

บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

ในการทำวิจัยเรื่อง การปรับแก้มเฉลี่ยสะสมระดับมัธยมศึกษาตอนปลายด้วยคะแนนความถนัดทางการเรียน : การเปรียบเทียบระหว่างวิธีการปรับเทียบอิกวิเปอร์เซ็นโตล เชิงเส้นตรงและไออาร์ที นั้น ผู้วิจัยได้ศึกษาแนวความคิดและงานวิจัยที่เกี่ยวข้อง และได้นำเสนอผลของการค้นคว้า โดยแบ่งเนื้อหาออกเป็น 3 ตอนดังนี้

- ตอนที่ 1 การคัดเลือกเข้าศึกษาในสถาบันอุดมศึกษาของรัฐ ระบบใหม่
- ตอนที่ 2 กรอบทฤษฎีการประมวลผลทางปัญญา, ความถนัดทางการเรียน, ผลสัมฤทธิ์ทางการเรียน และงานวิจัยที่เกี่ยวข้อง
- ตอนที่ 3 การปรับเทียบคะแนนระหว่างแบบสอบและงานวิจัยที่เกี่ยวข้อง

ตอนที่ 1 การคัดเลือกบุคคลเข้าศึกษาในสถาบันอุดมศึกษาของรัฐ ระบบใหม่

ทบวงมหาวิทยาลัยโดยความเห็นชอบของอธิการบดีทุกมหาวิทยาลัยและสถาบันในสังกัด และกำกับของทบวงมหาวิทยาลัยได้กำหนดเกณฑ์การคัดเลือกบุคคลเข้าศึกษาในสถาบันอุดมศึกษา ตั้งแต่ปีการศึกษา 2542 ดังรายละเอียดต่อไปนี้

1. วัตถุประสงค์

การคัดเลือกบุคคลเข้าศึกษาในสถาบันอุดมศึกษาระบบใหม่ (สำนักทดสอบกลาง, 2541) มีวัตถุประสงค์หลัก 2 ประการ คือ

- 1.1 เพื่อให้มหาวิทยาลัย/สถาบัน ได้ผู้เรียนที่มีความรู้ ความสามารถ และความถนัดตรงตามสาขาวิชาที่เรียน
- 1.2 เพื่อส่งเสริมให้การเรียนการสอนในระดับมัธยมศึกษาตอนปลายเป็นไปตามปรัชญาและวัตถุประสงค์ของหลักสูตร

2. องค์ประกอบของการคัดเลือก

การคัดเลือกบุคคลเข้าศึกษาในสถาบันอุดมศึกษาระบบใหม่ จะพิจารณาคัดเลือกผู้สมัครจากองค์ประกอบดังต่อไปนี้ (สำนักทดสอบกลาง, 2541)

2.1 ผลการเรียนตลอดหลักสูตรระดับมัธยมศึกษาตอนปลายหรือเทียบเท่า ที่นักเรียนแต่ละคนได้รับจากสถานศึกษาที่สำเร็จการศึกษา ให้ค่าน้ำหนักร้อยละ 10

2.2 ผลการสอบวิชาความรู้พื้นฐานวิชาการ (วิชาหลัก) และวิชาพื้นฐานเฉพาะวิชาชีพ (วิชาเฉพาะ) ให้ค่าน้ำหนักรวมกันร้อยละ 90

ในกรณีคณะ/ประเภทวิชาที่มีการสอบวิชาเฉพาะด้วยการกำหนดสัดส่วนค่าน้ำหนักระหว่างคะแนนวิชาหลัก และคะแนนวิชาเฉพาะให้เป็นไปตามที่แต่ละมหาวิทยาลัย/สถาบันกำหนด

2.3 ผลการสอบสัมภาษณ์และตรวจร่างกาย มหาวิทยาลัย/สถาบันจะทำการสอบสัมภาษณ์ และตรวจร่างกายเพื่อหาข้อมูลประกอบการพิจารณาความพร้อมและความเหมาะสมเป็นขั้นตอนสุดท้ายก่อนการรับเข้าศึกษา โดยไม่คิดเป็นค่าน้ำหนักคะแนน

ค่าน้ำหนักคะแนนของแต่ละองค์ประกอบ มีรายละเอียดดังตารางที่ 1

ตารางที่ 1 น้ำหนักคะแนนขององค์ประกอบการศึกษาในสถาบันอุดมศึกษา

องค์ประกอบ	น้ำหนักคะแนน	
	คณะ/ประเภทวิชาที่มีการสอบวิชาพื้นฐานวิชาชีพ	คณะ/ประเภทวิชาที่ไม่มีการสอบวิชาพื้นฐานวิชาชีพ
1. ผลการเรียนระดับมัธยมศึกษาตอนปลาย	10 %	10 %
2. ผลการสอบวิชาความรู้พื้นฐานวิชาการ (วิชาหลัก)	} 90 %	90 %
3. ผลการสอบวิชาพื้นฐานวิชาชีพ (วิชาเฉพาะ)		-
4. ผลการสอบสัมภาษณ์และตรวจร่างกาย		-

3. วิธีการนำแต่ละองค์ประกอบมาใช้

ในการพิจารณาคัดเลือกนิสิตนักศึกษาจะนำองค์ประกอบแต่ละอย่างมาใช้อย่างนี้

1) ผลการเรียนระดับมัธยมศึกษาตอนปลายหรือเทียบเท่า จะพิจารณาจากผลการเรียนตลอดหลักสูตรระดับมัธยมศึกษาตอนปลายของนักเรียนแต่ละคนที่ได้รับจากโรงเรียน โดยโรงเรียนจะออกเอกสารแสดงผลการเรียนเฉลี่ยตลอดหลักสูตรให้แก่ นักเรียนแต่ละคน พร้อมกับระเบียบแสดงผลการเรียน(รบ.1-ป) เพื่อให้ นักเรียนนำไปใช้สมัครเข้าเรียนต่อในระดับอุดมศึกษา ในเอกสารนี้จะแสดงข้อมูลเกี่ยวกับผลการเรียนเฉลี่ยและหน้าที่ความรับผิดชอบ และ/หรือความสามารถพิเศษของนักเรียน รวมทั้งค่าสถิติต่าง ๆ

ผลการเรียนระดับมัธยมศึกษาตอนปลายหรือเทียบเท่าที่นำมาใช้กับการคัดเลือกเข้าศึกษาในสถาบันอุดมศึกษาระบบใหม่ ในระยะแรกทบวงมหาวิทยาลัยและสถาบันอุดมศึกษาได้เห็นพ้องต้องกันว่าจะใช้องค์ประกอบของผลการเรียน โดยแบ่งเป็น 2 ส่วน คือ

องค์ประกอบที่ 1 ความสามารถของนักเรียนซึ่งวัดได้จากคะแนนเฉลี่ยสะสมตลอดหลักสูตรมัธยมศึกษาตอนปลายหรือเทียบเท่า (GPAX) โดยกำหนดค่าน้ำหนักคะแนนในส่วนนี้เป็นร้อยละ 5

ในการนำคะแนนเฉลี่ยสะสมตลอดหลักสูตร (GPAX) มาใช้นั้น จะทำการแปลงคะแนน GPAX ของนักเรียนทุกคนที่สมัครเข้ารับการคัดเลือกเข้าศึกษา ให้เป็นคะแนนในสเกลของเส้นโค้งปรกติมาตรฐานเดียวกัน (Standard Normal Curve) แล้วจึงนำไปคำนวณเป็นคะแนนในค่าน้ำหนักร้อยละ 5 ของคะแนนทั้งหมดที่นำมาใช้ในการตัดสินใจคัดเลือกนักศึกษาของแต่ละคณะ/ประเภทวิชา

องค์ประกอบที่ 2 ความสามารถของนักเรียนเมื่อเทียบกับกลุ่มในแต่ละโรงเรียน ซึ่งวัดได้จากค่าเปอร์เซ็นต์ไทล์ (Percentile Rank) กำหนดค่าน้ำหนักคะแนนในส่วนนี้เป็นร้อยละ 5

ในการนำค่าเปอร์เซ็นต์ไทล์ของนักเรียนแต่ละคนมาใช้นั้น จะกำหนดค่าเปอร์เซ็นต์ไทล์เป็นช่วง (Range) แล้วกำหนดคะแนนแต่ละช่วงเป็นสัดส่วนจากคะแนนเต็มร้อยละ 5 ของคะแนนทั้งหมดที่นำมาใช้ในการตัดสินใจคัดเลือกนักศึกษาของแต่ละคณะ/ประเภทวิชา

จากนั้นจะนำคะแนนทั้ง 2 องค์ประกอบ ซึ่งมีค่าน้ำหนักร้อยละ 10 ของคะแนนเต็ม ไปรวมกับคะแนนสอบวิชาความรู้พื้นฐานวิชาการ (วิชาหลัก) และคะแนนสอบวิชาพื้นฐานวิชาชีพ (วิชาเฉพาะ) ของผู้สมัครแต่ละคน

2) ผลการสอบวิชาหลักและวิชาเฉพาะสาขา

2.1) จำนวนวิชาที่จัดสอบ

ทบวงมหาวิทยาลัยจะจัดสอบวิชาหลักและวิชาเฉพาะสาขา ซึ่งเป็นความรู้พื้นฐานที่แต่ละคณะ/ประเภทวิชากำหนดให้สอบ เพื่อวัดพื้นฐานความรู้ของผู้ประสงค์จะเข้าศึกษาในคณะ/ประเภทวิชานั้น ๆ

วิชาหลักที่จะจัดสอบมีทั้งหมด 15 วิชา แบ่งเป็น 2 ประเภท คือ

ประเภทที่ 1 วิชาที่ผู้ประสงค์จะเข้าศึกษาในสถาบันอุดมศึกษาทุกคน จะต้องสอบจำนวน 2 วิชา ได้แก่ ภาษาไทยและสังคมศึกษา ผลของการสอบวิชาทั้งสองได้กำหนดให้นำไปใช้ใน 2 ลักษณะ คือ

- นำคะแนนไปรวมกับคะแนนวิชาอื่นที่คณะ/ประเภทวิชานั้นกำหนดให้สอบ เพื่อตัดสินผลการคัดเลือกนิสิตนักศึกษาเช่นเดียวกับระบบปัจจุบัน

- ใช้ผลการสอบเป็นคุณสมบัติในการสมัครเข้าศึกษาในคณะ/ประเภทวิชา โดยกำหนดให้คะแนนทั้งสองวิชารวมกันไม่ต่ำกว่าร้อยละ 30 แต่ไม่นำคะแนนไปรวมกับคะแนนวิชาอื่น

ประเภทที่ 2 วิชาที่เลือกอีก 13 วิชา เป็นวิชาที่แต่ละคณะ/ประเภทวิชา จะเลือกกำหนดให้ผู้ประสงค์จะเข้าศึกษาในคณะ/ประเภทวิชานั้น ๆ สอบ

ทั้งนี้ข้อสอบแต่ละรายวิชาจะครอบคลุมเนื้อหาความรู้ของวิชานั้นทั้งหมดในระดับมัธยมศึกษาตอนปลาย ใช้เวลาสอบวิชาละ 2 ชั่วโมง

สำหรับวิชาเฉพาะสาขา ในระยะแรกเพื่อมิให้เกิดความสับสนและเพิ่มความเครียดให้แก่นักเรียน จะจัดสอบวิชาเฉพาะสาขา จำนวน 27 วัน เท่ากับที่สอบในปัจจุบัน ส่วนจำนวนรายวิชาเฉพาะสาขาที่นักเรียนแต่ละคนจะต้องสอบ ขึ้นอยู่กับการกำหนดของคณะ/ประเภทวิชาที่จะเข้าศึกษา

2.2) กำหนดการสอบ

ทบวงมหาวิทยาลัยกำหนดจัดสอบวิชาความรู้พื้นฐานวิชาการ (วิชาหลัก) ปีละ 2 ครั้ง ครั้งละ 3 วัน ในเดือนมีนาคมและตุลาคมของทุกปี และจัดสอบวิชาเฉพาะสาขา ปีละ 1 ครั้ง ในเดือนตุลาคม โดยจะเริ่มสอบครั้งแรกในเดือนตุลาคม 2541 และจะแจ้งผลการสอบให้ผู้เข้าสอบทราบทุกครั้ง ภายหลังจากการสอบประมาณ 1 เดือน เพื่อให้ผู้เข้าสอบนำไปใช้ตรวจสอบคุณสมบัติและประกอบการตัดสินใจเลือกสมัครเข้าศึกษาในคณะ/ประเภทวิชาที่ประสงค์จะเข้าศึกษาต่อไป โดยผู้สมัครจะสอบกี่ครั้ง ครั้งละกี่วิชาก็ได้ และคะแนนที่สอบได้สามารถเก็บไว้ใช้ได้ 3 ปี

2.3) คุณสมบัติของผู้สมัครสอบ

ผู้มีสิทธิ์สมัครสอบวิชาหลักและวิชาเฉพาะสาขาจะต้องเป็นผู้ที่กำลังศึกษาอยู่ในชั้นมัธยมศึกษาปีที่ 6 หรือเทียบเท่า หรือเป็นผู้ที่สำเร็จการศึกษาชั้นมัธยมศึกษาตอนปลายหรือเทียบเท่า สำหรับนักเรียนในสังกัดกรมการศึกษานอกโรงเรียนจะต้องเป็นผู้ที่กำลังศึกษาอยู่ในภาคการศึกษาสุดท้ายของหลักสูตร

2.4) สถานที่จัดสอบ

ทบวงมหาวิทยาลัยมีนโยบายที่จะอำนวยความสะดวกให้กับผู้สมัครสอบให้มากที่สุด จึงได้กระจายสนามสอบไปทั่วประเทศ จำแนกเป็นสนามสอบในส่วนภูมิภาค 12 สนามสอบ และสนามสอบกรุงเทพมหานครอีก 1 สนามสอบ ในระยะแรกนี้ทบวงมหาวิทยาลัยจะได้พิจารณาร่วมกับสนามสอบภูมิภาคในการกำหนดจังหวัดที่มีความพร้อมและผู้สมัครสอบสามารถเดินทางไปสอบได้สะดวก สนามสอบละ 3-4 จังหวัด ให้เป็นหน่วยสอบของสนามสอบนั้น ๆ ซึ่งคาดว่าจะมีจำนวนประมาณ 40 จังหวัด ส่วนสนามสอบกรุงเทพมหานคร ได้แบ่งเป็นเขตการสอบ โดยมีมหาวิทยาลัย/สถาบัน ร่วมบริหารการสอบจำนวน 8 เขตการสอบ

สำหรับการสอบวิชาเฉพาะสาขา จะจัดสอบเฉพาะที่สนามสอบ (หน่วยสอบของมหาวิทยาลัย/สถาบัน) ในส่วนภูมิภาค 12 แห่ง และในกรุงเทพมหานคร 1 แห่ง

3) ผลการสอบสัมภาษณ์และการตรวจร่างกาย

มหาวิทยาลัย/สถาบัน จะทำการสอบสัมภาษณ์และตรวจร่างกายผู้ที่ได้รับการประกาศให้เป็นผู้มีสิทธิ์เข้ารับการสอบสัมภาษณ์และตรวจร่างกาย ผลการสอบสัมภาษณ์และตรวจร่างกาย จะนำมาใช้ 2 ลักษณะ คือ

3.1) เป็นข้อมูลเพื่อช่วยเหลือ ส่งเสริม และเตรียมความพร้อมให้แก่นิสิต นักศึกษา

3.2) เป็นข้อมูลเพื่อตรวจสอบคุณสมบัติของผู้สมัครเข้าศึกษาว่ามีคุณสมบัติครบถ้วนตามที่กำหนดหรือไม่ เพื่อใช้ประกอบการพิจารณาคัดเลือกนิสิตนักศึกษาครั้งสุดท้ายก่อนรับเข้าศึกษา

ตอนที่ 2 กรอบทฤษฎีการประมวลผลทางปัญญา ความถนัดทางการเรียน ผลสัมฤทธิ์ทางการเรียน และงานวิจัยที่เกี่ยวข้อง

1. กรอบทฤษฎีการประมวลผลทางปัญญา

กรอบทฤษฎีนี้เกิดจากบูรณาการของ 3 กลุ่มทฤษฎี คือ Psychometric , Piagetion และ Information Processing(กรมวิชาการ, 2537) ซึ่งเป็นกรอบทฤษฎีที่ใช้กันอย่างกว้างขวางในหมู่นักการศึกษาในปัจจุบัน กรอบทฤษฎีนี้จะอธิบายกระบวนการเรียนรู้หรือพัฒนาการทางสติปัญญาในรูปของการประมวลผลข้อมูล อันได้แก่ การจัดโครงสร้างข้อมูลเพื่อการจัดเก็บ (Organization) การเรียกใช้ (Retrieval) การสังเคราะห์อย่างต่อเนื่อง (Assembling and Reassembling) และการปรับและปรุงแต่ง (Tuning) เพื่อนำไปใช้ในการคิด การแก้ปัญหาและการเรียนรู้ต่อไป นอกจากนี้กรอบทฤษฎีการประมวลผลทางปัญญายังให้แนวสำหรับการวัดและประเมินประสิทธิภาพของโครงสร้างความรู้ กระบวนการและกรอบการวิเคราะห์ ในรูปของผลสัมฤทธิ์ทั่วไป ซึ่งการวัดผลสัมฤทธิ์โดยใช้กรอบทฤษฎีนี้จะบ่งชี้ถึงมาตรฐานการจัดการศึกษาและกระบวนการเรียนรู้ของผู้เรียนอีกด้วย

2. ความถนัดทางการเรียน

คำว่า "ความถนัด" นั้น มีนักวัดผลทางการศึกษาได้ให้นิยามไว้มากมาย ดังเช่น วอร์เรน (Warren, 1934) ได้ให้ความหมายของความถนัดว่า เป็นสภาวะหรือคุณลักษณะที่แสดงถึงความสามารถของบุคคลที่จะเรียนรู้ ซึ่งเกิดจากการฝึกฝนความรู้ ทักษะหรือมีปฏิกิริยาตอบสนองเฉพาะอย่าง ส่วน แครร์รอล (Carroll, 1963 อ้างถึงใน ล้วน สายยศและอังคณา สายยศ, 2527) กล่าวว่า ความถนัด หมายถึง ระยะเวลาที่ใช้เพื่อให้เกิดการเรียนรู้ กล่าวคือ ถ้าบุคคลใช้เวลาในการเรียนรู้มากแสดงว่ามีความถนัดน้อย แต่หากใช้เวลาในการเรียนรู้น้อยแสดงว่ามีความถนัดในเรื่องนั้นมาก นอกจากนี้ อนาสตาซี (Anastasi, 1990) ให้ความหมายไว้ว่า ความถนัดเป็นความสามารถของบุคคลที่วัดจากความรู้ที่สะสมในอดีต ซึ่งใช้ในการพยากรณ์ความสำเร็จในการเรียนรู้ ซึ่งสอดคล้องกับ สจิวร์ต (Stuart, 1995) ที่ให้ความหมายว่าเป็นทักษะที่เกิดจากการเรียนรู้ทั้งในและนอกชั้นเรียน ซึ่งสามารถทำนายความสามารถของแต่ละบุคคลได้ในความคิดของ คอร์ซินี (Corsini, 1994) นั้น ความถนัด หมายถึง ความเหมาะสมที่บุคคลจะทำงานอย่างใดอย่างหนึ่งตามความสามารถของตนเองที่มีอยู่

ส่วนในประเทศไทยได้มีนักการศึกษาที่สนใจเกี่ยวกับเรื่องนี้ คือ ชาวาล แพริตกุล (2516) โดยให้ความหมายไว้ว่า ความถนัด หมายถึง ชีตความสามารถสูงสุดของบุคคลในการเรียนรู้หรือฝึกฝนทักษะต่าง ๆ

จากที่กล่าวมาข้างต้นสามารถสรุปได้ว่า ความถนัด(Aptitude) หมายถึง ความสามารถที่บุคคลได้รับประสบการณ์ ฝึกฝนตนเอง และมีการสั่งสมไว้มากจนเกิดเป็นทักษะพิเศษเด่นชัดด้านใดด้านหนึ่ง พร้อมทั้งจะปฏิบัติกิจกรรมด้านนั้นได้อย่างดี

ความถนัดเป็นผลจากสิ่งแวดล้อม การฝึกอบรม ประสบการณ์เรียนรู้และทักษะ ซึ่งมีโอกาสเปลี่ยนแปลงได้ แต่ต้องอาศัยระยะเวลาพอสมควร ความถนัดสามารถนำไปใช้ในการพยากรณ์ผลภายหน้า ซึ่งอาจจะเป็นผลสัมฤทธิ์อย่างใดอย่างหนึ่งเช่น ผลสัมฤทธิ์ทางการเรียน ผลสัมฤทธิ์ในการทำงาน ผลสัมฤทธิ์ในการประกอบอาชีพ การใช้แบบทดสอบความถนัดมักใช้เป็นชุด ซึ่งประกอบด้วยฉบับย่อย ๆ ที่วัดองค์ประกอบความสามารถหรือความถนัดด้านต่าง ๆ เพื่อให้เกิดความแม่นยำในการพยากรณ์ เพราะผลสัมฤทธิ์ในกิจกรรมหนึ่งนั้นมิใช่มาจากองค์ประกอบเดียว แต่เป็นผลมาจากหลาย ๆ องค์ประกอบ

ในการวัดความถนัดสามารถจำแนกได้ 2 อย่าง คือ

1. ความถนัดทางการเรียน (Scholastic aptitude) เป็นการทดสอบความสามารถของบุคคลที่จะเรียนรู้สิ่งใหม่ในอนาคตได้หรือไม่ เป็นการวัดศักยภาพที่จะเรียนรู้ (potential to learn) และเป็นการใช้ผลการสอบทำนายความสำเร็จของการเรียนรู้ในอนาคต

2. ความถนัดพิเศษ (Special aptitude) เป็นการทดสอบความสามารถของบุคคลเฉพาะอย่าง เช่น ความสามารถทางเครื่องกล ความสามารถทางดนตรี ความสามารถทางศิลปะ เป็นต้น

ดังนั้น จากกล่าวได้ว่า ความถนัดทางการเรียน (scholastic aptitude) หมายถึง สมรรถภาพทางสมองที่เกิดจากการฝึกฝนมาตั้งแต่อดีต และเป็นปัจจัยสำคัญที่จะทำให้คนสามารถเรียนรู้และแก้ปัญหาใหม่ ๆ ได้

3. ผลสัมฤทธิ์ทางการเรียน (Academic Achievement)

ผลสัมฤทธิ์ (achievement) หมายถึง ผลสำเร็จที่เกิดจากการปฏิบัติงานอย่างใดอย่างหนึ่งที่อาศัยความพยายามทั้งร่างกายและสมอง (สำนักงานคณะกรรมการการศึกษาแห่งชาติ, 2527 อ้างถึงใน สุขุม มูลเมือง, 2539) นอกจากนั้น กรมวิชาการ (2537) ได้ให้นิยามของผลสัมฤทธิ์ว่าเป็นทักษะทางปัญญาหรือทักษะเชิงกระบวนการเฉพาะสาขาวิชา (domain-

specific process skills) ที่บุคคลได้รับการฝึกฝนอย่างต่อเนื่อง เพื่อใช้ในการวิเคราะห์วางแผน กำกับดูแลและประเมินผลการปฏิบัติภารกิจ ผลสัมฤทธิ์ครอบคลุมถึงปริมาณและคุณภาพของ ข้อมูลความรู้ทางวิชาการที่บุคคลได้รับและนำมาจัดองค์ประกอบอย่างเป็นระบบในหน่วย ความจำถาวร สามารถนำข้อมูลออกมาใช้แก้ปัญหาได้

ดังนั้นผลสัมฤทธิ์ทางการเรียน จึงหมายถึง ผลสำเร็จที่เกิดจากการเรียนตาม จุดมุ่งหมายของหลักสูตรใดหลักสูตรหนึ่งโดยเฉพาะ ผลสัมฤทธิ์ทางการเรียนตามกรอบทฤษฎี ของ Bloom ประกอบด้วยผลด้านพุทธิพิสัย (Cognitive) จิตพิสัย (Affective) และทักษะพิสัย (Psychomotor) แต่โดยทั่วไปแล้วมักยึดเอาผลสัมฤทธิ์ทางด้านพุทธิพิสัยเป็นหลัก ส่วน ผลสัมฤทธิ์ทางด้านจิตพิสัยและทักษะพิสัย ได้รับความสนใจค่อนข้างน้อย

ตัวบ่งชี้ของผลสัมฤทธิ์ทางการเรียน มีทั้งเป็นข้อมูลเชิงคุณภาพและข้อมูลเชิงปริมาณ

1) ตัวบ่งชี้เชิงคุณภาพ ได้แก่ แฟ้มสะสมผลงาน (Portfolio) ระเบียบสะสม (Transcript) ข้อมูลดังกล่าวมีบทบาทสำคัญมากในการพิจารณาคัดเลือกบุคคลเข้าศึกษาต่อ ในสถาบันอุดมศึกษาในต่างประเทศ และมีการริเริ่มนำแนวความคิดนี้มาใช้กับระบบการสอบ คัดเลือกเข้ามหาวิทยาลัยของไทยในปัจจุบัน

2) ตัวบ่งชี้เชิงปริมาณ ได้แก่ แด้มเฉลี่ยสะสม (Grade point average) ผลการทดสอบแบบสอบมาตรฐานประเภทผลสัมฤทธิ์ทั่วไป และคะแนนสอบไล่ (Exit examination) สำหรับระดับมัธยมศึกษาตอนปลาย อย่างเช่น การสอบไล่ ม.8 ในสมัยก่อน

ในทางปฏิบัติจะมีการเลือกใช้ตัวบ่งชี้ตัวใดตัวหนึ่ง หรือหลาย ๆ ตัวประกอบกันก็ได้ ขึ้นอยู่กับจุดประสงค์การนำไปใช้ (กรมวิชาการ, 2540)

แด้มเฉลี่ยสะสมระดับมัธยมศึกษาตอนปลายของนักเรียนแต่ละคนคำนวณจากผลการ เรียนในทุกวิชาตลอดหลักสูตร ผลการเรียนแต่ละรายวิชาตามหลักการวัดและประเมินผลได้ จากการวัดและประเมินผลระหว่างภาคเรียน (Formative evaluation) และปลายภาคเรียน (Summative evaluation) การประเมินระหว่างภาคเรียนอาจเป็นการสังเกตพฤติกรรม การ พิจารณาผลงานและการทดสอบ ส่วนการประเมินปลายภาคเรียนมักจะเป็นการสอบข้อเขียน ซึ่งใช้ข้อสอบที่ครูผู้สอนสร้างเอง (Teacher-made test) แด้มเฉลี่ยสะสมจะเป็นตัวบ่งชี้ที่ ครอบคลุมทั้งคุณลักษณะของผู้เรียนในหลาย ๆ ด้าน และคุณภาพของกระบวนการเรียน การสอนของโรงเรียน นอกจากนี้แด้มเฉลี่ยสะสมยังได้จากการวัดด้วยวิธีการที่หลากหลาย วัด หลาย ๆ ครั้ง และมีการบันทึกผลอย่างเป็นระบบ

นักวัดผลทางจิตวิทยาและการศึกษานหลายคนมีความเห็นว่า ความถนัด (Aptitude) และผลสัมฤทธิ์ (Achievement) มีความสัมพันธ์กันอย่างใกล้ชิดและเครื่องมือทั้งสองชนิดต่างวัดความสามารถที่ได้จากการเรียนรู้เช่นเดียวกัน แต่จะมีความแตกต่างกันในด้านลักษณะของเครื่องมือวัดหรือตัวบ่งชี้ ในแง่ของขอบเขตของมวลความรู้ ประสบการณ์ อันเป็นที่มาของตัวบ่งชี้ และลักษณะการนำเครื่องมือวัดไปใช้ สำหรับด้านมวลความรู้ประสบการณ์แบบทดสอบผลสัมฤทธิ์จะมีขอบเขตของมวลความรู้ ประสบการณ์ที่ชัดเจนและจำกัด ส่วนแบบทดสอบความถนัดจะมีขอบเขตของมวลความรู้ ประสบการณ์ที่กว้างขวางและไม่ชัดเจนในด้านของลักษณะการนำเครื่องมือวัดไปใช้ แบบทดสอบผลสัมฤทธิ์จะนำไปใช้เพื่อวัดความสามารถในปัจจุบัน ส่วนแบบทดสอบความถนัดจะใช้วัดเพื่อทำนายความสามารถหรือความสำเร็จในการเรียนหรือการทำงานในอนาคต โดยสัมพันธ์กับความตรงของแบบทดสอบผลสัมฤทธิ์ มักคำนวณโดยอาศัยเกณฑ์ซึ่งเป็นแบบความตรงเชิงเนื้อหา (Content validity) แต่แบบทดสอบความถนัดเป็นแบบความตรงเชิงทำนาย (Predictive validity)

หากพิจารณาจากทฤษฎีโครงสร้างความสามารถทางสมองสมัยใหม่ (กรมวิชาการ, 2537) เช่น ทฤษฎีโครงสร้างทางสมองของแคทเทลล์ (Catell's Intelligence Theory) สามารถอธิบายความแตกต่างระหว่างผลสัมฤทธิ์ทั่วไปและความถนัดทางการเรียนได้ว่า ผลสัมฤทธิ์ทั่วไป (General Achievement) เป็นตัวบ่งชี้ถึงความรู้ความสามารถที่ตกผลึก (Crystallized Ability) อันได้แก่ ปริมาณและคุณภาพของข้อมูลความรู้ทางวิชาการที่บุคคลได้รับ และจัดเป็นระบบในหน่วยความจำถาวร รวมทั้งทักษะกระบวนการและกรอบการวิเคราะห์ที่บุคคลจะนำออกมาใช้ในการแก้ปัญหาและศึกษาหาความรู้เพิ่มเติมต่อไป องค์ประกอบที่สำคัญของแบบทดสอบผลสัมฤทธิ์ทั่วไป (General Achievement Test) คือ

- 1) การจัดองค์ประกอบของโครงสร้างความรู้ (Organization of knowledge structure)
- 2) ทักษะทางปัญญา (Cognitive processes)
- 3) กรอบการวิเคราะห์ (Schemata)

ในขณะที่ความถนัดทางการเรียน (Scholastic Aptitude) เป็นตัวบ่งชี้ที่เน้นความสามารถที่เลื่อนไหล (Fluid Ability) คือ ความสามารถในการนำความรู้และทักษะทางสติปัญญามาจัดองค์ประกอบใหม่ ให้เป็นกรอบการวิเคราะห์ กระบวนการที่สามารถนำมาใช้แก้ปัญหาสำหรับสถานการณ์ใหม่ ๆ ได้ องค์ประกอบสำคัญของแบบทดสอบความถนัดทางการเรียน (กรมวิชาการ, 2539; Deptula & Freedman, 1992; Brownstein & others, 1994) คือ

1) ความสามารถทางภาษา (Verbal Ability)

ข้อสอบด้านนี้วัดความรู้ ความเข้าใจและเหตุผลเชิงภาษาของผู้เรียน โดยครอบคลุมพื้นฐานความรู้ความเข้าใจทางภาษา ความสามารถในการวิเคราะห์ เปรียบเทียบ และประเมินความสัมพันธ์รูปแบบต่าง ๆ ระหว่างศัพท์ คำนวน ข้อมูล และแนวความคิดที่สื่อในรูปข้อความหรือบทความ เหตุผลเชิงภาษา (Verbal Reasoning) และการอ่านอย่างมีวิจารณญาณ (Critical Reading) ลักษณะของข้อสอบแบ่งเป็น 3 แบบ คือ

ก. การเติมข้อความให้ได้ใจความสมบูรณ์ (Sentence Completion) เป็นการวัดความรู้ความเข้าใจความหมายของศัพท์และสำนวนภาษา และการใช้ศัพท์และสำนวนภาษาเหล่านั้นในประโยคหรือข้อความที่กำหนดให้ได้อย่างได้ใจความสมบูรณ์ และสมเหตุสมผล

ข. อุปมาอุปไมยทางภาษา (Word Analogy) เป็นการวัดความสามารถในการวิเคราะห์ความสัมพันธ์ระหว่างคู่คำหรือสำนวนที่กำหนดให้ และความสามารถในการเปรียบเทียบและประเมินความสัมพันธ์ระหว่างคู่คำหรือสำนวนต่าง ๆ

ค. การอ่านอย่างมีวิจารณญาณ (Critical Reading) เป็นการวัดความสามารถในการวิเคราะห์และประเมินข้อความหรือบทความที่กำหนดให้ ในองค์ประกอบต่าง ๆ ดังต่อไปนี้

- (1) การจับใจความหลักและเจตนาารมณ์สำคัญ
- (2) ความเข้าใจข้อมูลและข้อเท็จจริงที่ระบุอย่างชัดเจน
- (3) ความเข้าใจแนวความคิดที่พาดพิง แต่ไม่ได้ระบุอย่างชัดเจน
- (4) การประยุกต์ใช้แนวความคิดในบทความกับสถานการณ์อื่น ๆ
- (5) การวิเคราะห์ตรรกะ เหตุผลและเทคนิควิธีการนำเสนอที่ผู้เขียนใช้
- (6) การวิเคราะห์จุดยืน เจตคติและอารมณ์ความรู้สึกของผู้เขียนจากลักษณะภาษาที่ใช้

2) ความสามารถทางการคิดคำนวณ (Numerical Ability)

ข้อสอบด้านนี้วัดความเข้าใจในความคิดรวบยอดและหลักการทางคณิตศาสตร์ระดับเบื้องต้น ทักษะการคำนวณระดับพื้นฐาน ความสามารถด้านเหตุผลเชิงปริมาณ (Quantitative Reasoning) การวิเคราะห์ เปรียบเทียบและประเมินข้อมูลเชิงปริมาณในรูปแบบต่าง ๆ เช่น ตาราง แผนภูมิ และประโยคสัญลักษณ์ทางคณิตศาสตร์และความสามารถในการแก้ปัญหาในรูปของจำนวนหรือปริมาณ แบ่งเป็น 4 องค์ประกอบย่อย คือ

ก. ความสามารถพื้นฐานทางคณิตศาสตร์ (Basic Mathematical Ability) ข้อสอบจะครอบคลุมทักษะการใช้วิธีการคิดพื้นฐาน เช่น การบวก ลบ คูณ หาร การถอดราก การยกกำลัง และความสามารถในการตีความและแก้ปัญหาโจทย์ ที่ต้องอาศัยพื้นฐานความเข้าใจในความคิดรวบยอดและหลักการในคณิตศาสตร์ระดับเบื้องต้น เช่น เศษส่วน ทศนิยม ร้อยละ กำไร/ขาดทุน ระยะทาง พื้นที่ ปริมาตร เวลา สมการ เรขาคณิตเบื้องต้น พีชคณิตเชิงเส้นขั้นพื้นฐาน (Basic Linear Algebra) และกราฟ เป็นต้น

ข. ความสามารถด้านการเปรียบเทียบ (Quantitative Comparison) ข้อสอบด้านนี้วัดความสามารถในการคำนวณคิดหาเหตุผล และเปรียบเทียบขนาดของปริมาณ ซึ่งต้องอาศัยความรู้ความเข้าใจเบื้องต้นทางพีชคณิต เลขคณิต และเรขาคณิต

ค. ความสามารถในการตีความข้อมูล (Data Interpretation) ข้อสอบด้านนี้วัดความสามารถในการอ่านและวิเคราะห์ข้อมูลเชิงปริมาณที่เสนอในรูปแบบต่าง ๆ เช่น ตาราง แผนภูมิ กราฟ สมการ หรือประโยคสัญลักษณ์ทางคณิตศาสตร์ การสังเคราะห์ข้อมูล และการเลือกใช้ข้อมูลที่เหมาะสมเพื่อตอบปัญหา

ง. การประเมินความเพียงพอของข้อมูล (Evaluation of Data Sufficiency) เพื่อวัดความสามารถในการวิเคราะห์และตีความปัญหาทางคณิตศาสตร์ การพิจารณาและจำแนกความเหมาะสมของข้อมูลในการแสวงหาคำตอบ และการประเมินความเพียงพอของข้อมูลสำหรับการแก้ปัญหา

3) ความสามารถเชิงวิเคราะห์ (Analytical Ability)

ข้อสอบด้านนี้วัดความสามารถในด้านต่าง ๆ ดังนี้

(1) การสังเกตเห็นและการวิเคราะห์ความสัมพันธ์ระหว่างข้อมูลสนเทศต่าง ๆ เช่น ระหว่างหลักฐานและสมมติฐาน ระหว่างสาเหตุและผลสรุป และระหว่างข้อเท็จจริงกับคำอธิบาย เป็นต้น

(2) การพิจารณาและประเมินความสอดคล้องระหว่างข้อความที่มีความสัมพันธ์เกี่ยวข้องกัน

(3) การสรุปความจากข้อมูลหรือสาระที่ซับซ้อน

(4) การใช้วิธีหรือขั้นตอนในการจัดประเด็นหรือทางเลือกที่ผิด เพื่อนำไปสู่การสรุปผลอย่างถูกต้อง

(5) การสรุปและสร้างหลักเกณฑ์จากความสัมพันธ์ที่วิเคราะห์ได้จากข้อมูลที่ เป็นนามธรรม สัญลักษณ์ หรือแผนภาพต่าง ๆ

(6) การวิเคราะห์ความสัมพันธ์ระหว่างกลุ่มหรือการจัดประเภทที่มีความเหลื่อมล้ำในลักษณะต่าง ๆ หรือที่เป็นอิสระไม่คาบเกี่ยวกัน

ข้อสอบแบ่งตามลักษณะได้ 3 แบบใหญ่ ๆ คือ

ก. การใช้เหตุผลเชิงตรรกและเชิงวิเคราะห์ทางภาษา (Verbal Analysis) โจทย์จะประกอบด้วยคำบรรยายสถานการณ์หรือกลุ่มข้อความที่ให้ข้อมูลเกี่ยวกับสถานการณ์ ผู้สอบจะต้องใช้ความสามารถด้านต่าง ๆ ชำ้้งต้น เพื่อตอบคำถามที่ตามมา

ข. การวิเคราะห์แผนภูมิเชิงตรรก (Logical Diagram) โจทย์จะกำหนดภาพวงกลมให้อันหมายถึงกลุ่มหรือการจัดประเภทของสิ่งของ ปรัชญาการณ หรือคุณลักษณะต่าง ๆ และให้การตัดกันหรือซ้อนกันของวงกลมแสดงถึงความสัมพันธ์ระหว่างกลุ่มหรือการจัดประเภท ผู้สอบจะต้องวิเคราะห์ความสัมพันธ์จากภาพที่กำหนดให้ และใช้ความสัมพันธ์ที่วิเคราะห์ได้ใน การตอบคำถามแต่ละข้อ

ค. การวิเคราะห์เชิงภาพหรือสัญลักษณ์ (Non-verbal Analysis) แบ่งออกเป็น 4 แบบ คือ

(1) อุปมาอุปไมยด้วยรูปภาพ (Picture Analogy) มีลักษณะคล้ายกับ อุปมาอุปไมยทางภาษา เพียงแต่ใช้รูปภาพหรือสัญลักษณ์แทนคำหรือสำนวนภาษา ผู้สอบ ต้องตีความแต่ละภาพ วิเคราะห์ความสัมพันธ์ระหว่างภาพที่กำหนดให้ และใช้ความสัมพันธ์ที่ วิเคราะห์ได้กับภาพอื่น ๆ

(2) อนุกรมภาพ (Figural Series) ข้อสอบจะกำหนดอนุกรมภาพให้ 1 ชุด ซึ่งมีการเปลี่ยนแปลงลักษณะต่อเนื่องจากภาพหนึ่งไปอีกภาพหนึ่ง ผู้สอบต้องวิเคราะห์ ความต่อเนื่องของอนุกรมภาพเพื่อหากฎเกณฑ์ของความเปลี่ยนแปลง แล้วใช้กฎเกณฑ์นั้นไป พยากรณ์ความเปลี่ยนแปลงของภาพอื่น

(3) การจัดประเภทภาพ (Figural Classification) ข้อสอบจะวัด ความสามารถในการวิเคราะห์องค์ประกอบของภาพที่กำหนดให้ เพื่อใช้เป็นข้อมูลในการ สังเคราะห์กฎเกณฑ์ที่ใช้ในการจัดหมวดหมู่ และใช้กฎเกณฑ์นั้นในการพิจารณาจัดภาพใหม่ ให้เข้ากลุ่มอย่างถูกต้อง

(4) อนุกรมภาพ 2 มิติ (Figural Matrices) ข้อสอบประเภทนี้ต้องใช้ ความสามารถหลายอย่างรวมกัน คือ ความสามารถในการวิเคราะห์เชิงอุปมาอุปไมย ความ สามารถในการวิเคราะห์กฎเกณฑ์การเปลี่ยนแปลงในอนุกรม และความสามารถในการ สังเคราะห์กฎเกณฑ์ที่ใช้ในการจัดกลุ่มจากตัวอย่าง โดยทั่วไปโจทย์จะกำหนดภาพให้ 3 แถว แถวละ 3 ภาพ เรียงเป็นจัตุรัส การเปลี่ยนแปลงจากภาพหนึ่งไปสู่อีกภาพหนึ่งมีทั้งในแนวนอน

และแนวตั้ง โดยไม่จำเป็นต้องใช้กฎเกณฑ์เดียวกัน ผู้สอบจะต้องวิเคราะห์ สังเคราะห์ และ พยากรณ์ลักษณะของภาพที่เปลี่ยนแปลงไปอย่างสอดคล้องกับกฎเกณฑ์ทั้ง 2 มิติ

จะเห็นได้ว่าผลสัมฤทธิ์ทางการเรียนและความถนัดทางการเรียน ต่างเป็นผลระยะยาว ของการเรียนการสอนตามหลักสูตรในภาพรวม และเป็นตัวทำนาย (Predictors) ที่สำคัญของ ความสามารถในการศึกษาและการทำงานในอนาคต ซึ่งสอดคล้องกับผลงานวิจัยมากมาย ดังเช่น สพอลดิง (Spaulding, 1959) ได้ศึกษาถึงการทำนายผลการเรียนเฉลี่ยในวิทยาลัย โดยใช้คะแนนสอบความถนัดทางการเรียนทั้ง 3 ด้าน คือ ด้านภาษา ด้านคณิตศาสตร์และด้าน การให้เหตุผล เป็นตัวทำนายผลการเรียนเฉลี่ยชั้นปีที่ 1 ของนักศึกษา ผลการศึกษาพบว่า ค่า สัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนสอบความถนัดทางการเรียนกับผลการเรียนเฉลี่ยในปี แรก มีค่าระหว่าง 0.520 - 0.633 ซึ่งสอดคล้องกับงานวิจัยของ เอล (Elle, 1967) ได้ศึกษา เกี่ยวกับความตรงเชิงทำนายของความถนัดทางการเรียนและผลสัมฤทธิ์ทางการเรียนเพื่อ พยากรณ์ผลการเรียนเฉลี่ยภาคเรียนแรกของนักศึกษา พบว่า ความถนัดทางการเรียน สามารถ ทำนายผลการเรียนเฉลี่ยภาคเรียนแรกได้ดีที่สุด เช่นเดียวกับกับงานวิจัยของ ยัง (Young, 1991) ซึ่งได้ทำการศึกษาประสิทธิภาพในการทำนายผลสัมฤทธิ์ทางการเรียนในมหาวิทยาลัย ขององค์ประกอบในการรับคัดเลือก ซึ่งได้แก่ คะแนนสอบความถนัดทางการเรียน และผลการ เรียนระดับมัธยมศึกษาตอนปลาย โดยใช้นักศึกษาจากมหาวิทยาลัย Stanford ที่เข้าศึกษาในปี ค.ศ.1982 จำนวน 1,564 คน เป็นกลุ่มตัวอย่าง ผลการศึกษาพบว่า ความถนัดทางด้านภาษา, ความถนัดทางคณิตศาสตร์ และผลการเรียนระดับมัธยมศึกษาตอนปลาย มีความสัมพันธ์ใน ระดับปานกลางกับแต้มเฉลี่ยสะสมตลอดหลักสูตรในระดับอุดมศึกษา โดยมีค่าสัมประสิทธิ์ สหสัมพันธ์เท่ากับ 0.46, 0.46 และ 0.35 ตามลำดับ และองค์ประกอบทั้งสามสามารถอธิบาย ความแปรปรวนของแต้มเฉลี่ยสะสมตลอดหลักสูตร ได้เท่ากับ 0.3324

ส่วนงานวิจัยของ เลย์แมน (Leyman, 1967) ได้ศึกษาเกี่ยวกับการทำนายผลการเรียน ชั้นปีที่ 1 สาขาพลศึกษา โดยมีตัวพยากรณ์ คือ ผลการเรียนเฉลี่ยระดับมัธยมศึกษาตอนปลาย คะแนนสอบคัดเลือกเข้ามหาวิทยาลัย คะแนนทดสอบสมรรถภาพทางร่างกาย พบว่า ผลการ เรียนเฉลี่ยระดับมัธยมศึกษาตอนปลายเป็นตัวพยากรณ์ที่ดีที่สุดในการทำนายผลการเรียนชั้น ปีที่ 1 ซึ่งผลสอดคล้องกับงานวิจัยของ มาร์แชล (Marshall, 1969) ที่ได้ศึกษาเรื่องการทำนาย ผลสัมฤทธิ์ทางการเรียนของนิสิตชั้นปีที่ 1 โดยใช้ผลการเรียนเฉลี่ยในระดับมัธยมศึกษาตอน ปลาย คะแนนสอบภาษาอังกฤษ และคะแนนสอบเข้ามหาวิทยาลัยเป็นตัวทำนายผลการเรียน

เฉลี่ยชั้นปีที่ 1 ผลปรากฏว่า ผลการเรียนรู้เฉลี่ยในระดับมัธยมศึกษาตอนปลายสามารถทำนายผลสัมฤทธิ์ทางการเรียนชั้นปีที่ 1 ได้ดีที่สุด โดยมีค่าสัมประสิทธิ์สหสัมพันธ์ เท่ากับ 0.60 และสอดคล้องกับงานวิจัยของ โกลด์แมนและเฮวิตท์ (Goldman & Hewitt, 1976) ได้ศึกษาเกี่ยวกับการทำนายความสำเร็จในการเรียนของนักศึกษาที่มีความแตกต่างทางเชื้อชาติ (ethnicity) โดยเก็บข้อมูลในมหาวิทยาลัยแคลิฟอร์เนีย ทั้ง 4 แห่ง ได้แก่ Los Angeles, Davis, Irvine และ San Diego ในระหว่าง ปี ค.ศ.1973 - 1974 ผลการศึกษาพบว่า ตัวแปรทำนายที่เหมาะสมสำหรับกลุ่มนักศึกษามิวดำ (black) คือ คะแนนความถนัดทางด้านภาษาและผลการเรียนระดับมัธยมศึกษาตอนปลาย, กลุ่มนักศึกษาทางทวีปตะวันออก (oriental) คือ ผลการเรียนรู้ระดับมัธยมศึกษาตอนปลายและความถนัดทางคณิตศาสตร์ สำหรับกลุ่มนักศึกษา chicano คือ ผลการเรียนรู้ระดับมัธยมศึกษาตอนปลาย และกลุ่มนักศึกษามิวนขาว (white) คือ ผลการเรียนรู้ระดับมัธยมศึกษาตอนปลาย หากพิจารณาโดยรวมจะเห็นได้ว่า ผลการเรียนรู้ระดับมัธยมศึกษาตอนปลาย, ความถนัดทางด้านภาษาและความถนัดทางด้านคณิตศาสตร์ มีประสิทธิภาพในการทำนายผลสัมฤทธิ์ทางการเรียนในระดับปานกลาง ส่วนตัวแปรทำนายที่ดีที่สุด คือ ผลการเรียนรู้ระดับมัธยมศึกษาตอนปลาย

ส่วนงานวิจัยของ บรอดนิก และรี (Brodnic & Ree, 1995) ที่ได้ทำการศึกษาเกี่ยวกับโมเดลเชิงโครงสร้างของความสามารถทางการเรียน, ฐานะทางเศรษฐกิจและสังคม และความสามารถทางสติปัญญา ผลการศึกษาพบว่า โมเดลเชิงโครงสร้างที่เหมาะสมนั้น ความสัมพันธ์ระหว่างตัวแปร มีดังนี้ ความสามารถทางการเรียน สามารถอธิบายความแปรปรวนของความสามารถทางสติปัญญา ได้เท่ากับ 0.452 หรือ 45.2% โดยที่ความสามารถทางสติปัญญานี้วัดได้จาก คะแนนความถนัดทางคณิตศาสตร์, คะแนนความถนัดทางด้านภาษาและคะแนนสอบ ACT ซึ่งมีค่าน้ำหนักองค์ประกอบเท่ากับ 0.698, 0.804, 0.870 ตามลำดับ ส่วนความสามารถทางการเรียนวัดได้จากผลการเรียนระดับมัธยมศึกษาตอนปลาย, ผลการเรียนรู้ระดับอุดมศึกษา วิชาคณิตศาสตร์ และผลการเรียนระดับอุดมศึกษาวิชาภาษาอังกฤษ มีค่าน้ำหนักองค์ประกอบเท่ากับ 0.468, 0.384 และ 0.487 ตามลำดับ ซึ่งสรุปได้ว่า ผลการเรียนรู้ระดับมัธยมศึกษาตอนปลายมีความสัมพันธ์กับความสามารถในการเรียนในระดับปานกลาง ส่วนคะแนนสอบความถนัดทางการเรียนมีความสัมพันธ์กับความสามารถทางสติปัญญาในระดับสูง

สำหรับงานวิจัยที่ศึกษาในประเทศมีหลายเรื่อง เช่น งานวิจัยของหรรษา สานราษฎร์ทอง (2510), อวีววรรณ พรหมสาขา ณ สกลนคร (2511), ทิพย์พรรณ นพวงศ์ ณ ออยุธยา (2511),

วิรัตน์ วสะยางกูร (2511), สายหยุด เขียวดอกน้อย (2511) และวาสนา พานิชการ (2511) ได้ศึกษาเกี่ยวกับการเปรียบเทียบความสัมพันธ์ระหว่างคะแนนสอบชั้นมัธยมศึกษาปีที่ 5, ผลการสอบคัดเลือกเข้ามหาวิทยาลัย กับผลสัมฤทธิ์ทางการเรียนชั้นปีที่ 1 และ 2 โดยทำการศึกษาในมหาวิทยาลัยแตกต่างกัน ผลการศึกษามีความสอดคล้องกัน คือ มีค่าสัมประสิทธิ์สหสัมพันธ์พหุคูณระหว่างตัวพยากรณ์กับตัวแปรเกณฑ์ในระดับปานกลาง, ทั้งคะแนนสอบชั้นมัธยมศึกษาปีที่ 5 และคะแนนสอบเข้ามหาวิทยาลัยต่างมีประสิทธิภาพในการทำนายผลการเรียนชั้นปีที่ 1 และ 2 ใกล้เคียงกัน หากใช้คะแนนสอบชั้นมัธยมศึกษาปีที่ 5 และคะแนนสอบเข้ามหาวิทยาลัยร่วมกันจะสามารถทำนายได้สูงขึ้น แต่หากต้องใช้อย่างใดอย่างหนึ่ง คะแนนสอบชั้นมัธยมศึกษาปีที่ 5 จะให้ผลที่ดีกว่า

ส่วนงานวิจัยของนางลักษณ์ ประเสริฐ (2516) ที่ได้ศึกษาองค์ประกอบที่ส่งผลต่อผลสัมฤทธิ์ทางการเรียนชั้นมัธยมศึกษาพบว่ามี 3 ด้าน คือ องค์ประกอบด้านความถนัด องค์ประกอบด้านเศรษฐกิจและครอบครัว และองค์ประกอบด้านโรงเรียน ส่วนงานวิจัยของสำนักงานคณะกรรมการการศึกษาแห่งชาติ (2523) ได้ศึกษาในเรื่องเดียวกัน ซึ่งพบว่า ในด้านผลสัมฤทธิ์ทางการเรียน นักเรียนในกรุงเทพมหานครจะทำคะแนนได้ดีที่สุด ส่วนภาคที่ทำคะแนนได้ในระดับต่ำ ได้แก่ ภาคใต้ ภาคตะวันออกเฉียงเหนือ ถ้าพิจารณาตามประเภทโรงเรียน พบว่า นักเรียนในโรงเรียนสังกัดกรมสามัญศึกษาทำคะแนนได้ดีกว่านักเรียนในโรงเรียนสังกัดสำนักงานคณะกรรมการการศึกษาเอกชน และเมื่อพิจารณาเขตอำเภอที่ตั้งนักเรียนจากโรงเรียนในเขตอำเภอเมืองสามารถทำคะแนนได้ดีกว่านักเรียนจากโรงเรียนที่อยู่ในอำเภออื่น

นอกจากนี้งานวิจัยของจิตราภา กุณฑบุตร (2522) ได้ทำการศึกษาเกี่ยวกับการคัดเลือกสมการที่เหมาะสมในการทำนายผลสัมฤทธิ์ทางการเรียนของนิสิตคณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย จากตัวแปรภูมิหลัง ซึ่งได้แก่ เพศ อายุ ภูมิฐานะ อาชีพบิดา-มารดา คะแนนสอบชั้นมัธยมศึกษาปีที่ 5 และเกรดเฉลี่ยปีแรกที่เข้าศึกษา ผลการวิจัยปรากฏว่า ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนสอบชั้นมัธยมศึกษาปีที่ 5 กับเกรดเฉลี่ยปีแรก มีค่าเท่ากับ -0.332 ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนสอบชั้นมัธยมศึกษาปีที่ 5 กับเกรดเฉลี่ยตลอดหลักสูตร มีค่าเท่ากับ 0.376 และค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างเกรดเฉลี่ยปีแรกกับเกรดเฉลี่ยตลอดหลักสูตร มีค่าเท่ากับ 0.782 ส่วนตัวแปรที่ใช้ทำนายผลสัมฤทธิ์ทางการเรียนที่ดีที่สุด คือ เกรดเฉลี่ยปีแรก

ส่วนงานวิจัยของจันทร์ ดิยะวงศ์ (2528) ซึ่งได้ทำการศึกษารูปแบบที่เหมาะสมในการทำนายผลสัมฤทธิ์ทางการเรียนของนิสิตครุศาสตร์ชั้นปีที่ 1 จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2527 พบว่า ตัวแปรที่มีความสัมพันธ์ในทางบวกกับผลสัมฤทธิ์ทางการเรียนของนิสิตครุศาสตร์ ชั้นปีที่ 1 ได้แก่ ความรู้เดิม, สถิติปัญญา, นิสัยในการเรียน, คุณวุฒิบิดาและความถนัดทางด้านภาษา เช่นเดียวกับ ยุวดี บุญศรีสวัสดิ์ (2528) ที่ได้ศึกษาความสัมพันธ์ระหว่างผลสัมฤทธิ์ทางการเรียนกับองค์ประกอบที่เกี่ยวข้อง พบว่า ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างผลสัมฤทธิ์ทางการเรียนกับความถนัดทางการเรียน มีค่าเท่ากับ 0.5297 มีนัยสำคัญทางสถิติที่ระดับ .01 และสอดคล้องกับงานวิจัยของชาติชาย สุภสร (2531) ได้ศึกษาความสัมพันธ์ระหว่างผลสัมฤทธิ์ทางการเรียนระดับมัธยมศึกษาตอนปลาย, ความถนัดทางการเรียน, แรงจูงใจใฝ่สัมฤทธิ์, นิสัยในการเรียนและทัศนคติในการเรียนกับคะแนนสอบคัดเลือกเข้ามหาวิทยาลัยของนักเรียนในสังกัดกรมสามัญศึกษา จังหวัดอุบลราชธานี พบว่า ตัวพยากรณ์ทุกตัวมีความสัมพันธ์กับคะแนนสอบเข้ามหาวิทยาลัย และความถนัดทางการเรียนมีความสัมพันธ์กับคะแนนสอบเข้ามหาวิทยาลัยทุกวิชา ในทิศทางบวก

จากผลการศึกษาที่กล่าวมาข้างต้น ชี้ให้เห็นว่า ผลการเรียนระดับมัธยมศึกษาตอนปลายและคะแนนสอบความถนัดทางการเรียน มีความสัมพันธ์ในทางบวกและสามารถทำนายผลสัมฤทธิ์ทางการเรียนในมหาวิทยาลัยได้ อันเป็นประเด็นในการสนับสนุนให้ใช้ผลการเรียนระดับมัธยมศึกษาตอนปลายและคะแนนสอบความถนัดทางการเรียน เป็นองค์ประกอบในการคัดเลือกนักเรียนเข้าศึกษาในสถาบันอุดมศึกษา

ตอนที่ 3 การเปรียบเทียบคะแนนระหว่างแบบสอบ (Test Equating)

จากสภาพการศึกษาในปัจจุบันซึ่งมีการใช้แบบสอบและผลการสอบในการวัดและประเมินผลสัมฤทธิ์ทางการเรียนของผู้เรียนอาจมีข้อจำกัดหลายประการในการบริหารการสอบ เช่น มีผู้เข้าสอบจำนวนมากเกินไป ซึ่งจำเป็นต้องแบ่งเป็นกลุ่มย่อยหลาย ๆ กลุ่ม หรือการจัดสอบในเวลาแตกต่างกันและแบบสอบที่ใช้ต่างฉบับกัน ระดับความยากแตกต่างกัน คะแนนจากข้อสอบและแบบสอบย่อมมีค่าไม่เท่ากันและมีความหมายต่างกัน ดังนั้นคะแนนที่ได้จากแบบสอบจึงควรที่จะนำมาเปรียบเทียบให้อยู่บนสเกลเดียวกัน เพื่อที่จะได้ประเมินหรือตัดสินได้อย่างถูกต้องและยุติธรรม

การพัฒนาการปรับเทียบคะแนนระหว่างแบบสอบได้รับความสนใจมาเป็นเวลานาน ซึ่ง แองกอฟ (Angoff, 1971) ได้ให้ความหมายของการปรับเทียบคะแนนระหว่างแบบสอบว่าเป็นกระบวนการแปลงระบบคะแนนของแบบสอบหนึ่งให้อยู่บนสเกลของแบบสอบฉบับอื่นที่มุ่งวัดในสิ่งเดียวกันเพื่อให้คะแนนสามารถเปรียบเทียบกันได้ สอดคล้องกับ ศิริชัย กาญจนวาสิ (2541) ได้ให้นิยามว่า เป็นกระบวนการแปลงระบบคะแนนของแบบสอบหนึ่งให้อยู่บนสเกลของแบบสอบฉบับอื่นที่มุ่งวัดสิ่งเดียวกันเพื่อให้คะแนนเทียบกันได้และใช้แทนกันได้ และสอดคล้องกับ ปีเตอร์เซนและคณะ (Petersen & others, 1982) ที่ให้นิยามของการปรับเทียบคะแนนระหว่างแบบสอบว่าเป็นกระบวนการเชิงประจักษ์ที่ใช้ในการแปลงคะแนนจากแบบสอบหนึ่งไปยังแบบสอบหนึ่ง โดยที่แบบสอบทั้งสองนั้นต้องวัดคุณลักษณะเดียวกัน ส่วน ลอร์ด (Lord, 1980) ได้กล่าวถึงการปรับเทียบคะแนนระหว่างแบบสอบว่า การปรับเทียบคะแนนทำให้เกิดความเท่าเทียมกันระหว่างคะแนนดิบจากแบบสอบสองฉบับ และในความคิดของ โคลเลน (Kolen & Brennan, 1995) การปรับเทียบคะแนนระหว่างแบบสอบ หมายถึงกระบวนการทางสถิติที่ใช้ในการปรับคะแนนจากแบบสอบต่างฉบับ เพื่อให้คะแนนสามารถใช้สลับเปลี่ยน (interchangeably) กันได้ โดยที่แบบสอบแต่ละฉบับมีความยากแตกต่างกันได้ แต่ถ้ามีเนื้อหาต่างกัน จะไม่สามารถปรับเทียบกันได้ ดังคำกล่าวที่ว่า *"Equating adjusts to differences in difficulty, not for differences in content"*

ดังนั้นจึงกล่าวได้ว่า การปรับเทียบคะแนนระหว่างแบบสอบเป็นกระบวนการทางสถิติที่นำมาใช้ในการปรับคะแนนของแบบสอบต่างฉบับที่วัดคุณลักษณะเดียวกันให้อยู่ในสเกลเดียวกัน คะแนนจากแบบสอบทั้งสองฉบับจึงเปรียบเทียบกันได้อย่างมีความหมาย

แบบสอบที่นำมาปรับเทียบคะแนน นอกจากจะวัดคุณลักษณะ (trait) เดียวกันแล้ว จะต้องมีความเที่ยงสูงและเป็นแบบสอบที่คู่ขนานกัน (parallel) ในด้านโครงสร้าง (structure), เวลาที่ใช้สอบ (timing), ชนิดของข้อสอบ (item types), รูปแบบ (formats) และเนื้อหา (subject matter) ซึ่งข้อสอบของแต่ละฉบับอาจมีความยากแตกต่างกันได้ โดยเทคนิคการปรับเทียบคะแนนระหว่างแบบสอบจะช่วยปรับคะแนนจากแบบสอบต่างฉบับให้มีความเท่าเทียมกัน (Levine, 1955 อ้างถึงใน Holland & Rubin, 1982) นอกจากนี้ Angoff (1971); Lord (1980) และ Hambleton & Swaminathan (1984) ได้กำหนดเงื่อนไขในการปรับเทียบคะแนนไว้ว่า แบบสอบต่างฉบับจะสามารถนำมาเปรียบเทียบกันได้ก็ต่อเมื่อแบบสอบเหล่านั้นมีคุณสมบัติ 4 ประการคือ

1) ความเป็นเอกมิติของแบบสอบ (unidimensionality of the tests) คือ แบบสอบทั้งสองฉบับวัดคุณลักษณะหรือความสามารถเดียวกัน เพียงคุณลักษณะเดียว

2) ความเสมอภาค (equity) คือ เมื่อทุกกลุ่มมีความสามารถเดียวกัน การแจกแจงคะแนนของแบบสอบ Y หลังจากที่มีการแปลงคะแนนแล้วจะมีการแจกแจงเหมือนกับการแจกแจงของคะแนนจากแบบสอบ X

3) ความไม่ผันแปรตามกลุ่ม (invariance across groups) คือ การแปลงคะแนนต้องเป็นไปในลักษณะเดียวกันไม่ว่าคะแนนจะมาจากกลุ่มตัวอย่างใดก็ตาม

4) ความสมมาตร (symmetry) คือ การแปลงคะแนนจากฉบับ X ไปยังฉบับ Y มีผลเช่นเดียวกันกับการแปลงคะแนนจากแบบสอบฉบับ Y ไปสู่อฉบับ X

การปรับเทียบคะแนนสามารถจำแนกตามความแตกต่างของความยากของข้อสอบและความสามารถของผู้สอบ โดยแบ่งได้เป็น 2 ประเภท ได้แก่ การปรับเทียบคะแนนตามแนวระดับ (Horizontal Equating) และการปรับเทียบคะแนนตามแนวตั้ง (Vertical Equating) (Hambleton & Swaminatan, 1984) การปรับเทียบคะแนนตามแนวระดับเป็นการปรับเทียบคะแนนระหว่างแบบสอบที่มีความยากใกล้เคียงกันและผู้สอบมีระดับความสามารถคล้ายคลึงกัน ส่วนการปรับเทียบคะแนนตามแนวตั้งเป็นการปรับเทียบคะแนนระหว่างแบบสอบที่มีความยากต่างกันและผู้สอบมีระดับความสามารถแตกต่างกัน สำหรับงานวิจัยนี้เป็นการปรับเทียบคะแนนตามแนวระดับ เนื่องจากแต่้มเฉลี่ยคะแนนระดับมัธยมศึกษาตอนปลาย ได้มาจากผลการเรียนรายวิชาที่มีเนื้อหาในหลักสูตรมัธยมศึกษาตอนปลาย จึงถือว่ามีความยากใกล้เคียงกัน ส่วนผู้สอบเป็นนิสิตชั้นปีที่ 1 คณะครุศาสตร์ ซึ่งถือว่ามีความสามารถในระดับเดียวกัน

เนื่องจากการปรับเทียบคะแนนระหว่างแบบสอบเป็นกระบวนการเชิงประจักษ์ ดังนั้นจึงต้องมีการออกแบบในการเก็บรวบรวมข้อมูลและการออกแบบวิธีการปรับเทียบคะแนน ดังมีรายละเอียดต่อไปนี้

3.1 การออกแบบการเก็บรวบรวมข้อมูล (Designs for Data Collection)

การเก็บรวบรวมข้อมูลสำหรับนำมาผลมาใช้ในการปรับเทียบสามารถจำแนกได้ 3 รูปแบบ (Petersen & others, 1982; Hambleton & Swaminathan, 1984; Kolen & Brennan, 1995; ศิริชัย กาญจนวาสี, 2541)

- 1) แบบแผนกลุ่มเดียว (Single Group Design) แบ่งได้เป็น 2 รูปแบบ คือ
 1.1) แบบแผนกลุ่มเดียวที่ไม่ได้รับการจัดให้สมดุลย์ (Single group design with uncounterbalancing)

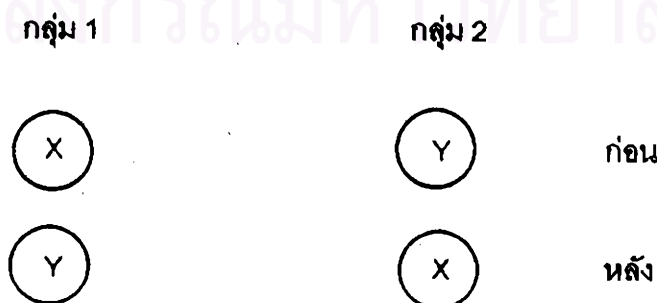
ภาพที่ 1 แบบแผนกลุ่มเดียวที่ไม่ได้รับการจัดให้สมดุลย์



รูปแบบนี้เป็นรูปแบบที่ผู้สอบกลุ่มเดียวทำแบบสอบทั้งสองฉบับโดยทำแบบสอบฉบับหนึ่งแล้วตามด้วยแบบสอบอีกฉบับหนึ่ง เช่น ทำแบบสอบฉบับ X ก่อน แล้วจึงทำแบบสอบฉบับ Y การที่ผู้สอบกลุ่มเดียวทำแบบสอบทั้งสองฉบับ ทำให้ระดับความสามารถของผู้สอบทั้งสองฉบับเท่ากัน ช่วยลดปัญหาความยากของข้อสอบต่างฉบับที่แตกต่างกันได้ แต่จะมีผลกระทบจากการเรียนรู้, ความคุ้นเคยต่อแบบสอบฉบับแรกทำให้ผู้สอบได้คะแนนสูงขึ้นหรืออาจเกิดความเมื่อยล้าในการทำแบบสอบ ซึ่งจะทำให้เกิดความไม่ตั้งใจทำแบบสอบ คะแนนที่ได้ก็จะลดลง

- 1.2) แบบแผนกลุ่มเดียวที่ได้รับการจัดให้สมดุลย์ (Single group design with counterbalancing)

ภาพที่ 2 แบบแผนกลุ่มเดียวที่ได้รับการจัดให้สมดุลย์



รูปแบบนี้ได้พัฒนาขึ้นเพื่อขจัดความลำเอียงที่เกิดในแบบแผนกลุ่มเดี่ยวที่ไม่ได้รับการจัดให้สมดุลย์ โดยกลุ่มผู้สอบออกเป็นสองกลุ่ม แต่ละกลุ่มจะได้ทำแบบสอบทั้งสองชุด ในลักษณะของการจัดลำดับก่อนหลังต่างกัน

2) แบบแผนกลุ่มสมมูล (Equivalent Group Design)

ภาพที่ 3 แบบแผนกลุ่มสมมูล



รูปแบบนี้ใช้กลุ่มผู้สอบที่มีความสามารถเท่าเทียมกันสองกลุ่มซึ่งได้มาโดยการสุ่ม แล้วให้ผู้สอบในแต่ละกลุ่มทำแบบสอบฉบับเดียว ซึ่งสามารถแก้ปัญหาการเรียนรู้หรือความเมื่อยล้าที่เกิดในแบบแผนกลุ่มเดี่ยวได้ แต่ปัญหาอาจเกิดขึ้นได้หากกลุ่มผู้สอบแต่ละกลุ่มมีความสามารถแตกต่างกัน ซึ่งทำให้เกิดความคลาดเคลื่อนในการเปรียบเทียบ การลดความแตกต่างระหว่างกลุ่มทำได้โดยใช้กลุ่มตัวอย่างขนาดใหญ่

3) แบบแผนข้อสอบร่วม (Anchor Test Design) แบ่งได้เป็น 2 รูปแบบ คือ

3.1) แบบแผนข้อสอบร่วมภายใน (Internal anchor test design)

ภาพที่ 4 แบบแผนข้อสอบร่วมภายใน



รูปแบบนี้เป็นการใช้กลุ่มผู้สอบสองกลุ่ม โดยแต่ละกลุ่มทำแบบสอบคนละชุด (X หรือ Y) โดยที่แบบสอบแต่ละชุดจะมีข้อสอบร่วมรวมอยู่ด้วย

3.2) แบบแผนข้อสอบร่วมภายนอก (External anchor test design)

ภาพที่ 5 แบบแผนข้อสอบร่วมภายนอก

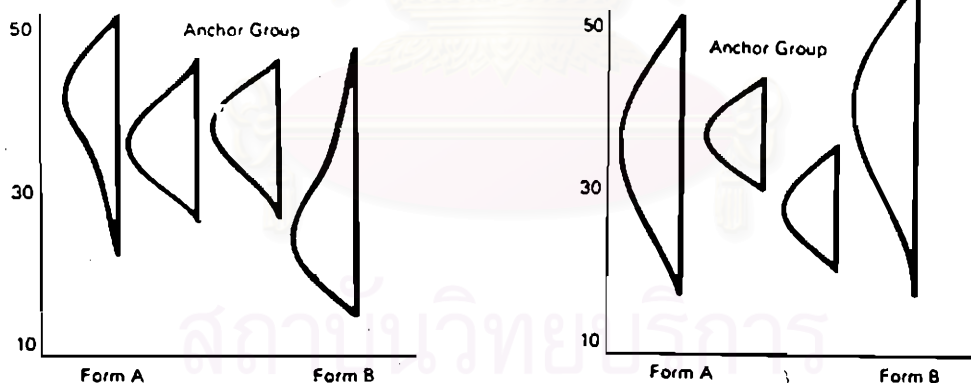


รูปแบบนี้เป็นการใช้กลุ่มผู้สอบสองกลุ่ม แล้วให้ผู้สอบแต่ละกลุ่มทำแบบสอบคนละชุดและจะมีข้อสอบร่วมซึ่งแยกอีกฉบับ

แบบแผนข้อสอบร่วมสามารถจำแนกตามความสามารถของผู้สอบ (Murphy, 1988) โดยแบ่งได้ 2 กรณี คือ

กรณีที่ 1 ผู้สอบมีความสามารถทัดเทียมกัน

กรณีที่ 2 ผู้สอบมีความสามารถแตกต่างกัน



ภาพที่ 6 แบบแผนข้อสอบร่วมจำแนกตามความสามารถของผู้สอบ

จากกรณีที่ 1 จะเห็นว่า การแจกแจงคะแนนของแบบสอบ A และ B มีลักษณะต่างกัน แต่การแจกแจงคะแนนของข้อสอบร่วมมีลักษณะเช่นเดียวกัน แสดงว่ากลุ่มผู้สอบทั้งสองฉบับมีความสามารถในระดับเดียวกันจึงทำข้อสอบร่วมได้คะแนนคล้ายคลึงกัน แต่แบบสอบ A ยากกว่าแบบสอบ B เพราะมีคะแนนเฉลี่ยสูงกว่า โดยคะแนนดิบเท่ากับ 30 ของแบบสอบ B เมื่อปรับเทียบแล้ว คะแนนแปลงจะมีค่าสูงกว่าคะแนนดิบเท่ากับ 30 ของแบบสอบ A

จากกรณีที่ 2 จะเห็นว่า การแจกแจงคะแนนของแบบสอบ A และ B มีลักษณะคล้ายคลึงกัน แต่การแจกแจงคะแนนของข้อสอบร่วมมีลักษณะต่างกัน ผู้สอบที่ทำแบบสอบ A จะทำข้อสอบร่วมได้คะแนนสูงกว่าผู้ที่ทำแบบสอบ B สรุปได้ว่าแบบสอบ A ยากกว่าแบบสอบ B หรืออีกกรณีหนึ่งคือ ผู้สอบกลุ่ม A มีความสามารถสูงกว่าผู้สอบกลุ่ม B โดยคะแนนดิบเท่ากับ 30 ของแบบสอบ A เมื่อปรับเทียบแล้วคะแนนแปลงจะมีค่าสูงกว่าคะแนนดิบเท่ากับ 30 ของแบบสอบ B

แบบแผนข้อสอบร่วมนี้จะช่วยลดความลำเอียงที่เกิดจากการเรียนรู้ ความคุ้นเคย ความเมื่อยล้า และความแตกต่างระหว่างกลุ่มผู้สอบในแบบแผนกลุ่มเดียวและแบบแผนกลุ่มผสมได้ แบบแผนข้อสอบร่วมจึงมีประสิทธิภาพดีกว่า แต่ประสิทธิภาพดังกล่าวขึ้นอยู่กับข้อสอบร่วม ซึ่งต้องวัดคุณลักษณะเดียวกัน มีเนื้อหาและระดับความยากเช่นเดียวกับแบบสอบที่ต้องการปรับเทียบและต้องมีความยาวที่เหมาะสม จากการศึกษาเกี่ยวกับแบบสอบร่วม พบว่า ข้อสอบที่บรรจุในแบบสอบร่วมควรทำหน้าที่วัดคุณลักษณะเดียวกันกับแบบสอบที่ต้องการปรับเทียบหรือมีความคู่ขนานกันสูง ซึ่งจะช่วยให้ผลการปรับเทียบมีความถูกต้องแม่นยำสูง ซึ่งสอดคล้องกับงานวิจัยของปีเตอร์เซน และคณะ (Petersen and others, 1982) ที่ได้ทำการปรับเทียบแบบสอบ SAT (Scholastic Aptitude Test) กับแบบสอบการเขียนภาษาอังกฤษ (TSWE : Test of Standard Written English) พบว่า หากความยากของแบบสอบร่วมกับแบบสอบที่ใช้ปรับเทียบต่างกัน ความคลาดเคลื่อนจะมากขึ้นหรือแบบสอบที่ใช้ปรับเทียบทั้งสองชุดมีความยากต่างกัน จะเกิดความคลาดเคลื่อนเพิ่มขึ้นและถ้าแบบสอบร่วมที่สร้างขึ้นมีลักษณะคล้ายคลึงกับแบบสอบฉบับปรับเทียบ การปรับเทียบคะแนนจะให้ผลที่ถูกต้องขึ้น

3.2 การออกแบบวิธีการปรับเทียบคะแนน (Designs for Equating Methods)

วิธีการที่ใช้ในการปรับเทียบคะแนนสามารถจำแนกตามแนวคิดพื้นฐานหรือทฤษฎีการวัดที่นำมาใช้โดยจำแนกได้เป็น 2 รูปแบบ คือ การปรับเทียบตามแนวทฤษฎีการวัดแบบดั้งเดิม (Classical Test Theory) และการปรับเทียบตามแนวทฤษฎีการตอบสนองข้อสอบ (Item Response Theory) (Angoff, 1971; Hambleton & Swaminathan, 1984; Kolen & Brennan, 1995; ศิริชัย กาญจนวาสี, 2541)

1) การปรับเทียบตามแนวทฤษฎีการวัดแบบดั้งเดิม มีวิธีการปรับเทียบที่นิยมใช้กันมาก 2 วิธี คือ

1.1) วิธีการปรับเทียบแบบอควิเปอร์เซ็นไทล์ (Equipercetile Equating)

มโนทัศน์ของการปรับเทียบแบบอควิเปอร์เซ็นไทล์ คือ การแจกแจงของคะแนนทั้งสองชุด ที่มีลักษณะคล้ายคลึงกันหรือแตกต่างกันเล็กน้อย การปรับเทียบคะแนนทำได้โดยใช้คะแนนที่อยู่ตำแหน่งเปอร์เซ็นไทล์เดียวกันของทั้งสองชุดมาเปรียบเทียบกัน มีการแสดงผลการปรับเทียบในรูปของกราฟ ระดับความยากของแบบสอบจะมีผลต่อเส้นกราฟที่ใช้ปรับเทียบ โดยแบบสอบที่มีความยากใกล้เคียงกัน เส้นกราฟจะใกล้เคียงเส้นตรง แต่ถ้าแบบสอบมีระดับความยากต่างกัน เส้นกราฟจะเป็นเส้นโค้ง เมื่อกราฟเป็นเส้นโค้งจะต้องมีการเกลารเส้นหรือปรับเส้นให้เรียบ ส่วนใหญ่ปรับด้วยมือ ซึ่งอาจจะทำให้เกิดความลำเอียงขึ้นได้ (Angoff, 1971)

ข้อจำกัดของวิธีการปรับเทียบแบบอควิเปอร์เซ็นไทล์ มีดังนี้ คือ

(1) ตามความจริงแล้วจะไม่สามารถคล้องกับหลักการที่ว่า การแจกแจงของคะแนนทั้งสองชุดต้องมีลักษณะคล้ายคลึงกัน ซึ่งอาจจะเนื่องมาจากกลุ่มผู้สอบมีความสามารถแตกต่างกันหรือแบบสอบทั้งสองมีความเที่ยงไม่เท่ากัน

(2) การเกลารหรือปรับเส้นกราฟให้เรียบ จะขึ้นอยู่กับประสบการณ์และการตัดสินใจของผู้ปรับ ทำให้เกิดความไม่เป็นปรนัย และมีความคลาดเคลื่อนสูง

(3) เมื่อกลุ่มตัวอย่างมีขนาดเล็ก จะทำให้เกิดความแปรปรวนเชิงสุ่ม

1.2) วิธีการปรับเทียบเชิงเส้นตรง (Linear Equating)

การปรับเทียบเชิงเส้นตรง เป็นรูปแบบที่ง่ายต่อการปฏิบัติ โดยมีหลักการว่า คะแนนจากแบบสอบทั้งสองฉบับจะมีการแจกแจงในลักษณะเดียวกัน ซึ่งอาจจะแตกต่างกันที่ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐาน คะแนนจากแบบสอบ X และแบบสอบ Y นั้นสามารถนำมาปรับเทียบเชิงเส้นตรงได้ ดังสมการต่อไปนี้

$$y = ax + b$$

โดยสัมประสิทธิ์ a และ b ประมาณค่าได้จาก (Kolen & Brennan, 1995)

$$\mu_y = a\mu_x + b$$

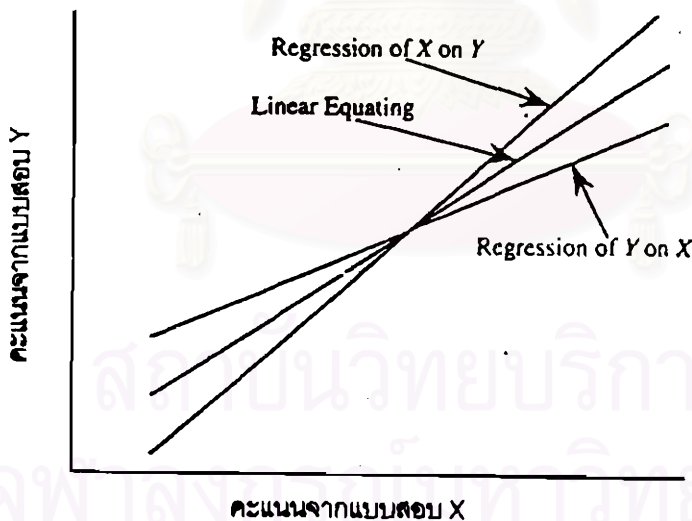
$$\sigma_y = a\sigma_x$$

เมื่อ μ_y และ μ_x เป็นค่าเฉลี่ยของคะแนน และ σ_y และ σ_x เป็นส่วนเบี่ยงเบนมาตรฐานของคะแนนจากแบบสอบ y และ x ตามลำดับ ดังนั้นจึงประมาณค่าคะแนนของแบบสอบอีกชุด จาก

$$y = \frac{\sigma_y}{\sigma_x} x + (\mu_y - \frac{\sigma_y}{\sigma_x} \mu_x)$$

สมการในการเปรียบเทียบเชิงเส้นตรงนี้คล้ายสมการถดถอยเชิงเส้นตรง แต่แตกต่างกันที่ค่า a (slope) โดยสมการเปรียบเทียบเชิงเส้นตรง ค่า $a = \frac{\sigma_y}{\sigma_x}$ ส่วนสมการถดถอยเชิงเส้นตรง ค่า $a = r_{xy} \frac{\sigma_y}{\sigma_x}$ นอกจากนั้นการทำนาย x จาก y กับการทำนาย y จาก x ก็มีลักษณะแตกต่างกัน ดังภาพที่ 7 ยกเว้นกรณี x และ y มีความสัมพันธ์กันอย่างสมบูรณ์ ($r_{xy} = 1$) การทำนาย x จาก y จึงจะเหมือนกับการทำนาย y จาก x

ภาพที่ 7 การเปรียบเทียบระหว่างการถดถอยเชิงเส้นตรงและการเปรียบเทียบเชิงเส้นตรง



ถ้าคะแนนจากแบบสอบ x และ y มีลักษณะการแจกแจงที่เหมือนกัน ความสัมพันธ์ระหว่างแบบสอบจะมีลักษณะเป็นเส้นตรง วิธีอิคิวเปอร์เซ็นไทล์จะให้ผลการเปรียบเทียบเช่นเดียวกับวิธีเชิงเส้นตรง แต่ถ้าการแจกแจงของคะแนนจากแบบสอบ x และ y แตกต่างกัน ความสัมพันธ์ระหว่างแบบสอบจะมีลักษณะเป็นเส้นโค้ง วิธีอิคิวเปอร์เซ็นไทล์จะให้ผลการเปรียบเทียบที่แน่นนอนกว่าวิธีเชิงเส้นตรง

2) การปรับตามทฤษฎีการตอบสนองข้อสอบ (Item Response Theory)

จากทฤษฎีการตอบสนองข้อสอบ ความสามารถของผู้สอบ (θ) จะเป็นอิสระจากชุดของข้อสอบและการประมาณค่าความสามารถของผู้สอบจะมีความคงเส้นคงวา เมื่อทราบค่าพารามิเตอร์ของข้อสอบ (Hambleton & Swaminathan, 1984 อ้างถึงใน พิชัย ละแมนชัย, 2538)

ยัง (Young,1990) ได้ประยุกต์ทฤษฎีการตอบสนองข้อสอบมาใช้ในการปรับผลการเรียนและเรียกวิธีการนี้ว่า The IRT-based GPA ซึ่งใช้โมเดลการตอบสนองแบบเกรด (Graded Response Model :GRM) โมเดลนี้ได้รับการพัฒนาโดย Samejima ตั้งแต่ปี ค.ศ.1969 โดยมีข้อตกลงเบื้องต้นว่าในแต่ละวิชาได้รายงานผลหรือคะแนนที่มีค่า 0,1, 2,3,4 และถือว่าผู้สอบที่มีความสามารถ (θ) สูง มีความน่าจะเป็นที่จะได้รับคะแนนหรือเกรดที่สูงกว่าผู้สอบที่มีความสามารถ (θ) ต่ำ ซึ่งความน่าจะเป็นนี้แปรผันตามความสามารถของผู้สอบ, คะแนนหรือเกรด และค่าอำนาจจำแนก การประยุกต์โมเดล GRM ที่เป็นการให้คะแนนแบบพหุวิภาค (Polytomous scoring) มาใช้กับข้อมูลนี้จะสามารถประมาณค่าความสามารถของผู้สอบได้ดีกว่า โมเดลสำหรับข้อมูลที่ให้คะแนนแบบทวิวิภาค (Dichotomous scoring) การประมาณค่าความสามารถของผู้สอบ ได้จาก

$$\pi_{nki} = \frac{\exp \{ \alpha_i [\theta_n - (\beta_i + \tau_k)] \}}{1 + \exp \{ \alpha_i [\theta_n - (\beta_i + \tau_k)] \}}$$

เมื่อ	π_{nki}	คือ	ความน่าจะเป็นของคนที่ n จะได้เกรด k ในวิชา i
	θ_n	คือ	ความสามารถของคนที่ n
	α_i	คือ	ค่าอำนาจจำแนกของวิชา i
	β_i	คือ	ค่าความยากของวิชา i
	τ_k	คือ	ค่าขอบเขตของเกรด k ซึ่งมีค่าระหว่าง k ถึง k + 1
	exp	คือ	ค่าคงที่มีค่าเท่ากับ 2.7183

ผลการเปรียบเทียบด้วยวิธีไออาร์ทีจะให้คะแนนความสามารถในรูปแบบที่ไม่เป็นเส้นตรง (Young, 1995) ส่วนโปรแกรมคอมพิวเตอร์ที่ใช้ในการประมาณค่าสำหรับโมเดล GRM ได้แก่ ORDINAL ของ Muraki (1988), MULTILOG ของ Thissen (1988;1991) สำหรับงานวิจัยนี้ได้ใช้โปรแกรม MULTILOG (Version 6.0) ในการวิเคราะห์ผลการเรียนรายวิชาในระดับมัธยมศึกษาตอนปลาย ซึ่งมีการตัดสินผลการเรียนแบบพหุวิภาคตามโมเดล GRM คือ 0, 1, 2, 3 และ 4 ดังนั้นผลการเรียนแต่ละรายวิชาจึงมี 5 ลำดับชั้นและผู้วิจัยได้คัดเลือกรายวิชาทั้งหมดของแต่ละแผนการเรียนให้เหลือ 50 วิชา เพื่อให้สะดวกในการวิเคราะห์กับเครื่องคอมพิวเตอร์แบบ PC ซึ่งจำกัดจำนวนไม่เกิน 50 วิชา

การเปรียบเทียบความเหมือนความแตกต่างของวิธีการเปรียบเทียบ (Kolen & Brennan, 1995)

ตารางที่ 2 การเปรียบเทียบระหว่างวิธีการปรับอควิเปอร์เซ็นไดส์, วิธีเส้นตรงและวิธีไออาร์ที

ประเด็น	วิธีอควิเปอร์เซ็นไดส์	วิธีเชิงเส้นตรง	วิธีไออาร์ที
1	แบบสอบที่นำมาเปรียบเทียบมีลักษณะโครงสร้างเหมือนกัน (same specification)	แบบสอบที่นำมาเปรียบเทียบมีลักษณะโครงสร้างเหมือนกัน (same specification)	แบบสอบที่นำมาเปรียบเทียบมีลักษณะโครงสร้างเหมือนกัน (same specification)
2	กลุ่มตัวอย่างควรมีขนาดใหญ่	เหมาะกับกลุ่มตัวอย่างขนาดเล็ก	กลุ่มตัวอย่างควรมีขนาดใหญ่
3	แบบสอบแต่ละฉบับสามารถมีความแตกต่างกันในด้านความยากได้มากกว่าวิธีเชิงเส้นตรง	แบบสอบแต่ละฉบับควรมีความยากคล้ายคลึงกัน	แบบสอบแต่ละฉบับสามารถมีความแตกต่างกันในด้านความยากได้มากกว่าวิธีเชิงเส้นตรง
4	กระบวนการวิเคราะห์ค่อนข้างซับซ้อน ยากที่จะทำให้คนทั่วไปเข้าใจได้	กระบวนการวิเคราะห์ค่อนข้างง่าย เป็นที่ยอมรับของคนทั่วไป	กระบวนการวิเคราะห์ค่อนข้างซับซ้อน ยากที่จะทำให้คนทั่วไปเข้าใจได้
5	ผลการเปรียบเทียบมีความถูกต้องทุกช่วงคะแนนแปลง	ผลการเปรียบเทียบมีความถูกต้องเฉพาะช่วงคะแนนที่อยู่ใกล้ค่าเฉลี่ย	ผลการเปรียบเทียบมีความถูกต้องทุกช่วงคะแนนแปลง

ปัจจัยที่ส่งผลต่อคุณภาพของการเปรียบเทียบ (Kolen & Brennan, 1995)

ก. หลักการทั่วไป

1. มีการกำหนดวัตถุประสงค์ของการเปรียบเทียบอย่างชัดเจน
2. มีการออกแบบการเก็บรวบรวมข้อมูลและวิธีการเปรียบเทียบ เพื่อให้การเปรียบเทียบสัมฤทธิ์ผลตามที่ต้องการ
3. มีกระบวนการประเมินคุณภาพหรือความเพียงพอของการเปรียบเทียบ

ข. การพัฒนาแบบสอบ (ทุกรูปแบบ)

1. แบบสอบที่นำมาเปรียบเทียบมีเนื้อหา, โครงสร้างและลักษณะทางสถิติเหมือนกัน
2. แบบสอบมีความยาวเหมาะสม เช่น 30 ข้อ ขึ้นไป
3. คำตอบหรือเฉลยมีความถูกต้องแน่นอน

ค. การพัฒนาแบบสอบ (กรณีรูปแบบข้อสอบร่วม)

1. ข้อสอบร่วมต้องเป็นตัวแทนของข้อสอบทั้งหมดในด้านเนื้อหาและลักษณะทางสถิติ
2. ข้อสอบร่วมมีจำนวนเพียงพอ เช่น แบบสอบ 40 ข้อ ควรมีข้อสอบร่วมอย่างน้อย 20 % และแบบสอบที่มีจำนวนข้อมากควรมีข้อสอบร่วมอย่างน้อย 30 ข้อ
3. ข้อสอบร่วมในแต่ละฉบับอยู่ในตำแหน่งเดียวกัน เช่น ข้อ 3,6,9,12,... เป็นต้น นอกจากนั้นคำถามและตัวเลือกของข้อสอบร่วมต้องเหมือนกันทุกฉบับ

ง. กลุ่มผู้สอบ

1. ผู้สอบแต่ละกลุ่มเป็นตัวแทนของผู้สอบทั้งหมด
2. จำนวนผู้สอบมีเพียงพอ
3. ในรูปแบบข้อสอบร่วม ผู้สอบแต่ละฉบับต้องมีความสามารถไม่แตกต่างกันมากนัก

จ. การบริหารการสอบ

1. แบบสอบที่ใช้ต้องเป็นความลับ ไม่รั่วไหล
2. การบริหารการสอบมีการดำเนินการอย่างเป็นมาตรฐาน

สำหรับงานวิจัยที่เกี่ยวกับการเปรียบเทียบคะแนนในประเทศไทยได้มีผู้ศึกษาเกี่ยวกับการใช้แบบสอบถาม เช่น ชูชีพ พงษ์สมบุญ (2528) ได้ศึกษาเปรียบเทียบประสิทธิภาพของการเปรียบเทียบระหว่างรูปแบบที่ใช้ผู้สอบถามกับรูปแบบที่ใช้แบบสอบถาม และเปรียบเทียบความคงที่ของวิธีการเปรียบเทียบระหว่างการเปรียบเทียบเชิงเส้นตรง การเปรียบเทียบแบบอิกვიเปอร์เซ็นไทล์และการเปรียบเทียบโดยใช้โค้งลักษณะข้อสอบ ผลการวิจัยพบว่าการเปรียบเทียบประสิทธิภาพของการเปรียบเทียบทั้ง 3 วิธี รูปแบบที่ใช้ผู้สอบถามมีประสิทธิภาพไม่แตกต่างจากรูปแบบที่ใช้แบบสอบถาม ส่วนการเปรียบเทียบความคงที่ของการเปรียบเทียบโดยรูปแบบผู้สอบถาม ปรากฏว่า การเปรียบเทียบแบบอิกვიเปอร์เซ็นไทล์และแบบโค้งลักษณะข้อสอบ มีความคงที่พอ ๆ กัน แต่ดีกว่าการเปรียบเทียบเชิงเส้นตรง ส่วนการเปรียบเทียบความคงที่โดยใช้แบบสอบถาม ปรากฏผลเช่นเดียวกับรูปแบบผู้สอบถาม

ภาวณี ศรีสุขวัฒนานันท์ (2528) ได้ศึกษาผลของการใช้รูปแบบการเปรียบเทียบ 3 วิธี คือ การเปรียบเทียบแบบอิกვიเปอร์เซ็นไทล์ การเปรียบเทียบเชิงเส้นตรง และการเปรียบเทียบตามโมเดลสามพารามิเตอร์ โดยใช้แบบสอบถามภายในที่มีความยาวต่างกัน ซึ่งคิดเป็นร้อยละของแบบสอบเปรียบเทียบ คือ 60, 40 และ 20 โดยการวิเคราะห์ความคลาดเคลื่อนมาตรฐานจากค่าดัชนีประสิทธิภาพสัมพัทธ์ (RE) ส่วนการวิเคราะห์กลุ่มสอบทานผล พิจารณาจากค่าดัชนีเปรียบเทียบความแตกต่าง (C) สรุปว่า ความยาวของแบบสอบถามมีผลกระทบต่อความแม่นยำและความเพียงพอของการเปรียบเทียบ โดยความยาวแบบสอบถามเพิ่มขึ้น ความคลาดเคลื่อนมาตรฐานของการเปรียบเทียบจะลดลง แต่ถ้าแบบสอบที่ใช้เปรียบเทียบมีความเที่ยงสูงใกล้เคียงกัน ผลของความยาวของแบบสอบถามจะไม่เด่นชัด ส่วนในสถานการณ์ของการสอบพบว่า กรณีแบบสอบคัดเลือก รูปแบบที่ให้ผลดีตามลำดับ คือ รูปแบบอิกვიเปอร์เซ็นไทล์ รูปแบบอิงทฤษฎีการตอบสนองข้อสอบและรูปแบบเชิงเส้นตรง ในกรณีแบบสอบวัดผลสัมฤทธิ์ รูปแบบที่ให้ผลดีตามลำดับ คือ รูปแบบเชิงเส้นตรง รูปแบบอิกვიเปอร์เซ็นไทล์และรูปแบบอิงทฤษฎีการตอบสนองข้อสอบ

สุนิสา รุยม่วงศรี (2534) ได้ตรวจสอบคุณภาพของการเปรียบเทียบเชิงเส้นตรง โดยใช้แบบสอบถามภายในที่มีความยาวต่างกัน 4 ขนาด คือ 10, 15, 20 และ 25 ข้อ โดยการเปรียบเทียบความคลาดเคลื่อนมาตรฐานของการเปรียบเทียบ (SEE) และดัชนีความแตกต่าง (C) จากการวิเคราะห์กลุ่มสอบทานผล สรุปผลการวิจัยได้ว่า ค่าความคลาดเคลื่อนมาตรฐานในแบบสอบถามที่ยาวกว่าให้ค่าน้อยกว่า

เทคนิคการปรับผลการเรียน

ยัง (Young : 1993) ได้ทำการรวบรวมงานวิจัยเกี่ยวกับเทคนิคการปรับผลการเรียนในรอบสองทศวรรษ ได้ทั้งหมด 13 เรื่อง ดังแสดงไว้ในตารางที่ 3 และมีรายละเอียดดังนี้

ตารางที่ 3 งานวิจัยเกี่ยวกับเทคนิคการปรับผลการเรียน

ผู้แต่งและปีที่วิจัย	แหล่งอ้างอิง	วิธีการปรับผลการเรียน
Boldt (1976)	LSAC	ปรับผลการเรียนระดับปริญญาตรี ด้วยวิธีเชิงเส้นตรง โดยพิจารณาจากค่าเฉลี่ยของ LSAT ในแต่ละสถาบัน
Braun & Szatrowski (1984a, 1984b)	JES	ปรับผลการเรียนแบบอิกวิเปอร์เซ็นไทล์
Elliott & Strenta (1975)	JEM	ปรับเชิงเส้นตรง ภายใน - ระหว่างภาควิชา
Goldman & Hewitt (1975)	JEM	คำนวณ GPA ที่คาดหวังจากผลการสอบคัดเลือก
Goldman, Schmidt, Hewitt & Fisher (1974)	AERJ	คำนวณ GPA ที่คาดหวังจากผลการสอบคัดเลือก
Goldman & Widawski (1976)	EPM	เปรียบเทียบผลการเรียนรายคู่ในแต่ละวิชา
McCornack & McLeod (1988)	JEM	วิเคราะห์ผลการเรียนแต่ละวิชา จากทั้งหมด 88 วิชา
Ramist, Lewis & McCamley (1990)	CB	เปรียบเทียบผลการเรียนกับคะแนนความถนัดทางการเรียน
Strenta & Elliott (1987)	JEM	หาค่าดัชนีการปรับผลการเรียน ระหว่างภาควิชา
Stricker, Rock & Burton (1991)	CB	ปรับผลการเรียนระดับมหาวิทยาลัยชั้นปีที่ 1
Stricker, Rock, Burton, Muraki & Jirele (1992)	ETS	เปรียบเทียบวิธีการปรับผลการเรียนหลาย ๆ วิธี โดยใช้ข้อมูลที่เดิม
Young (1990a)	JEM	ปรับตามทฤษฎี IRT โดยใช้โมเดล GRM
Young (1992)	JRE	ปรับตามโมเดล GLS โดยให้นักเรียนและวิชาเป็น main effects

ในปี 1976 โบลท์ ได้ทำการปรับผลการเรียนเฉลี่ยระดับปริญญาตรี โดยใช้ดัชนีบ่งชี้คุณภาพของสถาบัน คือ คะแนนเฉลี่ยสอบคัดเลือกเข้าโรงเรียนกฎหมาย(LSAT)ของนักศึกษา ในช่วงปี ค.ศ.1972 - 1973 และ ค.ศ.1973 - 1974 เพื่อใช้เปรียบเทียบความถูกต้องในการทำนายผลการเรียนระดับปริญญาโทในโรงเรียนสอนกฎหมาย ของ GPA ปกติ และ GPA ที่ปรับตามวิธีเชิงเส้นตรง โดยใช้เทคนิคการวิเคราะห์การถดถอยพหุคูณ แล้วประเมินค่าความตรงในการทำนายด้วยการเปรียบเทียบค่าสัมประสิทธิ์ความตรงและสัมประสิทธิ์เศษเหลือของผลการเรียนที่ปรับแล้วและผลการเรียนปกติ ผลการวิจัยพบว่า ผลการเรียนระดับปริญญาตรีที่ปรับด้วยวิธีเชิงเส้นตรงสามารถทำนายผลการเรียนระดับปริญญาโทได้ดีกว่าผลการเรียนปกติ

ในปี 1975 โกลด์แมน และเฮวิทท์ ได้ทำการศึกษาเกี่ยวกับมาตรฐานการตัดสินผลการเรียนที่แตกต่างกันในแต่ละสาขาวิชาของมหาวิทยาลัยแคลิฟอร์เนียที่ริเวอร์ไซด์ โดยเก็บข้อมูลจากนักศึกษาที่แบ่งกลุ่มตามสาขาวิชา ได้ 12 กลุ่ม กลุ่มละ 24 คน แล้วนำค่าสัมประสิทธิ์การถดถอยพหุคูณระหว่างผลการเรียนในมหาวิทยาลัยกับผลการเรียนระดับมัธยมศึกษาตอนปลาย, คะแนนความถนัดด้านภาษาและคะแนนความถนัดทางการคิดคำนวณ โดยวิเคราะห์แยกกันในแต่ละสาขา ผลการศึกษาพบว่า การตัดสินผลการเรียนของผู้สอนนั้นจะตัดสินตามระดับความสามารถของนักศึกษา นอกจากนี้โกลด์แมนและเฮวิทท์ ยังได้ศึกษาเพิ่มเติมอีก 4 วิทยาเขต คือ ที่ Davis, Irvine, Los Angeles และ San Diego ซึ่งก็ได้ผลเช่นเดิม

ในปี 1976 โกลด์แมนและวิตดอลกิ ได้พัฒนาดัชนีมาตรฐานการตัดสินผลการเรียน โดยใช้เทคนิคการเปรียบเทียบรายคู่ จากผลการเรียนใน 17 ภาควิชาของมหาวิทยาลัยแคลิฟอร์เนีย ที่ริเวอร์ไซด์ เนื่องจากว่าสถาบันที่เปิดสอนระดับปริญญาโทหลายแห่งมีการคัดเลือกนักศึกษาจาก GPA ขั้นต่ำ โกลด์แมนและวิตดอลกิจึงได้คิดวิธีการให้น้ำหนักผลการเรียนเพื่อให้เกิดความยุติธรรมและสามารถประมาณค่าผลสัมฤทธิ์ทางการเรียนของนักศึกษาได้ถูกต้องที่สุด โดยเก็บข้อมูลจากใบประมวลผลการเรียน (Transcript) ในระดับปริญญาตรีของนักศึกษาจำนวน 475 คน ในปี ค.ศ. 1973 แล้วนำมาเปรียบเทียบการตัดสินผลการเรียนในวิชาเดียวกัน ทีละ 2 ภาควิชา โดยใช้ผลการเรียนของนักศึกษาทั้งหมดเฉลี่ยในแต่ละวิชามาเปรียบเทียบกัน แล้วพิจารณามาตรฐานการตัดสินผลการเรียนจากความแตกต่างของค่าเฉลี่ย ซึ่งจะได้เมทริกซ์ความแตกต่าง ขนาด 17×17 ซึ่งแต่ละเซลล์จะเป็นตัวบ่งชี้มาตรฐานการตัดสินผลการเรียน

นอกจากนี้ยังมีงานวิจัยอีกฉบับของ โกลด์แมน ซึ่งพบว่าความไม่คงที่ของผลการเรียนในระดับมหาวิทยาลัย (GPA) จะทำให้ค่าสหสัมพันธ์ระหว่าง GPA และคะแนนสอบเข้ามหาวิทยาลัยลดลง ซึ่งสอดคล้องกับการศึกษาของ วิลลิงแฮม (1985) และ สเตรนาตา กับ เอลเลียต (1988) ที่พบว่า ผลการเรียนที่ไม่สามารถเปรียบเทียบกันได้ จะทำให้ความสัมพันธ์ระหว่าง GPA กับคะแนนสอบเข้ามหาวิทยาลัย มีค่าน้อยกว่าความสัมพันธ์ระหว่าง GPA กับ HSGPA และทำให้ประสิทธิภาพในการทำนายของ GPA ในแต่ละปี ลดลง

ในปี 1984 บรินและซาโทรสกี ได้ปรับผลการเรียนของนักศึกษาในโรงเรียนกฎหมายด้วยวิธีอควิเปอร์เซ็นไทล์ เพื่อให้อยู่ในเกณฑ์เดียวกัน โดยการสุ่มตัวอย่างสถาบันแบบแบ่งชั้น แล้วนำผลการเรียนมาเปรียบเทียบระหว่างสถาบัน การประมาณค่าความตรงของ LSAT และ GPA โดยใช้วิธีกำลังสองน้อยที่สุด (Ordinary Least Squares) สำหรับผลการเรียนที่ยังไม่ปรับและวิธีกำลังสองถ่วงน้ำหนักทั่วไป (WLS) สำหรับผลการเรียนที่ปรับแล้ว พบว่าค่าความตรงในการทำนายของผลการเรียนที่ปรับแล้วมีค่าสูง และมีความแตกต่างระหว่างสถาบันน้อยมาก

ในปี 1987 สเตรนาตาและเอลเลียต ได้ศึกษาเพิ่มเติมต่อจากงานของ โกลด์แมนและวิตอสกี โดยเก็บข้อมูลจากนักศึกษาระดับปริญญาโทของวิทยาลัยดาร์ทเมาท์ จำนวน 1,029 คน เพื่อศึกษาความสัมพันธ์ในการตัดสินผลการเรียนของผู้สอน จาก 18 วิชา โดยการเปรียบเทียบรายคู่ ผลการวิจัยพบว่า ค่าสหสัมพันธ์ระหว่างคะแนน SAT (SATV + SATM) กับ GPA ปกติ มีค่าเท่ากับ .43 ส่วนค่าสหสัมพันธ์ระหว่าง SAT กับ GPA ที่ปรับแล้ว มีค่าสูงกว่า คือ .48 โดยสามารถอธิบายความแปรปรวนของ GPA ได้เพิ่มขึ้น 4.5% นอกจากนี้ความสัมพันธ์ของ SAT กับผลการเรียนเฉลี่ยในวิชาบังคับยังสูงกว่า ความสัมพันธ์ของ SAT กับผลการเรียนเฉลี่ยรวมทุกวิชาอีกด้วย (.47 กับ .43)

และในปี 1988 สเตรนาตาและเอลเลียต ได้ศึกษาเกี่ยวกับวิธีการเปรียบเทียบรายคู่ โดยเพิ่มองค์ประกอบในการปรับที่ทำให้เกิดความแตกต่างในการตัดสินผลการเรียนของแต่ละวิชา โดยนำผลการเรียนในวิชาหนึ่งมาเปรียบเทียบกับผลการเรียนอีกวิชา ทั้งในภาควิชาเดียวกันและระหว่างภาควิชา ผลการศึกษาพบว่า ค่าสหสัมพันธ์ระหว่าง SAT กับ GPA ปกติ มีค่าเท่ากับ .43 เมื่อใช้วิธีการปรับภายในภาควิชาจะทำให้ค่าสหสัมพันธ์ เท่ากับ .45 และเมื่อใช้วิธีการปรับระหว่างภาควิชา ค่าจะสูงขึ้นเป็น .49 และค่าสหสัมพันธ์ระหว่าง SAT กับ GPA ที่

ปรับแล้ว โดยรวมมีค่าเท่ากับ .50 การใช้ SATV + SATM ในการปรับผลการเรียนจะทำให้สามารถอธิบายความแปรปรวนของ GPA ได้ถึง 6.5% นอกจากนั้นยังได้ศึกษาถึงประสิทธิภาพในการทำนายของ SAT กับ HSGPA และได้ศึกษาอิทธิพลของผลการเรียนที่ปรับแล้วในการทำนายความสามารถของนักศึกษาเพศหญิงและนักศึกษาที่เป็นชนกลุ่มน้อยอีกด้วย

ในปี 1988 แมคคอร์แนคและแมคลีฮอด ได้ศึกษาการตัดสินใจผลการเรียนของผู้สอนของมหาวิทยาลัยซานดิเอโกในสองภาคเรียน โดยเลือกเฉพาะวิชาที่มีผู้เรียนไม่น้อยกว่า 50 คน และมีแต่ละเพศอย่างน้อย 15 คน ซึ่งมีทั้งหมด 88 รายวิชา โดยนำกลุ่มตัวอย่าง จำนวน 30,410 คน ในปี 1985 กับกลุ่มตัวอย่างจำนวน 26,709 คน ในปี 1986 มาเปรียบเทียบกันและศึกษาประสิทธิภาพในการทำนายของผลการเรียนเมื่อมีความแตกต่างของเพศ ผลการศึกษาพบว่า การทำนายด้วยเฉลี่ยสะสมจะได้ผลดีในเพศหญิง แต่นักพิจารณาเฉพาะรายวิชาพบว่า ไม่มีความแตกต่างระหว่างเพศ

ในปี 1990 รมิสท์ และคณะ ได้ทำการศึกษาเกี่ยวกับผลการเรียนชั้นปีที่ 1 ของนักศึกษาจากวิทยาลัย 38 แห่ง ในปี ค.ศ.1982 และ 1985 พบว่าค่าสหสัมพันธ์ของตัวพยากรณ์ในการสอบคัดเลือกกับผลการเรียนชั้นปีที่ 1 ลดลงจากเมื่อก่อน สาเหตุเนื่องจากผลการเรียนชั้นปีที่ 1 มีความเที่ยงลดลง การหาค่าความเที่ยงได้จากการแบ่งผลการเรียนของนักศึกษาแต่ละคน โดยสุ่มออกเป็น 2 กลุ่ม แล้วหาความสัมพันธ์กัน ความเที่ยงที่คำนวณได้อยู่ในช่วง .71 - .91 มีค่าเฉลี่ย .82 แต่ผลการเรียนชั้นปีที่ 1 ยังคงมีความเที่ยงสูงกว่าผลการเรียนตลอด 4 ปี อย่างไรก็ตามงานวิจัยนี้ยังไม่ได้ปรับผลการเรียนให้อยู่บนสเกลเดียวกัน ทั้งภายในและระหว่างสถาบันแต่อย่างใด ข้อค้นพบอย่างหนึ่งที่ได้จากการศึกษานี้ คือ การเปลี่ยนแปลงของค่าสัมประสิทธิ์ความตรงของ SAT และ HSGPA มีความสัมพันธ์สูงกับการเปลี่ยนแปลงของความสามารถในการเปรียบเทียบกันได้ของผลการเรียนชั้นปีที่ 1 และการใช้ผลการเรียนแต่ละวิชาเป็นเกณฑ์แทนผลการเรียนชั้นปีที่ 1 จะช่วยลดปัญหาเรื่องความตรงของตัวพยากรณ์ในการสอบคัดเลือกได้ จุดเด่นของงานวิจัยนี้คือ การศึกษาความตรงในการทำนายของ SAT ซึ่งถือว่าเป็นประโยชน์อย่างยิ่งต่อวิธีวิทยาการในการปรับผลการเรียน

ในปี 1990 ยัง ได้ศึกษาการปรับผลการเรียนโดยใช้ทฤษฎี IRT ตามโมเดล GRM ที่เรียกว่า IRT-based GPA เพื่อหาความตรงในการทำนาย ของ SAT-V, SAT-M และ HSGPA ที่จะทำนาย GPA ด้วยวิธีการวิเคราะห์การถดถอยพหุคูณ โดยแบ่งกลุ่มตัวอย่างเป็น

3 กลุ่ม คือ วิทยาศาสตร์, สังคมศาสตร์ และ มนุษยศาสตร์ โดยใช้การวิเคราะห์องค์ประกอบ ซึ่งจะได้กลุ่มตัวอย่างที่มีความเป็นเอกมิติตามข้อตกลงเบื้องต้นของทฤษฎี IRT ผลการวิจัยสรุปได้ว่า

1. ตัวแปรพยากรณ์ที่วัดจากการสอบคัดเลือก (SAT-V, SAT-M และ HSGPA) สามารถทำนาย GPA ที่ปรับตามทฤษฎี IRT ได้ดีกว่า GPA ปกติ

2. การหาประสิทธิภาพในการทำนาย IRT - based GPA ซึ่งได้จากการวิเคราะห์การถดถอยนั้น สามารถทำนายได้สูงสุดในกลุ่มวิทยาศาสตร์ รองมาคือ กลุ่มสังคมศาสตร์ และทำนายได้ต่ำสุดในกลุ่มมนุษยศาสตร์

3. ความสามารถในการทำนายของตัวพยากรณ์ พบว่าในกลุ่มวิทยาศาสตร์ ตัวทำนายที่ดี คือ SAT-M , กลุ่มมนุษยศาสตร์ ตัวทำนายคือ SAT-V และในกลุ่มสังคมศาสตร์ มีความสามารถในการทำนายเท่ากันทั้ง SAT-V, SAT-M และ HSGPA

และในปี 1991 ยัง ได้ศึกษาเกี่ยวกับการใช้ IRT-based GPA เพื่อหาความตรงในการทำนายที่แตกต่างกันตามเชื้อชาติ และเพศ นอกจากนี้วิธีการปรับผลการเรียนของ ยัง อีกอย่างหนึ่ง คือ การปรับตามโมเดลเชิงเส้นตรง โดยให้นักเรียนและวิชาเป็น main effect (ไม่มี interaction term) แล้วพิจารณาการอธิบายความแปรปรวนสูงสุดในผลการเรียน ซึ่งชี้ให้เห็นว่า ตัวพยากรณ์ที่วัดในการสอบคัดเลือกสามารถทำนาย IRT-based GPA ได้เพิ่มขึ้น วิธีการนี้มีข้อจำกัดในการคำนวณที่ต้องวิเคราะห์เมทริกซ์ของผลการเรียนในแต่ละรายวิชาของนักเรียน

ส่วนงานวิจัยของ สตรีคเกอร์ และคณะ ในปี 1991 พวกเขาได้ทำการปรับผลการเรียนชั้นปีที่ 1 เพื่อขจัดความแตกต่างระหว่างเพศ เกี่ยวกับประสิทธิภาพในการทำนาย และลดความแตกต่างของผลการเรียนที่ได้รับในแต่ละวิชาซึ่งไม่สอดคล้องกับความสามารถของนักศึกษา การปรับความแตกต่างระหว่างวิชานี้ไม่มีผลกระทบต่อความแตกต่างในการทำนายของแต่ละเพศ อย่างไรก็ตามการปรับผลการเรียนของนักศึกษาทั้งเจ็ดสถาบัน พบว่า มีความแตกต่างในเรื่องความสามารถของนักศึกษาระหว่างสถาบัน

ในปี 1994 เคลเลอร์ และคณะ (Keller, Crouse .& Trusheim : 1994) ได้ทำการศึกษาเกี่ยวกับอิทธิพลของการปรับผลการเรียนระดับอุดมศึกษาที่ส่งผลต่อความตรงในการทำนายของคะแนนสอบความถนัดทางการเรียน ผลการศึกษาพบว่า ตัวแปรทำนาย ได้แก่ ผลการเรียนระดับมัธยมศึกษาตอนปลายสามารถอธิบายความแปรปรวนในผลการเรียนระดับ

อุดมศึกษาที่ยังไม่ปรับ (Unadjusted FGPA) และผลการเรียนระดับอุดมศึกษาที่ปรับแล้ว (Adjusted FGPA) เท่ากับ 0.250 และ 0.329 ตามลำดับ หากให้ผลการเรียนระดับมัธยมศึกษาตอนปลายร่วมกับคะแนนสอบความถนัดทางการเรียนเป็นตัวแปรทำนาย สามารถอธิบายความแปรปรวนในผลการเรียนระดับอุดมศึกษาที่ยังไม่ปรับและผลการเรียนระดับอุดมศึกษาที่ปรับแล้ว เท่ากับ 0.290 และ 0.416 ตามลำดับ สรุปได้ว่า ประสิทธิภาพในการทำนายจะมีค่าสูงขึ้น หากใช้ผลการเรียนระดับอุดมศึกษาที่ปรับแล้วเป็นเกณฑ์ (criterion) และเมื่อใช้ผลการเรียนระดับมัธยมศึกษาตอนปลายร่วมกับคะแนนสอบความถนัดทางการเรียนเป็นตัวแปรทำนายจะสามารถพยากรณ์ผลการเรียนในระดับอุดมศึกษาได้ดีกว่าใช้ผลการเรียนระดับมัธยมศึกษาตอนปลายเพียงอย่างเดียว

ในปี 1995 ยัง (Young : 1995) ได้ทำการศึกษาเปรียบเทียบวิธีการปรับผลการเรียนระหว่างวิธี IRT กับวิธีการปรับเชิงเส้นตรง (กำลังสองน้อยที่สุด - LS) ที่จะทำให้ประสิทธิภาพในการทำนายผลการเรียนในโรงเรียนกฎหมายสูงขึ้น โดยทำการเก็บข้อมูลในโรงเรียนกฎหมาย 4 แห่งของสหรัฐอเมริกา ผลการศึกษาพบว่า วิธีการปรับทั้งสองวิธีทำให้ประสิทธิภาพในการทำนายของคะแนนสอบคัดเลือกเข้าโรงเรียนกฎหมาย (LSAT) และผลการเรียนระดับปริญญาตรีสูงขึ้น โดยเฉพาะในโรงเรียนที่ 4 มีความสัมพันธ์ระหว่างคะแนนสอบเข้าโรงเรียนกฎหมายกับผลการเรียนในโรงเรียนกฎหมายสูงกว่าแห่งอื่น ๆ ซึ่งเป็นเพราะความแปรปรวนของ LSAT ในโรงเรียนที่ 4 มีค่ามากที่สุดนั่นเอง

จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวกับการปรับผลการเรียนดังกล่าวมา พอสรุปได้ว่า วิธีการปรับผลการเรียนที่ดำเนินการในต่างประเทศนั้นมีหลากหลายวิธี ซึ่งผลการเรียนหรือแต้มเฉลี่ยสะสมที่ปรับจากแต่ละวิธีจะมีความคงที่และมีประสิทธิภาพในการทำนายผลสัมฤทธิ์ทางการเรียนได้ดีกว่าผลการเรียนหรือแต้มเฉลี่ยสะสมปกติ แต่ยังไม่สามารถสรุปได้ว่า วิธีการปรับผลการเรียนแบบใดให้ผลดีที่สุด ดังนั้นในการตัดสินใจเลือกวิธีการปรับผลการเรียนมาประยุกต์ใช้ในประเทศไทยนั้น คงต้องวางแผนและดำเนินการให้มีความเหมาะสมกับบริบทของไทยซึ่งมีระบบการศึกษาแตกต่างกับต่างประเทศ ส่วนวิธีการปรับผลการเรียนที่นำมาใช้ในปัจจุบันก็ยังมีข้อสงสัยเกี่ยวกับความเชื่อถือได้ เนื่องจากมาตรฐานการตัดสินผลการเรียนของแต่ละโรงเรียนมีความแตกต่างกัน การพิจารณาเฉพาะแต้มเฉลี่ยสะสมจึงอาจก่อให้เกิดความไม่ยุติธรรมได้ ควรที่จะมีการปรับด้วยแบบสอบมาตรฐานอย่างใดอย่างหนึ่ง เช่น แบบสอบความถนัดทางการเรียน