



บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

ในการนำเสนอวรรณคดีที่เกี่ยวข้อง จะจำแนกเป็น 5 ส่วนดังนี้

1. ประวัติการฝึกหัดครู และการคัดเลือกนักศึกษาครู
2. ทฤษฎีการให้คะแนนข้อสอบ
3. การวิเคราะห์ถดถอยพหุคูณ
4. มโนทัศน์ว่าด้วยทฤษฎีการตอบสนองข้อสอบ
5. งานวิจัยที่เกี่ยวข้อง

รายละเอียดค่าค้นดังนี้

1. ประวัติการฝึกหัดครู และการคัดเลือกนักศึกษาครูในวิทยาลัยครู

การฝึกหัดครู ได้เริ่มดำเนินการมาตั้งแต่ พ.ศ. 2435 โดยกรมศึกษาธิการได้ตั้งโรงเรียนฝึกหัดครูแห่งแรกขึ้นที่โรงเรียนเต็ก เรียกว่า โรงเรียนฝึกหัดอาจารย์ ต่อมาได้ย้ายไปอยู่ที่เทพศิรินทร์ราชวาส และในปี พ.ศ. 2446 ก็ได้ขยายหลักสูตรสูงขึ้นไปจนถึงชั้นฝึกหัดครูมัธยม ภายหลังโรงเรียนได้รวมเข้ากับโรงเรียนฝึกหัดครูฝั่งตะวันตกที่บ้านสมเด็จเจ้าพระยา จนเมื่อได้มีการสถาปนาจุฬาลงกรณ์มหาวิทยาลัยขึ้นเมื่อปี พ.ศ. 2459 การฝึกหัดครู จึงกลายเป็นส่วนหนึ่งของจุฬาลงกรณ์มหาวิทยาลัย โรงเรียนฝึกหัดครูที่กล่าวมานี้ เป็นโรงเรียนฝึกหัดครูชาย ส่วนโรงเรียนฝึกหัดครูหญิงนั้น ก็ได้จัดตั้งขึ้นแห่งแรกในปี พ.ศ. 2456 ที่โรงเรียนเบญจมราชาลัย

ต่อมาในปี พ.ศ. 2464 การฝึกหัดครูได้ย้ายมาขึ้นกับกระทรวงศึกษาธิการตามเดิม โดยเปิดหลักสูตรฝึกหัดครูมูล และครูประถมขึ้นที่โรงเรียนฝึกหัดครูวัดบวรนิเวศ ซึ่งต่อมาโรงเรียนนี้ได้เปลี่ยนชื่อและย้ายไปอีกหลายแห่ง จนกระทั่งในปี พ.ศ. 2499 จึงได้ย้ายไปที่ อำเภอบางเขน และเปลี่ยนชื่อเป็นโรงเรียนฝึกหัดครูพระนคร สำหรับหลักสูตรฝึกหัดครูมัธยมก็ได้เปิดสอนที่โรงเรียนสวนกุหลาบวิทยาลัย ต่อมากระทรวงศึกษาธิการ ได้มอบให้จุฬาลงกรณ์ มหาวิทยาลัย

เป็นผู้จัดโดยได้ตั้งแผนกฝึกหัดครูมัธยมขึ้น ในคณะอักษรศาสตร์ และวิทยาศาสตร์ เมื่อปี พ.ศ. 2473

การฝึกหัดครู ซึ่งแต่เดิมมีฐานะเป็นแผนก ขึ้นอยู่กับกรมวิสามัญศึกษาที่ได้รับการยกฐานะขึ้นเป็นกรมการฝึกหัดครู เมื่อปี พ.ศ.2497 โดยทำหน้าที่ 2 ประการ คือการผลิตครูในสถาบันฝึกหัดครูให้ได้ทั้งปริมาณและคุณภาพ และช่วยกรมต่าง ๆ อบรมครูประจำการให้มีวุฒิสูงขึ้น การฝึกหัดครู ได้รับการปรับปรุงให้ก้าวหน้ายิ่งขึ้นตามลำดับ ได้มีการยกระดับวิद्यฐานะของครูให้สูงขึ้นจนถึงขั้นปริญญาและเทียบเท่าสาขาวิชาอื่น มีการปรับปรุงหลักสูตร วิธีการสอน อุปกรณ์การสอน การวัดผลการศึกษา การส่งเสริมฐานะครู ตลอดจนการบริหารให้เหมาะสมกับสภาพความเปลี่ยนแปลงของบ้านเมืองตลอดมา (บำรุง กลักเจริญ, 2506)

ระยะก่อนสงครามโลกครั้งที่ 2 การพัฒนาการฝึกหัดครูเป็นไปอย่างช้า ๆ มีโรงเรียนฝึกหัดครู 2-3 แห่ง จำนวนนักเรียนก็น้อย รัฐบาลต้องให้ทุนอุดหนุนแก่นักเรียนฝึกหัดครูตลอดมา การจัดโรงเรียนก็จัดแบบเดินเรียนแยกเป็นโรงเรียนหญิง โรงเรียนชาย ไม่มีการจัดแบบสหศึกษา การจัดโรงเรียนประจำเพิ่งมีขึ้นครั้งแรก ใน พ.ศ.2449 ตั้งแต่นั้นมาโรงเรียนฝึกหัดครูที่ตั้งขึ้น ส่วนมากก็เป็นโรงเรียนประจำ สำหรับระดับของการศึกษาดังแต่ระยะแรกจนถึง พ.ศ.2475 ก็มีการจัดฝึกหัดครู ประกาศนียบัตรมณฑล ครูมูล ครูประถม และครูมัธยม ภายหลังการเปลี่ยนแปลงการปกครองแล้ว จึงได้มีการฝึกหัดครูประจำบาล และครูประกาศนียบัตรเพิ่มขึ้นระดับทั้งสองนี้มีวุฒิต่ำกว่าครูมูลลงไปอีก ทั้งนี้ เพราะต้องการผลิตรู้ออกไปสอนในชนบท ซึ่งการขาดแคลนยังมีอยู่มาก

นับตั้งแต่สงครามโลกครั้งที่ 2 เป็นต้นมา การฝึกหัดครูของไทยได้เจริญก้าวหน้ามาก โดยเฉพาะอย่างยิ่ง ภายหลังที่ได้มีการจัดตั้งกรมการฝึกหัดครูขึ้นแล้วรัฐบาลเองก็มีนโยบายเร่งขยายการศึกษาเพื่อสนองความต้องการของประชาชนและของประเทศ นอกจากนี้ด้วยความช่วยเหลือจากต่างประเทศ ได้มีการปรับปรุงงานการฝึกหัดครู เป็นการใหญ่ทั้งในด้านปริมาณและคุณภาพ มีการยกระดับการฝึกหัดครูให้สูงขึ้นถึงระดับปริญญา มีการปรับปรุงหลักสูตรฝึกหัดครู วิธีสอน อุปกรณ์การสอน การวัดผลการศึกษา การส่งเสริมฐานะครู ตลอดจนให้บริการครูให้เหมาะสมแก่กาลสมัย (ประยูร ศรีประสาธน์ 2515: 19-36)

ประวัติการคัดเลือกนักศึกษาของกรมการฝึกหัดครู

สำหรับในด้านการคัดเลือกผู้ที่เหมาะสมเพื่อเข้าเรียนในสถาบันฝึกหัดครูนั้นในระยะแรก รัฐบาลต้องให้ทุนอุดหนุนแก่ผู้เรียน (ประยูร ศรีประสาธน์ 2515: 19) โดยที่ผู้เรียนเป็นผู้ที่มีความสามารถสูง แต่ระยะต่อมาสถาบันฝึกหัดครูเร่งผลิตครูจำนวนมาก ประกอบกับต้องรับภาระ นักศึกษาส่วนหนึ่งที่ไม่มีโอกาสเข้าศึกษาในระดับอุดมศึกษาแห่งอื่นหรือวิชาชีพอื่น อย่างไรก็ตาม แม้ว่าจะให้มีการสอบคัดเลือกโดยใช้แบบวัดหลาย ๆ ชุด เช่น ความรู้ทางวิชาการ บุคลิกภาพ และทัศนคติต่อวิชาชีพครูแต่ไม่ได้ช่วยให้ผลผลิตครูมีประสิทธิภาพเท่าที่ควร ดังที่ได้เสนอไว้ในบทหน้า และนอกเหนือไปกว่านั้น ความเป็นมาตรฐานของเครื่องมือวัดที่วิทยาลัยครู แต่ละแห่งสร้างขึ้นนั้นก็ยังเป็นที่น่าสงสัย อีกทั้งต้องรับผู้เข้าเรียนทางวิชาชีพครูด้วยเหตุผลพิเศษอื่น ๆ อีกโศกหนึ่ง

การสอบคัดเลือกนักเรียนฝึกหัดครูในสถาบันการศึกษา สังกัดกรมการฝึกหัดครูนั้น สามารถรวบรวมหลักฐานในระยะแรก ๆ ได้จากการสอบถามผู้ที่ได้ผ่านการสอบคัดเลือกเข้าเรียนในโรงเรียนฝึกหัดครู พอสรุปได้ว่า ในระยะก่อน พ.ศ. 2498 การคัดเลือกเข้าเรียนในโรงเรียนฝึกหัดครู ระดับ ป.ป. จะแบ่งเป็น 2 ประเภท ประเภททุนจังหวัด ให้แต่ละจังหวัดเป็นผู้จัดสอบคัดเลือก คัดเลือกจากผู้มีผลการเรียนดี และการทำงานดี ส่งเข้าเรียนในโรงเรียนฝึกหัดครูบ้านสมเด็จเจ้าพระยา และโรงเรียนฝึกหัดครูสวนสุนันทา ส่วนประเภททุนส่วนกลางนั้น ให้ใช้ ข้อสอบวิชาหลัก 4 วิชา คือ ภาษาไทย ภาษาอังกฤษ คณิตศาสตร์ วิทยาศาสตร์ และสอบสัมภาษณ์ ผู้ที่ได้รับการคัดเลือกเข้าเรียน จะเรียนที่โรงเรียนฝึกหัดครูพระนคร และโรงเรียนฝึกหัดครูเพชรบุรีวิทยาลงกรณ์ ส่วนการคัดเลือกเข้าเรียนระดับ ป.ม. ให้ส่วนกลางเป็นผู้จัดสอบวิชาหลัก 4 วิชา และสัมภาษณ์ด้วยเช่นเดียวกับระดับ ป.ป. และส่งเรียนที่โรงเรียนฝึกหัดครูส่วนกลาง เท่านั้น

ปีการศึกษา 2498 เปิดหลักสูตรการฝึกหัดครูประกาศนียบัตรวิชาการศึกษา (ป.กศ.) เป็นรุ่นแรก วิทยาลัยครูในส่วนกลางเป็นผู้สอบคัดเลือกนักเรียนเข้าเรียนเองส่วนหนึ่ง อีกส่วนหนึ่งเป็นนักเรียนที่ได้คะแนนชั้น ม.6 ยอดเยี่ยม เป็นที่ 1 ที่ 2 ของจังหวัด จังหวัดจะให้ทุนการศึกษาส่งมาเรียน ในวิทยาลัยครูส่วนกลาง โดยส่งเข้าเรียนที่วิทยาลัยครูบ้านสมเด็จเจ้าพระยาสำหรับนักเรียนชาย และส่งเข้าเรียนที่วิทยาลัยครูสวนสุนันทา สำหรับนักเรียนหญิง

สำหรับในวิทยาลัยครูส่วนภูมิภาคยังให้โควตาแก่จังหวัด เพื่อให้จังหวัดดำเนินการรับสมัครและสอบคัดเลือกให้ อาจมีบางจังหวัดให้วิทยาลัยครูดำเนินการสอบข้อเขียนให้ (ออกข้อสอบและตรวจข้อสอบ ทำผลการสอบ) วิชาที่ใช้ในการคัดเลือกในระยะนั้น ได้แก่ คณิตศาสตร์ วิทยาศาสตร์ ภาษาไทย ภาษาอังกฤษ สังคมศึกษา และลัทธิภาษณ์

ปีการศึกษา 2508 ผู้ที่สอบคัดเลือกในระดับประกาศนียบัตรวิชาการศึกษาระดับชั้นสูง (ป.กศ.) ของวิทยาลัยครูทั่วประเทศ ใช้ข้อสอบเดียวกัน โดยกรมการฝึกหัดครูเป็นผู้จัดทำข้อสอบส่งไปให้โดยในส่วนกลางยังคงดำเนินการเช่นปี 2507 ส่วนภูมิภาคให้จังหวัดเป็นผู้รับสมัคร และทำการสอบสัมภาษณ์ให้เรียบร้อยก่อน จึงสอบคัดเลือกข้อเขียนที่กรมการฝึกหัดครูจัดส่งไปให้จังหวัด วิชาที่ใช้ในการสอบคัดเลือกข้อเขียนคือ ความถนัดทางการเรียน คณิตศาสตร์ ภาษาไทย ภาษาอังกฤษ และวิทยาศาสตร์

สำหรับการคัดเลือกในระดับประกาศนียบัตรวิชาการศึกษาระดับชั้นสูง (ป.กศ.ชั้นสูง) ให้วิทยาลัยครูส่วนกลางคัดเลือกส่วนหนึ่ง จากนักเรียน ป.กศ.ปีที่ 2 ที่มีผลการเรียนดี และการงานดีของวิทยาลัย จำนวน 35% ของจำนวนที่จะรับเข้าเรียน ป.กศ.ชั้นสูงให้ โดยไม่ต้องสอบคัดเลือก และอีกส่วนหนึ่งใช้วิธีสอบคัดเลือกโดยกรมการฝึกหัดครู เป็นผู้ดำเนินการสอบคัดเลือกให้ เช่นเดียวกับการคัดเลือกนักเรียนเข้าเรียนระดับ ป.กศ.

ในส่วนภูมิภาค ใช้การคัดเลือกนักศึกษา ป.กศ. ปีที่ 2 ของวิทยาลัยชั้นเรียนในระดับ ป.กศ.ชั้นสูงในวิทยาลัยครูที่เปิดสอนในระดับ ป.กศ.ชั้นสูง ทั้งในส่วนกลาง และส่วนภูมิภาค

ปีการศึกษา 2517 มีการเปิดสอนในระดับปริญญาตรีขึ้นในวิทยาลัยครูทั้งส่วนกลาง และส่วนภูมิภาค รวม 17 แห่ง โดยกรมการฝึกหัดครูเป็นผู้ทำข้อสอบคัดเลือกให้สอบเฉพาะวิชาเอก 2 ฉบับ ฉบับแรกใช้เวลา 3 ชั่วโมง ฉบับที่สองใช้เวลาสอบ 2 ชั่วโมง กับการสอบสัมภาษณ์ ซึ่งวิทยาลัยครูเป็นผู้ดำเนินการเอง

ปีการศึกษา 2518 การสอบคัดเลือกนักเรียนเข้าเรียนระดับ ป.กศ. และ ป.กศ.ชั้นสูง กรมการฝึกหัดครูให้วิทยาลัยครูเป็นผู้จัดทำข้อสอบคัดเลือกทำผลการสอบเอง สำหรับระดับปริญญาตรี กรมการฝึกหัดครูจัดทำข้อสอบให้อย่างเต็ม

ปีการศึกษา 2519 ระดับ ป.กศ. ได้มีการเปลี่ยนแปลงวิธีการรับนักศึกษาใหม่ โดยไม่มีการสอบข้อเขียน แต่มอบให้สภาตำบลเป็นผู้ทำการคัดเลือกจากนักเรียน ม.ศ.3 ที่สนใจ จะสมัครเรียนครู และพิจารณาผลการสอบไล่ของชั้น ม.ศ.3 ที่อยู่ในเกณฑ์สูง รวมกับการสอบสัมภาษณ์ตามเกณฑ์กรมการฝึกหัดครูกำหนด ส่วนระดับ ป.กศ.ชั้นสูง และระดับปริญญาตรี ยังคงให้การสอบคัดเลือกโดยสอบข้อเขียนและสอบสัมภาษณ์ เช่นเดิม และวิทยาลัยครูแต่ละแห่ง เป็นผู้ทำข้อสอบคัดเลือกเองทั้งสองระดับ

ปีการศึกษา 2520 การคัดเลือกนักเรียนครูในระดับ ป.กศ. กลับมาใช้การสอบคัดเลือกเช่นเดิม แต่กรมการฝึกหัดครูได้เปลี่ยนแปลงเครื่องมือที่ใช้ในการสอบคัดเลือกใหม่ โดยกรมการฝึกหัดครู จัดทำแบบตรวจสอบคุณลักษณะพื้นฐานของการเป็นครู ซึ่งเป็นเรื่องวัดทางด้านความรู้สึทางจิตใจและบุคลิกภาพมาใช้แทนแบบทดสอบวัดความถนัดทางการเรียน แบบตรวจสอบคุณลักษณะพื้นฐานของการเป็นครูนี้ ประกอบด้วยแบบตรวจสอบย่อย 2 ชุด (Subtest) คือแบบตรวจสอบบุคลิกภาพของการเป็นครู กับแบบตรวจสอบความสนใจในอาชีพครู แบบตรวจสอบดังกล่าว กรมการฝึกหัดครู ได้ดำเนินการสร้างและจัดส่งให้กับวิทยาลัยครูทุกแห่งพร้อมกันนี้ ก็ให้มีการปรับปรุงการสอบสัมภาษณ์โดยกำหนดให้เป็นแบบมีโครงสร้าง (Structure Interview) มากขึ้น ส่วนการสอบคัดเลือกในระดับ ป.กศ. ชั้นสูง และระดับปริญญาตรี ก็ยังคงใช้แบบทดสอบที่แต่ละวิทยาลัยครูจัดสร้างขึ้นใช้เอง

ปีการศึกษา 2521 การคัดเลือกนักศึกษาครูในระดับ ป.กศ., ป.กศ.ชั้นสูง และระดับปริญญาตรี ก็ยังคงใช้การสอบคัดเลือกและดำเนินการเช่นเดียวกับปีการศึกษา 2520 เว้นแต่ให้ขยายการใช้แบบตรวจสอบคุณลักษณะพื้นฐานการเป็นครูมาในระดับ ป.กศ.ชั้นสูงด้วย โดยนำมาใช้แทนแบบทดสอบความถนัดทางการเรียน

ปีการศึกษา 2522 การคัดเลือกนักเรียนครูระดับ ป.กศ. ก็มีการเปลี่ยนแปลงหลักการในการใช้เครื่องมือสอบคัดเลือกอีก โดยกรมการฝึกหัดครูเปิดโอกาสให้วิทยาลัยครูที่รับนักเรียนเข้าศึกษาในระดับนี้เลือกที่จะใช้แบบตรวจสอบคุณลักษณะพื้นฐานของความเป็นครู หรือจะไม่ใช้ก็ได้ ส่วนในระดับ ป.กศ.ชั้นสูง ยังคงกำหนดให้ใช้และขยายไปใช้กับระดับปริญญาตรีอีกด้วย

ปีการศึกษา 2524 การคัดเลือกนักศึกษาครู ทั้ง 3 ระดับ ดำเนินการเช่นเดียวกับปี พ.ศ.2522 และต่อมาในปีการศึกษา 2525 เป็นต้นมา ใช้หลักเกณฑ์การคัดเลือกโดยแบ่งเป็นสองส่วนคือ ส่วนแรกคัดเลือกจากนักเรียนชั้นสูงสุดระดับมัธยมศึกษา ซึ่งมีผลการเรียนดี และอยู่ในระดับร้อยละ 20 แรกของจำนวนนักเรียนในแต่ละโปรแกรมการเรียน โดยคิดส่วนนี้ร้อยละ 25 ของจำนวนนักเรียนแต่ละสาขาวิชาเอกที่จะรับเข้าศึกษาในวิทยาลัยครู

2. ทฤษฎีการให้คะแนนข้อสอบ

วิธีการกำหนดค่าน้ำหนักคะแนนรายข้อ ให้แบบสอบฉบับหนึ่ง ตัวแปรของคะแนนรวมให้แก่ คะแนนของข้อสอบรายข้อ และการให้คะแนนก็เหมือนกับการคิดค่าน้ำหนักคะแนนของแบบสอบหลายฉบับที่นำมารวมกัน (Wang and Stanley, 1970: 672; Stanley and Wang, 1970: 30) ซึ่งมีแนวคิดดังนี้

ถ้ากำหนดให้

$$X_1, X_2, \dots, X_n = \text{คะแนนของข้อสอบแต่ละข้อ}$$

$$W_1, W_2, \dots, W_n = \text{ค่าน้ำหนักคะแนนของข้อสอบแต่ละข้อ}$$

ดังนั้น

$$\begin{aligned} \text{Var}(X_1 + X_2 + \dots + X_n) &= \text{Var}(X_1) + \text{Var}(X_2) + \dots \\ &+ \text{Var}(X_n) + 2\text{Cov}(X_1X_2) + 2\text{Cov}(X_1X_3) + \dots + \\ &+ 2\text{Cov}(X_{n-1}X_n) \end{aligned} \quad (1)$$

เมื่อเมตริกมีขนาด $n \times n$ แล้ว ความแปรปรวนที่ข้อสอบข้อที่ i (c_i) มีส่วนในความแปรปรวนทั้งหมด มีดังนี้

$$C_1 = \text{Cov}(X_1X_1) + \text{Cov}(X_1X_2) + \dots + \text{Cov}(X_1X_{i-1}) + \quad (2)$$

$$\text{Var}(X_1) + \text{Cov}(X_1X_{i+1}) + \dots + \text{Cov}(X_1X_n)$$

ในกรณีที่ข้อสอบมีค่าน้ำหนักคะแนน ดังนั้น ความแปรปรวนของคะแนนรวมเป็นดังนี้

$$\text{Var}(W_1X_1 + W_2X_2 + \dots + W_nX_n) = W_1^2 \text{Var}(X_1) + \quad (3)$$

$$W_2^2 \text{Var}(X_2) + \dots + W_n^2 \text{Var}(X_n) + 2W_1W_2\text{Cov}(X_1X_2) +$$

$$2W_1W_3\text{Cov}(X_1X_3) + \dots + 2W_{n-1}W_n\text{Cov}(X_{n-1}X_n)$$

และความแปรปรวนของข้อสอบรายข้อที่มีส่วนร่วมในความแปรปรวนทั้งหมดของคะแนนรวมเป็นดังนี้

$$C_i = W_iW_1\text{Cov}(X_iX_1) + W_iW_2\text{Cov}(X_iX_2) + \dots \quad (4)$$

$$+ W_iW_{i-1}\text{Cov}(X_iX_{i-1}) + W_i^2 \text{Var}(X_i) + \dots$$

$$+ W_iW_{i+1}\text{Cov}(X_iX_{i+1}) + W_iW_n\text{Cov}(X_iX_n)$$

แต่เนื่องจาก $\text{Cov}(X_iX_j) = \rho_{ij}\sigma_i\sigma_j$ ดังนั้น

$$C_i = W_iW_1\rho_{i1}\sigma_i\sigma_1 + W_iW_2\rho_{i2}\sigma_i\sigma_2 + \dots + \quad (5)$$

$$W_iW_{i-1}\rho_{i,i-1}\sigma_i\sigma_{i-1} + W_i^2\sigma_i^2 + W_iW_{i+1}\rho_{i,i+1}\sigma_i\sigma_{i+1} + \dots$$

$$+ W_iW_n\rho_{in}\sigma_i\sigma_n$$

จากสมการที่ 5 จะเห็นได้ว่า ข้อสอบแต่ละข้อมีส่วนร่วมในความแปรปรวนร่วมของคะแนนรวมทั้งหมดมาน้อยเพียงใดขึ้นอยู่กับปัจจัยต่อไปนี้

1. ค่าน้ำหนักคะแนนของข้อสอบแต่ละข้อที่กำหนดขึ้น
2. ค่าความแปรปรวนของคะแนนของข้อสอบ
3. ความสัมพันธ์ระหว่างข้อสอบนั้น กับข้อสอบต่าง ๆ ที่เหลือของแบบสอบ
4. ค่าเบี่ยงเบนมาตรฐานของข้อสอบอื่น ๆ ของแบบสอบ

แต่ในกรณีที่คะแนนของข้อสอบแต่ละข้อเป็นคะแนนมาตรฐาน คือหารทุก ๆ ค่าด้วยค่าเบี่ยงเบนมาตรฐานของข้อสอบแต่ละข้อ (C_i) สมการที่ 5 จะมีลักษณะดังนี้

$$C_i + W_1 W_1 P_{i1} + W_1 W_2 P_{i2} + \dots + W_1 W_{i-1} P_{i,i-1} + \quad (6)$$

$$W_1^2 + W_1 W_{i+1} P_{i,i+1} + \dots + W_1 W_n P_{in}$$

ดังนั้น ในกรณีที่ข้อสอบแต่ละข้อเป็นคะแนนมาตรฐาน ค่าความแปรปรวนร่วมของข้อสอบแต่ละข้อจะมากน้อยเพียงใดขึ้นอยู่กับปัจจัย 2 อย่างคือ

1. ค่าน้ำหนักคะแนนของข้อสอบข้อนั้น ๆ ที่กำหนดขึ้น
2. ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างข้อสอบแต่ละข้อ

ปกติแล้ว ถ้าคะแนนของข้อสอบแต่ละข้อเป็นคะแนนมาตรฐาน คะแนนของข้อสอบนั้น โดยประมาณแล้วเท่ากับ 1 บวกกับ ผลบวกของค่าสหสัมพันธ์ระหว่างข้อสอบแต่ละข้อ กล่าวคือ

$$C_i = 1 + \sum_{j=1}^n P_{ij} \quad (7)$$

$$\text{และ } C_i = 1 + (n-1)\bar{P}_{ij} \quad (8)$$

เมื่อ $\bar{P}_{ij} = \bar{X}$ ของค่าสหสัมพันธ์ของข้อสอบที่เหลือ (n-1)

ดังนั้น จึงกล่าวได้ว่าปกติแล้ว คำนี้นักคะแนนรายข้อที่กำหนดให้ไม่เท่ากับค่าที่กำหนดให้โดยตรง เว้นแต่จะทำให้คะแนนเป็นคะแนนมาตรฐาน และค่าความสัมพันธ์ซึ่งกันและกันของข้อสอบมีค่าเท่า ๆ กัน ตามทฤษฎีแล้ว การกำหนดค่านี้นักคะแนนรายข้อทำให้แบบสอบมีความแปรปรวนของคะแนนมากกว่าการคิดคะแนนวิธีประเพณีนิยม ซึ่งจะสังเกตได้จากสมการที่ 5 และ 6 ดังนั้น จึงเชื่อว่าการคิดค่านี้นักคะแนนเป็นวิธีหนึ่งที่จะทำให้แบบสอบมีความเที่ยงและความตรงสูงกว่าค่าความเที่ยงและความตรงที่คิดจากคะแนนวิธีประเพณีนิยมได้ (Wang and Stanley, 1970: 663) ดังนั้น จึงมีการศึกษาเรื่องดังกล่าวนี้เรื่อยมา

3. มโนทัศน์การวิเคราะห์ถดถอยพหุคูณ (Multiple Regression Analysis)

การวิเคราะห์ถดถอยพหุคูณเป็นเทคนิควิธีที่ขยายจากแนวความคิดเดิมของความสัมพันธ์ระหว่างตัวแปรอิสระและตัวแปรตามกับการพยากรณ์ค่าของตัวแปรดังกล่าว จากสมการถดถอยพหุคูณ ศึกษาถึงบทบาทหรืออิทธิพลของตัวแปรอิสระมากกว่า 1 ตัวแปรขึ้นไปที่มีต่อตัวแปรตาม โดยใช้วิธีการหรือหลักความสัมพันธ์ (correlation) และการถดถอย (regression) หน้าที่ประการสำคัญของการถดถอยพหุคูณก็คือ การช่วย "อธิบาย" ความแปรปรวน หรือการกระจายของตัวแปรตามโดยประเมินจากการรู้ความแปรปรวนของตัวแปรอิสระ งานวิจัยทางการศึกษาส่วนใหญ่พยายามที่จะแสวงหาคำอธิบายความแปรปรวนของสัมฤทธิ์ผลทางการเรียนไทยศึกษาสิ่งที่มีความสัมพันธ์ต่อสัมฤทธิ์ผลนั้นหลายประการ

วัตถุประสงค์ของการวิเคราะห์ถดถอยพหุคูณนี้ เพื่อที่จะคำนวณค่านี้นัก หรือสัมประสิทธิ์การถดถอย (beta weight หรือ regression coefficient) ของตัวแปรอิสระแต่ละตัวแปร ตลอดจนค่าคงที่เฉพาะของสมการ (beta zero หรือ intercept) เพื่อให้ผลที่คำนวณได้จากสมการถดถอย ต่อไปนี้

$$Y = b_0 + b_1x_1 + b_2x_2 + \dots + b_kx_k$$

เป็นค่าพยากรณ์ที่มีค่ากำลังสองของความคลาดเคลื่อนในการพยากรณ์น้อยที่สุด จากความรู้ค่าตัวแปรอิสระ $x_1, x_2, x_3, \dots, x_k$ ในกรณีที่มีตัวแปรอิสระ k ตัวแปร, เราสามารถทำนายหรือหาค่าของตัวแปรตาม (y) ได้

3.1 ลักษณะของการวิเคราะห์ถดถอยพหุคูณ

ตัวแปรสำหรับการวิเคราะห์

การศึกษาการถดถอยพหุคูณเพื่อการพยากรณ์ ต้องใช้ตัวแปรอิสระจำนวนหนึ่งจากตัวแปรที่ต้องการศึกษาทั้งหมด วิธีการนี้จะใช้ให้พิจารณาเท่าที่ข้อมูลที่ทำการศึกษาเป็นข้อมูลที่มีลักษณะต่อเนื่อง (continuous data) เป็นข้อมูลที่จัดเรียงลำดับ (ranked data) หรือเป็นข้อมูลที่จัดเป็นพวกเป็นหมู่ (categorical data) โดยที่ในแต่ละกลุ่มหรือระดับของข้อมูลที่จัดเป็นพวกเป็นหมู่นั้นประกอบด้วย ตัวเลขสองประเภท คือ เลข 1 และเลข 0 ในการศึกษาข้อมูลที่จัดเป็นพวกเป็นหมู่ โดยเฉพาะอย่างยิ่งข้อมูลที่สามารถจัดแบ่งได้หลาย ๆ ระดับ อาจใช้ข้อมูลในแต่ละระดับ เป็นตัวแปรอิสระสำหรับการวิเคราะห์ถดถอยพหุคูณ โดยเลือกระดับของตัวแปรจำนวนเท่ากับขั้นแห่งความเป็นอิสระ (degree of freedom) ของตัวแปรนั้น ๆ ตัวอย่างเช่น ศึกษาตัวแปรเกี่ยวกับอายุที่ประกอบด้วยกลุ่มอายุ 5 ระดับต้องใช้ระดับของกลุ่มอายุในการศึกษาจำนวน 4 ระดับใช้ข้อมูลทั้ง 4 ระดับนี้ เป็นตัวแทนสำหรับตัวแปรที่เกี่ยวกับอายุทั้งหมดในการวิเคราะห์ถดถอยพหุคูณอีกวิธีหนึ่งอาจทำได้โดยการศึกษาระดับของตัวแปรที่ต้องการศึกษาทุกระดับ และกำหนดตัวเลขให้คุณลักษณะของระดับที่ต้องการศึกษาในแต่ละระดับมีค่าเป็น 1 กับคุณลักษณะอื่นที่ไม่ต้องการศึกษามีค่าเป็น 0

3.2 ข้อตกลงเบื้องต้นของการวิเคราะห์ถดถอยพหุคูณ

1. การวิเคราะห์ถดถอยในทุกลักษณะ จะสมมติว่า คะแนนของตัวแปรตาม (y) มีการแจกแจงเป็นโค้งปกติในแต่ละค่าของตัวแปรอิสระ (x) ข้อสมมตินี้เป็นสิ่งจำเป็นสำหรับการทดสอบนัยสำคัญทางสถิติ F

2. คะแนนตัวแปรตาม (y) ที่ได้มีความแปรปรวน (variance) เท่ากันในทุก ๆ ค่าตัวแปรอิสระ (x) กล่าวคือ ค่า y ณ x ใดถือว่าเป็นตัวแทนสุ่มมาจากประชากรปกติโดยที่ทุก ๆ ประชากรมีการกระจายร่วมกันอยู่คือ $\sigma_{y,x}^2$

3. ค่าความผิดพลาดแต่ละค่า (error terms) ที่ได้จากการคำนวณในสมการถดถอยเป็นค่าที่ได้โดยการสุ่ม (random) และมีการแจกแจงเป็นโค้งปกติ ข้อสมมตินี้ใช้ประกอบในการคาดคะเนทางสถิติ (Statistical estimation)

3.3 คุณสมบัติของสมการถดถอยเส้นตรง

คุณสมบัติของสมการถดถอยแบบที่เป็นเส้นตรงมีดังต่อไปนี้

- 1) ค่าเฉลี่ยของแปรอิสระ (\bar{x}) และค่าเฉลี่ยของตัวแปรตาม (\bar{y}) จะอยู่บนเส้นถดถอย
- 2) ผลรวมของผลต่าง (deviation) ระหว่างจุดที่สังเกตได้กับค่าเฉลี่ยของประชากรเป็นศูนย์
- 3) ผลรวมของกำลังสองของผลต่างระหว่างจุดที่สังเกตได้กับค่าเฉลี่ยมีค่าต่ำสุด

3.4 ข้อดีของการวิเคราะห์ถดถอยพหุคูณ

- 1) วิธีการนี้ มีส่วนสัมพันธ์อย่างใกล้ชิดและสอดคล้องกับวัตถุประสงค์ของวิทยาศาสตร์ กล่าวคือ การอธิบายปรากฏการณ์หนึ่งอันใดที่เกิดขึ้น แม้ว่าปรากฏการณ์นั้นจะมีลักษณะปนกันหรือสับสนเพียงใดก็ตาม

2) สามารถที่จะใช้ให้กับตัวแปรอิสระจำนวนหรือประเภทใดก็ได้วิธีการวิเคราะห์ให้ผลในรูป ๆ ด้านตามที่การวิเคราะห์ความแปรปรวน (analysis of variance) จะทำได้และให้ผลมากกว่าช่วยทำให้คิดว่าตัวแปรอิสระจะมีลักษณะหรือมาตราการวัด (scale) ใดๆก็ตาม โดยเฉพาะอย่างยิ่งการวิเคราะห์ที่ทดลองทุกชุดเป็นวิธีที่พิเศษสำหรับการวิเคราะห์ข้อมูลที่มีผลจากการทดลอง (nonexperimental data)

3) การวิเคราะห์ที่ทดลองทุกชุด ใช้วิธีการทางสถิติหลายวิธีในการพิจารณาแปลความหมายข้อมูล ตัวอย่างเช่น วัดความสัมพันธ์ (correlation) ทั้งหมดที่เกิดขึ้นในข้อมูลระหว่างตัวแปรอิสระและระหว่างตัวแปรอิสระกับตัวแปรตาม ค่า R^2 ใช้ในการคาดคะเนอัตราส่วนของความแปรปรวนในตัวแปรตาม ซึ่งเป็นผลมาจากการรู้ความแปรปรวนของตัวแปรอิสระ นอกจากนั้นการทดสอบนัยสำคัญทางสถิติ F ช่วยให้เห็นความแตกต่างในค่าของ R^2 หรือลักษณะความมีนัยสำคัญของการเพิ่มความรู้ในการกระจายของตัวแปรตามอันเป็นผลมาจากการนำตัวอิสระบางตัวแปรเพิ่มในสมการถดถอย เป็นต้น

3.5 จุดอ่อนของการวิเคราะห์ที่ทดลองทุกชุด

ข้อบกพร่องของการวิเคราะห์ที่ทดลองทุกชุดอยู่ที่ความไม่เที่ยง (unreliability) ของค่าสัมประสิทธิ์ของการถดถอย (regression coefficient) เมื่อมีการนำตัวแปรอิสระเข้าในสมการถดถอยหรือออกจากสมการถดถอย ค่าสัมประสิทธิ์ถดถอยดังกล่าวจะไม่คงที่ และจะเปลี่ยนแปลงไปตามขนาดของกลุ่มตัวอย่างและจำนวนตัวแปรอิสระ ปัญหานี้อาจลดลงได้ หรือเกิดขึ้นน้อย สำหรับการใช้กลุ่มตัวอย่างขนาดใหญ่ และใช้ตัวแปรอิสระจำนวนน้อย ๆ ในกรณีที่ตัวแปรอิสระมีค่าสหสัมพันธ์ระหว่างกันและกันสูงนั้น ตัวแปรอิสระดังกล่าวจะวัดหรืออธิบายในสิ่งเดียวกัน ผลที่ตามมาคือค่าความผิดพลาดในการวัด (standard error of measurement) ของสัมประสิทธิ์การถดถอยจะเพิ่มขึ้นอย่างเห็นได้ชัด และจะไม่มี ความหมายใด ๆ ในการแปลผลการอธิบายความแปรปรวนของตัวแปรตาม ผลอีกประการหนึ่งของการที่ตัวแปรอิสระมีความสัมพันธ์ระหว่างกันสูง ทำให้เกิดปัญหา multicollinearity ซึ่งนอกจากจะก่อให้เกิดความลำบากในการอธิบายความแปรปรวนของตัวแปรตามแล้ว ยังทำให้ค่าสัมประสิทธิ์สหสัมพันธ์ส่วนย่อย (partial correlation coefficient) ที่ใช้สำหรับการคัดเลือกตัวแปรอิสระเพื่อเข้าสู่สมการถดถอยเกิดความคลาดเคลื่อนอีกด้วย

3.6 การวิเคราะห์ถดถอยพหุคูณแบบสเต็ปไวส์ (Stepwise Multiple Regression Analysis)

ในการเลือกสมการถดถอยเพื่อการพยากรณ์ที่ดีที่สุดนั้น มีเกณฑ์การเลือก 2 ประการ ที่เป็นข้อคิดสำหรับการพิจารณา (Draper N.R. and Smith, H. 1966: 163)

ประการแรกคือ การให้สมการที่สร้างมีประสิทธิภาพสูงและเป็นประโยชน์ในการพยากรณ์มากที่สุด ซึ่งในสมการนั้น ควรประกอบด้วยตัวแปรอิสระมากที่สุดเท่าที่จะทำได้ เพื่อค่าของตัวแปรตามที่คำนวณได้มีความเชื่อถือได้สูงสุด

ประการที่สอง เนื่องจากปัญหาของการวิจัยทั้งหลายมักจะเนื่องมาจากองค์ประกอบของเวลาและค่าใช้จ่ายการวิเคราะห์ข้อมูลโดยใช้ตัวแปรอิสระจำนวนมาก ๆ เป็นการสิ้นเปลืองและแทบเป็นไปไม่ได้ ดังนั้น จึงควรที่จะใช้ตัวแปรอิสระให้น้อยที่สุดเท่าที่จะทำได้ เนื่องจากข้อพิจารณาทั้งสองประการดังกล่าวมีส่วนขัดแย้งกันในตัวเอง วิธีการที่จะประนีประนอมสำหรับเป็นข้อสรุปในการพิจารณาเลือกสมการพยากรณ์นั้นก็คือ ต้องพยายามเลือกสมการถดถอยเพื่อพยากรณ์ที่ดีที่สุด เพื่อให้ตัวแปรอิสระที่เหมาะสมจำนวนน้อย แต่ให้ผลการพยากรณ์ที่เชื่อถือได้สูงเท่าที่จะเป็นไปได้ การเลือกตัวแปรในสมการถดถอยเพื่อการพยากรณ์ดังกล่าวสามารถทำได้หลายวิธี แต่ในแต่ละวิธีความเห็นหรือการตัดสินใจของบุคคลที่เกี่ยวข้องมีส่วนสำคัญไม่น้อย

ในบรรดาวิธีการเลือกสมการถดถอยเพื่อการพยากรณ์ต่าง ๆ นั้น วิธีการถดถอยพหุคูณแบบสเต็ปไวส์ (Stepwise Multiple Regression procedure) เป็นวิธีการที่ให้ความถูกต้องในการพยากรณ์สูงสุด และเป็นวิธีที่คัดเลือกกลุ่มของตัวแปรอิสระที่เหมาะสมจำนวนพอดี (optimum set of independent variables) ในการพยากรณ์ตัวแปรตามหรือตัวแปรที่เป็นผลของการศึกษา วิธีการนี้ได้รับความนิยมและนำมาใช้อย่างแพร่หลาย (Halinski, R.S. and Feldt, L.S. 1970)

เพื่อที่จะแสดงการใช้การถดถอยพหุคูณแบบ Stepwise ในการพิจารณาเลือกตัวแปร สมมติว่ามีตัวแปรอิสระทั้งหมด 20 ตัวแปร และต้องการสร้างสมการพยากรณ์สมการหนึ่งจากตัวแปรอิสระจำนวน 5 ตัวแปร ที่สามารถให้การพยากรณ์ที่ดี วิธีการเลือกสมการที่ประกอบด้วยตัวแปร 5 เหอม ในสมการนี้วิธีหนึ่งคือการสร้างสมการพยากรณ์ที่มีตัวแปร ทุกตัวแปรอยู่ใน

สมการ โดยการจับตัวแปรนั้นๆ ให้มีความประสานสัมพันธ์ (Combination) กัน ในลักษณะต่าง ๆ ซึ่งถ้าพิจารณาการจับในลักษณะดังกล่าวแล้วอาจทำได้ถึง 15,000 วิธี และเป็นการแน่นอนว่า วิธีการเลือกตัวแปรจากลักษณะการจับทั้งหมดโดยวิธีดังกล่าวย่อมทำได้ยากและสิ้นเปลือง วิธีการที่ง่ายและสะดวกกว่าจึงมีความจำเป็นอย่างยิ่ง การวิเคราะห์ถดถอยพหุคูณแบบ Stepwise เป็นวิธีการที่มีประสิทธิภาพและทำได้รวดเร็ว วิธีนี้จะเลือกตัวแปรอิสระและจับตัวแปรดังกล่าวในสมการพยากรณ์ที่นับว่าเป็นค่าตอบที่ค่อนข้างเหมาะสม วิธีการเริ่มทวนในคอนแรก สมการพยากรณ์หรือสมการถดถอย (regression equation) จะประกอบด้วยตัวแปรอิสระเพียงตัวเดียวที่มีค่าความสัมพันธ์สูงสุดกับตัวแปรตาม ถ้าเราให้ X_1 เป็นตัวแปรดังกล่าวที่มีค่าสัมพันธ์สหสัมพันธ์สูงสุดกับ Y สมการแรกในขบวนการวิเคราะห์ถดถอยพหุคูณแบบ Stepwise จะเป็น

$$Y_i = b_{01} + b_{11}X_{1i} + e_{1i} \quad (i = 1, 2, \dots, N)$$

$$\text{หรือ } \hat{Y}_i = b_{01} + b_{11}X_{1i}$$

เมื่อ Y_i เป็นข้อมูลที่ i ของตัวแปรตาม Y

\hat{Y}_i เป็นค่าพยากรณ์ที่คำนวณได้จากค่าตัวแปร X

N เป็นจำนวนข้อมูลหรือจำนวนกลุ่มตัวอย่างทั้งหมด

X_1 เป็นตัวแปรอิสระใช้ในการพยากรณ์ตัวแปรแรกซึ่งมาจากกลุ่มของตัวแปรอิสระทั้งหมด (potential predictors) จำนวน p ตัวแปร (X_1, X_2, \dots, X_p)

e_{1i} เป็นข้อมูลที่ i ของตัวแปรอิสระตัวแปรที่ 1

b_{01} และ b_{11} เป็นค่าคำนวณกำลังสองน้อยที่สุดของจุดตัดแกน y (intercept) และค่าของความลาด (slope) สำหรับสมการถดถอยตามลำดับ

e_{1i} เป็นส่วนของความแปรปรวนใน Y ที่ไม่สามารถรู้ได้จากความสัมพันธ์ของ Y และ X_1

หลักกำลังสองน้อยที่สุดในการวิเคราะห์ถดถอยกำหนดว่าค่าคำนวณของความลาด และจุดตัดแกน y ที่ได้เป็นค่าที่ให้ผลรวมของส่วนที่เหลือกำลังสอง $(\sum_{i=1}^N e_{1i}^2)$ มีค่าน้อยกว่าค่าที่ได้จากการใช้หลักการหรือวิธีการอื่น

ความสัมพันธ์ของความแปรปรวนในตัวแปรตาม ความแปรปรวนในส่วนที่รู้และ ความแปรปรวนในส่วนที่เหลือ หรือส่วนที่ไม่รู้อาจเขียนได้ในรูปของ

$$\sum_{i=1}^N (y_i - \bar{y})^2 = \sum_{i=1}^N (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^N e_{1i}^2$$

หรือ

$$\left[\begin{array}{c} \text{ความแปรปรวน} \\ \text{ในตัวแปรตาม} \end{array} \right] = \left[\begin{array}{c} \text{ความแปรปรวนในตัวแปรตาม} \\ \text{ที่วัดได้จากผลการของสมการ} \\ \text{ถดถอย} \end{array} \right] + \left[\begin{array}{c} \text{ความแปรปรวน} \\ \text{ที่ไม่สามารถรู้ได้} \\ \text{จากสมการถดถอย} \end{array} \right]$$

หรือเขียนแทนว่า

$$Q_y = Q_f + Q_r$$

เมื่อพิจารณาคุณสมบัติของความสัมพันธ์ ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่าง y และ \hat{y} อาจเรียกว่า สัมประสิทธิ์สหสัมพันธ์พหุคูณ (coefficient of multiple correlation) ซึ่งเป็นค่าที่แสดงขนาดของความสัมพันธ์ระหว่างตัวแปรตาม และค่าพยากรณ์ที่ได้จากความสัมพันธ์เชิงเส้นตรงของตัวแปรอิสระในการวิเคราะห์ถดถอย ค่าสัมประสิทธิ์สหสัมพันธ์ไม่ใช่สิ่งที่ควรแก่การสนใจมากนักแต่เราสนใจว่าตรรกะที่แสดงความสัมพันธ์ระหว่าง y และ \hat{y} ซึ่งเรียกว่าสัมประสิทธิ์การทำนาย (coefficient of determination, R^2) การคำนวณตรรกะนี้ดังกล่าว อาจทำได้โดยการยกกำลังสองค่าสัมประสิทธิ์สหสัมพันธ์พหุคูณ อย่างไรก็ตาม Fisher ระบุลักษณะการกระจายของ R^2 แตกต่างไปจากการกระจายของสัมประสิทธิ์สหสัมพันธ์พหุคูณกำลังสอง เว้นเสีย

แต่ว่าจะไม่มีความสัมพันธ์เชิงเส้นตรงระหว่างตัวแปรในสมการ (Fisher, R.A. 1928: 654-673)

อนึ่ง สัมประสิทธิ์การทำนาย R^2 สามารถพิจารณาได้ในรูปของอัตราส่วนของความแปรปรวนในตัวแปรตามที่สามารถวัดหรือรู้ได้จากความแปรปรวนของตัวแปรอิสระ และ $1 - R^2$ เป็นค่าของอัตราส่วนในความแปรปรวนของตัวแปรตามที่ไม่สามารถวัดได้ อนึ่ง Kendall และ Stuart (Kendall, M.G. and Stuart, A 1961) ให้ความหมายสัมประสิทธิ์การทำนายในรูปของอัตราส่วนของความแปรปรวนที่เราสามารถวัดได้ ดังนี้

$$R^2 = \frac{Q_r}{Q_y} = [1 - \frac{Q_r}{Q_y}]$$

จากสมการแสดงความสัมพันธ์ดังกล่าว จะเห็นว่า Q_r และ R^2 จะมีค่าสูงสุดเมื่อ Q_r มีค่าน้อยที่สุด และ Q_y เป็นค่าคงที่เฉพาะสำหรับกลุ่มตัวอย่างแต่ละกลุ่ม ขั้นตอนที่สองของขบวนการเลือกตัวแปรโดยวิธีการถดถอยแบบ Stepwise เกี่ยวข้องกับการเลือกตัวแปรที่สองเข้าในสมการถดถอย โดยที่ตัวแปรดังกล่าวจะเลือกจากตัวแปรอิสระที่เหลือจำนวน $p-1$ ตัวแปร ซึ่งยังไม่ได้รับการพิจารณาคัดเลือกเข้าสมการถดถอยในขั้นตอนแรก เกณฑ์การเลือกจะพิจารณาตัวแปรอิสระที่เพิ่มความเข้าใจในการพยากรณ์ตัวแปรตามมากที่สุด อีกนัยหนึ่งอาจกล่าวได้ว่า ถ้าการเพิ่มของ x_2 ยังผลให้สมการถดถอย ซึ่งประกอบด้วยตัวแปรสองตัวแปรให้ค่าการเพิ่มของ R^2 สูงกว่าการเพิ่มตัวแปรอิสระอื่นในสมการแล้ว x_2 จะได้รับการคัดเลือกเป็นตัวแปรที่จะเข้าร่วมการพยากรณ์ในสมการ

$$y_1 = b_{02} + b_{12}x_{11} + b_{22}x_{21} + c_{21}$$

ถ้าค่า R^2 จากขั้นตอนแรกและขั้นตอนที่สองเขียนแทนด้วย R_1^2 และ R_2^2 ตามลำดับ แล้ว $R_2^2 - R_1^2$ จะมีค่ามากกว่าค่าอื่น ๆ ที่ได้จากการเลือกตัวแปรอิสระอื่นในขั้นการเลือกดังกล่าว อีกวิธีหนึ่งในการหาว่าตัวแปรอิสระใดที่ให้การเพิ่มค่า R^2 สูงกว่าในขั้นตอนนี้อาจ

ทำโดยการหาค่ากำลังสองของสหสัมพันธ์ส่วนย่อย (squared partial correlation coefficient) ระหว่างค่า y และค่าแต่ละค่าของตัวแปรอิสระที่เหลือจำนวน $p-1$

ตัวแปรโดยไม่พิจารณา x_1 (partialed out x_1) ถ้า x_2 ให้ค่าการเพิ่ม R^2 สูงสุด ก็จะได้ค่ากำลังสองของสหสัมพันธ์ส่วนย่อยสูงสุด

หลังจากที่มีการคัดเลือกตัวแปรอิสระเข้าในสมการพยากรณ์แต่ละครั้งจะมีการทดสอบนัยสำคัญของตัวแปรอิสระที่ร่วมกันอยู่ในสมการถดถอย โดยการทดสอบสมมติฐานว่า ความแปรปรวนใน y จะสามารถรู้หรืออธิบายได้จากความแปรปรวนที่ได้จากการทำนาย (R^2) มีค่าแตกต่างจากศูนย์หรือไม่ การทดสอบสมมติฐานนี้ บางครั้งเรียกว่า การทดสอบสถิติส่วนรวม F (overall-F test) โดยสูตร

$$F_{d_f, d_r} = \frac{R^2 Q_y / d_f}{(1-R^2) Q_y / d_r} = \frac{d_r R^2}{(1-R^2) d_f}$$

เมื่อ d_f เป็นจำนวนตัวแปรอิสระในสมการถดถอย (p)

d_r มีค่าเป็น $N-p-1$

ขั้นตอนอีกประการหนึ่งเพื่อการศึกษาว่า ตัวแปรใดควรพิจารณาเข้าหรือควร จะตัดออกจากสมการถดถอย อาจกระทำโดยการทดสอบสถิติส่วนย่อย F (partial-F test) เนื่องจากอิทธิพลของตัวแปรอิสระหนึ่ง ๆ ที่มีต่อตัวแปรตาม อาจจะมีมากเมื่อตัวแปรนั้น เป็นตัวแปรพยากรณ์ (predictor) เพียงตัวแปรเดียวในสมการถดถอย แต่ตัวแปรที่ได้รับการพิจารณาแล้วว่า เป็นตัวแปรเดียวที่ดีที่สุด ได้รับการคัดเลือกเข้าสู่สมการตั้งแต่ตอนต้น อาจจะไม่ มีประโยชน์ในการที่จะให้ข้อมูลเพิ่มเติมเมื่อพิจารณาประกอบในหาค่าความสัมพันธ์กับตัวแปรอิสระที่ได้รับเลือกเข้าไปในสมการถดถอยใหม่ กล่าวคือ ตัวแปรอิสระเดียวกันนั้น อาจจะมีอิทธิพลต่อตัวแปรตามน้อย เมื่อตัวแปรอิสระนั้นได้รับการพิจารณาเข้าสู่สมการตามหลังตัวแปรอิสระอื่น ทั้งนี้เนื่อง จากความจริงเกี่ยวกับความสัมพันธ์ของตัวแปรอิสระที่ศึกษากับตัวแปรอิสระที่อยู่ในสมการพยากรณ์

การทดสอบสถิติส่วนย่อย F สามารถกระทำกับสัมประสิทธิ์การถดถอยทุกค่า ไม่ว่าตัวแปรอิสระจะได้รับเลือกในขั้นตอนใดของขบวนการเลือกตัวแปรสู่สมการถดถอย ทั้งนี้เพื่อจะศึกษาผลเกี่ยวข้องกับอิทธิพลร่วมของตัวแปรอิสระในการพยากรณ์เกี่ยวเนื่องกับการศึกษาสถิติส่วนย่อย F การทดสอบนัยสำคัญของการเพิ่มความแปรปรวนอันเป็นผลของการเพิ่มตัวแปรอิสระตัวที่ m ในสมการพยากรณ์ที่มีตัวแปรอิสระอยู่แล้ว m ตัวแปรอาจทำได้โดยการคำนวณอัตราส่วน

$$F_{1, N-n-1} = \frac{(R_n^2 - R_m^2) (N-n-1)}{(1-R_n^2)} = \frac{(R_n^2 - R_m^2) / 1}{(1-R_n^2) (N-n-1)}$$

เมื่อ R_n^2 คือสัมประสิทธิ์การพยากรณ์สำหรับสมการที่มีตัวแปรอิสระ n ตัวแปร และ

R_m^2 คือ สัมประสิทธิ์การพยากรณ์สำหรับสมการที่มีตัวแปรอิสระ m ตัวแปร

ผลต่างของ R_n^2 และ R_m^2 คือ อัตราส่วนของความแปรปรวนที่เพิ่มขึ้นโดยการนำตัวแปรที่ n เข้าสู่สมการพยากรณ์ อนึ่ง การตรวจสอบโดยใช้สถิติ F ดังกล่าวจะเท่ากับตัวแปรทุกตัวแปรที่พิจารณาเลือกเข้าสู่สมการพยากรณ์ในทุก ๆ ขั้นตอน ตัวแปรใดที่ให้ค่าการทดสอบต่ำกว่าเกณฑ์ที่ตั้ง จะถูกกำจัดออกจากสมการพยากรณ์

ขบวนการคัดเลือกตัวแปรเข้าในสมการพยากรณ์และการทดสอบตัวแปรนั้น จะสิ้นสุดลงเมื่อ ไม่มีตัวแปรใดที่ได้รับการพิจารณาเข้าสู่สมการพยากรณ์ และไม่มีตัวแปรใดที่ถูกกำจัดออกจากสมการพยากรณ์

วิธีการคำนวณสำหรับการถดถอยพหุคูณแบบ Stepwise ส่วนใหญ่ใช้หลักการแก้ระบบสมการเส้นตรง (System of Linear Equations) นั่นคือ ใช้วิธีเปลี่ยนแถว (row) และสัณฐาน (Column) ของข้อมูลตัวแปรต่าง ๆ ตามวิธีของ กอส์ (Gauss Elimination) ด้วยวิธีการดังกล่าว จะให้ข้อมูลที่จำเป็นในการเลือกตัวแปรอิสระตัวต่อไปเพื่อที่จะนำเข้าสู่สมการพยากรณ์

4. มโนทัศน์ว่าด้วยทฤษฎีการตอบสนองข้อสอบ

4.1 ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory; IRT)

IRT เป็นทฤษฎีหนึ่งในทฤษฎีคุณลักษณะแฝง (Latent trait theory) บางคนจึงเรียก IRT ว่าเป็นทฤษฎีคุณลักษณะภายใน Ferguson (1942) และ Lawley (1943) ได้รับการยกย่องว่าเป็นผู้ริเริ่ม IRT (Warm 1979: 19; Lord and Novick 1968: 369) โดยมีหลักการว่า ผลการตอบของผู้สอบจากแบบสอบใด ๆ นั้นขึ้นอยู่กับความสามารถ (Latent trait or ability or skill) ของผู้สอบ ซึ่งสามารถเขียนอยู่ในรูปฟังก์ชันทางคณิตศาสตร์ ซึ่งฟังก์ชันเหล่านี้ประกอบด้วยค่าพารามิเตอร์ต่าง ๆ ได้แก่ ค่าความยาก อ่านาจจำแนกและการเดา ซึ่งแล้วแต่ว่าฟังก์ชันจะมีพารามิเตอร์กี่ตัว

ต่อมาในปี 1952 ลอร์ดได้เสนอ IRT ขึ้นใหม่ให้ชื่อว่าทฤษฎีโค้งคุณลักษณะข้อกระหง (Item Characteristic Curve Theory) โดยเสนอว่าลักษณะของโค้งผลการตอบ (Item characteristic curve) ของข้อสอบแต่ละข้อมีลักษณะเป็น normal ogive ต่อมาจึงเรียกว่า Normal Ogive Model ซึ่งโมเดลนี้มีการคำนวณยุ่งยากมาก ดังนั้นจึงทำให้ ลอร์ดซึ่งให้ความสนใจใน IRT ไประยะหนึ่ง ในปี 1960 George Rasch ได้เสนอ IRT ในรูปพารามิเตอร์เดียว คือ ความยาก และในปี 1965 ลอร์ดก็หันกลับมาสนใจพัฒนา IRT ใหม่ (Warm, 1979: 19)

ปี 1960 เบิร์นบอม (Birnbaum) ได้เสนอ Logistic Model ที่ใช้ 2 พารามิเตอร์ขึ้น ซึ่งเป็นโมเดลที่ง่ายกว่าของลอร์ด จึงทำให้ IRT เป็นที่นิยมใช้กันแพร่หลาย และมีการพัฒนาขึ้นเรื่อย ๆ จนกระทั่งใช้ได้กับพารามิเตอร์ 1 ตัว 2 ตัว และ 3 ตัว (Warm 1979: 19-21)

4.2 ข้อตกลงเบื้องต้นของ IRT (Lord and Novick, 1968: 359)

4.2.1 แบบสอบที่มีมิติเดียว (Unidimensional tests) หมายความว่า ข้อสอบแต่ละข้อในแบบสอบจะวัดความสามารถหรือคุณลักษณะเดียวกัน (a Single ability or latent trait) หรือมีความเป็นเอกพันธ์กัน (homogeneous)

4.2.2 ข้อสอบแต่ละข้อจะต้องเป็นอิสระจากกัน (Local independence) หมายความว่า การตอบข้อสอบข้อหนึ่งถูก จะไม่มีผลต่อการตอบข้อสอบอื่น ๆ

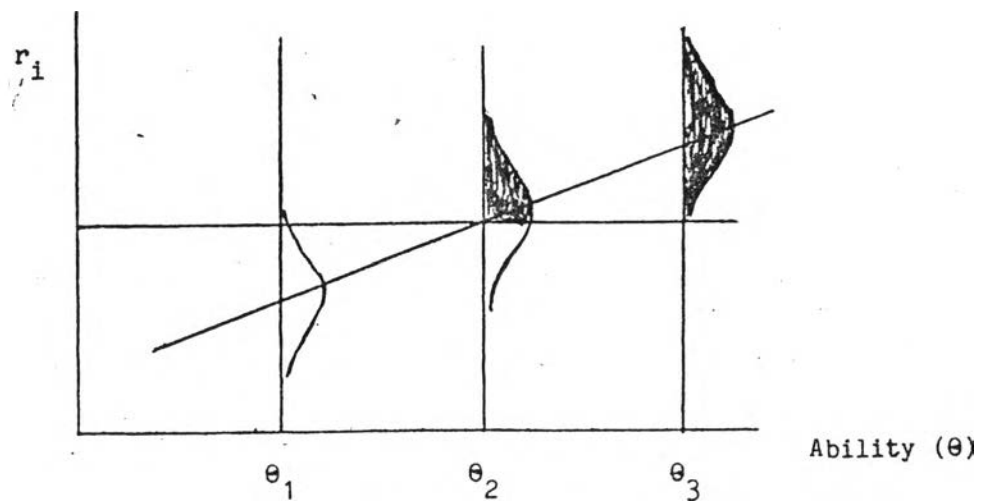
4.2.3 โอกาสที่ผู้สอบจะตอบข้อสอบถูก ขึ้นอยู่กับลักษณะของโค้งผลการตอบ (Item characteristic curve: ICC) ของแต่ละโมเดลที่ใช้ ไม่ขึ้นกับการแจกแจงความสามารถของกลุ่มตัวอย่าง

4.3 โมเดลต่าง ๆ ใน IRT

โมเดลต่าง ๆ ใน IRT ได้มาจากมโนทัศน์ในเรื่องความสัมพันธ์ระหว่างความสามารถ (ability) กับการกระทำ (Performance) ถ้ามีความสามารถ (θ) สูง ก็น่าจะมี performance (y) สูงขึ้นด้วย ถ้ามีเกณฑ์ (r_1) ใดหนึ่ง เป็นตัวบอกว่า y แต่ไหนจึงจะต้องทำข้อสอบ 1 ถูก ดังนั้น $y > r_1$ แสดงว่าข้อสอบ 1 ถูก ถ้า $y < r_1$ แสดงว่าทำข้อสอบ 1 ผิด

แสดงความสัมพันธ์ระหว่างความสามารถ (θ) กับ performance (y)

Performance (y)



ดังนั้น ถ้านำโอกาสที่จะตอบถูก (พ.ท. ส่วนที่แรงๆ) ในระดับความสามารถต่าง ๆ มาเขียนกราฟใหม่ จะได้เป็นรูป ICC ต่าง ๆ ซึ่ง ICC เหล่านี้จะแตกต่างกันไปตามฟังก์ชันทางคณิตศาสตร์ และจำนวนพารามิเตอร์ที่ใช้อธิบาย ICC ของแต่ละโมเดล ซึ่งเขียนเป็นรูปฟังก์ชันทั่วไปคือ

$$P(\theta) = f(\theta)$$

$P(\theta)$ แทนความน่าจะเป็นในการตอบถูก

θ แทนความสามารถที่แท้จริง

r แทนความสัมพันธ์
 ซึ่ง $r(\theta)$ จะเป็นไปตามโมเดลต่าง ๆ ดังนี้

4.3.1 The Latent linear model (Hambleton and Cook 1977: 78)

$$P_g(\theta) = b_g + a_g \theta$$

$P_g(\theta)$ คือ ความน่าจะเป็นที่คนที่มีความสามารถ θ จะตอบข้อสองที่ g ถูก
 a_g คือ ค่าอำนาจจำแนก
 b_g คือ ค่าความยาก

4.3.2 The latent distance model (Torgerson 1958: 374)

$$P_g(\theta) = \begin{cases} b_g - a_g & ; \theta \leq \theta_g \\ b_g + a_g & ; \theta > \theta_g \end{cases}$$

โมเดลนี้ใช้กับ Guttman scale และ scale อื่น ๆ

4.3.3 Normal ogive model: (Lord and Novick 1968: 366)

$$P_g(\theta) = \int_{-\infty}^{a_g(\theta - b_g)} \phi(t) dt \quad ; g = 1, 2, 3, \dots, n$$

$P_g(\theta), a_g, b_g$ มีความหมายเช่นเดียวกับในข้อ 1
 θ คือ ระดับความสามารถของผู้สอบ
 $\phi(t)$ คือ normal density function

4.3.4 Logistic Model จะมี ICC เหมือนกับของ Normal Ogive model เมื่อเพิ่มค่าคงที่ 1.7 เข้าไปใน model ซึ่ง Logistic Model คำนวณง่ายกว่าของ Normal Ogive มาก และยังสามารถแบ่ง model ย่อยออกเป็น 3 โมเดล ดังนี้ (Hambleton and Cook 1977: 8-82)

4.3.4.1 โมเดลที่ใช้พารามิเตอร์ 3 ตัว (Three-Parameter Logistic Model) มีสูตรการคำนวณดังนี้

$$P_g(\theta) = C_g + (1 - C_g) \frac{e^{D a_g (\theta - b_g)}}{1 + e^{D a_g (\theta - b_g)}} ; g = 1, 2, 3, \dots, n$$

C_g คือ ค่าการเดา

D คือ scaling factor มีค่า 1.7

e คือ ค่าคงที่มีค่า 2.7182818

4.3.4.2 โมเดลที่ใช้พารามิเตอร์ 2 ตัว (Two-Parameter Logistic Model) ถือว่าแบบสอบไม่มีการเดา มีสูตรการคำนวณดังนี้

$$P_g(\theta) = \frac{e^{D a_g (\theta - b_g)}}{1 + e^{D a_g (\theta - b_g)}} ; g = 1, 2, 3, \dots, n$$

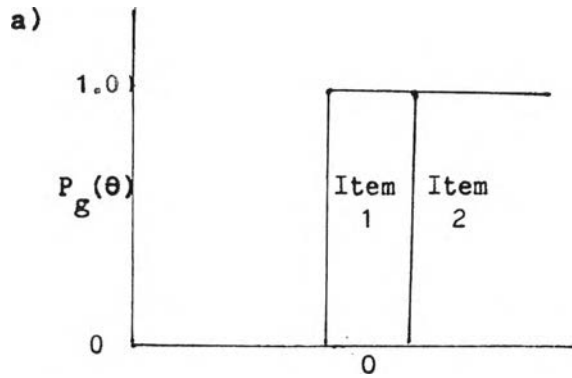
4.3.4.3 โมเดลที่ใช้พารามิเตอร์ 1 ตัว (One-Parameter logistic Model) เป็นโมเดลที่ Georg Rasch เสนอในปี 1960 ซึ่งตรงกับ Logistic Model ของเบิร์นบอม ที่คิดค่าความยากของแบบสอบเพียงอย่างเดียว จึงนิยมเรียกโมเดลนี้ว่า ราสชโมเดล (Rasch Model)

$$P_g(\theta) = \frac{e^{\theta - b_g}}{1 + e^{\theta - b_g}}$$

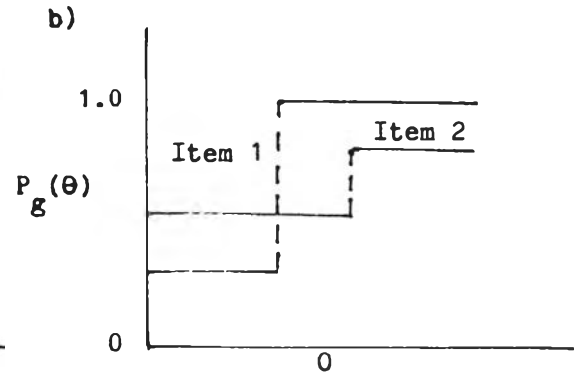
4.4 พิสัยของค่าพารามิเตอร์

θ , a , b มีค่าระหว่าง $-\infty$ ถึง ∞ แต่ในทางปฏิบัติ θ มีค่าระหว่าง -3 ถึง 3 a มีค่าอยู่ระหว่าง 0.5 ถึง 2.5 b มีค่าอยู่ระหว่าง -2 ถึง 2 และ c มีค่าอยู่ระหว่าง 0 ถึง 0.3 (Ree 1979: 372)

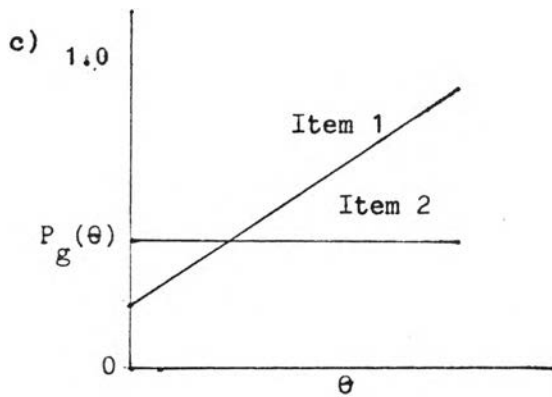
แสดง ICC ของโมเดลต่าง ๆ (Hambleton and Cook 1979: 79)



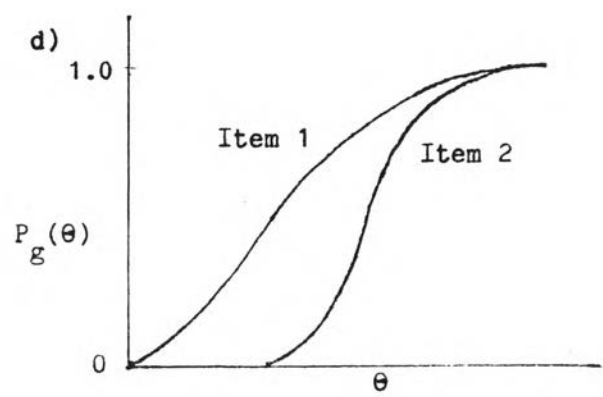
Perfect Scale curves
(Guttman's Perfect Scales)



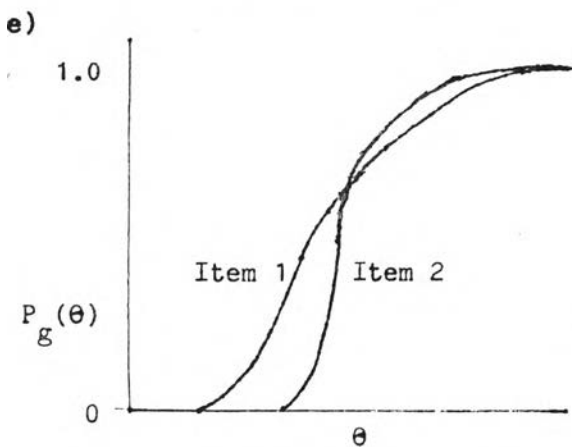
Latent distance curves



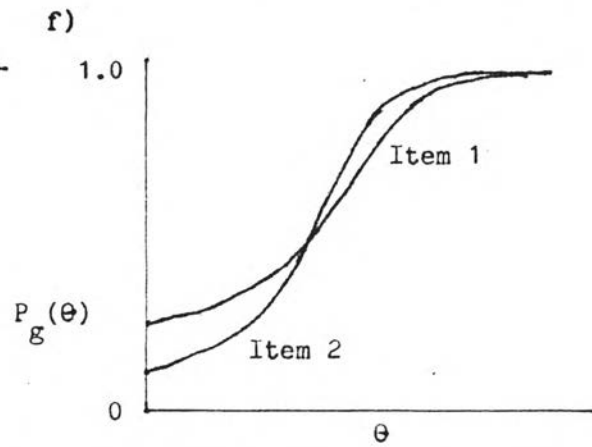
Latent Linear curves



One-parameter logistic curves



Two-parameter logistic curves



Three-parameter logistic curves

รูป a แสดงถึงโอกาสที่จะทำให้ข้อสอบแต่ละข้อได้ถูกเป็น 0 หรือ 1 เป็น Step function ใช้กับ Guttman's scale จากรูปแสดงว่าข้อสอบที่ 2 ยากกว่าข้อ 1

รูป b เป็น step function เหมือน a แต่เป็นรูปทั่วไปที่มีโอกาสที่จะตอบถูก ไม่เป็น 0 และ 1 จากรูปแสดงว่า ข้อกระทงที่ 2 ยากกว่า ข้อ 1

รูป c เป็นรูปที่แสดงความสัมพันธ์ของโอกาสที่จะถูกกับระดับความสามารถเป็นเส้นตรง ซึ่งข้อสอบที่ 1 จำแนกได้ดีกว่าข้อสอบที่ 2

รูป d แสดงถึงโอกาสที่จะทำให้ข้อสอบได้ถูกขึ้นอยู่กับความยากของข้อสอบเพียงอย่างเดียว ทุกข้อสอบมีค่าอำนาจจำแนกเท่ากันหมดคือ เป็น 1 ในรูปแสดงถึงข้อสอบที่ 2 ยากกว่าข้อที่ 1

รูป e แสดงถึงโอกาสที่จะทำให้ข้อสอบได้ถูกขึ้นอยู่กับค่าอำนาจจำแนกและค่าความยาก ในรูปข้อสอบที่ 2 จำแนกคนได้ดีกว่าข้อสอบที่ 1

รูป f แสดงถึงโอกาสที่จะทำให้ข้อสอบได้ถูก ขึ้นอยู่กับค่าอำนาจจำแนก ค่าความยากและการเดา ในรูปข้อสอบที่ 2 มีการเดาน้อยกว่าข้อสอบที่ 1 และมีค่าอำนาจจำแนกสูงกว่าข้อสอบที่ 1

4.5 การประมาณค่าพารามิเตอร์ในลอจิสติกแบบ 3 พารามิเตอร์ (Three-parameter Logistic Model)

การประมาณค่าอำนาจจำแนก (a) ความยาก (b) และค่าการเดา (c) ของข้อสอบ และค่าความสามารถแท้ (θ) ของผู้ตอบข้อสอบ ค่าเนินการด้วยวิธีแมกซิมัมไลคลิฮูด (Maximum Likelihood Estimates) กล่าวคือต้องหาค่าอนุพันธ์ของค่าฟังก์ชันไลคลิฮูด ซึ่งเขียนอยู่ในรูปของลอการิทึม (Logarithm of the Likelihood function) ตัวอย่างเช่น จะหาความสามารถแท้ (θ) ของผู้สอบ N คน ก็จะมีอนุพันธ์เทียบกับความสามารถแท้ (θ) N สมการ ซึ่งเมื่อหาค่าอนุพันธ์เทียบกับค่าความสามารถแท้ (θ) แล้ว ได้ดังนี้ (Frederic M. Lord, 1980 : 61)

$$\frac{dP_1}{d\theta} = \frac{Da_1(1 - c_1)}{e^{DL_1} + 2 + e^{-DL_1}}$$

$$\text{เมื่อ } D = 1.7 \text{ และ } L_1 = a_1(\theta - b_1)$$

หลังจากการหาค่าอนุพันธ์แล้วทำการหาค่าราคสมการโดยการกำหนดให้ อนุพันธ์มีค่าเท่ากับศูนย์และทำนองเดียวกันในการหาค่า a , b , และ c ของข้อสอบแต่ละข้อก็จะ หาค่าอนุพันธ์เทียบกับ a , b , และ c ตามลำดับ และกำหนดให้มีค่าเท่ากับศูนย์เช่นเดียวกัน หลังจากนั้นจะต้องใช้กระบวนการทดลองกำหนดตัวเลขให้แก่ตัวแปรในสมการแบบซ้ำ ๆ

(Iterative Procedure) จนกระทั่งได้ค่าพารามิเตอร์คงที่ (numerical Values Converge)

ในกรณีประมาณค่าพารามิเตอร์ดังกล่าวสามารถใช้คอมพิวเตอร์ประมวลผล ซึ่งมีโปรแกรมสำเร็จรูป ชื่อ LOGIST 5 (M.S. Wingersky, M.A. Barton and F.M. Lord, 1982)

5. งานวิจัยที่เกี่ยวข้อง

5.1 งานวิจัยเกี่ยวกับการให้คะแนนข้อสอบ

ในการวัดผลนั้น วิธีที่นิยมกันทั่วไปในการให้คะแนน ได้แก่ การให้คะแนนข้อที่ถูกได้ 1 คะแนน และข้อที่ผิดได้ 0 คะแนน แต่เพื่อต้องการให้ค่าความตรงและค่าความเที่ยงสูงขึ้น จึงมีผู้พยายามคิดค้นหาวิธีจะให้คะแนนอีกหลายวิธี และได้มีงานวิจัยอีกเป็นจำนวนมากที่แสดงว่า ผลของความพยายามนี้ทั้งประสบผลสำเร็จและล้มเหลว ซึ่งจะให้กล่าวต่อไป

5.1.1 การให้คะแนนตัวเลือกแตกต่างกัน

(1) วิธีอาศัยความเห็นของผู้เชี่ยวชาญ

จากการศึกษาของผู้วิจัยพบว่า การให้คะแนนตัวเลือกของข้อสอบโดยผู้เชี่ยวชาญมี 2 วิธี คือ 1) อาศัยความเห็นของผู้เชี่ยวชาญว่าตัวเลือกแต่ละตัวมีความถูกต้องมากน้อยเพียงใด และ 2) ใช้ค่าเฉลี่ยของลำดับที่ของความถูกต้องของตัวเลือกแต่ละตัวที่ตัดสินโดยผู้เชี่ยวชาญ จากการศึกษาของ Nedelsky ในปี 1954 (อ้างจาก Wang, and Stanley, 1970: 671) ที่ใช้นักศึกษาที่สอบผ่านการสอบรวมยอด 306 คน เป็นกลุ่มตัวอย่าง โดยกำหนดค่าน้ำหนักคะแนนโดยวิธีที่ 1 ปรากฏว่าแบบสอบแบบเลือกตอบมีค่าความเที่ยงสูงกว่าค่าความเที่ยงที่คำนวณจากคะแนนวิธีประเมินนิยม และในปี 1968 Jacobs และ Vanderventer (อ้างจาก Wang and Stanley, 1970, :693) ให้ใช้วิธีการให้คะแนนวิธีเดียวกัน ปรากฏว่า การให้คะแนนดังกล่าวทำให้แบบสอบแบบเลือกตอบมีค่าความเที่ยงจากการ

สอบซ้ำ ค่าความตรงเชิงพยากรณ์และค่าความตรงร่วมสมัย (Concurrent Validity) มีค่าสูงขึ้นกว่าค่าความเที่ยงและความตรงชนิดเดียวกันที่คำนวณจากคะแนนที่ให้โดยวิธีประเพณีนิยม

ส่วนการให้คะแนนวิธีที่ 2 นั้น Hambleton, Roberts และ Traub ได้ทำการศึกษาในปี 1970 (อ้างจาก Wang and Stanley, 1970: 693) ปรากฏว่าการให้คะแนนวิธีดังกล่าว ทำให้แบบสอบเลือกตอบมีค่าความเที่ยงและค่าความตรงเชิงพยากรณ์สูงขึ้นกว่าค่าความเที่ยงและค่าความตรงที่คำนวณจากคะแนนที่ให้โดยวิธีประเพณีนิยมแต่ไม่มีนัยสำคัญทางสถิติ

(2) วิธีใช้ค่าความสัมพันธ์ระหว่างตัวเลือกกับสิ่งที่มีงทดสอบ

การให้คะแนนโดยอาศัยแนวคิดเกี่ยวกับความสัมพันธ์ระหว่างตัวเลือกกับสิ่งที่มีงทดสอบ เท่าที่พบมี 3 วิธีคือ 1) ใช้ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวเลือกกับคะแนนรวม 2) ใช้ค่ามาตรวัดที่คำนวณจากวิธี Guttman's Scale Analysis และ 3) ใช้ค่ามาตรวัดที่คำนวณจากวิธีการเปรียบเทียบรายคู่ของ Thurstone (Thurstone's Paired Comparison) จากการศึกษาวรรณคดีที่เกี่ยวข้องพบว่า Davis และ Fifer ได้ใช้ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวเลือกแต่ละตัวกับคะแนนรวม เป็นคะแนนแบบสอบคู่ขนานวิชาคณิตศาสตร์ ซึ่งยาว 45 ข้อ ในปี 1959 (อ้างจาก Wang and Stanley, 1970: 692) และปรากฏว่า การให้คะแนนวิธีดังกล่าวทำให้แบบสอบมีค่าความเที่ยงสูงมากกว่าค่าความเที่ยงที่คำนวณจากคะแนนที่ให้โดยวิธีประเพณีนิยมอย่างมีนัยสำคัญคือ เพิ่มขึ้นจาก 0.68 เป็น 0.76 ซึ่งถ้าจะทำให้แบบสอบนั้นมีค่าความเที่ยงเท่ากับค่าที่เพิ่มขึ้น เมื่อให้คะแนนวิธีประเพณีนิยมจะต้องเพิ่มข้อสอบอีก 45-67 ข้อ

ส่วนการให้คะแนนโดยใช้มาตรวัดจากวิธี Guttman's Scale Analysis นั้น Hendrickson ได้ทำการศึกษาเกี่ยวกับกลุ่มตัวอย่างที่เป็นเพศชาย 5,000 คน และเพศหญิง 5,000 คน ที่สอบ Scholastic Aptitude Test ในปี 1970 และ Raffeld ได้ทำการศึกษาเกี่ยวกับกลุ่มตัวอย่างที่เป็นนักเรียน Grade 9 จำนวน 1,354 คน ที่สอบวิชาการอ่านเข้าใจความ กิจที่ ไวยากรณ์ การสะกดคำ และคณิตศาสตร์ (อ้างจาก Wang and Stanley, 1970: 693; Raffeld, 1975: 285) จากการศึกษา Hendrickson พบ

ว่า ความเที่ยงของแบบสอบสูงขึ้นกว่าค่าความเที่ยงที่คำนวณจากคะแนนที่ให้โดยวิธีประเพณีนิยมอย่างมีนัยสำคัญ ซึ่งถ้าให้คะแนนวิธีประเพณีนิยม และต้องการให้แบบสอบนั้นมีค่าความเที่ยงสูงเท่ากับค่าความเที่ยงที่เพิ่มขึ้น จะต้องเพิ่มข้อสอบอีกประมาณ 49% ในทำนองเดียวกัน Raffield พบว่า ความเที่ยงและความตรงเชิงพยากรณ์ของแบบสอบที่ให้คะแนนโดยใช้มาตราวัดวิธี Guttman's Scale Analysis มีค่าสูงกว่าค่าความเที่ยงและความตรงเชิงพยากรณ์ที่คำนวณจากคะแนนที่ให้โดยวิธีประเพณีนิยมอย่างมีนัยสำคัญ เมื่อใช้คะแนนปลายปีเป็นเกณฑ์สำหรับการพยากรณ์

แต่อย่างไรก็ตามเมื่อ Patnaik และ Traub (patnaik and Traub, 1973: 281) ทำการศึกษาผลของการให้คะแนนต่อค่าความเที่ยงและความตรงเชิงพยากรณ์ในปี 1973 กับนักศึกษาระดับบัณฑิตศึกษา จำนวน 61 คน โดยใช้แบบสอบวัดเชาวน์ปัญญาที่เป็นแบบสอบแบบเลือกตอบ ชนิดมี 5 ตัวเลือก และใช้มาตราวัดจากวิธีการเปรียบเทียบรายคู่ของ Thurstone เป็นคะแนนของตัวเลือกแต่ละตัว ผลปรากฏว่า การให้คะแนนดังกล่าวทำให้แบบสอบมีค่าความเที่ยงแบบแบ่งครึ่งสูงกว่าค่าความเที่ยงที่คำนวณจากคะแนนที่ให้โดย วิธีประเพณีนิยมอย่างมีนัยสำคัญ ($p < .05$) และมีค่าความตรงเชิงพยากรณ์ที่ใช้ผลการสอบสัมฤทธิ์ผลวิชาต่าง ๆ เป็นเกณฑ์ต่ำกว่าค่าความตรงชนิดดังกล่าวอย่างมีนัยสำคัญ ($p < .05$) การที่ค่าความเที่ยงของแบบสอบเพิ่มขึ้น แต่ค่าความตรงเชิงพยากรณ์ลดลงจากค่าความตรงเชิงพยากรณ์ที่คำนวณจากคะแนนที่ให้โดยวิธีประเพณีนิยมนั้น Patnaik และ Traub อธิบายว่าเป็นเพราะว่าการให้คะแนนที่ไม่ใช้วิธีประเพณีนิยม ทำให้แบบสอบค่าสหสัมพันธ์ระหว่างข้อสอบเพิ่มขึ้น เพราะคะแนนมีการกระจายเพิ่มขึ้นมากกว่าเดิม ด้วยเหตุนี้จึงทำให้คะแนนจากแบบสอบมีความสัมพันธ์กับเกณฑ์ต่ำ เช่นเดียวกับกรณีค่าสัมประสิทธิ์สหสัมพันธ์สหคูณของหัวแปรที่มีความสัมพันธ์ซึ่งกันและกันสูง มีความสัมพันธ์กับเกณฑ์ต่ำ (Patnaik and Traub, 1977: 285)

(3) วิธีอาศัยความเชื่อมั่นของผู้สอบต่อตัวเลือกแบบสอบ

วิธีการให้คะแนนโดยอาศัยความเชื่อมั่นของผู้สอบต่อตัวเลือกของแบบสอบ อาจแบ่งออกได้ 3 วิธีคือ 1) วิธีคิดคะแนนตามความมั่นใจในการตอบ 2) วิธีคิดคะแนนโดยหักค่าการเดา และ 3) วิธีคิดคะแนนโดยวิธีการกำจัดตัวเลือก สำหรับการคิดคะแนนวิธีที่ 1 นั้น แต่ละข้อถ้าผู้สอบตอบถูกในการเลือกตอบครั้งแรกจะให้ 3 คะแนน ถ้าเลือก 2

ครึ่ง ตอบถูกได้ 2 คะแนน ถ้าเลือก 3 ครั้ง ตอบถูกได้ 1 คะแนน และถ้าเลือก 4 ครั้งตอบถูก จะได้ 0 คะแนน (ข้อสอบมี 4 ตัวเลือก) วิธีที่ 2 ให้คะแนนโดยอาศัยสูตรการคำนวณที่หักค่า การเดา และวิธีที่ 3 นั้น ผู้สอบเลือกตอบหลาย ๆ ตัวเลือกในแต่ละข้อ ถ้าตอบถูกจะได้ 1 คะแนน แต่ถ้าตอบผิดจะได้ -2 คะแนน (ข้อสอบมี 4 ตัวเลือก)

จากการศึกษาของ Diamond (Diamond, 1975: 129) ในปี 1975 เกี่ยวกับการให้คะแนนทั้ง 3 วิธี กับกลุ่มตัวอย่าง 84 คน ในระดับมัธยมศึกษาด้วย แบบสอบวิชาวิทยาศาสตร์ก่อนข้างง่าย และมีอำนาจจำแนกต่ำ ผลปรากฏว่า แบบสอบที่ให้คะแนนทั้ง 3 วิธี ถึงแม้ว่ามีค่าความเที่ยงแบบ Flanagan สูงกว่าค่าความเที่ยงที่คำนวณจากคะแนนที่ให้โดย วิธีประเพณีนิยม แต่ค่าความตรงเชิงพยากรณ์ของแบบสอบที่ใช้ผลการสอบประจำภาคเป็นเกณฑ์ ไม่แตกต่างจากค่าความตรงเชิงพยากรณ์ที่คำนวณจากคะแนนที่ให้โดยวิธีประเพณีนิยม ผลการศึกษานี้มีความคล้ายคลึงกับผลการศึกษาของ Hanna ในปีเดียวกัน (Hanna, 1975: 145) ซึ่งได้ ให้คะแนนตัวเลือกแบบสอบแบบเลือกตอบชนิด 4 ตัวเลือก 2 ฉบับ ซึ่งยาว 10 ข้อและ 30 ข้อ และใช้กับกลุ่มตัวอย่าง 85 คน โดยวิธีให้คะแนนตามความมั่นใจในการตอบ ผลการศึกษพบว่า แบบสอบที่ให้คะแนนโดยวิธีประเพณีนิยมเล็กน้อย แต่ค่าความตรงเชิงพยากรณ์ซึ่งใช้คะแนนสอบ ศีความและการเขียนรายงานเป็นเกณฑ์กลับมีค่าต่ำกว่าค่าความตรงเชิงพยากรณ์ที่ได้จากคะแนนที่ ให้โดยวิธีประเพณีนิยม ทั้งนี้ Hanna จึงสรุปว่าการให้คะแนนวิธีประเพณีนิยมให้ผลการสอบดีกว่า วิธีการให้คะแนนตามความมั่นใจในการตอบ เพราะการสอบนั้นควรคำนึงถึงความตรงของแบบสอบ มากกว่าความเที่ยง

นอกจากนี้ ในปี 1975 Hakstian และ Kansup (Hakstian and Kansup, 1975b: 231) ยังทำการศึกษาผลการให้คะแนนโดยวิธีให้คะแนนตามความมั่นใจ ในการตอบ และวิธีให้คะแนนโดยวิธีการกำจัดตัวเลือก โดยใช้นักเรียนระดับ grade 9 จำนวน 1,023 คน เป็นกลุ่มตัวอย่าง ผลการศึกษพบว่า สอดคล้องกับผลการศึกษาของ Diamond และ Hanna กล่าวคือ การให้คะแนนทั้ง 2 วิธี ทำให้แบบสอบมีค่าความเที่ยงสูงขึ้นเล็กน้อย แต่ให้ค่า ความตรงเชิงพยากรณ์ต่ำกว่าค่าความตรงที่คำนวณจากคะแนนที่ให้โดยวิธีประเพณีนิยม

อนึ่ง เพื่อเป็นการตรวจสอบผลการศึกษาที่หาไว้ในปี 1975

Hanna ได้ทำการศึกษาเกี่ยวกับผลการให้คะแนน โดยวิธีให้คะแนนตามความมั่นใจในการตอบอีก ในปี 1977 ครั้งนี้ Hanna, 1977: 643) ใช้แบบสอบถามเลือกตอบยาว 18 ข้อ 2 ฉบับ คือ วิชาวิทยาศาสตร์และคณิตศาสตร์ และใช้นักเรียน grade 5 และ 6 จำนวน 1,391 คน เป็น ตัวอย่าง ผลการศึกษาปรากฏว่า มีความสอดคล้องกับผลการศึกษาในปี 1975 คือ แบบสอบถามมีค่าความเที่ยงสูงกว่าค่าความเที่ยงที่คำนวณจากคะแนนที่ให้โดยวิธีประเมินนิยมอย่างมีนัยสำคัญ แต่ให้ค่าความตรงไม่แตกต่างจากค่าความตรงที่คำนวณจากคะแนนที่ให้โดยวิธีประเมินนิยม

5.1.2 การให้ค่าน้ำหนักคะแนนแบบข้อสอบแตกต่างกัน

นอกจากให้คะแนนตัวเลือกของข้อสอบแต่ละข้อแตกต่างกันแล้ว

นักทดสอบยังสนใจที่จะให้คะแนนเฉพาะข้อสอบแตกต่างกันด้วย เพื่อศึกษาว่าการให้คะแนนดังกล่าวนี้จะทำให้แบบสอบถามมีค่าความเที่ยงและความตรงสูงกว่าค่าความเที่ยงและความตรงที่ได้จากคะแนนที่ให้โดยวิธีประเมินนิยมหรือไม่

จากการศึกษาวรรคศัพท์ที่เกี่ยวข้อง พบว่าการให้คะแนนข้อสอบแตกต่างกันมีหลายวิธีดังนี้

1) วิธีอาศัยความเห็นของผู้เชี่ยวชาญ

การให้คะแนนข้อสอบแตกต่างกัน โดยอาศัยความเห็นของผู้เชี่ยวชาญที่เกี่ยวข้อง เป็นวิธีที่ Corey ให้นามการศึกษาในปี 1930 และ Ode11 ให้ตรวจสอบผลการศึกษานั้น ในปี 1931 (อ้างจาก Wang and Stanley, 1970: 688) Corey ให้ผู้เชี่ยวชาญทางจิตวิทยา 6 คน พิจารณาให้คะแนนข้อสอบแตกต่างกันตามความสำคัญของสิ่งที่มีผู้ทดสอบ โดยไม่ต้องปรึกษากัน ปรากฏว่าคะแนนของแบบสอบถามมีคะแนนรายข้อแตกต่างกันทั้ง 6 ข้อ มีค่าสัมประสิทธิ์สหสัมพันธ์กับคะแนนที่ให้โดยวิธีประเมินนิยม ระหว่าง .836 ถึง .960 จึงนับว่าสูงมาก ผลของการศึกษาของ Ode11 ในปีต่อมาสอดคล้องกับผลการศึกษาของ Corey ทุกประการ ทำให้เขาสรุปว่า การให้คะแนนรายข้อแตกต่างกัน ไม่มีประโยชน์ในการสอบ

2) วิธีใช้ค่าถดถอยพหุคูณ

ในปี 1942 Guilford, Lovell และ Williams (Guilford and Others, 1942: 15) ได้ทำการศึกษาผลของการให้คะแนนรายข้อแตกต่างกัน โดยอาศัยค่าประมาณของค่าถดถอยพหุคูณ ซึ่งคำนวณจากสูตรพิเศษที่คิดขึ้น คือ

$$w_i = \frac{r_{ic}\sigma_c}{\sigma_i}$$

r_{ic} = ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างข้อสอบกับเกณฑ์

σ_c = ค่าความเบี่ยงเบนมาตรฐานของเกณฑ์

σ_i = ค่าความเบี่ยงเบนมาตรฐานของรายข้อ

ผลจากการให้คะแนนรายข้อแตกต่างกันด้วยสูตรดังกล่าว แล้ว กับแบบสอบแบบเลือกตอบ 3 ฉบับ ที่มีความยาว 20 ข้อ 50 ข้อ และ 100 ข้อ Guilford และคณะพบว่า ความเที่ยงแบบแบ่งครึ่ง และความตรงเชิงพยากรณ์ของแบบสอบ ซึ่งใช้คะแนนแบบสอบสัมฤทธิ์ผลยาว 100 ข้อ เป็นเกณฑ์ มีค่าไม่แตกต่างกันจากค่าความเที่ยงและค่าความตรงเชิงพยากรณ์ที่คำนวณจากคะแนนที่ให้โดยประเพณีนิยมอย่างมีนัยสำคัญ ดังนั้น Guilford และคณะ จึงสรุปว่า การให้คะแนนรายข้อแตกต่างกันได้ผลไม่สู้ดีกว่ากับความยุ่งยากในการคำนวณ

นอกจากนี้แล้ว ยังมีนักทดสอบอีกจำนวนมากที่ให้คะแนนรายข้อแตกต่างกันด้วยวิธีต่าง ๆ ปรากฏว่า ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนที่ได้จากการให้คะแนนรายข้อต่างกับคะแนนที่ให้โดยวิธีประเพณีนิยม มีความสัมพันธ์กันสูงมาก เช่น Douglas และ Spencer (1923) พบว่า ค่าดังกล่าว สัมพันธ์กันระหว่าง .98 - .991 Holzinger (1923) west (1924) Peatman (1930) Ruch และ Meyer (1931) Pothoff และ Barnett (1932) และคนอื่น ๆ อีกต่างก็พบว่า ความสัมพันธ์ระหว่างคะแนนดังกล่าวมีความสัมพันธ์กันสูงระหว่าง .93 - .99 (อ้างจาก Wang and Stanley 1970: 689)

ดังนั้น นักทดสอบจำนวนมากจึงมีความเห็นว่า การให้คะแนนรายข้อของแบบสอบต่างกันไม่มีประโยชน์ เพราะว่าคะแนนที่ได้มีความสัมพันธ์กับคะแนนที่ได้โดยวิธีประเมินนิยมมาก ด้วยเหตุนี้ Wang and Stanley (Wang and Stanley, 1970:699) ผู้ซึ่งศึกษาและรวบรวมวรรณคดีที่เกี่ยวกับการให้ค่าน้ำหนักคะแนนแบบสอบวิธีต่าง ๆ จนกระทั่งถึงปี 1970 ได้สรุปว่า การให้คะแนนวิธีต่าง ๆ ไม่มีผลต่อคุณภาพของแบบสอบ หรือถ้ามีผลก็จะมีเพียงเล็กน้อยเท่านั้น เพราะว่าในทางปฏิบัติการให้คะแนนเหล่านั้นมักนำไปใช้ไม่ได้ แต่ก็ยังพอมีความหวังอีก 2 วิธี ที่จะนำมาใช้ในการให้คะแนนข้อสอบ คือ

1. วิธีให้คะแนนโดยอาศัยรูปแบบความสามารถแท้
2. วิธีให้คะแนนโดยอาศัยรูปแบบการถดถอยพหุคูณ
- 3) วิธีให้คะแนนโดยอาศัยรูปแบบความสามารถแท้

Lord (Lord, 1968: 989) ได้ทำการวิเคราะห์ผลการสอบวิชา Scholastic Aptitude Test โดยให้คะแนนของข้อสอบต่างกันตามแนวคิดของ Bimbaum โดยอาศัยรูปแบบโลจิสติกใช้พารามิเตอร์ 3 ตัว ทั้งนี้โดยการกำหนดค่าน้ำหนักคะแนนของผู้สอบแต่ละคน และแต่ละข้อต่างกันตามความสามารถของแต่ละบุคคล ซึ่งจะทราบได้จากโค้งลักษณะของข้อสอบ (Item Characteristic Curve) แนวคิดดังกล่าวนี้ ต่อมา Lord (Lord, 1977 : 123) ได้ย้่าว่า เป็นแนวคิดที่สำคัญกล่าวคือเขาคิดว่า "ข้อสอบข้อที่ให้คะแนนมากสำหรับคนเก่ง ไม่จำเป็นต้องให้คะแนนมากสำหรับคนไม่เก่ง" และสูตรในการคำนวณหาค่าน้ำหนักคะแนนที่เหมาะสมแต่ละข้อคือ

$$W_i = \frac{a_i}{1 - c_i} \cdot \frac{P_i - c_i}{P_i}$$

เมื่อ a_i = ค่าอำนาจจำแนกของข้อสอบข้อที่ i

c_i = ค่าการเดาของข้อสอบที่ i

p_i = ค่าความยากง่ายของข้อสอบที่ i

ผลของการวิเคราะห์ผลการสอบ Lord พบว่า

1. เมื่อความสามารถของผู้สอบมีมาก โท้งแสดงค่าน้ำหนักคะแนนจะขนานกับแกนนอน
2. เมื่อความสามารถของผู้สอบต่ำ โท้งแสดงค่าน้ำหนักคะแนนจะขึ้นอยู่กับความยากของข้อสอบ

ดังนั้น เขาจึงสรุปว่า ในกรณีของผู้สอบที่มีความสามารถสูงมาก การให้คะแนนข้อสอบรายข้อวิธีประเพณีนิยมจะดีกว่า เพราะว่าค่าน้ำหนักคะแนนระดับสูงจะคงที่ แต่ถ้าผู้สอบที่มีความสามารถต่ำหรือปานกลาง การให้คะแนนข้อสอบต่างกันมีความถูกต้องมากกว่า

อย่างไรก็ตาม เนื่องจากผลการวิจัยเกี่ยวกับการให้คะแนนตัวเลือกหรือข้อสอบต่างกันได้ผลไม่แน่นอน ทอห้จะสรุปได้ด้วยความเชื่อมั่นว่า การให้คะแนนดังกล่าวทำให้แบบสอบมีความเที่ยงและความตรงเพิ่มขึ้นได้จริง ในปี 1975 (Lord, 1975 :10) นักวิจัยที่มีชื่อเสียงมากคนหนึ่ง ในปัจจุบันนี้ ได้สรุปความคิดเห็นเกี่ยวกับเรื่องนี้ที่น่าสนใจมากไว้ดังนี้คือ

1. การให้คะแนนวิธีประเพณีนิยม และผู้สอบทำไม่ครบหมดทุกข้อ ผลของการสอบจะเป็นตัวทำนายได้ไม่ดีเท่ากับการให้คะแนนโดยวิธีให้คะแนนข้อสอบต่างกัน
2. การให้คะแนนโดยวิธีให้คะแนนข้อสอบต่างกันแล้ว ทำให้แบบสอบมีค่าความเที่ยงหรือค่าความตรงเพิ่มขึ้นแม้เพียงเล็กน้อย เช่น .90 เป็น .91 หรืออย่างน้อยประมาณ 1/9 จากจำนวนข้อสอบ
3. การศึกษาเปรียบเทียบค่าความเที่ยงของแบบสอบ โดยการให้คะแนนข้อสอบต่างกันไม่มีประโยชน์ เว้นเสียแต่ว่าจะสามารถพิสูจน์ว่า วิธีหนึ่งมาเปรียบเทียบกับวัตถุประสงค์เฉพาะ (traits) เที่ยงกันอย่างแท้จริง

ต่อมาในปี 1980 Lord (Lord, 1980: 74) ได้ศึกษาเปรียบเทียบผลของการให้คะแนนข้อสอบแต่ละข้อแตกต่างกันอีก โดยใช้แบบสอบทางภาษาชื่อ SCAT II Form 2A ซึ่งมี 4 ตัวเลือก จำนวน 40 ข้อ เป็นเครื่องทื่อ เขาพบว่าการให้คะแนนข้อสอบแต่ละข้อโดยวิธีใช้ค่าน้ำหนักคะแนนที่เหมาะสม (w) ให้สาระของการสอบมากที่สุด แต่สำหรับผู้สอบที่มีความสามารถในระดับต่ำ การให้คะแนนวิธีประเพณีนิยม ให้สาระค้งกล่าวน้อยที่สุด ถ้าการให้คะแนนค้งกล่าวโดยอาศัยรูปแบบโลจิสติกใช้พารามิเตอร์ 3 ตัวแล้ว การให้คะแนนวิธีประเพณีนิยมให้สาระการสอบได้ค้งกว่า ทั้งนี้เพราะรูปแบบค้งกล่าวไม่ได้ใช้ค่าการเคาค่วย และในปี 1982 Hsu และ Choosak (Hsu and C. Khampalikit, 1982: 7) ได้เปรียบเทียบผลการให้คะแนนโดยวิธีใช้ค่าน้ำหนักคะแนนที่เหมาะสม (w) กับการให้คะแนนวิธีประเพณีนิยมแก่ข้อสอบแต่ละข้อสำหรับวิชาภาษาจีนและวิชาประวัติศาสตร์ เขาพบว่าการให้คะแนนวิธีแรกทำให้แบบสอบมีค่าอำนาจจำแนกและค่าความเที่ยงสูงกว่าค่าอำนาจจำแนก และค่าความเที่ยง ที่คำนวณจากคะแนนที่ให้โดยวิธีประเพณีนิยม แต่ไม่ให้เกิดผลเกี่ยวกับผลที่เกิข้ขึ้นกับค่าความตรงของแบบสอบค้งกล่าว

อนึ่ง สำหรับการให้คะแนนโดยวิธีอาศัยรูปแบบการถดถอยพหุคูณ นั้น เป็นแนวคิดใหม่ที่มีงกำหนดคะแนนแก่ผู้สอบแต่ละคนแตกต่างกันตามความสามารถของแต่ละบุคคล รายละเอียดของวิธีนี้ยังไม่อาจหาได้ ส่วนการให้คะแนนรายข้อหรือแต่ละตัวเลือก โดยวิธีใช้ค่าถดถอยพหุคูณนั้น นักทดสอบใช้กันมานานแล้ว แต่ปรากฏว่ามักทำให้แบบสอบมีค่าความเที่ยงและความตรงเพิ่มขึ้นอย่างมีนัยสำคัญเฉพาะแบบสอบที่มีขนาดสั้นเท่านั้น (Wang and Stanley, 1970: 670)

5.2 งานวิจัยเกี่ยวกับการสอบวัด และการคัดเลือก

วิทยาลัยและมหาวิทยาลัยในสหรัฐอเมริกา ใช้แบบทดสอบเป็นเครื่องมือในการคัดเลือกนิสิตเข้าศึกษาตั้งแต่ศตวรรษที่ 17 เมื่อ ค.ศ. 1636 มหาวิทยาลัยฮาร์เวต ใช้แบบทดสอบที่เลือกรายบุคคลเป็นแห่งแรก จนถึง ค.ศ. 1900 มีการก่อตั้งคณะกรรมการสอบคัดเลือกเข้าศึกษาในวิทยาลัย (The College Entrance Examination Board หรือ CEEB) เพื่อสร้างแบบทดสอบวัดผลสัมฤทธิ์ทางวิชาการ สำหรับใช้คัดเลือกนักเรียนเข้าศึกษาในวิทยาลัย และเป็นที่ยอมรับของสถาบันชั้นอุดมศึกษาอื่น ๆ ต่อมา (Robert L, Ebel 1969: 8)

ชวาล แพร์ตกุล (2507 : 24) กล่าวว่า แบบทดสอบคัดเลือกเป็นแบบทดสอบชนิดหนึ่งที่ใช้ผลการสอบเพื่อพยากรณ์ โดยนำคะแนนที่ได้จากการสอบไปเป็นเกณฑ์ช่วยชี้ว่าบุคคลผู้นั้นควรเรียนอะไรจึงจะประสบความสำเร็จ หรืออาจไม่ประสบความสำเร็จ ค.ศ. 1910-1920 เป็นระยะที่แบบทดสอบด้านจิตวิทยารุ่งเรืองมาก มีผู้นำแบบทดสอบความถนัด (Aptitude Test) ใช้วัดความแตกต่างบุคคลมาเป็นเกณฑ์ในการคัดเลือกนิสิตเข้าค่าย และนิยมใช้กันอย่างกว้างขวาง (Ahmen and Glock 1958: 18-19)

อย่างไรก็ตามเครื่องมือที่ใช้วัดเพื่อคัดเลือกและพยากรณ์ผลสำเร็จทางการเรียนรู้ของผู้เรียน ส่วนใหญ่คือแบบทดสอบวัดผลสัมฤทธิ์และแบบทดสอบความถนัดส่วนแบบตรวจสอบความสนใจและแบบตรวจสอบบุคลิกภาพนั้น มักจะใช้ประโยชน์ด้านจิตวิทยา และการแนะแนว แบบทดสอบต่าง ๆ ที่ใช้เป็นเครื่องมือในการคัดเลือกนักเรียน นักศึกษาเข้าเรียนในสถาบันการศึกษาต่าง ๆ ใ้รับผิดตามผลเพื่อศึกษาประสิทธิภาพในการคัดเลือกอยู่เสมอ ทั้งผลการวิจัยทั้งในประเทศไทยและจากต่างประเทศที่รวบรวมมา ดังต่อไปนี้

ชวาล แพร์ตกุล (Chawal Paeratakul: 1961) ได้ทำการวิจัยที่มหาวิทยาลัยอินเดียนา (Indiana University) เรื่องประสิทธิภาพของแบบสอบปฐมนิเทศนักศึกษาเข้าใหม่ แบบสอบที่นำมาวิจัยเป็นแบบสอบที่มหาวิทยาลัยใช้ทดสอบนักศึกษา ก่อนเข้าเรียน โดยวัดความรู้ทางวิทยาศาสตร์ คณิตศาสตร์ คำศัพท์ เรียงความ และแบบสอบมาตรฐาน STEF (Sequential Tests of Education Progress) ผลการวิจัยพบว่า แบบสอบมาตรฐานเป็นตัวทำนายสัมฤทธิ์ผลทางการเรียนในภาคเรียนแรกของนักศึกษาได้ดีที่สุด มีค่าสัมประสิทธิ์สหสัมพันธ์พหุคูณเท่ากับ .57

พจน์ สะเพียรชัย (Poj Sapianchaiy, 1964 : 3210) ได้ศึกษาถึงประสิทธิภาพของแบบทดสอบคัดเลือกเข้าวิทยาลัยวิชาการศึกษา ประสานมิตร ปีการศึกษา 2504 ซึ่งใช้แบบทดสอบวัดผลสัมฤทธิ์จำนวนห้าฉบับ ได้แก่ สังคมศึกษา ภาษาไทย ภาษาอังกฤษ วิทยาศาสตร์ และคณิตศาสตร์ ผลการวิจัยปรากฏว่า ค่าสหสัมพันธ์ของแบบทดสอบแต่ละฉบับกับคะแนนเฉลี่ยในแต่ละภาคเรียนมีค่าค่อนข้างต่ำ และค่าสหสัมพันธ์ของแบบทดสอบทั้งชุดกับคะแนนเฉลี่ยในแต่ละภาคมีค่าตั้งแต่ .45 - .58

ภิญโญ สาร (2505) วิจัยหาความสัมพันธ์ระหว่างคะแนนสอบคัดเลือกของสภาการศึกษาแห่งชาติปีการศึกษา 2505 กับคะแนนสอบไล่กลางปีของนิสิตชั้นปีที่ 1 จุฬาลงกรณ์มหาวิทยาลัย โดยตัวพยากรณ์กับตัวเกณฑ์เป็นวิชาเดียวกัน พบว่าวิชาภาษาอังกฤษ ให้ค่าสหสัมพันธ์สูงสุดมีค่าอยู่ระหว่าง .55 ถึง .82 ภาษาฝรั่งเศส ให้ค่าสหสัมพันธ์ต่ำสุด เท่ากับ .43

โครงการวิจัยการเลือกสรร วิทยาลัยวิชาการศึกษา ใต้รายงานไว้เมื่อ พ.ศ.2511 ถึงการสร้างแบบวัดความถนัดทางการเรียน สำหรับใช้สอบคัดเลือกนักเรียนฝึกหัดครู ระดับประกาศนียบัตรวิชาการศึกษา ชุดที่ 1 ประกอบด้วยแบบสอบ 5 ฉบับ ได้นำไปทดลองใช้สอบกับนักเรียนชั้นมัธยมศึกษาปีที่ 3 ที่สุ่มมา 11 จังหวัด นำผลการสอบมาวิเคราะห์รายข้อหาสัมประสิทธิ์แห่งความตรงเชิงพยากรณ์ โดยใช้คะแนนสัมฤทธิ์ผลทางการเรียนเป็นเกณฑ์ ผลการวิเคราะห์พบว่า มีแบบสอบที่ใช้ได้ 2 ฉบับ คือ ฉบับคำศัพท์ และฉบับการคำนวณ เพราะสามารถใช้แบบสอบคัดเลือกนักเรียนได้ผลดีเท่ากับใช้ทั้ง 5 ฉบับ จึงไม่จำเป็นต้องใช้ทั้ง 5 ฉบับ

ในรายงานการสร้างแบบทดสอบความถนัดทางการเรียนระดับประกาศนียบัตร วิชาการศึกษาชั้นสูง ชุดที่ 1 พ.ศ. 2511 ของโครงการวิจัยการเลือกสรร ปรากฏว่าแบบทดสอบความถนัดทางการเรียน 4 ฉบับ ได้นำไปทดลองใช้กับกลุ่มตัวอย่าง นักศึกษา ประกาศนียบัตรวิชาการศึกษาชั้นสูง ปีที่ 1 ในวิทยาลัยครูที่เป็นกลุ่มตัวอย่าง 3 แห่งนั้น มีค่าความเชื่อมั่นและความเที่ยงตรงสูง พอที่จะนำไปทำการคัดเลือกที่มีประสิทธิภาพดี เมื่อเลือกใช้แบบทดสอบเพียง 3 ฉบับ คือ แบบทดสอบคำศัพท์ แบบทดสอบการคำนวณ และแบบทดสอบอนุกรมอุปไมย ปรากฏว่าความเที่ยงตรงรวม (ค่าสหสัมพันธ์ทุกคู่) สูงกว่าการใช้แบบทดสอบทั้งหมด 4 ฉบับ คือ ค่าความเที่ยงตรงรวม .3353 กล่าวคือ ถ้าใช้แบบทดสอบชุดนี้ในการคัดเลือก เมื่อรับไว้ร้อยละ 5 ของจำนวนผู้เข้าสอบคัดเลือก โอกาสจะเรียนให้สำเร็จสูงถึง 85%

ในปีเดียวกันนี้ โครงการวิจัยการเลือกสรร ให้ศึกษาความเที่ยงตรงเชิงพยากรณ์ของแบบทดสอบวิชาสามัญและวิชาความถนัดทางการเรียนที่ใช้คัดเลือกนิสิตเข้าเรียน ในระดับชั้นปีที่ 1 วิทยาลัยวิชาการศึกษาประสานมิตร ปีการศึกษา 2510 และยังให้ศึกษาความเที่ยงตรงเชิงพยากรณ์ของผลการสอบคัดเลือกนิสิตปีที่ 3 ปีการศึกษา 2511 อีกด้วย เกณฑ์ที่ใช้ในการหาความเที่ยงตรงเชิงพยากรณ์คือ ผลสัมฤทธิ์ทางการเรียน โดยใช้ผลเฉลี่ยระดับคะแนน 4

ภาคเรียนสำหรับนิสิตปีที่ 1 และใช้คะแนนผลการเรียนวิชาเอก สำหรับนิสิตปีที่ 3 ผลการวิจัยพบว่า แบบทดสอบวิชาสามัญที่ใช้สอบคัดเลือกนิสิตปีที่ 1 เข้าเรียนปีการศึกษา 2510 มีประสิทธิภาพในการคัดเลือกประมาณ 93 % ส่วนแบบทดสอบความถนัดทางการเรียนนั้น ค่าสถิติไม่ได้แสดงถึงประสิทธิภาพในการคัดเลือก และในการสอบคัดเลือกนิสิตชั้นปีที่ 3 ปีการศึกษา 2511 ที่ใช้แบบทดสอบความถนัดทางการเรียนและแบบทดสอบวิชาเอกนั้น แบบทดสอบวิชาเอกมีประสิทธิภาพในการคัดเลือกให้ดีกว่าแบบทดสอบความถนัดทางการเรียน ถ้าหากพิจารณาแต่การสอบเฉพาะวิชาเอกต่าง ๆ ประสิทธิภาพของการคัดเลือกมีประมาณ 68% ถึง 98% (รับนิสิตได้ประมาณ 5% ของจำนวนผู้สมัครสอบเข้า) ส่วนแบบทดสอบความถนัดทางการเรียนมีประสิทธิภาพในการคัดเลือกต่ำ

ลอมสมัย พิทักษ์ (ลอมสมัย พิทักษ์ 2512: 75) ได้ศึกษาความสัมพันธ์ระหว่างคะแนนสอบคัดเลือก คะแนนผลสัมฤทธิ์ทางการเรียนและทัศนคติต่ออาชีพครูของนักเรียน นักศึกษาระดับ ป.กศ. ผลการวิจัยปรากฏว่า คะแนนจากแบบทดสอบคัดเลือกทุกฉบับมีความสัมพันธ์กับคะแนนเฉลี่ยสะสมไม่สูงนัก และแบบทดสอบคัดเลือกวัดผลสัมฤทธิ์ทางการเรียนมีความสัมพันธ์กับคะแนนเฉลี่ยสะสมสูงกว่าแบบทดสอบคัดเลือกวัดความถนัดทางการเรียน

ในปี พ.ศ. 2514 นิภา เมธาวิชัย ศึกษาความสัมพันธ์ระหว่างผลการทดสอบความถนัดทางวิชาการและผลการสอบคัดเลือกวิชาเอกกับความสำเร็จในการศึกษาของนักศึกษาวិทยาลัยวิชาการศึกษา บางแสน โดยศึกษาว่า การทำนายคะแนนสัมฤทธิ์ผลทางการเรียนวิชาเอกที่เหมาะสมควรใช้คะแนนสอบคัดเลือกวิชาเอกหรือใช้คะแนนทดสอบความถนัดทางวิชาการหรือใช้คะแนนทั้ง 2 ชนิดร่วมกันในการคัดเลือกนักศึกษาวิเคราะห์หาความสัมพันธ์ โดยวิธีหาค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน และแบบสหคูณแต่ละแบบเปรียบเทียบความแตกต่างโดยใช้เรโซเนท (F-ratio) ได้พบว่า แบบสอบคัดเลือกวิชาเอกส่วนใหญ่ทำนายความสำเร็จในการศึกษาของนักศึกษาบางสาขาวิชา ได้มากกว่าแบบทดสอบความถนัดทางวิชาการทั้ง 5 ฉบับ และแบบทดสอบคัดเลือกวิชาเอกร่วมกันแบบทดสอบความถนัดทางวิชาการชุดเดียวกับฉบับ หรือแบบทดสอบความถนัดทางวิชาการร่วมกัน 2 ฉบับ ทำนายความสำเร็จในการศึกษาของนักศึกษาให้บางสาขาวิชาเท่านั้น เฉพาะแบบทดสอบความถนัดทางวิชาการร่วมกันตั้งแต่ 4 ฉบับขึ้นไป ไม่สามารถทำนายความสำเร็จในการศึกษาของนักศึกษาได้

ในปี พ.ศ. 2515 จรรยา มีวาสนา ให้นำแบบสอบคัดเลือกเข้าเรียน ฝึกหัดครูระดับประกาศนียบัตรวิชาการศึกษา ของวิทยาลัยครูสวนสุนันทา มาวิเคราะห์โดยใช้เทคนิค 27 % ในการแบ่งกลุ่มสูง กลุ่มต่ำ เพื่อวิเคราะห์แบบสอบเป็นรายข้อ แบบสอบคัดเลือกที่นำไปวิเคราะห์ประกอบด้วย 5 แบบสอบวิชา คือ วิชาความรู้ทั่วไป คณิตศาสตร์ ภาษาไทย ภาษาอังกฤษ และวิทยาศาสตร์ พบว่าแบบสอบคัดเลือกชุดนี้มีสัมประสิทธิ์แห่งความเที่ยงระหว่าง .3510 - .7682 และสัมประสิทธิ์แห่งความเที่ยงเชิงพยากรณ์ เมื่อใช้อันดับคะแนนเฉลี่ย (G.P.A.) เป็นเกณฑ์ เรียงตามลำดับวิชาที่ทำนายได้ดีที่สุดไปถึงไม่ดีที่สุด คือ แบบสอบวิชาภาษาอังกฤษ คณิตศาสตร์ วิทยาศาสตร์ ภาษาไทย สำหรับแบบสอบวิชาความรู้ทั่วไปนั้นใช้ทำนายเกณฑ์ไม่ได้เลย ในปีเดียวกันนี้ มีผู้ศึกษาหาความสัมพันธ์ระหว่างคะแนนสอบคัดเลือกกับคะแนนสัมฤทธิ์ผลทางการเรียนอีก 2 ราย คือ สุจิตต์ รักษ์เผ่า (2515) ได้ศึกษาความสัมพันธ์ระหว่างผลการสอบคัดเลือก วิชาความถนัดทางการเรียน และผลการสอบคัดเลือกวิชาเอกกับความสำเร็จในการศึกษาของนักศึกษาวิทยาลัยวิชาการศึกษา พิษณุโลก และฤกษ์จิต จินตณสินธิ (2515) ได้ศึกษาความสัมพันธ์แบบเดียวกันแต่ตัวอย่างประชากร เป็นนักศึกษาชั้น 2 ของวิทยาลัยวิชาการศึกษาพิษณุโลก พบว่าค่าสหสัมพันธ์ส่วนมากมีความสัมพันธ์ผกผัน หรือกลับกันโดยมีค่าอยู่ระหว่าง .3301 - .1972 แบบสอบคัดเลือกวิชาเอกทำนายได้ดีที่สุดสำหรับวิชาเอกคณิตศาสตร์ ส่วนวิชาเอกที่ทำนายได้เร็วที่สุดคือ วิชาเอกสังคมศึกษา สำหรับแบบสอบคัดเลือกวิชาความถนัดทางการเรียนกับอันดับคะแนนเฉลี่ย (G.P.A.) ของนักศึกษามีค่าความสัมพันธ์ระหว่าง -.0677 ถึง .6177 วิชาที่มีความสัมพันธ์ดีที่สุดกับแบบสอบวัดความถนัดทางการเรียนคือวิชาวิทยาศาสตร์ และวิชาฟิสิกส์ มีความสัมพันธ์ต่ำสุด

พ.ศ. 2516 วรรณา ปุณณโชติ และนิสิตบัณฑิตวิทยาลัย ได้วิเคราะห์แบบสอบคัดเลือกบุคคลเข้าศึกษาในสถาบันการศึกษาชั้นอุดมศึกษา ประจำปีการศึกษา 2516 รวม 16 วิชา ใช้กลุ่มตัวอย่างผู้เข้าสอบ 200 คน ในแต่ละหมวดวิชา ให้พบว่าระดับ ความยากของแบบสอบทุกวิชาอยู่ในช่วง .24 ถึง 1.00 ซึ่งก็อยู่ในเกณฑ์ ค่าสัมประสิทธิ์แห่ง ความเที่ยงของแบบสอบแต่ละวิชาที่คำนวณโดยสูตร กูเดอร์ ริชาร์ดสัน สูตรที่ 21 (Kuder Richardson Formula 21) มีค่าอยู่ระหว่าง .86 ถึง .955 จึงได้ข้อสังเกตว่า สำหรับแบบสอบที่มีความเที่ยงต่ำ ผลการศึกษาค่าสัมประสิทธิ์ แห่งความตรงเชิงพยากรณ์ของแบบสอบคัดเลือก โดย

ใช้คะแนนรวมที่สอบเข้าหลาย ๆ วิชาเป็นเกณฑ์ พบว่าแบบสอบทุกวิชามีความตรงเชิงพยากรณ์ อยู่ในช่วง .244 ถึง .933

ฉัตรชัย วัฒนานันท์ (ฉัตรชัย วัฒนานันท์ 2516: 63) ได้ศึกษาประสิทธิภาพของ แบบทดสอบที่ใช้คัดเลือกนักศึกษาที่เข้าเรียนในชั้นปีที่ 3 ในวิทยาลัยวิชาการศึกษา ประสานมิตร ปีการศึกษา 2514 จำนวน 12 สาขาวิชาเอก โดยใช้แบบทดสอบวัดผลสัมฤทธิ์ทางวิชาการหนึ่งฉบับ และแบบทดสอบวัดความถนัดทางการเรียนห้าฉบับ เป็นตัวพยากรณ์ผลสัมฤทธิ์ทางการเรียน ผลการวิจัยปรากฏว่า คะแนนแบบทดสอบทั้งหกฉบับสามารถพยากรณ์คะแนนเฉลี่ยได้ดีเพียงสองวิชาเอกคือ คณิตศาสตร์ และวิทยาศาสตร์ และแบบทดสอบวัดความถนัดทางการเรียนทั้งห้าฉบับ ไม่สามารถพยากรณ์ผลการเรียนในวิชาการศึกษา 3 สาขา คือ ประถมศึกษา มัธยมศึกษา และบริหารการศึกษากัน

ยรรยง ยรรยงเมธ (ยรรยง ยรรยงเมธ 2518 : 34) ได้ศึกษาความสัมพันธ์ระหว่างแบบทดสอบคัดเลือกนักศึกษา ป.กศ. ของวิทยาลัยครูนครสวรรค์ ปีการศึกษา 2518 จำนวน 6 ฉบับ ได้แก่ ความรู้ทั่วไป สังคมศึกษา วิทยาศาสตร์ ภาษาอังกฤษ ภาษาไทย และคณิตศาสตร์ กับคะแนนเฉลี่ย ผลการวิจัยปรากฏว่า เมื่อใช้แบบทดสอบห้าวิชา (โดยตัดวิชาสังคมศึกษาออกไป) ร่วมกันพยากรณ์ก็สามารถพยากรณ์ผลได้ไม่ต่างจากใช้ทั้ง 6 ฉบับร่วมกัน

วิรัตน์ หังใจรบ (วิรัตน์ หังใจรบ 2518 : 296) ได้ศึกษาถึงความสัมพันธ์ของแบบทดสอบวัดความถนัดทางการเรียน ที่หันเหคิต่ออาชีพพยาบาลและผลสัมฤทธิ์ทางการเรียนพยาบาลของนักเรียนพยาบาลศาสตร์ มหาวิทยาลัยมหิดล ปีการศึกษา 2517 ผลการวิจัยปรากฏว่า แบบทดสอบวัดความถนัดมีประสิทธิภาพต่ำในการใช้คัดเลือกนักศึกษาที่มีผลสัมฤทธิ์ทางการเรียนสูงและมีหันเหคิต่อวิชาชีพพยาบาล นอกจากนี้ผลสัมฤทธิ์ทางการเรียนพยาบาลและหันเหคิต่อวิชาชีพพยาบาลมีความสัมพันธ์กันน้อยมาก

ฝ่ายวิจัยและประเมินผล กองแผนงาน กรมการฝึกหัดครู ได้ติดตามผลการใช้แบบตรวจสอบคุณสมบัติและพื้นฐานของความเป็นครู ในการสอบคัดเลือกบุคคลเข้าศึกษาในวิทยาลัยครูระดับป.กศ. ชั้นสูง ปีการศึกษา 2522 ผลการวิจัยปรากฏว่า คะแนนบุคลิกภาพ

ของความเป็นครู และคะแนนความสนใจในอาชีพไม่สามารถพยากรณ์ผลสัมฤทธิ์ทางการเรียนได้ (ฝ่ายวิจัยและประเมินผล 2524: 167)

และในปีต่อมา จินตนา เวชมี (จินตนา เวชมี 2524 : 57) ได้ศึกษาความสัมพันธ์ระหว่างผลการสอบคัดเลือกกับผลสัมฤทธิ์ทางการเรียนภาคเรียนที่ 1 ของนักศึกษา วิทยาลัยครูพระนครศรีอยุธยา ระดับ ป.กศ.ชั้นสูง ปีที่ 1 ก.บ. ปีที่ 1 และ ก.บ. ปีที่ 3 ปีการศึกษา 2523 โดยใช้คะแนนจากแบบทดสอบวิชาเอก แบบทดสอบวิชาเอก แบบทดสอบวัดความสามารถพื้นฐานทางวิชาการ และแบบทดสอบวัดบุคลิกภาพพื้นฐานของผู้ที่จะเป็นครู เป็นตัวพยากรณ์ผลสัมฤทธิ์ทางการเรียนภาคเรียนที่ 1 ผลการวิจัยปรากฏว่า คะแนนจากแบบทดสอบแต่ละวิชาสามารถพยากรณ์ได้เพียงบางวิชาเอกเท่านั้น โดยแบบทดสอบวัดบุคลิกภาพพื้นฐานของผู้ที่จะเป็นครูสามารถพยากรณ์ผลการเรียนวิชาเอกพลศึกษาและเกษตรศาสตร์ ส่วนแบบทดสอบวัดความสามารถทางวิชาการสามารถพยากรณ์ผลการเรียนได้มากที่สุดซึ่งได้แก่ วิชาเอกพลศึกษา ดนตรีศึกษา นาฏศิลป์ และภาษาอังกฤษ

สำหรับงานวิจัยต่างประเทศที่เกี่ยวข้องและสำคัญดังนี้

ห่มหาวิทยาลัยมิชิแกน วอลเลซ (Wallace 1951 : 587 - 597) ศึกษาประสิทธิภาพของแบบทดสอบสัมฤทธิ์และแบบทดสอบความถนัดรวม 6 ฉบับ ที่ใช้ในการสอบคัดเลือก เมื่อ ค.ศ. 1951 พบว่าค่าสหสัมพันธ์หุคูณ ระหว่างคะแนนจากแบบทดสอบกับที่เป็นเกณฑ์มีเท่ากับ .5541

ต่อมาในปี 1954 แชพเพิล (Chappell 1954 : 724 - 725) ศึกษาประสิทธิภาพของแบบทดสอบสัมฤทธิ์ผล 4 ฉบับ ที่ใช้คัดเลือกนักเรียนของมหาวิทยาลัยซูริ พบว่าแบบทดสอบผู้คนมีสัมประสิทธิ์สหสัมพันธ์หุคูณกับคะแนนการเรียนของนิสิตเป็น .5170 ในปีเดียวกันแอนเดอร์สันและสตีแกน (Anderson and Stegaman 1954 : 722) ได้ศึกษาความเที่ยงตรงเชิงพยากรณ์ของแบบทดสอบเจ็ดฉบับที่มีเพื่อนักศึกษาเป็นแรกของฟอร์ทเฮร์ ผลการวิจัยปรากฏว่า แบบทดสอบทั้งเจ็ดฉบับ มีความสัมพันธ์กับคะแนนเฉลี่ยเป็นเกณฑ์สูง โดยแบบทดสอบภาษาอังกฤษ (Barratt Ryan English Test) เป็นตัวพยากรณ์ที่ดีที่สุด ส่วนแบบทดสอบอ่าน (Schrammel Gray High School and College Reading Test) เป็นตัวพยากรณ์ที่รองลงมา

แอดเกอร์สัน (Anderson 1956 : 5-9) ทำความเที่ยงตรงของแบบทดสอบคัดเลือกประเภทวิศวกรรมในทางวิชาการ โดยศึกษานักศึกษาปีที่ 1 มหาวิทยาลัยนอร์เทิร์นอิลลินอยส์ (Northern Illinois) ใช้คะแนนเฉลี่ยของภาคเรียนแรกเป็นเกณฑ์ ได้ว่าสหสัมพันธ์หุคูณ .718 ซึ่งสอดคล้องกับผลการวิจัยของ ชวาล แร้ทกุล ศึกษาบัณฑิตเข้ามหาวิทยาลัยอินเดียมา เมื่อ ค.ศ. 1961 ปรากฏว่า แบบทดสอบคัดเลือกมีค่าสหสัมพันธ์หุคูณกับคะแนนเฉลี่ยของภาคเรียนแรกได้เท่ากับ .5625 เป็นอันดับแรก และมีค่าสหสัมพันธ์หุคูณกับคะแนนเฉลี่ยของภาคเรียนที่สองรองลงมา การวิจัยของฮิวส์ (Hughes 1963 : 1448) ได้ผลสอดคล้องกับการศึกษาข้างต้นว่าแบบทดสอบที่ใช้ในการสอบคัดเลือกเข้ามหาวิทยาลัยเท็กซัส ต่างมีสหสัมพันธ์กับผลการเรียนของนิสิตอย่างมีนัยสำคัญ

ในปี 1959 ดามิโกและทอเจอร์ (Damico and togers 1959 : 611-616) ใช้แบบทดสอบ เอ็ม.เอ.ที. (The Multiple Aptitude Test) ประกอบด้วยแบบทดสอบความถนัด 9 แบบ ทำนายคะแนนในสาขาวิชาต่าง ๆ ของนักศึกษามหาวิทยาลัยฟลินต์ (Flint Junior College) เมื่อ ค.ศ. 1959 ปรากฏว่า ได้ค่าสหสัมพันธ์หุคูณระหว่าง .38 ถึง .85

ฮิวส์ (Huges, 1963: 1448) ได้ศึกษาถึงความสัมพันธ์ของแบบทดสอบคัดเลือกกับผลสัมฤทธิ์ทางการเรียนของนักศึกษามหาวิทยาลัยเท็กซัสห้าวิชาเอกผลการวิจัยปรากฏว่า ผลการสอบคัดเลือกมีความสัมพันธ์กับผลสัมฤทธิ์ทางการเรียนอย่างมีนัยสำคัญ และสามารถนำไปใช้ในการแนะแนวได้อย่างมีประสิทธิภาพ

ในปี ค.ศ.1965 แคมป์เบลล์ (campbell 1965 : 373) ได้ใช้แบบสอบวัดความถนัดทางวิชาการ (General Scholastic Aptitude Test) และแบบสอบความถนัดทางคณิตศาสตร์ (Mathematic Aptitude Test) กับแบบสอบเรียงความ (Written Expression) ใช้ศึกษานักศึกษ่วิทยาลัย โดยใช้คะแนนเฉลี่ยจากวิชาต่าง ๆ ความหลังสูตร 2 ปี เป็นเกณฑ์ ผลปรากฏว่า ค่าสหสัมพันธ์หุคูณระหว่างแบบสอบแต่ละชุดมีค่าอยู่ระหว่าง .45 ถึง .55

เอล (Elle 1967: 2875-2876) ได้ศึกษาความตรงเชิงพยากรณ์ของแบบสอบวัดความถนัดทางการเรียนและแบบสอบวัดสัมฤทธิ์ผลทางการเรียนโดยใช้อันดับคะแนนเฉลี่ย (G.P.A.) ในภาคเรียนแรกของนักศึกษาวิทยาลัยโอเรกอนใต้ (Southern Oregon College) เป็นเกณฑ์ พบว่า แบบสอบวัดความถนัดทางการเรียนเป็นตัวทำนายเกณฑ์ได้ดีที่สุด ส่วนแบบวัดสัมฤทธิ์ผลทางการเรียนที่ทำนายเกณฑ์ได้ดี คือ แบบสอบวิชาสังคมศึกษาและวิชาภาษาอังกฤษ ต่อมาได้มีการศึกษาค่าสัมประสิทธิ์แห่งความตรงเชิงพยากรณ์ของแบบสอบโอทิส ควิก สกอริง อะบิลิตี้ เทสต์ (Otis Quick Scoring Mental Ability Test) ซึ่งวิจัยโดย อาหมัด (Ahmad 1968: 947-949) โดยนำแบบทดสอบฉบับนี้ไปทดสอบนักศึกษาในปากีสถานตะวันตก นำข้อมูลไปวิเคราะห์หาสัมประสิทธิ์แห่งความตรงเชิงพยากรณ์ของแบบสอบเมื่อใช้อันดับคะแนนเฉลี่ย (G.P.A.) เป็นเกณฑ์ พบว่า แบบสอบฉบับนี้ สามารถทำนายความสำเร็จทางการเรียนของนักศึกษาปากีสถานตะวันตกได้

จากเอกสารและงานวิจัยที่เกี่ยวข้องตั้งแต่ส่วนที่ 1-5 ดังกล่าวแล้ว พบข้อประเด็นที่สำคัญ 3 ประการคือ

ประการที่ 1 สาขาวิชาหลัก หรือจำเป็นสำหรับนำมาสร้างเป็นแบบทดสอบคัดเลือก ส่วนใหญ่นั้น ได้แก่ วิชาเอก และวิชาสามัญพื้นฐานทั่วไป คือ คณิตศาสตร์ วิทยาศาสตร์ ภาษาอังกฤษ ภาษาไทย และสังคมศึกษา

ประการที่ 2 ปรากฏว่า การตัดสินใจคะแนนสอบคัดเลือกนั้น ส่วนใหญ่นำไปเน้นในแง่การศึกษาสิ่งที่ทำนายผลสัมฤทธิ์ทางการเรียน โดยใช้คะแนนดิบแต่อย่างเดียวยังไม่ได้แสวงหาระบบการคิดคะแนนตัดสินที่น่าจะก่อให้เกิดความยุติธรรมแก่ผู้สอบ และประสิทธิภาพการทำนายผลสัมฤทธิ์ทางการเรียนด้วยการคิดคะแนนวิธีอื่น

ประการที่ 3 การคิดคะแนนแบบทดสอบที่นักวัดผลคาดหวังว่าจะช่วยให้แบบทดสอบมีคุณภาพดีขึ้น และสามารถนำไปปฏิบัติได้ คือวิธีให้คะแนนอาศัยรูปแบบความสามารถแท้และให้คะแนนโดยอาศัยรูปแบบการลดอรรถประโยชน์