



วรรณคดีที่เกี่ยวข้อง

การวิเคราะห์ข้อสอบ (Item Analysis) คือเทคนิคสำหรับตรวจสอบว่าข้อคำถามเป็นรายข้อ ว่าแต่ละข้อมีคุณลักษณะตรงตามวัตถุประสงค์ที่เราต้องการหรือไม่ และมีมากเท่าใดกับแต่ละข้อมีคุณค่าเด่นค้อยในทางใดบ้าง และเป็นจำนวนเท่าใด การวิเคราะห์นี้จะกระทำได้ก็หลังจากที่ได้นำข้อสอบนั้นไปทดลองใช้จริง ๆ กับเด็กมาแล้วเท่านั้น ก็จะต้องใช้ผลสอบจากเด็กมาเป็นหลักในการวิเคราะห์ ความมุ่งหมายขั้นแรกของการวิเคราะห์ข้อสอบในวงการศึกษาก็เพื่อจะทราบคุณภาพของข้อคำถามอยู่ 2 ประการ คือ¹

1. ระดับความยาก (Item Difficulty) ของข้อสอบ หมายถึง สัดส่วนหรือเปอร์เซ็นต์ของผู้สอบ ที่ตอบคำถามข้อนั้นถูกต้อง การประมาณค่าความยากโดยอาศัยการคำนวณจากกลุ่มสูงและกลุ่มต่ำเท่านั้นไม่รวมพวกตรงกลาง ทำให้เกิดความคลาดเคลื่อนเนื่องจากการสุ่มตัวอย่าง (Sampling Error) และค่าไม่เท่ากับที่จะคำนวณโดยใช้ประชากรทั้งหมด อีเบล² (Ebel) ได้แนะนำ ถ้าแบบสอบมีค่าอำนาจจำแนกต่ำ ควรมีการคิดคะแนนแก้การเดาคำตอบ (Correct for Chance Success)

2. อำนาจจำแนก (Power of Discrimination) หมายถึงความสามารถของคำถามที่จะแบ่งเด็กออกเป็น 2 กลุ่ม คือกลุ่มที่เก่งมีความรู้ดี กับกลุ่มค้อที่มีความรู้-

¹ชวาล แพทย์กุล, เทคนิคการวัดผล (พระนคร : บริษัทสำนักพิมพ์วัฒนาพานิช จำกัด, 2518), หน้า 292 - 299.

²Robert L. Ebel, Measuring Educational Achievement (New Jersey : Prentice - Hall Inc., 1965), pp. 365 - 366.

อน การคำนวณหาค่าอำนาจจำแนกมีหลายวิธี เช่นวิธีของ จอห์นสัน (Johnson) ซึ่งคำนวณโดยใช้ผลต่างของผู้ทดสอบแต่ละข้อถูกของกลุ่มสูงกลุ่มต่ำ หาค่ายจำนวนผู้ทดสอบทั้งหมดในกลุ่มสูงหรือกลุ่มต่ำ นอกจากนี้ อำนาจจำแนกยังใช้เป็นตัวชี้บอกความตรงของข้อระหว่ง ในทางปฏิบัติเราหาความตรงในรูปของ สัมประสิทธิ์สหสัมพันธ์ระหว่างข้อระหว่งกับคะแนนรวมของแบบสอบ (Item - Total Score Correlation) ซึ่งถือเป็นเกณฑ์ภายใน (Internal Criterion)³ สูตรที่ใช้ในการคำนวณ มีด้วยกันหลายสูตร แต่ละสูตรมีข้อตกลงเบื้องต้นในการใช้ต่างกัน ที่นิยมคือ สูตรของเพียร์สัน (Pearson Product Moment Correlation) สูตรสหสัมพันธ์แบบสองค่าคัม (Biserial Correlation) และสูตรสหสัมพันธ์แบบจุดสองค่าคัม (Point Biserial Correlation)

1. สัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน เป็นค่าความสัมพันธ์ระหว่างข้อมูล 2 ชุด มีข้อตกลงเบื้องต้นว่า⁴

1.1 ข้อมูลของตัวออกมาเป็นคะแนนซึ่งเป็นตัว เลขที่บอกปริมาณ (Interval Scale) และลักษณะต่อเนื่องกัน (Continuous)

1.2 ข้อมูลทั้ง 2 ชุด ต้องมีการแจกแจงเป็นโค้งปกติ (Normal Distribution)

1.3 ข้อมูลทั้ง 2 ชุด จะต้องมีความสัมพันธ์กันเป็นเส้นตรง (Linearity) หรือแนวโน้มที่จะเป็นเส้นตรง

1.4 ถ้าข้อมูลมีน้อยกว่า 30 จะต้องอ้างอิงไปถึงการแจกแจงของคะแนนทั้ง 2 ชุด

³Ibid., p. 352.

⁴J.P. Guilford, Fundamental Statistics in Psychology and Education (5th ed.; Tokyo : McGraw - Hill Kogakusha, 1973), pp. 94 - 95.

2. สัมประสิทธิ์สหสัมพันธ์แบบสองลำดับ (Biserial Correlation) เป็นค่าความสัมพันธ์ระหว่างข้อมูล 2 ชุด มีข้อตกลงเบื้องต้นว่า

2.1 ข้อมูลชุดหนึ่งเป็นตัวแปรที่มีลักษณะต่อเนื่องกัน (Continuous Variable) และอีกชุดหนึ่งเป็นตัวแปรที่มีลักษณะแบ่งออกเป็น 2 รายการ (Dichotomous Variable)

2.2 ข้อมูลชุดที่แบ่งออกเป็น 2 รายการนั้นจะต้องมีลักษณะการแจกแจงเป็นโค้งปกติ (Normal Distribution) ส่วนอีกชุดที่มีลักษณะต่อเนื่องนั้น จะมีการแจกแจงเป็นโค้งปกติหรือไม่ก็ตาม จะต้องมียอดโค้งอันเดียว (Unimodal) และมีลักษณะสมมาตร (Symmetry)

2.3 ความสัมพันธ์ของข้อมูลทั้ง 2 ชุด จะต้องเป็นเส้นตรง (Linearity)

2.4 ตัวอย่างประชากร (N) จะต้องมีจำนวนมากพอ

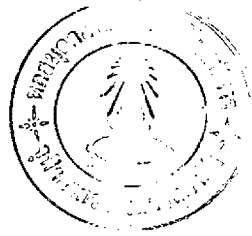
3. สัมประสิทธิ์สหสัมพันธ์แบบจุดสองลำดับ (Point Biserial Correlation) เป็นความสัมพันธ์ระหว่างข้อมูล 2 ชุด มีข้อตกลงเบื้องต้นว่า

3.1 ข้อมูลชุดหนึ่งเป็นตัวแปรที่มีลักษณะต่อเนื่อง (Continuous Variable) และอีกชุดหนึ่งมีลักษณะแบ่งออกเป็น 2 รายการ อย่างแท้จริง (Genuine or Distinct Dichotomous Variable)

3.2 ข้อมูลที่แบ่งเป็น 2 รายการนั้นมีลักษณะการแจกแจงเป็นโค้งปกติ (Normal Distribution) หรือการแจกแจงของคะแนนไม่เป็นแบบปกติ แต่พิสูจน์ได้ว่ามาจากประชากรที่มีการแจกแจงแบบปกติ

3.3 ความสัมพันธ์ของข้อมูลทั้งสองจะต้องเป็นเส้นตรง (Linearity) หรือมีแนวโน้มที่จะเป็นเส้นตรง

ค่าสหสัมพันธ์แบบสองลำดับจะมีค่าใกล้เคียงกับสหสัมพันธ์แบบเพียร์สัน แต่จะให้ค่าความคลาดเคลื่อนมากกว่า โดยเฉพาะเมื่อระดับความยาก (P) เข้าใกล้ 1 หรือ 0 ส่วนค่าสหสัมพันธ์แบบจุดสองลำดับจะให้ค่าสหสัมพันธ์ต่ำกว่าแบบสองลำดับ เมื่อคำนวณจาก



ข้อมูลชุดเดียวกัน และในขณะเดียวกันจะให้ค่าประมาณใกล้เคียงกับสหสัมพันธ์แบบเพียร์สัน⁵

ในการวิเคราะห์รายข้อ ความหนักแน่นจะต้องวิเคราะห์จากคะแนนคำตอบของผู้เข้าสอบทุกคน แต่ในทางปฏิบัติ เราอาจทำการวิเคราะห์โดยใช้ผลการสอบของคำตอบเพียงบางส่วนที่ปลายสุด คือใช้เทคนิคแบ่งกลุ่มสูง กลุ่มต่ำที่จะได้ค่าประมาณใกล้เคียงกับการที่จะนำข้อมูลทั้งหมดมาวิเคราะห์ อีเบล⁶ (Ebel) ได้แสดงถึงเทคนิคที่นำมาใช้ในการแบ่งกลุ่มสูง และกลุ่มต่ำมีหลายวิธี เช่น เทคนิค 25%, 27%, 33% ถ้าการแจกแจงของคะแนนเป็นปกติ (Normal Distribution) เคลลี (Kelly, 1939) ได้แสดงว่าหากใช้กลุ่มสูงและกลุ่มต่ำ กลุ่มละ 27% มาทำการคำนวณก็จะได้ค่าประมาณใกล้เคียงกับที่จะทำการวิเคราะห์ทั้งหมด และกลุ่มตัวอย่างประชากรจะต้องมากพอ เคอตัน⁷ (Cureton, 1957) ได้แสดงให้เห็นว่าถ้าการแจกแจงของคะแนนไม่เป็นแบบการแจกแจงปกติจะต้องใช้เทคนิค 33% จึงจะได้ค่าใกล้เคียง และ อัดัมส์⁸ (Adams) ชี้ให้เห็นว่า ถ้ากลุ่มตัวอย่างมีจำนวนน้อย เช่น 50 คน หรือน้อยกว่า จะต้องใช้การแบ่งครึ่งกลุ่มตัวอย่างประชากร (เทคนิค 50%) จึงจะทำให้ผลการวิเคราะห์มีความเชื่อมั่นได้สูง กรณีในการพิจารณาข้อกระทง ซึ่งดูจากระดับความยากและอำนาจจำแนกของ

⁵ Ibid., 296 - 300.

⁶ Ebel, op. cit., 349.

⁷ Anne Anastasi, Psychological Testing (2d ed.; New York: The Mac Millan Company, 1968), p. 433.

⁸ Georgia Sachs Adams, Measurement and Evaluation in Education, Psychological and Guidance (New York : Holt, Rinehart and Winston, 1966), p. 355.

ข้อสอบแต่ละข้อนั้น นั้นแนลลี⁹ (Nunnally) กำหนดว่า ข้อกระทงที่ดีควรมีระดับความยากอยู่ระหว่าง 20% ถึง 80% ส่วนอีเบล¹⁰ (Ebel) ได้กำหนดค่าอำนาจจำแนกของข้อกระทงในแบบสอบผลสัมฤทธิ์ทางการเรียนว่า ควรมีค่าตั้งแต่ .30 ขึ้นไป และ ดร.ชวาล แพทย์กุล¹¹ กำหนดว่าข้อกระทงที่ดีควรมีอำนาจจำแนก .20 ขึ้นไป

สำหรับเกณฑ์ในการพิจารณาตัวเลือกของข้อกระทง (Alternative Response) ตัวเลือกที่เป็นตัวถูกพิจารณาจากระดับความยากระหว่าง 20% ถึง 80% และค่าอำนาจจำแนก .20 ขึ้นไป ส่วนตัวเลือกที่เป็นตัวลวงพิจารณาจากระดับความยาก 5% ขึ้นไป และจำนวนผู้ตอบในกลุ่มต่ำมีมากกว่ากลุ่มสูง หรืออาจพิจารณาจากเปอร์เซ็นต์ของตัวเลือกตัวลวงนั้น ๆ จะต้องน้อยกว่าเกณฑ์ คือครึ่งหนึ่งของค่าเฉลี่ยของเปอร์เซ็นต์รวมผู้เลือกตัวลวงทุกตัว ส่วนตัวเลือกที่จะต้องมีการปรับปรุงแก้ไข ได้แก่ ตัวเลือกที่มีคนเลือกน้อย หรือไม่มีคนเลือกเลย ตัวเลือกที่มีคนเลือกตอบทั้งหมดหรือเกือบทั้งหมด รวมทั้งตัวเลือกชนิดตัวลวงที่มีคนในกลุ่มสูง เลือกตอบเป็นจำนวนมากกว่าคนในกลุ่มต่ำ และตัวเลือกชนิดตัวถูก ที่มีคนในกลุ่มต่ำเลือกตอบเป็นจำนวนมากกว่าคนในกลุ่มสูง¹²

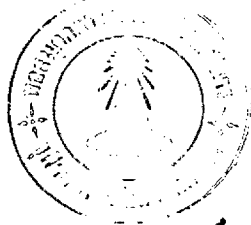
ในการปรับปรุงแบบสอบโดยทั่ว ๆ ไป หลังการวิเคราะห์แล้ว เมื่อจะรวบรวมข้อคำถามที่ดีให้เป็นข้อสอบทั้งฉบับ ถ้าเป็นข้อสอบสำหรับวัดผลสัมฤทธิ์ในการศึกษา (Achievement Test) มีหลักว่า จะต้องเลือกข้อสมมติให้มีระดับความยากเป็นสัดส่วนดังนี้คือ ง่าย 1 ส่วน ปานกลาง 2 ส่วน และยาก 1 ส่วน อยู่ในช่วง 20% ถึง 80%

⁹Jum C. Nunnally, Educational Measurement and Evaluation (New York : McGraw - Hill Book Company, 1959), p. 134.

¹⁰Ebel, op. cit., p. 373.

¹¹ชวาล แพทย์กุล, เรื่องเดิม, หน้า 314.

¹²Nunnally, op. cit., pp. 134 - 135.



แต่ถ้าเป็นข้อสอบประเภทแข่งขัน (Selection Test) ก็เลือกเอาแต่ข้อยาก ๆ ตั้งแต่ระดับความยาก 50% ลงมาถึง 5% โดยแบ่งเป็นสัดส่วน 1 : 2 : 1 เช่นเดียวกัน หรืออาจจะตั้งแต่ 25% หรือ 30% ลงมาถึงได้ถ้าในการแข่งขันนั้นต้องการคัดไว้เพียงจำนวนน้อย¹³

หลังการวิเคราะห์ข้อกระทงแล้ว จะมีข้อกระทงบางส่วนที่อยู่ในเกณฑ์เก็บไว้ใช้ได้ต่อไป ข้อที่ไม่อยู่ในเกณฑ์จะได้รับการปรับปรุงแก้ไข ตลอดจนสร้างข้อกระทงขึ้นทดแทนข้อกระทงที่ถูกคัดทิ้งไป ซึ่งจะ เป็นผลให้แบบสอบมีคุณภาพดีขึ้นในบางครั้ง การวิเคราะห์ข้อกระทงจะทำให้แบบสอบสั้นลง ซึ่งแบบสอบชุดใหม่นี้อาจมีความตรงและความเที่ยงมากกว่าแบบสอบชุดเดิม¹⁴

ความเที่ยง (Reliability) ของแบบสอบ

หมายถึง ความคงที่ ความสอดคล้อง ความถูกต้องของแบบสอบหรืออาจหมายถึง ผลลัพธ์คงที่ไม่เปลี่ยนแปลงของเครื่องมือวัด ถ้าหากนำเอาเครื่องมือนี้ไปทดลองอีกครั้งหนึ่ง

ข้อตกลงเบื้องต้น (Assumption) ของความเที่ยง คือการวัดความคลาดเคลื่อนที่เกิดขึ้นด้วยวิธีสุ่มตัวอย่าง สัมประสิทธิ์ความเที่ยงแสดงสัดส่วนของความแปรปรวนที่เกิดขึ้นจริงของคะแนนสอบ

เทคนิคใหญ่ ๆ ในการหาความเที่ยงซึ่งเป็นที่นิยมกัน มี 5 วิธี คือ

1. การสอบซ้ำ (Test - Retest) หาสัมประสิทธิ์แห่งความคงที่ (Coefficient of Stability) คือการหาค่าสัมประสิทธิ์สหสัมพันธ์ของการสอบ 2 ครั้ง

¹³ ชาวล แพร์ทกุล, เรื่องเดิม, หน้า 317 - 318.

¹⁴ Anastasi, *op. cit.*, p. 169.

โดยใช้แบบสอบชุดเดิม และผู้สอบกลุ่มเดิมด้วยสูตรของเพียร์สัน¹⁵

ข้อบกพร่องที่เกิดขึ้นจากการใช้เทคนิคนี้ คือ ค่าคอมครั้งที่สองไม่เป็นอิสระจากครั้งแรก เนื่องจากการจำข้อสอบได้ของนักเรียน การวิพากษ์วิจารณ์ค่าคอมในกลุ่มนักเรียนด้วยกันทำให้เกิดการเรียนรู้เป็นต้น และถ้าช่วงเวลาระหว่างการสอบทั้งสองครั้งห่างเกินไป ความคลาดเคลื่อนในการวัดอาจเป็นผลมาจากการเรียนรู้ไม่ใช่ความสามารถจริง ๆ ของนักเรียนเอง¹⁶ ด้วยปัญหาดังกล่าวการพิจารณาใช้เทคนิคนี้ จึงต้องมีข้อตกลงเบื้องต้น¹⁷ คือคุณสมบัติที่วัดด้วยแบบสอบนั้นคงที่ตลอดเวลา การปฏิบัติทั้งสองครั้งไม่แตกต่างกัน และการเรียนรู้ที่เกิดขึ้นระหว่างวิธีดำเนินการสอบสองครั้งไม่แตกต่างกัน โดยทั่วไปการใช้เทคนิคนี้เหมาะกับการวัดลักษณะทางกายภาพ (Physical Characteristic) และพฤติกรรมบางอย่าง เช่น ความเร็วในการตอบสนอง ความแข็งแรงของกล้ามเนื้อ¹⁸ เป็นต้น

2. ใช้ฟอร์มขนาน (Parallel Test Form or Equivalent Form)
วิธีหาความเที่ยงที่นิยมกันแบบหนึ่งก็คือ การใช้แบบสอบขนาน ซึ่งมีลักษณะที่สำคัญ คือ แต่ละฉบับสร้างไว้ด้วยวัตถุประสงค์อย่างเดียวกัน วัดเนื้อหาวิชาเดียวกัน และมีความ

¹⁵Frederick G. Brown, Principals of Educational and Psychological Testing (Illinois : The Dryden Press Inc., 1970), p. 60.

¹⁶Robert L. Ebel, Essentials of Educational Measurement (Englewood Cliffs, New Jersey : Prentice - Hall Inc., 1972), p. 412.

¹⁷Brown, op. cit., p. 66.

¹⁸Robert L. Thorndike and Elizabeth Hagen, Measurement and Evaluation (2d ed.; New York : John Wiley & Sons, Inc., 1965), p. 177.

ยากง่ายเท่ากัน และชนิดของข้อกระทงเหมือนกัน จะแตกต่างกันเฉพาะลักษณะของคำถาม และจะหาค่าความเที่ยงได้โดยนำแบบสอบทั้งสองฟอร์ม ไปสอบกลุ่มตัวอย่างกลุ่มเดียวกัน สองครั้ง ช่วงเวลาระหว่างการสอบอาจสั้นหรือยาวขึ้นอยู่กับว่าต้องการดูความคงที่ใน เวลาที่เปลี่ยนไปหรือไม่ แต่โดยปกติจะเว้นช่วงเวลาไม่นานนัก¹⁹ ค่าสหสัมพันธ์ที่ได้ เรียกว่าสัมประสิทธิ์แห่งความสมนัย (Coefficient of Equivalent) และหาได้ ด้วยสูตรของ เพียร์สัน²⁰ แต่โดยมากจะหาแบบสอบที่มีคุณสมบัติเท่าเทียมกันดังกล่าวมาแล้วได้ยาก หรือไม่มีเวลาพอที่จะทำการสอบครั้งที่สอง นอกจากนี้ยังมีปัญหาเรื่องการหา ข้อสอบได้อีกด้วย ข้อตกลงเบื้องต้นก็คือ แบบสอบจะต้องสมนัยกันในเนื้อหา แบบฟอร์ม ความยาวและความยาก อย่างไรก็ตามก็ผู้สร้างแบบสอบทั่วไปก็ยอมรับว่าการหาความเที่ยง โดยวิธีนี้มีความเหมาะสมและดีกว่าวิธีสอบซ้ำ²¹

3. การแบ่งครึ่งแบบสอบ (Split - Half Reliability) ในกรณีนี้ แบบสอบชุดเดียวกันถูกแบ่งออกเป็น 2 ส่วน แต่ละส่วนจะมีความเท่าเทียมกันทั้งในระดับ ความยากและเนื้อหา วิธีแบ่งแบบง่าย ๆ ก็คือ แบ่งเป็นข้อคู่และข้อคี่ เพื่อนำแบบสอบ ไปสอบแล้ว ผู้เข้าสอบแต่ละคนจะมีคะแนนเป็น 2 จุด หาค่าความเที่ยงของครึ่งแบบสอบได้ โดยหาสหสัมพันธ์ระหว่างข้อคู่และข้อคี่ แล้วนำค่าความเที่ยงนั้นมาหาค่าความเที่ยงของ แบบสอบฉบับเต็มด้วยสูตรของ สเปียร์แมน บราวน์²² (Spearman Brown) หรืออาจ ใช้วิธีย่อของ รูลอน (Rulon) โดยนำสัดส่วนของความแปรปรวนของผลทางระหว่าง

¹⁹Ibid., p. 178.

²⁰Brown, loc. cit.

²¹Jum C. Nunnally, Jr., Test and Measurement (New York : McGraw - Hill Book Company, Inc., 1959), p. 109.

²²Thorndike, op. cit., pp. 179 - 180.

คะแนนทั้ง 2 ชุด (The Variance of the difference between the half test score) กับความแปรปรวนของผลรวมระหว่างคะแนนทั้ง 2 ชุด (The Variance of the sum of the half test score) แล้วหักออกจากจำนวนเต็มหนึ่ง

ข้อตกลงเบื้องต้น แบบสอบนั้นจะต้องมีความยาวมากพอ มีความเป็นเอกพันธ์ (Homogeneity) เมื่อแบ่งแล้วแต่ละครั้งมีคุณสมบัติเท่าฉบับเต็ม และมีความเท่าเทียมกัน และให้นำหนักคะแนนของทุกข้อแบบเดียวกัน (simple unit - weight right only scoring)²³ เทคนิคนี้จะไม่ใช่ในกรณีที่เป็นแบบสอบใช้ความเร็ว (Speed Test) เพราะการจำกัดเวลาในการสอบมีผลทำให้ค่าประมาณความเที่ยงสูงไปกว่า เพราะคะแนนรวมจากทั้งข้อถูกและข้อผิดของแต่ละคนจะคล้ายกัน ไม่ว่าเขาจะทำข้อสอบถูกต้อง เกือบทั้งฉบับหรือผิดเกือบทั้งหมด และด้วยเหตุที่เป็นการสอบครั้งเดียว ฉะนั้นจึงไม่มีความคลาดเคลื่อนเนื่องจากการเว้นช่วงเวลาในการสอบ (error due to instability over time) ค่าประมาณความเที่ยงจึงค่อนข้างสูง²⁴

4. ความสอดคล้องภายใน (Internal Consistency) คูเดอร์ และ ริชาร์ดสัน ได้คิดวิธีหาความเที่ยงชนิดนี้ขึ้น เป็นการหาความเที่ยงของแบบสอบด้วยการสอบเพียงครั้งเดียว ข้อกระทงทุกข้อในแบบสอบจะต้องมีความเป็นเอกพันธ์ (Homogeneity) กล่าวคือ ข้อกระทงทุก ๆ ข้อในแบบสอบจะวัดลักษณะ ความสามารถหรือบุคลิกภาพอย่างเดียวกัน²⁵ ดังนั้นเทคนิคนี้จึงต้องมีข้อตกลงเบื้องต้นว่า ลักษณะของแบบสอบจะต้องเป็นเอกพันธ์ (Homogeneity) การให้คะแนนของแบบสอบจะต้องเป็นแบบ ถ้าถูกให้ 1 คะแนน และผิดให้ 0 คะแนน สูตรเบื้องต้นก็คือ สูตรคูเดอร์ริชาร์ดสัน

²³Ebel, op. cit., pp. 413 - 414.

²⁴Nunnally, loc. cit.

²⁵Ebel, op. cit., p. 415.

สูตรที่ 20 (Kuder - Richardson, Formula 20) ถ้าระดับความยาก (Difficulty) ของข้อสอบแต่ละข้อในแบบสอบไม่แตกต่างกันมาก อาจจะใช้สูตรคูเดอร์ ริชาร์ดสัน สูตรที่ 21 (Kuder - Richardson, Formula 21) คำนวณไคค่าใกล้เคียงกัน แต่อาจอกระทงในแบบสอบมีระดับความยากแตกต่างกัน การใช้สูตรคูเดอร์ ริชาร์ดสัน สูตรที่ 21 จะให้ค่าประมาณความเที่ยงต่ำกว่าที่เป็นจริง (Underestimate Reliability Coefficient)

และเนื่องจากการคำนวณโดยใช้สูตรคูเดอร์ ริชาร์ดสัน สูตรที่ 20 และวิธีแบ่งครึ่งแบบสอบมีข้อจำกัดในการใช้แบบเดียวกันคือ เป็นการคำนวณจากการสอบครั้งเดียว จึงไม่มีความแปรปรวนเนื่องจากเวลา และอีกประการหนึ่งคือ จะไม่ใช้วิธีนี้กับข้อสอบประเภทใช้ความเร็ว (Speed Test) ดังนั้น สัมประสิทธิ์ความเที่ยงที่คำนวณจาก 2 วิธีดังกล่าว จึงมีค่าใกล้เคียงกัน²⁶

5. การหาค่าความเที่ยงโดยการใชสูตรของ ฮอยท์²⁷ (Hoyt's Analysis of Variance) เป็นการประมาณค่าความเที่ยงของแบบสอบ โดยวิธีวิเคราะห์ความแปรปรวน มีข้อตกลงเบื้องต้นว่า

5.1 สมมุติว่าคะแนนของแต่ละคนแบ่งออกเป็น 4 ส่วน ซึ่งไม่ได้ขึ้นต่อกัน

คือ

ส่วนที่เกี่ยวข้องกันแต่ละคน และแต่ละข้อ
 ส่วนที่สัมพันธ์กับข้อสอบ
 ส่วนที่สัมพันธ์กันแต่ละคน

002909

²⁶Thorndike, op. cit., p. 181.

²⁷Palmer O. Johnson, Statistical Method in Research Englewood Cliffs, N.J.: Prentice - Hall, Inc., 1961), pp. 134 - 136.



ส่วนที่เป็นความคลาดเคลื่อน และถือว่าความคลาดเคลื่อนของแต่ละ
ข้อมีการแจกแจงเป็นปกติ

5.2 สามารถใช้ได้กับแบบสอบที่ให้คะแนนแต่ละข้อเท่ากับ 1 หรือมากกว่า
1 คะแนน ก็ได้

ผลจากการคำนวณแบบ Hoyt's Analysis of Variance จะได้ผลเหมือนกับ
การคำนวณจากสูตรคูเคอร์ ริชาร์ดสัน สูตรที่ 20

ความคลาดเคลื่อนมาตรฐานในการวัด (Standard Error of Measure-
ment) หมายถึงความคงที่ของคะแนนเมื่อมีการวัดหลาย ๆ ครั้ง (The accuracy
or precision of an individual score)²⁸ หรืออาจกล่าวได้ว่า ความคลาด
เคลื่อนมาตรฐานในการวัด เป็นค่าประมาณของส่วนเบี่ยงเบนมาตรฐานของชุดของคะแนน
ของบุคคลหนึ่ง ๆ เมื่อมีการวัดซ้ำ ๆ กันเกิดขึ้นหลายครั้ง (an estimate of
standard deviation that would be obtained for a series of mes-
surement of the same individual) โดยมีสมมุติฐานว่าบุคคลจะไม่เปลี่ยนแปลง
เนื่องจากการวัด ค่าความคลาดเคลื่อนมาตรฐานหาได้จากสูตร ผลคูณระหว่างส่วนเบี่ยง-
เบนมาตรฐานของคะแนนกับรากที่สองของผลต่างระหว่างจำนวนเต็มหนึ่งกับค่าความเที่ยง
ของคะแนนชุดนั้น เช่น ถ้าคำนวณได้ 2.2 หมายความว่า ถ้ามีการวัดครั้งที่สอง คะแนน
สอบในครั้งหลังของบุคคลใด ๆ ในกลุ่มนั้น จะเบี่ยงเบนไปจากคะแนนสอบในครั้งที่ 1
ของเขา อยู่ในช่วง ± 2.2 และโอกาสเช่นนี้เกิดขึ้นเพียง 7 ใน 10 คุณสมบัติของ
ค่าความคลาดเคลื่อนมาตรฐานในการวัดไม่เพียงแต่ช่วยทำให้เข้าใจความหมายของความ
เที่ยงที่ขึ้น ยังทำให้ทราบความแปรปรวนของคะแนนสอบอีกด้วย²⁹

²⁸Henry E. Garrett, Testing for Teacher (New York :
American Book Company, 1959), pp. 29 - 30.

²⁹Thorndike, op. cit., pp. 182 - 183.

องค์ประกอบที่มีอิทธิพลต่อค่าความเที่ยง³⁰ (Factors Affecting Reliability)

1. ความยาวของแบบสอบ (Length of Test) ถ้าความยาวของแบบสอบเพิ่มขึ้น ค่าความเที่ยงจะสูงขึ้นด้วย เพราะจะเป็นการเพิ่มช่วงกว้างของคะแนน ถ้าคะแนนกระจายมากทำให้ความเที่ยงสูง
2. พิสัยความแตกต่างระหว่างบุคคล เช่น สัมฤทธิ์ผลในการเรียน ถ้าความแปรปรวนภายในในกลุ่มมีมาก ค่าความเที่ยงจะสูง ถ้าความแปรปรวนในกลุ่มมีน้อย ค่าความเที่ยงจะต่ำ
3. สภาพของการบริหารการสอบและการให้คะแนน มีผลทำให้ค่าความเที่ยงสูงขึ้นหรือต่ำลง เช่น ความเร็วในการทำงาน สิ่งเร้า และสภาพอารมณ์ในขณะทำการสอบ การทุจริตในการสอบ และอุบัติเหตุในขณะทำการสอบ เป็นต้น
4. โครงสร้างของแบบสอบ (The Construction of the Test) เช่น รูปแบบของแบบสอบ (Form), ความยากของแบบสอบ (Test Difficulty) ความเป็นปรนัยของการตรวจให้คะแนน ข้อสอบที่มีตัวเลือก (Alternatives) 5 ตัว จะให้ค่าความเที่ยงสูงกว่าชนิด 3 ตัวเลือก รวมทั้งข้อคำถามที่ไม่เป็นอิสระต่อกัน (Inter-dependent) และรูปแบบของคำถาม (Item Wording) ต่างก็มีผลต่อความเที่ยงทั้งสิ้น
5. เทคนิคที่ใช้ในการคำนวณหาสัมประสิทธิ์ความเที่ยง มีผลทำให้ค่าความเที่ยงสูงหรือต่ำได้ เช่นแบบสอบใช้ความเร็ว (Speed Test) จะไม่ใช่เทคนิคความสอดคล้องภายใน (Internal Consistency) และสำหรับเทคนิคการใช้ฟอร์มคู่ขนาน (Parallel Forms) จะใช้ข้อเมื่อแบบสอบทั้งสองมีความเท่าเทียมกันจริง ๆ และใช้เวลา

³⁰HH. Remmer and N.L. Gage, Educational Measurement and Evaluation (New York : Horper & Brothers, 1955), pp. 136 - 138.

ในการสอบแยกจากกัน (Time Independently) เช่น การใช้เทคนิคนี้จะทำให้ค่าความเที่ยงของแบบสอบมาตรฐาน (Standardize Test) บางฉบับสูงกว่าที่เป็นจริง

ค่าความเที่ยงควรจะเป็นเท่าใดจึงจะเป็นที่พอใจนั้น เคลลี³¹ (Kelley) กล่าวว่า แบบสอบควรมีความเที่ยงสูงแค่ไหนนั้นแล้วก็จะนำไปใช้ประโยชน์ในแง่ใด เขาได้กำหนดจุดมุ่งหมายแต่ละประเภทไว้ดังนี้

1. เพื่อประเมินผล ระดับผลสัมฤทธิ์ของคนกลุ่มใดกลุ่มหนึ่ง ควรมีค่าความเที่ยงตั้งแต่ .50 ขึ้นไป
 2. เพื่อประเมินผลความแตกต่างของระดับผลสัมฤทธิ์ของกลุ่ม ในการกระทำ 2 อย่าง หรือมากกว่านั้น ควรมีค่าความเที่ยงตั้งแต่ .90 ขึ้นไป
 3. เพื่อประเมินผลสัมฤทธิ์ของคนเป็นรายบุคคล ค่าความเที่ยงตั้งแต่ .94 ขึ้นไป
 4. เพื่อประเมินผลความแตกต่างในระดับผลสัมฤทธิ์ของแต่ละบุคคล ในการกระทำ 2 อย่าง หรือมากกว่านั้น ค่าความเที่ยงตั้งแต่ .98 ขึ้นไป
- อย่างไรก็ดี ค่าความเที่ยงตามที่กำหนดตามรายการข้างต้นอาจเปลี่ยนแปลงได้

ความทรง (Validity) ของแบบสอบ

ความทรงนี้มีนักวัดหลายท่านได้ให้ความหมายไว้เช่น

ลินควิส (Linguist) ให้ความหมายไว้ว่า ความทรงคือความแน่นอนของแบบสอบที่สามารถวัดได้ตรงกับวัตถุประสงค์ของการวัด ซึ่งจะต้องเปรียบเทียบระหว่างสิ่งที่

³¹E.F. Linquist, Educational Measurement (Washington D.C.: American Council on Educational 1951), p. 609.

แบบสอบวัดออกมาถึงสิ่งที่ผู้วัดต้องการว่าตรงกันหรือไม่ เพียงใด

เคอร์ตัน (Curton) และ กิลฟอร์ด (Guilford) ได้จำกัดความหมายในวงแคบว่า ความตรงของแบบสอบ คือค่าสหสัมพันธ์ระหว่างคะแนนจากแบบสอบกับคะแนนซึ่งเป็นเกณฑ์ที่ยอมรับว่าเชื่อถือได้

เอกเจตัน (Edjerton) กล่าวว่า "เมื่อเราพูดถึงความตรง เราพิจารณาประโยชน์ของเครื่องมือวัดว่า มีข้อเป้าหมายที่คงเอาไว้เพียงใด"

ครอนบาค (Cronbach) กล่าวว่า ยิ่งแบบสอบมีความสมบูรณ์ เชื่อถือได้ในการวัดความคะแนนมากเพียงใด แบบสอบย่อมมีความตรงตามทฤษฎีมากขึ้นเพียงนั้น

และ อีเบล (Ebel) อ้างถึง Webster's New Collegiate Dictionary ได้ให้ความหมายไว้ว่า แบบสอบที่มีความตรงจะต้องสร้างขึ้นบนรากฐานของความเป็นจริงที่สามารถที่จะพิจารณาโต้แย้ง หรือสนับสนุนได้อย่างมีเหตุผล³²

เดวิส³³ (Davis) ได้สรุปไว้ว่า แบบสอบแต่ละฉบับมีจุดประสงค์ต่างกัน จึงมีความตรงหลายแบบ แต่ละแบบจะขึ้นอยู่กับเกณฑ์ที่นำมาเปรียบเทียบ ซึ่งแตกต่างกันออกไป

วิธีการหาค่าความตรง คือการคำนวณค่าสหสัมพันธ์ระหว่างคะแนนของแบบสอบกับเกณฑ์ซึ่งมีค่าอยู่ระหว่าง + 1 กับ - 1 ค่า + 1 บอกให้ทราบถึงความสัมพันธ์อย่างสมบูรณ์ของคะแนนกับเกณฑ์ เช่นตำแหน่ง (Rank Order) ของคนหนึ่ง ๆ จะเหมือนกันในคะแนนแบบสอบและเกณฑ์ ถ้าค่าเป็น 0 แสดงว่าไม่มีความสัมพันธ์กันเลย

³²Ebel, op. cit., pp. 296 - 298.

³³Frederick B. Davis, Educational Measurement and their Interpretation (Belmont, California : Wadsworth Publishing Company, Inc., 1964), pp. 24 - 39.

และถ้าเป็นลมแสดงว่ามีความสัมพันธ์ในทางกลับกัน เช่น ใต้คะแนนต่ำในระบบสอบ แต่จะได้คะแนนสูงในชุดที่เป็นเกณฑ์ เป็นต้น

ความทรงมี 4 ชนิด แบ่งโดยผู้เชี่ยวชาญด้านวัดผลการศึกษาและพฤติกรรมศาสตร์ และจัดพิมพ์โดย สถาบัน Technical Recommendation for Achievement Test ได้แก่

1. ความตรงตามทำนาย (Predictive Validity) หาได้โดยการนำคะแนนที่สอบได้จากแบบสอบนี้ไปหาค่าสหสัมพันธ์กับ เกณฑ์ซึ่งอาจเป็นคะแนนสอบจากการสอบใน ภายหน้าหรืออาจเป็นผลงานที่ได้ทำมาในระยะ เวลาหนึ่งหลังจากทดสอบ ข้อสังเกตที่นำมา หาค่าความตรงแบบนี้ มักเป็นประเภทแบบสอบวัดความฉับไว แบบสอบเข้ามหาวิทยาลัย แบบ สอบวัดบุคลิกภาพ ความตรงแบบนี้ขึ้นอยู่กับผลจากการสอบใช้พยากรณ์การกระทำในอนาคต ได้แม่นยำเพียงใด ถ้าสามารถพยากรณ์การกระทำในอนาคตได้มาก ก็ถือว่ามีความตรง ตามทำนายสูงนั่นเอง

2. ความตรงตามสภาพ (Concurrent Validity) มักเป็นแบบสอบวัดผล สัมฤทธิ์แบบสอบที่มีความตรงตามสภาพปัจจุบัน หมายถึงแบบสอบนั้นต้องวัดตรงกับสภาพ เป็นจริงในปัจจุบันของคน ๆ นั้น วิธีหาค่าความตรงหาได้โดยหาค่าสหสัมพันธ์ระหว่าง แบบสอบกับเกณฑ์ภายนอก เช่น ข้อสอบวิชาเดียวกันที่ถือว่าเป็นมาตรฐาน หรือเปรียบเทียบ กับคะแนนเฉลี่ยสะสมจากการ เรียนที่ได้จนถึงเวลาสอบแบบสอบนั้น และเวลาที่หาค่า สัมพันธ์ต้องใกล้เคียงกับเวลาที่สอบ

ความตรงทั้ง 2 แบบนี้ รวมเรียกชื่อว่าการตรงตามเกณฑ์เทียบ (Cri- terion - Related Validity) ซึ่งมักจะหาอยู่ในรูปสัมประสิทธิ์สหสัมพันธ์ ต่างกัน ที่ระยะเวลาที่เว้นช่วงระหว่างแบบสอบทั้งสองชุดที่นำมาหาค่าสหสัมพันธ์เท่านั้น ถ้าระยะ เวลายาวนานเกินไป จะทำให้ความแม่นยำในการทำนายลดลง³⁴

³⁴Anastasi, op. cit., pp. 138 - 144.

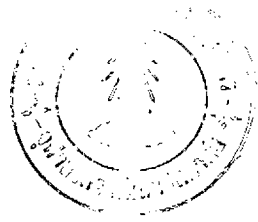
3. ความตรงตามเนื้อเรื่อง (Content Validity) หมายถึงความสามารถของข้อสอบที่จะวัดคุณแทนที่ดีของเนื้อหาทั้งหมด ตัวแทนของพฤติกรรมที่ต้องการวัด และวัดตัวแทนของจุดมุ่งหมายที่ต้องการสอน ดังนั้นเนื้อหาของข้อสอบและ เนื้อหาที่สอบจะต้องมีความสัมพันธ์กัน ความตรงตามเนื้อเรื่องมีความสำคัญยิ่งในการวัดผลการศึกษา ข้อสอบสัมฤทธิ์ผลทางการศึกษาถ้าขาดความตรงตามเนื้อหา ก็ไม่อาจนำมาใช้ในการสอบได้ การสร้างแบบสอบให้มีความตรงตาม เนื้อหา นั้น จะต้อง สุ่มตัวอย่าง ข้อสอบที่มีปริมาณมากพอ ซึ่งต้องอาศัยการ คัดเลือก โจทย์อย่างมีเหตุผลของครูผู้สอนหรือผู้ออกข้อสอบด้วย มีผู้เสนอแนะ การปรับปรุงความตรงตามเนื้อหาของแบบสอบว่า ควร จะสร้างตารางวิเคราะห์หลักสูตร (Table of Specification) ซึ่งจะช่วยกำหนดน้ำหนัก เนื้อหาวิชาแต่ละตอนและ น้ำหนักของจุดมุ่งหมายที่จะวัด³⁵

4. ความตรงตามทฤษฎี (Construct Validity) ความตรงตามทฤษฎีนี้ เราใช้ผลการสอบเพื่ออธิบายคุณสมบัติบางอย่างทางจิตวิทยา แบบสอบที่มีความตรงตามทฤษฎีนั้นหมายถึงแบบสอบนั้นสามารถวัดคุณลักษณะต่าง ๆ (Traits Characteristics) ในตัวบุคคล เช่น ความปัญญา (Intelligence), ความถนัด (Aptitude), ความสนใจ (Interest), บุคลิกภาพ (Personality), ความกังวล (Anxiety), อាកารทางประสาท (Neuroticism) ความเร็วในการเดิน (Speed of Walking) เป็นต้น³⁶.

การทำให้แบบสอบมีความตรงตามทฤษฎีนั้น จะต้องดูทฤษฎีที่อยู่เบื้องหลัง กรอบความคิดอธิบายว่า การดูว่าแบบสอบมีความตรงตามทฤษฎีหรือไม่นั้นมี 3 ตอน คือ ต้องดูว่าเบื้องต้นประกอบอะไรอยู่ในแบบสอบ แล้ววางสมมุติฐานจากทฤษฎีที่เกี่ยวข้องและทดสอบ

³⁵ Anastasi, op. cit., pp. 135 - 137.

³⁶ Ibid., pp. 114 - 116.



สมมุติฐานโดยการทดลองอย่างถูกวิธี³⁷ เช่น การทดสอบสมมุติฐานเกี่ยวกับทัศนคติทาง
 การศึกษาของคน 3 กลุ่ม คือกลุ่มที่จบมหาวิทยาลัยแล้ว กลุ่มที่กำลังศึกษา และกลุ่มที่อยู่
 นอกมหาวิทยาลัย ผู้วิจัยได้ใช้แบบสอบถามเป็นเครื่องมือชี้ให้เห็นความแตกต่างของทัศนคติ
 ทางการศึกษาของบุคคลทั้ง 3 กลุ่ม วิธีการเช่นนี้เรียกว่า วิธีการที่ใช้กลุ่ม -
 รู้ลักษณะ (Known Group Method) บางครั้งใช้วิธีหาความแตกต่างของค่า
 เฉลี่ย (Mean) ของกลุ่มตัวอย่างประชากร 2 กลุ่ม คือกลุ่ม เอ และ บี ถ้าค่าเฉลี่ย
 ของคะแนนทั้ง 2 กลุ่ม แตกต่างกันอย่างมีนัยสำคัญ แสดงว่าแบบสอบถามมีความตรงตาม
 ทฤษฎี³⁸ นอกจากนี้ สหสัมพันธ์ระหว่างแบบสอบถามนี้กับแบบสอบถามที่มีความตรงตามทฤษฎีอยู่
 แล้ว มักเป็นหลักฐานว่าแบบสอบถามนี้วัดพฤติกรรมคนเดียวกันนั้นได้โดยประมาณ หรือหา
 ความตรงด้วยวิธีความสอดคล้องภายใน คือหากาสหสัมพันธ์ระหว่างคะแนนแต่ละข้อ
 กับคะแนนรวมของแบบสอบถาม หรือหากำจากการคำนวณค่าสหสัมพันธ์ระหว่างคะแนนของ
 แบบสอบถามย่อยกับคะแนนรวม ซึ่งข้อกระทงแต่ละข้อหรือแบบสอบถามย่อยแต่ละชุดนั้นจะต้องวัด
 สิ่งเดียวกัน³⁹

³⁷Fred N. Kerlinger, Foundation of Behavioral Research
 (2d ed.; New York : Holt, Rinehart and Winston, Inc., 1964),
 pp. 461 - 462.

³⁸Ibid., pp. 467 - 468.

³⁹Anastasi, op. cit., pp.

องค์ประกอบที่มีอิทธิพลต่อค่าความตรง⁴⁰ (Factor Affecting Validity)
สรุปได้ดังนี้

1. องค์ประกอบในตัวแบบสอบเอง หมายถึง ความไม่สมบูรณ์ของแบบสอบ เช่น คำสั่งที่ไม่ชัดเจน คำสั่งหรือประโยคที่ซับซ้อน ซึ่งเป็นการวัดความเข้าใจในการอ่านหรือเขาวีมากกว่าสิ่งที่ต้องการวัด แบบสอบที่ยากเกินไปหรือง่ายเกินไป แบบสอบที่แนะนำคำตอบ และแบบสอบที่ไม่ได้วัดพฤติกรรมที่ต้องการวัด เช่น ข้อสอบคณิตศาสตร์ที่เขียนเป็นภาษาอังกฤษ เป็นต้น
2. องค์ประกอบจากกระบวนการสอน เช่น ครูบางคนสอนให้นักเรียนท่องจำแต่ออกข้อสอบให้คิด
3. องค์ประกอบจากผู้สอบเอง เช่น อารมณ์ของผู้สอบขณะทำการสอบ หรือความโน้มเอียงที่จะตอบอย่างใดอย่างหนึ่ง เช่น ข้อสอบแบบถูกผิด บางคนตอบแบบสอบไม่ได้ก็จะใส่เครื่องหมายถูกไว้ก่อน เป็นต้น

รายงานการวิจัยที่เกี่ยวข้อง

รายงานการวิจัยที่เกี่ยวข้องกับการวิเคราะห์ข้อสอบ ทั้งในประเทศไทยและต่างประเทศ มีเป็นจำนวนมาก ดังตัวอย่างที่ยกมาดังนี้

⁴⁰ เสริมศักดิ์ วิชาลาภรณ์ และ เอนก กรีสแสง, หลักเบื้องต้นของการวัดผลการศึกษา (พระนคร : โรงพิมพ์อักษรสัมพันธ์, 2518), หน้า 200 - 202.

ปี พ.ศ. 2506 พัทธา เกกานนท์⁴¹ ได้วิเคราะห์แบบสอบความพร้อม ซึ่งคัดแปลงมาจาก เมโทรโพลิแตน (Metropolitan Readiness Test - Verbal) และแบบสอบเขาว์ซึ่งคัดแปลงจากแบบสอบโอทิส (Otis Test) ที่ใช้ในการสอบคัดเลือกนักเรียนในชั้นประถมศึกษาปีที่ 1 ของโรงเรียนสาธิต จุฬาลงกรณ์มหาวิทยาลัย ปีการศึกษา 2506 ผลการวิเคราะห์พบว่า ข้อสอบความพร้อมในเรื่องความเข้าใจคำ 19 ข้อ มีระดับความยากระหว่าง 18% ถึง 95% และอำนาจจำแนกระหว่าง .10 ถึง .53 ในเรื่องความเข้าใจประโยค 14 ข้อ มีระดับความยาก 50% ถึง 92% อำนาจจำแนกระหว่าง .13 ถึง .50 ในเรื่องความเข้าใจการเปรียบเทียบ 19 ข้อ มีระดับความยากอยู่ระหว่าง 27% ถึง 82% อำนาจจำแนกระหว่าง .33 ถึง .59 ในเรื่องการนับจำนวน 24 ข้อ มีระดับความยากระหว่าง 37% ถึง 90% อำนาจจำแนกระหว่าง .17 ถึง .73 และในเรื่องความสามารถในการเลียนแบบ 10 ข้อ มีระดับความยากอยู่ระหว่าง .32 ถึง .59

ปี พ.ศ. 2507 ยุพิน ไชยวงศ์เกียรติ และเพื่อน⁴² วิเคราะห์ข้อสอบคัดเลือกเข้าศึกษาต่อระดับประกาศนียบัตรวิชาการศึกษา ปีการศึกษา 2506 ของกรมการฝึกหัดครู

⁴¹ พัทธา เกกานนท์, "ระดับความยากง่าย และระดับการ फैนจำแนกเป็นรายข้อของข้อสอบแบบความพร้อมซึ่งคัดแปลงมาจาก Metropolitan Readiness Test (Verbal) และแบบสอบเขาว์ของ Otis (ซึ่งคัดแปลงมาจาก Otis Test Form AS) และตำแหน่งเปอร์เซ็นต์ของคะแนนนักเรียนที่สอบเข้าศึกษาชั้นประถมศึกษาปีที่ 1 ของโรงเรียนสาธิต จุฬาลงกรณ์มหาวิทยาลัย ปีการศึกษา 2506" (วิทยานิพนธ์คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย, 2506)

⁴² ยุพิน ไชยวงศ์เกียรติ, สุมาลี สะอาดแก้ว, วันเพ็ญ พิศาลพงศ์ และมณีนิพนธ์ นครพนม, "การวิเคราะห์ข้อสอบคัดเลือกเข้าเรียนประกาศนียบัตรวิชาการศึกษาในสถาบันฝึกหัดครู จังหวัดพระนครและธนบุรี ปีการศึกษา 2506" (วิทยานิพนธ์คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย, 2507)

ซึ่งแบบสอบแยกเป็น 4 วิชา คือ ภาษาไทย ภาษาอังกฤษ คณิตศาสตร์ และวิทยาศาสตร์ พบว่าหมวดคณิตศาสตร์ ภาษาไทย ภาษาอังกฤษ และวิทยาศาสตร์ มีข้อกระทงที่ดี คือมีระดับความยากระหว่าง 20% ถึง 80% และอำนาจจำแนก .20 ขึ้นไป จำนวน 44, 70 และ 55 ข้อ ตามลำดับ ข้อที่มีอำนาจจำแนกต่ำกว่า .20 จำนวน 13, 34, 17 และ 34 ข้อ ตามลำดับ ข้อยากคือข้อที่มีระดับความยากมากกว่า 80% ขึ้นไป แต่มีอำนาจจำแนก .20 ขึ้นไปจำนวน 1, 13, 9 และ 1 ข้อ ตามลำดับ และข้อกระทงที่ยากคือข้อที่มีระดับความยากต่ำกว่า 20% แต่มีอำนาจจำแนก .20 ขึ้นไป จำนวน 2, 3, 2 และ 10 ข้อ ตามลำดับ สำหรับผลการวิเคราะห์ตัวเลือกพบว่า ส่วนใหญ่ยังไม่ดี คงแก้ไขและปรับปรุง

ปี พ.ศ. 2511 จีรพัทธ์ จันทรศรีวงศ์⁴³ ได้พัฒนาข้อสอบสมรรถภาพทางสมอง เพื่อใช้กับนักเรียนไทยชั้นประถมศึกษาตอนปลาย โดยอาศัยแนวข้อสอบมาตรฐานของ ฟินเนอร์ (Pintner General Ability Test of Mental Maturity (Long Form; Level 2 and 3) ผลการวิจัยพบว่า หลังจากได้ปรับปรุงแก้ไขแล้ว ข้อสอบฉบับสุดท้ายมีดัชนีความยากอยู่ระหว่าง 21% ถึง 91% อำนาจจำแนกระหว่าง .28 ถึง .83 ความเที่ยงของแบบสอบชั้นประถมศึกษาปีที่ 5, 6 และ 7 เป็น .661, .603 และ .753 ตามลำดับ

ปี พ.ศ. 2512 ศักดา บุญไวโรจน์⁴⁴ ได้ทำการวิเคราะห์ข้อสอบความถนัด

⁴³ จีรพัทธ์ จันทรศรีวงศ์, "การพัฒนาข้อสอบสมรรถภาพทางสมอง เพื่อใช้กับนักเรียนไทยชั้นประถมศึกษาตอนปลาย" (วิทยานิพนธ์ปริญญาโทมหาบัณฑิต แผนกวิชาวิจัยการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, 2511)

⁴⁴ ศักดา บุญไวโรจน์, "การวิเคราะห์ข้อสอบความถนัดเชิงกลขององค์การโทรศัพท์แห่งประเทศไทย" (วิทยานิพนธ์ปริญญาโทมหาบัณฑิต แผนกวิชาวิจัยการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, 2514)

เชิงกลขององค์กรโทรศัพท์แห่งประเทศไทย โดยใช้คะแนนจากผู้สมัคร สอบคัดเลือก เข้าทำงานในองค์กรโทรศัพท์แห่งประเทศไทย ปี 2513 จำนวน 629 คน ผลปรากฏว่าแบบสอบถามอันดับ ความแรงและเครื่องกล มีอำนาจจำแนกระหว่าง $-.16$ ถึง $.63$ ระดับความยากระหว่าง $.13$ ถึง $.90$ สัมประสิทธิ์ความเที่ยงเท่ากับ $.8803$ สัมประสิทธิ์ความตรงเท่ากับ $.7072$ แบบสอบถามอันดับความสัมพันธ์ด้านรูปร่าง มีอำนาจจำแนกระหว่าง $-.12$ ถึง $.70$ ระดับความยากระหว่าง $.29$ ถึง $.91$ สัมประสิทธิ์ความตรงเท่ากับ $.8869$ แบบสอบถามอันดับความสามารถด้านตัวเลข มีอำนาจจำแนกระหว่าง 0 ถึง $.66$ ระดับความยากระหว่าง 0 ถึง $.83$ สัมประสิทธิ์ความเที่ยงเท่ากับ $.8574$ และสัมประสิทธิ์ความตรงเท่ากับ $.5884$

ปี พ.ศ. 2517 ประสิทธิ์ สาริกัลยะ⁴⁵ วิเคราะห์แบบสอบวิชาภาษาอังกฤษ ฉบับ 16 ซึ่งอาจารย์คณะครุศาสตร์ เป็นผู้สร้างขึ้น เพื่อใช้สอบคัดเลือกนิสิตที่จบปริญญาตรีของกรม เข้าศึกษาต่อปริญญาโทคณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ผลการวิเคราะห์พบว่า คะแนนจากกลุ่มตัวอย่างมีลักษณะการ แจกแจง เป็นโค้งปกติ ค่าสัมประสิทธิ์สหสัมพันธ์ของแบบสอบตอนที่ 1 และตอนที่ 2 เท่ากับ $.88$ ค่ามัธยฐานของระดับความยากเท่ากับ 65.5 ค่ามัธยฐานของอำนาจจำแนกเท่ากับ $.42$ ข้อกระทงที่อยู่ในเกณฑ์ดี ก็คือมีระดับความยากระหว่าง 20% ถึง 80% และอำนาจจำแนก $.20$ ขึ้นไป มีจำนวนร้อยละ 72 ความเที่ยงชนิดความสอดคล้องภายในของแบบสอบตอนที่ 1, 2 และทั้งฉบับ เท่ากับ $.86, .94$ และ $.72$ ตามลำดับ จัดว่าอยู่ในระดับสูง ความตรงตามทำนายของแบบสอบ เท่ากับ $.11$ ซึ่งค่อนข้างต่ำ ความตรงตามทฤษฎีของแบบสอบพบว่า คะแนนเฉลี่ยของกลุ่มตัวอย่างที่เลือกการสอบภาษาอังกฤษเป็นวิชาเอก สูงกว่ากลุ่มที่เลือกวิชาอื่น ๆ เป็นวิชาเอกอย่างมีนัยสำคัญ

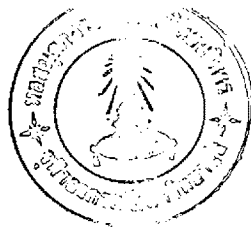
⁴⁵ ประสิทธิ์ สาริกัลยะ, "การวิเคราะห์แบบสอบภาษาอังกฤษ 16" (วิทยานิพนธ์ปริญญาโทมหาบัณฑิต แผนกวิชาวิจัยการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, 2517)

ปี พ.ศ. 2518 ฉวีวรรณ โปษีศรี⁴⁶ วิเคราะห์แบบสอบถามคัดเลือกของโรงเรียนเทคนิค ในโครงการ เงินกู้เพื่อพัฒนาอาชีวศึกษา ปี 2517 โดยใช้คะแนนของผู้เข้าสอบจำนวน 2,074 คน ผลการวิเคราะห์ปรากฏว่า ในแบบสอบถามทั้งหมด 5 ชุด คือ แบบสอบถามทัศนศาสตร์ ภาษาอังกฤษ ภาษาไทย วิทยาศาสตร์ และความถนัดทางช่าง มีข้อสอบที่มีระดับความยากและอ่านง่ายในเกณฑ์ดี เรียงตามลำดับ คือ ร้อยละ 46.58 40.35 และ 43 ตามลำดับ แบบสอบถามมีสัมประสิทธิ์ความเที่ยงสูงระหว่าง .9011 ถึง .9569 และแต่ละวิชาสามารถจำแนกบุคคลในกลุ่มที่ได้รับการทดสอบได้ สัมประสิทธิ์ความตรงเชิงพยากรณ์ เมื่อใช้แบบสอบถาม 5 วิชา ร่วมพยากรณ์ ตัวเกณฑ์มีค่าสูงสุดเป็น .6693 ที่ระดับความมีนัยสำคัญ .05

ในปีเดียวกัน มาโนช กลั่นฤทธิ⁴⁷ วิเคราะห์แบบสอบถามคัดเลือกวิชาภาษาอังกฤษที่ใช้สอบคัดเลือกนิสิตที่จบปริญญาตรี ต้องการเข้าศึกษาต่อปริญญาโท คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ใช้วิเคราะห์ซอกกระทง 2 แบบ คือ เปิดตารางวิเคราะห์ข้อของจุฬแฟน กับคำนวณตามสูตรของ จอห์นสัน ผลการวิเคราะห์วิธีแรกค่าเฉลี่ยระดับความยากเท่ากับ .35 ค่าเฉลี่ยอ่านง่ายเท่ากับ .265 ซอกกระทงที่อยู่ในเกณฑ์คือ 57 ข้อ คิดเป็นร้อยละ 57 สำหรับวิธีที่ 2 ได้ค่าเฉลี่ยระดับความยากเท่ากับ .36 ค่าเฉลี่ยอ่านง่ายเท่ากับ .235 ซอกกระทงที่อยู่ในเกณฑ์มีจำนวน 51 ข้อ คิดเป็นร้อยละ 51

⁴⁶ฉวีวรรณ โปษีศรี, "การวิเคราะห์แบบสอบถามคัดเลือกของโรงเรียนเทคนิคในโครงการ เงินกู้เพื่อพัฒนาอาชีวศึกษา ปี 2517" (วิทยานิพนธ์ปริญญาโท ศึกษาศาสตร์ แผนกวิชาวิจัยการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, 2518)

⁴⁷มาโนช กลั่นฤทธิ, "การวิเคราะห์แบบสอบถามคัดเลือกวิชาภาษาอังกฤษ" (วิทยานิพนธ์ปริญญาโท ศึกษาศาสตร์ แผนกวิชาวิจัยการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, 2518)



การเปรียบเทียบการวิเคราะห์ชอกระหง 2 แบบ โดยการทดสอบ คอคราน คิว (Cochran Q Test) แยกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับความมีนัยสำคัญ .05 หากความเที่ยงชนิดความถี่ที่ภายในด้วยสูตรคูเคอร์ ริชาร์ดสัน สูตรที่ 20 จากการวิเคราะห์ 2 วิธีดังกล่าว ได้ค่าความเที่ยงเป็น .773 และ .770 ค่าความคลาดเคลื่อนมาตรฐานในการวัดเป็น 4.374 และ 4.402 ตามลำดับความถี่เชิงพยากรณ์ของแบบ สอบ เท่ากับ .85 ที่ระดับความมีนัยสำคัญเท่ากับ .01

และในโอกาสนี้ รจน ภัทรานนท์⁴⁸ ได้วิเคราะห์แบบสอบผลสัมฤทธิ์วิชาการศึกษาศึกษาวิทยาลัยครูเพชรบุรี 2 ฉบับ คือวิชาพัฒนาการเด็ก จำนวน 80 ข้อ และวิชาหลักการสอน จำนวน 100 ข้อ ผลการวิจัยพบว่า แบบสอบวิชาพัฒนาการเด็ก ชอกระหงทั้งหมดมีระดับความยากอยู่ระหว่าง 6% ถึง 100% และมีอำนาจจำแนกระหว่าง -.06 ถึง .48 ชอกระหงที่ตีสมควร เก็บไว้ใช้ต่อไปมีจำนวน 27 ข้อ ส่วนแบบสอบวิชาหลักการสอน ชอกระหงทั้งหมดมีระดับความยากระหว่าง 4% ถึง 99% มีอำนาจจำแนกระหว่าง -.18 ถึง .56 มีชอกระหงที่ตีสมควร เก็บไว้ใช้ต่อไป 40 ข้อ ตัวเลือกส่วนใหญ่ของทั้ง 2 แบบสอบ ยังไม่ตีต้องปรับปรุงและแก้ไข ผลการวิเคราะห์ครั้งที่ 2 ซึ่งได้จากการเลือกชอกระหงในแบบสอบฉบับเดิมที่มีค่าอำนาจจำแนก .20 ขึ้นไป พบว่าแบบสอบวิชาพัฒนาการเด็กฉบับใหม่จำนวน 29 ข้อ มีระดับความยากอยู่ระหว่าง 31% ถึง 83% อำนาจจำแนกระหว่าง .16 ถึง .50 มีชอกระหงที่ตีสมควร เก็บไว้ใช้ต่อไปจำนวน 27 ข้อ ส่วนในแบบสอบวิชาหลักการสอนฉบับใหม่จำนวน 42 ข้อ มีระดับความยากระหว่าง 23% ถึง 86% มีอำนาจจำแนกระหว่าง .16 ถึง .68 ชอกระหงที่ตีสมควร เก็บไว้ใช้ต่อไป 38 ข้อ ผลการคำนวณค่าความเที่ยงของแบบสอบฉบับเดิมและฉบับใหม่ ด้วยวิธีวิเคราะห์

⁴⁸รจน ภัทรานนท์, "การวิเคราะห์แบบสอบผลสัมฤทธิ์วิชาการศึกษาศึกษาวิทยาลัยครูเพชรบุรี" (วิทยานิพนธ์ปริญญาโท สาขาบริหารการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, 2518)

ความแปรปรวนของ ฮอยท์ (Hoyt) แบบสอบวิชาพัฒนาการ เด็กฉบับเดิมและฉบับใหม่ เท่ากับ .4867 และ .4033 ตามลำดับ แบบสอบวิชาหลักการสอนเท่ากับ .7140 และ .8972 ตามลำดับ ผลการคำนวณค่าสัมประสิทธิ์ความเที่ยงของแบบสอบฉบับใหม่เมื่อเทียบ ความยาวของแบบสอบให้เท่ากับความยาวของแบบสอบฉบับเดิม ในวิชาการพัฒนาการ เด็กและวิชาหลักการสอนเท่ากับ .6509 และ .9453 ตามลำดับ ความตรงความทำนาย ของแบบสอบวิชาพัฒนาการ เด็กฉบับเดิมและฉบับใหม่ เท่ากับ .4853 และ .4624 และ ในวิชาหลักการสอนทั้งฉบับเดิมและฉบับใหม่ เท่ากับ .5675 และ .5999 ตามลำดับ

ตัวอย่างรายงานการวิจัยในต่างประเทศ

ค.ศ. 1960 ฮิลการ์ด⁴⁹ (Hilgard) ได้สร้างแบบสอบนำไปทดสอบกับ นักเรียนที่แอสตันฟอร์ด (Stanford) และซานโฮเซ (San Jose) พบว่าความยากของ ข้อสอบมีค่าเท่ากันทั้ง 2 แห่ง ข้อสอบค่อนข้างยาก อ่านาจำแนก .31 ซึ่งค่อนข้างต่ำ และมีความสัมพันธ์กับเกณฑ์ค่อนข้างต่ำ เนื่องจากมีการเจตนา

ค.ศ. 1961 บลูม⁵⁰ (Bloom) ได้รวบรวมการวิเคราะห์แบบสอบเกี่ยวกับการ พยากรณ์ความสำเร็จในการเรียนชั้นมัธยม พบว่า การใช้คะแนนสะสมเฉลี่ยเป็นตัวพยากรณ์ ไม่ค่อยได้ผลดี แต่การใช้แบบสอบความถนัดวิชานั้นเป็นเกณฑ์และคะแนนเฉลี่ยสะสมวิชาเดียว

⁴⁹ Ernest R. Hilgard, "Stability of Item Analysis in Two Institution," Journal of Educational Psychology, (Vol. 51 No. 4 (1960), pp. 1955 - 1958.

⁵⁰ Benjamin Segel Bloom, Perter Frank R., "The Use of Academic Predicted Scale for Counselling and Selecting College Entrance (New York : The Fen Press of Otenceoe, Inc., 1961), pp.

กันเป็นคัพพยากรณได้ผลดีกว่า วิชาอะไรควรใช้คะแนนวิชาเดียวกันนั้นเป็นคัพพยากรณ ไม่ใช้วิชาหนึ่งไปพยากรณความสำเร็จของอีกวิชาหนึ่ง โดยที่เนื้อหาทั้ง 2 วิชา ไม่เกี่ยวข้องหรือเนื้อหาต่างกัน ด้วยเหตุนี้จึงนิยมใช้ แบบสอบความถนัดทางวิชาการไปพยากรณ ผลสัมฤทธิ์ทางการเรียนมากยิ่งขึ้น

ค.ศ. 1965 แคมป์เบลล์⁵¹ (Campbell) ได้ใช้แบบสอบวัดความถนัดทาง วิชาการ (General Scholastic Aptitude Test) และแบบสอบความถนัดทาง คณิตศาสตร์ (Mathematic Aptitude Test) กับแบบสอบเรียงความ (Written Expression) ใช้ศึกษาแก่นักศึกษา 406 คน ของมหาวิทยาลัยมิเนโซตา (Minnesota) เพื่อพยากรณความสำเร็จในการศึกษา หลักสูตรวิศวกรรมโยธาใช้คะแนนเฉลี่ยจากวิชา ต่าง ๆ ตามหลักสูตร 2 ปี เป็นเกณฑ์ ผลปรากฏว่าความสัมพันธ์พหุคูณระหว่างแบบสอบ แต่ละชุดมีค่าอยู่ระหว่าง .45 ถึง .55

ค.ศ. 1966 เพนน์ และ ทัทเทิล⁵² (Payne and Tuttle) ได้ทำการศึกษา ตรงตามทำนายของแบบสอบค่านเทกนอล (Miller Analogue Test) ซึ่งใช้คัดเลือก นักศึกษาระดับปริญญาโททางการศึกษา โดยทำการศึกษาความสัมพันธ์ระหว่างแบบสอบกับเกณฑ์

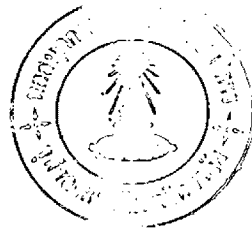
⁵¹John Paul Campbell, "The Use and Evaluation of an Interactive Multiple Regression Technique for Enhancing the Prediction of Academic Success by Criterion Group," Educational and Psychological Measurement, Vol. 34, No. 2 (1974), p. 373.

⁵²David A. Payne and Cynthia E. Tuttle, "The Prediction of Relationship of the Miller Analogies Test to Objective and Subjective Criterion of Success in a Graduate School of Education," Educational and Psychological Measurement, Vol. 26, No. 2 (1966), pp. 427 - 430.

2 แบบ คือ คะแนนเฉลี่ยสะสมในสาขาวิชาต่าง ๆ ได้แก่ จิตวิทยาทางการศึกษา วิชาผลและสถิติ พื้นฐานทางวัฒนธรรม และเกรคเฉลี่ย เกณฑ์อีกตัวหนึ่งได้แก่แบบสอบรวมก่อนจบการศึกษา (Comprehensive Examination) ซึ่งแบ่งเป็นแบบสอบย่อย 3 ชุดตามสาขาดังกล่าวมาแล้ว เป็นแบบสอบชนิดเลือกตอบและจับคู่ กลุ่มตัวอย่างที่ใช้ในการวิจัยจำนวน 219 คน เป็นชาย 115 คน และหญิง 109 คน และใช้เวลาศึกษาระหว่างเดือนมิถุนายน 1958 ถึงเดือนมีนาคม 1963 ปรากฏว่า การแจกแจงของเวลาระหว่างเวลาสอบ MAT กับเวลาที่สำเร็จการศึกษา เบี่ยงขวา ค่าความเที่ยง (Reliability) หาโดยสูตรคูเคอร์ริชาร์ดสันสูตรที่ 20 มีค่า .47, .82, .65 และ .84 ตามลำดับ แบบสอบ MAT ทำนายผลการสอบ Comprehensive Examination ได้ดีกว่าเกรคเฉลี่ย และค่าความตรงบางสาขาต่ำมาก แสดงให้เห็นว่าการให้เกรคของแต่ละแผนกมีความเอนเอียง (bias) ค่าสหสัมพันธ์ระหว่างเกรคเฉลี่ยกับคะแนนรวมของแบบสอบ เท่ากับ .56 เป็นการประเมินการวัดผลที่แตกต่างกัน

ค.ศ. 1968 เวสต์บรูก และ โจนส์⁵³ (Westbrook and Jones) ได้วิเคราะห์แบบสอบที่นักศึกษาปริญญาโท มหาวิทยาลัย North Carolina State ที่เรียนวิชาวัดผลการศึกษาได้ช่วยกันสร้างไว้ ผลการวิเคราะห์พบว่า ในแบบสอบฟอร์มเอ (Form A) มีค่าสัมประสิทธิ์ความเที่ยงแบบสอบซ้ำ (Test - Retest) เป็น .82 ค่าสัมประสิทธิ์ความเที่ยงแบบแบ่งครึ่ง (Split - Half) เป็น .73 ค่าสัมประสิทธิ์ความเที่ยงคำนวณโดยสูตร คูเคอร์ ริชาร์ดสัน (Kuder - Richardson Formula) มีค่า .62 แบบสอบฟอร์มบี (Form B) มีค่าสัมประสิทธิ์ความเที่ยงแบบสอบซ้ำเท่ากับ

⁵³Bert W. Westbrook and Charles I. Jones, "The Reliability and Validity of a class Constructed Measure of Achievement in Test and Measure," Educational and Psychological Measurement, Vol 28, No. 3 (1968), pp. 484 - 486.



.71 ค่าสัมประสิทธิ์ความเที่ยงแบบฟอร์มคู่ขนาน (Pararell Form) ของแบบฟอร์ม เอ กับฟอร์มบี ในการสอบครั้งแรกและครั้งหลังเป็น .63 และ .62 ตามลำดับ ค่าสัมประสิทธิ์ความตรงของแบบฟอร์มทั้งสองเป็น .75 และ .59 ตามลำดับ

ค.ศ. 1969 เบิร์ก (Burke) และเพื่อน⁵⁴ ได้ศึกษาความตรงตามเกณฑ์ เทียบ (Criterion - Related Validity) ของเครื่องมือวัดความสามารถทางภาษาอังกฤษ ของนักศึกษาต่างชาติที่เข้าศึกษาต่อในมหาวิทยาลัย Southern California เครื่องมือดังกล่าวประกอบด้วยแบบผสมสัมฤทธิ์ผล 6 ชุดด้วยกัน ได้แก่ 1. California Reading Test Advanced Form, 2. U.S.C. English Classification Examination, 3. Brown - Carlsen Listening Comprehension Test, 4. English Theme, 5. Speech and Interview และ 6. Larry - Ward English Examination for foreign Students ทั้ง 6 ชุด จะสอบทุก ๆ คน เหนอเป็นเวลา 3 ปี ติดต่อกัน ส่วนเกณฑ์มี 5 ตัว ได้แก่ 1. Grade point Average, 2. Academic Standing, 3. Final Grade in English Communication Program for foreign Student (ECPFS') 4. The Cooperative School and College Ability Test (SCAT) Verbal score and 5. The SCAT Quantitative Score.

⁵⁴ Jack D. Burke, William B. Michael, Robert B. Kaplan, and Robert A. Jones "The Criterion Related Validity of English language Screening Instrument for foreign Students Entering the University of Southern California," Educational and Psychological Measurement, Vol. 29, No. 3 (1969), pp. 503 - 506.

กลุ่มตัวอย่าง เป็นนักศึกษาต่างชาติ จำนวน 178 คน พุฒภาษาประจำชาติที่ไม่ใช่ภาษาอังกฤษ 60% เป็นนักศึกษาระดับปริญญาโท วิชาเอกทางสังคมศาสตร์และวิจิตร มาจากกลุ่มที่เรียนภาษาอังกฤษในระดับความสามารถที่แตกต่างกัน 4 กลุ่ม ผลการวิจัยสรุปผลได้ว่า ค่าสหสัมพันธ์พหุคูณของแบบสอบทั้ง 6 ชุด ที่ใช้ทำนาย GPA สูงกว่า Academic Standing และทำนายตัวที่ 5 และ 6 เป็นตัวทำนายที่ดี ค่าสหสัมพันธ์ (Inter correlation) ระหว่างตัวทำนายด้วยกันเองมีค่าค่อนข้างสูง แต่ในระหว่างตัวเกณฑ์ด้วยกันค่อนข้างต่ำ ค่าสหสัมพันธ์ระหว่างตัวทำนายแต่ละตัวกับเกณฑ์ตัวที่ 2 มีค่าสูง แบบสอบทั้ง 6 ชุด รวมกันทำนายเกณฑ์แต่ละตัว ให้ค่าความตรงสูงกว่าที่จะทำนายโดยแบบสอบชุดใดชุดหนึ่งเพียงชุดเดียว

ในปีเดียวกัน Anastasi⁵⁵ ได้ศึกษาลักษณะของแบบสอบ รายงานว่า แบบสอบใช้ความเร็ว (Speed Test) ซึ่งเป็นข้อสอบค่อนข้างง่าย ใช้สูตรคูเดอร์ ริชาร์ดสัน (Kuder Richardson) ทาคความเที่ยงของแบบสอบไม่ได้ เพราะผู้เข้ารับการสอบทุกคนไม่มีโอกาสตอบข้อกระทงทุกข้อที่ทำได้เนื่องจากเวลาจำกัดมาก ทำให้ค่าความเที่ยงคลาดเคลื่อนจากความเป็นจริง ควรใช้วิธีสอบซ้ำ (Test - Retest) ส่วนแบบสอบใช้ความสามารถ (Power Test) นั้นเป็นแบบสอบค่อนข้างยากกว่าแบบสอบใช้ความเร็ว และได้เวลานานพอที่ทุกคนตอบทันทุกข้อกระทง

ค.ศ. 1970 Costin⁵⁶ ได้ศึกษาเกี่ยวกับตัวเลือกของแบบสอบชนิดเลือกตอบ (Multiple Choice) โดยได้นำแบบสอบวิชาจิตวิทยา ซึ่งเขาและ

⁵⁵Anastasi, op. cit., pp. 89 - 91.

⁵⁶Frank Costin, "The Optimal Number of Alternative in Multiple Choice Achievement Test : Some Empirical Evidence for a Mathematical," Educational and Psychological Measurement, Vol. 30, No. 2 (1970), pp. 353 - 357.

ร่วมงานได้สร้างไว้ และวิเคราะห์หาค่าความยากและอำนาจจำแนกของข้อสอบแต่ละข้อไว้แล้ว นำกลังแปลงเป็นแบบสอบชนิดเลือกตอบ 4 ตัวเลือก และชนิด 3 ตัวเลือกรวม 2 ฉบับ แล้วนำไปทดสอบกับนักศึกษามหาวิทยาลัย อิลลินอยส์ (Illinois)

จำนวน 207 คน และคำนวณหาค่าความยากและอำนาจจำแนก และความเที่ยงของแบบสอบทั้ง 2 ฉบับ พบว่าแบบสอบชนิด 3 ตัวเลือกมีค่าเฉลี่ยของอำนาจจำแนก และค่าความเที่ยงสูงกว่าแบบเลือกตอบ 4 ตัวเลือก

ในปีเดียวกัน หว่อง และ คิสเนีย⁵⁷ (Hwang and Disney) ได้ศึกษาความตรงตามทำนายของแบบสอบ TOEFL (Test of English as a second Language) ซึ่งเป็นแบบสอบวัดความสามารถทางภาษาอังกฤษ สำหรับนักศึกษาต่างชาติที่ต้องการสมัคร เข้าศึกษาต่อในมหาวิทยาลัยในสหรัฐอเมริกา โดยใช้เกรดเฉลี่ยรวมและเกรดเฉลี่ยในวิชาภาษาอังกฤษเป็นเกณฑ์ ตัวอย่างที่ใช้เป็นนักศึกษาชาวจีน ของมหาวิทยาลัยโอเรกอน (Oregon) จำนวน 63 คน มีประสบการณ์เรียน 7 - 12 ปี เรียนในสาขาต่าง ๆ 5 กลุ่ม ได้แก่การศึกษา (Education), สังคม (Social and Professional Service) วิทยาศาสตร์ (Natural Sciences) สถาปัตยกรรมศาสตร์ (Architecture) มีจำนวน 21, 16, 11, 9 และ 6 คน ตามลำดับ พบว่า ค่าสหสัมพันธ์ระหว่างแบบสอบ TOEFL กับเกรดเฉลี่ยสะสมแยกแต่ละกลุ่มและทุกกลุ่มรวมกัน เท่ากับ .21, .05, .22, .32, .69 และ .19 ตามลำดับ แต่ไม่มีนัยสำคัญที่ระดับ .05 และค่าสหสัมพันธ์ระหว่าง TOEFL กับเกรดเฉลี่ยวิชาภาษาอังกฤษ เท่ากับ .66 มีนัยสำคัญที่ระดับ .05

⁵⁷Kwo - Yann Hwang and Henry E. Disney, "Predictive Validity of the Test of English as a Foreign Language for Chinese Graduate Student at an American University," Educational and Psychological Measurement, Vol. 30, No. 2 (1970), pp. 475 - 477.

ค.ศ. 1971 ทัทแมน และ ทัทแฮม⁵⁸ (Tatman and Tatham) ได้วิเคราะห์แบบสอบวิชาพีชคณิต 3 (Algebra III) ของมหาวิทยาลัยออกตาวา ที่ใช้คัดเลือกนักศึกษาในการเลือกเรียนวิชาต่าง ๆ พบว่าแบบสอบมีค่าสัมประสิทธิ์ความเที่ยง ซึ่งกำหนดโดยวิธีวิเคราะห์ความแปรปรวนของฮอยท์ (Hoyt's Analysis of Variance) เท่ากับ .8995 และค่าความเที่ยงที่คำนวณโดยสูตรของ ซอเป (Saupe's Method) เท่ากับ .865 ค่าความคลาดเคลื่อนมาตรฐานในการวัด (Standard Error of Measurement) เท่ากับ 2.52 ความตรงตามทำนายของแบบสอบศึกษาจากผลการเรียนวิชาแคลคูลัสและเรขาคณิตวิเคราะห์ (Calculus and Geometric Analysis) ของนักศึกษา 113 คน พบว่า นักศึกษากลุ่มที่ประสบผลสำเร็จในการเรียน (เกรด C ขึ้นไป) และนักศึกษากลุ่มที่ไม่ประสบความสำเร็จในการเรียน (เกรด D ลงมา) มีคะแนนการสอบวิชาพีชคณิต 3 (Algebra III) แตกต่างกันอย่างมีนัยสำคัญ ซึ่งแสดงว่าคะแนนจากแบบสอบพีชคณิต 3 สามารถทำนายสัมฤทธิ์ผลในการเรียนคณิตศาสตร์ของนักศึกษาได้

ค.ศ. 1973 ฮอปกินส์ (Hopkins) และเพื่อน⁵⁹ ได้ศึกษาเปรียบเทียบระหว่างวิธีคำนวณคะแนนแบบธรรมดา (Conventional score) กับวิธีที่ให้น้ำหนักคะแนนความเชื่อมั่น (Confident Weighting) ด้วย กล่าวคือจะต้องกำหนดลงไป

⁵⁸Cliffard B. Tatman and Elaine J. Tatham, "A note on the Predictive Validity of the Cooperative Algebra III," Educational and Psychological Measurement, Vol. 31, No. 2 (1971), pp. 517 - 518.

⁵⁹Kenneth D. Hopkins, A Ralph Hakstian and B.R. Hopkins, "Validity and Reliability Consequence of Confidence Weighting," Educational and Psychological Measurement, Vol. 33, No. 1 (1973), pp. 135 - 140.

คว่ำว่าค่าตอบนั้น ๆ มีความมั่นใจเพียงใด โดยเขียนเป็นอักษร H, M และ L ซึ่งหมายถึงมั่นใจสูง กลาง และต่ำ ตามลำดับ กำหนดคะแนนดังนี้คือ ถ้าค่าตอบนั้นถูกต้อง และเขียนอักษร H หรือ M หรือ L จะได้คะแนน 3, 2 และ 1 ตามลำดับ ในทางตรงข้ามถ้าค่าตอบนั้นผิด การเขียน H, M และ L จะได้คะแนน -3, -2 และ -1 ตามลำดับ ตัวอย่างในการทดลองเป็นนักศึกษาปริญญาโท. จำนวน 63 คน ของมหาวิทยาลัย Colorado เครื่องมือที่ใช้เป็นแบบสอบไล่วิชาสถิติ แบบเลือกตอบ (Multiple Choice) จำนวน 65 ข้อ และใช้แบบสอบอีกชุดหนึ่ง ซึ่งครอบคลุมเนื้อหาเกี่ยวกับชุดแรก แต่ขนาดสั้นกว่า เป็นเกณฑ์ พบว่าค่าความเที่ยงโดยวิธีแบ่งครึ่งแบบสอบ (Speit - Half Method) ของแบบสอบชนิดนี้ นำหนักคะแนนความมั่นใจ ใจด้วยเท่ากับ .915 สูงกว่าแบบธรรมดา ซึ่งเท่ากับ .833 เพียงเล็กน้อย และไม่มีนัยสำคัญที่ระดับความเชื่อมั่น .10 ค่าเฉลี่ย (mean) ของคะแนนจากทั้ง 2 วิธี ต่างกันเล็กน้อย และมีแนวโน้มที่นักศึกษาที่มีความสามารถสูงจะตอบด้วยความมั่นใจ สูงกว่า ซึ่งแสดงด้วยค่าสหสัมพันธ์ระหว่างผลรวมของน้ำหนักคะแนน (sum of the confident weights) กับเกณฑ์เท่ากับ .419 ที่ระดับความเชื่อมั่น .01 และกับคะแนน GRE (Graduate Record Examination) เท่ากับ .32 ที่ระดับความเชื่อมั่น .10 ค่าความตรง (Validity) ของวิธีที่สองต่ำกว่าวิธีแรก และถ้าใช้คะแนนจากทั้ง 2 วิธี รวมกันทำนายเกณฑ์จะได้ค่าสหสัมพันธ์พหุคูณเท่ากับ .719 ถ้าใช้วิธีที่ 1 อย่างเดียวจะได้ค่าความตรงเท่ากับ .701

จากตัวอย่างรายงานการวิจัยดังกล่าว จะเห็นว่าได้มีการวิเคราะห์แบบสอบทั้งในประเทศและต่างประเทศเป็นจำนวนมาก ส่วนใหญ่จะเป็นวิเคราะห์หาค่าสถิติที่สำคัญ เช่น ระดับความยาก อรรถาธิบายจำแนก ความเที่ยง ความตรงของแบบสอบ รวมทั้งการหาเทคนิควิธีต่าง ๆ เพื่อตรวจสอบคุณภาพข้อกระทงและปรับปรุงคุณภาพของแบบสอบให้ดียิ่งขึ้น โดยแบบสอบที่นำมาวิเคราะห์นั้น มีทั้งแบบสอบวัดความถนัด แบบสอบวัดคุณลักษณะวิชาต่าง ๆ และแบบสอบคัดเลือกรับเข้าศึกษาต่อ