET จำนวน 5 กลุ่มสาระ ได้แก่ ภาษาไทย ภาษาอังกฤษ สังคมศึกษา คณิตศาสตร์ วิทยาศาสตร์ กลุ่มละ 7% รวม 35% สอบ A-NET/วิชาเฉพาะ คือ สังคมศึกษา / ฝรั่งเศส /เยอรมัน/ บาลี/อาหรับ/จีน/ญี่ปุ่น (เลือก 1 วิชา) 35%

2) สาขาความสัมพันธ์ระหว่างประเทศ กำหนดใช้ GPA 3 กลุ่มสาระ ได้แก่ ภาษาไทย 5% สังคมศึกษา 10% ภาษาต่างประเทศ/คณิตศาสตร์ 5% รวม 20% สอบ O-NET จำนวน 5 กลุ่มสาระ ได้แก่ ภาษาไทย ภาษาอังกฤษ สังคมศึกษา คณิตศาสตร์ วิทยาศาสตร์ กลุ่มละ 7% รวม 35% สอบ A-NET/วิชาเฉพาะ คือ สังคมศึกษา 17.5% และภาษาอังกฤษ/ ฝรั่งเศส /เยอรมัน/ บาลี/อาหรับ/จีน/ ญี่ปุ่น (เลือก 1 วิชา) 17.5% รวม 35%

3) สาขารัฐประศาสนศาสตร์ กำหนดใช้ GPA 3 กลุ่มสาระ ได้แก่ ภาษาไทย 5% สังคมศึกษา 10% คณิตศาสตร์ 5% รวม 20% สอบ O-NET จำนวน 5 กลุ่มสาระ ได้แก่ ภาษาไทย ภาษาอังกฤษ สังคมศึกษา คณิตศาสตร์ วิทยาศาสตร์ กลุ่มละ 7% รวม 35% สอบ A-NET 1 กลุ่มสาระ คือ คณิตศาสตร์ 35%

3.9.4 สังคมสงเคราะห์ กำหนดใช้ GPA 5 กลุ่มสาระ ได้แก่ ภาษาไทย สังคม ศึกษา ภาษาต่างประเทศ กลุ่มละ 5% คณิตศาสตร์ และวิทยาศาสตร์ กลุ่มละ 2.5% รวม 20% สอบ O-

NETจำนวน 3 กลุ่มสาระ ได้แก่ ภาษาไทย 20% สังคมศึกษา 20% และภาษาอังกฤษ 15% รวม 55% และสอบ A-NET คณิตศาสตร์/วิทยาศาสตร์ /ฝรั่งเศส/เยอรมัน/บาลี/อาหรับ/จีน/ญี่ปุ่น (เลือก 1 วิชา) 15%

3.9.5 นิติศาสตร์ กำหนดใช้ GPA 3 กลุ่มสาระ ได้แก่ ภาษาไทย 7% สังคมศึกษา 6% ภาษต่างประเทศ 7% รวม 20% สอบ O-NETจำนวน 3 กลุ่มสาระ ได้แก่ ภาษาไทย 18% สังคมศึกษา 19% และภาษาอังกฤษ 18% รวม 55% และสอบ A-NET คณิตศาสตร์/วิทยาศาสตร์ /ฝรั่งเศส/เยอรมัน/บาลี/อาหรับ/จีน/ญี่ปุ่น (เลือก 1 วิชา) 15%

3.9.6 มนุษยศาสตร์/ศิลปศาสตร์/อักษรศาสตร์ มี 9 รูปแบบ ได้แก่

รูปแบบที่ 1 กำหนดใช้ GPA 4 กลุ่มสาระ ได้แก่ ภาษาไทย สังคมศึกษา ภาษต่างประเทศ และคณิตศาสตร์ กลุ่มละ 5% รวม 20% สอบ O-NETจำนวน 3 กลุ่มสาระ ได้แก่ ภาษาไทย 10% สังคมศึกษา 10% และภาษาอังกฤษ 15% รวม 35% และสอบ A-NET 3 กลุ่มสาระ ได้แก่ ภาษาไทย 10% ภาษาอังกฤษ 10% คณิตศาสตร์/วิทยาศาสตร์ (เลือก 1 วิชา) 15% รวม 35%

รูปแบบที่ 2 กำหนดใช้ GPA 3 กลุ่มสาระ ได้แก่ ภาษาไทย 5% สังคมศึกษา 5% และภาษต่างประเทศ 10% รวม 20% สอบ O-NETจำนวน 3 กลุ่มสาระ ได้แก่ ภาษาไทย 10% สังคมศึกษา 10% และภาษาอังกฤษ 15% รวม 35% และสอบ A-NET 3 กลุ่มสาระ ได้แก่ ภาษาไทย 10% ภาษาอังกฤษ 10% และฝรั่งเศส/เยอรมัน/บาลี/อาหรับ/จีน/ญี่ปุ่น (เลือก 1 วิชา) 15% รวม 35%

รูปแบบที่ 3 กำหนดใช้ GPA 4 กลุ่มสาระ ได้แก่ ภาษาไทย สังคมศึกษา ภาษต่างประเทศ และคณิตศาสตร์ กลุ่มละ 5% รวม 20% สอบ O-NET จำนวน 5 กลุ่มสาระ ได้แก่ ภาษาไทย ภาษอังกฤษ สังคมศึกษา คณิตศาสตร์ วิทยาศาสตร์ กลุ่มละ 7% รวม 35% และสอบ A-NET 3 กลุ่มสาระ ได้แก่ ภาษาไทย 10% ภาษาอังกฤษ 10% คณิตศาสตร์/วิทยาศาสตร์ (เลือก 1 วิชา) 15% รวม 35%

รูปแบบที่ 4 กำหนดใช้ GPA 3 กลุ่มสาระ ได้แก่ ภาษาไทย 5% สังคมศึกษา 5% และภาษต่างประเทศ 10% รวม 20% สอบ O-NET จำนวน 5 กลุ่มสาระ ได้แก่ ภาษาไทย สังคมศึกษา ภาษาอังกฤษ คณิตศาสตร์ วิทยาศาสตร์ กลุ่มละ 7%รวม 35% และสอบ A-NET 3 กลุ่มสาระ ได้แก่ ภาษาไทย 10% ภาษาอังกฤษ 10% ฝรั่งเศส/เยอรมัน/บาลี/อาหรับ/จีน/ญี่ปุ่น (เลือก 1 วิชา) 15% รวม 35%

รูปแบบที่ 5 กำหนดใช้ GPA 3 กลุ่มสาระ ได้แก่ ภาษาไทย 5% สังคมศึกษา 5% และภาษต่างประเทศ 10% รวม 20% สอบ O-NET จำนวน 3 กลุ่มสาระ ได้แก่ ภาษาไทย 10%



ภาษาอังกฤษ 15% และสังคมศึกษา 10% รวม 35% และสอบ A-NET ภาษาไทย/สังคมศึกษา/ ภาษาอังกฤษ/ฝรั่งเศส/เยอรมัน/บาลี/อาหรับ/จีน/ญี่ปุ่น (เลือก 1 วิชา) 35%

รูปแบบที่ 6 กำหนดใช้ GPA 4 กลุ่มสาระ ได้แก่ ภาษาไทย สังคมศึกษา ภาษาต่างประเทศ และคณิตศาสตร์ กลุ่มละ 5% รวม 20% สอบ O-NET จำนวน 3 กลุ่มสาระ ได้แก่ ภาษาไทย 23% ภาษาอังกฤษ 24% และสังคมศึกษา 23% รวม 70% ไม่มีสอบ A-NET

รูปแบบที่ 7 กำหนดใช้ GPA 3 กลุ่มสาระ ได้แก่ ภาษาไทย 5% สังคมศึกษา 5% และภาษาต่างประเทศ 10% รวม 20% สอบ O-NET จำนวน 3 กลุ่มสาระ ได้แก่ ภาษาไทย 23% ภาษาอังกฤษ 24% และสังคมศึกษา 23% รวม 70% ไม่มีสอบ A-NET

รูปแบบที่ 8 กำหนดใช้ GPA 4 กลุ่มสาระ ได้แก่ ภาษาไทย สังคมศึกษา ภาษาต่างประเทศ และคณิตศาสตร์ กลุ่มละ 5% รวม 20% สอบ O-NET จำนวน 5 กลุ่มสาระ ได้แก่ ภาษาไทย ภาษาอังกฤษ สังคมศึกษา คณิตศาสตร์ วิทยาศาสตร์ กลุ่มละ 14% รวม 70% ไม่มีสอบ A-NET

รูปแบบที่ 9 กำหนดใช้ GPA 3 กลุ่มสาระ ได้แก่ ภาษาไทย 5% สังคมศึกษา 5% และภาษาต่างประเทศ 10% รวม 20% สอบ O-NET จำนวน 5 กลุ่มสาระ ได้แก่ ภาษาไทย สังคมศึกษา ภาษาอังกฤษ คณิตศาสตร์ วิทยาศาสตร์ กลุ่มละ 14% รวม 70% ไม่มีสอบ A-NET (<http://www.cuas.or.th/>)

1.5 ระบบการรับเข้าศึกษาในปีการศึกษา 2553

หลังจากที่มีการนำระบบการรับเข้าศึกษา (Admissions) มาใช้ได้ระยะหนึ่ง ทำให้เกิดปัญหาตามมาไม่น้อย เช่น ไม่สามารถลดการกวาดวิชาลงได้ มีผลการประเมินสถานศึกษาและงานวิจัยที่ชี้ให้เห็นถึงมาตรฐานของผลการเรียน (GPA และ GPAX) ของโรงเรียนต่างๆ ไม่เท่าเทียมกัน เป็นต้น ทำให้ที่ประชุมอธิการบดีแห่งประเทศไทย (ทปอ.) ต้องทบทวนระบบการรับเข้าศึกษาใหม่ โดยได้กำหนดองค์ประกอบและค่าน้ำหนักในการคัดเลือกเข้าศึกษาในสถานศึกษา ระดับอุดมศึกษาด้วยระบบกลาง (Admissions) ปีการศึกษา 2553 ประกอบด้วย ผลการเรียนเฉลี่ยสะสมตลอดหลักสูตร (GPAX) ร้อยละ 20 ผลการทดสอบทางการศึกษาระดับพื้นฐาน (O-NET) ร้อยละ 30 และผลการทดสอบความถนัดทางการเรียน (Aptitude Test) ร้อยละ 50 ซึ่งจำแนกเป็นการทดสอบความถนัดทั่วไป (General Aptitude Test : GAT) ร้อยละ 10-50 และการทดสอบความถนัดเฉพาะด้าน/วิชาชีพ (Professional Aptitude Test : PAT) ร้อยละ 0-40 และได้กำหนดรายละเอียดเกี่ยวกับ GAT และ PAT พร้อมค่าน้ำหนักในแต่ละกลุ่มสาขาวิชา ดังนี้

ตาราง 2 องค์ประกอบ Admissions ปี 2553 รายกลุ่มสาขาวิชา

กลุ่มสาขาวิชา	องค์ประกอบ (%)				รวม (100%)
	GPAX	ONET	GAT	PAT	
1.กลุ่มวิทยาศาสตร์สุขภาพ					
1.1 เกษษศาสตร์ สหเวชศาสตร์ สาธารณสุขศาสตร์ เทคนิคการแพทย์ วิทยาศาสตร์การกีฬา พลศึกษาและสุข ศึกษา	20	30	20	P2 = 30	100
1.2 ทันตแพทยศาสตร์ พยาบาลศาสตร์	20	30	30	P2 = 20	100
2.กลุ่มวิทยาศาสตร์กายภาพ					
2.1 วิทยาศาสตร์ ทรัพยากรธรรมชาติ เทคโนโลยีสารสนเทศ	20	30	10	P1 = 10 P2 = 30	100
3. กลุ่มวิศวกรรมศาสตร์					
4. กลุ่มสถาปัตยกรรมศาสตร์	20	30	10	P4 = 40	100
5. กลุ่มเกษตรศาสตร์					
เกษตรศาสตร์ อุตสาหกรรมเกษตร วนศาสตร์ เทคโนโลยีการเกษตร	20	30	20	P1 = 10 P2 = 20	100
6. กลุ่มบริหาร พาณิชยศาสตร์ การบัญชี การจัดการท่องเที่ยว และเศรษฐศาสตร์					
6.1 บริหารธุรกิจ พาณิชยศาสตร์ การ บัญชี เศรษฐศาสตร์	20	30	30	P1 = 20	100
6.2 การท่องเที่ยว และ โรงแรม					
6.2.1 รูปแบบที่ 1	20	30	50	-	100
6.2.2 รูปแบบที่ 2	20	30	40	P7 = 10	100
7. กลุ่มครุศาสตร์/ศึกษาศาสตร์					
	20	30	20	P5 = 30	100

ตาราง 2 องค์ประกอบ Admissions ปี 2553 รายกลุ่มสาขาวิชา (ต่อ)

กลุ่มสาขาวิชา	องค์ประกอบ (%)				รวม (100%)
	GPAX	ONET	GAT	PAT	
8. กลุ่มศิลปกรรม ดุริยางคศิลป์และ นาฏยศิลป์ จิตรศิลป์ ศิลปะประยุกต์ ดุริยางค ศิลป์และนาฏยศิลป์	20	30	10	P6 = 40	100
9. กลุ่มมนุษยศาสตร์และสังคมศาสตร์ นิเทศศาสตร์ วารสารศาสตร์ อักษรศาสตร์ ศิลปะศาสตร์ มนุษยศาสตร์ รัฐศาสตร์ นิติศาสตร์ สังคมวิทยา สังคมสงเคราะห์ศาสตร์					
รูปแบบที่ 1 พื้นฐานวิทยาศาสตร์	20	30	40	P1 = 10	100
รูปแบบที่ 2 พื้นฐานศิลปะศาสตร์					
รูปแบบที่ 1	20	30	50	-	100
รูปแบบที่ 2	20	30	40	P7 = 10	100

1.6 สถาบันทดสอบทางการศึกษาแห่งชาติ

สถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) เป็นหน่วยงานของรัฐ จัดตั้งขึ้นตามพระราชกฤษฎีกาจัดตั้งสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) พ.ศ. 2548 เหตุผลในการประกาศใช้พระราชกฤษฎีกาดังกล่าว เนื่องจากมีความจำเป็นต้องมีสถาบันทดสอบทางการศึกษาเพื่อเป็นองค์กรกลางในการศึกษาวิจัย พัฒนา และให้บริการทางการประเมินผลทางการศึกษาและทดสอบทางการศึกษา รวมทั้งเป็นศูนย์กลางความร่วมมือด้านการทดสอบทางการศึกษาในระดับชาติและระดับนานาชาติ และเพื่อที่จะให้สถาบันที่จัดตั้งขึ้นมีการบริหารและการจัดการที่มีความเป็นอิสระ คล่องตัว และมีประสิทธิภาพ สมควรจัดตั้งสถาบันทดสอบทางการศึกษาแห่งชาติขึ้นเป็นองค์การมหาชน ตามกฎหมายว่าด้วยองค์การมหาชน (CUAS:http://www.cuas.or.th/download_pdf/cuasinfo.pdf)

การจัดตั้งสถาบันทดสอบทางการศึกษาแห่งชาติ มีประกาศในพระราชกฤษฎีกาที่ประกาศ ในราชกิจจานุเบกษา เล่มที่ 122 ตอนที่ 76 ก วันที่ 2 กันยายน 2548 และมีผลบังคับใช้ วันที่ 3 กันยายน 2548

สถาบันทดสอบทางการศึกษาแห่งชาติเป็นองค์กรที่รับผิดชอบในการประเมินคุณภาพ การศึกษาของประเทศ มุ่งให้เกิดการประเมินคุณภาพความรู้ และคุณลักษณะระดับต่างๆ อย่างเป็น มาตรฐานและทัดเทียมกับนานาชาติ โดยมีพันธกิจในการวัดและประเมินมาตรฐานวิชาการและ วิชาชีพเพื่อการเรียนรู้ การศึกษาต่อ และการประกอบอาชีพในระดับต่างๆ ประเมินมาตรฐานการจัด การศึกษาของประเทศเพื่อพัฒนาไปสู่ความสามารถในการแข่งขันระดับนานาชาติ จัดระบบการ บริหารการทดสอบที่เป็นมาตรฐานและมุ่งอำนวยความสะดวกให้กับผู้ใช้บริการ สนับสนุน / ส่งเสริม ประสานทรัพยากร เพื่อการพัฒนาการทดสอบให้เป็นมาตรฐานสากล พัฒนาศูนย์ข้อมูล สารสนเทศเกี่ยวกับการทดสอบระดับประเทศและระดับนานาชาติ ทั้งนี้สถาบันทดสอบทาง การศึกษาแห่งชาติ (องค์การมหาชน) มีภารกิจหลักในการทดสอบเพื่อประเมินมาตรฐาน การจัดการ ศึกษาในระดับการศึกษาขั้นพื้นฐาน โดยทดสอบนักเรียนใน 4 ช่วงชั้น คือ ช่วงชั้นที่ 1 (ป.1- ป.3) ช่วงชั้นที่ 2 (ป.4-ป.6) ช่วงชั้น ที่ 3 (ม.1 – ม.3) และช่วงชั้นที่ 4 (ม.4 – ม.6)

1.7 ปัญหาของระบบการรับเข้าศึกษา (Admissions)

การคัดเลือกบุคคลเข้าศึกษาในสถาบันอุดมศึกษาในระบบกลาง (Admissions) ถึงแม้ว่าการศึกษาส่วนใหญ่จะเห็นว่าเป็นระบบที่ดี แต่ก็มีประเด็นปัญหาที่หลายๆ ฝ่ายเห็นว่าควร ได้รับการปรับปรุงแก้ไขอีกมาก ดังที่ ไทย ทิพย์สุวรรณกุล (2543 : 23-29) ได้ศึกษาจากบทความ วารสารและหนังสือพิมพ์สามารถจำแนกลักษณะปัญหาที่พบในการคัดเลือกบุคคลเข้าศึกษา ใน ระบบการรับเข้าศึกษา (Admissions) ดังนี้

1. แนวคิดของการรับเข้าศึกษาระบบใหม่

การคัดเลือกบุคคลเข้าศึกษาในสถาบันอุดมศึกษาในระบบกลาง (Admissions) สำนักงานคณะกรรมการการอุดมศึกษา กำหนดให้พิจารณาคัดเลือกผู้สมัคร โดยมีแนวคิด 3 องค์ประกอบ คือ 1) ผลการเรียนรู้เฉลี่ยสะสมตลอดหลักสูตรมัธยมศึกษาตอนปลายหรือเทียบเท่า มีค่าน้ำหนักร้อยละ 10 และผลการเรียนเฉลี่ยสะสมตลอดหลักสูตรมัธยมศึกษาตอนปลายตามกลุ่มสาระการเรียนรู้ (3 – 5 กลุ่ม) ให้ค่าน้ำหนักร้อยละ 20 2) ผลการสอบทางการศึกษาขั้นพื้นฐาน (O-NET) ให้ค่าน้ำหนักร้อยละ 35 - 70 และผลการสอบทางการศึกษาแห่งชาติขั้นสูง (A-NET) และ/หรือวิชาเฉพาะ ไม่เกิน 3 วิชา ให้ค่าน้ำหนักร้อยละ 0 - 35 และ 3) ผลการสอบสัมภาษณ์และตรวจร่างกาย ใช้เป็นข้อมูลประกอบการพิจารณา ความพร้อมและความเหมาะสม ก่อนรับเข้าศึกษา ไม่คิดค่าน้ำหนักคะแนน

จากองค์ประกอบทั้ง 3 ข้อข้างต้น สำนักงานคณะกรรมการการอุดมศึกษาคาดว่าจะทำ ให้มหาวิทยาลัย/สถาบันได้ผู้เรียนที่มีความรู้ ความสามารถและความถนัดตรงตามสาขาวิชาที่เรียน ตลอดจนส่งเสริมให้การเรียนการสอนในระดับมัธยมศึกษาตอนปลายเป็นไปตามปรัชญาและ

วัตถุประสงค์ของหลักสูตร และเป็นแนวคิดที่ดี เพื่อให้ให้นักเรียนที่มีผลการเรียนดีและตั้งใจเรียน สมัครเข้าเรียนในระดับมหาวิทยาลัย แต่ปัญหาที่เกิดขึ้นก็คือ มีข้อโต้แย้งเกี่ยวกับการนำผลการเรียนตลอดหลักสูตร ซึ่งวัดได้จากผลการเรียนเฉลี่ยสะสมตลอดหลักสูตร (GPAX) และผลการเรียนสะสมตามกลุ่มสาระการเรียนรู้ (GPA) มาเป็นองค์ประกอบของการรับเข้าศึกษา โดยไม่ผ่านการเปรียบเทียบมาตรฐานของผลการเรียน ถึงแม้ในปีการศึกษา 2553 องค์ประกอบของการรับเข้าศึกษาจะเปลี่ยนแปลงไปบ้าง โดยใช้ผลการเรียนเฉลี่ยสะสมตลอดหลักสูตรมัธยมศึกษาตอนปลาย ผลการสอบทางการศึกษาขั้นพื้นฐาน และผลการทดสอบความถนัดทั่วไป/เฉพาะทาง ก็คงจะยังมีปัญหาเกี่ยวกับผลการเรียนเฉลี่ยสะสมตลอดหลักสูตร (GPAX) เหมือนเดิม เนื่องจากผลการเรียนเฉลี่ยสะสมตลอดหลักสูตร (GPAX) ที่ได้จากแต่ละโรงเรียนนั้น จะให้มีการประเมินผลในเกณฑ์เดียวกันนั้นเป็นเรื่องยาก เพราะคุณภาพของโรงเรียนมัธยมศึกษาแต่ละแห่งมีความแตกต่างกัน

แม้ว่าการนำผลการเรียนเฉลี่ยสะสมตลอดหลักสูตร (GPAX) มาเป็นเกณฑ์ในการคัดเลือก จะเป็นแนวทางที่ดีเป็นการส่งเสริมการเรียนการสอนตลอดหลักสูตรมัธยมศึกษาตอนปลาย ให้นักเรียนมีความตั้งใจเรียนโดยสมัคร เป็นการศึกษาจากสถาบันกวดวิชาต่างๆ ให้น้อยลง แต่มาตรฐานโรงเรียนมัธยมศึกษาในแต่ละโรงเรียนไม่เท่าเทียมกัน ทั้งในด้านคุณภาพของนักเรียนที่เข้าเรียน คุณภาพการสอน และการวัดประเมินผล รวมถึงคุณภาพของปัจจัยที่เอื้อต่อการเรียนการสอนอื่นๆ มาตรฐานการวัดประเมินผลที่แตกต่างกัน ความน่าเชื่อถือของผลการสอบแต่ละโรงเรียน ซึ่งรวมถึงวิธีการที่ใช้ในการให้ระดับผลการเรียนด้วย ทำให้หลักเกณฑ์ดังกล่าวขาดความน่าเชื่อถือ ทั้งยังพบว่า ครู/โรงเรียนอาจมีการเปลี่ยนแปลงการให้เกรด หรือปล่อยเกรดเพื่อให้นักเรียนของตนได้เปรียบในการคำนวณผลการเรียนเฉลี่ย จนทำให้คะแนนนักเรียนเพื่อ ไม่ตรงกับความสามารถที่ควรจะเป็น นักเรียนยังคงต้องกวดวิชาเพิ่มมากขึ้นกว่าเดิมเพื่อให้ได้เกรดเฉลี่ยเป็นที่น่าพอใจ เพราะการนำผลการเรียนเฉลี่ยสะสมของนักเรียนมาใช้ในการคำนวณทำให้นักเรียนไม่มีโอกาสแก้ตัวใหม่ ทำให้นักเรียนมีความเครียดมากขึ้นเพราะต้องทุ่มเทให้กับการเรียนตลอดเวลาเพื่อทำคะแนนสะสมเป็นที่น่าพอใจ

เช่นเดียวกับคำกล่าวของ ศาสตราจารย์ ดร.สมหวัง พิธิยานุวัฒน์ ผู้อำนวยการสำนักงานรับรองมาตรฐานและประเมินคุณภาพการศึกษา (สมศ.) ในการสัมมนาทางวิชาการเกี่ยวกับการปฏิรูปวิธีการรับนักศึกษาในระดับอุดมศึกษาว่า "...การวัดและประเมินการเรียนระดับมัธยมศึกษาจะต้องมีมาตรฐาน การปรับปรุงเรื่อง GPA เป็นสิ่งที่ต้องดำเนินการ เรื่องนี้ก็น่าจะอาศัยการวิจัยค้นคว้ามาก เพราะ สภาพปัจจุบันขณะนี้ GPA มาจากโรงเรียนที่มีคุณภาพต่างกัน ซึ่งทางวิชาการรับไม่ได้เลย ถ้ามองในแง่การวัดผลด้วยแล้ว ทำอย่างไรคุณภาพของโรงเรียนจะใกล้เคียงกันเป็นเรื่องที่ต้องทำในระยะยาว สำหรับในระยะสั้นจะต้องใช้เทคนิคของการแปลงค่า GPA โดยปรับให้

ใกล้เคียงกันกับคุณภาพโรงเรียนที่แตกต่างกัน...” ทำนองเดียวกับ ดร.ประพัฒน์พงษ์ เสนาฤทธิ์ รองอธิบดีกรมสามัญศึกษา ได้กล่าวไว้ในการสัมมนาในครั้งเดียวกันนี้ว่า “...หลายฝ่ายยังเชื่อว่า คุณภาพมาตรฐานโรงเรียนมัธยมศึกษามีความต่างกัน ทบวงมหาวิทยาลัย นักวิชาการ เป็นห่วงเรื่อง วิชาการ สาเหตุที่คุณภาพมาตรฐานของโรงเรียนไม่เท่ากัน มีปัจจัยหลายอย่างเช่น 1. มีตัวป้อนแตกต่างกัน คือ เด็กที่เข้าสู่โรงเรียนมีความแตกต่างกัน จะเห็นว่าโรงเรียนที่มีเด็กไปสมัครเป็นจำนวนมากก็ต้องมีการสอบคัดเลือก ทำให้ได้เด็กเก่งจำนวนหนึ่ง ในขณะที่บางโรงเรียนเด็กมาสมัครน้อยก็รับหมด ซึ่งถ้าไปพิจารณาภูมิหลังของเด็กเหล่านั้นแล้ว จะเห็นว่าภูมิหลังที่ต่างกัน ตัวป้อนต่างกันตั้งแต่เริ่มต้นแล้ว เพราะฉะนั้นจะให้ผลผลิตออกมาเท่ากันจึงเป็นไปได้ 2. ปัญหาเรื่อง จำนวนครู แต่ละโรงเรียนมีจำนวนครูไม่เท่ากัน และขีดความสามารถก็ไม่เท่ากัน แม้ว่าจะขณะนี้จะมีครูที่จบปริญญาตรี ปริญญาโท เป็นส่วนใหญ่ แต่ที่น่าเป็นห่วงคือ เมื่อมีการทดสอบความรู้ ความสามารถของครูในวิชาหลักๆ ครูหลายคนมีความรู้ไม่ถึงครึ่งหนึ่งของเนื้อหาที่สอบ นอกจากนี้ในเรื่องวัสดุอุปกรณ์ต่างๆ ก็มีความแตกต่างกันระหว่างโรงเรียนดังนั้นถ้าจะสร้างคุณภาพของโรงเรียนมัธยมศึกษาให้ใกล้เคียงกันต้องปรับวิธีการจัดสรรงบประมาณให้โรงเรียนเหล่านี้ใหม่ และพัฒนาขีดความสามารถในการวัดผลและประเมินผลของนักเรียน เพื่อสร้างความเชื่อมั่นให้เกิดขึ้นว่า ค่า GPA ที่ให้เด็กไปนั้น มหาวิทยาลัยยอมรับได้...” (สำนักงานคณะกรรมการการศึกษาแห่งชาติ. 2543 : 39-51)

2. ระบบบริหารการจัดสอบ

ในการจัดสอบของระบบการรับเข้าศึกษา (Admissions) สำนักงานคณะกรรมการการอุดมศึกษาได้กระจายสถานที่จัดสอบโดยใช้สนามสอบทั่วประเทศไทย มีมหาวิทยาลัย/สถาบันอุดมศึกษาของรัฐในสังกัดสำนักงานคณะกรรมการการอุดมศึกษาทำหน้าที่เป็นศูนย์สอบ

ด้านเนื้อหาวิชา จำนวนวิชา และชั่วโมงในการสอบ จะเห็นได้ว่า ลักษณะของข้อสอบรวมทั้งเนื้อหา ความยากง่ายของข้อสอบไม่แตกต่างจากระบบการสอบคัดเลือกแบบเดิม แต่เวลาในการสอบ ระบบการสอบคัดเลือกแบบเดิมใช้เวลา 3 ชั่วโมง แต่ระบบการรับเข้าศึกษา (Admissions) ใช้เวลา 2 ชั่วโมง

ด้านจำนวนครั้งในการจัดสอบในระบบการรับเข้าศึกษา (Admissions) กำหนดให้มีการสอบ 2 ครั้ง เพื่อเลือกครั้งที่ดีที่สุดไปใช้ในการคัดเลือกเข้าศึกษาต่อ โดยครั้งแรกในช่วงเดือนตุลาคม ครั้งที่สองช่วงเดือนมีนาคม ในการจัดสอบครั้งแรกนั้นทำให้โรงเรียนต้องเร่งสอนให้นักเรียนเรียนจบหลักสูตรก่อนการสอบครั้งแรก ซึ่งทำให้การเรียนการสอนในระดับมัธยมศึกษาตอนปลาย ไม่เป็นไปตามหลักสูตร และเป็นโอกาสเปิดโอกาสให้เด็กที่เรียนเก่ง เด็กตกค้าง เด็กนอก

ระบบ หรือเด็กกรณีพิเศษอื่นๆ เช่น เด็กที่เรียนจบจากต่างประเทศ สามารถเป็นผู้สอบได้ด้วยมากกว่าจะเป็นการเปิดโอกาสให้นักเรียนที่เรียนในระบบปกติ

3. ประเด็นอื่นๆ

การกำหนดน้ำหนักคะแนนของแต่ละวิชา และความซ้ำซ้อนในการสอบวัดความรู้ในกลุ่มวิชาหลักกับวิชาเฉพาะ

การเก็บคะแนนสอบไว้ใช้ได้ 2 ปี เป็นแนวคิดที่ดี แต่อาจมีปัญหาเกี่ยวกับความเป็นคู่ขนานของแบบสอบที่ใช้ในการสอบแต่ละครั้ง คะแนนที่ได้จากการสอบแต่ละครั้งมีความเที่ยงตรงและความเชื่อมั่นไม่เท่ากัน ซึ่งควรมีการสร้างข้อสอบที่เป็นระบบ มีความเที่ยงตรงและมีความเชื่อมั่นเท่ากัน ควรมีการเทียบคะแนนผลการสอบแต่ละครั้ง เพราะอย่างยิ่งไม่มีการรับรองในความเป็นคู่ขนานของแบบสอบที่ใช้ในแต่ละครั้ง

ในการคัดเลือกไม่ควรคำนึงเฉพาะผลการเรียนและคะแนนสอบเท่านั้น น่าจะมีการให้ความสำคัญต่อคุณสมบัติอื่นๆ ของนักเรียนด้วย เช่น การทำกิจกรรมดีเด่น ความสามารถ/ความถนัดเฉพาะด้านของนักเรียนแต่ละคน ประกอบการพิจารณาในการคัดเลือกด้วย

จากปัญหาและบทวิเคราะห์ดังกล่าว คณะกรรมการการศึกษา สภาผู้แทนราษฎร จึงแสดงความคิดเห็นและข้อเสนอแนะต่อหน่วยงานที่รับผิดชอบการจัดสอบคัดเลือกบุคคลเข้าศึกษาในสถาบันอุดมศึกษาดังนี้

1. ผลการวิเคราะห์ของสำนักงานคณะกรรมการการศึกษาแห่งชาติ ปรากฏว่าคะแนนเฉลี่ยสะสม (GPAX) ของผู้สมัครมีความสัมพันธ์กับคะแนนสอบคัดเลือกเข้ามหาวิทยาลัย การสอบคัดเลือกระบบใหม่มีความสอดคล้องกับในร่างพระราชบัญญัติการศึกษาแห่งชาติ มาตรา 25 เป็นการกระจายโอกาสและสร้างความเสมอภาคทางการศึกษาระดับอุดมศึกษาเพิ่มขึ้น การสอบปีละ 2 ครั้ง ช่วยลดความเครียดของผู้สมัครและผู้ปกครอง ที่สำคัญที่สุด คือ การนำค่าคะแนนเฉลี่ยสะสม ตลอดช่วงชั้นที่ 4 มาคำนวณค่าในการตัดสินใจรับเข้าเรียนสมัครเข้าเรียนสมัครและเกิดการปรับปรุงประสิทธิภาพทางการเรียนการสอนในระดับมัธยมศึกษา จากเหตุผลดังกล่าว จึงเห็นสมควรให้นำผลการประเมินในระดับมัธยมศึกษาตอนปลาย อาจเพิ่มสัดส่วนจากร้อยละ 10 เป็นร้อยละ 15 หรือ ร้อยละ 20 ก็ได้

2. ควรส่งเสริมให้มีการวิจัยหาสูตรการปรับค่าคะแนนเฉลี่ยสะสมของผู้สมัครจากโรงเรียนต่างขนาด เพื่อให้เกิดโอกาสที่เท่าเทียมกัน

3. ควรเร่งรัดการประกันคุณภาพการศึกษาและเกณฑ์มาตรฐานการศึกษาของสถานศึกษา มีการจัดระบบข้อมูล การประเมินมาตรฐานภายในสถานศึกษาทุกสังกัด การประเมินคุณภาพและมาตรฐานจากคณะบุคคลภายนอก และการวางแผนตรวจเยี่ยมสถานศึกษา ทั้งนี้เพื่อ

สร้างความเท่าเทียมกันในโอกาสและลดความเหลื่อมล้ำทางคุณภาพ โรงเรียนที่มีขนาดต่างกัน สถานภาพต่างกัน

4. สืบเนื่องมาจากการแสวงหามาตรฐานของ โรงเรียนทั่วประเทศ อย่างน้อยเป็นมาตรฐานขั้นต่ำที่โรงเรียนต้องมีในระบบการประกันคุณภาพ และมาตรฐานการศึกษา กรมวิชาการ กระทรวงศึกษาธิการเคยสร้างข้อสอบมาตรฐานชั้นมัธยมศึกษาปีที่ 6 เป็นข้อสอบกลาง ที่สามารถวัดความรู้ของนักเรียนได้ทั่วประเทศ ได้มีการทดลองใช้หาประสิทธิภาพของข้อสอบมาตรฐานนี้ไว้แล้ว แต่มิได้นำมาใช้จริง สมควรที่กระทรวงศึกษาธิการจะสนใจและใช้เป็นทางเลือกหนึ่งในการกำหนดมาตรฐานของ โรงเรียนมัธยมศึกษาทั่วประเทศ

5. การคัดเลือกนักเรียนเข้ามหาวิทยาลัยระบบใหม่ไม่ควรส่งผลกระทบต่อหลักสูตรการเรียนการสอนของนักเรียนระดับมัธยมศึกษาตอนปลาย ซึ่งจะทำให้ปรัชญาการศึกษาของการศึกษาระดับมัธยมศึกษาตอนปลาย มีความเสียหายเกิดขึ้นได้เนื่องจาก เมื่อจะมีการกำหนดหลักสูตรการเรียนการสอนระดับมัธยมศึกษาตอนปลายก็จะต้องคำนึงถึงวิธีการคัดเลือกนักเรียนเข้ามหาวิทยาลัยด้วย ซึ่งไม่สอดคล้องกับวัยและความพร้อมของนักเรียนระดับมัธยมศึกษาตอนปลาย อีกทั้งก่อให้เกิดกระบวนการกวัดวิดมากขึ้นกว่าอดีต

ในขณะที่ ธีระพร วีระถาวร (2541 : 7) ได้พิจารณาเกณฑ์ต่างๆ ของการสอบคัดเลือกระบบใหม่ และมีความเห็นว่าผู้ที่เกี่ยวข้องน่าจะพิจารณาอย่างรอบคอบในประเด็นต่อไปนี้

1. มาตรฐานการเรียนการสอนของโรงเรียนต่างๆ ในปัจจุบันยังมีความแตกต่างกันมาก การควบคุมมาตรฐานการให้คะแนนและการปรับตัวดวงน้ำหนัก เมื่อต้องการเปรียบเทียบกันจึงมีความสำคัญอย่างมาก

2. การนำค่าทางสถิติมาใช้ เช่น ค่าเปอร์เซ็นต์ไทล์ จะต้องมีการปรับค่าเมื่อนำมาเปรียบเทียบกัน เพราะโรงเรียนต่างๆ ยังมีมาตรฐานที่แตกต่างกันมาก

3. ลักษณะการกระจายค่าของ GPAX ของผู้สมัครสอบทุกคนมีความน่าจะเป็นน้อยมากที่จะมีการแจกแจงปกติ เพราะความแตกต่างของสาขาวิชา คุณภาพของการคิดคะแนน GPAX และคุณภาพของข้อสอบที่ใช้วัดผลในแต่ละโรงเรียน

4. การนำค่า GPAX ของผู้สมัครสอบทุกคนมาหาค่าเฉลี่ยไม่ทำให้ได้ค่าเฉลี่ยที่จะเป็นตัวแทนที่ดีของค่า GPAX ของผู้สมัครสอบแต่ละคน เพราะความแตกต่างของสาขาวิชา และความแตกต่างของคุณภาพการเรียนการสอนของแต่ละโรงเรียน

5. การควบคุมคุณภาพของข้อสอบความรู้พื้นฐานวิชาการ (วิชาหลัก) และพื้นฐาน วิชาชีพ (วิชาเฉพาะ) เป็นสิ่งจำเป็นอย่างยิ่ง และต้องมีกระบวนการตรวจสอบคุณภาพเป็นระยะๆ เนื่องจากผู้สมัครสอบสามารถเก็บคะแนนที่สอบไว้ใช้ได้ถึง 2 ปี

6. การจะเพิ่มน้ำหนักขององค์ประกอบแรกของการสอบคัดเลือกระบบใหม่จาก 10% เป็นจำนวนที่มากขึ้น ควรจะต้องมีผลสรุปจากงานวิจัยถึงผลดีและผลเสียของระบบใหม่อย่างชัดเจน ก่อนที่เริ่มดำเนินการปกติ

นอกจากนี้ อรุณี อ่อนสวัสดิ์ และคณะ (2551) ได้ศึกษาข้อดีและข้อด้อยของระบบการรับบุคคลเข้าศึกษาในสถาบันอุดมศึกษาในระบบกลาง (Admissions) ได้ข้อค้นพบดังนี้

1. ความคิดเห็นของผู้มีส่วนได้ส่วนเสียที่มีต่อการใช้ระบบการรับบุคคลเข้าศึกษาในสถาบันอุดมศึกษาในระบบกลาง ในระยะ 3 ปีที่ผ่านมา พบว่า

1.1 ข้อดีของระบบกลางการรับบุคคลเข้าศึกษา

1.1.1 ในระยะยาว ระบบแอดมิชชันจะลดความเหลื่อมล้ำของเด็กชนบทกับเด็กในเมือง เป็นการกระจายโอกาสทางการศึกษา และเปิดโอกาสให้มหาวิทยาลัยพัฒนาศักยภาพการจัดการเรียนการสอน เพื่อตอบสนองนักศึกษาที่มีความสามารถแตกต่างกันให้ประสบความสำเร็จในการศึกษา

1.1.2 ลดความสูญเปล่าทางการศึกษา เพราะนักเรียนทราบคะแนนล่วงหน้าและสามารถเลือกเรียนในคณะหรือสาขาวิชาที่เหมาะสมกับความสามารถของตนเอง ในขณะที่เดียวกันทำให้มหาวิทยาลัยได้นักศึกษาที่มีความพร้อมและตั้งใจที่จะเข้าศึกษาในสาขานั้นอย่างแท้จริง

1.1.3 ส่งเสริมให้นักเรียนสนใจศึกษาตามหลักสูตรเพราะต้องตั้งใจศึกษาทุกวิชาในหลักสูตรมัธยมศึกษาตอนปลาย ทำให้นักเรียนอยู่ในระบบโรงเรียนแทนการกวดวิชา และสอบเทียบเพื่อเร่งเข้ามหาวิทยาลัยก่อนเวลาอันควร

1.2 ข้อด้อยของระบบกลางการรับบุคคลเข้าศึกษา

1.2.1 การเปลี่ยนแปลงจากระบบเอ็นทรานซ์เป็นระบบแอดมิชชัน ขาดการรับฟังความคิดเห็นของผู้มีส่วนได้ส่วนเสียก่อนใช้จริง

1.2.2 ขาดสารสนเทศการวิจัยที่สามารถนำมาเป็นข้อมูลสนับสนุนเกี่ยวกับองค์ประกอบและน้ำหนักองค์ประกอบที่ควรนำมาใช้ในระบบแอดมิชชัน

1.2.3 ระบบมีความซับซ้อนเกินไป เพราะการรับเข้าศึกษาใช้หลายองค์ประกอบในสัดส่วนที่ต่างกัน ในแต่ละสาขาและแตกต่างกันในแต่ละมหาวิทยาลัย

1.2.4 การนำคะแนน ONET มาเป็นองค์ประกอบหนึ่งในการรับเข้าศึกษา เป็นการใช้ผิดวัตถุประสงค์ เพราะ การทดสอบทางการศึกษาแห่งชาติ มีวัตถุประสงค์เพื่อวัดประสิทธิภาพการจัดการศึกษาของโรงเรียน ไม่ใช่เพื่อการคัดเลือกบุคคลเข้าศึกษาในมหาวิทยาลัย

1.2.5 โรงเรียนมีมาตรฐานต่างกัน ผลการเรียนรู้เฉลี่ยสะสมในโรงเรียนต่างกัน ย่อมมีมาตรฐานต่างกัน

1.2.6 การนำผลการเรียนเฉลี่ยสะสมตลอดหลักสูตร และผลการเรียนเฉลี่ยสะสมตามกลุ่มสาระการเรียนรู้ มาใช้โดยไม่ผ่านการปรับเทียบให้เป็นเป็นมาตรฐานเดียวกันก่อน ทำให้เกิดความไม่เป็นธรรมกับนักเรียนของโรงเรียนบางแห่งที่เข้มงวดการให้เกรด

1.2.7 ระบบแอดมิชชั่น เป็นระบบที่ยังไม่นิ่งมีการเปลี่ยนแปลงเพื่อปรับปรุงอยู่บ่อยครั้งทำให้นักเรียนที่อยู่ในช่วงของการเปลี่ยนแปลงกติกาได้รับผลกระทบ เช่น การเปลี่ยนจากการสอบ ONET 5 กลุ่มสาระมาเป็น 8 กลุ่มสาระ ทำให้นักเรียนที่จบการศึกษาก่อนหน้านี้จะสมัครเข้ามหาวิทยาลัยปีถัดไปไม่ได้ และยังมีข่าวว่าจะมีการเปลี่ยนแปลงองค์ประกอบและปรับน้ำหนักองค์ประกอบในการรับเข้าอีก ในปีการศึกษา 2553

1.2.8 ระบบแอดมิชชั่น ไม่เปิดโอกาสให้เด็กแก้ตัว เช่น นักเรียนที่ใช้ชีวิตผิดพลาดแล้วจะกลับตัวกลับใจไม่ได้เพราะผลการเรียนที่ตามมาแต่แรก และคะแนน ONET ที่สอบได้ครั้งเดียว

1.2.9 ความไม่พร้อมและความประมาทของระบบใหม่ที่สร้างความผิดพลาด จนเกิดความไม่น่าเชื่อถือ

1.2.10 ขาดกระบวนการนำร่อง

1.2.11 ไม่สามารถลดการกวดวิชาได้

1.2.12 นักศึกษาที่รับเข้าในระบบ Admissions เรียนไม่ผ่านในวิชาต่างๆ สูงกว่าระบบ Entrance ถึง 4 เท่าหรือนักศึกษามีผลการเรียนอ่อนลง

2. ความสัมพันธ์ระหว่างคะแนนองค์ประกอบต่างๆ กับผลการเรียนเฉลี่ยสะสมในระดับมหาวิทยาลัยในกลุ่มต่างๆ พบว่า ทุกกลุ่มสาขาวิชา คะแนน ONET ทุกวิชา คะแนน ANET ทุกวิชา ผลการเรียนเฉลี่ยสะสมกลุ่มสาระการเรียนรู้ทุกกลุ่มสาระ ผลการเรียนเฉลี่ยสะสมตลอดหลักสูตร สัมพันธ์กับผลการเรียนเฉลี่ยสะสมระดับมหาวิทยาลัยปี 2 อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ตัวแปรทุกตัวจึงมีคุณสมบัติเป็นตัวพยากรณ์ผลการเรียนเฉลี่ยสะสมระดับมหาวิทยาลัยได้ แต่เนื่องจากตัวแปรพยากรณ์ต่างมีความสัมพันธ์กันเองสูง จึงควรทำการวิเคราะห์สหสัมพันธ์พหุแบบขั้นบันได ซึ่งผลการวิเคราะห์พบว่า

2.1 กลุ่มวิทยาศาสตร์สุขภาพ ควรใช้คะแนน ONET สังคมศึกษา คะแนน ONET คณิตศาสตร์ ผลการเรียนเฉลี่ยสะสมกลุ่มสาระการเรียนรู้สังคมศึกษา ในสัดส่วน 21:40: 39 สามารถอธิบายผลการเรียนในมหาวิทยาลัยได้ร้อยละ 18.1

2.2 กลุ่มวิทยาศาสตร์กายภาพ ควรใช้ คะแนน ONET วิทยาศาสตร์ คะแนน ANET คณิตศาสตร์ และ GPAX ในสัดส่วน 44 : 18 : 38 สามารถอธิบายผลการเรียนในมหาวิทยาลัยได้ร้อยละ 11.3

2.3 กลุ่มวิศวกรรมศาสตร์ ควรใช้ คะแนน ONET ภาษาไทย คะแนน ONET ภาษาอังกฤษ คะแนน ANET คณิตศาสตร์ และ GPA กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ในสัดส่วน 20: 24: 37: 19 สามารถอธิบายผลการเรียนในมหาวิทยาลัยได้ร้อยละ 25.7

2.4 กลุ่มสถาปัตยกรรมศาสตร์ ควรใช้ คะแนน ONET ภาษาอังกฤษ และ GPAX ในสัดส่วน 39: 61 สามารถอธิบายผลการเรียนในมหาวิทยาลัยได้ร้อยละ 33.3

2.5 กลุ่มเกษตรศาสตร์ ควรใช้คะแนน ONET ภาษาไทย คะแนน ANET วิทยาศาสตร์ และผลการเรียนเฉลี่ยสะสมกลุ่มสาระการเรียนรู้ภาษาไทย ในสัดส่วน 23: 37: 40 สามารถอธิบายผลการเรียนในมหาวิทยาลัยได้ร้อยละ 15.8

2.6 กลุ่มบริหาร พาณิชยศาสตร์ การบัญชี การจัดการการท่องเที่ยวและ เศรษฐศาสตร์ ควรใช้ คะแนน ONET ภาษาไทย คะแนน ONET วิทยาศาสตร์ และ GPAX ใน สัดส่วน 36: 15: 49 สามารถอธิบายผลการเรียนในมหาวิทยาลัยได้ร้อยละ 25.7

2.7 กลุ่มครุศาสตร์/ศึกษาศาสตร์ (การสอน) ควรใช้ คะแนน ONET ภาษาอังกฤษ GPA กลุ่มสาระการเรียนรู้ภาษาไทย และ GPAGroup สาระการเรียนรู้สังคมศึกษา ศาสนา และวัฒนธรรม ในสัดส่วน 12: 6: 82 สามารถอธิบายผลการเรียนในมหาวิทยาลัยได้ ร้อยละ 15

2.8 กลุ่มศิลปกรรม วิจิตรศิลป์และประยุกต์ศิลป์ ควรใช้ คะแนน ONET วิทยาศาสตร์ และ GPAX ในสัดส่วน 43:57 สามารถอธิบายผลการเรียนในมหาวิทยาลัยได้ร้อยละ 50.7

2.9 กลุ่มมนุษยและสังคมศาสตร์ ควรใช้ คะแนน ONET ภาษาไทย คะแนน ONET สังคมศึกษา คะแนน ONET ภาษาอังกฤษ และ GPAX ในสัดส่วน 26 : 17 : 22 : 35 สามารถอธิบาย ผลการเรียนในมหาวิทยาลัยได้ร้อยละ 20.8

3. สรุปผลการสังเคราะห์เอกสารแนวทางการรับบุคคลเข้าศึกษาในสถาบันอุดมศึกษาของ ต่างประเทศ

ระบบการรับนักศึกษาเข้าศึกษาในสถาบันอุดมศึกษาของต่างประเทศ มีความ หลากหลายขึ้นอยู่กับนโยบายของประเทศและแต่ละมหาวิทยาลัย พอดีประมวลองค์ประกอบที่ ใช้ได้ดังนี้

3.1 ผลการเรียนระดับมัธยมศึกษา ประเทศที่ใช้ได้แก่ สิงคโปร์ บรูไน อินเดีย อังกฤษ ฮองกง สกอตแลนด์ ไอร์แลนด์ สวิสเซอร์แลนด์ นิวซีแลนด์ สหรัฐอเมริกา และแคนาดา ซึ่ง ใช้เป็นเงื่อนไขในการรับสมัคร ไม่ได้นำไปรวมกับคะแนนสอบอื่นๆ

3.2 การทดสอบความถนัด หลายประเทศใช้ผลการทดสอบความถนัดทั่วไป (General Scholastic/ Aptitude Test) ซึ่งจะมีหน่วยงานกลางรับผิดชอบดำเนินการ ประเทศที่ใช้ ได้แก่ สวิตเซอร์แลนด์ ออสเตรเลีย นิวซีแลนด์ อิสราเอล สวีเดน และ สหรัฐอเมริกา เป็นต้น

3.3 การทดสอบวิชาเฉพาะสาขา มหาวิทยาลัยบางแห่งกำหนดให้มีการสอบความรู้เฉพาะสาขา (Special requirements) เพื่อสอบวัดความรู้ในเนื้อหาวิชา หรือ ความถนัดในสาขาวิชาชีพ เช่น ความสามารถในการพูดภาษาต่างประเทศ หรือความสามารถในการวาดภาพในสาขาสถาปัตยกรรมศาสตร์ เป็นต้น

3.4 การเขียนเรียงความ มหาวิทยาลัยต่างประเทศ ได้แก่ ประเทศเยอรมันให้ผู้สมัครเขียนเรียงความ ประมาณ 500 คำ เกี่ยวกับเรื่องต่างๆ ไป เช่น แนะนำตัว ทำไมจึงอยากเรียนสาขานี้ ทำไมจึงอยากมาเรียนต่อที่มหาวิทยาลัยนี้ และ ความคิดเห็นอื่น ๆ มีเป้าหมายเพื่อทดสอบความสามารถในการสื่อสารความคิดและการใช้เหตุผล

3.5 จดหมายแนะนำ (Letter of recommendation) มหาวิทยาลัยบางแห่งต้องการจดหมายแนะนำจากครู เช่น ประเทศสหรัฐอเมริกา และประเทศเกาหลี

3.6 การบำเพ็ญประโยชน์ มหาวิทยาลัยบางแห่งนำการมีส่วนร่วมในกิจกรรมการบำเพ็ญประโยชน์ของผู้สมัครมาพิจารณาการคัดเลือกเข้าเรียน เช่น มหาวิทยาลัยบอสตันในประเทศสหรัฐอเมริกา

3.7 การพิจารณาคูณลักษณะอื่นๆ เช่น พิจารณาผลงานนักเรียนจากแฟ้มสะสมงาน (Portfolio) ในสาขาวิชาที่ต้องการจะเรียนต่อ กิจกรรมทางสังคม โดยครูในโรงเรียนเป็นผู้รับรองการทำกิจกรรมดังกล่าว ซึ่งกิจกรรมควรมีความเกี่ยวข้องกับสาขาที่จะเลือกเรียน โดยบางประเทศจะมีองค์กรต่างๆ ที่รับอาสาสมัครนักเรียนเพื่อทำกิจกรรมทางสังคมและออกไปรับรองให้ (พร้อมกับการรับรองจากโรงเรียนด้วย)

3.8 การสัมภาษณ์ มหาวิทยาลัยในต่างประเทศหลายแห่งให้ผู้สมัครเข้าศึกษาในระดับอุดมศึกษามารับการสัมภาษณ์ ก่อนตัดสินใจรับหรือไม่รับ เช่น ประเทศอังกฤษ สหรัฐอเมริกา จีน ญี่ปุ่น

4. แนวทางการรับบุคคลเข้าศึกษาในสถาบันอุดมศึกษาในระบบกลางในอนาคตก

ผู้วิจัยมีข้อเสนอแนะในการคัดเลือกเข้าศึกษาในสถาบันอุดมศึกษาของประเทศไทยในอนาคต ดังนี้

4.1 ด้านองค์ประกอบและน้ำหนักองค์ประกอบ

4.1.1 ควรพิจารณาใช้องค์ประกอบและน้ำหนักองค์ประกอบในแต่ละกลุ่มสาขาตามผลการวิจัยข้อ 2.1-2.9 ซึ่งแต่ละสาขา มีตัวพยากรณ์ 2 ถึง 4 ตัว พยากรณ์ได้ร้อยละ 11.3 ในสาขา

วิทยาศาสตร์สุขภาพ ถึง 50.7 ในสาขาศิลปกรรม วิจารณ์ศิลป์และประยุกต์ศิลป์ นำหนักองค์ประกอบตามผลวิจัยอาจเป็นตัวเลขที่ใช้เป็นแนวทางเบื้องต้น ในทางปฏิบัติคณะกรรมการที่เกี่ยวข้องควรพิจารณากำหนดน้ำหนักให้จ่ายต่อการสื่อสารและการจำ ซึ่งอาจทำให้อำนาจการพยากรณ์ไม่ต่างจากเดิมมากนัก

4.1.2 องค์ประกอบที่ควรใช้ในการรับเข้าศึกษาในอนาคตควรนำ GPAX มาใช้เป็นองค์ประกอบในการรับเข้าศึกษา เนื่องจากผลการวิจัยพบว่า GPAX มีความสัมพันธ์เชิงบวกกับ GPA ในระดับมหาวิทยาลัยสูง และการนำ GPAX มาเป็นองค์ประกอบด้วยนั้น เป็นการพิจารณาความรู้ความสามารถของผู้เรียนตลอดหลักสูตรมัธยมศึกษาตอนปลาย ที่สอดคล้องกับวัตถุประสงค์ของการจัดการศึกษาที่ต้องการให้ผู้เรียน เรียนครบตามหลักสูตร เป็นการส่งเสริมสนับสนุนให้ผู้เรียนมีความกระตือรือร้นในการเรียนในห้องเรียนมากยิ่งขึ้น และเมื่อพิจารณาการรับเข้าศึกษาของต่างประเทศก็พิจารณา GPAX เป็นองค์ประกอบหนึ่งในการรับเข้าศึกษาด้วยเช่นกัน แต่เนื่องจากการเข้าศึกษาในแต่ละสาขานั้นความรู้ความสามารถของผู้เข้าศึกษาที่มีความสามารถต่างกัน ดังนั้นจึงควรกำหนดผลการเรียนเฉลี่ยสะสมกลุ่มสาระการเรียนรู้ (GPA) มาพิจารณาเป็นเงื่อนไขในการรับเข้าศึกษา ทั้งนี้ อาจพิจารณาในสองถึงสี่กลุ่มสาระการเรียนรู้ เพื่อจะทำให้ได้ผู้เรียนที่มีความรู้ความสามารถตรงกับสาขาวิชาที่จะเรียนที่สุด ส่วนจะใช้ผลการเรียนเฉลี่ยสะสมกลุ่มสาระการเรียนรู้ (GPA) ใด ควรให้คณะ / สาขา ได้มีส่วนในการกำหนด พร้อมทั้งเกณฑ์ที่เหมาะสมด้วย

4.1.3 ควรพิจารณาหาตัวพยากรณ์อื่นๆ ที่วัดคุณลักษณะทางสติปัญญาที่ต่างไปจากความรู้ที่ได้รับจากทางโรงเรียน โดยตรง แต่เป็นคุณลักษณะที่เอื้อต่อการเรียนอุดมศึกษา เช่น ความถนัดทั่วไป และความถนัดเฉพาะ ตามที่ประเทศสหรัฐอเมริกาใช้ และประเทศต่างๆ อาทิ สหราชอาณาจักร สิงคโปร์ อิสราเอล สวีเดน พัฒนาใช้ในเวลาต่อมา หรืออาจมีการพิจารณาผลงานนักเรียนจากแฟ้มสะสมงาน (Portfolio) ในสาขาวิชาที่ต้องการจะเรียนต่อ

4.1.4 ควรพิจารณาหาตัวพยากรณ์อื่นๆ ที่วัดคุณลักษณะที่ไม่เกี่ยวข้องกับสติปัญญา ได้แก่ บุคลิกภาพที่เอื้อต่อสาขาวิชาชีพที่เลือก ความสามารถในการปรับตัว แรงจูงใจใฝ่สัมฤทธิ์ หรือการร่วมกิจกรรมทางสังคม เป็นต้น

4.2 ด้านการบริหารจัดการ

4.2.1 รูปแบบการรับควรมีมากกว่า 1 รูปแบบ เพื่อรองรับนักเรียนหรือนักศึกษาที่มีความจำเป็นต่างกัน ดังเช่นปัจจุบันที่มีการรับเข้าระบบกลาง การรับตรงหรือโควตาของแต่ละมหาวิทยาลัยจัดว่าเป็นเรื่องที่ดี แม้ในการรับเข้าระบบกลางก็ควรมีทางเลือกมากกว่า 1 ทาง เนื่องจากผู้สนใจที่จะเข้าเรียนอุดมศึกษามีความหลากหลาย เช่น สำเร็จการศึกษาในประเทศหรือต่างประเทศ กรณีในประเทศก็มีจบจากอาชีวศึกษา จากพระปริยัติธรรม นักศึกษาผู้ใหญ่ หรือการศึกษานอก

ระบบ และการศึกษาตามอัชฌาศัย หรือแม้แต่นักศึกษาที่ออกกลางคันมาสอบใหม่ หากอยู่ในช่วงรอยต่อของการเปลี่ยนแปลงแนวทางการรับเข้าควรมีวิธีการสอบลักษณะอื่น ไว้สำรอง สอดคล้องกับนักวิชาการที่กล่าวว่า วิธีการรับแต่ละแห่งไม่ควรมีวิธีรับเข้าที่ตายตัวหรือเหมือนกันหมด ควรปรับให้สอดคล้องกับแต่ละพื้นที่

4.2.2 การรับสมัครในการสอบแบบทดสอบที่ใช้ยื่นสมัครแอดมิชชั่น เช่น คะแนน ONET หรือแบบทดสอบความถนัด ควรมีหน่วยงานรับสมัครสอบที่เป็นมิตรกับนักเรียน ถึงแม้จะมีการรับสมัครทางระบบอินเทอร์เน็ต โรงเรียนควรจัดให้นักเรียนมีคอมพิวเตอร์ใช้ที่โรงเรียน ดูแลให้นักเรียนสมัครอย่างถูกต้อง ทันเวลา และหน่วยรับสมัครควรมีหน่วยรับเรื่องร้องเรียนทางโทรศัพท์ เพื่อตอบข้อซักถาม เมื่อเกิดอุปสรรค

4.2.3 การดำเนินการสอบควรเพิ่มสนามสอบให้กระจายพื้นที่ไปยังท้องที่ห่างไกลมากขึ้น เพราะมีนักเรียนที่ด้อยโอกาสอีกมาก ที่เข้าถึงระบบแอดมิชชั่นได้น้อยหรือยากลำบากกว่า เพราะเดินทางไกลกว่า เสียค่าใช้จ่ายในการเดินทางและที่พักมากกว่าทั้งๆ ที่ยากจนกว่า

4.2.4 สถาบันทดสอบทางการศึกษาแห่งชาติ ควรจัดข้อสอบหลายๆ ชุด เพื่อไม่ให้ผู้สอบสามารถลอกข้อสอบกันได้

4.2.5 เนื่องจากผลการเรียนเฉลี่ยสะสมตลอดหลักสูตร (GPAX) และผลการเรียนเฉลี่ยสะสมกลุ่มสาระการเรียนรู้ (GPA) ที่มาจากต่างโรงเรียนมีมาตรฐานต่างกัน ควรปรับเทียบผลการเรียนก่อนนำมาใช้เป็นองค์ประกอบในการรับเข้าศึกษา โดยนำผลการสอบทางการศึกษาแห่งชาติขั้นพื้นฐาน (ONET) ที่สถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) ดำเนินการจัดสอบวัดความรู้ความสามารถของผู้เรียนตามหลักสูตร มาเป็นตัวปรับเทียบผลการเรียน โดยไม่ได้นำผลการสอบทางการศึกษาแห่งชาติขั้นพื้นฐานมาใช้เป็นองค์ประกอบในการรับเข้าศึกษา ซึ่งสอดคล้องกับวัตถุประสงค์ของการทดสอบทางการศึกษาแห่งชาติขั้นพื้นฐานที่ต้องการวัดคุณภาพการจัดการศึกษาของโรงเรียน

จากที่กล่าวมาอาจสรุปได้ว่า ปัญหาของระบบการรับเข้าศึกษาที่พบกันมากที่สุดคือ ประเด็นข้อได้แย้งเกี่ยวกับการนำผลการเรียนสะสมตลอดหลักสูตร (GPAX) และผลการเรียนสะสมตามกลุ่มสาระการเรียนรู้ (GPA) มาใช้นั้น นักเรียน ครู และผู้ปกครองในโรงเรียนที่มีชื่อเสียงเห็นว่าไม่มีความเป็นธรรม โดยอ้างว่ามาตรฐานของโรงเรียนมัธยมศึกษาไม่เท่าเทียมกัน มาตรฐานของผลการเรียนสะสมตลอดหลักสูตร (GPAX) และผลการเรียนสะสมตามกลุ่มสาระการเรียนรู้ (GPA) จึงไม่เท่าเทียมกันด้วย ดังจะเห็นได้จากผลการสอบเข้ามหาวิทยาลัยในระบบเดิม (Entrance) ที่นักเรียนจากโรงเรียนที่มีชื่อเสียงซึ่งมีผลการเรียนสะสมตลอดหลักสูตร และผลการเรียนสะสมตามกลุ่มสาระการเรียนรู้ต่ำ แต่สามารถสอบเข้ามหาวิทยาลัยได้ด้วยคะแนนที่สูงกว่านักเรียนจากโรงเรียนเล็ก ๆ ที่มีผลการเรียนสะสม

ตลอดหลักสูตรและผลการเรียนสะสมตามกลุ่มสาระการเรียนรู้สูง นอกจากนี้ผลการประเมินคุณภาพโรงเรียนของสำนักงานรับรองมาตรฐานและประเมินคุณภาพการศึกษา (สมศ.) ก็พบว่า มาตรฐานของโรงเรียนมัธยมศึกษายังมีความแตกต่างกันอยู่ ถึงแม้ว่าผู้บริหารระดับสูงในกระทรวงศึกษาธิการจะยืนยันว่า มาตรฐานของโรงเรียนมัศึกษามีความเท่าเทียมกันก็ตาม และผลการวิจัยของศิริชัย กาญจนวาสี และคณะ (Kanjanavasee, and others. 2009 : 117-124) ที่พบว่า มีโรงเรียนที่ปล่อยเกรดจำนวน 1,225 โรงเรียน คิดเป็นร้อยละ 47.43 ส่วนมากเป็นโรงเรียนขนาดเล็กและขนาดกลาง และโรงเรียนที่คัดเกรด จำนวน 1,235 โรงเรียน คิดเป็นร้อยละ 47.81 เป็นกลุ่มโรงเรียนขนาดใหญ่ ส่วนโรงเรียนที่ให้เกรดเป็นปกติมีเพียง 123 โรงเรียน คิดเป็นร้อยละ 4.76 เท่านั้น

ปัจจัยที่สำคัญที่มีผลต่อการเบี่ยงเบนของผลการเรียนสะสมตลอดหลักสูตร (GPAX) ของนักเรียน ก็คือ มาตรฐานการประเมินผลการเรียนของโรงเรียน เนื่องจากผลการเรียนสะสมตลอดหลักสูตร (GPAX) ซึ่งเป็นผลการเรียนเฉลี่ยที่เกิดจากการประเมินผลการเรียนรู้ในทุกวิชา ทุกชั้นปีของหลักสูตรการศึกษาขั้นพื้นฐาน ช่วงชั้นที่ 4 จึงเป็นผลจากทั้งนโยบายของโรงเรียน การประเมินผลของครูทุกคนในโรงเรียนที่สอนในช่วงชั้นดังกล่าว ส่วนผลการเรียนสะสมตามกลุ่มสาระการเรียนรู้ (GPA) นั้นขึ้นกับมาตรฐานการประเมินผลการเรียนรู้ในแต่ละกลุ่มวิชาของแต่ละโรงเรียน ซึ่งกลุ่มวิชาที่ต่างกันมักจะมีมาตรฐานการประเมินผลที่แตกต่างกัน เมื่อระบบการรับเข้าศึกษานำผลการเรียนสะสมตลอดหลักสูตร (GPAX) และผลการเรียนสะสมตามกลุ่มสาระการเรียนรู้ (GPA) มาใช้จึงควรมีการเปรียบเทียบผลการเรียนก่อน เพื่อความเป็นธรรมและสอดคล้องกับความสามารถที่แท้จริงของผู้เรียน ซึ่งจะลดปัญหาการออกกลางคันของนักศึกษา เนื่องจากสาเหตุไม่สามารถเรียนในสาขาสอบเข้าไปเรียนได้ ซึ่งหลักการเปรียบเทียบผลการเรียนนั้น จะได้กล่าวถึงในหัวข้อต่อไป

2. การเชื่อมโยงคะแนนระหว่างแบบสอบ

2.1 ความเป็นมา

คะแนนจากการประเมินผลทางการศึกษาจะมีความเชื่อมโยงกันในเชิงสถิติ ดังนั้นคะแนนจากการประเมินหนึ่งสามารถแสดงในอีกหน่วยของการประเมินหนึ่งได้ หรือคะแนนจากการประเมิน ทั้งสองสามารถแสดงในมาตรคะแนนเดียวกันได้ การเชื่อมโยงกัน (Linking) ของคะแนนจึงเป็นความสัมพันธ์ระหว่างคะแนนของแบบสอบสองชุด ซึ่งโดยทั่วไปคะแนนที่นำมาเชื่อมโยงกันมักจะมาจากแบบสอบที่วัดในคุณลักษณะเดียวกัน แต่บางครั้งคะแนนที่นำมาเชื่อมโยงกันอาจจะมาจากแบบสอบที่วัดคุณลักษณะที่ต่างกันได้

หากพิจารณาการเชื่อมโยงกันของคะแนนจากแบบสอบต่างชุดกัน อาจเกิดขึ้นได้ในหลากหลายสถานการณ์ ซึ่งอาจเป็นแบบสอบคู่ขนานที่ตั้งใจสร้างขึ้นเพื่อวัดคุณลักษณะเดียวกัน

หรืออาจเป็นแบบสอบที่วัดคุณลักษณะเดียวกันแต่เป็นแบบสอบคนละภาษา หรือแบบสอบที่วัดคุณลักษณะเดียวกันแต่เป็นแบบสอบคนละประเภทกัน หรืออาจเป็นแบบสอบที่วัดคุณลักษณะเดียวกันแต่มีระดับความยากต่างกัน หรือแม้แต่แบบสอบที่วัดคุณลักษณะต่างกันก็อาจนำมาเชื่อมโยงกันได้ เพียงแต่เป้าหมายและวิธีการในการเชื่อมโยงอาจแตกต่างกันไป วัตถุประสงค์ของการเชื่อมโยงคะแนนระหว่างแบบสอบจึงมีความเกี่ยวข้องกับสถานการณ์ที่เป็นเงื่อนไขเหล่านี้ เริ่มจากแนวคิดการเปรียบเทียบกันได้ (comparability) ระหว่างแบบสอบของฟลานาแกน (Flanagan, 1951) ที่ยึดเงื่อนไขการแจกแจงที่คล้ายคลึงกันของคะแนนมาตรฐานจากแบบสอบต่างชุดกันของผู้สอบกลุ่มเดียวกัน แนวคิดการปรับเทียบคะแนน (equating) และการปรับเทียบสเกล (calibration) ของ แองกอฟฟ์ (Angoff, 1971) ที่มองการเชื่อมโยงคะแนนระหว่างแบบสอบที่วัดคุณลักษณะเดียวกันและมีคุณลักษณะทางสถิติเหมือนกัน (เป็นแบบสอบคู่ขนาน) เป็นการปรับเทียบคะแนน (equating) ส่วนการเชื่อมโยงคะแนนระหว่างแบบสอบที่วัดคุณลักษณะเดียวกันแต่มีความยากง่ายและความเชื่อมั่นต่างกัน เป็นการปรับเทียบสเกล (calibration) ตลอดจนแนวคิดที่กำลังได้รับความสนใจในปัจจุบัน คือ การหาความสอดคล้องของ แบบสอบ (concordance) ของโดแรนส์ (Dorans, 2000) ที่ผ่อนคลายข้อตกลงเบื้องต้นเรื่องข้อสอบคู่ขนาน จึงสามารถหาความเชื่อมโยงของคะแนนระหว่างแบบสอบที่วัดคุณลักษณะที่คล้ายคลึงกันได้ จึงนับว่าเป็นแนวทางใหม่ที่ น่าสนใจ และกำลังมีการศึกษาวิจัยเพิ่มขึ้นเรื่อยๆ ทั้งนี้เพราะความโดดเด่นในด้านการนำไปใช้ในสถานการณ์จริง

2.1.1 การเปรียบเทียบกันได้ของฟลานาแกน

ฟลานาแกน (Flanagan, 1951) ได้อธิบายเกี่ยวกับการเชื่อมโยงและการสร้างมาตรฐานคะแนนไว้อย่างลึกซึ้ง เขาใช้คำว่า “การเปรียบเทียบกันได้” (comparability) ในการเชื่อมโยงคะแนนจากแบบสอบต่างชุดกันที่ได้รับมาตรฐานวัดมาเป็นมาตรฐานเดียวกันแล้วมีการแจกแจงของคะแนนคล้ายคลึงกันของประชากรผู้สอบกลุ่มเดียวกัน เขากล่าวว่า “คะแนนจากแบบสอบสองชุดหรือมากกว่า อาจกล่าวได้ว่าสามารถเปรียบเทียบกันได้ ถ้าในกลุ่มประชากรเดียวกัน แบบสอบเหล่านั้นให้คะแนนที่มีการแจกแจงเหมือนกันในกลุ่มประชกรนั้น” เขาได้อธิบายถึงการเปรียบเทียบกันได้ของคะแนนจากแบบสอบชนิดต่างๆ ไว้ดังนี้

แบบสอบต่างชนิดกันที่วัดเรื่องเดียวกันสามารถนำคะแนนมาเปรียบเทียบกันได้ ถ้าแบบสอบเหล่านั้นสามารถใช้แทนกันได้อย่างสมบูรณ์ โดยแบบสอบต่างชนิดกันที่สามารถนำคะแนนมาเปรียบเทียบกันได้นั้นเรียกว่า **แบบสอบที่เท่าเทียมกัน** (Equivalent forms) การเท่าเทียมกันของแบบสอบนั้นไม่เพียงแต่จะมีกระบวนการสร้างที่เหมือนกันมากๆ แล้ว ยังมีคะแนนมาตรฐานที่เท่าเทียมกันด้วย ซึ่งจะทำให้การเปรียบเทียบกันระหว่างคะแนนของแบบสอบมีความถูกต้องยิ่งขึ้น

ฟลานาแกน ยังกล่าวไว้ว่า “แบบสอบที่สามารถใช้แทนกันได้จริงๆนั้น ต้องเป็นแบบสอบที่วัดในฟังก์ชันเดียวกัน ด้วยความถูกต้องที่เท่ากัน และรายงานคะแนนในมาตราที่เปรียบเทียบกันได้ ซึ่งก็คือ แบบสอบที่เท่าเทียมกันนั่นเอง” เขาได้เน้นอย่างมากว่า คะแนนจากแบบสอบต่างชนิดกันที่สามารถใช้แทนกันได้ ต้องถูกสร้างขึ้นมาเหมือนกัน และมาตราของคะแนนต้องสามารถใช้แสดงคะแนนได้ชัดเจน

ในสถานการณ์การเปรียบเทียบคะแนนจากแบบสอบที่วัดต่างคุณลักษณะกัน ฟลานาแกนก็ได้กล่าวถึงเหมือนกัน เขากล่าวว่า “ถ้าผลลัพธ์จากแบบสอบหลายๆฉบับ หลายๆประเภท ถูกนำมาเปรียบเทียบกัน อย่างผลลัพธ์จากชุดของแบบสอบที่ประกอบด้วยแบบสอบหลายๆฉบับ คะแนนที่ได้จากแบบสอบทั้งหลายจำเป็นต้องมีคุณลักษณะบางประการที่สามารถนำมาเปรียบเทียบกันได้” (Flanagan, 1951) ตัวอย่างหนึ่งที่เขาใช้คือ การสร้างมาตรวัดสำหรับแบบสอบในชุดนั้น (เช่น แบบสอบการอ่านและแบบสอบคณิตศาสตร์) ให้มีคุณสมบัติทางการแจกแจงเหมือนกันสำหรับประชากรของผู้สอบกลุ่มหนึ่ง

จะเห็นว่าการเปรียบเทียบกันได้ทั้งสองสถานการณ์ของฟลานาแกน เป็นความสัมพันธ์ระหว่างคะแนนจากแบบสอบที่มีประชากรผู้สอบกลุ่มเดียวกัน ซึ่งเขาบอกว่า การเปรียบเทียบกันได้ของคะแนนจากประชากรทุกกลุ่มนั้น เป็นการเปรียบเทียบกันได้ทั่วไป (general comparability) ระหว่างแบบสอบที่ต่างกัน หรือแม้แต่ระหว่างแบบสอบแบบต่างๆที่วัดเรื่องใดเรื่องหนึ่ง ซึ่งเป็นไปได้ยากมากทั้งในเชิงเหตุผลและในทางปฏิบัติ นอกจากนี้ ฟลานาแกนยังให้ข้อเสนอแนะไว้ว่า คะแนนที่เปรียบเทียบกันได้สำหรับรูปแบบต่างๆของแบบสอบควรมีความจำเพาะของประชากรน้อยกว่าคะแนนที่สามารถเปรียบเทียบกันได้สำหรับแบบทดสอบที่ต่างกัน

ในการอภิปรายถึงประเด็นของการเชื่อมโยงคะแนนจากแบบสอบที่มีความแตกต่างกันในเรื่องของความยากง่าย (อย่างเช่น ระดับความยากของชุดแบบทดสอบผลสัมฤทธิ์ทางการเรียนที่เหมาะสมกับผู้เรียนในระดับชั้นต่างกัน) ฟลานาแกนก็ใช้คำว่า “การเปรียบเทียบกันได้” ในกระบวนการนี้เช่นกัน การพิจารณาความสัมพันธ์ระหว่างการเปรียบเทียบกันได้กับความเชื่อมั่น ฟลานาแกนได้ชี้ให้เห็นว่า ในการสร้างคะแนนที่สามารถเปรียบเทียบกันได้นั้น การแจกแจงของคะแนนจริงควรจะเหมือนกัน ซึ่งเขาได้กล่าวไว้ว่า “ถ้าความเชื่อมั่นของแบบสอบสองฉบับเท่ากัน สำหรับประชากรผู้สอบกลุ่มหนึ่งแล้ว เมื่อนำการแจกแจงของคะแนนจากแบบสอบทั้งสองมาเปรียบเทียบกัน ลักษณะของการแจกแจงจะคล้ายคลึงกัน” ถ้าความเชื่อมั่นของแบบสอบไม่เท่ากัน ฟลานาแกนแนะนำว่า การแจกแจงของคะแนนจริงควรจะปรับสเกลให้สามารถเปรียบเทียบกันได้

สำหรับการใช้วิธีการถดถอยในการหาความสัมพันธ์ของคะแนนจากแบบสอบต่างกันนั้น Flanagan เห็นว่ากระบวนการถดถอย “ไม่ได้ทำให้คะแนนที่สามารถเปรียบเทียบกันได้” เนื่องจาก

การวิเคราะห์ถดถอย คะแนนจากแบบสอบหนึ่งอาจพยากรณ์คะแนนจากอีกแบบสอบหนึ่งได้ แต่ไม่สามารถแปลความหมายของคะแนนพยากรณ์เหล่านั้น ได้อย่างมีเหตุผล เช่น คะแนนการอ่านสัมพันธ์กับคะแนนคณิตศาสตร์ เป็นต้น

โดยสรุปแล้ว ฟลานาแกน ได้แยกกระบวนการเชื่อมโยงคะแนนจากแบบสอบแบบต่างๆที่สร้างขึ้นเพื่อวัดคุณลักษณะเดียวกัน ออกจากการเชื่อมโยงคะแนนจากแบบสอบที่วัดคุณลักษณะต่างกัน ฟลานาแกนได้อธิบายถึงกระบวนการเชื่อมโยงคะแนนจากแบบสอบที่วัดคุณลักษณะเดียวกันแต่มีความยากง่ายต่างกัน นำเอาความเชื่อมั่นของแบบสอบมาใช้กับแนวคิดการเปรียบเทียบกันได้ของคะแนน แยกฟังก์ชันการเชื่อมโยงที่สมมาตรออกจากฟังก์ชันการถดถอยที่ไม่สมมาตร และเห็นว่าฟังก์ชันการเชื่อมโยงทั้งหมดนั้นขึ้นอยู่กับกลุ่มประชากรของผู้สอบ สถานการณ์และประเด็นต่างๆที่ ฟลานาแกน ได้พิจารณายังคงเป็นที่สนใจกันอยู่ในปัจจุบัน และวิธีการเชื่อมโยงคะแนนแบบต่างๆที่ใช้กันอยู่ก็ล้วน ได้รับผลมาจากแนวคิดของฟลานาแกน

2.1.2 การปรับเทียบคะแนนและการปรับเทียบสเกลของแองกอฟฟ์

แองกอฟฟ์ (Angoff, 1971) ได้สรุปและจัดระบบงานเกี่ยวกับการเชื่อมโยงและการกำหนดมาตรฐานคะแนน (scaling) โดยเขาใช้คำว่า “การปรับเทียบคะแนน” (equating) ในการเชื่อมโยงคะแนนของแบบสอบแบบต่างๆที่สร้างขึ้นเพื่อวัดคุณลักษณะเดียวกัน แต่ไม่เหมือนกับแนวคิดของฟลานาแกน เพราะ แองกอฟฟ์เห็นว่า ความสัมพันธ์ของการปรับเทียบคะแนนนั้นควรเป็นอิสระจากกลุ่มประชากรของผู้สอบ

แองกอฟฟ์ได้ใช้คำว่า “การปรับเทียบสเกล” (calibration) ในกรณีที่เป็นการเชื่อมโยงคะแนนจากแบบสอบที่วัดคุณลักษณะเดียวกัน แต่มีความยากง่ายหรือความเชื่อมั่นต่างกัน ดังนั้นการเชื่อมโยงคะแนนระหว่างแบบทดสอบต่างๆจากชุดของแบบสอบวัดผลสัมฤทธิ์ทางการเรียนของโรงเรียนประถมศึกษา ที่ออกแบบสำหรับนักเรียนระดับชั้นที่ต่างกัน ซึ่งมักจะเรียกว่า “การกำหนดมาตรฐานวัดในแนวตั้ง” (vertical scaling) ควรจะเป็นการปรับเทียบสเกลตามวิธีการของแองกอฟฟ์

แองกอฟฟ์ได้จำกัดการใช้คำว่า “การเปรียบเทียบกันได้” ให้แคบลง โดยแองกอฟฟ์ ใช้ในความหมายของการเชื่อมโยงคะแนนระหว่างแบบสอบที่วัดคุณลักษณะต่างกัน ถึงแม้ว่าฟลานาแกนจะใช้คำนี้ในความหมายของการเชื่อมโยงคะแนนจากแบบสอบใดๆที่มีการแจกแจงคล้ายคลึงกันของคะแนนที่ถูกปรับสเกลแล้วก็ตาม

ในสถานการณ์ต่างๆที่ฟลานาแกน ได้กล่าวถึง แองกอฟฟ์ได้อธิบายว่า การเปรียบเทียบกันได้จะเกิดขึ้นเมื่อคะแนนจากแบบสอบฉบับต่างๆในชุดของแบบสอบถูกปรับสเกลให้มีการแจกแจงเหมือนกัน ในการประเมินความความเก่งและความอ่อนของผู้สอบเทียบกับกลุ่มปกติวิสัย (norm group)

2.1.3 การเชื่อมโยงทางสถิติ 4 แบบของมิสเลวีและลินน์

มิสเลวี (Mislevy, 1992) และลินน์ (Linn, 1993) ได้พัฒนากรอบแนวคิดการเชื่อมโยงกันในเชิงสถิติของคะแนนจากแบบสอบ สรุปได้เป็น 4 แบบด้วยกันคือ การปรับเทียบคะแนน (equating) การปรับเทียบสเกล (calibration) การใช้ตัวกลางเชิงสถิติ (statistical moderation) และการสร้างภาพอนาคต หรือการพยากรณ์ (projection or prediction) โดยการปรับเทียบคะแนนและการปรับเทียบสเกลนั้นใช้ในความหมายตามที่แองกอฟฟ์นิยามไว้ คือ การปรับเทียบคะแนนเป็นการเชื่อมโยงคะแนนของแบบสอบแบบต่างๆที่สร้างขึ้นเพื่อวัดคุณลักษณะเดียวกันและมีคุณลักษณะทางสถิติเหมือนกัน ส่วนการปรับเทียบสเกล เป็นการเชื่อมโยงคะแนนจากแบบสอบที่วัดคุณลักษณะเดียวกัน แต่มีความยากง่ายหรือความเชื่อมั่นต่างกัน

ส่วนการสร้างภาพอนาคต และการใช้ตัวกลางทางสถิติ ตามกรอบแนวคิดของมิสเลวีและลินน์นั้น เป็นการเชื่อมโยงกันเมื่อแบบสอบนั้นวัดในคุณลักษณะที่ต่างกัน โดยการสร้างภาพอนาคตเป็นการพยากรณ์คะแนนในแบบสอบฉบับหนึ่งจากคะแนนของแบบสอบอีกฉบับหนึ่ง โดยใช้วิธีการวิเคราะห์การถดถอย การนำการสร้างภาพอนาคตไปใช้ผู้สอบส่วนหนึ่งต้องได้ทำแบบสอบทั้งสองฉบับที่นำมาเชื่อมโยงกัน

ในขณะที่คำว่า “การใช้ตัวกลางทางสถิติ” เป็นคำที่เคียฟ (Keeve, 1990) ใช้ ซึ่งเป็นการเชื่อมโยงคะแนนระหว่างแบบสอบสองฉบับ โดยเชื่อมโยงผ่านแบบสอบฉบับที่สามหรือตัวแปรกลาง (moderator) ดังนั้นผู้สอบกลุ่มหนึ่งจะต้องได้ทำแบบสอบฉบับหนึ่งกับฉบับกลาง และผู้สอบอีกกลุ่มหนึ่งจะต้องทำแบบสอบอีกฉบับหนึ่งกับฉบับกลาง

มิสเลวีและลินน์ได้เน้นย้ำว่า การสร้างภาพอนาคต และการใช้ตัวกลางทางสถิตินั้นไม่เป็นอิสระจากกลุ่มผู้สอบ สำหรับการใช้ตัวกลางทางสถิติ การแปลงรูปคะแนนที่เขาพัฒนาขึ้นยังคงมีความสมมาตร แต่การสร้างภาพอนาคตนั้น การแปลงรูปคะแนนจะไม่มี ความสมมาตร

จะเห็นว่าตามกรอบแนวคิดของมิสเลวีและลินน์คะแนนจากแบบสอบใดๆก็สามารถนำมาเชื่อมโยงกันแบบการสร้างภาพอนาคต และการใช้ตัวกลางทางสถิติได้ トラบเท่าที่ข้อมูลที่เกี่ยวข้องมารวมมา มีความเหมาะสม โดยปราศจากเงื่อนไขของแบบสอบว่าจะวัดคุณลักษณะเดียวกันหรือไม่ก็ตาม

2.1.4 ปัจจัยที่มีผลต่อการเชื่อมโยงคะแนนของฟูเออร์และคณะ

ในปี 1999 The National Research Council ได้ทำการศึกษาความเป็นไปได้ของการเชื่อมโยงคะแนนของแบบสอบวัดผลสัมฤทธิ์ทางการเรียนที่จัดสอบโดยรัฐ กับคะแนนของแบบสอบ NAEP (National Assessment of Educational Progress) การรายงานขั้นสุดท้ายที่จัดทำโดย

ฟูเออร์ และคณะ (Feuer and others, 1999) ได้กล่าวถึงปัญหาของการเชื่อมโยงคะแนนจากแบบสอบต่างๆ และการปรับปรุงและพัฒนากรอบแนวคิดต่อเนื่องมาจากกรอบแนวคิดของมิสเลวีและลินน์

เพื่อให้สามารถศึกษาถึงผลของความแตกต่างกันของเนื้อหาของแบบสอบที่นำมาเชื่อมโยงกัน ฟูเออร์และคณะได้เน้นที่กระบวนการพัฒนาแบบสอบซึ่งมี 3 ขั้นตอนดังนี้

1. กำหนดเค้าโครงของแบบสอบ (framework definition) เป็นการกำหนดขอบเขตและส่วนประกอบของแบบสอบ เช่น ขอบเขตเนื้อหาและทักษะที่ต้องการวัด เป็นต้น

2. กำหนดคุณลักษณะของแบบสอบ หรือพิมพ์เขียว (test specification or blueprint) เป็นการกำหนดสัดส่วนของเนื้อหา และรูปแบบของข้อสอบ จำนวนงานหรือข้อสอบ กฎเกณฑ์การให้คะแนน ฯลฯ

3. คัดเลือกข้อสอบ (item selection) เป็นการคัดเลือกข้อสอบที่เป็นตัวแทนของคุณลักษณะที่ต้องการวัดให้มีความเชื่อถือได้มากที่สุดเท่าที่จะเป็นไปได้

จากกระบวนการพัฒนาแบบสอบดังกล่าว ฟูเออร์และคณะได้นิยามวิธีการเชื่อมโยงคะแนนใหม่ โดยให้การปรับเทียบคะแนน (equating) เป็นการเชื่อมโยงคะแนนระหว่างแบบสอบ เมื่อแบบสอบนั้นสร้างโดยใช้เค้าโครงของแบบสอบและคุณลักษณะของแบบสอบอันเดียวกัน การปรับเทียบสเกล (calibration) เป็นการเชื่อมโยงคะแนนระหว่างแบบสอบที่สร้างโดยใช้เค้าโครงของแบบสอบอันเดียวกัน แต่มีคุณลักษณะของแบบสอบต่างกัน ตัวอย่างเช่น การปรับสเกลในแนวตั้ง ส่วนการใช้ตัวกลาง หรือการฉายภาพ เป็นการเชื่อมโยงคะแนนระหว่างแบบสอบที่มีเค้าโครงแบบสอบต่างกัน

จากผลการศึกษา ฟูเออร์และคณะได้สรุปว่ามีปัจจัยมากมายที่มีผลต่อความเที่ยงตรงของการเชื่อมโยงคะแนน ซึ่งอาจพอสรุปได้ดังนี้ คือ 1) ความคล้ายคลึงกันของเนื้อหา ความยากง่าย และรูปแบบของข้อกระทง 2) การเปรียบเทียบกันได้ของความคลาดเคลื่อนการวัดที่เกี่ยวข้องกับคะแนน 3) เส้นใยการบริหารจัดการการสอบ 4) การใช้ผลที่ได้จากการสอบและสิ่งที่เกิดจากการใช้ผลของการสอบ รวมทั้งผู้ที่เกี่ยวข้องกับผลของการสอบ และ 5) ความถูกต้องและความคงที่ของการเชื่อมโยง รวมทั้งความคงที่ในระหว่างกลุ่มย่อยๆของผู้สอบ

2.1.5 การหาความสอดคล้องของโดแรนส์

โดแรนส์ (Dorans, 2004) ได้ทำการแยกการเชื่อมโยงคะแนนจากแบบสอบที่วัดคุณลักษณะต่างกันออกจากการเชื่อมโยงคะแนนจากแบบสอบที่วัดคุณลักษณะที่คล้ายคลึงกัน เมื่อปี 2000 โดย โดแรนส์ใช้คำว่า “การหาความสอดคล้อง” (concordance) ในกรณีที่เป็นการเชื่อมโยงคะแนนระหว่างแบบสอบที่วัดคุณลักษณะคล้ายคลึงกัน มีการแจกแจงคะแนนคล้ายกัน โดแรนส์เสนอแนะว่า การหาความสอดคล้อง จะใช้เมื่อแบบสอบวัดคุณลักษณะคล้ายคลึงกัน ซึ่งการพิจารณาความคล้ายคลึงกันของเนื้อหาแบบสอบนั้นดูจากคะแนนของแบบสอบเหล่านั้นมีความสัมพันธ์กัน

สูง และความสัมพันธ์ของการเชื่อมโยงแตกต่างกันน้อยมากระหว่างกลุ่มผู้สอบ เขาอธิบายว่า เมื่อแบบสอบมีความสัมพันธ์กันตามเกณฑ์ดังกล่าว เราก็คาดหมายได้ว่าคะแนนจากแบบสอบทั้งสองสามารถนำมาเปรียบเทียบกันได้ ซึ่งจะมีประโยชน์สำหรับการตัดสินใจได้อย่างถูกต้อง และเหมาะสมสำหรับกรณีที่มีกลุ่มผู้สอบหลายกลุ่ม เขาสรุปว่า สำหรับวิธีการวิเคราะห์การถดถอยนั้นควรนำมาใช้ในกรณีที่การเชื่อมโยงคะแนนระหว่างแบบสอบที่ไม่มีความสัมพันธ์กันตามวิธีการหาความสอดคล้อง

2.1.6 สถานการณ์การเชื่อมโยงคะแนนของโคลเลนและเบรนแนน

ตามที่ พูเออร์และคณะ ได้ให้ข้อเสนอแนะว่า สถานการณ์การเชื่อมโยงคะแนนนั้น มีหลากหลายสถานการณ์ ซึ่งล้วนแต่มีความสำคัญต่อการประเมินการเชื่อมโยงกันระหว่างแบบสอบ โคลเลนและเบรนแนน (Kolen & Brennan, 2004) ได้เสนอกรอบแนวคิดในการจัดกลุ่มสถานการณ์การเชื่อมโยงเหล่านี้ ซึ่งเรียกว่า ลักษณะของสถานการณ์การเชื่อมโยง โดยโคลเลนและเบรนแนน ได้เสนอลักษณะของสถานการณ์การเชื่อมโยงไว้ 4 ประการคือ

1. การอนุมาน (inferences) เป็นขอบเขตของการนำคะแนนจากแบบสอบทั้งสองไปใช้สรุปและอ้างอิงในเรื่องที่คล้ายคลึงกัน
2. คุณลักษณะที่วัด (Constructs) เป็นขอบเขตของทักษะและเนื้อหาที่แบบสอบทั้งสองวัดเหมือนกัน
3. ประชากร (Populations) เป็นขอบเขตของกลุ่มผู้สอบที่แบบสอบทั้งสองออกแบบเพื่อใช้กับกลุ่มผู้สอบเหมือนกัน
4. เงื่อนไขการวัด (measurement conditions) เป็นขอบเขตของการดำเนินการสอบที่แบบสอบทั้งสองมีเงื่อนไขร่วมกัน ซึ่งรวมถึงความยาวของแบบสอบ รูปแบบของแบบสอบ เงื่อนไขการบริหารการสอบ ฯลฯ

โคลเลนและเบรนแนน ยังได้อธิบายถึงลักษณะของสถานการณ์เหล่านี้เกี่ยวข้องกับการจัดกลุ่มของการเชื่อมโยงคะแนนอย่างไร ตัวอย่างเช่น การปรับเทียบคะแนน (Equating) คะแนนจากแบบสอบต่างๆที่จะนำมาปรับเทียบคะแนนกัน จะต้องเป็นแบบสอบที่มีการอนุมานเหมือนกัน และวัดในคุณลักษณะเดียวกัน ถูกออกแบบมาเพื่อใช้กับประชากรเดียวกัน และมีเงื่อนไขการบริหารจัดการการสอบเหมือนกัน ส่วนการหาความสอดคล้องนั้น คะแนนที่นำมาเชื่อมโยงกันมาจากแบบสอบที่มีการอนุมานและวัดในคุณลักษณะที่คล้ายคลึงกัน ถูกออกแบบมาเพื่อใช้กับประชากรคล้ายกัน และมีเงื่อนไขการบริหารจัดการการสอบที่คล้ายคลึงกัน

จากกรอบแนวคิดในการเชื่อมโยงคะแนนที่ได้กล่าวมาจะเห็นว่า กระบวนการปรับเทียบคะแนน (Equating) นั้นจะถูกแยกออกมาอย่างชัดเจน คือเป็นการเชื่อมโยงคะแนนระหว่างแบบสอบ

ที่สร้างขึ้นโดยมีเนื้อหาและคุณลักษณะทางสถิติเหมือนกัน ซึ่งคะแนนที่ได้จากการเปรียบเทียบคะแนนนั้นสามารถใช้แทนกันได้ ถึงแม้การเปรียบเทียบคะแนนจะถูกออกแบบมาเพื่อใช้กับกลุ่มผู้สอบกลุ่มเดียวกัน แต่ก็สามารถใช้กับกลุ่มผู้สอบที่มีความคล้ายคลึงกันได้ (Kolen, 2004)

การปรับเทียบสเกล (Calibration) เป็นกระบวนการเชื่อมโยงคะแนนจากแบบสอบที่วัดในคุณลักษณะเดียวกัน แต่มีความยากง่ายและความเชื่อมั่นต่างกัน แบบสอบที่สามารถนำคะแนนมาปรับเทียบสเกลร่วมกันจะสร้างมาจากกรอบเนื้อหาเดียวกัน เพื่อให้แน่ใจได้ว่าแบบสอบทั้งสองวัดในคุณลักษณะเดียวกัน ตัวอย่างหนึ่งของการปรับเทียบสเกลก็คือ การปรับเทียบสเกลระหว่างแบบสอบฉบับย่อ (Short version) กับฉบับยาว (Long version) ซึ่งฉบับย่อจะมีจำนวนข้อสอบในแต่ละกลุ่มเนื้อหาเป็นสัดส่วนเท่ากับฉบับยาว อีกตัวอย่างหนึ่งก็คือ การปรับเทียบสเกลในแนวตั้ง อย่างเช่น การเชื่อมโยงคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์ระหว่างระดับชั้นต่างๆ เป็นต้น

ถ้าเป็นชุดของแบบสอบที่ประกอบด้วยแบบสอบหลายๆฉบับ มักจะปรับสเกลคะแนนของแต่ละฉบับให้มีลักษณะการแจกแจงเหมือนกันสำหรับประชากรผู้สอบกลุ่มหนึ่ง กระบวนการปรับสเกลคะแนนดังกล่าวเรียกว่า การปรับสเกลของชุดแบบสอบ (Test battery scaling) ตามวิธีการเชื่อมโยงคะแนนของฟลานาแกน (Flanagan, 1951)

เมื่อนำคะแนนจากแบบสอบที่วัดคุณลักษณะต่างกันมาเชื่อมโยงกัน วิธีการและสถานการณ์การเชื่อมโยงจะยิ่งมีความซับซ้อนมากขึ้น มิสเลวีและลินน์เสนอให้ใช้การวิเคราะห์ถดถอยที่เรียกว่า การฉายภาพหรือการพยากรณ์ ส่วนการเชื่อมโยงคะแนนระหว่างแบบสอบที่มีคุณลักษณะการแจกแจงที่คล้ายคลึงเรียกว่า การใช้ตัวกลางเชิงสถิติ

ในกรณีที่แบบสอบวัดในคุณลักษณะที่แตกต่างกัน การเชื่อมโยงคะแนนจะเกิดขึ้นในหลายๆสถานการณ์ ซึ่งการปรับสเกลของชุดแบบสอบก็เป็นตัวอย่างหนึ่งในสถานการณ์เหล่านั้น ซึ่งวิธีการที่มีความชัดเจนจะให้ผลลัพธ์ที่เป็นประโยชน์ต่อผู้ใช้แบบสอบในการเปรียบเทียบความเก่งและความอ่อนของผู้สอบ ตัวอย่างที่สองคือ การเชื่อมโยงคะแนนระหว่างแบบสอบของแต่ละรัฐกับแบบสอบ NAEP ดังที่ ฟุเออร์และคณะ (Feuer and others, 1999) ได้ทำการศึกษา แต่การเชื่อมโยงแบบนี้ไม่น่าจะมีประโยชน์มากนัก เพราะว่าแบบสอบของแต่ละรัฐมีความแตกต่างจากแบบสอบ

NAEP มาก ทั้งในส่วนของคุณลักษณะที่ต้องการวัด การใช้แบบสอบ เงื่อนไขการบริหารการสอบ ฯลฯ อีกตัวอย่างหนึ่งคือ การเชื่อมโยงคะแนนระหว่างชุดแบบสอบ ACT กับ SAT I ซึ่งจุดมุ่งหมายของแบบสอบทั้งสองเหมือนกัน ผลที่ตามมาก็เหมือนกัน มีเงื่อนไขการบริหารการสอบที่คล้ายคลึงกัน และเนื้อหาคล้ายคลึงกัน ดังนั้นการเชื่อมโยงคะแนนระหว่างแบบสอบจากชุดแบบสอบทั้งสองชุดนี้น่าจะมีประโยชน์มากกว่าจากกรอบแนวคิดในการเชื่อมโยงคะแนนที่กล่าวมา มีเพียงแนวคิด

ของ โดแรนส์ (Dorans, 2004) ที่แยกแยะการเชื่อมโยงคะแนนจากแบบสอบที่วัดคุณลักษณะคล้ายคลึงกันออกจากแบบสอบที่วัดคุณลักษณะแตกต่างกัน

ฟูเออร์และคณะ (Feuer and others, 1999) ได้แยกแยะประเด็นต่างๆของการเชื่อมโยงคะแนนออกมาอย่างชัดเจน หนึ่งในนั้นคือการเชื่อมโยงในเชิงสถิติ ซึ่งควรจะใช้ในกรณีของการเชื่อมโยงคะแนนจากแบบสอบที่วัดในคุณลักษณะที่ต่างกัน ส่วน โคลเลนและเบรนแนน (Kolen & Brennan, 2004) ได้ระบุ 4 ลักษณะที่ต้องพิจารณาเกี่ยวกับสถานการณ์ของการเชื่อมโยงคะแนนที่สามารถใช้ในการประเมินความเชื่อมโยงคะแนนในบริบทต่างๆได้อย่างหลากหลาย

2.1.7 ทิศทางการวิจัยเกี่ยวกับการหาความสอดคล้อง

โคลเลน (Kolen, 2004) ได้ให้ความหมายของ “การหาความสอดคล้อง” (concordance) ว่าเป็นการเชื่อมโยงคะแนนจากการประเมินที่วัดคุณลักษณะที่คล้ายคลึงกัน (ไม่จำเป็นต้องเป็นคุณลักษณะเดียวกัน) และเป็นคะแนนที่นำมาใช้เพื่อการตัดสินใจเรื่องใดเรื่องหนึ่งเฉพาะ เช่น ในการคัดเลือกนักเรียนเข้าศึกษาต่อในระดับอุดมศึกษา อาจพิจารณาจากคะแนน O-NET, A-NET หรืออาจใช้เกรดเฉลี่ยร่วมกับคะแนนสอบความถนัดแทนก็ได้ ซึ่งถือว่าเป็นสถานการณ์ที่สอดคล้องกัน เพื่อจุดมุ่งหมายเฉพาะการคัดเลือกนักเรียนเข้าศึกษาต่อเท่านั้น หรือการเชื่อมโยงคะแนนที่เลือกมาจากแบบสอบวัดผลสัมฤทธิ์ต่างสำนักพิมพ์กัน แต่ครอบคลุมเนื้อหาคล้ายคลึงกัน หรือการเชื่อมโยงคะแนนระหว่างแบบสอบที่ปรับปรุงใหม่กับแบบสอบเดิม โดยฉบับปรับปรุงใหม่นั้นสร้างจากคุณลักษณะของข้อสอบที่แตกต่างจากฉบับดั้งเดิม

คุณภาพและประโยชน์ของการหาความสอดคล้องนั้น ขึ้นอยู่กับความคล้ายคลึงกันของแบบสอบและสถานการณ์การทดสอบ ซึ่งโคลเลนและเบรนแนน (Kolen & Brennan, 2004) ได้ระบุไว้ 4 ลักษณะ ดังที่ได้กล่าวมาแล้ว การพิจารณาความคล้ายคลึงกันของแบบสอบสองฉบับจึงควรมองถึงการอนุมาน เนื้อหาและคุณลักษณะที่ต้องการวัด กลุ่มประชากรเป้าหมาย และเงื่อนไขของการวัด

กระบวนการตัดสินใจจะประกอบด้วยการวิเคราะห์ลักษณะของการเชื่อมโยงดังกล่าว อย่างเช่น จุดมุ่งหมายของแบบสอบสองฉบับจะสามารถนำมาเปรียบเทียบกันได้หรือไม่ นั่นก็ต้องการการวิเคราะห์การอนุมาน การเปรียบเทียบกันอย่างเป็นระบบในเรื่องเนื้อหาของแบบสอบซึ่งเป็นสิ่งสำคัญอีกอย่างหนึ่งในการเปรียบเทียบแบบสอบ ก็จะได้จากการวิเคราะห์คุณลักษณะที่ต้องการวัด การวิเคราะห์อย่างละเอียดถี่ถ้วนถึงเงื่อนไขในการวัด ไม่ว่าจะเป็นการวิเคราะห์การบริหารการสอบ หรือผลประโยชน์ที่เกี่ยวข้องกับผลการทดสอบ เป็นต้น

ประเด็นหนึ่งที่ต้องทำการวิจัยเกี่ยวกับการหาความสอดคล้องก็คือ การวิเคราะห์ในเชิงสถิติที่มีประโยชน์ต่อการประเมินความคล้ายคลึงกันของแบบสอบที่จะนำมาหาความสอดคล้องกัน ไม่ว่าจะใช้การประมาณค่าสหสัมพันธ์ระหว่างคะแนนดิบหรือระหว่างคะแนนจริงของแบบสอบ หรือใช้

โมเดลสมการโครงสร้าง (SEM) หรือใช้ทฤษฎีสรุปร่างทั่วไป (G-Theory) เป็นต้น อีกประเด็นหนึ่งที่น่าสนใจก็คือ ความเชื่อมั่นของแบบสอบและรูปแบบของความคลาดเคลื่อนมาตรฐานการวัดของแบบสอบที่จะนำมาเชื่อมโยงกันควรมีเกณฑ์อย่างไร รวมทั้งการหาวิธีการทางสถิติที่จะช่วยวิเคราะห์หาขอบเขตการเชื่อมโยงที่ไม่แปรเปลี่ยนไปตามกลุ่มประชากร ซึ่งวิธีการต่างๆ เหล่านี้รวมทั้งดัชนีทางสถิติและกราฟจะช่วยในการตัดสินใจเกี่ยวกับการเชื่อมโยงคะแนน โดยการหาความสอดคล้อง

การเชื่อมโยงคะแนนระหว่างแบบสอบนั้นมียากมายหลายสถานการณ์ ทำให้เกิดวิธีการเชื่อมโยงขึ้นหลายแบบ การเชื่อมโยงคะแนนแต่ละแบบก็อาจมีวัตถุประสงค์แตกต่างกันไป และอาจมีเทคนิควิธีย่อยๆ อีกมากมาย เพื่อให้การเชื่อมโยงคะแนนสอดคล้องกับสถานการณ์เหล่านั้นมากที่สุด และตอบสนองวัตถุประสงค์ของการเชื่อมโยงนั้นมากที่สุด

2.2 การปรับเทียบคะแนน (Score-Equating)

การปรับเทียบคะแนนระหว่างแบบสอบ เป็นการปรับคะแนนของผู้สอบแต่ละคนที่สอบแบบสอบต่างฉบับกันให้สามารถเปรียบเทียบกันได้โดยตรง เพื่อให้เกิดความยุติธรรมระหว่างผู้สอบ มีผู้ที่ให้ความหมายของการปรับเทียบคะแนนไว้หลายท่าน จะขอนำเสนอดังนี้

แองกอฟฟ์ (Angoff, 1971) กล่าวถึงการปรับเทียบคะแนนไว้ว่า เป็นกระบวนการแปลงคะแนนจากแบบสอบฉบับหนึ่งไปสู่แบบสอบอีกฉบับหนึ่งที่วัดคุณลักษณะเดียวกัน เพื่อให้คะแนนแปลงจากแบบสอบทั้งสองฉบับเทียบเท่ากัน และสามารถใช้คะแนนสอบจากแบบสอบหนึ่งแทนแบบสอบอีกฉบับหนึ่งได้

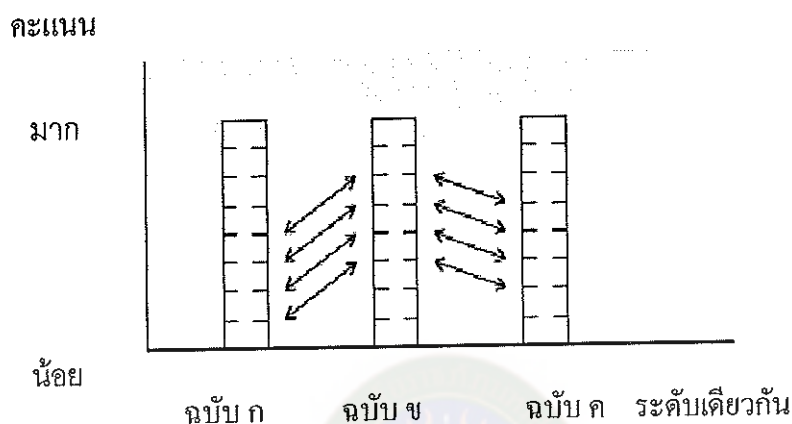
คร็อกเกอร์และอัลจินา (Crocker & Algina, 1986 : 457) กล่าวถึงการปรับเทียบคะแนนไว้ว่า เป็นกระบวนการที่ทำให้คะแนนจากแบบสอบสองชุดมีความสมมูลกันสามารถนำมาเปรียบเทียบกันได้

ศิริชัย กาญจนวาสี (2550 :158) ได้ให้ความหมายว่า “การปรับเทียบคะแนน หมายถึง การแปลงระบบคะแนนของแบบสอบต่างฉบับที่วัดเนื้อหาเดียวกัน โดยแปลงคะแนนของแบบสอบฉบับหนึ่งให้อยู่บนมาตรฐานหรือสเกลของแบบสอบอีกฉบับหนึ่ง ที่ถือเป็นคะแนนที่สมมูลกัน (equivalent scores) เพื่อทำให้คะแนนจากแบบสอบต่างฉบับกันสามารถเปรียบเทียบกันได้โดยตรง”

จากที่กล่าวมาแล้วพอสรุปได้ว่า การปรับเทียบคะแนนหมายถึงการแปลงระบบคะแนนสอบของแบบสอบต่างฉบับกันที่วัดเนื้อหาเดียวกันและมีค่าสถิติของแบบสอบเท่ากัน ให้สามารถเปรียบเทียบกันได้

การสอบที่ต้องการเทียบคะแนนจากแบบสอบฉบับหนึ่งไปสู่มาตรฐานหรือสเกลคะแนนของฉบับอื่นๆ สามารถจำแนกได้เป็น 2 สถานการณ์ (Hambleton & Swaminathan, 1987 : 197) คือ

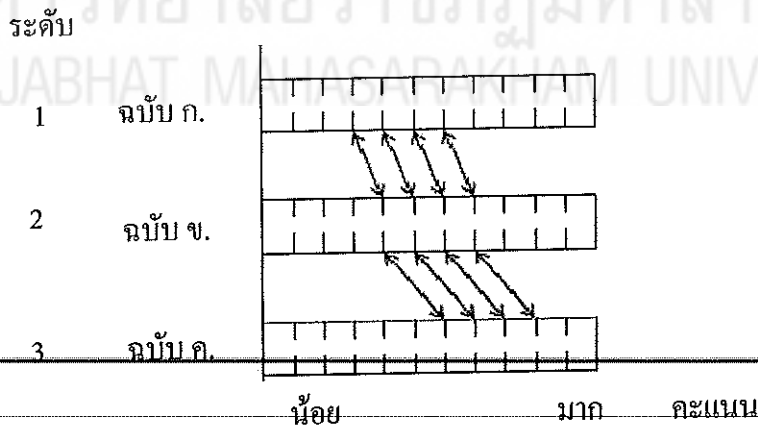
1. การเปรียบเทียบคะแนนเมื่อแบบสอบมีระดับความยากใกล้เคียงกันและใช้สอบกับกลุ่มผู้สอบที่มีระดับความสามารถใกล้เคียงกัน ซึ่งเรียกว่าการเปรียบเทียบคะแนนในแนวนอน (Horizontal Equating) สามารถแสดงได้ดังภาพ 2



ภาพ 2 การเปรียบเทียบคะแนนตามแนวนอน (Horizontal Equating)

ที่มา ศิริชัย กาญจนวาสี. (2541: 6). การเปรียบเทียบคะแนนระหว่างแบบสอบ.

2. การเปรียบเทียบคะแนนเมื่อแบบสอบมีระดับความยากแตกต่างกันและใช้สอบกับกลุ่มผู้สอบที่มีระดับความสามารถแตกต่างกัน ซึ่งเรียกว่า การเปรียบเทียบคะแนนในแนวตั้ง (Vertical Equating) สามารถแสดงได้ดังภาพ 3



ภาพ 3 การเปรียบเทียบคะแนนตามแนวตั้ง (Vertical Equating)

ที่มา ศิริชัย กาญจนวาสี. (2541:7). การเปรียบเทียบคะแนนระหว่างแบบสอบ.

การปรับเทียบคะแนนนั้นมีหลากหลายวิธีที่แตกต่างกันไปขึ้นอยู่กับแบบแผน (Design) ในการปรับเทียบ วิธีการ (Method) ที่ใช้ในการปรับเทียบ และข้อมูล (Data) ที่นำมาปรับเทียบ

แบบแผน (Design) ในการปรับเทียบนั้นเป็นรูปแบบการบริหารการสอบของแบบสอบที่นำมาปรับเทียบ หรือการออกแบบการเก็บรวบรวมข้อมูลจากแบบสอบ เช่น การออกแบบให้กลุ่มผู้สอบสองกลุ่มที่มาจากประชากรเดียวกัน แต่ละกลุ่มทำแบบสอบเพียงหนึ่งชุด (Random groups-one test administered to each group) เป็นต้น

วิธีการ (Method) ที่ใช้ในการปรับเทียบนั้น หมายถึง เทคนิควิธีทางสถิติที่ใช้ในการปรับเทียบ เช่น บางวิธีที่ใช้สถิติเชิงวิเคราะห์ (ค่าเฉลี่ย ความแปรปรวน สหสัมพันธ์) บางวิธีใช้กราฟในการวิเคราะห์ เป็นต้น

ข้อมูล (Data) ที่นำมาปรับเทียบนั้นจะมีอยู่ 2 ลักษณะคือ ข้อมูลของแบบสอบ (Score data) และข้อมูลรายข้อ (Item data) โดยการนำข้อมูลของแบบสอบ (คะแนน) มาปรับเทียบนั้น มีรากฐานจากทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory) ในขณะที่การนำข้อมูลรายข้อมาใช้ในการปรับเทียบนั้น มีรากฐานจากทฤษฎีการตอบสนองข้อสอบ (Item Response Theory) ลักษณะข้อมูลที่ใช้ในการปรับเทียบนี้ จึงจัดให้เป็นส่วนย่อยของแบบแผนการเก็บรวบรวมข้อมูล และวิธีการปรับเทียบได้ ดังนั้นจึงอาจจำแนกวิธีการปรับเทียบออกตามเกณฑ์ใหญ่ ๆ 2 เกณฑ์ต่อไปนี้

ก. จำแนกตามเทคนิควิธีการปรับเทียบคะแนน

ข. จำแนกตามแบบแผนการเก็บรวบรวมข้อมูล

ก. จำแนกตามเทคนิควิธีการปรับเทียบคะแนน

เทคนิควิธีการปรับเทียบคะแนนระหว่างแบบสอบเป็นการกำหนดเงื่อนไข และกฎเกณฑ์การแปลงคะแนน จากแบบสอบต่างชุดที่นำมาปรับเทียบกัน ให้ใช้คะแนนแทนกันได้ และสามารถเปรียบเทียบคะแนนกันได้ด้วยความหมาย

การปรับเทียบคะแนนเมื่อพิจารณาจากข้อมูลที่นำมาปรับเทียบ สามารถจำแนกได้ 2 รูปแบบคือ การปรับเทียบคะแนนตามทฤษฎีการทดสอบแบบดั้งเดิม และการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบ ซึ่งทั้ง 2 รูปแบบมีหลักการในการปรับเทียบคะแนนที่เหมือนกัน เพียงแต่วิธีการในการปรับเทียบคะแนนต่างกัน รูปแบบการปรับเทียบคะแนนที่มีผู้นิยมใช้ดังนี้

1. การปรับเทียบคะแนนตามทฤษฎีการทดสอบแบบดั้งเดิม แบ่งเป็น

1.1 การปรับเทียบคะแนนรูปแบบอิกวิเปอร์เซ็นต์ไทล์ (Equipercentile Equating)

1.2 การปรับเทียบคะแนนรูปแบบอิกวิเปอร์เซ็นต์ไทล์โดยใช้แบบสอบร่วม

1.3 การปรับเทียบคะแนนเชิงเส้นตรง (Linear Equating)

ซึ่งมีรายละเอียดดังนี้

1.1 การเปรียบเทียบคะแนนรูปแบบอควิเปอร์เซ็นไทล์ (Equipercntile Equating)

การเปรียบเทียบคะแนนรูปแบบอควิเปอร์เซ็นไทล์ มีนิยามว่า คะแนนจากแบบสอบสองฉบับจะถือว่าเป็นคะแนนที่สมมูลกัน เมื่อคะแนนของแต่ละฉบับมีตำแหน่งเปอร์เซ็นไทล์เดียวกัน การเปรียบเทียบคะแนนรูปแบบอควิเปอร์เซ็นไทล์ประกอบด้วย 2 ขั้นตอน (Petersen, Kolen and Hoover, 1989 : 247-248) ดังนี้

ขั้นตอนที่หนึ่ง การแจกแจงความถี่สะสมสัมพัทธ์สำหรับแบบสอบที่จะนำมาเทียบคะแนน โดยมีขั้นตอนย่อยๆ คือ

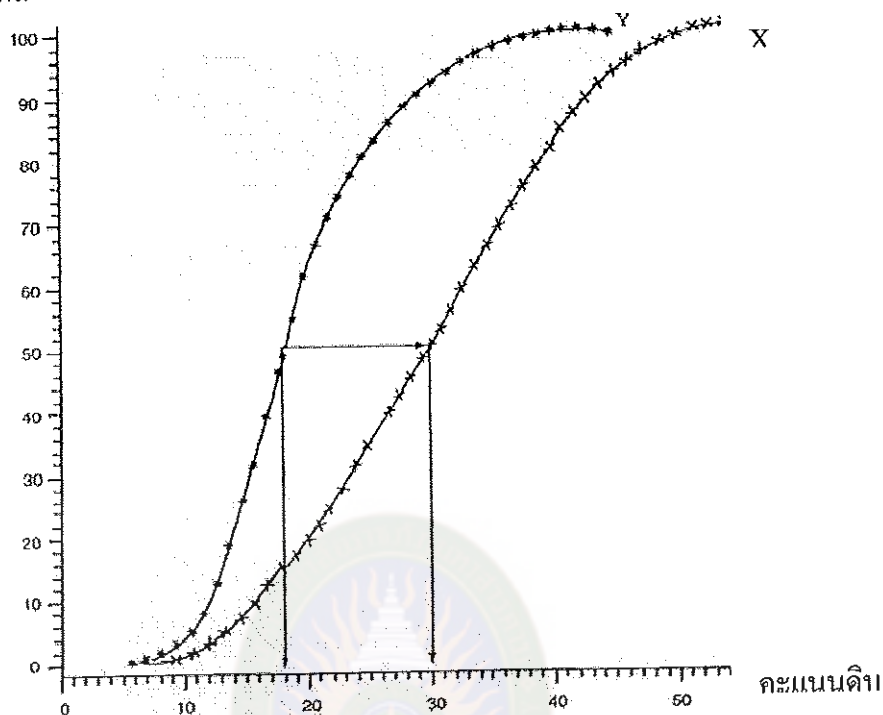
1. นำคะแนนของกลุ่มผู้สอบที่มีความสามารถทั้งเก่ง ปานกลาง และอ่อน โดยสุ่มเป็นสองกลุ่มย่อย ให้กลุ่มหนึ่งทำแบบสอบ X และอีกกลุ่มทำแบบสอบ Y มาทำการแจกแจงคะแนน X และ Y

2. คำนวณหาจุดกลางเปอร์เซ็นไทล์ของแต่ละการแจกแจง

3. อ่านและทำเครื่องหมายค่าคะแนนของแบบสอบฉบับ X และฉบับ Y ของการแจกแจงที่สมมูลกันบนกระดาษกราฟ โดยแกนนอนเป็นคะแนนดิบและแกนตั้งเป็นตำแหน่งเปอร์เซ็นไทล์ประมาณ 30 จุด แล้วลากเส้นเชื่อมต่อกันเป็นกราฟ ดังภาพประกอบ 4

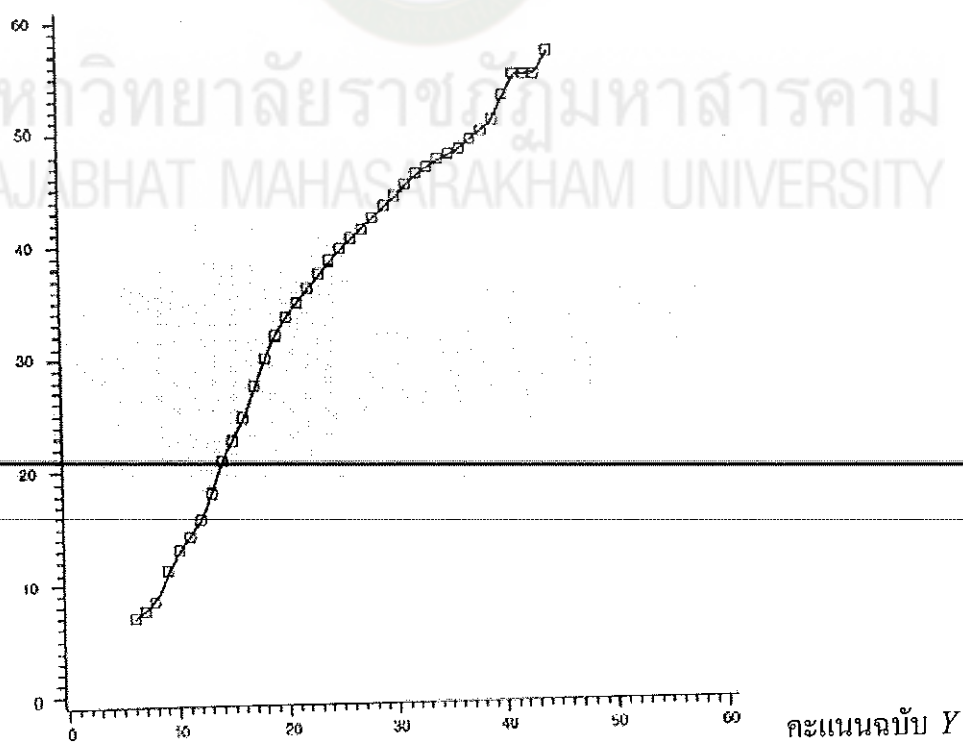
ขั้นตอนที่สอง เทียบมาตราคะแนนโดยนำคะแนนจากกราฟในขั้นตอนที่หนึ่งมาเขียนกราฟใหม่ลงบนกระดาษกราฟ โดยแกนนอนเป็นคะแนนฉบับ Y แกนตั้งเป็นคะแนนฉบับ X ทำการปรับเส้นกราฟให้เรียบ เส้นกราฟที่ได้จะอ่านค่า X ที่สมมูลกับ Y จากนั้นสร้างตารางสำเร็จเพื่ออ่านค่าคะแนนแปลงจากกราฟ ดังภาพประกอบ 5

เปอร์เซ็นต์ไทล์



ภาพ 4 กระบวนการเปรียบเทียบคะแนนรูปแบบอิกวิเปอร์เซ็นต์ไทล์

คะแนนฉบับ X



ภาพ 5 การแปลงคะแนนของการเปรียบเทียบคะแนนรูปแบบอิกวิเปอร์เซ็นต์ไทล์

1.2 การเปรียบเทียบคะแนนรูปแบบอิกวิเปอร์เซ็นไทล์โดยใช้แบบสอบร่วม

ลอร์ดและเลวิน (Lord and Levine อ้างถึงใน วรเนตร พิชิตเกริกพล. 2537 :17) ได้เสนอขั้นตอนการเปรียบเทียบคะแนนอิกวิเปอร์เซ็นไทล์โดยใช้แบบสอบร่วมสำหรับกลุ่มที่ไม่ได้สุ่มซึ่งมีความสามารถไม่แตกต่างกันมากนัก โดยที่กลุ่ม α สอบแบบสอบฉบับ X และฉบับ U ส่วนกลุ่ม β สอบแบบสอบฉบับ Y และฉบับ U ขั้นตอนในการเปรียบเทียบคะแนนคือประมาณค่าความถี่ในแบบสอบฉบับ X และ Y สำหรับกลุ่ม t ($\alpha + \beta$) มีขั้นตอนดังนี้

1. รวมคะแนนของฉบับ U ที่สอบโดยกลุ่ม α, β และกลุ่ม t
2. หาสัดส่วนความถี่ $f_u / f_{i\alpha}, f_u / f_{i\beta}$ ที่ทุกช่วงคะแนน i
3. กระจายความถี่ของฉบับ U ที่ทุกช่วงคะแนนของฉบับ X และ Y
4. คูณความถี่ของฉบับ U ในแต่ละช่วงคะแนนของฉบับ X ด้วยสัดส่วน $f_u / f_{i\alpha}$
5. คูณความถี่ของฉบับ U ในแต่ละช่วงคะแนนของฉบับ Y ด้วยสัดส่วน $f_u / f_{i\beta}$
6. หาดำแหน่งเปอร์เซ็นไทล์ของความถี่ที่ประมาณค่าใหม่ของทั้งสองฉบับ
7. ดำเนินการเทียบตามขั้นตอนที่สองของการเปรียบเทียบคะแนน โดยวิธี

อิกวิเปอร์เซ็นไทล์ดังกล่าวแล้วข้างต้น

ในการเปรียบเทียบคะแนนรูปแบบอิกวิเปอร์เซ็นไทล์มีข้อจำกัดที่ควรคำนึงอยู่หลายประการ ดังนี้ (Angoff, 1971 ; Potthoff, 1982. อ้างถึงใน พรพิมล นาคเวช. 2537 : 14 – 15)

1. การสร้างกราฟเปอร์เซ็นไทล์เพื่อเทียบคะแนนระหว่างแบบสอบ ควรสร้างด้วยความระมัดระวังและไม่ลำเอียง
2. รูปแบบอิกวิเปอร์เซ็นไทล์มีความไวต่อความแปรปรวนเชิงสุ่มมาก โดยเฉพาะกลุ่มตัวอย่างที่มีขนาดเล็กจึงควรใช้กลุ่มผู้สอบที่มีขนาดใหญ่และมีความสามารถค่อนข้างกระจาย
3. แบบสอบที่นำมาเทียบคะแนนควรมีความเที่ยงสูงพอกัน ถ้าความเที่ยงต่างกันมากผลการเปรียบเทียบคะแนนจะขาดความคงที่
4. การเปรียบเทียบคะแนนระหว่างแบบสอบสองชุดทำได้เฉพาะในช่วงพิสัยของคะแนนที่มีความถี่ของคะแนนสังเกตเพียงพอ ส่วนช่วงที่มีความถี่ของคะแนนน้อยจะมีความคลาดเคลื่อนสูงมาก

1.3 การเปรียบเทียบคะแนนเชิงเส้นตรง (Linear Equating)

การเปรียบเทียบคะแนนเชิงเส้นตรงมีนิยามว่า สำหรับกลุ่มผู้สอบที่กำหนดคะแนนจากแบบสอบสองฉบับจะถือว่าเท่าเทียมกันถ้าตรงกับคะแนนมาตรฐานเดียวกัน (Angoff, 1971; Petersen and other. 1982 อ้างถึงใน สุนิสา จัยม่วงศรี. 2537 : 13) การเปรียบเทียบคะแนนเชิงเส้นตรง

เป็นวิธีการปรับเทียบคะแนนที่ง่ายและสะดวกที่สุดในการนำไปใช้ (Hills, Subhiyah and Hirsch, 1988 : 229)

แองกอฟฟ์ (Angoff, 1984 อ้างถึงในสุนิสา จุ้ยม่วงศรี, 2537 : 13) ได้เสนอแบบแผนการรวบรวมและการจัดกระทำข้อมูลในทางสถิติของการปรับเทียบคะแนนเชิงเส้นตรงไว้ 6 รูปแบบ ซึ่งแต่ละรูปแบบจะมีวิธีการประมาณค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของแบบสอบแตกต่างกันออกไปตามเงื่อนไขของรูปแบบการรวบรวมข้อมูล แต่ทุกรูปแบบจะตัดสินคะแนนสมมูลจากค่าคะแนนมาตรฐานเดียวกันคือ

$$\frac{Y - M_y}{S_y} = \frac{X - M_x}{S_x} \quad (2.1)$$

เมื่อ X คือ คะแนนจากแบบสอบฉบับ X

Y คือ คะแนนจากแบบสอบฉบับ Y

M_x คือ ค่าเฉลี่ยจากคะแนนแบบสอบฉบับ X

M_y คือ ค่าเฉลี่ยจากคะแนนแบบสอบฉบับ Y

S_x คือ ค่าส่วนเบี่ยงเบนมาตรฐานของคะแนนแบบสอบฉบับ X

S_y คือ ค่าส่วนเบี่ยงเบนมาตรฐานของคะแนนแบบสอบฉบับ Y

จากสมการ (2.1) สามารถเขียนใหม่ได้ดังนี้

$$\begin{aligned} (Y - M_y) \frac{S_x}{S_y} &= X - M_x \\ X &= M_x + (Y - M_y) \frac{S_x}{S_y} \\ &= M_x + Y \frac{S_x}{S_y} - M_y \frac{S_x}{S_y} \\ &= \frac{S_x}{S_y} Y + \left[(M_x - M_y) \frac{S_x}{S_y} \right] \end{aligned}$$

จะได้สมการปรับเทียบคะแนนเชิงเส้นตรง

$$X = AY + B$$

เมื่อ

$$A = \frac{S_x}{S_y}$$

$$B = M_x - AM_y$$

ในแต่ละรูปแบบการรวบรวมข้อมูลของแองกอฟฟ์ ได้เสนอรายละเอียดแยกแยะระหว่างกรณีที่เป็นแบบสอบเทียบคะแนนมีความเที่ยงเท่ากัน และเมื่อแบบสอบเทียบคะแนนมีความเที่ยงไม่เท่ากัน โดยแต่ละรูปแบบมีรายละเอียดดังนี้

รูปแบบที่ 1 กลุ่มผู้สอบสองกลุ่มที่ถูกสุ่มมาจากประชากรเดียวกัน โดยการสุ่มผู้สอบแยกเป็นสองกลุ่มย่อยแต่ละกลุ่มทำแบบสอบเพียงหนึ่งฉบับ กลุ่มทั้งสองควรมีความคล้ายคลึงกันทางความรู้ ความสามารถเพื่อให้คะแนนสอบที่ได้ไม่เกิดจากผลของความแตกต่างของกลุ่ม ข้อดีของรูปแบบนี้คือ แต่ละกลุ่มทำแบบสอบฉบับเดียวจึงสามารถหลีกเลี่ยงปัญหาเรื่องการเรียนรู้ การฝึกฝน และความเมื่อยล้าจากการทำแบบสอบแต่ปัญหาที่ตามมาคือการใช้กลุ่มสองกลุ่ม ความแตกต่างที่เกิดขึ้น แม้เพียงเล็กน้อยย่อมส่งผลกระทบต่อความลำเอียงในการปรับเทียบคะแนนได้วิธีการที่จะช่วยลดความแตกต่างระหว่างกลุ่มทำได้โดยการใช้กลุ่มตัวอย่างสุ่มขนาดใหญ่

รูปแบบที่ 2 กลุ่มผู้สอบสองกลุ่มที่ถูกสุ่มมาจากประชากรเดียวกัน โดยการสุ่มผู้สอบแยกเป็นสองกลุ่มย่อย แต่ละกลุ่มย่อยได้รับการทดสอบจากแบบสอบทั้ง 2 ฉบับ ในลักษณะที่กลุ่มย่อยหนึ่งให้ทำแบบสอบฉบับที่ 1 แล้วตามด้วยแบบสอบฉบับที่ 2 ส่วนอีกกลุ่มย่อยหนึ่งให้ทำแบบสอบฉบับที่ 2 แล้วตามด้วยแบบสอบฉบับที่ 1 สลับกันเพื่อให้เกิดความสมดุล การใช้รูปแบบนี้กลุ่มผู้สอบจึงได้รับอิทธิพลจากลำดับการสอบ การเรียนรู้ การฝึกฝน และความเมื่อยล้าที่สมดุลกันในทั้งสองกลุ่มย่อย

รูปแบบที่ 3 กลุ่มผู้สอบสองกลุ่มที่ถูกสุ่มมาจากประชากรเดียวกัน โดยกลุ่มผู้สอบแยกเป็นสองกลุ่มย่อย ในแต่ละกลุ่มทำแบบสอบเพียงฉบับเดียว และทำแบบสอบร่วมที่เหมือนกันอีกส่วนหนึ่งกล่าวคือกลุ่ม α ทำแบบสอบฉบับ X และแบบสอบร่วม (U) ส่วนกลุ่ม β ทำแบบสอบฉบับ Y และแบบสอบร่วม (U) โดยแบบสอบร่วมอาจจะรวมอยู่ในแบบสอบทั้งสองชุดซึ่งเรียกว่า แบบสอบร่วมภายใน (U) ส่วนกลุ่ม β ทำแบบสอบฉบับ Y และแบบสอบร่วม (U) โดยแบบสอบร่วมอาจจะรวมอยู่ในแบบสอบทั้งสองชุดซึ่งเรียกว่าแบบสอบร่วมภายใน (Internal Anchor test) หรืออาจจะจัดเป็นฉบับที่แยกจากแบบสอบฉบับ X และ Y เรียกว่า แบบสอบร่วมภายนอก (External Anchor test) ซึ่งแบบสอบร่วมภายนอกกลุ่มผู้สอบจะต้องทำและจับเวลาแยกออกจากตัวแบบสอบที่ต้องการนำมาเทียบคะแนนแบบสอบร่วมจะต้องถูกจัดให้สอบในลำดับที่เหมือนกันทั้งสองกลุ่ม รูปแบบที่ 3 จะมีข้อได้เปรียบรูปแบบที่หนึ่งเพราะแบบสอบร่วมจะทำหน้าที่กำหนดสเกลร่วมของคะแนนจากแบบสอบแต่ละชุดที่สอบ โดยผู้สอบทั้งสองกลุ่มเพื่อนำไปปรับค่าความยากของข้อสอบหรือค่าความสามารถของผู้สอบจากแบบสอบต่างชุดให้อยู่ในสเกลร่วมกัน แบบสอบร่วมจะลดความแปรปรวนของความคลาดเคลื่อนในการปรับเทียบคะแนน

รูปแบบที่ 4 กลุ่มที่ไม่ได้มาจากการสุ่มสองกลุ่มแต่ละกลุ่มทำแบบสอบเพียงฉบับเดียว และทำแบบสอบร่วมที่เหมือนกันอีกส่วนหนึ่ง กลุ่มผู้สอบที่ไม่ได้มาจากการสุ่มเป็นกลุ่มที่เกิดขึ้นในสถานการณ์จริงซึ่งต้องสอบแบบสอบต่างชุดในเวลาต่างกัน ผู้สอบจึงไม่ได้ถูกสุ่มมาจากประชากรเดียวกัน เช่นการจัดสอบด้วยแบบสอบฉบับใหม่ในปี พ.ศ. 2546 แต่ต้องการเทียบคะแนนไปยังแบบสอบฉบับเก่าที่สอบในปี พ.ศ. 2545 แบบสอบร่วมจึงต้องมีความคล้ายคลึงกับแบบสอบเทียบคะแนนทั้งสองฉบับให้มากที่สุดเพื่อลดอคติของการปรับเทียบคะแนนมาจากความแตกต่างในความสามารถระหว่างกลุ่ม

รูปแบบที่ 5 เป็นวิธีการปรับเทียบคะแนน โดยใช้แบบสอบร่วมที่เกี่ยวข้องกับคะแนน มีลักษณะการเทียบเป็นลักษณะเฉพาะ คือ

5.1 แบบสอบฉบับ X และ ฉบับ Y เทียบไปสู่แบบสอบร่วมเป็นการบริหารแบบสอบร่วมซึ่งอาจจะสอบแบบสอบร่วมก่อน แล้วตามด้วยแบบสอบฉบับ X และ ฉบับ Y หรือสอบแบบสอบฉบับ X และฉบับ Y ก่อนแล้วตามด้วยแบบสอบร่วม การปรับเทียบคะแนนโดยวิธีนี้คือการปรับเทียบคะแนนจากแบบสอบฉบับ X ไปยังแบบสอบร่วม และเทียบจากฉบับ Y ไปยังแบบสอบร่วมซึ่งคะแนนที่เทียบแล้วของทั้งสองแบบสอบที่ระดับคะแนนของแบบสอบร่วมเดียวกันถือว่าเท่าเทียมกัน โดยแบบสอบร่วม U ไม่มีลักษณะคู่ขนานกับแบบสอบฉบับ X และฉบับ Y แล้วกลุ่ม α และกลุ่ม β ต้องมาจากการสุ่มจากประชากรเดียวกัน

5.2 แบบสอบร่วมเป็นตัวพยากรณ์แบบสอบฉบับ X และฉบับ Y ซึ่งจะมีสมการแปลงคะแนนแตกต่างจากข้อ 5.1

5.3 แบบสอบฉบับ X และฉบับ Y เป็นตัวพยากรณ์แบบสอบร่วม ซึ่งจะมีสมการแปลงคะแนนแตกต่างจากข้อ 5.1 และ 5.2

รูปแบบที่ 6 เป็นวิธีการปรับเทียบคะแนนที่ขึ้นกับลักษณะข้อสอบได้แก่วิธีการของเทอร์สโตนและแฟน (Thurstone, 1925; Fan, 1957 อ้างถึงในสุนิสสา ชัยม่วงศรี, 2537 : 16) โดยให้แบบสอบเทียบคะแนนทั้งสองฉบับมีข้อสอบชุดหนึ่งซึ่งมีลักษณะร่วมกันคือค่าความยาก

2. การปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบ

ลอร์ด (Lord, 1980 อ้างถึงใน สุนิสสา ชัยม่วงศรี, 2537 : 22-24) ได้กล่าวถึงการปรับเทียบคะแนนโดยใช้ทฤษฎีการตอบสนองข้อสอบไว้ว่า แบบสอบสองฉบับใดๆที่ใช้ในการปรับเทียบคะแนนจะต้องเป็นแบบสอบที่วัดคุณลักษณะเดียวกันและเทียบมาตรฐานคะแนนที่สมมูลกันที่ระดับความสามารถเดียวกัน จะต้องมีข้อกำหนด 3 ประการดังนี้

1. ความเสมอภาค (Equity) หมายถึง สำหรับทุกๆระดับความสามารถ (θ) การแจกแจงความถี่แบบมีเงื่อนไขของคะแนนแปลงต้องเหมือนกับการแจกแจงความถี่แบบมีเงื่อนไขของคะแนนแบบสอบที่ต้องการเทียบ

2. ความไม่ผันแปรตามกลุ่ม (Invariance across Groups) หมายถึง คะแนนที่แปลงแล้วจะคงที่โดยไม่แปรเปลี่ยนไปตามกลุ่มประชากรที่นำมาเทียบจึงสามารถนำผลไปใช้กับกลุ่มอื่นๆได้

3. ความสมมาตร (Symmetry) หมายถึง คะแนนแปลงที่ได้จากการปรับเทียบคะแนนจะต้องเหมือนกัน ไม่ว่าจะเป็นการปรับเทียบคะแนนจากแบบสอบฉบับ X ไปยังแบบสอบฉบับ Y หรือจากแบบสอบฉบับ Y ไปยังแบบสอบฉบับ X

การปรับเทียบคะแนนรูปแบบทฤษฎีการตอบสนองข้อสอบสามารถแบ่งได้ 2 วิธีการใหญ่ๆ ดังนี้

2.1 การปรับเทียบคะแนนโดยใช้คะแนนจริง (True-Score Equating)

แบบสอบที่มีความยากต่างกันถึงแม้ว่าจะนำไปสอบกับประชากรที่มีความสามารถเท่ากันหรือคุณลักษณะเดียวกัน แต่คะแนนที่ได้จะมีการแจกแจงแตกต่างกัน ดังนั้นถ้าสามารถหาคะแนนจริงได้ ก็จะสามารถเทียบมาตรฐานคะแนนระหว่างความสามารถ (θ) กับคะแนนจริงได้จากความสัมพันธ์เชิงคณิตศาสตร์ดังนี้

$$\xi = \xi(\theta) = \sum_{i=1}^m P_i(\theta_x)$$

$$\eta = \eta(\theta) = \sum_{i=1}^n P_i(\theta_y)$$

เมื่อ	ξ	คือ	คะแนนจริงของผู้สอบที่ได้จากแบบสอบฉบับ X
	η	คือ	คะแนนจริงของผู้สอบที่ได้จากแบบสอบฉบับ Y
	m	คือ	จำนวนข้อของแบบสอบฉบับ X
	n	คือ	จำนวนข้อของแบบสอบฉบับ Y
	θ_x	คือ	ค่าความสามารถของผู้สอบที่ได้จากแบบสอบฉบับ X
	θ_y	คือ	ค่าความสามารถของผู้สอบที่ได้จากแบบสอบฉบับ Y

ดังนั้น ถ้าแทนค่าความสามารถ (θ) ใด ๆ ลงในสมการที่ θ เท่ากันจะหาคะแนนสมมูลของแบบสอบฉบับ X และฉบับ Y ได้โดยคำนวณจากค่า $P_i(\theta)$ และ $P_j(\theta)$ และในทางตรงกันข้าม ก็สามารถประมาณค่าความสามารถ (θ) เมื่อทราบคะแนนจริง

การหาค่า $P_i(\theta)$ และ $P_j(\theta)$ ได้จากการประมาณค่าพารามิเตอร์, สองพารามิเตอร์ หรือสามพารามิเตอร์จากนั้นนำคะแนนจริงของแบบสอบฉบับ X และฉบับ Y หาค่าความสัมพันธ์กันโดยใช้ค่าความสามารถ (θ) ที่ระดับเดียวกัน

2.2 การเปรียบเทียบคะแนนโดยใช้คะแนนสังเกต (Observed – Score Equating)

ปัญหาของการเปรียบเทียบคะแนนโดยใช้คะแนนจริงคือไม่สามารถทราบคะแนนจริงของแต่ละคนได้นอกจากใช้วิธีประมาณจากผลการสอบ ซึ่งคะแนนที่ได้เป็นเพียงค่าประมาณเท่านั้น ยังไม่มีคุณสมบัติเป็นคะแนนจริงและไม่สามารถทำการเปรียบเทียบคะแนนที่ต่ำกว่าคะแนนการคาดได้ (ศิริชัย กาญจนวาสี. 2541 : 30) ลอร์ด (Lord. 1980 อ้างถึงใน ภาวิณี ศรีสุขวัฒนานนท์. 2529 : 43) จึงเสนอการเทียบคะแนนโดยใช้คะแนนสังเกต โดยอาศัยข้อมูลจากการตอบแบบสอบร่วม (Anchor test)

การแจกแจงทางทฤษฎีของคะแนนที่สังเกตได้ของผู้สอบที่มีความสามารถ θ จากการทำแบบสอบฉบับหนึ่ง สามารถแสดงได้ดังนี้ (Hambleton & Swaminathan. 1985 : 215)

$$\sum_{r=0}^n f(r|\theta)r^r = \prod_{i=1}^n [Q_i(\theta) + rP_i(\theta)]$$

ขั้นตอนในการเปรียบเทียบคะแนนที่สังเกตได้โดยใช้ทฤษฎีการตอบสนองข้อสอบสามารถสรุปเป็นขั้นตอนได้ดังนี้ (ศิริชัย กาญจนวาสี. 2545, หน้า 164)

1. นำคะแนนจากกลุ่มผู้สอบแบบสอบฉบับ X มาวิเคราะห์ข้อมูลเพื่อประมาณค่าพารามิเตอร์ของผู้สอบและข้อสอบ แล้วนำค่า θ_x ของผู้สอบแต่ละคนมาคำนวณความถี่สัมพัทธ์อย่างมีเงื่อนไขทางทฤษฎีของการแจกแจงคะแนนดิบ $f_x(r|\theta_x)$
2. แปลงค่า θ_x ให้สมมูลกับ θ_y ด้วยสูตร $T_r(\theta_y) = \alpha\theta_x + \beta$ แล้วนำ $T_r(\theta_x)$ ของกลุ่มผู้สอบมาคำนวณความถี่ของคะแนนดิบทางทฤษฎี $f_x(r)$ จะได้ตารางการแจกแจงความถี่ของคะแนนดิบทางทฤษฎีของกลุ่มผู้สอบ ฉบับ X
3. นำคะแนนจากกลุ่มผู้สอบแบบสอบฉบับ Y มาวิเคราะห์ข้อมูลเพื่อประมาณค่าพารามิเตอร์ของข้อสอบและผู้สอบ แล้วนำค่า θ_y ของผู้สอบแต่ละคนมาคำนวณความถี่สัมพัทธ์อย่างมีเงื่อนไขทางทฤษฎีของการแจกแจงคะแนนดิบ $f_y(r|\theta_y)$
4. นำค่า θ_y ของกลุ่มผู้สอบมาคำนวณความถี่ของคะแนนดิบทางทฤษฎี $f_y(r)$ จะได้ตารางการแจกแจงความถี่ของคะแนนดิบทางทฤษฎีของกลุ่มผู้สอบ ฉบับ Y
5. เทียบคะแนนระหว่างคะแนนดิบของแบบสอบฉบับ X และ Y (จากข้อ 2 และ 4)

โดยใช้วิธีอิกวิเปอร์เซ็นไทล์

ข. จำแนกตามแบบแผนการเก็บรวบรวมข้อมูล (Designs for Data Collection)

การจัดการทดสอบเพื่อการเก็บรวบรวมข้อมูลสำหรับนำมาใช้ในการเปรียบเทียบคะแนนสามารถจำแนกเป็นแบบแผนการเก็บรวบรวมข้อมูลได้ 4 รูปแบบ ดังนี้

1. รูปแบบสอบกลุ่มเดียว เป็นการใช้กลุ่มผู้สอบกลุ่มเดียวกันแต่ทำแบบสอบ 2 ฉบับ โดยให้ผู้สอบทำแบบสอบฉบับหนึ่งก่อนแล้วตามด้วยแบบสอบอีกฉบับหนึ่ง

2. รูปแบบผู้สอบกลุ่มเท่าเทียมกัน เป็นการใช้กลุ่มผู้สอบที่มีลักษณะเท่าเทียมกัน โดยการจัดกลุ่มผู้สอบให้มีความคล้ายคลึงกันมากที่สุดเท่าที่จะทำได้ แล้วให้ผู้สอบแต่ละกลุ่มทำแบบสอบฉบับเดียวกัน

3. รูปแบบผู้สอบต่างกลุ่มโดยใช้แบบสอบร่วม เป็นการใช้กลุ่มผู้สอบที่มีลักษณะไม่เท่าเทียมกัน และให้ผู้สอบแต่ละกลุ่มทำแบบสอบฉบับเดียว โดยแบบสอบแต่ละฉบับมีข้อสอบร่วมกันจำนวนหนึ่ง ซึ่งเรียกว่า แบบสอบร่วม

4. รูปแบบผู้สอบต่างกลุ่มโดยมีผู้สอบร่วม เป็นการใช้กลุ่มผู้สอบที่มีลักษณะไม่เท่าเทียมกัน และให้ผู้สอบแต่ละกลุ่มทำแบบสอบต่างฉบับกัน แต่จะมีผู้สอบจำนวนหนึ่งจากทั้งสองกลุ่มได้ทำแบบสอบทั้งสองฉบับ

การเปรียบเทียบคะแนนจะมีคุณภาพดีที่สุดเมื่อคะแนนที่ได้จากแบบสอบที่นำมาเทียบคะแนนเป็นไปตามเงื่อนไขต่างๆของการเปรียบเทียบคะแนนที่กำหนดไว้ในแต่ละรูปแบบ แต่ในสถานการณ์จริงอาจมีข้อจำกัดบางประการทำให้ไม่สามารถได้ข้อมูลตรงตามเงื่อนไขที่กำหนดได้ ดังนั้นการเปรียบเทียบคะแนนจึงจำเป็นต้องมีการประเมินคุณภาพของวิธีการเปรียบเทียบคะแนน ซึ่งถือว่าการตรวจสอบคุณภาพของการเปรียบเทียบคะแนน ซึ่งมีวิธีการตรวจสอบคุณภาพของการเปรียบเทียบคะแนนดังนี้

1. ความเพียงพอในการเปรียบเทียบคะแนน

การเปรียบเทียบคะแนนแต่ละวิธี จะมีคุณภาพที่ดีที่สุดเมื่อคะแนนที่ได้จากแบบทดสอบเป็นไปตามเงื่อนไขต่าง ๆ ที่กำหนดไว้ในแต่ละรูปแบบของการเปรียบเทียบคะแนน แต่ในสถานการณ์จริง อาจมีข้อจำกัดทำให้ไม่สามารถได้ข้อมูลตามเงื่อนไขที่กำหนด ดังนั้นการเปรียบเทียบคะแนนจึงจำเป็นต้องมีการตรวจสอบความเพียงพอ(Adequacy) ของการเปรียบเทียบคะแนน ซึ่งเป็นการประเมินประสิทธิภาพของวิธีการเปรียบเทียบคะแนน (พิชัย ละแมนชัย, 2538 : 44) และได้มีผู้เสนอแนวความคิดไว้หลายวิธีดังนี้

1.1 ดัชนีตรวจสอบความเพียงพอของเจเกอร์ (Jaeger's Adequacy Index)

1.2 ดัชนีความแตกต่าง (Discrepancy Index)

1.3 ดัชนีเปรียบเทียบเปอร์เซ็นต์ไทล์ (The Percentile Comparison Index)

1.4 ดัชนีความแตกต่างตามแนวคิดของ ภาวิณี ศรีสุขวัฒนานันท์

ซึ่งมีรายละเอียดดังนี้

1.1 ดัชนีตรวจสอบความเพียงพอของเจเกอร์ (Jaeger)

เจเกอร์ (Jaeger, 1981 : 26) ได้เสนอดัชนี 5 ตัว เพื่อตรวจสอบความเพียงพอในการเปรียบเทียบโดยใช้วิธีการเปรียบเทียบคะแนนเชิงเส้นตรงดังนี้

1.1.1 ดัชนีความคล้ายคลึงของการแจกแจงคะแนนสะสมของแบบทดสอบชุดเก่าและชุดใหม่โดยการปรับความแตกต่างระหว่างค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐาน การทดสอบความเหมือนของการแจกแจงใช้การทดสอบด้วย The Kolmogorov-Smirnov Two-Sample Test

1.1.2 รูปแบบของการแจกแจงคะแนนดิบกับคะแนนแปลง (Shape of the Raw Score Transformation) เหตุผลของการใช้ดัชนีตัวนี้ คือ ถ้าการปรับเทียบคะแนนด้วยวิธีเชิงเส้นตรงสามารถอธิบายความแตกต่างในการแจกแจงของคะแนนดิบทั้งสองชุดอย่างเพียงพอ ก็เป็นเหตุผลอย่างเพียงพอเช่นกันที่จะยอมรับว่าการแปลงคะแนนดิบจากแบบทดสอบชุดใหม่ไปยังแบบทดสอบชุดเก่าเป็นเส้นตรงอย่างแน่นอน

1.1.3 ความคงเส้นคงวาของผลลัพธ์ของการปรับเทียบคะแนนตามวิธีเชิงเส้นตรงกับการเทียบที่ตำแหน่งเปอร์เซ็นต์ไทล์ (Consistency of linear and Equipercntile Equating Results) การวิเคราะห์นี้อาศัยข้อตกลงที่เป็นสมมุติฐานเบื้องต้นที่ว่า ถ้าวิธีเชิงเส้นตรงมีความเพียงพอแล้ว ฟังก์ชันของวิธีการเทียบที่ตำแหน่งเปอร์เซ็นต์ไทล์จะแปรผันไปโดยสุ่มรอบ ๆ ฟังก์ชันของวิธีเชิงเส้นตรงที่สมนัยกัน

1.1.4 ความคล้ายคลึงของการแจกแจงความยากของข้อสอบ (Similarity of Item Difficulty Distributions) โดยอาศัยหลักการที่ว่า การใช้วิธีเชิงเส้นตรงมีความเพียงพออย่างแท้จริงกับแบบทดสอบที่มีคุณสมบัติเป็นคู่ขนานกัน ถ้ามีการเบี่ยงเบนจากความเป็นคู่ขนานมากเท่าใด แสดงว่า ต้องการวิธีการปรับเทียบคะแนนที่ซับซ้อนขึ้น เพราะการแจกแจงของแบบทดสอบที่ไม่ใช่คู่ขนานจะมีความแตกต่างเกิดขึ้นในระดับ โมเมนต์ที่สูงขึ้น

1.1.5 ความคล้ายคลึงของค่าอำนาจจำแนกของข้อสอบ (Similarity of Item Discrimination Distributions) เหตุผลทำนองเดียวกับดัชนีตัวที่ 1.1.4

1.2 ดัชนีความแตกต่าง (Discrepancy Index)

ดัชนีความแตกต่างใช้เพื่อเปรียบเทียบคะแนนแปลงที่ได้จากการเทียบคะแนน 2 วิธี ดัชนีนี้ใช้สรุปภาพรวมของความแตกต่าง ซึ่งแต่เดิมใช้ดัชนีเพื่อประเมินความคลาดเคลื่อนทางสถิติ แต่นำมาใช้กับการเทียบคะแนนอาจจะไม่ชัดเจนนัก แต่สามารถประยุกต์ใช้ดัชนีเหล่านี้ได้โดยใช้ค่าที่แท้จริง (true value) เป็นเกณฑ์ ดัชนีกลุ่มนี้สามารถจำแนกได้ดังนี้ คือ ดัชนีถ่วงน้ำหนักและดัชนีไม่ถ่วงน้ำหนัก

1.2.1 ดัชนีถ่วงน้ำหนัก ด้วยความถี่ของคะแนนดิบของแบบทดสอบฉบับเป้าหมาย ประกอบด้วย

$$1) \text{RMAD} = \left(\frac{\sum_i f_i (A_i - B_i)^2}{\sum_i f_i} \right)^{\frac{1}{2}}$$

$$2) \text{MAD} = \left(\frac{\sum_i f_i |A_i - B_i|}{\sum_i f_i} \right)$$

$$3) \text{MAD} = \frac{\sum_i f_i (A_i - B_i)}{\sum_i f_i}$$

โดยที่ A_i คือคะแนนสมมูล i จากแบบทดสอบฉบับใหม่ที่เทียบคะแนนไปยังสเกลของฉบับเก่าด้วยการเทียบคะแนนวิธีที่หนึ่ง

B_i คือคะแนนเกณฑ์ ซึ่งเป็นคะแนนที่เป็นจริง หรือคะแนนสมมูล i จากแบบทดสอบฉบับใหม่ที่เทียบคะแนนไปยังสเกลของฉบับเก่าด้วยการเทียบคะแนนวิธีอื่น

1.2.2 ดัชนี ไม่ถ่วงน้ำหนัก เป็นการพิจารณาความแตกต่างที่เกิดขึ้นตลอดช่วงคะแนนของมาตรวัด โดยไม่กำหนดน้ำหนักในแต่ละระดับคะแนน ประกอบด้วย

$$1) \text{RMSD} = \left(\frac{\sum_i (A_i - B_i)^2}{k} \right)^{\frac{1}{2}}$$

$$2) \text{MAD}_u = \left(\frac{\sum_i |A_i - B_i|}{k} \right)$$

$$3) \text{MSD}_u = \frac{\sum_i (A_i - B_i)}{k}$$

โดยที่ k คือ จำนวนของระดับคะแนน (Score Point)

นอกจากนี้ดัชนี RMSD ยังสามารถนำมาแยกเป็นความแปรปรวนของความแตกต่างและความลำเอียงยกกำลังสอง คือ

$$\sum_i f_i (d_i)^2 / n = \sum_i f_i (d_i - \bar{d})^2 / n + \bar{d}^2$$

โดยที่ d_i คือ $(A_i - B_i)$

\bar{d} คือ ความแตกต่างเฉลี่ย

การแปลความหมายของดัชนีความแตกต่าง

การใช้ดัชนีเหล่านี้เป็นเกณฑ์ในการเทียบคะแนนสามารถแปลความหมายตามขนาดดัชนี ซึ่งเป็นค่าที่เรียงลำดับได้ว่าดัชนีที่มีค่าน้อยแสดงว่าความแตกต่างนั้นมีน้อย แต่ดัชนีขนาดเท่าใดที่ทำให้ผลการเทียบคะแนนเป็นที่น่าพอใจยังไม่ได้มีการระบุไว้

ปีเตอร์เซ็น มาร์โค และ สตีเวอาร์ท (Petersen, Marco and Stewart. 1982 : 91) ได้เสนอวิธีประเมินความเพียงพอของรูปแบบการปรับเทียบคะแนน โดยใช้ดัชนีความแตกต่าง ซึ่งมีแนวคิดว่าจะแนบคียบแต่ละตัวจากแบบทดสอบชุด X จะมีคะแนนเกณฑ์ (t) และคะแนนแปลง (t') ที่ได้จากการประมาณค่า โดยใช้วิธีการปรับเทียบคะแนนเฉพาะอันหนึ่ง ค่าความแตกต่าง d ระหว่าง t และ t' เป็นความคลาดเคลื่อนในการปรับเทียบคะแนน ถ้าความแตกต่างมีค่าน้อยแสดงว่า วิธีการปรับเทียบคะแนนวิธีนั้น มีความเหมาะสมมากกว่า

ดัชนีความแตกต่าง คือ ค่าความคลาดเคลื่อนรวมในการปรับเทียบคะแนนของวิธีการปรับเทียบ คะแนนที่ใช้ นั้น ซึ่งคำนวณจากค่าเฉลี่ยกำลังสองของความแตกต่างระหว่างคะแนนแปลงกับคะแนนเกณฑ์ และถ่วงน้ำหนักด้วยความแปรปรวนของคะแนนเกณฑ์ ซึ่งเท่ากับค่าความแปรปรวนของความแตกต่างรวมกับค่ากำลังสองของความลำเอียง

นั่นคือ ความคลาดเคลื่อนรวม (Total error) หรือ ดัชนีความแตกต่าง(C)

$$C = \frac{\sum f_j d_j^2}{nS_t^2} = \frac{\sum_{j=1}^n f_j (d_j - \bar{d})^2}{nS_t^2} + \frac{\bar{d}^2}{S_t^2}$$

(Total error) = (Variance of difference) + (Squared bias)

เมื่อ $d_j = t_j - t'_j$

t'_j = คะแนนแปลงสำหรับคะแนนคียบ X_j

t_j = คะแนนเกณฑ์สำหรับคะแนนคียบ X_j

$\bar{d} = \sum f_j d_j / n$

S_t = ความเบี่ยงเบนมาตรฐานของคะแนนเกณฑ์ t_j

f_j = ความถี่ของคะแนนคียบ X_j

$n = \sum f_j$ = จำนวนคะแนนที่ใช้

ดัชนีเปรียบเทียบความแตกต่างนี้เป็นค่ามาตรฐานแสดงในรูปสัดส่วนของความเบี่ยงเบนมาตรฐานของคะแนนเกณฑ์ ค่าความคลาดเคลื่อนรวมที่ได้นี้ จึงสามารถเปรียบเทียบกันได้โดยตรงระหว่างค่าที่ได้จากการใช้รูปแบบและสถานการณ์การเทียบที่แตกต่างกัน (Petersen, Marco and Stewart. 1982 : 91)

1.3 ดัชนีเปรียบเทียบเปอร์เซ็นต์ไทล์ (The Percentile Comparison Index)

โคเลนและวิทนี (Kolen and Whitney.1982 : 284) ได้แนะนำให้ใช้ดัชนีเปรียบเทียบเปอร์เซ็นต์ไทล์ เพื่อประเมินความเพียงพอของการปรับเทียบคะแนนทั้งนี้โดยอาศัยกลุ่มตัวอย่างตรวจสอบ (Cross-Validation Sample Examinees) ซึ่งเป็นกลุ่มอิสระอีกกลุ่มหนึ่งที่สุ่มมาจากประชากรเดียวกันกับกลุ่มผู้สอบที่ใช้สร้างตารางการปรับเทียบคะแนน วิธีการหาค่าดัชนีเปรียบเทียบเปอร์เซ็นต์ไทล์นี้ คือ การวัดความไม่สอดคล้อง (Dissimilarity) ระหว่างการแจกแจงคะแนนที่ได้จากกลุ่มตรวจสอบผล คือ คะแนนจากแบบทดสอบชุด X กับคะแนนซึ่งได้มาจากการแปลงคะแนนจากแบบทดสอบชุด Y ไปยังสเกลของแบบทดสอบชุด X ตามตารางการปรับเทียบคะแนนที่ได้จากวิธีการที่ใช้เทียบโดยหาค่าสมมูล ณ ที่ตำแหน่งเปอร์เซ็นต์ไทล์เดียวกัน ดัชนีดังกล่าวคำนวณได้จาก

$$C = \frac{\sum (X_i - X_i^*)^2}{nk}$$

เมื่อ n คือ จำนวนคะแนนดิบในกลุ่มตรวจสอบผล

k คือ จำนวนข้อสอบในแบบทดสอบรวมที่ใช้

X_i คือ คะแนนจากแบบทดสอบชุด X ซึ่งเป็นคะแนนเกณฑ์

X_i^* คือ คะแนนจากแบบทดสอบชุด Y ที่แปลงให้อยู่ในมาตราคะแนนของ X

ถ้าค่า C น้อยกว่า แสดงว่าวิธีการปรับเทียบคะแนนที่นำมาใช้สร้างตารางการแปลงคะแนนนั้นมีความเพียงพอที่จะนำมาใช้ให้ได้ผลการเทียบที่คงที่มากกว่า

1.4 ดัชนีความแตกต่างตามแนวคิดของ ภาวิณี ศรีสุขวัฒนานันท์

ภาวิณี ศรีสุขวัฒนานันท์ (2529 : 67) ได้เสนอวิธีการประเมินความเพียงพอโดยใช้ดัชนีความแตกต่างซึ่งคัดแปลงมาจากสูตรของโคเลน (Kolen, 1982 อ้างถึงใน ภาวิณี ศรีสุขวัฒนานันท์, 2529 : 67) ซึ่งได้ใช้ข้อมูลคะแนนจากผู้สอบเองเป็นเกณฑ์ในการหาความแตกต่าง ข้อมูลเหล่านี้ได้มาจากการออกแบบด้วยการใช้กลุ่มตัวอย่างตรวจสอบผล ซึ่งผู้สอบในกลุ่มตัวอย่างได้รับการทดสอบด้วยแบบทดสอบทั้งสองชุด ดังนั้น การใช้คะแนนของตนเองเป็นเกณฑ์จึงมีความอิสระ ไม่ขึ้นกับกระบวนการแปลงคะแนนอื่น ๆ และใช้แนวคิดของปีเตอร์เซ็น และคณะ คือ ใช้ค่าความแปรปรวนเป็นตัวถ่วงน้ำหนักเพื่อให้ค่าที่ได้มีหน่วยเป็นมาตรฐาน สูตรที่คัดแปลงนำมาใช้ คือ

$$C = \frac{\sum (X_i - X_i^*)^2}{nS_x^2}$$

- เมื่อ X_i คือ คะแนนเกณฑ์หรือคะแนนจากการสอบชุด X ของคนที่ i
 X_i^* คือ คะแนนที่ได้เทียบด้วยตารางแปลงคะแนนที่สมนัยกันของคนที i
 n คือ จำนวนในกลุ่มตัวอย่างตรวจสอบผลที่นำมาวิเคราะห์
 S_x^2 คือ ค่าความแปรปรวนของคะแนน X

เกณฑ์ที่ใช้ในการประเมินค่าดัชนีความแตกต่างใช้เกณฑ์ของปีเตอร์เซ็น และคณะ (Petersen, Marco and Stewart. 1982 : 93 – 94) โดยกำหนดระดับการยอมรับดังนี้

ระดับน่าพอใจอย่างยิ่ง	เมื่อ	$C \leq (0.05 SD_x)^2$
ระดับน่าพอใจ	เมื่อ	$(0.05 SD_x)^2 < C \leq (0.10 SD_x)^2$
ระดับปานกลาง	เมื่อ	$(0.10 SD_x)^2 < C \leq (0.15 SD_x)^2$
ระดับไม่น่าพอใจ	เมื่อ	$(0.15 SD_x)^2 < C \leq (0.20 SD_x)^2$
ระดับไม่น่าพอใจอย่างยิ่ง	เมื่อ	$C > (0.20 SD_x)^2$

2. ความคลาดเคลื่อนของการปรับเทียบคะแนน

การประเมินความคลาดเคลื่อนของการปรับเทียบคะแนนที่สมบูรณ์ คือ การวิเคราะห์แหล่งความลำเอียง (Bias) และความแปรผัน (Variability) แล้วนำมาเสนอให้เห็นเป็นภาพเดียวกัน เพื่อช่วยในการตัดสินใจว่า การนำวิธีการปรับเทียบคะแนนรูปแบบนั้น ๆ ไปใช้ มีความเหมาะสมหรือไม่อย่างไร (Braun and Holland 1982 : 32)

ความลำเอียงของการปรับเทียบคะแนน (Bias Error) เป็นองค์ประกอบหนึ่งของความคลาดเคลื่อน ของการปรับเทียบคะแนน ซึ่งจำเป็นที่จะต้องพิจารณาอย่างรอบคอบ ถ้าต้องการนำผลของการปรับเทียบคะแนนไปใช้ ในกรณีที่มีการสุ่มตัวอย่างอย่างง่าย สักส่วนตัวอย่างย่อมเป็นตัวอย่างประมาณที่ไม่เอนเอียงของสัดส่วนประชากรที่สมนัยกัน และเมื่อนำไปใช้กับการปรับเทียบคะแนนของแบบทดสอบคู่ขนานสองชุด ด้วยการใช้รูปแบบการปรับเทียบที่ตำแหน่งเปอร์เซ็นต์ไทล์ (Equipercentile Method) ย่อมจะไม่มี ความลำเอียง แต่ถ้าในกรณีที่มีการแจกแจงของแบบทดสอบสองชุดที่นำมาเทียบนั้นมีการแจกแจงคนละรูปร่าง แตกต่างกันตั้งแต่โมเมนต์ที่สองขึ้นไป การใช้ฟังก์ชันการปรับเทียบคะแนนแบบเส้นตรงจะทำให้เกิดความลำเอียงได้

บรุน และฮอลแลนด์ (Braun and Holland.1982 : 36-39 อ้างถึงใน ภาวิณี ศรีสุขวัฒนานันท์. 2529 : 62) ได้เสนอบทวิเคราะห์ไว้ว่า ความลำเอียงเชิงสถิติ มีความหมายสองประการ คือ

1. ความแตกต่างของค่าเฉลี่ยของตัวประมาณ ตลอดจนการทำการสุ่มซ้ำๆ จากประชากรเดียวกัน กับค่าของประชากรที่ถูกประมาณ

2. ความแตกต่างของฟังก์ชันของการปรับเทียบคะแนนโดยประมาณ กับค่าที่แท้จริงของฟังก์ชันการปรับเทียบคะแนน

แหล่งความลำเอียงของการปรับเทียบคะแนนที่สำคัญมี 2 แหล่ง คือ

1. ความผันแปรของประชากร (Population Variability) เนื่องจากการปรับเทียบคะแนนไม่ได้จัดกระทำกับประชากรกลุ่มเดียวโดยตลอด ฟังก์ชันของการปรับเทียบคะแนนที่สร้างขึ้นเพื่อใช้กับประชากรหนึ่งอาจนำไปใช้กับอีกประชากรหนึ่งไม่ได้ เช่น ถ้า $X_p^*(y)$ เทียบคะแนนจากฟอร์ม Y ไปยังฟอร์ม X ในประชากร P และ $Y_Q^*(z)$ เทียบคะแนนจาก Z ไปยังมาตรา Y ในประชากร Q แล้วจะสรุปได้ว่า ฟังก์ชันเชิงซ้อน $X_p^*(Y_Q^*(z))$ เทียบคะแนนจาก Z ไปยัง X ในประชากร P หรือ Q ย่อมไม่ได้

ในแบบแผนการปรับเทียบคะแนนที่ใช้แบบทดสอบร่วมได้แก้ไขปัญหานี้ โดยการสร้างประชากรสังเคราะห์ (Synthesis Population) ซึ่งเป็นผลรวมของประชากร P และ Q ตามสัดส่วนแต่ยังไม่รับประกันผลว่า จะไม่มีความลำเอียงในทุก ๆ ประชากรที่แปรเปลี่ยนไป

2. ความคลาดเคลื่อนของรูปแบบ (Model Error) แบ่งเป็น 2 พวก คือ

ก. ความคลาดเคลื่อนของรูปแบบที่ทดสอบได้ (Testable Model Error) เช่นความเป็นเส้นตรงของฟังก์ชันการถดถอย สามารถทดสอบได้จากข้อมูลจำนวนมาก ซึ่งจะเห็นผลได้ว่า มีความเพียงพหรือไม่

ข. ความคลาดเคลื่อนของรูปแบบที่ทดสอบไม่ได้ (Non-testable Model Error) เป็นกรณีที่ไม่สามารถหาข้อมูลมาทดสอบเพื่อพิสูจน์ว่าสมมุติฐานที่ว่าไว้เพียงพหรือไม่

การใช้วิธีการปรับเทียบคะแนนทุกวิธี กลุ่มตัวอย่างที่ได้จากการสุ่มจากประชากรเดียวกันหรือหลายประชากร จะมีความผันแปรเชิงสุ่มเกิดขึ้นกับกลุ่มตัวอย่างที่ทำการศึกษา ดังนั้นสามารถวิเคราะห์ความคลาดเคลื่อนที่เกิดขึ้นในเชิงทฤษฎี สำหรับการปรับเทียบคะแนนระหว่างแบบสอบจากกลุ่มแทนผลได้ดังนี้

ให้ E เป็นความคลาดเคลื่อนของการปรับเทียบคะแนน

$$E = |X_i - X_i^*|$$

เมื่อ X_i คือ คะแนนของคนที่ i ที่ได้จากการสอบแบบสอบ X

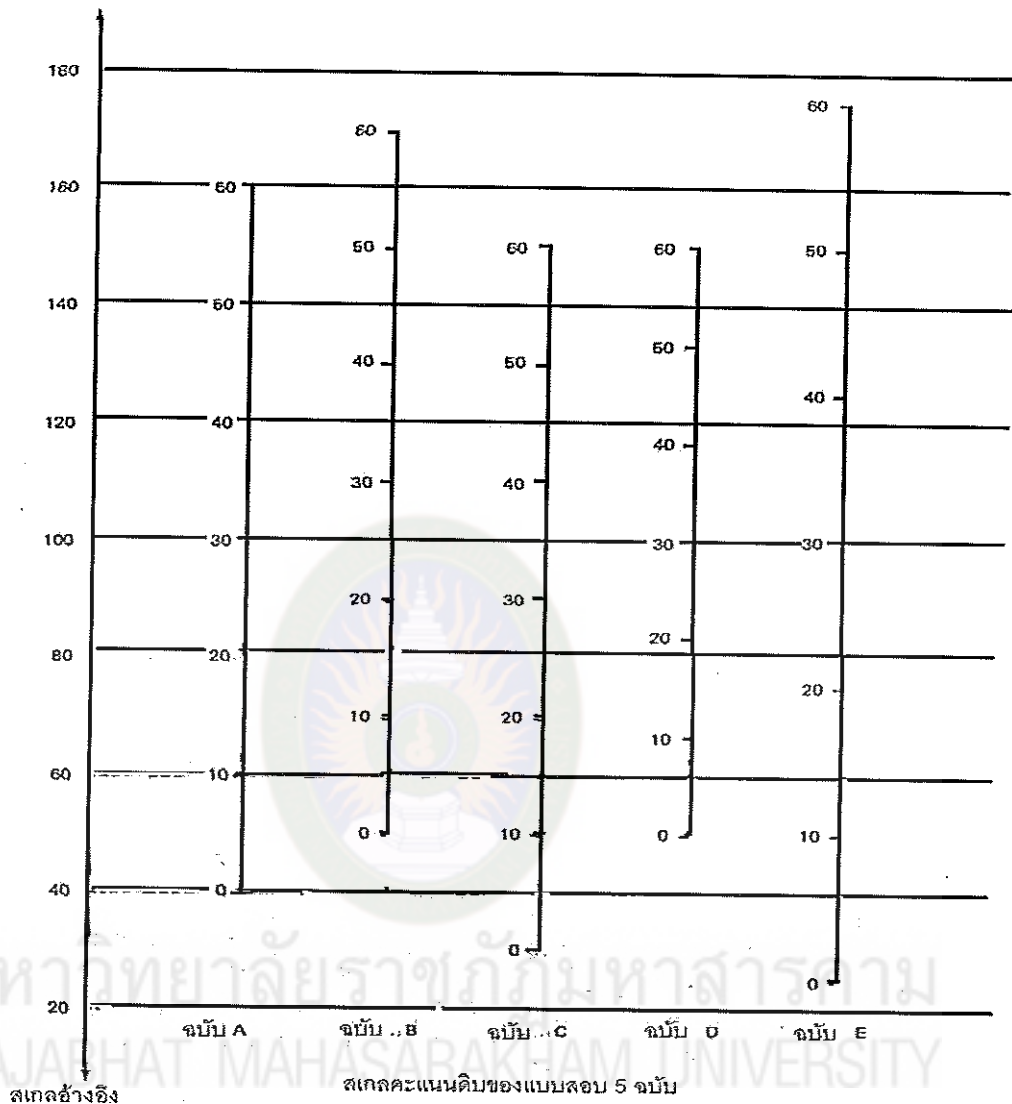
X_i^* คือ คะแนนของคนที่ i ที่ได้จากตารางเทียบคะแนน

2.3 การปรับเทียบสเกล (Calibration)

ในการปรับเทียบคะแนน (Equating) นั้น เงื่อนไขทางสถิติของแบบสอบก่อนข้างเข้มงวด คือ แบบสอบที่จะนำมาปรับเทียบคะแนนต้องวัดในคุณลักษณะเดียวกัน มีระดับความยากง่าย การกระจายของความยากของข้อสอบ และสหสัมพันธ์ระหว่างข้อสอบต้องเท่ากัน แต่ถ้าหาก

แบบสอบไม่มีคุณลักษณะดังกล่าวก็ไม่สามารถนำมาปรับเทียบคะแนนได้ อย่างไรก็ตามก็ยังคงมีการเชื่อมโยงคะแนนอีกรูปแบบหนึ่งที่ค่าสถิติของแบบสอบไม่จำเป็นต้องเท่ากัน เพียงแต่ให้เป็นแบบสอบที่วัดคุณลักษณะเดียวกันหรือคล้ายกัน ก็สามารถนำมาปรับเทียบกันได้ วิธีการปรับเทียบนี้เรียกว่า “การปรับเทียบสเกล” (Calibration) ตัวอย่างของการปรับเทียบสเกลในการวัดทางกายภาพ เช่น ในการวัดระดับความร้อน สเกลในการวัดอุณหภูมิจะเริ่มต้นตั้งแต่ $-460^{\circ} F$ หรือ $-273^{\circ} C$ ขึ้นไป แต่เทอร์โมมิเตอร์ที่สร้างขึ้นมาใช้วัดอุณหภูมิเฉพาะในบางที่ เช่น ในห้องนอน อาจจะมีสเกล $40^{\circ} - 100^{\circ} F$ เท่านั้น เทอร์โมมิเตอร์ที่สร้างขึ้นมาใช้ในกรณีนี้จะถูกปรับสเกลให้ละเอียดมากขึ้น เพื่อให้มีความถูกต้องมากขึ้น ในขณะที่พิสัยของการวัดแคบลง เช่นเดียวกับเทอร์โมมิเตอร์ที่สร้างขึ้นมาเพื่อวัดอุณหภูมิของร่างกาย เพื่อให้มีความถูกต้องสูง อาจมีการปรับสเกลให้ละเอียดขึ้น และมีพิสัยที่แคบลง ซึ่งจะมีพิสัยอยู่ระหว่าง $94^{\circ} F - 108^{\circ} F$ เท่านั้น จากที่กล่าวมาจะเห็นว่าเทอร์โมมิเตอร์ในแต่ละแบบที่สร้างขึ้นมาและปรับสเกลให้เป็นแบบใดนั้น ก็เพื่อให้มีความเหมาะสมกับจุดมุ่งหมายที่ต้องการวัด

สำหรับในกรณีของแบบสอบนั้น เราอาจสร้างแบบสอบที่ยาวชุดหนึ่งที่วัดคุณลักษณะใดคุณลักษณะหนึ่ง แล้วกำหนดสเกลของแบบสอบนี้ โดยแบบสอบนี้และสเกลที่กำหนดให้กับแบบสอบจะเป็นเกณฑ์อ้างอิงให้กับแบบสอบอื่นๆ ที่สร้างขึ้นภายหลัง หรือเป็นเกณฑ์ในการปรับเทียบสเกลของแบบสอบอื่นๆ อย่างเช่นในภาพประกอบ 6 ผลของการปรับเทียบสเกลของแบบสอบฉบับ A (ข้อสอบ 60 ข้อ) กับสเกลอ้างอิง แบบสอบฉบับ A มีพิสัยสเกลคะแนนอยู่ระหว่าง 40 ถึง 160 คะแนน (สเกลคะแนนดิบของแบบสอบทั้ง 5 ฉบับในภาพประกอบ 5 เลื่อนมาเพื่อแสดงความสัมพันธ์เชิงเส้นกับสเกลอ้างอิง) ส่วนแบบสอบฉบับ B (ข้อสอบ 60 ข้อ) มีพิสัยสเกลคะแนนอยู่ระหว่าง 50 ถึง 170 เมื่อเปรียบเทียบพิสัยของแบบสอบทั้งสองนี้ จะเห็นว่า แบบสอบ B จะมีความยากมากกว่าแบบสอบ A สำหรับกลุ่มผู้สอบที่กำหนดให้ ซึ่งมีค่าเฉลี่ยของความสามารถ 120 คะแนน ในสเกลอ้างอิง เมื่อพิจารณาคะแนนดิบจากแบบสอบ A ค่าเฉลี่ยของผู้สอบกลุ่มนี้จะเป็น 40 คะแนน ในขณะที่คะแนนเฉลี่ยของผู้สอบกลุ่มนี้ในแบบสอบ B จะต่ำกว่าคือ 35 คะแนน



ภาพ 6 ความสัมพันธ์ระหว่างสเกลคะแนนดิบของแบบสอบ 5 ฉบับกับคะแนนบนสเกลอ้างอิง
 ที่มา Angoff, W. H. (1971 : 566). Scales, norm, and equivalent score. In R. L. Thorndike (Ed.)

สำหรับแบบสอบ C นั้นจะเป็นแบบสอบที่ง่ายมากๆ เมื่อกลุ่มผู้สอบที่มีความสามารถเฉลี่ย 120 คะแนนในสเกลอ้างอิง จะมีค่าเฉลี่ยในแบบสอบ C เป็น 45 คะแนน ดังนั้นจึงเป็นที่คาดหวังว่าแบบสอบนี้จะแยกแยะผู้สอบที่มีความสามารถสูงได้น้อยที่สุด และเป็นที่คาดหวังว่าแบบสอบนี้จะแยกแยะผู้สอบที่มีความสามารถต่ำได้ดีที่สุด

สำหรับแบบสอบ D นั้นมีพิสัยของคะแนนแคบที่สุด ซึ่งยากที่จะจำแนกผู้สอบได้ทั้งในกลุ่มที่มีความสามารถสูงและกลุ่มที่มีความสามารถต่ำ อย่างไรก็ตามด้วยข้อจำกัดของพิสัยคะแนนของแบบสอบนี้จะทำให้แบบสอบนี้วัดได้ถูกต้องมากที่สุด เนื่องจากส่วนเบี่ยงเบนมาตรฐานของคะแนน

คิบมีค่าต่ำ เมื่อเทียบกับสเกลอ้างอิงพิสัยของคะแนนของแบบสอบ D จะอยู่ระหว่าง 50-150 หรือมีพิสัยเพียง 100 คะแนนเท่านั้น ซึ่งแตกต่างจากแบบสอบ E ที่มีจำนวนข้อสอบ 60 ข้อเหมือนกัน แต่มีพิสัยถึง 150 คะแนนเมื่อเทียบกับสเกลอ้างอิง

จากการปรับเทียบสเกล (Calibration) ที่กล่าวมาจะเห็นว่า แม้แบบสอบจะมีความยากแตกต่างกัน (แบบสอบ A B และ C) หรือมีความถูกต้องต่างกัน (แบบสอบ D และ E) การปรับเทียบสเกลก็ไม่ได้ทำให้สารสนเทศที่ได้จากแบบสอบนั้นเปลี่ยนไป แต่ทำให้สามารถเปรียบเทียบคะแนนจากแบบสอบหลายๆฉบับได้

จากหลักการเชื่อมโยงคะแนน (Linking) ระหว่างแบบสอบที่ได้กล่าวมาทั้งหมดนี้ จะเห็นว่าเทคนิควิธีในการเชื่อมโยงคะแนนระหว่างแบบสอบนั้นมีหลายวิธี ซึ่งแต่ละวิธีก็จะถูกนำมาใช้ในบริบทที่ต่างกัน อย่างเช่น การปรับเทียบคะแนน (Equating) จะถูกนำมาใช้ในกรณีที่แบบสอบที่นำคะแนนมาเชื่อมโยงกันนั้นเป็นแบบสอบคู่ขนาน ซึ่งวัดในคุณลักษณะเดียวกัน มีค่าสถิติของแบบสอบเท่ากัน ทั้งค่าความยากง่าย ค่าอำนาจจำแนก และค่าความเชื่อมั่นของแบบสอบ ส่วนการปรับเทียบสเกล หรือการปรับเทียบมาตรา (Calibration) นั้นเป็นวิธีการเชื่อมโยงคะแนนระหว่างแบบสอบที่วัดในคุณลักษณะที่คล้ายคลึงกัน โดยค่าสถิติของแบบสอบไม่จำเป็นต้องเท่ากัน นั่นคือคะแนนที่ได้จากแบบสอบอาจจะอยู่ในมาตรวัด (scale) ที่ต่างกันได้ และการหาความสอดคล้อง (Concordance) ซึ่งเป็นวิธีการเชื่อมโยงคะแนนระหว่างแบบสอบที่วัดในคุณลักษณะที่ต่างกัน ค่าสถิติของแบบสอบก็ไม่จำเป็นต้องเท่ากัน เป็นต้น ทั้งนี้วิธีการเชื่อมโยงคะแนนแบบต่างๆ ก็จะมีเทคนิคในการเชื่อมโยงคะแนนที่แตกต่างกันไปอีกมากมายหลายวิธี

วัตถุประสงค์หลักของการเชื่อมโยงคะแนนระหว่างแบบสอบก็คือ เพื่อให้สามารถเปรียบเทียบคะแนนระหว่างแบบสอบได้ หรือเพื่อให้สามารถนำคะแนนจากแบบสอบฉบับหนึ่งมาใช้แทนคะแนนจากแบบสอบอีกฉบับหนึ่งได้อย่างยุติธรรม และมีเหตุผลที่เชื่อถือได้ นอกจากนี้ยังสามารถประยุกต์การเชื่อมโยงคะแนนมาใช้ปรับคะแนนจากแบบสอบต่างฉบับกันให้มีมาตรฐานเดียวกันสามารถเปรียบเทียบกันได้อีกด้วย

ด้วยหลักการของการเชื่อมโยงคะแนนดังกล่าว จึงสามารถนำมาใช้ในการปรับเทียบผลการเรียน ทั้งที่เป็นผลการเรียนเฉลี่ยสะสมตลอดหลักสูตร (GPAX) และผลการเรียนเฉลี่ยสะสมตามกลุ่มสาระการเรียนรู้ (GPA) ของโรงเรียนต่างๆ ที่มีมาตรฐานต่างกัน ให้มีมาตรฐานเดียวกันเพื่อใช้ในการระบบการรับเข้าศึกษาได้ และเนื่องจากผลการเรียนเหล่านี้ได้จากการประเมินผลด้วยแบบสอบที่มีค่าสถิติของแบบสอบแตกต่างกัน (ต่างโรงเรียนใช้แบบสอบต่างกัน) แต่ก็เป็นกรวัดในคุณลักษณะเดียวกัน (เนื้อหาเดียวกัน) วิธีการเชื่อมโยงคะแนนที่นำมาใช้จึงเป็นวิธีการปรับเทียบสเกล หรือ การปรับเทียบมาตรา (Calibration)

ดังนั้นในการวิจัยครั้งนี้ผู้วิจัยจึงใช้ การปรับเทียบผลการเรียน (Grade Calibration) ในการปรับผลการเรียนเฉลี่ยสะสมตลอดหลักสูตร (GPAX) ของโรงเรียนต่างๆ ให้อยู่ในมาตรฐานเดียวกัน เพื่อนำไปใช้ในระบบกลางการรับเข้าศึกษา (Admissions) โดยใช้เทคนิคการปรับสเกล (Scaling Calibration) ซึ่งจะได้กล่าวในรายละเอียดต่อไป

3. วิธีการปรับเทียบผลการเรียน

ปิยะธิดา ปัญญา (2551 : 109-127) ได้พัฒนาวิธีการปรับเทียบผลการเรียน โดยอาศัยหลักการเชื่อมโยงเชิงเส้น ได้วิธีการปรับเทียบผลการเรียน (Grade Calibration Methods) ได้ 3 วิธีดังนี้

3.1 วิธีการปรับเทียบผลการเรียนแบบเชื่อมโยงส่วนเดียว (One - Section Linking Grade Calibration: OLC)

วิธีการปรับเทียบผลการเรียนแบบเชื่อมโยงส่วนเดียว เป็นวิธีที่พัฒนาขึ้นจากแนวคิดที่ว่า ผลการเรียนของนักเรียนแต่ละคนภายในโรงเรียนหนึ่งๆ ได้รับความเอนเอียงจากระบบการตัดสินผลการเรียนแบบสุ่ม (Random Effect) นั่นคือ เมื่อแต่ละโรงเรียนใช้ระบบการตัดสินผลการเรียนแบบกำหนดระดับผลสัมฤทธิ์เป็นตัวอักษร ผู้เรียนที่ได้รับการตัดสินผลการเรียนร่วมกันในแต่ละกลุ่มสาระการเรียนรู้ จะมีทั้งผู้ที่ได้เปรียบและผู้ที่เสียเปรียบในการตัดสินผลการเรียนแต่ละครั้ง แต่เนื่องจากผลการเรียนเฉลี่ยสะสมในแต่ละกลุ่มสาระการเรียนรู้ และผลการเรียนเฉลี่ยสะสมตลอดหลักสูตร เป็นผลการเรียนเฉลี่ยของนักเรียนแต่ละคนที่เกิดจากการตัดสินผลการเรียนตลอดหลักสูตร หรือ 6 ภาคการศึกษา หากความเอนเอียงของผลการเรียนของนักเรียนแต่ละคนเกิดขึ้นอย่างสุ่มภายใน 6 ภาคการศึกษานี้ ค่าเฉลี่ยของผลการเรียนที่เอนเอียงไปจากความสามารถที่ควรจะเป็นของนักเรียนแต่ละคน เนื่องจากอิทธิพลของระบบการตัดสินผลการเรียนจะหักล้างกันเป็นศูนย์ ดังนั้นผลการเรียนของนักเรียนแต่ละคนจึงสามารถนำไปเปรียบเทียบกับนักเรียนคนอื่นๆ ภายในโรงเรียนเดียวกันได้ แต่ในระหว่างโรงเรียน มาตรฐานของผลการเรียนของแต่ละโรงเรียนนั้นยังแตกต่างกัน จึงไม่สามารถนำผลการเรียนของนักเรียนในโรงเรียนหนึ่งไปเปรียบเทียบกับนักเรียนในโรงเรียนอื่นได้ นั่นคือ มีความเอนเอียงของผลการเรียนในส่วนหนึ่งของโรงเรียน ดังนั้นจึงต้องทำการปรับเทียบผลการเรียน ระหว่างโรงเรียนก่อน ตามวิธีการปรับเทียบผลการเรียนแบบเชื่อมโยงส่วนเดียว ซึ่งมีขั้นตอนในการปรับเทียบ 2 ขั้นตอน ดังนี้

3.1.1 การปรับเทียบมาตรฐานของผลการเรียนระหว่างโรงเรียน

เนื่องจากการปรับเทียบสเกล (Calibration) นั้นต้องอาศัยสเกลอ้างอิง (reference scale) เพื่อเป็นสเกลในการปรับเทียบ โดยสเกลดังกล่าวเป็นคะแนนจากแบบสอบที่นักเรียนทุกคนสอบร่วมกัน ดังนั้นในการปรับเทียบผลการเรียนครั้งนี้ จึงใช้คะแนนจากผลการทดสอบทาง

การศึกษาแห่งชาติขั้นพื้นฐาน (O-NET) ที่นักเรียนทุกคนต้องสอบเป็นสเกลอ้างอิง และจากการศึกษาความสัมพันธ์ระหว่างผลการเรียนเฉลี่ยสะสม (GPA) และคะแนน O-NET พบว่ามีความสัมพันธ์กันเป็นเชิงเส้นตรง ซึ่งมีเหตุผลสนับสนุนที่สำคัญ 2 ประการคือ

ประการแรก การทดสอบทางการศึกษาแห่งชาติขั้นพื้นฐาน (O-NET) เป็นการวัดผลการจัดการศึกษาขั้นพื้นฐาน ซึ่งจัดสอบโดยสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) ที่เป็นหน่วยงานกลาง โดยวัตถุประสงค์เบื้องต้นของการทดสอบเพื่อประโยชน์ในการประกันคุณภาพการศึกษา และเป็นดัชนีบ่งชี้เพื่อใช้ในการปรับปรุงคุณภาพการเรียนการสอนของโรงเรียน โดยวัตถุประสงค์ของการทดสอบทางการศึกษาแห่งชาติขั้นพื้นฐาน (O-NET) ต้องการวัดความรู้ความสามารถของนักเรียนตามหลักสูตร เช่นเดียวกับผลการเรียนเฉลี่ยสะสมกลุ่มสาระการเรียนรู้ (GPA) หรืออาจกล่าวได้ว่า คะแนน O-NET และผลการเรียนเฉลี่ยสะสมกลุ่มสาระการเรียนรู้ (GPA) วัดในคุณลักษณะเดียวกัน ดังนั้นจึงควรมีความสัมพันธ์กันเชิงเส้นตรง

ประการที่สอง จากการวิเคราะห์ความสัมพันธ์ระหว่างผลการเรียนเฉลี่ยสะสมของนักเรียนกับคะแนน O-NET ประจำปีการศึกษา 2548 ของศูนย์ปฏิบัติการ GPA สำนักทดสอบทางการศึกษา สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน พบว่า ร้อยละ 68.19 ของโรงเรียนทั้งหมดที่มีผลการเรียนเฉลี่ยสะสมของนักเรียนสัมพันธ์กับคะแนน O-NET สูง โดยมีค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สันตั้งแต่ 0.60 ขึ้นไป ซึ่งความสัมพันธ์ในระดับดังกล่าวมีลักษณะเป็นความสัมพันธ์เชิงเส้นตรงชัดเจน ส่วนที่เหลืออีกร้อยละ 31.81 ของโรงเรียนทั้งหมดมีค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างผลการเรียนเฉลี่ยสะสมกับคะแนน O-NET ต่ำกว่า 0.60 แสดงให้เห็นว่า ผลการเรียนของนักเรียนในโรงเรียนประมาณ 1 ใน 3 ของโรงเรียนทั้งหมดยังมีความเอนเอียงไปจากความสามารถที่ควรจะเป็นของผู้เรียน จึงทำให้ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่าง ผลการเรียนเฉลี่ยสะสมกับคะแนน O-NET มีค่าต่ำ

เมื่อผลการเรียนเฉลี่ยสะสม (GPA) กับคะแนน O-NET (ONET) มีความสัมพันธ์กันเชิงเส้นตรง จะสามารถเขียนแทนความสัมพันธ์ในรูปสมการเส้นตรงได้ดังนี้

$$Y = a + bX$$

ดังนั้นความสัมพันธ์ระหว่างผลการเรียนเฉลี่ยสะสม (GPA) กับคะแนน O-NET (ONET) จึงเขียนได้เป็น

$$GPA = a + b(ONET) \quad (2.1)$$

โดย a เป็นจุดตัดของเส้นตรงบนแกน GPA หรือ GPA-Intercept และ b เป็นความชันของเส้นตรง (slope)

ถ้ากำหนดให้ \overline{GPA}_j เป็นค่าเฉลี่ยของผลการเรียนเฉลี่ยสะสมของนักเรียนทุกคนในโรงเรียน j แล้ว ผลการเรียนของนักเรียนแต่ละคนในโรงเรียน j ซึ่งเขียนแทนด้วย GPA_{ij} จะมีความสัมพันธ์กับ \overline{GPA}_j ดังนี้

$$\overline{GPA}_j = \frac{\sum GPA_{ij}}{n_j}$$

การเขียนสมการแสดงผลการเรียนของนักเรียนแต่ละคนในโรงเรียน j ให้ประกอบไปด้วย 2 ส่วน ที่ต้องการลดความเอนเอียงสามารถเขียนได้ดังนี้

เมื่อผลการเรียนของนักเรียนแต่ละคนในโรงเรียน j เขียนแทนด้วย

$$GPA_{ij} = GPA_{ij}$$

ต้องการแยกผลการเรียนของนักเรียนออกเป็นสองส่วน โดย นำ \overline{GPA}_j บวกเข้าและลบออก

$$\begin{aligned} GPA_{ij} &= GPA_{ij} + (-\overline{GPA}_j + \overline{GPA}_j) \\ &= (GPA_{ij} - \overline{GPA}_j) + \overline{GPA}_j \\ GPA_{ij} &= \Delta GPA_{ij} + \overline{GPA}_j \end{aligned}$$

จะได้ผลการเรียนของนักเรียนแต่ละคนในโรงเรียน j ที่ประกอบไปด้วย 2 ส่วน คือ GPA ส่วนของโรงเรียนที่เป็นค่าเฉลี่ยของโรงเรียน (\overline{GPA}_j) และ GPA ส่วนของนักเรียน ที่เป็นผลต่างของนักเรียนที่ต่างไปจากค่าเฉลี่ยของโรงเรียน (ΔGPA_{ij}) ดังนี้

$$GPA_{ij} = \overline{GPA}_j + \Delta GPA_{ij} \quad (2.2)$$

$$\text{เมื่อ} \quad \Delta GPA_{ij} = GPA_{ij} - \overline{GPA}_j \quad (2.3)$$

ซึ่ง ΔGPA_{ij} สามารถมีค่าเป็นบวกหรือลบก็ได้ กล่าวคือ ถ้านักเรียนคนที่มีผลการเรียนสูงกว่าค่าเฉลี่ย ค่า ΔGPA_{ij} จะมีค่าเป็นบวก ส่วนนักเรียนที่มีผลการเรียนต่ำกว่าค่าเฉลี่ย ค่า ΔGPA_{ij} จะมีค่าเป็นลบ สมมติ เช่น ค่าเฉลี่ยผลการเรียนของนักเรียนในโรงเรียนแห่งหนึ่ง มีค่า 2.50 นักเรียนคนหนึ่ง ในโรงเรียนแห่งนี้มีผลการเรียน 3.25 ดังนั้นค่า ΔGPA_{ij} ของนักเรียนคนนี้จะมีความหมายคือ $3.25 - 2.50 = 0.75$ ในขณะที่นักเรียนอีกคนหนึ่ง ในโรงเรียนแห่งนี้เหมือนกันมีผลการเรียน 2.00 ดังนั้นค่า ΔGPA_{ij} ของนักเรียนคนนี้จะมีความหมายคือ $2.00 - 2.50 = -0.50$ เป็นต้น

ในการเปรียบเทียบผลการเรียนแบบเชื่อมโยงส่วนเดียวนี้ มีแนวคิดที่ว่ามาตรฐานของผลการเรียนของแต่ละโรงเรียนแตกต่างกัน แต่ระบบการตัดสินผลการเรียนไม่ทำให้ผลการเรียนของนักเรียนเอนเอียงไปจากความสามารถที่ควรจะเป็น จึงนำเอาเฉพาะค่าผลการเรียนเฉลี่ยของแต่ละโรงเรียน (\overline{GPA}_j) มาทำการเปรียบเทียบเท่านั้น ซึ่งถ้าให้

\overline{GPA}_j เป็นผลการเรียนเฉลี่ยของแต่ละโรงเรียน ซึ่งเป็นค่าเฉลี่ยของผลการเรียนเฉลี่ยสะสมของนักเรียนทุกคนในโรงเรียน j

\overline{ONET}_j เป็นคะแนน O-NET เฉลี่ยของแต่ละโรงเรียน ซึ่งเป็นค่าเฉลี่ยของคะแนน O-NET ของนักเรียนทุกคนในโรงเรียน j

เมื่อนำผลการเรียนเฉลี่ยของแต่ละโรงเรียน (\overline{GPA}_j) และคะแนน O-NET เฉลี่ยของแต่ละโรงเรียน (\overline{ONET}_j) มาวิเคราะห์การถดถอยโดยใช้โมเดลการวิเคราะห์สมการถดถอยอย่างง่ายจะได้

$$\overline{GPA}_j = b_0 + b_1 \overline{ONET}_j + \varepsilon_j$$

ถ้าผลการเรียนเฉลี่ยของแต่ละโรงเรียนอยู่บนมาตรฐานเดียวกัน คือ ไม่มีความเอนเอียงของผลการเรียนในส่วนของโรงเรียน ค่าความคลาดเคลื่อนของผลการเรียนแต่ละโรงเรียน (ε_j) จะเป็นศูนย์ นั่นคือ ผลการเรียนเฉลี่ยของแต่ละโรงเรียนทุกโรงเรียนจะอยู่บนเส้นถดถอยเดียวกัน ดังนั้นผลการเรียนเฉลี่ยที่ควรจะเป็นของโรงเรียน j (\widehat{GPA}_j) จะมีค่าเป็น

$$\widehat{GPA}_j = b_0 + b_1 \overline{ONET}_j \quad (2.4)$$

เมื่อนำผลการเรียนเฉลี่ยที่ควรจะเป็น (\widehat{GPA}_j) ในสมการ (2.4) ไปแทนผลการเรียนเฉลี่ย (\overline{GPA}_j) ในสมการ (2.2) จะได้ผลการเรียนที่ควรจะเป็นของนักเรียนแต่ละคน (\widehat{GPA}_{ij}) ดังนี้

$$\widehat{GPA}_{ij} = b_0 + b_1 \overline{ONET}_j + \Delta GPA_{ij}$$

หรือ

$$\widehat{GPA}_{ij} = b_0 + b_1 \overline{ONET}_j + GPA_{ij} - \overline{GPA}_j \quad (2.5)$$

เมื่อ \widehat{GPA}_{ij} = ผลการเรียนที่ปรับแล้วของนักเรียนคนที่ i ในโรงเรียนที่ j

GPA_{ij} = ผลการเรียนเดิมของนักเรียนคนที่ i ในโรงเรียนที่ j

\overline{ONET}_j = คะแนน O-NET ของโรงเรียนที่ j

= ค่าเฉลี่ยของคะแนน O-NET ของนักเรียนทุกคนในโรงเรียนที่ j

\overline{GPA}_j = ผลการเรียนของโรงเรียนที่ j

= ค่าเฉลี่ยของผลการเรียนของนักเรียนทุกคนในโรงเรียนที่ j

b_0 = ค่าคงที่หรือ \overline{GPA}_j Intercept ของเส้นตรงความสัมพันธ์ระหว่าง \overline{GPA}_j กับ \overline{ONET}_j

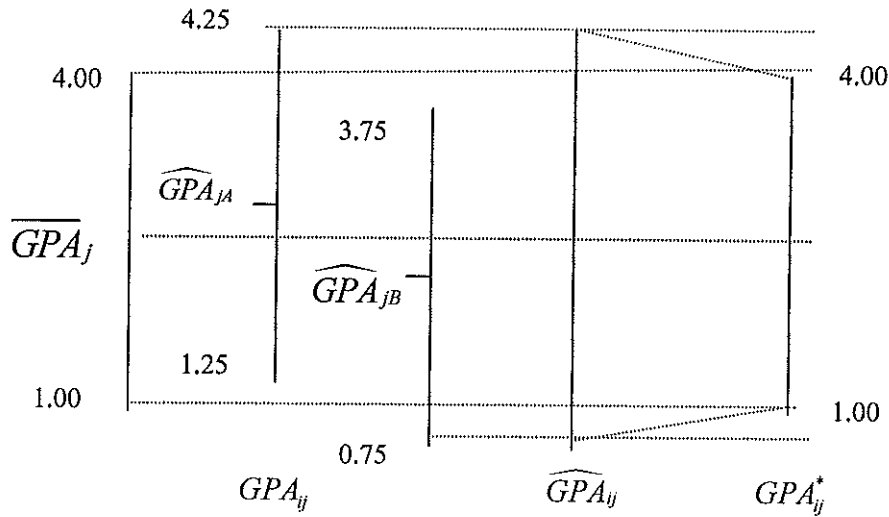
b_1 = ค่าความชัน (Slope) ของเส้นตรงความสัมพันธ์ระหว่าง \overline{GPA}_j กับ \overline{ONET}_j

ในกรณีของผลการเรียนเฉลี่ยสะสมตลอดหลักสูตร ($GPAX$) ใช้คะแนนรวม $ONET$ ($TONET$) แทนคะแนน $ONET$ และดำเนินการเช่นเดียวกับกรณีของผลการเรียนเฉลี่ยสะสมตามกลุ่มสาระการเรียนรู้ (GPA)

3.1.2 การปรับสเกลตามระบบการตัดสินผลการเรียน

การปรับผลการเรียนตามสมการ (2.5) นี้ จะทำให้ค่าสูงสุดและค่าต่ำสุดของผลการเรียนที่ปรับแล้ว (\widehat{GPA}_{max} และ \widehat{GPA}_{min}) มีค่าไม่เป็นไปตามธรรมชาติของผลการเรียน ซึ่งโดยปกติผลการเรียนเฉลี่ยสะสมของนักเรียนที่สำเร็จการศึกษาตามหลักสูตร จะมีค่าไม่ต่ำกว่า 1.00 และมีค่าสูงสุดไม่เกิน 4.00 แต่หากนักเรียนคนหนึ่งมีผลการเรียนเฉลี่ยสะสม 4.00 อยู่ในโรงเรียนที่มีมาตรฐานของผลการเรียนสูงหรือได้รับอิทธิพลเพดาน หรือกีดเกรดในการประเมินผลการเรียน อิทธิพลดังกล่าวจะทำให้ผลการเรียนเดิมของโรงเรียนหรือค่าเฉลี่ยของผลการเรียนของนักเรียนทุกคนในโรงเรียนนั้น (\overline{GPA}_j) มีค่าต่ำกว่าที่ควรจะเป็น เมื่อทำการปรับเทียบด้วยสมการ (2.5) จะทำให้ผลการเรียนของโรงเรียนสูงขึ้นเท่ากับผลการเรียนที่ควรจะเป็นของโรงเรียน (\widehat{GPA}_j) ซึ่งจะส่งผลให้ผลการเรียนที่ควรจะเป็นของนักเรียนคนนั้น (\widehat{GPA}_{ij}) มีค่าสูงกว่าผลการเรียนเดิม (สูงกว่า 4.00) ในทางตรงข้ามถ้านักเรียนคนหนึ่งมีผลการเรียนเฉลี่ยสะสม 1.00 และอยู่ในโรงเรียนที่มีมาตรฐานผลการเรียนต่ำหรือปล่อยเกรด ผลการเรียนเดิมของโรงเรียนหรือค่าเฉลี่ยของผลการเรียนของนักเรียนทุกคนในโรงเรียนนั้น (\overline{GPA}_j) จะมีค่าสูงกว่าที่ควรจะเป็น เมื่อทำการปรับเทียบด้วยสมการ (2.5) จะทำให้ผลการเรียนของโรงเรียนนี้ต่ำลงเท่ากับผลการเรียนที่ควรจะเป็นของโรงเรียน (\widehat{GPA}_j) ซึ่งจะส่งผลให้ผลการเรียนของนักเรียนคนนั้น (\widehat{GPA}_{ij}) มีค่าต่ำกว่าผลการเรียนเดิม (ต่ำกว่า 1.00) ดังนั้นจึงต้องปรับให้ผลการเรียนที่ผ่านการปรับเทียบแล้ว มีค่าสูงสุดและค่าต่ำสุดเท่ากับค่าสูงสุดและค่าต่ำสุดของผลการเรียนเดิม เช่นเดียวกับการปรับมาตรฐานรองศาอุณหภูมิจึงนี้

สมมติโรงเรียน A เป็นโรงเรียนที่มีการกีดเกรด และโรงเรียน B เป็นโรงเรียนที่มีการปล่อยเกรด โดยทั้งสองโรงเรียนมีผลการเรียนเฉลี่ย (\overline{GPA}_j) เท่ากัน และผลการเรียนเฉลี่ยสะสมของนักเรียนในโรงเรียน (GPA_{ij}) มีค่าต่ำสุดเท่ากับ 1.00 และสูงสุดเท่ากับ 4.00 เหมือนกัน เมื่อนำผลการเรียนมาปรับเทียบเพื่อให้ผลการเรียนของโรงเรียนทั้งสองมีมาตรฐานเดียวกันตามสมการ (2.5) จะได้ผลการเรียนหลังการปรับเทียบดังนี้



ภาพ 7 ผลการเรียนหลังการปรับเทียบ

จากภาพ 7 จะเห็นว่า หลังการปรับเทียบให้ผลการเรียนทั้งสองโรงเรียนมีมาตรฐานเดียวกัน ผลการเรียนสูงสุดของโรงเรียน A ซึ่งเดิมถูกกคกเดมมีค่า 4.00 จะถูกปล่อยให้มีค่าสูงกว่าเดิม คือ 4.25 และมีค่าเฉลี่ยของผลการเรียนทั้งโรงเรียนเป็น \widehat{GPA}_{jA} ในขณะที่ผลการเรียนต่ำสุดของโรงเรียน B ซึ่งเดิมปล่อยเกรดมีค่า 1.00 จะถูกกคกให้ต่ำกว่าเดิมคือ 0.75 และค่าเฉลี่ยของผลการเรียนทั้งโรงเรียนเป็น \widehat{GPA}_{jB} ดังนั้นสเกลใหม่หลังการปรับเทียบจะมีค่าสูงสุด 4.25 และต่ำสุด 0.75 ซึ่งไม่สอดคล้องกับธรรมชาติของผลการเรียน จึงต้องปรับสเกลใหม่ (\widehat{GPA}_{ij}) ให้มีค่าต่ำสุดและสูงสุด สอดคล้องกับธรรมชาติของผลการเรียนเดิม (GPA_{ij}) ซึ่งจะได้สเกลใหม่เป็น GPA_{ij}^* ที่มีค่าต่ำสุดและสูงสุดเท่ากับค่าต่ำสุดและสูงสุดสอดคล้องกับธรรมชาติของผลการเรียนเดิม (GPA_{ij}) ดังภาพ 7 โดยในการปรับสเกล \widehat{GPA}_{ij} ให้เป็นสเกล GPA_{ij}^* เพื่อไม่ให้ตำแหน่งของผลการเรียนของนักเรียนทั้งหมดเปลี่ยนไป ดังนั้นสัดส่วนของระยะสเกลเดิมต่อพิสัยเดิม จะต้องเท่ากับสัดส่วนของระยะสเกลใหม่ต่อพิสัยใหม่ หรือ

$$\frac{\text{ระยะสเกลเดิม}}{\text{พิสัยสเกลเดิม}} = \frac{\text{ระยะสเกลใหม่}}{\text{พิสัยสเกลใหม่}}$$

หากผลการเรียนที่ผ่านการปรับเทียบแล้วเป็น \widehat{GPA}_{ij} เมื่อปรับให้สเกลของผลการเรียนเป็น GPA_{ij}^* จะได้

$$\frac{GPA_{ij}^* - GPA_{\min}}{GPA_{\max} - GPA_{\min}} = \frac{\widehat{GPA}_{ij} - \widehat{GPA}_{\min}}{\widehat{GPA}_{\max} - \widehat{GPA}_{\min}}$$

$$\text{หรือ} \quad GPA_{ij}^* - GPA_{\min} = \frac{(\widehat{GPA}_{ij} - \widehat{GPA}_{\min})(GPA_{\max} - GPA_{\min})}{(\widehat{GPA}_{\max} - \widehat{GPA}_{\min})}$$

$$\text{ดังนั้น} \quad GPA_{ij}^* = \frac{(\widehat{GPA}_{ij} - \widehat{GPA}_{\min})(GPA_{\max} - GPA_{\min})}{(\widehat{GPA}_{\max} - \widehat{GPA}_{\min})} + GPA_{\min} \quad (2.6)$$

- เมื่อ GPA_{ij}^* = ผลการเรียนของนักเรียนแต่ละคนที่ปรับตามระบบการตัดสินผลการเรียน
 \widehat{GPA}_{ij} = ผลการเรียนของนักเรียนแต่ละคนที่ยังไม่ได้ปรับตามระบบการตัดสินผลการเรียน
 \widehat{GPA}_{\max} = ผลการเรียนที่ยังไม่ได้ปรับตามระบบการตัดสินผลการเรียนที่มีค่าสูงสุด
 \widehat{GPA}_{\min} = ผลการเรียนที่ยังไม่ได้ปรับตามระบบการตัดสินผลการเรียนที่มีค่าต่ำสุด
 GPA_{\max} = ผลการเรียนที่ยังไม่ได้ผ่านการปรับเทียบผลการเรียนที่มีค่าสูงสุด
 GPA_{\min} = ผลการเรียนที่ยังไม่ได้ผ่านการปรับเทียบผลการเรียนที่มีค่าต่ำสุด

ตัวอย่าง จากสถานการณ์ที่กล่าวมาข้างต้น ถ้านักเรียนคนหนึ่งของโรงเรียน A มีผลการเรียนเฉลี่ยสะสม 3.00 (ผลการเรียนที่ถูกกดเกรด) เมื่อทำการปรับเทียบผลการเรียนให้มีมาตรฐานเดียวกันแล้ว จะทำให้มีผลการเรียนเป็น 3.25 และเมื่อนำมาปรับสเกลตามสมการ (2.6) จะได้ดังนี้

$$GPA_{ij}^* = \frac{(3.25 - 0.75)(4.00 - 1.00)}{(4.25 - 0.75)} + 1.00 = 3.14$$

ส่วนนักเรียนในโรงเรียน A ที่มีผลการเรียนสูงสุด 4.00 เมื่อปรับเทียบผลการเรียนให้มีมาตรฐานเดียวกันจะมีผลการเรียน 4.25 และเมื่อนำมาปรับเทียบสเกลตามสมการ (2.6) จะได้

$$GPA_{ij}^* = \frac{(4.25 - 0.75)(4.00 - 1.00)}{(4.25 - 0.75)} + 1.00 = 4.00$$

ในขณะที่นักเรียนในโรงเรียน B (ปล่อยเกรด) ที่มีผลการเรียนสูงสุด 4.00 เมื่อปรับเทียบผลการเรียนให้มีมาตรฐานเดียวกันจะมีผลการเรียน 3.75 และเมื่อนำมาปรับเทียบสเกลตามสมการ (2.6) จะได้

$$GPA_{ij}^* = \frac{(3.75 - 0.75)(4.00 - 1.00)}{(4.25 - 0.75)} + 1.00 = 3.57$$

ส่วนนักเรียนในโรงเรียน B ที่มีผลการเรียนต่ำสุด 1.00 เมื่อปรับเทียบผลการเรียนให้มีมาตรฐานเดียวกันจะมีผลการเรียน 0.75 และเมื่อนำมาปรับเทียบสเกลตามสมการ (2.6) จะได้

$$GPA_{ij}^* = \frac{(0.75 - 0.75)(4.00 - 1.00)}{(4.25 - 0.75)} + 1.00 = 1.00$$

จากตัวอย่างจะเห็นว่า นักเรียนในโรงเรียนที่ถูกกวดเกรด ถ้าเดิมมีผลการเรียนเป็น 3.00 เมื่อเปรียบเทียบผลการเรียนแบบเชื่อมโยงส่วนเดียวแล้ว จะมีผลการเรียนเป็น 3.25 และเมื่อนำมาปรับสเกลตามระบบการตัดสินผลการเรียน จะมีผลการเรียนเป็น 3.14 (เพิ่มขึ้นจากผลการเรียนเดิมก่อนการเปรียบเทียบจาก 3.00 เป็น 3.14) ในขณะที่นักเรียนที่มีผลการเรียนสูงสุด 4.00 เมื่อเปรียบเทียบผลการเรียนแบบเชื่อมโยงส่วนเดียวแล้ว จะมีผลการเรียนเป็น 4.25 และเมื่อนำมาปรับสเกลตามระบบการตัดสินผลการเรียน จะคงมีผลการเรียนเป็น 4.00 (ซึ่งก็เป็นผลการเรียนที่อยู่ในลำดับที่สูงที่สุดในระบบการตัดสินผลการเรียน ซึ่งเดิมที่นักเรียนคนนี้ก็มีการเรียนอยู่ในลำดับสูงสุดอยู่แล้ว)

ส่วนนักเรียนในโรงเรียนที่ปล่อยเกรด ถ้าเดิมมีผลการเรียนสูงสุดเป็น 4.00 เมื่อเปรียบเทียบผลการเรียนแบบเชื่อมโยงส่วนเดียวแล้ว จะมีผลการเรียนเป็น 3.75 และเมื่อนำมาปรับสเกลตามระบบการตัดสินผลการเรียน จะมีผลการเรียนเป็น 3.57 ในขณะที่นักเรียนที่มีผลการเรียนต่ำสุดของโรงเรียนนี้ (1.00) เมื่อเปรียบเทียบผลการเรียนแบบเชื่อมโยงส่วนเดียวแล้วจะมีผลการเรียนเป็น 0.75 และเมื่อนำมาปรับสเกลตามระบบการตัดสินผลการเรียน จะมีผลการเรียนเป็น 1.00 เช่นเดิม (ซึ่งก็เป็นผลการเรียนที่อยู่ในลำดับที่ต่ำที่สุดในระบบการตัดสินผลการเรียน ซึ่งเดิมที่นักเรียนคนนี้ก็มีการเรียนอยู่ในลำดับต่ำสุดอยู่แล้ว)

ตัวอย่างการปรับสเกลผลการเรียนกลุ่มสาระการเรียนรู้ภาษาไทยที่ปรับเทียบด้วยวิธีการเปรียบเทียบผลการเรียนแบบเชื่อมโยงส่วนเดียว (One - Section Linking Grade Calibration: OLC)

ถ้านักเรียนคนหนึ่งของโรงเรียน A ได้คะแนน O-NET 78.75 คะแนน มีผลการเรียนเฉลี่ยสะสม 3.58 (ผลการเรียนที่ถูกกวดเกรด) โดยโรงเรียนมีผลการเรียนเฉลี่ยสะสมสูงสุดเป็น 3.58 และต่ำสุดเป็น 1.00 เมื่อทำการปรับเทียบผลการเรียนให้มีมาตรฐานเดียวกันแล้ว จะทำให้มีผลการเรียนเป็น 4.93 และเมื่อนำมาปรับสเกลตามสมการ (2.6) จะได้เป็น 4.00

และถ้านักเรียนคนหนึ่งของโรงเรียน B ได้คะแนน O-NET 36.25 คะแนน มีผลการเรียนเฉลี่ยสะสม 1.16 (ผลการเรียนที่ถูกปล่อยเกรด) โดยโรงเรียนมีผลการเรียนเฉลี่ยสะสมสูงสุดเป็น 4.00 และต่ำสุดเป็น 1.00 เมื่อทำการปรับเทียบผลการเรียนให้มีมาตรฐานเดียวกันแล้ว จะทำให้มีผลการเรียนเป็น 0.86 และเมื่อนำมาปรับสเกลตามสมการ (2.6) จะได้เป็น 1.00 เป็นต้น

โดยสรุปแล้ววิธีการปรับเทียบผลการเรียนแบบเชื่อมโยงส่วนเดียว (OLC) พัฒนาขึ้นตามแนวคิดที่ว่า ความเอนเอียงของผลการเรียนเกิดจากมาตรฐานของผลการเรียนระหว่างโรงเรียน ไม่เท่าเทียมกัน แต่มาตรฐานของผลการเรียนภายในโรงเรียนเป็นมาตรฐานเดียวกัน จึงทำการปรับเทียบผลการเรียนเฉพาะในระหว่างโรงเรียนเท่านั้น ซึ่งมีขั้นตอนการปรับเทียบ 2 ขั้นตอน คือ ขั้นตอนการปรับเทียบมาตรฐานผลการเรียนตามสมการ (2.5) และขั้นตอนการปรับสเกลตามสมการ (2.6) ในกรณีของ

ผลการเรียนเฉลี่ยสะสมตลอดหลักสูตร ($GPAX$) ก็ดำเนินการเช่นเดียวกับกรณีของผลการเรียนเฉลี่ยสะสมตามกลุ่มสาระการเรียนรู้ (GPA)

3.2 วิธีการปรับเทียบผลการเรียนแบบเชื่อมโยงเสริมกันสองส่วน (Two-Section Additive Linking Grade Calibration: TALC)

วิธีการปรับเทียบผลการเรียนแบบเชื่อมโยงเสริมกันทั้งสองส่วน (TALC) เป็นวิธีที่พัฒนาขึ้นจากแนวคิดที่ว่า ผลการเรียนที่เอนเอียงไปของนักเรียนเป็นผลมาจากมาตรฐานการประเมินผลการเรียนที่แตกต่างกันระหว่างโรงเรียน ซึ่งเป็นความเอนเอียงระหว่างโรงเรียนหรือความเอนเอียงในส่วนของโรงเรียนเสริมกับความเอนเอียงของผลการเรียนที่เป็นผลมาจากระบบการตัดสินผลการเรียนแบบกำหนดระดับผลสัมฤทธิ์เป็นตัวอักษร ซึ่งเป็นความเอนเอียงภายในโรงเรียนหรือความเอนเอียงในส่วนของผู้เรียน ดังนั้นการปรับเทียบให้ผลการเรียนของนักเรียนทุกคนมีมาตรฐานเดียวกัน จึงต้องปรับทั้งผลการเรียนในส่วนของโรงเรียน และผลการเรียนในส่วนของผู้เรียน โดยอาศัยคะแนนผลการสอบทางการศึกษาแห่งชาติขั้นพื้นฐาน (O-NET) เป็นสเกลอ้างอิงในการปรับเทียบ ซึ่งมีขั้นตอนในการปรับเทียบ 2 ขั้นตอนดังนี้

3.2.1 การปรับเทียบผลการเรียนให้มีมาตรฐานเดียวกัน

เมื่อความเอนเอียงของผลการเรียนเกิดขึ้นในสองส่วนคือ ส่วนของโรงเรียน ซึ่งทำให้มาตรฐานของผลการเรียนแต่ละโรงเรียนไม่เท่าเทียมกัน และส่วนของผู้เรียนซึ่งทำให้มาตรฐาน ผลการเรียนของนักเรียนแต่ละคนภายใน โรงเรียนเดียวกันไม่เท่าเทียมกัน ดังนั้นการปรับเทียบให้ ผลการเรียนของนักเรียนทั้งหมดมีมาตรฐานเดียวกัน จึงต้องปรับเทียบทั้งในส่วนของโรงเรียนและในส่วนของผู้เรียน ถ้าให้ GPA_{ij} เป็นผลการเรียนของนักเรียนคนที่ i ของโรงเรียนที่ j เราอาจแยกผลการเรียนของนักเรียนแต่ละคนออกเป็น 2 ส่วนเพื่อใช้ในการปรับเทียบ คือ 1. \overline{GPA}_j ซึ่งเป็นค่าเฉลี่ยผลการเรียนของโรงเรียนที่ j ใช้ในการปรับเทียบผลการเรียนในส่วนของโรงเรียน และ 2. ΔGPA_{ij} ซึ่งเป็นผลการเรียนของนักเรียนแต่ละคนที่ต่างไปจากค่าเฉลี่ยของโรงเรียน ใช้ในการปรับเทียบผลการเรียนในส่วนของผู้เรียน ดังนั้นผลการเรียนของนักเรียนแต่ละคน (GPA_{ij}) จึงเขียนให้อยู่ในรูป \overline{GPA}_j และ ΔGPA_{ij} ได้เช่นเดียวกับในสมการ (2.2) คือ

$$GPA_{ij} = \overline{GPA}_j + \Delta GPA_{ij} \quad (2.7)$$

เมื่อ GPA_{ij} = ผลการเรียนของนักเรียนคนที่ i ของโรงเรียนที่ j
 \overline{GPA}_j = ผลการเรียนของโรงเรียน j
 = ค่าเฉลี่ยของผลการเรียนของนักเรียนทุกคนใน โรงเรียนที่ j

$$\begin{aligned}\Delta GPA_{ij} &= \text{ผลการเรียนของนักเรียนแต่ละคนที่ต่างจากค่าเฉลี่ย} \\ &= GPA_{ij} - \overline{GPA}_j\end{aligned}$$

ในส่วนของผลการเรียนของโรงเรียน (ผลการเรียนเฉลี่ยของโรงเรียน) ซึ่งมีความเอนเอียงไปทำให้มาตรฐานผลการเรียนของแต่ละโรงเรียนไม่เท่าเทียมกัน การปรับให้ผลการเรียนของโรงเรียน (\overline{GPA}_j) มีมาตรฐานเดียวกันนั้น เนื่องจากผลการเรียนเฉลี่ยสะสมกับคะแนนการทดสอบทางการศึกษาแห่งชาติ (O-NET) มีความสัมพันธ์เชิงเส้นตรง จึงสามารถหามาตรฐานร่วมกันโดยการนำผลการเรียนเฉลี่ยของโรงเรียน (\overline{GPA}_j) และคะแนน O-NET เฉลี่ยของโรงเรียน (\overline{ONET}_j) มาวิเคราะห์การถดถอยโดยใช้โมเดลการวิเคราะห์สมการถดถอยอย่างง่าย ซึ่งจะได้สมการถดถอยดังนี้

$$\overline{GPA}_j = b_{00} + b_{01} \overline{ONET}_j + \varepsilon_j \quad (2.8)$$

เมื่อ \overline{GPA}_j = ผลการเรียนเฉลี่ยของโรงเรียน j
 = ค่าเฉลี่ยของผลการเรียนเฉลี่ยสะสมของนักเรียนทุกคนในโรงเรียน j
 \overline{ONET}_j = คะแนน O-NET เฉลี่ยของโรงเรียน j
 = ค่าเฉลี่ยของคะแนน O-NET ของนักเรียนทุกคนในโรงเรียน j
 b_{00} = ค่าคงที่หรือ \overline{GPA}_j -Intercept ของเส้นตรงความสัมพันธ์ระหว่าง \overline{GPA}_j กับ \overline{ONET}_j
 b_{01} = ค่าความชัน (Slope) ของเส้นตรงความสัมพันธ์ระหว่าง \overline{GPA}_j กับ \overline{ONET}_j
 ε_j = ความคลาดเคลื่อนของผลการเรียน โรงเรียน j

เพื่อให้ผลการเรียนของทุกโรงเรียนมีมาตรฐานเดียวกัน จึงตัดความคลาดเคลื่อนของผลการเรียนของแต่ละโรงเรียน ซึ่งเป็นส่วนที่ทำให้ผลการเรียนเอนเอียงออกไป ดังนั้นจะได้ผลการเรียนที่ควรจะเป็นของแต่ละโรงเรียน (\widehat{GPA}_j) ดังนี้

$$\widehat{GPA}_j = b_{00} + b_{01} \overline{ONET}_j \quad (2.9)$$

ในส่วนของผลการเรียนของนักเรียนแต่ละคนที่ต่างไปจากค่าเฉลี่ยของนักเรียนทุกคนในโรงเรียน (ΔGPA_{ij}) ความเอนเอียงของผลการเรียนในส่วนนี้ เกิดขึ้นในกลุ่มของนักเรียนภายในโรงเรียนเดียวกัน ซึ่งได้รับการประเมินผลการเรียนร่วมกัน แต่เนื่องจากระบบการตัดสินผลการเรียนแบบกำหนดระดับผลสัมฤทธิ์เป็นตัวอักษร ทำให้ผลการเรียนของนักเรียนบางคนเอนเอียงไปจากผลการเรียนที่ควรจะเป็น ดังนั้นการปรับเทียบผลการเรียนในส่วนนี้จึงเป็นการปรับเทียบภายในโรงเรียน

เดียวกัน และเนื่องจากผลการเรียนเฉลี่ยสะสมและคะแนน O-NET มีความสัมพันธ์เชิงเส้นตรง จึงใช้การวิเคราะห์การถดถอยระหว่างผลการเรียนที่ต่างไปจากค่าเฉลี่ยของนักเรียนแต่ละคน (ΔGPA_{ij}) กับคะแนน O-NET ของนักเรียนแต่ละคนที่ต่างไปจากคะแนน O-NET เฉลี่ยของโรงเรียน ($\Delta ONET_{ij}$) ซึ่งจะได้สมการถดถอยดังนี้

$$\Delta GPA_{ij} = b_{0j} + b_{1j} \Delta ONET_{ij} + \varepsilon_{ij} \quad (2.10)$$

เมื่อ ΔGPA_{ij} = ผลการเรียนที่ต่างไปจากค่าเฉลี่ยของโรงเรียนของนักเรียนแต่ละคน

$$\begin{aligned} \Delta ONET_{ij} &= \text{คะแนน O-NET ที่ต่างไปจากค่าเฉลี่ยของโรงเรียนของนักเรียนแต่ละคน} \\ &= ONET_{ij} - \overline{ONET}_j \end{aligned}$$

$ONET_{ij}$ = คะแนน O-NET ของนักเรียนคนที่ i ในโรงเรียนที่ j

\overline{ONET}_j = คะแนน O-NET ของโรงเรียนที่ j

= ค่าเฉลี่ยของคะแนน O-NET ของนักเรียนทุกคนในโรงเรียน j

b_{0j} = ค่าคงที่หรือ ΔGPA_{ij} -Intercept ของเส้นตรงความสัมพันธ์ระหว่าง

ΔGPA_{ij} กับ $\Delta ONET_{ij}$ ของโรงเรียนที่ j

b_{1j} = ค่าความชัน (Slope) ของเส้นตรงความสัมพันธ์ระหว่าง ΔGPA_{ij} กับ

$\Delta ONET_{ij}$ ของโรงเรียนที่ j

ε_{ij} = ความคลาดเคลื่อนของผลการเรียนของนักเรียนคนที่ i ของโรงเรียนที่ j

เพื่อให้ผลการเรียนของนักเรียนแต่ละคนภายในโรงเรียนนั้นมีมาตรฐานเดียวกัน จึงตัดความคลาดเคลื่อนของผลการเรียนของนักเรียนแต่ละคน (ε_{ij}) ซึ่งเป็นส่วนของผลการเรียนที่เอนเอียงออกไป จะทำให้ได้ผลการเรียนที่ต่างไปจากค่าเฉลี่ยของนักเรียนแต่ละคนที่ควรจะเป็นคือ

$$\widehat{\Delta GPA}_{ij} = b_{0j} + b_{1j} \Delta ONET_{ij}$$

หรือ

$$\widehat{\Delta GPA}_{ij} = b_{0j} + b_{1j} (ONET_{ij} - \overline{ONET}_j) \quad (2.11)$$

จากสมการ (2.7), (2.9) และ (2.11) ผลการเรียนที่ควรจะเป็นของนักเรียนแต่ละคนที่ปรับให้มีมาตรฐานเดียวกันทั้งภายในโรงเรียนและระหว่างโรงเรียนแล้ว จะเขียนได้เป็น

$$\begin{aligned} \widehat{GPA}_{ij} &= \widehat{GPA}_j + \widehat{\Delta GPA}_{ij} \\ &= (b_{00} + b_{01} \overline{ONET}_j) + [b_{0j} + b_{1j} (ONET_{ij} - \overline{ONET}_j)] \end{aligned}$$

หรือ

$$\widehat{GPA}_{ij} = (b_{00} + b_{01} \overline{ONET}_j) + [b_{0j} + b_{1j} (ONET_{ij} - \overline{ONET}_j)] \quad (2.12)$$

เมื่อ \widehat{GPA}_{ij} = ผลการเรียนที่ปรับแล้วของนักเรียนคนที่ i ของโรงเรียน j

$ONET_{ij}$ = คะแนน O-NET ของนักเรียนคนที่ i ของโรงเรียน j

\overline{ONET}_j = คะแนน O-NET ของโรงเรียน j

= ค่าเฉลี่ยของคะแนน O-NET ของนักเรียนทุกคนในโรงเรียน j

b_{00} = ค่าคงที่หรือ \overline{GPA}_j -Intercept ของเส้นตรงความสัมพันธ์ระหว่าง \overline{GPA}_j กับ \overline{ONET}_j

b_{01} = ค่าความชัน (Slope) ของเส้นตรงความสัมพันธ์ระหว่าง \overline{GPA}_j กับ \overline{ONET}_j โดย \overline{GPA}_j เป็นผลการเรียนเฉลี่ยของโรงเรียน j ซึ่งเท่ากับค่าเฉลี่ยของผลการเรียนของนักเรียนทุกคนในโรงเรียน j

b_{0j} = ค่าคงที่ หรือ ΔGPA_{ij} -Intercept ของเส้นตรงความสัมพันธ์ระหว่าง ΔGPA_{ij} กับ $\Delta ONET_{ij}$ ของโรงเรียน j

b_{1j} = ค่าความชัน (Slope) ของเส้นตรงความสัมพันธ์ระหว่าง ΔGPA_{ij} กับ $\Delta ONET_{ij}$ ของโรงเรียน j โดย $\Delta GPA_{ij} = GPA_{ij} - \overline{GPA}_j$ เป็นผลการเรียนของนักเรียนแต่ละคนภายในโรงเรียนที่ต่างไปจากค่าเฉลี่ยของโรงเรียน j นั้น และ $\Delta ONET_{ij} = ONET_{ij} - \overline{ONET}_j$ เป็นคะแนน O-NET ของนักเรียนแต่ละคนภายในโรงเรียนที่ต่างไปจากคะแนน O-NET เฉลี่ยของโรงเรียน j นั้น

ในกรณีของผลการเรียนเฉลี่ยสะสมตลอดหลักสูตร ($GPAX$) ใช้คะแนนรวม $ONET$ ($TONET$) แทนคะแนน $ONET$ และดำเนินการเช่นเดียวกับกรณีของผลการเรียนเฉลี่ยสะสม ตามกลุ่มสาระการเรียนรู้ (GPA)

3.2.2 การปรับสเกลตามระบบการตัดสินผลการเรียน

การปรับเทียบผลการเรียนให้มีมาตรฐานเดียวกันตามขั้นตอนที่ 1 มีโอกาสที่จะทำให้ค่าสูงสุดและค่าต่ำสุดของผลการเรียนที่ปรับแล้ว (\widehat{GPA}_{\max} และ \widehat{GPA}_{\min}) มีค่าไม่สอดคล้องกับค่าสูงสุดและค่าต่ำสุดตามระบบการตัดสินผลการเรียน คือ ค่าสูงสุดอาจมีค่าเกิน 4.00 ซึ่งจะเกิดขึ้นในกรณีที่นักเรียนคนนั้นอยู่ในโรงเรียนที่เกรดแรกเมื่อเทียบกับโรงเรียนทั้งหมด และได้รับอิทธิพลเพดาน (ceiling effect) ในการประเมินผลภายในโรงเรียน นักเรียนคนนี้จึงมีผลการเรียนเฉลี่ยสะสม 4.00 เท่ากับนักเรียนคนอื่นที่มีความรู้ความสามารถรองลงมา ในขณะที่ผลการทดสอบทางการศึกษาแห่งชาติขั้นพื้นฐาน (O-NET) ของเขาสูงกว่านักเรียนคนอื่นๆ เพราะตามสถิติของสถาบันทดสอบทางการศึกษาแห่งชาติ คะแนน O-NET ไม่มีอิทธิพลเพดาน หลังการปรับเทียบตามขั้นตอนที่ 1 ผลการเรียนของเขาจะถูกปรับเพิ่มขึ้นตามคะแนน O-NET ทำให้ผลการเรียนเฉลี่ยสะสมของเขา

มีค่าสูงกว่า 4.00 ซึ่งไม่สอดคล้องกับระบบการตัดสินผลการเรียน ดังนั้นจึงต้องทำการปรับสเกล เช่นเดียวกับวิธีการปรับเทียบผลการเรียนแบบเชื่อมโยงส่วนเดียว (OLC) โดยใช้สมการ (2.6) เช่นเดียวกัน

โดยสรุปแล้ววิธีการปรับเทียบผลการเรียนแบบเชื่อมโยงเสริมกันสองส่วน (TALC) พัฒนาขึ้นตามแนวคิดที่ว่า ความเอนเอียงของผลการเรียนเกิดจาก 1. มาตรฐานของผลการเรียน ระหว่างโรงเรียน และ 2. มาตรฐานของผลการเรียนภายในโรงเรียนไม่เท่าเทียมกัน โดยความเอนเอียงของผลการเรียนทั้งสองส่วนนี้มีลักษณะเสริมกัน การพัฒนาวิธีการปรับเทียบผลการเรียนวิธีนี้ ผู้วิจัยมองผลการเรียนเป็น “ผล” ของปัจจัยในสองส่วนแบบเสริมกัน จึงทำการปรับเทียบผลการเรียน ทั้งในส่วนของโรงเรียนและในส่วนของผู้เรียน โดยยึดโมเดลเชิงบวก (Additive Model) แต่ไม่ใช่เป็นการวิเคราะห์พหุระดับเพื่อหาปัจจัยที่มีอิทธิพลเสริมกัน โดยวิธีการปรับเทียบผลการเรียนแบบเชื่อมโยงเสริมกันทั้งสองส่วนมีขั้นตอนการปรับเทียบ 2 ขั้นตอน คือ ขั้นตอนการปรับเทียบผลการเรียน ให้มีมาตรฐานเดียวกันตามสมการ (2.12) และขั้นตอนการปรับสเกลตามสมการ (2.6)

3.3 วิธีการปรับเทียบผลการเรียนแบบเชื่อมโยงร่วมกันสองส่วน (Two-Section Interactive Linking Grade Calibration: TILC)

วิธีการปรับเทียบผลการเรียนแบบเชื่อมโยงร่วมกันสองส่วน เป็นวิธีที่พัฒนาขึ้นจากแนวคิดที่ว่า ผลการเรียนที่เอนเอียงไปของนักเรียนเกิดจาก 1. มาตรฐานของผลการเรียนที่แตกต่างกัน ระหว่างโรงเรียน ซึ่งเป็นความเอนเอียงในส่วนของโรงเรียน และ 2. มาตรฐานของผลการเรียนภายในโรงเรียนเดียวกันที่ไม่เท่าเทียมกัน ซึ่งเป็นความเอนเอียงในส่วนของผู้เรียน โดยความเอนเอียงในส่วนของผู้เรียนนี้เป็นผลมาจากระบบการตัดสินผลการเรียนร่วมกับความเอนเอียงในส่วนของโรงเรียน หรือมีลักษณะเป็น โมเดลเชิงปฏิสัมพันธ์ (Interactive model) เช่น นักเรียนเก่งในโรงเรียนที่มีการปล่อยเกรดผลการเรียนของเขาจะได้รับอิทธิพลเพดานของผลการเรียน ซึ่งมีค่าไม่เกิน 4.00 นั่นคือ เมื่อพิจารณาในส่วนของโรงเรียน ผลการเรียนของเขาจะเหมือนกับได้รับการปล่อยเกรด (ค่าเฉลี่ยของโรงเรียนสูง ทั้งที่ความสามารถของนักเรียนส่วนใหญ่ต่ำ) แต่หากเทียบผลการเรียนของเขา กับโรงเรียนอื่นๆ ที่ไม่ปล่อยเกรด ความสามารถของเขาจะเหมาะสมกับผลการเรียนที่ได้ ในขณะที่เดียวกันหากพิจารณาเทียบกับนักเรียนภายในโรงเรียนเดียวกัน ผลการเรียนของเขาจะได้รับความเอนเอียงจากระบบการตัดสินผลการเรียน เพราะผลการเรียนที่เขาได้รับจะเท่ากับนักเรียนที่มีความสามารถต่ำกว่า ดังนั้น ผลการเรียนของนักเรียนคนนี้จะจึงมีลักษณะปล่อยเกรด เมื่อพิจารณาในส่วนของโรงเรียน แต่จะไม่เอนเอียงเมื่อพิจารณาในส่วนของผู้เรียนเทียบกับนักเรียนทั้งหมดจากทุกโรงเรียน และมีลักษณะกดเกรดเมื่อพิจารณาในส่วนของผู้เรียนเทียบกับนักเรียนภายในโรงเรียน ความเอนเอียงของผลการเรียนของนักเรียนจึงเป็นจากความเอนเอียงในส่วนของโรงเรียนร่วมกับความเอนเอียงในส่วนของผู้เรียน ใน

กรณีเช่นนี้ หากเปรียบเทียบผลการเรียนด้วยวิธี OLC จะไม่เหมาะสมเพราะมีความเอนเอียงในส่วนของผู้เรียนด้วย และหากเปรียบเทียบผลการเรียนด้วยวิธี TALC ก็ไม่เหมาะสมเช่นกัน เพราะความเอนเอียงในส่วนของผู้เรียนนั้นเป็นผลร่วมกันระหว่างระบบการตัดสินผลการเรียนซึ่งเกิดขึ้นภายในโรงเรียน แต่อิทธิพลพหุคูณของผลการเรียนซึ่งเป็นค่าสูงสุดของผลการเรียนที่ใช้กับโรงเรียนทั้งหมด เช่น สมมติว่านักเรียนที่เก่งอยู่ในโรงเรียนที่ปล่อยเกรด มีผลการเรียน 4.00 ซึ่งถ้าเทียบความสามารถกับเด็กในโรงเรียนเดียวกัน นักเรียนคนนี้อาจได้ระดับผลการเรียนที่มากกว่า 4.00 แต่เนื่องจากเกณฑ์การตัดสินผลการเรียนกำหนดเกณฑ์การตัดเกรดไว้สูงสุดเท่ากับ 4.00 ถ้าเทียบความรู้ความสามารถของเด็กคนนี้กับนักเรียนทั่วไป (นักเรียนจากโรงเรียนทั้งหมด) ก็ได้ผลการเรียน 4.00 ซึ่งสอดคล้องกับความสามารถของเขา ถึงแม้ว่าเดิมทีจะเป็นผลการเรียนที่ติดตัวมาจากโรงเรียนที่ปล่อยเกรดก็ตาม จากตัวอย่างแสดงให้เห็นว่า ความเอนเอียงของผลการเรียนในส่วนของโรงเรียนส่งผลร่วมมายังความเอนเอียงในส่วนของนักเรียนด้วย ดังนั้นการเปรียบเทียบผลการเรียนแบบเชื่อมโยงร่วมกันสองส่วนจึงใช้การเปรียบเทียบผลการเรียนในส่วนของผู้เรียนร่วมกันระหว่างโรงเรียน ซึ่งมีขั้นตอนในการเปรียบเทียบ 2 ขั้นตอนดังนี้

3.3.1 การเปรียบเทียบผลการเรียนให้มีมาตรฐานเดียวกัน

เมื่อความเอนเอียงของผลการเรียนเกิดขึ้นในสองส่วนคือ 1. ส่วนของโรงเรียนซึ่งทำให้มาตรฐานของผลการเรียนแต่ละโรงเรียนไม่เท่าเทียมกัน และ 2. ส่วนของผู้เรียนซึ่งทำให้มาตรฐานผลการเรียนของนักเรียนแต่ละคนภายในโรงเรียนเดียวกันไม่เท่าเทียมกัน โดยได้รับผลร่วมระหว่างโรงเรียนด้วย ดังนั้นการเปรียบเทียบให้ผลการเรียนของนักเรียนทั้งหมดให้มีมาตรฐานเดียวกันจึงต้องเปรียบเทียบทั้งในส่วนของโรงเรียนและในส่วนของผู้เรียน ถ้าให้ GPA_{ij} เป็นผลการเรียนของนักเรียนคนที่ i ของโรงเรียนที่ j เราอาจแยกผลการเรียนของนักเรียนแต่ละคนออกเป็น 2 ส่วนเพื่อใช้ในการเปรียบเทียบ คือ 1. \overline{GPA}_j ซึ่งเป็นค่าเฉลี่ยผลการเรียนของโรงเรียนที่ j ใช้ในการเปรียบเทียบผลการเรียนในส่วนของโรงเรียน และ 2. ΔGPA_{ij} ซึ่งเป็นผลการเรียนของนักเรียนแต่ละคนที่ต่างไปจากค่าเฉลี่ยของโรงเรียน ใช้ในการเปรียบเทียบผลการเรียนในส่วนของผู้เรียน ดังนั้นผลการเรียนของนักเรียน (GPA_{ij}) จึงเขียนให้อยู่ในรูป \overline{GPA}_j และ ΔGPA_{ij} ได้เช่นเดียวกับในสมการ (2.3) คือ

$$GPA_{ij} = \overline{GPA}_j + \Delta GPA_{ij} \quad (2.13)$$

เมื่อ GPA_{ij} = ผลการเรียนของนักเรียนคนที่ i ของโรงเรียนที่ j
 \overline{GPA}_j = ผลการเรียนของโรงเรียน j
 = ค่าเฉลี่ยของผลการเรียนของนักเรียนทุกคนในโรงเรียนที่ j
 ΔGPA_{ij} = ผลการเรียนของนักเรียนแต่ละคนที่ต่างจากค่าเฉลี่ย

ในส่วนของผลการเรียนของโรงเรียน (ผลการเรียนเฉลี่ยของโรงเรียน) ซึ่งมีความเอนเอียงไปทำให้มาตรฐานผลการเรียนของแต่ละโรงเรียนไม่เท่าเทียมกัน การปรับให้ผลการเรียนของโรงเรียน (\overline{GPA}_j) มีมาตรฐานเดียวกันนั้น เนื่องจากผลการเรียนเฉลี่ยสะสมควรมีความสัมพันธ์เชิงเส้นตรงกับคะแนนการทดสอบทางการศึกษาแห่งชาติ (O-NET) จึงสามารถหามาตรฐานร่วมกัน โดยการนำผลการเรียนเฉลี่ยของโรงเรียน (\overline{GPA}_j) และคะแนน O-NET เฉลี่ยของโรงเรียน (\overline{ONET}_j) มาวิเคราะห์การถดถอยโดยใช้โมเดลการวิเคราะห์สมการถดถอยอย่างง่าย ซึ่งจะได้สมการถดถอยเช่นเดียวกับวิธี TALC คือ

$$\overline{GPA}_j = b_{00} + b_{01} \overline{ONET}_j + \varepsilon_j \quad (2.14)$$

เมื่อ \overline{GPA}_j = ผลการเรียนเฉลี่ยของโรงเรียน j
 = ค่าเฉลี่ยของผลการเรียนเฉลี่ยสะสมของนักเรียนทุกคนในโรงเรียน j
 \overline{ONET}_j = คะแนน O-NET เฉลี่ยของโรงเรียน j
 = ค่าเฉลี่ยของคะแนน O-NET ของนักเรียนทุกคนในโรงเรียน j
 b_{00} = ค่าคงที่หรือ \overline{GPA}_j -Intercept ของเส้นตรงความสัมพันธ์ระหว่าง \overline{GPA}_j กับ \overline{ONET}_j
 b_{01} = ค่าความชัน (Slope) ของเส้นตรงความสัมพันธ์ระหว่าง \overline{GPA}_j กับ \overline{ONET}_j
 ε_j = ความคลาดเคลื่อนของผลการเรียนโรงเรียน j

เพื่อให้ผลการเรียนของทุกโรงเรียนมีมาตรฐานเดียวกัน จึงตัดความคลาดเคลื่อนของผลการเรียนของแต่ละโรงเรียน ซึ่งเป็นส่วนที่ทำให้ผลการเรียนเอนเอียงออกไป ดังนั้นจะได้ผลการเรียนที่ควรจะเป็นของแต่ละโรงเรียน (\widehat{GPA}_j) ดังนี้

$$\widehat{GPA}_j = b_{00} + b_{01} \overline{ONET}_j \quad (2.15)$$

ในส่วนของผลการเรียนของนักเรียนแต่ละคนที่ต่างไปจากค่าเฉลี่ยของนักเรียนทุกคนในโรงเรียน (ΔGPA_{ij}) ความเอนเอียงของผลการเรียนในส่วนนี้ เกิดขึ้นในกลุ่มของนักเรียนภายในโรงเรียนเดียวกัน ซึ่งได้รับการประเมินผลการเรียนรู้อันร่วมกัน แต่เนื่องจากระบบการตัดสินผลการเรียนแบบกำหนดระดับผลสัมฤทธิ์เป็นตัวอักษร ทำให้ผลการเรียนของนักเรียนบางคนเอนเอียงไปจากผลการเรียนที่ควรจะเป็น ในขณะที่เดียวกันความเอนเอียงของผลการเรียนส่วนนี้ก็ได้รับผลร่วมจากความเอนเอียงในส่วนของโรงเรียนด้วย กล่าวคือ ความเอนเอียงในส่วนของผู้เรียนเกิดขึ้นกับ

ผู้เรียนทั้งหมดในทุกโรงเรียนที่ใช้ระบบการตัดสินผลการเรียนร่วมกัน เมื่อต้องการปรับให้มีมาตรฐานเดียวกัน จึงต้องปรับเทียบกับนักเรียนทั้งหมดทั่วประเทศ เนื่องจากพิจารณาว่า ความเอนเอียงในส่วนของโรงเรียนส่งผลร่วมมายังส่วนของนักเรียน ที่เป็นนักเรียนทั้งหมดของทุกโรงเรียน จึงถือว่าความเอนเอียงในส่วนของผู้เรียนได้รับผลร่วมมาจากความเอนเอียงในส่วนของ โรงเรียนด้วย ดังนั้น ΔGPA_{ij} จึงแทนผลการเรียนของนักเรียนแต่ละคนที่ต่างไปจากค่าเฉลี่ยของนักเรียนทั้งหมด หรือ $\Delta GPA_{ij} = GPA_{ij} - \overline{GPA..}$ เมื่อ $\overline{GPA..}$ เป็นผลการเรียนเฉลี่ยรวม (grand mean) ของนักเรียนทั้งหมด การปรับเทียบผลการเรียนในส่วนนี้จึงเป็นการปรับเทียบระหว่างนักเรียนทั้งหมดของทุกโรงเรียน และเนื่องจากผลการเรียนเฉลี่ยสะสมและคะแนน O-NET มีความสัมพันธ์เชิงเส้นตรง จึงใช้การวิเคราะห์การถดถอยระหว่างผลการเรียนที่ต่างไปจากค่าเฉลี่ยของนักเรียนแต่ละคน (ΔGPA_{ij}) กับคะแนน O-NET ของนักเรียนแต่ละคนที่ต่างไปจากคะแนน O-NET เฉลี่ยรวม (grand mean) ของนักเรียนทั้งหมด ($ONET_{ij} - \overline{ONET..}$) ซึ่งจะได้สมการถดถอยดังนี้

$$\Delta GPA_{ij} = b_{10} + b_{11}(ONET_{ij} - \overline{ONET..}) + \varepsilon_{ij} \quad (2.16)$$

เมื่อ ΔGPA_{ij} = ผลการเรียนที่ต่างไปจากผลการเรียนเฉลี่ยรวม (grand mean) ของนักเรียนทั้งหมด ($GPA_{ij} - \overline{GPA..}$)

$ONET_{ij}$ = คะแนน O-NET ของนักเรียนคนที่ i ในโรงเรียนที่ j

$\overline{ONET..}$ = ค่าเฉลี่ยของคะแนน O-NET ของนักเรียนทั้งหมด (grand mean)

b_{10} = ค่าคงที่หรือ ΔGPA_{ij} -Intercept ของเส้นตรงความสัมพันธ์ระหว่าง ΔGPA_{ij} กับ $ONET_{ij} - \overline{ONET..}$

b_{11} = ค่าความชัน (Slope) ของเส้นตรงความสัมพันธ์ระหว่าง ΔGPA_{ij} กับ $ONET_{ij} - \overline{ONET..}$

ε_{ij} = ความคลาดเคลื่อนของผลการเรียนของนักเรียนคนที่ i ของโรงเรียนที่ j

เพื่อให้ผลการเรียนของนักเรียนแต่ละคนมีมาตรฐานเดียวกัน จึงตัดความคลาดเคลื่อนของผลการเรียนของนักเรียนแต่ละคน (ε_{ij}) ซึ่งเป็นส่วนของผลการเรียนที่เอนเอียงออกไป จะทำให้ได้ผลการเรียนที่ต่างไปจากค่าเฉลี่ยของนักเรียนแต่ละคนที่ควรจะเป็นคือ

$$\widehat{\Delta GPA}_{ij} = b_{10} + b_{11}(ONET_{ij} - \overline{ONET..}) \quad (2.17)$$

จากสมการ (2.13), (2.15) และ (2.17) ผลการเรียนที่ควรจะเป็นของนักเรียนแต่ละคนที่ปรับให้มีมาตรฐานเดียวกันทั้งภายในโรงเรียนและระหว่างโรงเรียนแล้ว จะเขียนได้เป็น

$$\begin{aligned}\widehat{GPA}_{ij} &= \widehat{GPA}_j + \Delta \widehat{GPA}_{ij} \\ &= (b_{00} + b_{01} \overline{ONET}_j) + [b_{10} + b_{11} (ONET_{ij} - \overline{ONET}_{..})]\end{aligned}$$

หรือ

$$\widehat{GPA}_{ij} = (b_{00} + b_{01} \overline{ONET}_j) + [b_{10} + b_{11} (ONET_{ij} - \overline{ONET}_{..})] \quad (2.18)$$

เมื่อ \widehat{GPA}_{ij} = ผลการเรียนที่ปรับแล้วของนักเรียนคนที่ i ของโรงเรียน j

$ONET_{ij}$ = คะแนน O-NET ของนักเรียนคนที่ i ของโรงเรียน j

\overline{ONET}_j = คะแนน O-NET ของโรงเรียน j

= ค่าเฉลี่ยของคะแนน O-NET ของนักเรียนทุกคนในโรงเรียน j

$\overline{ONET}_{..}$ = ค่าเฉลี่ยของคะแนน O-NET ของนักเรียนทั้งหมด (grand mean)

b_{00} = ค่าคงที่หรือ \widehat{GPA}_j -Intercept ของเส้นตรงความสัมพันธ์ระหว่าง \widehat{GPA}_j กับ \overline{ONET}_j

b_{01} = ค่าความชัน (Slope) ของเส้นตรงความสัมพันธ์ระหว่าง \widehat{GPA}_j กับ \overline{ONET}_j

โดย \widehat{GPA}_j เป็นผลการเรียนเฉลี่ยของโรงเรียน j ซึ่งเท่ากับค่าเฉลี่ยของผลการเรียนของนักเรียนทุกคนในโรงเรียน j

b_{10} = ค่าคงที่หรือ $\Delta \widehat{GPA}_{ij}$ -Intercept ของเส้นตรงความสัมพันธ์ระหว่าง $\Delta \widehat{GPA}_{ij}$ กับ $ONET_{ij} - \overline{ONET}_{..}$

b_{11} = ค่าความชัน (Slope) ของเส้นตรงความสัมพันธ์ระหว่าง $\Delta \widehat{GPA}_{ij}$ กับ $ONET_{ij} - \overline{ONET}_{..}$

ในกรณีของผลการเรียนเฉลี่ยสะสมตลอดหลักสูตร (GPAX) ใช้คะแนนรวม ONET (TONET) แทนคะแนน ONET และดำเนินการเช่นเดียวกับกรณีของผลการเรียนเฉลี่ยสะสมตามกลุ่มสาระการเรียนรู้ (GPA)

3.3.2 การปรับสเกลตามระบบการตัดสินผลการเรียน

การปรับเทียบผลการเรียนให้มีความมาตรฐานเดียวกันตามเงื่อนไขที่ 1 มีโอกาสที่จะทำให้

ค่าสูงสุดและค่าต่ำสุดของผลการเรียนที่ปรับแล้ว (\widehat{GPA}_{\max} และ \widehat{GPA}_{\min}) มีค่าไม่สอดคล้องกับค่าสูงสุดและค่าต่ำสุดตามระบบการตัดสินผลการเรียน คือ ค่าสูงสุดอาจมีค่าเกิน 4.00 ซึ่งจะเกิดขึ้นในกรณีที่นักเรียนคนนั้นอยู่ในโรงเรียนที่กวดเกรดเมื่อเทียบกับโรงเรียนทั้งหมด และได้รับอิทธิพลเพดาน (ceiling effect) ในการประเมินผลภายในโรงเรียนและเมื่อเทียบกับนักเรียนทั้งหมด นักเรียนคนนี้จึงมีผลการเรียนเฉลี่ยสะสม 4.00 เท่ากับนักเรียนคนอื่นที่มีความรู้ความสามารถรองลงมาทั้งภายในโรงเรียนเดียวกัน และระหว่างโรงเรียน ในขณะที่ผลการทดสอบทางการศึกษาแห่งชาติขึ้น

พื้นฐาน (O-NET) ของเขาสูงกว่านักเรียนคนอื่นๆ เพราะตามสถิติของสถาบันทดสอบทางการศึกษาแห่งชาติ คะแนน O-NET ไม่มีอิทธิพลเพดาน หลังการปรับเทียบตามขั้นตอนที่ 1 ผลการเรียนของเขาจะถูกปรับเพิ่มขึ้นตามคะแนน O-NET ทำให้ผลการเรียนเฉลี่ยสะสมของเขามีค่าสูงกว่า 4.00 ซึ่งไม่สอดคล้องกับระบบการตัดสินผลการเรียน ดังนั้นจึงต้องทำการปรับสเกลเช่นเดียวกับวิธีการปรับเทียบผลเรียนแบบเชื่อมโยงส่วนตัว (OLC) และวิธีการปรับเทียบผลเรียนแบบเชื่อมโยงเสริมกันสองส่วน (TALC) โดยใช้สมการ (2.6) เช่นเดียวกัน

โดยสรุปแล้ววิธีการปรับเทียบผลเรียนแบบเชื่อมโยงร่วมกันสองส่วน (TILC) พัฒนาขึ้นตามแนวคิดที่ว่า ความเอนเอียงของผลเรียนของนักเรียนแต่ละคน เป็นผลมาจากความเอนเอียงของผลเรียนในส่วนของโรงเรียน และความเอนเอียงของผลเรียนในส่วนของผู้เรียน โดยทั้งสองส่วนส่งผลต่อความเอนเอียงของผลเรียนร่วมกัน การพัฒนาวิธีการปรับเทียบผลเรียนวิธีนี้ ผู้วิจัยมองผลเรียนเป็น “ผล” ของปัจจัยในทั้งสองส่วนที่มีปฏิสัมพันธ์กัน จึงทำการปรับเทียบผลเรียนทั้งในส่วนของโรงเรียนและในส่วนของผู้เรียน โดยยึดโมเดลเชิงปฏิสัมพันธ์ (Interactive Model) แต่ไม่ใช่เป็นการวิเคราะห์หุ้ระดับเพื่อหาปัจจัยที่มีอิทธิพลร่วมกัน โดยการปรับเทียบผลเรียนแบบเชื่อมโยงร่วมกันทั้งสองส่วน มีขั้นตอนในการปรับเทียบ 2 ขั้นตอน คือ ขั้นตอนการปรับเทียบผลเรียนให้มีมาตรฐานเดียวกันตามสมการ (2.18) และขั้นตอนการปรับสเกลตามสมการ (2.6)

4. งานวิจัยที่เกี่ยวข้อง

4.1 งานวิจัยในต่างประเทศ

การศึกษาวิจัยของต่างประเทศที่น่าสนใจในตอนนี้เป็นงานวิจัยที่เกี่ยวข้องกับการศึกษาเชิงเปรียบเทียบประสิทธิภาพ ของวิธีการปรับเทียบคะแนนระหว่างแบบทดสอบโดยแต่ละท่านได้ทำการศึกษาตัวแปรที่แตกต่างกัน เช่น วิธีการปรับเทียบคะแนน แบบแผนการเก็บข้อมูล เกณฑ์การตรวจสอบความแตกต่างของประสิทธิภาพแต่ละวิธี ซึ่งงานวิจัยที่เกี่ยวข้องมีดังนี้

ฮิวลิน ลิตตัก และ ดรอสโกว์ (Hulin, Lissak, and Drasgow, 1982 : 153-158) ได้ใช้การจำลองข้อมูลเพื่อศึกษา ขนาดของกลุ่มตัวอย่าง ความยาวของแบบทดสอบ และการประมาณค่าพารามิเตอร์ด้วยโปรแกรม LOGIST โดยมีการเปลี่ยนความยาวแบบทดสอบและเปลี่ยนขนาดของกลุ่มตัวอย่าง กับ IRT โมเดล 3 พารามิเตอร์ และพารามิเตอร์ความสามารถมีการแจกแจงแบบโค้งปกติ ผลการวิจัยพบว่าความยาวของแบบทดสอบและขนาดของกลุ่มตัวอย่างมีผลต่อการประมาณค่าเมื่อประมาณค่าพารามิเตอร์ 3 พารามิเตอร์โดยใช้ข้อสอบ 60 ข้อ กับกลุ่มผู้สอบจำนวน 1,000 คน พบว่าให้ผลออกมาในลักษณะที่คงที่ ถ้าลดข้อสอบลงครึ่งหนึ่งคือเหลือ 30 ข้อ และเพิ่มขนาดของ

กลุ่มตัวอย่างเป็นสองเท่าจะให้ผลใกล้เคียงกัน และพบว่า โปรแกรม LOGIST ใช้กับกลุ่มตัวอย่างขนาด 15 คนไม่ได้

คาลด์เวลล์ (Caldwell, 1984 : 2847) ได้ทำการศึกษาเพื่อวัดประสิทธิผลสัมพัทธ์ในโมเดล การเปรียบเทียบความยากของแบบทดสอบ รูปแบบที่ใช้คือ รูปแบบเชิงเส้นตรง แบบแผนที่ 4 และ ราสซ์โมเดล ในการเปรียบเทียบได้พิจารณาถึง การเพิ่มค่าเฉลี่ยและการลดส่วนเบี่ยงเบนมาตรฐาน สำหรับการสอบครั้งที่ 2 ซึ่งคาดว่า โมเดลเชิงเส้นตรงจะมีประสิทธิผลน้อยกว่าราสซ์โมเดล เนื่องจาก โมเดลเชิงเส้นตรงมีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานบรรจุอยู่ในสูตร ส่วนราสซ์โมเดลไม่มี และ คาดว่าประสิทธิผลสัมพัทธ์ของราสซ์โมเดลจะเพิ่มขึ้นขณะที่ความแตกต่างของค่าเฉลี่ยเพิ่มขึ้น ได้มีการประเมินแบบทดสอบรวมทั้งสองแบบ แบบทดสอบร่วมชุดหนึ่งประกอบด้วยข้อสอบที่มีความยากในระดับปานกลางส่วนอีกชุดหนึ่งมีความยากสูงสุด แบบทดสอบที่มีค่าความยากปานกลาง ได้รับการพิจารณาว่าให้ประสิทธิผลต่ำสุด และแบบทดสอบร่วมที่มีความยากปานกลางได้รับการพิจารณาว่าให้ประสิทธิผลที่ดีกว่า ผลการวิจัยพบว่าเมื่อพิจารณาคะแนนทั้งหมด ราสซ์โมเดล เหนือกว่ารูปแบบเชิงเส้นตรงทั้งในแบบทดสอบร่วมที่มีความยากระดับปานกลางและระดับสูง แต่เมื่อพิจารณาคะแนนจุดตัดที่แสดงถึงความสามารถต่ำสุด ราสซ์โมเดลเหนือกว่าในลักษณะที่ให้ ความลำเอียงเพียงเล็กน้อยแต่เชิงเส้นตรง ให้ค่าความคลาดเคลื่อนในระดับต่ำดัชนีความแตกต่างของ ราสซ์โมเดลมีค่าเฉลี่ยเกือบเป็นศูนย์และมีการแปรเปลี่ยนในช่วงกว้าง รูปแบบเชิงเส้นตรงให้ค่า ความลำเอียงด้านลบ แต่มีการแปรเปลี่ยนที่น้อยกว่า

สแก็กส์ (Skaggs, 1984 : 1259) ได้ทำการตรวจสอบวิธีการเปรียบเทียบคะแนน 4 วิธีภายใต้ เงื่อนไขที่พารามิเตอร์ข้อสอบและพารามิเตอร์ความสามารถแปรเปลี่ยนอย่างเป็นระบบ วิธีการ เปรียบเทียบที่ใช้ได้แก่ IRT 2 แบบ คือ ราสซ์โมเดล กับ โมเดล 3-พารามิเตอร์ และแบบมาตรฐานเดิม อีก 2 แบบ คือการเปรียบเทียบแบบเส้นตรงและแบบอิกวิเปอร์เซ็นไทล์ ข้อมูลที่ใช้ในการศึกษาใช้การ จำลองขึ้นมา ตัวแปรที่ใช้ในการศึกษาได้แก่ ค่าความยาก อำนาจจำแนก และระดับโอกาสการเดา และค่าเฉลี่ยของความสามารถ เกณฑ์ที่ใช้ในการเปรียบเทียบคือคะแนนจริงจาก โมเดล 3-พารามิเตอร์ ข้อแตกต่างระหว่างวิธีการเปรียบเทียบทั้ง 4 แบบกับเกณฑ์นำมาวิเคราะห์ หากสถิติเชิงบรรยายและ แสดงด้วยกราฟและจุดประสงค์หลักของการศึกษาคือเพื่อทดสอบความแกร่งของราสซ์โมเดลเมื่อ มีการละเมิดข้อตกลง ผลการวิจัยพบว่าราสซ์โมเดลยัง ไม่มีความแกร่งเมื่อมีการละเมิดในข้อตกลง กรณีนี้เป็นจริงเมื่อค่าเฉลี่ยของอำนาจจำแนกมีค่าไม่เท่ากัน และ โมเดล 3-พารามิเตอร์ ก็ประสบ ปัญหาเช่นกันเมื่ออำนาจจำแนกไม่เท่ากัน โดยภาพรวมการเปรียบเทียบคะแนนแบบอิกวิเปอร์เซ็นไทล์

ให้ผลดีเกือบทั้งหมด การวิจัยครั้งนี้ได้ให้ข้อเสนอแนะว่าการใช้ราสซ์โมเดลและ โมเดล 3-พารามิเตอร์ ต้องใช้ด้วยความระมัดระวัง

วูดครัฟ (Woodruff, 1986 : 245-257) ได้ศึกษาวิธีปรับเทียบคะแนนแบบเส้นตรงโดยใช้แบบแผนข้อสอบรวมที่มีประชากรแตกต่างกันและแบบจำลองคะแนนจริงสัมพันธ์ โดยใช้กลุ่มตัวอย่างจำนวน 4,000 คน ทดสอบกับแบบทดสอบ 5 ฉบับ ที่เป็นฟอร์มใหม่ มีความยาวที่แตกต่างกันคือ ตั้งแต่จำนวน 25 ถึง 115 ข้อ และใช้แบบทดสอบที่เป็นฟอร์มเก่าร่วมกับแบบทดสอบเชื่อมโยงมาทดสอบกับกลุ่มตัวอย่างจำนวน 4,000 คน นำคะแนนที่ได้มาศึกษาวิธีปรับเทียบคะแนน 2 วิธี ผลการศึกษาพบว่า มีความแตกต่างของกลุ่มผู้สอบเล็กน้อย โดยดูจากความเบี่ยงเบนมาตรฐานจากแบบทดสอบ 5 ฉบับ ที่ใช้แบบทดสอบหลักและใช้วิธีการปรับเทียบคะแนน 2 วิธี ให้ความเพียงพอในการปฏิบัติ แม้ว่าแบบทดสอบทั้ง 5 ฉบับ จะวัดในลักษณะที่ต่างกัน

โคป (Cope, 1987 : 143-149) ได้เปรียบเทียบการปรับเทียบคะแนนเชิงเส้นตรง 5 วิธีระหว่างวิธีของแองกอฟ (Angoff) ตามวิธีที่ 5 ซึ่งมี 3 วิธี กับวิธีของทักเกอร์ (Tucker) และ เลวิน (Levine) ในกรณีที่แบบทดสอบมีความเชื่อมั่นเท่ากัน โดยใช้แบบทดสอบรวม แบบทดสอบที่ใช้ปรับเทียบคะแนนเป็นแบบทดสอบเพื่อให้ประกาศนียบัตรในวิชาชีพ จำนวน 5 ชุด สอบกับกลุ่มประชากรที่ไม่เท่าเทียมกัน 5 กลุ่ม ๆ ละ 1 ชุด ปีละ 2 ครั้ง การเปรียบเทียบผลการเทียบระหว่างแบบทดสอบแต่ละคู่ (Single link) ภายใต้วิธีการเชิงเส้นตรงที่แตกต่างกัน 5 วิธี ทั้งนี้ไม่มีการเปรียบเทียบความถูกต้องของวิธีการต่างๆ เพราะว่าสเกลคะแนนดิบของแบบทดสอบสองชุดถูกคาดว่าจะต่างกัน อีกแบบหนึ่งคือการเปรียบเทียบผลการปรับเทียบคะแนนแบบโซ่ ซึ่งเป็นการเปรียบเทียบคะแนนแบบทดสอบแต่ละชุดกับตัวเอง โดยผ่านแบบทดสอบชุดอื่นที่เลือก ทั้งนี้เพื่อทำการเปรียบเทียบความถูกต้องของวิธีการเหล่านั้นด้วย

การวัดความคลาดเคลื่อนของการปรับเทียบคะแนน (Equating Error) ใช้สถิติ 2 ตัว คือ ค่าประมาณความลำเอียง (Estimate Bias) และค่าประมาณที่ถ่วงน้ำหนักแล้วของรากที่สองของค่าเฉลี่ยยกกำลังสองของความคลาดเคลื่อนในการปรับเทียบคะแนน (Estimated Weighted Root Squared Error : RMSE) ซึ่งใช้เปรียบเทียบระหว่างวิธีที่แตกต่างกัน คำนวณจากสมการ

$$\text{BIAS} = \bar{X}' - \bar{X}$$

เมื่อ \bar{X} คือ ค่าเฉลี่ยของคะแนนดิบของแบบทดสอบชุดหนึ่ง
 \bar{X}' คือ ค่าเฉลี่ยของคะแนนดิบที่เท่าเทียมกัน โดยการเทียบแบบทดสอบชุดนั้น
 ไปยังตัวเอง

$$\text{และ } RMSE = \left[\frac{\sum N_i (X'_i - X_i)^2}{\sum N_i} \right]^{\frac{1}{2}}$$

เมื่อ X_i คือ คะแนนดิบของคนที่ i
 X'_i คือ คะแนนดิบที่เท่าเทียมกัน
 N_i คือ จำนวนผู้สอบซึ่งได้คะแนน X_i

ในกรณีที่มีผู้สอบจำนวนมาก การคำนวณ RMSE จะง่ายขึ้น โดยใช้สมการ

$$RMSE = \left[(m-1)^2 \text{Var}(x) + (\text{BIAS})^2 \right]^{\frac{1}{2}}$$

เมื่อ m คือ ความชันของฟังก์ชันการเทียบ $x' = mx + b$
 $\text{Var}(x)$ คือ ความแปรปรวนของ x

ผลการวิจัยสรุปว่า ในสถานการณ์ที่ศึกษาภายใต้รูปแบบที่ 5 ตามวิธีของแองกอฟให้คะแนนดิบ (Raw Score Equivalents) ใกล้เคียงกับวิธีการของทักเกอร์และเลวิน โดยการปรับเทียบคะแนนแบบสายโซ่ทั้ง 5 วิธี ให้ค่าประมาณความลำเอียงและรากที่สองของค่าเฉลี่ยยกกำลังสองของความคลาดเคลื่อนที่ใกล้เคียงกัน ส่วนการปรับเทียบคะแนนจากวิธีการของทักเกอร์และเลวิน ผู้วิจัยกล่าวว่าผลการวิจัยนี้สนับสนุนวิธีการปรับเทียบคะแนนเชิงเส้นตรงตามรูปแบบที่ 5 ซึ่งมีข้อตกลงที่มีข้อจำกัดน้อยกว่าวิธีของทักเกอร์และเลวิน ดังนั้นในทางปฏิบัติควรใช้วิธีการตามรูปแบบที่ 5 วิธีการใดวิธีหนึ่งใน 3 วิธี

แมคแคน (MacCann, 1989 : 263-276) ได้เปรียบเทียบวิธีปรับเทียบคะแนนสองวิธีที่ใช้ข้อตกลงว่ามีความเชื่อมั่นเท่าเทียมกันและใช้แบบทดสอบที่เป็นคะแนนจริงสัมพันธ์ ทำการทดสอบกับกลุ่มตัวอย่างทั้งที่เป็นกลุ่มคล้ายคลึงกันและกลุ่มที่มีความแตกต่างกัน โดยมีขนาดกลุ่มตัวอย่างที่แตกต่างกันในแต่ละวิชาที่สอบ และแต่ละกลุ่มที่ใช้เป็นกลุ่มปรับเทียบคะแนน ตั้งแต่จำนวน 326 คน ถึง 5,610 คน ใช้แบบทดสอบ 5 วิชา แบ่งเป็นส่วนย่อยที่แตกต่างกันในแต่ละวิชา และใช้แบบทดสอบเชื่อมโยงที่มีความยาวคิดเป็นสัดส่วนแตกต่างกัน ทั้งที่มากกว่าและน้อยกว่าแบบทดสอบที่นำมาใช้ปรับเทียบคะแนน แล้วจึงนำมาเปรียบเทียบความคลาดเคลื่อนมาตรฐานของแต่ละวิธี (สี่วิธี) คือ วิธีของแมคแคน สองวิธี และวิธีของทักเกอร์กับวิธีของเลวิน พบว่า วิธีที่ 1 ให้ผลใกล้เคียงกับวิธีของทักเกอร์ ส่วนวิธีที่ 2 ให้ผลใกล้เคียงกับวิธีของเลวิน และทั้งสี่วิธี ให้ผลความคลาดเคลื่อนมาตรฐาน ในระดับที่น่าพอใจเป็นส่วนใหญ่ ผู้วิจัยสรุปว่า วิธีที่ 1 และวิธีที่ 2 มีข้อดี

การวิเคราะห์ร่วมให้ความคลาดเคลื่อนน้อยที่สุด และการปรับเทียบแบบเส้นตรง ให้ความคลาดเคลื่อนมากที่สุดจากการเก็บข้อมูลจริงพบว่า มีความสอดคล้องกับการจำลองข้อมูลร้อยละ 70

กลอวากิ (Glowacki, 1991 : 1722) ได้ตรวจสอบโมเดลของการปรับเทียบคะแนนที่มีความเหมาะสมกับการสอบของบัณฑิตวิทยาลัย แห่งมหาวิทยาลัยอลาบามา ปัญหาในการวิจัย คือ โมเดลของการปรับเทียบที่ตรวจสอบมีการแจกแจงของคะแนนดิบหรือคะแนนที่ผ่านจากการสอบแบบทดสอบการอ่าน และคณิตศาสตร์ แตกต่างกันหรือไม่ โมเดลที่ใช้ตรวจสอบคือ โมเดลเชิงเส้นตรง อิกวิเปอร์เซ็นไทล์และทฤษฎีการตอบสนองข้อสอบ ชนิด 1 2 และ 3 พารามิเตอร์ ผลการวิจัยพบว่า วิธีการปรับเทียบคะแนนทั้ง 5 โมเดล ในการสอบการอ่านและคณิตศาสตร์ให้ผลที่คล้ายคลึงกัน แสดงว่าโมเดลทั้งหมดสามารถนำมาใช้กับการปรับเทียบคะแนนได้ โดยไม่มีโมเดลใดที่ดีที่สุด

แฮริส (Harris, 1991 : 221-235) ได้ศึกษาผลของแบบแผนการรวบรวมข้อมูล 2 วิธี คือ แบบแผนที่ 1 ของแองกอฟ(ผู้สอบกลุ่มเดียวจัดให้สมดุล) และแบบแผนที่ 2 ของแองกอฟพี (ผู้สอบกลุ่มเท่าเทียมกัน) ที่มีต่อการเทียบคะแนนตามแนวตั้ง โดยใช้วิธีปรับเทียบคะแนน 2 วิธี คือ วิธีอิกวิเปอร์เซ็นไทล์ และวิธีตามทฤษฎีการตอบสนองข้อสอบแบบสามพารามิเตอร์ แบบทดสอบที่ใช้คือ แบบทดสอบวิชาคณิตศาสตร์ 2 ฉบับ คือ EACT มีจำนวน 60 ข้อ และฉบับ PACT มีจำนวน 40 ข้อ แบบทดสอบ ทั้ง 2 ฉบับ มีเนื้อหาคาบเกี่ยวกัน แต่มีเนื้อหาเฉพาะและระดับความยากต่างกัน กลุ่มตัวอย่างเป็นนักเรียนระดับ 11 และ 12 ซึ่งแบ่งเป็นกลุ่มย่อยหลายกลุ่ม การวิเคราะห์ข้อมูลใช้ ค่าเฉลี่ยกำลังสองของความแตกต่าง (RMSD)

ผลการวิจัยสรุปได้ว่า การปรับเทียบคะแนนตามแนวตั้ง โดยใช้แบบแผนการเก็บข้อมูลทั้ง 2 แบบ มีความเพียงพอ และให้ผลคล้ายคลึงกัน โดยรูปแบบที่ใช้ผู้สอบกลุ่มเดียว ให้ผลคงที่กว่ารูปแบบที่ใช้ผู้สอบกลุ่มเท่าเทียมเล็กน้อย

อเยอร์วี (Ayerve, 1992 : 1841) ได้ศึกษาเปรียบเทียบประสิทธิภาพของการปรับเทียบคะแนนตามแนวตั้งด้วยวิธีอิกวิเปอร์เซ็นไทล์ และวิธี IRT โมเดล 3 พารามิเตอร์ โดยใช้การจำลองข้อมูลภายใต้เงื่อนไขของขนาดกลุ่มตัวอย่าง ความยาวของแบบทดสอบและความยาวของแบบทดสอบร่วม โดยเปรียบเทียบใน 3 กรณี คือ 1. เปรียบเทียบประสิทธิภาพของวิธีการปรับเทียบคะแนนทั้งสองภายใต้เงื่อนไขทุกเงื่อนไข 2. เปรียบเทียบประสิทธิภาพของวิธีการปรับเทียบคะแนนแต่ละวิธีภายใต้เงื่อนไขของตัวแปรอิสระแต่ละตัว 3. ตรวจสอบผลของตัวแปรอิสระแต่ละตัวในการปรับเทียบคะแนนแต่ละวิธี ขนาดกลุ่ม ตัวอย่างที่ใช้ มี 3 ขนาด คือ 200 500 และ 1,000 คน ความยาวของแบบทดสอบที่ใช้มี 2 ขนาด คือ 30 ข้อ และ 60 ข้อ ส่วนความยาวของแบบทดสอบร่วมที่ใช้มี 2

ขนาด คือ 5 ข้อ และ 10 ข้อ การวิเคราะห์ข้อมูล ใช้ค่าเฉลี่ยกำลังสองของความคลาดเคลื่อนซึ่งถ่วงน้ำหนักแล้ว (Weighted Mean Square Error :WMSE) และค่าเฉลี่ยกำลังสองของความคลาดเคลื่อนซึ่งไม่ถ่วงน้ำหนัก (Unweighted Mean Square Error :UMSE)

ผลการวิจัยพบว่า โดยส่วนรวมแล้ว วิธีการปรับเทียบแบบ วิธีอิกวิเปอร์เซ็นไทล์และวิธี IRT ไม่แตกต่างกันอย่างมีนัยสำคัญ สำหรับการปรับเทียบคะแนนโดยวิธีอิกวิเปอร์เซ็นไทล์ พบว่าความยาวของแบบทดสอบกับความยาวของแบบทดสอบรวมเป็นองค์ประกอบที่สำคัญ ขณะที่การปรับเทียบคะแนนตามวิธี IRT 3 พารามิเตอร์ ขนาดของกลุ่มตัวอย่างเป็นองค์ประกอบที่สำคัญ คือ กลุ่มตัวอย่าง 200 คน มีแนวโน้มที่จะให้ผลการปรับเทียบคะแนนที่ไม่ถูกต้อง ในขณะที่กลุ่มตัวอย่างขนาดใหญ่ คือ 500 และ 1,000 จะให้ผลการปรับเทียบคะแนนที่ถูกต้องมากกว่า

ยัง (Young,1995 : 558-571) ได้ปรับคะแนนเฉลี่ยสะสมของนักเรียนศึกษาชั้นปีที่ 1 ของโรงเรียนกฎหมายด้วยวิธีปรับคะแนนเฉลี่ยสะสม 2 วิธี คือ วิธีปรับคะแนนเฉลี่ยสะสมโดยใช้ทฤษฎีการตอบสนองข้อสอบภายใต้ Graded Response Model กับวิธีปรับคะแนนเฉลี่ยสะสมด้วยวิธีเชิงเส้นตรง เพื่อเปรียบเทียบว่าคะแนนเฉลี่ยสะสมที่ปรับด้วยวิธีใด จะเป็นเกณฑ์ที่ดีที่สุดสำหรับตรวจสอบความตรง เชิงพยากรณ์ของคะแนน LSAT และคะแนนเฉลี่ยสะสมระดับมหาวิทยาลัย (UGPA) ข้อมูลที่ใช้ได้จากโรงเรียนกฎหมาย 4 แห่งในสหรัฐอเมริกา แยกวิเคราะห์เป็นรายโรงเรียนผลการวิเคราะห์พบว่า โดยทั่วไปเมื่อใช้คะแนนเฉลี่ยสะสมปี 1 ของโรงเรียนกฎหมายที่ปรับด้วยการวิเคราะห์แบบเชิงเส้นตรงเป็นเกณฑ์จะให้ความเที่ยงตรงเชิงพยากรณ์ดีกว่าวิธีปรับคะแนนเฉลี่ยสะสมโดยใช้ทฤษฎีการตอบสนองข้อสอบภายใต้ Graded Response Model

หยาง และคณะ (Yang, et al. 2002 : 2) ได้ทำการศึกษาความไม่แปรเปลี่ยนข้ามเขต (regional invariance) ของฟังก์ชันการปรับเทียบคะแนนระหว่างแบบสอบแบบเลือกตอบและแบบสอบทั้งฉบับ (มีทั้งแบบเลือกตอบและแบบเขียนตอบ) ของแบบทดสอบ AP (Advanced Placement Program) โดยเปรียบเทียบดัชนีการปรับเทียบได้ (equitability index) ของ Dorans-Holland ระหว่างวิธีการปรับเทียบแบบเชิงเส้นและไม่เป็นเชิงเส้น (อิกวิเปอร์เซ็นไทล์) และแบ่งเขตพื้นที่ในการศึกษาออกเป็น 3 เขตตามโซนเวลา คือ เขตโซนเวลาหลัก ซึ่งเป็นพื้นที่ส่วนใหญ่ของสหรัฐอเมริกา เขตก่อนโซนเวลาหลัก ได้แก่ เอเชียและยุโรป และเขตหลังโซนเวลาหลัก ได้แก่ ฮาวาย และอลาสกา ผลการศึกษาพอสรุปได้ว่า โดยภาพรวมการเชื่อมโยงแบบไม่เป็นเชิงเส้นให้ความไม่แปรเปลี่ยนข้ามเขตดีกว่าการเชื่อมโยงเชิงเส้น แต่ก็ให้ข้อสังเกตไว้ว่า ในเขตก่อนโซนเวลาหลักและเขตหลังโซนเวลาหลักนั้นมียุทธศาสตร์ตัวอย่างน้อย (เพียงร้อยละ 0.8 และ 0.6 ของกลุ่มตัวอย่างทั้งหมด) จึงอาจจะไม่เพียงพอที่จะสนับสนุนข้อค้นพบดังกล่าว ในขณะที่การเชื่อมโยงเชิงเส้นจะมีความผันแปรข้ามเขตน้อยกว่า

วอน คาไวร์ และคณะ (Von Davier, et al. 2002 : 13) ได้นำวิธีการใหม่ คือใช้การปรับเทียบคะแนนภายใน (self-equating) ร่วมกับดัชนีรากที่สองของผลต่างกำลังสองเฉลี่ย RMSD(x) มาใช้วินิจฉัยความไม่แปรเปลี่ยนตามประชากร (population invariance) ของแบบแผนการปรับเทียบคะแนนแบบใช้แบบสอบร่วมของกลุ่มไม่เท่าเทียม (non-equivalent-groups anchor test : NEAT) และใช้การปรับเทียบแบบลูกโซ่ (Chain Equating : CE) กับการปรับเทียบแบบอควิเปอร์เซ็นไทล์ ที่เรียกว่า Post-Stratification Equating (PSE) ซึ่งผลการศึกษาพบว่า วิธีการของพวกเขาสามารถใช้ได้ดีกับวิธีการปรับเทียบคะแนนทั้งสองวิธี แต่ก็มีปัจจัยที่สำคัญ 2 ปัจจัยที่ต้องคำนึงถึงคือ ความสัมพันธ์ระหว่างแบบสอบร่วมกับแบบสอบที่นำมาปรับเทียบ และความแตกต่างระหว่างประชากรผู้สอบ ซึ่งทั้งสองปัจจัยมีบทบาทที่สำคัญในแบบแผนการปรับเทียบคะแนนแบบ NEAT

ฮานิกและฮวง (Hanick & Huang. 2002 : 13) ได้ทำการศึกษาอิทธิพลของการลดจำนวนข้อสอบร่วมในการปรับเทียบคะแนน โดยใช้แบบสอบที่มีลิขสิทธิ์/ใบรับรอง จำนวน 200 ข้อ โดยมีเนื้อหา 6 เรื่อง มีข้อสอบร่วม (common item) เนื้อหาละ 10 ข้อ รวม 60 ข้อ วิธีการปรับเทียบใช้การปรับเทียบเชิงเส้นของ Tucker การปรับเทียบเชิงเส้นโดยใช้คะแนนสังเกตของ Levine และการปรับเทียบเชิงเส้นโดยใช้คะแนนจริงของ Levine กลุ่มตัวอย่างที่ใช้จำนวน 10,000 คน ผลการศึกษาพบว่า สำหรับแบบสอบที่มีลิขสิทธิ์หรือใบรับรองนั้นจะมีความแปรปรวนสูงต่อการลดจำนวนข้อสอบร่วม ซึ่งสำหรับการปรับเทียบคะแนนที่มีข้อสอบร่วม 30 ข้อ สามารถลดจำนวนข้อสอบร่วมลงได้ 5 - 10 ข้อ ประสิทธิภาพการปรับเทียบคะแนนก็ยังอยู่ในเกณฑ์ที่น่าพอใจ

วูดครัฟและซีโอเม็ก (Woodruff & Ziomek. 2004 : 2) ได้ศึกษาความเฟ้อ (inflation) ของเกรดเฉลี่ยในโรงเรียนมัธยมในอเมริกา โดยเปรียบเทียบเกรดเฉลี่ย (GPA) และคะแนนสอบ ACT (The American College Testing Program) ในช่วงปี 1991-2003 กลุ่มตัวอย่างปีละ 637,261 คน ถึง 781,460 คน พบว่า ในช่วงเวลา 13 ปีที่ทำการศึกษ เกรดเฉลี่ยในระดับมัธยม (เกรด 11 และ 12) ของสหรัฐอเมริกา มีความเฟ้อเพิ่มขึ้นร้อยละ 12.5

4.2 งานวิจัยภายในประเทศ

สุกมาศ อังศุโชติ (2543) ได้ศึกษาการปรับเทียบคะแนนเฉลี่ยสะสมระดับมัธยมศึกษาตอนปลาย การเปรียบเทียบความตรงเชิงพยากรณ์ กลุ่มตัวอย่างในการปรับเทียบคะแนนเฉลี่ยสะสมระดับมัธยมศึกษาตอนปลาย ได้แก่ นักเรียนจำนวน 5,919 คน ที่สำเร็จการศึกษาในปีการศึกษา 2539 จากโรงเรียนสังกัดกรมสามัญศึกษา กระทรวงศึกษาธิการ ที่มีคุณภาพต่างกัน 3 ระดับในเขตกรุงเทพมหานครและภาคกลาง จำนวน 28 โรงเรียน และใช้เฉพาะกลุ่มตัวอย่างที่ผ่านการสอบ

คัดเลือกเข้าศึกษาในมหาวิทยาลัยของรัฐในปีการศึกษา 2540 จำนวน 1,029 คน ในการตรวจสอบความตรงเชิงพยากรณ์ ข้อมูลที่ใช้ในการวิจัยประกอบด้วย คะแนนเฉลี่ยสะสมระดับมัธยมศึกษาตอนปลาย คะแนนสอบเข้ามหาวิทยาลัย และคะแนนเฉลี่ยสะสมระดับมหาวิทยาลัยปี 1 และ ปี 2 ของผู้ที่สอบผ่านการคัดเลือก ผลการวิจัยพบว่า โรงเรียนที่มีระดับคุณภาพในกลุ่มเดียวกันหรือต่างกลุ่มกันมีมาตรฐานการให้คะแนนเฉลี่ยสะสมต่างกันเป็น 4 ลักษณะคือ 1) ให้คะแนนเฉลี่ยสะสมสูงกว่าที่ควรจะเป็นในทุกช่วงคะแนนเฉลี่ยสะสม 2) ให้คะแนนเฉลี่ยสะสมต่ำกว่าที่ควรจะเป็นในทุกช่วงคะแนนเฉลี่ยสะสม 3) ให้คะแนนเฉลี่ยสะสมสูงกว่าที่ควรจะเป็นในทุกช่วงคะแนนเฉลี่ยสะสมต่ำแต่ให้คะแนนเฉลี่ยสะสมต่ำกว่าที่ควรจะเป็นในช่วงคะแนนเฉลี่ยสะสมสูง และ 4) ให้คะแนนเฉลี่ยสะสมต่ำกว่าที่ควรจะเป็นในช่วงคะแนนเฉลี่ยสะสมต่ำแต่ให้คะแนนเฉลี่ยสะสมสูงกว่าที่ควรจะเป็นในช่วงคะแนนเฉลี่ยสะสมสูง และเมื่อใช้คะแนนเฉลี่ยสะสมระดับมหาวิทยาลัยปี 1 และ ปี 2 ที่ปรับโดยใช้สมการถดถอยและปรับด้วยโมเดลหลายฟาเซทของราสซ์เป็นเกณฑ์ให้ผลสอดคล้องกัน คือที่ระดับนัยสำคัญ .05 คะแนนเฉลี่ยสะสมระดับมัธยมศึกษาตอนปลาย ที่ปรับด้วยวิธีปรับ 3 วิธี คือ วิธี RASC - FACET วิธี ANGOFF-4C2 วิธี IRT-GRM ให้ความตรงเชิงพยากรณ์สูงกว่าคะแนนเฉลี่ยสะสมระดับมัธยมศึกษาตอนปลายเดิม และวิธี CON-CFA ให้ความตรงเชิงพยากรณ์ต่ำกว่าคะแนนเฉลี่ยสะสมระดับมัธยมศึกษาตอนปลายเดิม ผลที่ต่างกันคือ เมื่อใช้คะแนนเฉลี่ยสะสมระดับมหาวิทยาลัยปี 1 ปี 2 ที่ปรับโดยใช้สมการถดถอยเป็นเกณฑ์ วิธี GLM-MODEL ให้ความตรงเชิงพยากรณ์สูงกว่าคะแนนเฉลี่ยสะสมระดับมัธยมศึกษาตอนปลายเดิม แต่จะไม่แตกต่างกัน เมื่อใช้คะแนนเฉลี่ยสะสมระดับมหาวิทยาลัยปี 1 ปี 2 ที่ปรับด้วยวิธี หลายฟาเซทของราสซ์เป็นเกณฑ์ ส่วนการเปรียบเทียบความตรงเชิงพยากรณ์ระหว่างวิธีพบว่า วิธี RASCH - FACET ให้ผลดีที่สุด รองลงมาคือ วิธี ANGOFF-4C2 และ วิธี IRT-GRM ซึ่งสองวิธีนี้ให้ความตรง เชิงพยากรณ์ไม่แตกต่างกัน อันดับที่ดีที่สุดคือวิธี GLM-MODEL และอันดับสุดท้ายคือ วิธี CON-CFA

สุนิสา จุ้ยม่วงศรี (2546) ได้ศึกษาผลของการเทียบคะแนนข้อสอบที่ตรวจให้คะแนนแบบหลายค่า เครื่องมือที่ใช้ในการวิจัยประกอบด้วยแบบทดสอบเลือกตอบวัดผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ด้านเหตุผล ได้แก่ ฉบับ X และฉบับ Y ซึ่งผู้วิจัยสร้างขึ้นสำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 2 แต่ละฉบับประกอบด้วยข้อสอบที่มี 5 ตัวเลือก จำนวน 30 ข้อ รวมข้อสอบรวมจำนวน 10 ข้อ แต่ละข้อตรวจให้คะแนน 6 ค่า คือ 0 1 2 3 4 และ 5 แบบทดสอบฉบับ X และฉบับ Y มีค่าความเชื่อมั่นเท่ากับ 0.937 และ 0.940 ตามลำดับ กลุ่มตัวอย่างที่ใช้ในการวิจัย เป็นนักเรียนชั้นมัธยมศึกษา ปีที่ 2 สังกัดกรมสามัญศึกษา ในเขตกรุงเทพมหานคร ซึ่งเลือกมาด้วยวิธีการแบ่งชั้นจำนวน 2,901 คน แบ่งเป็นกลุ่มตัวอย่างเทียบคะแนน เพื่อนำผลการสอบไปวิเคราะห์เทียบคะแนนแบบทดสอบจำนวน 2,312 คน โดยการสุ่มอย่างมีระบบให้แต่ละคนทำแบบทดสอบฉบับ X หรือ

ฉบับ Y เพียงหนึ่งฉบับ ได้ผู้สอบฉบับ X จำนวน 1,155 คนและฉบับ Y จำนวน 1,157 คน อีกกลุ่มใช้กลุ่มตัวอย่างสอบทานผลเพื่อนำผลการตรวจสอบคุณภาพของการเทียบคะแนนในแต่ละวิธี จำนวน 589 คน ประเมินคุณภาพของวิธีการเทียบคะแนน โดยกลุ่มสอบทานผลแต่ละคนต้องทำแบบทดสอบทั้งสองฉบับ โดยทำแบบทดสอบแต่ละฉบับสลับกัน ผลการวิจัยพบว่า 1) คะแนนสมมูลของแบบทดสอบฉบับ Y ส่วนใหญ่สูงกว่าคะแนนของแบบทดสอบฉบับ X จากการเทียบคะแนนทั้ง 5 วิธี 8 รูปแบบ โดยเฉพาะการเทียบคะแนนเชิงเส้นตรงให้ค่าสูงกว่าตลอดช่วงมาตรวัดของ X 2) การแจกแจงคะแนนสะสมระหว่างคะแนนแปลงจากการเทียบคะแนนแต่ละวิธีกับคะแนนจากแบบทดสอบเป้าหมาย (ฉบับ X) มีความคล้ายคลึงกัน 3) ค่าดัชนีคะแนนความแตกต่าง RMSD ทั้งถ่วงน้ำหนักและไม่ถ่วงน้ำหนัก MAD ทั้งถ่วงน้ำหนักและไม่ถ่วงน้ำหนัก MSD ไม่ถ่วงน้ำหนัก ให้ผลที่คล้ายคลึงกันคือ ฟังก์ชันการเทียบคะแนนจริงวิธีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ความยากมีความคล้ายคลึงกับฟังก์ชันการเทียบคะแนนพื้นฐานมากที่สุด โดยเฉพาะฟังก์ชันการเทียบคะแนนอิกวิเปอร์เซ็นไทล์ สำหรับผลจากดัชนี MSD ถ่วงน้ำหนัก สรุปได้ว่าฟังก์ชันการเทียบคะแนนสังเกต วิธีปรับค่าพารามิเตอร์พร้อมกัน มีความคล้ายคลึงกับฟังก์ชันการเทียบคะแนนพื้นฐานมากที่สุด โดยเฉพาะฟังก์ชันการเทียบคะแนนอิกวิเปอร์เซ็นไทล์ และดัชนี RMSD MAD และ MSD ทั้งถ่วงน้ำหนักและไม่ถ่วงน้ำหนักรวม 6 ดัชนี ให้ผลที่สอดคล้องกัน คือฟังก์ชันการเทียบคะแนนสังเกตวิธีโค้งลักษณะแบบทดสอบแตกต่างจากฟังก์ชันการเทียบคะแนนเชิงเส้นตรงมากที่สุด และ 4) ค่าดัชนีความแตกต่างจากการเทียบคะแนน แต่ละวิธีให้ผลอยู่ในระดับที่ยอมรับได้ทั้งสิ้น คือมีค่าต่ำกว่าค่าตั้งสองของ 0.05 ของค่าส่วนเบี่ยงเบนมาตรฐานของผลสอบจากฉบับ X และพบว่า ความคลาดเคลื่อนของการเทียบคะแนนในแต่ละวิธีจากกลุ่มสอบทานผล ไม่แตกต่างกัน

อุทัยวรรณ พงศ์อร่าม (2545) ได้ศึกษานาขนาดกลุ่มตัวอย่างที่เหมาะสมสำหรับการปรับเทียบคะแนนด้วยวิธีอิกวิเปอร์เซ็นไทล์และวิธีเชิงเส้นตรงตามแบบจำลองคะแนนจริงสัมพัทธ์ ที่มีแบบแผนการปรับเทียบและความยาวของแบบทดสอบแตกต่างกัน กลุ่มตัวอย่างที่ใช้ในการวิจัยเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 1 ของโรงเรียนมัธยมศึกษา ตั้งกัณฑ์กรมสามัญศึกษา จังหวัดชุมพร ภาคเรียนที่ 1 ปีการศึกษา 2544 จำนวน 3,000 คน ซึ่งได้มาโดยการสุ่มสองขั้นตอน แบ่งเป็นสองกลุ่มคือกลุ่มตัวอย่างปรับเทียบคะแนน และกลุ่มตัวอย่างสอบทานผล เครื่องมือที่ใช้ในการวิจัยเป็นแบบทดสอบวัดผลสัมฤทธิ์วิชาคณิตศาสตร์ ค.101 ชั้นมัธยมศึกษาปีที่ 1 จำนวนสองฉบับ การวิเคราะห์ข้อมูลใช้วิธีการปรับเทียบคะแนนด้วยวิธีอิกวิเปอร์เซ็นไทล์ และวิธีเชิงเส้นตรงตามแบบแผนกลุ่มสมมูลและแบบแผนข้อสอบร่วมเพื่อสร้างตารางแปลงคะแนนสมมูลและวิเคราะห์ดัชนีความแตกต่าง (C) หาขนาดกลุ่มตัวอย่างที่เหมาะสม โดยใช้เกณฑ์ตัดสินจากค่าดัชนีความแตกต่าง

(C) ที่อยู่ในระดับน่าพอใจและน่าพอใจมาก ผลการวิจัยพบว่า 1) ขนาดของกลุ่มตัวอย่างที่เหมาะสมในการเปรียบเทียบคะแนนวิธีเชิงเส้นตรง ในแบบแผน กลุ่มสมมูล ที่มีความยาวของแบบทดสอบเป็น 60, 50, และ 40 ข้อ เป็น 100 คน ส่วน 30 ข้อ เป็น 200 คน สำหรับการเปรียบเทียบคะแนนวิธีอิควิเปอร์เซ็นไทล์ในแบบแผน กลุ่มสมมูลที่มีความยาวของแบบทดสอบเป็น 60 ข้อ มีขนาดกลุ่มตัวอย่างที่เหมาะสมเป็น 200 คน และ 50 กับ 40 ข้อ เป็น 300 คน ส่วน 30 ข้อ เป็น 400 คน 2) ขนาดของกลุ่มตัวอย่างที่เหมาะสมในการเปรียบเทียบคะแนนวิธีเชิงเส้นตรง ในแบบแผนข้อสอบร่วม ที่มีความยาวของแบบทดสอบเป็น 60, 50 และ 40 ข้อ เป็น 100 คน ส่วน 30 ข้อ เป็น 300 คน สำหรับการเปรียบเทียบคะแนนวิธีอิควิเปอร์เซ็นไทล์ในแบบแผนข้อสอบร่วมที่มีความยาวของแบบทดสอบเป็น 60 และ 50 ข้อ เป็น 200 คน ส่วน 40 และ 30 ข้อ เป็น 500 คน 3) การเปรียบเทียบคะแนนด้วยวิธีเชิงเส้นตรงและวิธีอิควิเปอร์เซ็นไทล์ทั้งแบบแผนกลุ่มสมมูลและแบบแผนข้อสอบร่วมขนาดของกลุ่มตัวอย่างที่เหมาะสมแตกต่างกันในแต่ละความยาวของแบบทดสอบ โดยการเปรียบเทียบคะแนนด้วยวิธีเชิงเส้นตรงใช้กลุ่มตัวอย่างที่น้อยกว่า และ 4) ตามแบบแผนการเปรียบเทียบคะแนนแบบกลุ่มสมมูล และแบบแผนข้อสอบร่วมทั้งวิธีเชิงเส้นตรงและวิธีอิควิเปอร์เซ็นไทล์ ขนาดของกลุ่มตัวอย่างที่เหมาะสมในแต่ละความยาวของแบบทดสอบไม่แตกต่างกัน

ภทราพร เกษสังข์ (2546) ได้ศึกษาพัฒนาการความสามารถทางคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาตอนต้น โดยการเทียบคะแนนในแนวตั้งที่ใช้วิธีการที่เหมาะสม กลุ่มตัวอย่างของการวิจัย เป็นนักเรียนมัธยมศึกษาปีที่ 1, 2 และ 3 สังกัดกองการมัธยมศึกษา กรมสามัญศึกษา ในจังหวัดชัยนาท จำนวน 4,041 คน ที่ได้จากการสุ่มกลุ่มตัวอย่างแบบแบ่งชั้น แบ่งกลุ่มตัวอย่างเป็นสองกลุ่ม กลุ่มแรก จำนวน 3,081 คน ใช้สำหรับศึกษาการเทียบคะแนน เพื่อนำผลไปเทียบคะแนนระหว่างแบบทดสอบ และศึกษาพัฒนาการความสามารถทางคณิตศาสตร์ กลุ่มที่สอง จำนวน 960 คน ใช้สำหรับสอบทานผล เพื่อนำผลไปตรวจสอบประสิทธิภาพของการเทียบคะแนน เครื่องมือวัดที่ใช้เป็น แบบทดสอบคณิตศาสตร์ เรื่องสมการและอสมการ ที่มีความเชื่อมั่นของระดับชั้นมัธยมศึกษาปีที่ 1, 2 และ 3 เท่ากับ 0.745, 0.945 และ 0.951 ตามลำดับ และแบบทดสอบวัดความถนัดคำนวณตัวเลขที่มีความเชื่อมั่นของระดับชั้นมัธยมศึกษาปีที่ 1, 2 และ 3 เท่ากับ 0.792, 0.847 และ 0.708 ตามลำดับ ผลการวิจัยพบว่า 1. การวิเคราะห์ความคลาดเคลื่อนมาตรฐานของการเทียบคะแนนตามรูปแบบที่ศึกษาสำหรับผู้ที่มีความสามารถต่างกันสามระดับ พบว่า มีค่าระหว่าง 0.1011 ถึง 0.1830 เป็นค่าระดับต่ำ ความคลาดเคลื่อนมาตรฐานของการเทียบคะแนนต่ำที่สุด เป็นการเทียบคะแนนในแนวตั้ง ตามทฤษฎีการตอบสนองข้อสอบรูปแบบโลจิสติกสองพารามิเตอร์ ที่แบบทดสอบร่วมภายใน 15 ข้อ มีความยากของข้อสอบร่วมเฉลี่ยอยู่ในระดับยากมาก ($\bar{X} = 2.2$) 2. การวิเคราะห์ความเพียงพอของการเทียบคะแนน พบว่า ค่าดัชนีความแตกต่างของการเทียบคะแนน

(C) อยู่ในระดับน่าพอใจมาก และระดับน่าพอใจ มีความเพียงพอของการเทียบคะแนนที่ดีที่สุด เป็นการเทียบคะแนนในแนวตั้งตามทฤษฎีการตอบสนองข้อสอบรูปแบบโลจิสติก X สองพารามิเตอร์ ที่แบบทดสอบรวมภายใน 15 ข้อ ที่ลักษณะความยากของข้อสอบรวมเฉลี่ยระดับยากมาก และ 3. การศึกษาพัฒนาการความสามารถทางคณิตศาสตร์ จากรูปแบบการเทียบคะแนนที่มีประสิทธิภาพ พบว่า ความสามารถทางคณิตศาสตร์ ของนักเรียนชั้นมัธยมศึกษาปีที่ 1 ไปชั้นมัธยมศึกษาปีที่ 2 กับนักเรียนชั้นมัธยมศึกษาปีที่ 2 ไปชั้นมัธยมศึกษาปีที่ 3 มีพัฒนาการสูงขึ้นและนักเรียนชั้นมัธยมศึกษาปีที่ 1 ไปชั้นมัธยมศึกษาปีที่ 3 มีพัฒนาการสูงขึ้นมาก

อดิศร ศรีบุญวงษ์ (2545) ได้ศึกษาการพัฒนาเกณฑ์ตัดสินคุณภาพการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบ โดยใช้ข้อมูลจำลองจากโปรแกรม IRT DATA ที่ตรวจสอบความถูกต้องด้วยการวิเคราะห์ซ้ำด้วยโปรแกรม BILOG การจำลองข้อมูลตามเงื่อนไขการปรับเทียบคะแนนด้วย โมเดลการตอบสนองข้อสอบแบบ 1 และ 3 พารามิเตอร์ วิเคราะห์คุณภาพของการปรับเทียบคะแนนด้วยการปรับเทียบกลับสู่แบบสอบเดิมและการใช้กลุ่มสอบทานผล เก็บรวบรวมข้อมูลโดยใช้กลุ่มสมมูลและการใช้ข้อสอบร่วม วิธีปรับเทียบคะแนน 4 วิธี คือ วิธี The Same Scaling Convention วิธี Mean and Mean วิธี Mean and Sigma และวิธี Characteristic Curve ที่วิเคราะห์ด้วยโปรแกรม EQUATE ผู้สอบจำนวน 100, 200, 300, ..., 3,000 คน และข้อสอบจำนวน 10, 20, 30, ..., 200 ข้อ จัดกลุ่มค่าดัชนีความแตกต่าง AMD, MAD และ RMS โดยใช้จุดตัดซึ่งเป็นค่าดัชนีแสดงความแตกต่างเมื่อทดสอบความแตกต่างระหว่างค่าเฉลี่ยของคะแนนจริงที่ปรับเทียบแล้วกับคะแนนจริงที่ยังไม่ได้ปรับเทียบ สำหรับผู้สอบกลุ่มเดียวกันด้วยสถิติ t-test แล้วพบค่าวิกฤตที่ระดับนัยสำคัญทางสถิติที่ระดับ .001, .01 และ .02 คำนวณค่าดัชนีความแตกต่างแล้วนำกลุ่มค่าดัชนีความแตกต่างแต่ละกลุ่มมาระบุระดับคุณภาพการปรับเทียบคะแนนเพื่อกำหนดเป็นเกณฑ์ผลการวิจัยพบว่า 1) ได้เกณฑ์ตัดสินคุณภาพการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบที่พัฒนาขึ้น มีทั้งหมด 7 เกณฑ์ คือ เกณฑ์รวมทุกเงื่อนไข เกณฑ์สำหรับ โมเดล 1 พารามิเตอร์ เกณฑ์สำหรับ โมเดล 3 พารามิเตอร์ เกณฑ์สำหรับการใช้กลุ่มสมมูล เกณฑ์สำหรับการใช้แบบสอบร่วม เกณฑ์สำหรับการปรับเทียบกลับสู่แบบสอบเดิม และเกณฑ์สำหรับการใช้กลุ่มสอบทานผล แต่ละเกณฑ์เป็นกลุ่มค่าดัชนี AMD, MAD และ RMS ที่บอกระดับคุณภาพการปรับเทียบคะแนน 4 ระดับ คือ การปรับเทียบน่าพอใจอย่างยิ่ง การปรับเทียบน่าพอใจ การปรับเทียบไม่น่าพอใจ และการปรับเทียบไม่น่าพอใจอย่างยิ่ง และ 2) ผลการตัดสินคุณภาพการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบที่พัฒนาขึ้นเปรียบเทียบกับผลการใช้ ตามเกณฑ์ของปีเตอร์เซ็นและคณะพบว่าไม่สอดคล้องกัน แต่เมื่อเทียบกับเกณฑ์ความเสมอภาคของลอว์ด์ ปรากฏว่า ผลการตัดสินคุณภาพการปรับเทียบมีความสอดคล้องกัน

อวยพร เรื่องตระกูล (2544) ได้ศึกษาการพัฒนาและวิเคราะห์คุณภาพของวิธีการวัดคะแนน พัฒนาการตามทฤษฎีการทดสอบแบบดั้งเดิม และทฤษฎีการตอบสนองข้อสอบ กลุ่มตัวอย่างเป็น นักเรียนชั้นมัธยมศึกษาปีที่ 3 จำนวน 2 กลุ่ม กลุ่มละ 698 คน และ 637 คน จากโรงเรียนสังกัดกรมสามัญศึกษาเขตกรุงเทพมหานคร 12 โรงเรียน เครื่องมือที่ใช้ในการวิจัยครั้งนี้เป็นแบบทดสอบวิชาคณิตศาสตร์ ระดับชั้นมัธยมศึกษาปีที่ 3 จำนวน 3 ฉบับที่คู่ขนานกัน ข้อมูล ที่ใช้ในการวิจัยมี 3 ชุด สองชุดแรกเป็นข้อมูลปฐมภูมิ ที่ได้จากการเก็บรวบรวมข้อมูลด้วยแบบทดสอบคณิตศาสตร์ที่ทำการวัดซ้ำ 3 ครั้ง ด้วยแบบทดสอบฉบับเดิม และแบบจัดรัสละติน ด้วยแบบทดสอบคู่ขนาน ส่วนข้อมูลชุดที่สาม เป็นข้อมูลทุติยภูมิที่ได้จากคะแนนสอบวิชาคณิตศาสตร์ ระดับชั้นมัธยมศึกษาปีที่ 2 ที่มีการวัดซ้ำ 5 ครั้ง มีการตรวจให้คะแนนแบบทวิภาค (Dichotomous) และให้คะแนนความรู้บางส่วน (Partial credit) การวิเคราะห์ข้อมูลเป็นการประมาณค่าคะแนนพัฒนาการ และการวิเคราะห์เปรียบเทียบคุณภาพของวิธีการวัดคะแนนพัฒนาการในด้านความตรงตามเกณฑ์สัมพัทธ์และค่าความคลาดเคลื่อนด้วยการวิเคราะห์โมเดลเชิงเส้นตรงระดับลดหลั่นและสถิติทดสอบ Hotelling ผลการวิจัยพบว่า 1) เมื่อเปรียบเทียบผลการประมาณค่าคะแนนพัฒนาการระหว่าง 3 กลุ่มวิธี พบว่า กลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม กลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนาขึ้นมีคุณภาพไม่แตกต่างกัน 2) เมื่อเปรียบเทียบระหว่างวิธีภายในกลุ่มทฤษฎีการทดสอบแบบดั้งเดิม พบว่า วิธีการวัดคะแนนพัฒนาการจากความแตกต่างระหว่างคะแนนดิบ และวิธีการวัดคะแนนพัฒนาการสัมพัทธ์มีคุณภาพสูงกว่าวิธีอื่นสำหรับข้อมูล 2 ชุดแรก และวิธีการวัดคะแนนพัฒนาการส่วนที่เหลือเทียบกับศักยภาพของผู้สอบ มีคุณภาพสูงกว่าวิธีอื่นสำหรับข้อมูลชุดที่ 3 3) เมื่อเปรียบเทียบระหว่างวิธีภายในกลุ่มทฤษฎีการตอบสนองข้อสอบ พบว่า วิธีการวัดคะแนนพัฒนาการ โดยใช้โมเดลราสซ์พหุมิติ สำหรับการเรียนรู้และการเปลี่ยนแปลงมีคุณภาพสูงกว่าวิธีการวัดคะแนนพัฒนาการจากความแตกต่างของความสามารถที่แท้จริง และ 4) เมื่อเปรียบเทียบระหว่างวิธีภายในกลุ่มทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา พบว่า วิธีการวัดคะแนนพัฒนาการจากความสามารถที่แท้จริงสัมพัทธ์เมื่อเทียบกับความสามารถที่แท้จริงก่อนเรียนมีคุณภาพสูงกว่าวิธีการวัดคะแนนพัฒนาการจากความสามารถที่แท้จริงสัมพัทธ์เมื่อเทียบกับศักยภาพการพัฒนา สำหรับข้อมูลที่มีการตรวจให้คะแนนแบบทวิภาค

กนิษฐา แสนแก้ว (2541) ได้ศึกษาการปรับแก้เฉลยสะสมระดับมัธยมศึกษาตอนปลายด้วยคะแนนสอบความถนัดทางการเรียน การเปรียบเทียบระหว่างวิธีการปรับแก้แบบอิกวิเปอร์เซ็นไทล์ เชิงเส้นตรงและ ไออาร์ที เพื่อเปรียบเทียบคุณภาพของวิธีการปรับแก้เฉลยสะสมระดับมัธยมศึกษาตอนปลายระหว่างวิธีการปรับแก้แบบอิกวิเปอร์เซ็นไทล์ เชิงเส้นตรงและ ไออาร์ที

โดยอาศัยกรณีของกลุ่มตัวอย่างที่มีแผนการเรียนแตกต่างกันและจบจากโรงเรียนที่มีคุณภาพ การศึกษาแตกต่างกัน การเปรียบเทียบการปรับเทียบกระทำโดยการวิเคราะห์การถดถอยพหุคูณ ระหว่างแเต็มเฉลี่ยสะสมที่ปรับแล้ว ซึ่งมีผลการเรียนเฉลี่ยสะสมระดับอุดมศึกษา ภาคเรียนที่ 1 และ คะแนนสอบเข้ามหาวิทยาลัยเป็นเกณฑ์เพื่อเปรียบเทียบในการพยากรณ์ของแต่ละวิธี กลุ่มตัวอย่างที่ใช้ในการวิจัยเป็นนิสิตชั้นปีที่ 2 คณะครุศาสตร์/ศึกษาศาสตร์ ที่เข้าศึกษาในปีการศึกษา 2540 ของ จุฬาลงกรณ์มหาวิทยาลัย มหาวิทยาลัยเกษตรศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒและ มหาวิทยาลัยศิลปากร โดยเรียนจบชั้นมัธยมศึกษาปีที่ 6 และผ่านการสอบคัดเลือกโดย ทบวงมหาวิทยาลัย จำนวน 559 คน เครื่องมือที่ใช้ในการวิจัยแบ่งเป็น 2 ส่วน คือแบบบันทึกข้อมูล ทักษะภูมิ และแบบทดสอบความถนัดทางการเรียน โดยวัดความรู้ความสามารถทั่วไปเกี่ยวกับ ความสามารถทางภาษา การคิดคำนวณ และการวิเคราะห์ ผลการวิจัยพบว่า 1) แเต็มเฉลี่ยสะสม ระดับมัธยมศึกษาตอนปลายที่ปรับแล้วทั้งสามวิธี มีความสัมพันธ์กับคะแนนสอบเข้ามหาวิทยาลัย และผลการเรียนเฉลี่ยระดับอุดมศึกษา ภาคเรียนที่ 1 สูงกว่าแเต็มเฉลี่ยสะสมระดับมัธยมศึกษาตอน ปลายปกติ 2) ผลการถดถอยพหุคูณพบว่า 2.1) เมื่อใช้ผลการเรียนสะสมระดับอุดมศึกษา ภาคเรียน ที่ 1 เป็นเกณฑ์ วิธีไออาร์ที่เป็นวิธีการปรับเทียบที่มีคุณภาพมากที่สุด รองลงมาคือ วิธีเชิงเส้นตรง และวิธีอควิเปอร์เซ็นไทล์ ตามลำดับ 2.2) เมื่อใช้คะแนนสอบเข้ามหาวิทยาลัยเป็นเกณฑ์ วิธีอควิ เอร์เซ็นไทล์เป็นวิธีการปรับเทียบที่มีคุณภาพมากที่สุด รองลงมาคือ วิธี ไออาร์ทีและ วิธีเชิง เส้นตรง ตามลำดับ

อรุณี อ่อนสวัสดิ์ และชอบ ลีซอ (2541) ได้ศึกษาผลของวิธีแปลงคะแนนผลการเรียนสะสม ในการสอบคัดเลือกบุคคลเข้าศึกษาในสถาบันอุดมศึกษาระบบใหม่ เพื่อศึกษาผลของวิธีแปลง ผล การเรียนสะสมเป็นคะแนนมาตรฐานกลาง และคะแนนผลการเรียนสะสมที่อิงตำแหน่งเปอร์เซ็นต์ ไทล์ รวมทั้งศึกษาความสัมพันธ์ระหว่างคะแนนรวมทั้ง 2 วิธี กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษา ปีที่ 6 ปีการศึกษา 2540 ที่เรียนอยู่ในโรงเรียนมัธยม กรมสามัญศึกษา เขตการศึกษา 7 จำนวน 1,176 คน จากโรงเรียนขนาดใหญ่ ขนาดกลางและขนาดเล็ก เก็บรวบรวมข้อมูลโดยใช้แบบทดสอบมาตรฐาน ความถนัดทางการเรียน 3 ฉบับ และขอหลักฐานผลการเรียนสะสม 6 ภาคเรียน ของนักเรียนที่เป็นกลุ่ม ตัวอย่างจากทางโรงเรียน วิเคราะห์ข้อมูลเพื่อหาค่าสถิติ การเรียงลำดับคะแนนและการหาค่า สหสัมพันธ์ เป็นต้น ผลการวิจัยพบว่า ลำดับที่ของค่ามัชฌิมเลขคณิตผลการเรียนสะสมและความ ถนัดทางการเรียนของนักเรียนแต่ละ โรงเรียน ไม่สอดคล้องกัน ผู้ได้ผลการเรียนเท่ากันจาก โรงเรียน เดียวกัน หรือต่างโรงเรียนกันจะได้คะแนนมาตรฐานกลางมาไม่เท่ากัน คะแนนผลการเรียน ความ ถนัดทางการเรียน และคะแนนมาตรฐานกลาง สัมพันธ์ซึ่งกันและกันสูง อย่างมีนัยสำคัญทางสถิติที่ ระดับ .001 มีนักเรียนส่วนใหญ่ของบางโรงเรียนได้คะแนนร้อยละ 5 แรกสูงกว่าคะแนนร้อยละ 5

หลัง ในขณะที่นักเรียนส่วนใหญ่ของบางโรงเรียนได้คะแนนร้อยละ 5 หลัง สูงกว่าคะแนนร้อยละ 5 แรก และมีนักเรียนส่วนใหญ่ของบางโรงเรียนได้คะแนนทั้งสองส่วน พอกัน คะแนนแปลงทั้ง 2 วิธี สัมพันธ์กันสูง อย่างมีนัยสำคัญทางสถิติที่ระดับ .001 วิธีแปลงคะแนนทั้ง 2 วิธี ให้โอกาสนักเรียนในโรงเรียนขนาดใหญ่ในสัดส่วนที่สูงกว่าโรงเรียนขนาดกลางและขนาดเล็ก

สุภาภรณ์ คงทวี (2541) ได้ศึกษาการปรับแก้คะแนนระดับมัธยมศึกษาตอนปลายด้วยคะแนนสอบเข้ามหาวิทยาลัย การเปรียบเทียบระหว่างวิธีการปรับเทียบแบบอิกวิเปอร์เซ็นไทล์เชิงเส้นตรงและไออาร์ที เพื่อเปรียบเทียบประสิทธิภาพของวิธีการปรับเทียบระหว่างอิกวิเปอร์เซ็นไทล์เชิงเส้นตรงและไออาร์ที ในการปรับแก้คะแนนระดับมัธยมศึกษาตอนปลายด้วยคะแนนสอบเข้ามหาวิทยาลัย กลุ่มตัวอย่างเป็นนิสิต/นักศึกษาชั้นปีที่ 1 ปีการศึกษา 2540 ซึ่งจบชั้นมัธยมศึกษาปีที่ 6 บริบูรณ์ในมหาวิทยาลัยของรัฐ ได้แก่ จุฬาลงกรณ์มหาวิทยาลัย มหาวิทยาลัยธรรมศาสตร์ มหาวิทยาลัยเกษตรศาสตร์ มหาวิทยาลัยศิลปากร และมหาวิทยาลัยศรีนครินทรวิโรฒ โดยวิเคราะห์แยกเป็น 6 กลุ่ม ได้แก่ กลุ่มสาขาวิชาวิทยาศาสตร์ วิศวกรรมศาสตร์ สถาปัตยกรรมศาสตร์ พาณิชยศาสตร์และการบัญชี มนุษยศาสตร์ สังคมศาสตร์ (แผนการสอบศิลป์) และมนุษยศาสตร์ สังคมศาสตร์ (แผนการสอบวิทยาศาสตร์) วิเคราะห์ข้อมูลโดยใช้สถิติบรรยาย วิธีการปรับเทียบอิกวิเปอร์เซ็นไทล์ วิธีเชิงเส้นตรงและวิธีไออาร์ที การวิเคราะห์การถดถอยอย่างง่ายและสถิติทดสอบที (t-test) ผลการวิจัยพบว่า 1) คะแนนที่ปรับแก้ด้วยวิธีอิกวิเปอร์เซ็นไทล์แก้คะแนนที่ปรับแก้ด้วยวิธีเชิงเส้นตรง และค่าพารามิเตอร์ความสามารถที่ปรับแก้ด้วยวิธีไออาร์ที มีความสัมพันธ์กับผลสัมฤทธิ์ทางการเรียนชั้นปีที่ 1 อย่างมีนัยสำคัญทางสถิติที่ระดับ .001 2) ผลการพยากรณ์ผลสัมฤทธิ์ทางการเรียนชั้นปีที่ 1 พบว่า 2.1) คะแนนที่ปรับแก้ด้วยวิธีอิกวิเปอร์เซ็นไทล์ แก้คะแนนที่ปรับแก้ด้วยวิธีเชิงเส้นตรง และค่าพารามิเตอร์ความสามารถที่ปรับแก้ด้วยวิธีไออาร์ที สามารถพยากรณ์ผลสัมฤทธิ์ทางการเรียนชั้นปีที่ 1 อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 2.2) เมื่อเปรียบเทียบกับผลการพยากรณ์ด้วย คะแนนเดิมและค่าพารามิเตอร์ความสามารถเดิม พบว่าในกลุ่มสาขาวิชาวิทยาศาสตร์ คะแนนที่ปรับแก้ด้วยวิธีอิกวิเปอร์เซ็นไทล์ วิธีเชิงเส้นตรง และค่าพารามิเตอร์ความสามารถที่ปรับแก้ด้วยวิธี ไออาร์ที มีประสิทธิภาพในการพยากรณ์ดีกว่าคะแนนเดิมและค่าพารามิเตอร์ความสามารถเดิม ที่ระดับนัยสำคัญทางสถิติ .01 ส่วนในกลุ่มสาขาวิชาวิศวกรรมศาสตร์ กับกลุ่มสาขาวิชามนุษยศาสตร์ สังคมศาสตร์ (แผนการสอบวิทยาศาสตร์) เฉพาะค่าพารามิเตอร์ความสามารถที่ปรับแก้ด้วยวิธี ไออาร์ทีเท่านั้นที่มีประสิทธิภาพในการพยากรณ์ดีกว่าค่าพารามิเตอร์ความสามารถเดิม ที่ระดับนัยสำคัญทางสถิติ .05 ขณะที่ในกลุ่มสาขาวิชาอื่นๆ คะแนนเดิมและค่าพารามิเตอร์ความสามารถเดิมกับคะแนนเดิมและค่าพารามิเตอร์ความสามารถที่ปรับแก้แล้ว มีประสิทธิภาพในการพยากรณ์ไม่แตกต่างกัน ยกเว้นสาขา

วิศวกรรมศาสตร์ที่เต็มเฉลี่ยสะสมเดิมมีประสิทธิภาพในการพยากรณ์ดีกว่าเต็มเฉลี่ยสะสมที่ปรับแล้ว ด้วยวิธีเชิงเส้นตรง อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 2.3) ผลการเปรียบเทียบประสิทธิภาพของวิธีการปรับเทียบระหว่างวิธีอิกวิเปอร์เซ็นไทล์ วิธีเชิงเส้นตรงและไออาร์ที ผลปรากฏว่าในสาขาวิชาวิทยาศาสตร์ วิธีเชิงเส้นตรงมีประสิทธิภาพดีกว่าวิธีอิกวิเปอร์เซ็นไทล์ สาขาวิศวกรรมศาสตร์ วิธีอิกวิเปอร์เซ็นไทล์และวิธีไออาร์ที มีประสิทธิภาพดีกว่าวิธีเชิงเส้นตรง อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ส่วนในกลุ่มวิชาอื่นๆ วิธีการปรับเทียบ 3 วิธีให้ประสิทธิภาพไม่แตกต่างกัน

สำราญ มีแจ้ง (2542) ได้ศึกษาการเปรียบเทียบคะแนนสอบคัดเลือกเข้าศึกษาในสถาบันอุดมศึกษาระบบใหม่ที่มีวิธีแปลงคะแนนผลการเรียนสะสมต่างกัน เพื่อเปรียบเทียบคะแนนสอบคัดเลือกเข้าศึกษาในสถาบันอุดมศึกษาระบบใหม่ในส่วนที่ได้จากการแปลงคะแนน ผลการเรียนสะสมตามวิธีของทบวง วิธีการแปลงเป็นคะแนนมาตรฐานและวิธีเทียบมาตรา กลุ่มตัวอย่างเป็นนักเรียนที่กำลังเรียนอยู่ในชั้นมัธยมศึกษาปีที่ 6 ของโรงเรียนมัธยมศึกษาในจังหวัดพิษณุโลก และจังหวัดเพชรบูรณ์ จำนวน 4,540 คน เก็บรวบรวมข้อมูลโดยใช้แบบสำรวจที่ผู้วิจัยสร้างขึ้นเอง วิเคราะห์ข้อมูลโดยหาค่าสถิติพื้นฐานของคะแนนแปลงจากผลการเรียนสะสม เปรียบเทียบคะแนนสอบคัดเลือกเข้าศึกษาในสถาบันอุดมศึกษาระบบใหม่ที่มีวิธีการแปลงคะแนนผลการเรียนสะสมแบบต่างๆระหว่างโรงเรียนที่มีขนาดต่างกัน โดยใช้ F-test หาค่าสหสัมพันธ์ระหว่างคะแนนสอบคัดเลือกเข้าศึกษาในสถาบันอุดมศึกษาระบบใหม่ในส่วนที่ได้จากการแปลงคะแนนผลการเรียนสะสมแบบต่างๆกับผลการเรียนสะสมจำแนกตามขนาดโรงเรียน จำแนกตามโรงเรียน และค่าสหสัมพันธ์ระหว่างคะแนนแปลง ผลการเรียนสะสมแต่ละวิธีโดยใช้สูตรสหสัมพันธ์แบบเพียร์สัน ผลการวิจัยพบว่า 1) คะแนนสอบคัดเลือกเข้าศึกษาในสถาบันอุดมศึกษาระบบใหม่ส่วนที่ได้จากการแปลงผลการเรียนสะสมตามวิธีของทบวง วิธีการแปลงเป็นคะแนนมาตรฐานและวิธีการเทียบมาตรา แตกต่างกันอย่างมีนัยสำคัญทางสถิติ 2) โรงเรียนที่มีขนาดต่างกันมีคะแนนสอบคัดเลือกเข้าศึกษาในสถาบันอุดมศึกษาในส่วนที่ได้จากการแปลงคะแนนผลการเรียนสะสมตามวิธีของทบวงไม่แตกต่างกัน ส่วนวิธีการแปลงเป็นคะแนนมาตรฐานและวิธีการเทียบมาตรา ทำให้คะแนนที่ได้จากการแปลงผลการเรียนสะสมตามของนักเรียนที่มาจากโรงเรียนขนาดต่างกันมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติ 3) คะแนนสอบคัดเลือกเข้าศึกษาในสถาบันอุดมศึกษาระบบใหม่ในส่วนที่ได้จากการแปลงคะแนนผลการเรียนสะสมในแต่ละวิธีมีความสัมพันธ์กับคะแนนผลการเรียนสะสมสูงในทางบวกอย่างมีนัยสำคัญทางสถิติทั้งภาพรวม ตามขนาดโรงเรียนและรายโรงเรียน และ 4) ความสัมพันธ์ระหว่างคะแนนสอบคัดเลือกเข้าศึกษาในสถาบันอุดมศึกษาในส่วนที่ได้จากการแปลงคะแนนผลการเรียนสะสมตามวิธีของทบวงกับวิธีการแปลงเป็นคะแนนมาตรฐานและวิธีการเทียบมาตรา มีความสัมพันธ์กันสูงในทางบวกอย่างมีนัยสำคัญทางสถิติ

ณรงค์ฤทธิ์ เตชะนานาเลิศ (2543) ได้ศึกษาการปรับค่าเกรดเฉลี่ยสะสมสำหรับการสอบเข้ามหาวิทยาลัย เพื่อศึกษาถึงความเหมาะสมในการใช้ค่าเกรดเฉลี่ย (GPAX) มาปรับให้เป็นคะแนนเพื่อนำไปพิจารณาร่วมกับคะแนนข้อเขียนในการสอบเข้ามหาวิทยาลัย โดยศึกษาความสอดคล้องกันระหว่างค่าเกรดเฉลี่ยสะสมและคะแนนสอบข้อเขียนเฉลี่ยของนักเรียนในแต่ละกลุ่มโรงเรียน และเปรียบเทียบประสิทธิภาพของวิธีการปรับค่าเกรดเฉลี่ยสะสมให้เป็นคะแนน เพื่อนำไปพิจารณาร่วมกับคะแนนสอบข้อเขียน โดยวิธีการปรับค่าให้เป็นมาตรฐาน (Norm) และวิธีการเทียบคะแนน (Equating) วิธีการเทียบคะแนน โดยใช้คะแนนผลการเรียนเฉลี่ยสะสม (GPAX) และคะแนนจากการสอบวัดความสามารถพื้นฐานของกลุ่มผู้สอบที่จะนำมาเป็นเกณฑ์ในการเทียบ ซึ่งมีวิธีการเทียบคะแนนจากค่าเฉลี่ย และคะแนนสอบวัดความสามารถพื้นฐานดังนี้ 1) นำข้อมูล GPA ของนักเรียนที่สมัครสอบคัดเลือกทุกคนไปคำนวณค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐาน 2) สร้างแบบสอบขึ้นมาฉบับหนึ่งนำไปสอบกับกลุ่มผู้สอบนั้น แล้วหาค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐาน และค่าคะแนนสูงสุด (X) ของคะแนนจากแบบสอบนั้น 3) นำข้อมูล GPA และคะแนนที่ได้จากการสอบในข้อ 1 และ ข้อ 2 ไปคำนวณหาค่าคงที่ของเกรด (Y) และค่าถ่วงน้ำหนัก (W) แล้วนำไปคำนวณค่าคะแนนเทียบ (X_e) โดยใช้ระดับผลการเรียน ($GPAX_i$) และคะแนนสอบ (X_i) ของผู้สอบที่ได้คะแนนจริงแต่ละคน ผลการวิจัยพบว่า คะแนนผลการเรียนเฉลี่ยสะสมไม่ได้สอดคล้องและสะท้อนถึงความสามารถที่แท้จริงของนักเรียน อันเป็นผลมาจากระบบการให้เกรดที่แตกต่างกันในแต่ละโรงเรียน และเมื่อเปรียบเทียบประสิทธิภาพของวิธีการปรับคะแนน พบว่า การปรับค่าเกรดเฉลี่ยสะสมด้วยวิธีการเทียบคะแนน (Equating) มีประสิทธิภาพสูงกว่าวิธีการปรับให้เป็นค่ามาตรฐาน (Norm)

น้ำฝน กันมา (2543) ได้ศึกษาวิธีการแปลงคะแนนสอบคัดเลือกเข้าศึกษาในสถาบันอุดมศึกษา โดยวิธีการของทบวงมหาวิทยาลัย วิธีการแปลงเป็นคะแนนมาตรฐานและวิธีการเทียบมาตราโดยใช้ข้อสอบร่วม เพื่อศึกษาวิธีการแปลงคะแนนสอบคัดเลือกเข้าศึกษาต่อในสถาบันอุดมศึกษา โดยวิธีการของทบวงมหาวิทยาลัย วิธีการแปลงเป็นคะแนนมาตรฐานและวิธีการเทียบมาตรา โดยใช้ข้อสอบร่วมกลุ่มตัวอย่างเป็นนักเรียนที่สำเร็จชั้นมัธยมศึกษาปีที่ 6 ภาคเรียนที่ 1 ปีการศึกษา 2542 ของโรงเรียนมัธยมศึกษา จังหวัดพิษณุโลก จำนวน 1,642 คน ผลการวิจัยพบว่า 1) การแจกแจงของคะแนนแปลงผลการเรียนสะสมโดยวิธีการของทบวงมหาวิทยาลัย วิธีการแปลงเป็นคะแนนมาตรฐานที่ และวิธีการเทียบมาตราโดยใช้ข้อสอบร่วม พบว่า 1.1) การแจกแจงของคะแนนแปลงผลการเรียนสะสมโดยวิธีการของทบวงมหาวิทยาลัย โดยภาพรวมมีค่าเฉลี่ยเท่ากับ 36.32 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 18.40 ค่าสูงสุดเท่ากับ 66.52 และค่าต่ำสุดเท่ากับ 2.17 1.2) การแจกแจงของคะแนนแปลงผล

การเรียนสะสมโดยวิธีการแปลงเป็นคะแนนมาตรฐานที่โดยภาพรวมมีค่าเฉลี่ยเท่ากับ 33.58 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 6.45 ค่าสูงสุดเท่ากับ 55.18 และค่าต่ำสุดเท่ากับ 17.37 และ 1.3) การแจกแจงของคะแนนแปลงผลการเรียนสะสมโดยวิธีการเทียบมาตราโดยใช้ข้อสอบร่วม โดยภาพรวมมีค่าเฉลี่ยเท่ากับ 36.70 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 6.70 ค่าสูงสุดเท่ากับ 58.99 และค่าต่ำสุดเท่ากับ 19.90 2) คะแนนแปลงผลการเรียนสะสมโดยวิธีการของทบวงมหาวิทยาลัย วิธีการแปลงเป็นคะแนนมาตรฐานที่ และวิธีการเทียบมาตราโดยใช้ข้อสอบร่วม ระหว่างโรงเรียนที่มีขนาดต่างกัน พบว่า 2.1) โรงเรียนที่มีขนาดต่างกันมีคะแนนแปลงผลการเรียนสะสม โดยวิธีการของทบวงมหาวิทยาลัยไม่แตกต่างกัน 2.2) โรงเรียนที่มีขนาดต่างกันมีคะแนนแปลงผลการเรียนสะสมโดยวิธีการแปลงเป็นคะแนนมาตรฐานที่ ของนักเรียน โปรแกรมวิทย์ไม่แตกต่างกัน แต่นักเรียนโปรแกรมศิลป์แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และเมื่อทดสอบความแตกต่างเป็นรายคู่ พบว่า โรงเรียนขนาดใหญ่พิเศษมีค่าเฉลี่ยของคะแนนแปลงผลการเรียนสะสมสูงกว่าโรงเรียนขนาดเล็ก 2.3) โรงเรียนที่มีขนาดต่างกันมีคะแนนแปลงผลการเรียนสะสม โดยวิธีการเทียบมาตราโดยใช้ข้อสอบร่วมของนักเรียน โปรแกรมวิทย์ ไม่แตกต่างกัน แต่นักเรียนโปรแกรมศิลป์แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และเมื่อทดสอบความแตกต่างเป็นรายคู่ พบว่า โรงเรียนขนาดใหญ่พิเศษมีค่าเฉลี่ยของคะแนนแปลงผลการเรียนสะสมสูงกว่าโรงเรียนขนาดเล็ก 3) คะแนนแปลงผลการเรียนสะสมในแต่ละวิธีมีความสัมพันธ์กับคะแนนผลการเรียนสะสมสูงในทางบวกอย่างมีนัยสำคัญทางสถิติทั้งภาพรวม ตามขนาดโรงเรียนและรายโรงเรียน และ 4) ความสัมพันธ์ระหว่างคะแนนแปลงผลการเรียนสะสม โดยวิธีการของทบวงมหาวิทยาลัย กับวิธีการแปลงเป็นคะแนนมาตรฐานและวิธีการเทียบมาตราโดยใช้ข้อสอบร่วมมีความสัมพันธ์กันสูงในทางบวกอย่างมีนัยสำคัญทางสถิติ

เลิศศักดิ์ ไชยศิลป์ (2546) ได้ศึกษาการปรับเทียบผลการสอบคัดเลือกเข้าศึกษาในมหาวิทยาลัยนเรศวรประเภทรับตรงที่ใช้วิธีรวมคะแนนต่างกัน โดยการรวมคะแนนดิบกับการรวมคะแนนมาตรฐาน เพื่อเปรียบเทียบลำดับที่ของผลการสอบคัดเลือกเข้าศึกษาในมหาวิทยาลัยนเรศวร ที่เกิดจากการรวมคะแนนดิบกับการรวม โดยแปลงเป็นคะแนนมาตรฐานที่เชิงเส้นและคะแนนมาตรฐานที่ปกติ และหาความสัมพันธ์ระหว่างคะแนนเฉลี่ยสะสม (GPA) คะแนนรวมดิบ คะแนนรวมมาตรฐานที่เชิงเส้น และคะแนนรวมมาตรฐานที่ปกติ กลุ่มตัวอย่างเป็นนักเรียนที่สมัครสอบคัดเลือกเข้าศึกษาเฉพาะคณะวิศวกรรมศาสตร์ สาขาวิชาวิศวกรรมโยธา เป็นอันดับ 1 ในมหาวิทยาลัยนเรศวร ระดับปริญญาตรี ประเภทรับตรง ประจำปีการศึกษา 2544 จำนวน 154 คน ผลการวิจัยพบว่า 1) ลำดับที่ที่เกิดจากการรวมคะแนนมาตรฐานที่เชิงเส้น ในโรงเรียนขนาดใหญ่ ขนาดกลางและขนาดเล็ก จำนวน 65, 30, และ 28 คน คิดเป็นร้อยละ 84.42, 73.17 และ 77.78 ตามลำดับ ถ้า

นำผู้เข้าสอบทั้ง 154 คน จากโรงเรียนทุกขนาดมาเรียงลำดับคะแนนใหม่โดยไม่คำนึงถึงขนาดโรงเรียนพบว่า มีผู้ได้ลำดับที่เปลี่ยนแปลงไปจำนวน 136 คน คิดเป็นร้อยละ 88.31 2) ลำดับที่เกิดจากการรวมคะแนนมาตรฐานที่ปกติ ในโรงเรียนขนาดใหญ่ ขนาดกลางและขนาดเล็ก จำนวน 64, 33, และ 28 คน คิดเป็นร้อยละ 83.12, 80.49 และ 77.78 ตามลำดับ ถ้านำผู้เข้าสอบทั้ง 154 คน จากโรงเรียนทุกขนาด มาเรียงลำดับคะแนนใหม่โดยไม่คำนึงถึงขนาดโรงเรียน พบว่า มีผู้ได้ลำดับที่เปลี่ยนแปลงไปจำนวน 143 คน คิดเป็นร้อยละ 92.86 3) การวิเคราะห์ความแตกต่างของลำดับที่ได้จากการรวมคะแนนดิบกับการรวมคะแนนโดยแปลงเป็นคะแนนมาตรฐานที่เชิงเส้นและคะแนนมาตรฐานที่ปกติของผู้เข้าสอบจากโรงเรียนแต่ละขนาด และจากโรงเรียนทุกขนาดคละกัน โดยใช้ Wilcoxon Matched Pairs Signed-Ranks Test พบว่า ลำดับที่ของผลการสอบคัดเลือกเปลี่ยนแปลงไปอย่างไม่มีนัยสำคัญ 4) การหาความสัมพันธ์ระหว่างลำดับที่ได้จากการรวมคะแนนดิบ กับการรวมคะแนนโดยแปลงเป็นคะแนนมาตรฐานที่เชิงเส้น และคะแนนมาตรฐานที่ปกติของผู้เข้าสอบ 154 คน จากโรงเรียนทุกขนาดคละกัน พบว่า มีความสัมพันธ์กัน เป็น .991 และ .984 ตามลำดับ 5) การหาความสัมพันธ์ระหว่างคะแนนผลการเรียนเฉลี่ยสะสม คะแนนรวมดิบ คะแนนรวมมาตรฐานที่เชิงเส้น และคะแนนรวมมาตรฐานที่ปกติ พบว่า คะแนนรวมดิบกับคะแนนรวมมาตรฐานที่เชิงเส้นและคะแนนรวมดิบกับคะแนนรวมมาตรฐานที่ปกติ ความสัมพันธ์กันในระดับสูง เป็น .992 และ .983 ตามลำดับ ขณะที่คะแนนผลการเรียนเฉลี่ยสะสม (GPA) มีความสัมพันธ์กับคะแนนรวมดิบ คะแนนรวมมาตรฐานที่เชิงเส้นและคะแนนรวมมาตรฐานที่ปกติ ในระดับปานกลาง เป็น .753, .705 และ .664 ตามลำดับ

ศุวิมล ว่องวานิช และคณะ (2546) ได้ศึกษาการวิเคราะห์ความเหมาะสมขององค์ประกอบที่ใช้ในการคัดเลือกนักศึกษาเข้ามหาวิทยาลัยตามระบบเอ็นทรานซ์ ซึ่งเป็นระบบที่ใช้ในช่วงปี 2543-2545 พบว่า คะแนนเฉลี่ยสะสมและคะแนนเปอร์เซ็นต์ไทล์ มีความเหมาะสมสัมพันธ์มากกว่า 0.8 คะแนนเฉลี่ยสะสมมัธยมปลายและคะแนนรวมเอ็นทรานซ์ มีความสัมพันธ์กันมากกว่าคะแนนเปอร์เซ็นต์ไทล์กับคะแนนรวมเอ็นทรานซ์ การใช้ผลการเรียนสะสมและคะแนนเปอร์เซ็นต์ไทล์ เป็นองค์ประกอบอย่างละร้อยละ 5 ทำให้เกิดการเปลี่ยนแปลงของผลการตัดสินใจประมาณ ร้อยละ 3-5 ใน 8 กลุ่มสาขาวิชา องค์ประกอบของ GPA, PR และคะแนนเอ็นทรานซ์วิชาต่างๆ สามารถอธิบายความแปรปรวนได้ร้อยละ 60-80 ความสามารถของผู้สมัครสอบประกอบด้วยองค์ประกอบ ประมาณ 1-3 องค์ประกอบ ขึ้นอยู่กับกลุ่มสาขาวิชาหรือวิชาที่ใช้ในการทดสอบ เช่น ความสามารถของผู้สมัครสอบในกลุ่มสาขาวิชามนุษย์และสังคมส่วนใหญ่ประกอบด้วย 2 องค์ประกอบ กลุ่มสาขาวิชาวิทยาศาสตร์หรือสาขาอื่นๆที่มีการสอบวิชาเฉพาะ มี 3 องค์ประกอบ ข้อค้นพบนี้ แสดงว่า การกำหนดองค์ประกอบแต่ละสาขาวิชา หรือคณะวิชาไม่จำเป็นต้องเหมือนกัน อาจจะเป็น 2 หรือ 3 องค์ประกอบแล้วแต่สาขาวิชา คะแนน GPA หรือ PR จะแยกเป็น 1 องค์ประกอบย่อย ภาษาไทย และสังคมจะ

รวมตัวอยู่ภายใต้องค์ประกอบเดียวกัน วิชาหลักอื่นๆ เป็นส่วนใหญ่ โดยเฉพาะในกลุ่มสาขามนุษย์และสังคม แสดงว่าภาษาไทย สังคมและวิชาหลักที่สาขามนุษย์และสังคมสอบนั้นต่างวัดคุณลักษณะแฝงร่วมกันในกลุ่มสาขามนุษย์และสังคม สามารถอธิบายได้จากวิชาที่สอบสูงกว่ากลุ่มสาขาวิชาวิทยาศาสตร์ ในกลุ่มสาขาวิศวกรรมศาสตร์ และสถาปัตยกรรมศาสตร์ วิชาภาษาไทยและสังคมศึกษา เป็นองค์ประกอบที่แตกต่างหากจากวิชาหลักอื่นๆ ผลการวิจัยของสุวิมล ว่องวานิช และคณะ ได้ให้ข้อเสนอแนะเพื่อการศึกษาวิจัยต่อว่า 1) การวิจัยนี้ทราบ่วิชาใดสัมพันธ์กับคะแนนรวมสูง แต่ไม่ทราบว่าองค์ประกอบใดเหมาะสมหรือไม่ ต้องติดตามผลการเรียนในมหาวิทยาลัยและหาความตรงเชิงพยากรณ์ต่อไป 2) วิชาต่างๆ สามารถอธิบายความสามารถของผู้สมัคร ได้ร้อยละ 50 – 60 และเมื่อใช้ GPA และ PR สามารถอธิบายความสามารถของผู้สมัคร ได้อีกประมาณร้อยละ 10 แต่ผลการวิจัยนี้ไม่สามารถระบุได้ว่า GPA, PR เหมาะสมหรือไม่ จนกว่าจะศึกษาความตรงเชิงพยากรณ์ก่อน 3) ระบบแอดมิชชัน เป็นระบบที่มีแนวคิดที่จะนำความสามารถทางวิชาการเฉพาะศาสตร์ ความรู้ความสามารถตามหลักสูตร ความถนัดทางการเรียนและคุณธรรมของผู้สมัครมาใช้ด้วยนั้น การให้นำหนักความสำคัญขององค์ประกอบใดต้องการผลวิจัยนำร่องก่อนการตัดสินใจ

มานะ กวางทอง (2546) ได้ศึกษาการเปรียบเทียบผลการแปลงเกรดเฉลี่ยเป็นคะแนนการสอบคัดเลือกเข้าศึกษาในสถาบันอุดมศึกษาที่ใช้อัตราส่วนระหว่างเกรดเฉลี่ยกับตำแหน่งเปอร์เซ็นต์ไทล์ต่างกัน เพื่อเปรียบเทียบผลการแปลงเกรดเฉลี่ยช่วงเดียวกันเป็นคะแนนการสอบคัดเลือก ของโรงเรียนขนาดต่างกัน เมื่อใช้อัตราส่วนระหว่างเกรดเฉลี่ยกับตำแหน่งเปอร์เซ็นต์ไทล์เดียวกัน โดยใช้ F-test และ เปรียบเทียบผลการแปลงเกรดเฉลี่ยช่วงเดียวกันเป็นคะแนนการสอบคัดเลือกของโรงเรียน รวมทุกขนาด เมื่อใช้อัตราส่วนระหว่างเกรดเฉลี่ยกับตำแหน่งเปอร์เซ็นต์ไทล์ต่างกัน โดยใช้ F-test แบบวัดซ้ำ กลุ่มตัวอย่างเป็นนักเรียนที่สำเร็จชั้นมัธยมศึกษาปีที่ 6 ภาคปลาย ปีการศึกษา 2543 ในจังหวัดพิษณุโลก จำนวน 2,444 คน ผลการวิจัยพบว่า 1) เปรียบเทียบผลการแปลงเกรดเฉลี่ยช่วงเดียวกันเป็นคะแนนการสอบคัดเลือกของโรงเรียนขนาดต่างกัน เมื่อใช้อัตราส่วนระหว่างเกรดเฉลี่ยกับตำแหน่งเปอร์เซ็นต์ไทล์เดียวกัน 1.1) สำหรับผู้ที่ได้เกรดเฉลี่ย 2.00-2.50, 2.51-3.00 และ 3.01-3.50 พบว่า โรงเรียนขนาดใหญ่พิเศษมีค่าเฉลี่ยของผลการแปลงเกรดเฉลี่ยเป็นคะแนนการสอบคัดเลือกต่ำกว่าโรงเรียนขนาดใหญ่ โรงเรียนขนาดกลางและโรงเรียนขนาดเล็ก 1.2) สำหรับผู้ที่ได้เกรดเฉลี่ย 3.51-4.00 พบว่า โรงเรียนขนาดต่างกันมีค่าเฉลี่ยของผลการแปลงเกรดเฉลี่ยเป็นคะแนนการสอบคัดเลือกไม่แตกต่างกัน 2) การเปรียบเทียบผลการแปลงเกรดเฉลี่ยช่วงเดียวกันเป็นคะแนนการสอบคัดเลือกของโรงเรียนรวมทุกขนาด เมื่อใช้อัตราส่วนระหว่างเกรดเฉลี่ยกับตำแหน่งเปอร์เซ็นต์ไทล์ต่างกัน 2.1) สำหรับผู้ที่ได้เกรดเฉลี่ย 2.00-2.50 และ 2.51-3.00 พบว่า ค่าเฉลี่ยของผลการแปลงเกรดเฉลี่ยเป็นคะแนนการสอบคัดเลือกของโรงเรียนรวมทุกขนาด มีค่าสูงขึ้นตามอัตราส่วนที่เพิ่มขึ้น

ของตำแหน่งเปอร์เซ็นต์ (2.2) สำหรับผู้ที่ได้เกรดเฉลี่ย 3.01-3.50 และ 3.51-4.00 พบว่า โรงเรียนรวมทุกขนาด เมื่อใช้อัตราส่วนระหว่างเกรดเฉลี่ยกับตำแหน่งเปอร์เซ็นต์ต่างกัน มีค่าเฉลี่ยของผลการแปลงเกรดเฉลี่ยเป็นคะแนนการสอบคัดเลือกไม่แตกต่างกัน

จิตตานันท์ ตีกุล (2548) ได้ศึกษาความสัมพันธ์ระหว่างผลการเรียนระดับมัธยมศึกษาตอนปลาย ผลการเรียนระดับอุดมศึกษา และคะแนนสอบ คัดเลือกเข้าศึกษาในสถาบันอุดมศึกษาของนักศึกษามหาวิทยาลัยเทคโนโลยีสุรนารี รุ่นปีการศึกษา 2540 - 2543 เพื่อศึกษาและเปรียบเทียบผลการเรียนระดับมัธยมศึกษาตอนปลาย ระดับอุดมศึกษา คะแนนคัดเลือกเข้าศึกษาในสถาบันอุดมศึกษา และเพื่อศึกษาความสัมพันธ์ระหว่างผลการเรียนระดับมัธยมศึกษาตอนปลายกับผลการเรียนระดับอุดมศึกษา ความสัมพันธ์ระหว่างผลการเรียนระดับมัธยมศึกษาตอนปลายกับคะแนนสอบคัดเลือกเข้าศึกษาในสถาบันอุดมศึกษา และความสัมพันธ์ระหว่างผลการเรียนระดับอุดมศึกษากับคะแนนสอบคัดเลือกเข้าศึกษาในสถาบันอุดมศึกษา กลุ่มตัวอย่างในการวิจัยครั้งนี้คือ นักศึกษามหาวิทยาลัยเทคโนโลยีสุรนารีที่ขึ้นทะเบียนเป็นนักศึกษาใหม่ รุ่นปีการศึกษา 2540 - 2543 และสำเร็จการศึกษาตามข้อบังคับมหาวิทยาลัยเทคโนโลยีสุรนารี ว่าด้วยการศึกษาขั้นปริญญาตรี พ.ศ. 2536 และ 2541 นับถึงเมื่อสิ้นภาคการศึกษาที่ 1 ปีการศึกษา 2547 ศึกษาจากข้อมูลทุติยภูมิของศูนย์บริการการศึกษา วิเคราะห์ข้อมูลด้วยสถิติเชิงบรรยาย การวิเคราะห์ความแปรปรวนทางเดียว โดยใช้อันดับที่ และการวิเคราะห์สหสัมพันธ์ Somers'd ผลการวิจัยพบว่า 1) ผลการเรียนระดับมัธยมศึกษาตอนปลายของนักศึกษาที่รับเข้าด้วยวิธี โควตาทุกรุ่นสูงกว่านักศึกษาที่รับเข้าด้วยวิธีการสอบคัดเลือกอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 สำหรับผลการเรียนระดับมัธยมศึกษาตอนปลายของนักศึกษาที่จบจากโรงเรียนขนาดกลางและขนาดใหญ่ทุกรุ่นสูงกว่าโรงเรียนขนาดใหญ่พิเศษอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 2) ผลการเรียนระดับอุดมศึกษาของนักศึกษาที่รับเข้าด้วยวิธี โควตาสูงกว่านักศึกษาที่รับเข้าด้วยวิธีการสอบคัดเลือกอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 เฉพาะรุ่นปีการศึกษา 2541 2542 และ 2543 ส่วนผลการเรียนระดับอุดมศึกษาของนักศึกษาที่จบจากโรงเรียนทั้งสี่ขนาดทุกรุ่นมีผลการเรียนระดับอุดมศึกษาไม่แตกต่างกัน 3) คะแนนสอบคัดเลือกเข้าศึกษาในสถาบันอุดมศึกษาของนักศึกษาที่รับเข้าด้วยวิธีสอบคัดเลือกทุกรุ่นส่วนใหญ่ต่ำกว่าค่าเฉลี่ยอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ยกเว้นนักศึกษาที่สอบเข้ากลุ่มสาขาวิชาเทคโนโลยีสารสนเทศ (วิทย์) ทุกรุ่น กลุ่มสาขาวิชาเทคโนโลยีสารสนเทศ (ศิลป์คำนวณ) รุ่นปีการศึกษา 2541 และ 2543 และกลุ่มสาขาวิชาสาธารณสุขศาสตร์รุ่นปีการศึกษา 2542 ที่คะแนนสอบเข้าศึกษาในสถาบันอุดมศึกษาใกล้เคียงกับค่าเฉลี่ย 4) ผลการเรียนระดับมัธยมศึกษาตอนปลายสัมพันธ์กับผลการเรียนระดับอุดมศึกษาอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 เฉพาะนักศึกษาที่รับเข้าด้วยวิธีสอบคัดเลือก รุ่นปีการศึกษา 2541, 2542 และ 2543 นักศึกษาที่รับเข้าด้วยวิธี โควตาจังหวัด รุ่นปีการศึกษา 2541

และ 2542 นักศึกษาที่จบจากโรงเรียนขนาดกลางรุ่นปีการศึกษา 2541 และ 2542 และนักศึกษาที่จบจากโรงเรียนขนาดใหญ่และขนาดใหญ่พิเศษรุ่นปีการศึกษา 2541, 2542 และ 2543 โดยความสัมพันธ์อยู่ในระดับต่ำถึงค่อนข้างต่ำ 5) ผลการเรียนระดับมัธยมศึกษาตอนปลายสัมพันธ์กับคะแนนสอบคัดเลือกเข้าศึกษาในสถาบันอุดมศึกษาอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 เพียงบางวิชา ในรุ่นปีการศึกษา 2541, 2542 และ 2543 โดยความสัมพันธ์เกือบทั้งหมดอยู่ในระดับค่อนข้างต่ำ 6) ผลการเรียนระดับอุดมศึกษาสัมพันธ์กับคะแนนสอบคัดเลือกเข้าศึกษาในสถาบันอุดมศึกษาอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 เพียงบางวิชาและบางกลุ่มสาขาวิชาที่สอบเข้า ในรุ่นปีการศึกษา 2541, 2542 และ 2543 โดยความสัมพันธ์เกือบทั้งหมดอยู่ในระดับค่อนข้างต่ำ

ปิยะธิดา ปัญญา (2551) ได้พัฒนาวิธีการเปรียบเทียบผลการเรียนโดยใช้หลักการเชื่อมโยงเชิงเส้น และตรวจสอบคุณภาพของวิธีการเปรียบเทียบที่พัฒนาขึ้น 3 วิธี ได้แก่ 1) วิธีการเปรียบเทียบผลการเรียนแบบเชื่อมโยงส่วนเดียว 2) วิธีการเปรียบเทียบผลการเรียนแบบเชื่อมโยงเสริมกันทั้งสองส่วน และ 3) วิธีการเปรียบเทียบผลการเรียนแบบเชื่อมโยงร่วมกันทั้งสองส่วน โดยพิจารณาค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างผลการเรียนที่เปรียบเทียบแล้วกับคะแนน O-NET และค่าดัชนีความแตกต่าง (Discrepancy Index) ของผลการเรียนที่เปรียบเทียบแล้วของแต่ละวิธีกับผลการเรียนเกณฑ์ ข้อมูลที่ใช้เป็นผลทดสอบในระบบกลางการรับบุคคลเข้าศึกษาในสถาบันอุดมศึกษา (Admissions) ประจำปีการศึกษา 2550 ซึ่งประกอบด้วยผลการเรียนเฉลี่ยสะสมตลอดหลักสูตร (GPAX) ผลการเรียนเฉลี่ยสะสมตามกลุ่มสาระการเรียนรู้ (GPA) และผลทดสอบทางการศึกษาแห่งชาติขั้นพื้นฐาน (O-NET) ผลการวิจัยปรากฏว่า 1) ผลการเรียนที่เปรียบเทียบด้วยวิธีการเชื่อมโยงส่วนเดียว วิธีการเชื่อมโยงเสริมกันทั้งสองส่วน และวิธีการเชื่อมโยงร่วมกันทั้งสองส่วนที่ผู้วิจัยพัฒนาขึ้น มีความสัมพันธ์กับคะแนน O-NET สูงขึ้นตามลำดับ โดยมีค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างผลการเรียนที่เปรียบเทียบแล้วกับคะแนน O-NET สูงกว่าสัมประสิทธิ์สหสัมพันธ์ระหว่างผลการเรียนเดิมกับคะแนน O-NET อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 2) ค่าดัชนีความแตกต่างระหว่างผลการเรียนเกณฑ์กับผลการเรียนที่เปรียบเทียบโดยวิธีการเชื่อมโยงร่วมกันทั้งสองส่วนมีค่าต่ำสุด รองลงมาเป็นผลการเรียนที่เปรียบเทียบด้วยวิธีการเชื่อมโยงเสริมกันทั้งสองส่วน และวิธีการเชื่อมโยงส่วนเดียว ตามลำดับ ทั้งในส่วนของคุณค่าของดัชนีรากที่สองของผลต่างกำลังสองเฉลี่ย (RMSD) และดัชนีค่าเฉลี่ยผลต่างสัมบูรณ์ (MAD) โดยค่าดัชนีความแตกต่างระหว่างผลการเรียนเกณฑ์กับผลการเรียนที่เปรียบเทียบแล้วต่ำกว่าค่าดัชนีความแตกต่างระหว่างผลการเรียนเกณฑ์กับผลการเรียนเดิมอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ทุกวิธี

ศิริชัย กาญจนวาที และคณะ (Kanjanaavasee and others. 2009 : 117-124) ได้พัฒนาวิธีการเปรียบเทียบผลการเรียนเฉลี่ยสะสมตลอดหลักสูตรมัธยมศึกษาตอนปลาย (GPA) กับคะแนนทดสอบทางการศึกษาแห่งชาติขั้นพื้นฐาน (O-NET) โดยใช้วิธีการเปรียบเทียบ 4 วิธี ได้แก่ วิธีการวิเคราะห์การถดถอย วิธีเชิงเส้นตรง วิธีอิกวิเปอร์เซนไทล์ และวิธีวิเคราะห์พหุระดับ กลุ่มตัวอย่างเป็นโรงเรียนมัธยมศึกษาตอนปลาย ในปีการศึกษา 2549 จำนวน 2,583 โรงเรียน ผลการวิจัยพบว่า 1) ระหว่างวิธีการเปรียบเทียบทั้ง 4 วิธี วิธีเชิงเส้นตรงและวิธีอิกวิเปอร์เซนไทล์ ให้ความถูกต้องและกลมกลืนกับข้อมูลเชิงประจักษ์มากกว่าอีกสองวิธี ผลจากการเปรียบเทียบด้วยวิธีเชิงเส้นตรงและวิธีอิกวิเปอร์เซนไทล์ สะท้อนให้เห็นว่า โรงเรียนที่มีคุณภาพสูงมีแนวโน้มที่จะกวดเกรดมากกว่าปล่อยเกรด ในขณะที่โรงเรียนที่มีคุณภาพปานกลางจะมีทั้งการกวดเกรดและปล่อยเกรด และโรงเรียนที่มีคุณภาพต่ำส่วนใหญ่จะปล่อยเกรด วิธีเชิงเส้นตรงและวิธีอิกวิเปอร์เซนไทล์มีความเหมาะสมมากที่สุดในการเปรียบเทียบผลการเรียนเฉลี่ยสะสม (GPA) โดยวิธีเชิงเส้นตรงนั้นจะง่ายและไม่ยุ่งยากในการนำไปใช้ 2) มีความสัมพันธ์กันสูงระหว่างผลการเรียนเฉลี่ยสะสม ตามกลุ่มสาระการเรียนรู้ กับผลการเรียนเฉลี่ยสะสมรวมทุกกลุ่มสาระการเรียนรู้ แสดงให้เห็นว่า ผลการเรียนเฉลี่ยสะสมทั้งสองชนิดสามารถนำมาใช้เปรียบเทียบได้ แต่ผลการเรียนเฉลี่ยสะสมรวมทุกกลุ่มสาระการเรียนรู้ (GPAX) จะเหมาะสมมากกว่า 3) ผลการเปรียบเทียบลักษณะของผลการเรียนทั้งที่ผ่านการเปรียบเทียบและยังไม่ผ่านการเปรียบเทียบ แสดงให้เห็นว่า โรงเรียนส่วนใหญ่ให้เกรดสูงกว่า หรือต่ำกว่าความสามารถที่แท้จริงของผู้เรียน (47.43% และ 47.81%) มีโรงเรียนจำนวนน้อยมากที่ให้เกรดสอดคล้องกับความสามารถของผู้เรียน (4.76%) 4) ในรายกลุ่มสาระการเรียนรู้ทั้ง 5 กลุ่มสาระ (ภาษาไทย สังคมศึกษา ภาษาอังกฤษ คณิตศาสตร์ และวิทยาศาสตร์) กลุ่มสาระสังคมศึกษา ให้เกรดสูงกว่าความสามารถของผู้เรียนร้อยละ 49.67 ในขณะที่กลุ่มสาระการเรียนรู้ภาษาไทย ให้เกรดต่ำกว่าความสามารถของผู้เรียนร้อยละ 48.20 ส่วนกลุ่มสาระการเรียนรู้ภาษาอังกฤษให้เกรดสอดคล้องกับความสามารถของผู้เรียนร้อยละ 5.38 5) ผลการจัดกลุ่มโรงเรียนตามมาตรฐานของผลการเรียนรู้อ่อนและหลังเปรียบเทียบผลการเรียนแตกต่างกันอย่างมีนัยสำคัญ โดยก่อนการเปรียบเทียบผลการเรียน โรงเรียนส่วนใหญ่อยู่ในระดับดี (49.3%) ในขณะที่หลังเปรียบเทียบผลการเรียน โรงเรียนส่วนใหญ่อยู่ในระดับปานกลาง (60.8%)

จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องจะพบว่า มีวิธีการเปรียบเทียบคะแนนมาจากหลากหลายแนวคิด และหลากหลายวิธี จุดมุ่งหมายในการเปรียบเทียบก็แตกต่างกัน ส่วนในการเปรียบเทียบผลการเรียนระหว่างโรงเรียนนั้น ปิยะธิดา ปัญญา (2551) ได้พัฒนาวิธีการเปรียบเทียบผลการเรียนขึ้นมา 3 วิธี โดยเป็นการเชื่อมโยงเชิงเส้น ซึ่งเป็นวิธีการที่ค่อนข้างซับซ้อน ผู้วิจัยจึงมีแนวคิดที่จะพัฒนาวิธีการเปรียบเทียบผลการเรียน โดยอาศัยหลักการปรับสเกลที่ไม่ยุ่งยากและง่ายกว่าขึ้น