บทที่ 2

วรรณคดีที่เกี่ยวข้อง

ในการวิจัยครั้งนี้ ผู้วิจัยได้นำเสนอวรรณคดีและงานวิจัยที่เกี่ยวข้อง โดยแบ่งเป็น 3 ตอนดังนี้

ตอนที่ 1 มโนทัศน์เบื้องต้นเกี่ยวกับความเที่ยง ความตรง

ตอนที่ 2 มโนทัศน์เกี่ยวกับทฤษฎีการให้น้ำหนักคะแนนตามวิธีของกิบบอนส์และคณะกับวิธีของลอร์ด

ตอนที่ 3 งานวิจัยที่เกี่ยวข้อง

ตอนที่ 1 มโนทัศน์เบื้องต้นเกี่ยวกับความเที่ยง ความตรง

ความเที่ยงของแบบสอบ

มีผู้ให้นิยามความเที่ยง (Reliability) ไว้หลายท่าน ดังนี้

ความเที่ยง หมายถึง ดัชนีของการทำนายค่าคะแนนสอบฉบับหนึ่ง จากแบบสอบคู่ขนานอีกฉบับหนึ่ง (G hiselli, 1964)

ความเที่ยง หมายถึง ค่าความคงที่ของคะแนนสอบเมื่อใช้วัดกลุ่มผู้สอบกลุ่มเดียวกันและด้วยแบบสอบชุดเดียวกัน อย่างน้อย 2 ครั้ง (Gronlund, 1968)

ความเที่ยง หมายถึง ระดับความคงที่หรือความคงเส้นคงวาของคะแนนสอบจากการทดสอบเรื่องเดียวกันในเวลาใดก็ตาม (เยาวดี วิบูลย์ศรี, 2526)

จากนิยามหรือความหมายดังกล่าว จึงสรุปได้ว่า ค่าความเที่ยง หมายถึงค่าความคงที่ของคะแนนสอบ จากการสอบด้วยแบบสอบชุดเดียวกัน หรือจากแบบสอบชุดอื่นที่เทียบเคียงกัน ในเวลาต่างกันจากกลุ่มบุคคลเดียวกัน

<u>วิธีประมาณค่าความเที่ยงของแบบสอบ</u>

โดยทั่วไปวิธีประมาณค่าความเที่ยงมีดังนี้ (Mehrens and Lehman, 1984)

1. แบบวัดซ้ำ (Measures of Stability)

2. แบบใช้ข้อสอบคู่ขนาน (Measure of Equivalence)

3. แบบใช้ข้อสอบคู่ขนานและวัดซ้ำ (Measure of Equivalence and Stability)

4. แบบวัดความคงที่ภายใน (Measure of Internal consistency) ซึ่งสามารถวัดได้หลายวิธีดังนี้

4.1 แบบแบ่งครึ่งข้อสอบ (Split-half)

4.2 แบบคูเดอร์-ริชาร์ดสัน (Kuder Richardson)

4.3 แบบวิเคราะห์ความแปรปรวนของฮอยท์ (Hoyt's Analysis of Variance)

4.4 แบบสัมประสิทธิ์แอลฟาของครอนบัค (Bronbach's Alpha Coefficient)

5. ค่าความเที่ยงของผู้ให้คะแนน (Score (judge) reliability)

ในที่นี้จะกล่าวถึง ความเที่ยงแบบสัมประสิทธิ์แอลฟา ซึ่งเป็นวิธีประมาณค่าความเที่ยงที่ใช้ในการวิจัยครั้งนี้

<u>ความเที่ยงแบบสัมประสิทธิ์แอลฟา</u> (Coefficient Alpha: α)

Cronbach (1951 อ้างจาก Mehren and Lehman 1984) ได้พัฒนาวิธีการประมาณค่าความเที่ยงจากวิธีของคูเดอร์-ริชาร์ดสันขึ้น เมื่อค่าคะแนนของแบบสอบไม่สามารถแบ่งเป็น 2 ส่วน (Dichotomous) ได้ เช่นในกรณีแบบสอบความเรียง (Essay) ค่าคะแนนสอบในแต่ละข้อคำถามจะแตกต่างกันออกไป สูตรในการคำนวณหาสัมประสิทธิ์แบบแอลฟาก็พัฒนามาจากสูตร K - R 20 แต่เปลี่ยนค่า $\sum pq$ เป็น $\sum S^2_i$ โดยที่ $\sum S^2_i$ หมายถึงผลรวมความแปรปรวนรายข้อของข้อกระทง ดังนั้นสูตรที่ใช้ในการคำนวณ คือ

$$\alpha = \frac{n}{(n-1)}\left[1 - \frac{\sum S^2_i}{S^2_x}\right]$$

โดยที่ α หมายถึง ค่าสัมประสิทธิ์แอลฟา

n หมายถึง จำนวนข้อกระทง

$\sum S^2_i$ หมายถึง ผลรวมความแปรปรวนของข้อกระทงในแต่ละข้อ

S^2_x หมายถึง ความแปรปรวนของคะแนนข้อสอบทั้งหมด

<u>องค์ประกอบที่มีอิทธิพลต่อค่าความเที่ยงของแบบสอบ</u>

นอกเหนือจากวิธีการประมาณความเที่ยงที่มีผลต่อค่าความเที่ยงของแบบสอบแล้ว องค์ประกอบอื่นที่มีอิทธิพลต่อค่าความเที่ยง มีหลายประการดังนี้ (Mehren and Lehman, 1984)

1. ความยาวของแบบสอบ (Test Length) โดยที่ถ้าแบบสอบมีความยาวเพิ่มขึ้น ความเที่ยงก็จะเพิ่มขึ้น ซึ่งแสดงให้เห็นความสัมพันธ์ระหว่างความยาวของแบบสอบกับความเที่ยง คือ สูตรของสเปียร์แมน บราวน์ (Spearman Brown Formula)

$$r_{tt} = \frac{kr}{1 + (k-1)r}$$

เมื่อ r_{tt} คือ ค่าความเที่ยงของแบบสอบที่ยาวขึ้น k เท่า

k คือ อัตราส่วนของข้อสอบใหม่ที่เพิ่มขึ้น

r คือ ค่าความเที่ยงเดิม

เช่น สมมติว่า แบบสอบชุดหนึ่งมีค่าความเที่ยง .70 ถ้าเพิ่มความยาวเป็น 2 เท่า ความเที่ยงใหม่จะเป็น

$$r_{tt} = \frac{2 \times .70}{1 + (1 \times .70)} = \frac{1.4}{1.7} = .823$$

2. ความเร็ว (Speed) การประมาณค่าความเที่ยงของแบบสอบประเภทความเร็ว (Speed Test) ด้วยวิธีวัดความคงที่ภายใน เป็นวิธีที่ไม่เหมาะสม เพราะแบบสอบชนิดนี้เป็นแบบสอบที่ง่ายแต่ให้เวลาในการทำน้อยนักเรียนไม่สามารถทำได้ทุกข้อ จะเกิดความแปรปรวนของคะแนนสูง อันจะมีผลทำให้ค่าความเที่ยงสูงขึ้น

3. ความเป็นเอกพันธ์ของกลุ่มผู้รับการทดสอบ (Group Homogeneity) โดยกลุ่มที่มีความเป็นเอกพันธ์จะมีพิสัยของคะแนนความสามารถของกลุ่มน้อยกว่า กลุ่มที่เป็นวิวิธพันธ์ (Group Heterogeneity) ซึ่งจะทำให้ค่าความเที่ยงของแบบสอบของกลุ่มที่มีความเป็นเอกพันธ์จะต่ำกว่า กลุ่มที่เป็นวิวิธพันธ์

4. ความยากของข้อสอบ (Difficulty of Items) ข้อสอบที่ยากเกินไปหรือง่ายเกินไป จะทำให้พิสัยของคะแนนลดลง ซึ่งจะมีผลทำให้ค่าความเที่ยงต่ำลง

5. ความเป็นปรนัย (Objectivity) ถ้าแบบสอบมีความเป็นปรนัยสูง ค่าความเที่ยงจะสูงขึ้น

6. เวลาในการทำแบบสอบ เช่น ถ้าเวลาทำในการสอบน้อยเกินไป ผู้สอบทำไม่ทันก็อาจเดาคำตอบ ทำให้คะแนนมีความแปรปรวนสูงมาก หรือ ถ้าให้เวลามาก จะทำให้คนเรียนเก่ง และอ่อน ๆ ทำให้เท่ากัน ทำให้คะแนนมีความแปรปรวนน้อย ค่าความเที่ยงจะลดลง โดย Cook (1932 อ้างถึงใน สุพัฒน์ สุกลสันต์, 2528) ศึกษาพบว่า เวลาในการสอบมีผลต่อความเที่ยงอย่างมีนัยสำคัญ กล่าวคือ ในแบบสอบสะกดคำฉบับหนึ่ง ที่ให้เวลาพอดีจะให้ ค่าความเที่ยง .91 - .92 แต่ถ้าให้เวลามากขึ้น ความเที่ยง จะลดลงเป็น .71 - .85 เท่านั้น

7. การเดา (Guessing) ถ้าแบบสอบมีการเดามาก ค่าความเที่ยงจะลดลง

นอกจากนี้ การบริหารแบบสอบ คำสั่งในการตอบ การตกลงให้คะแนน หรือองค์ประกอบภายในตัวผู้สอบ เช่น ความวิตกกังวลแรงจูงใจ ประสบการณ์ในการสอน ล้วนแล้วแต่มีผลต่อค่าความเที่ยงทั้งสิ้น

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

018426

ความตรงของแบบสอบ (Validity)

มีผู้ให้นิยามเกี่ยวกับความตรงไว้หลายท่าน คือ

ความตรง หมายถึง ค่าความถูกต้องของคะแนนสอบที่วัดในสิ่งที่ต้องการวัด (Ebel, 1972)

ความตรง หมายถึง ดัชนีที่มุ่งถึงระดับความสามารถของแบบสอบที่บรรลุจุดมุ่งหมายเฉพาะอย่างใดอย่างหนึ่ง (คณะกรรมการมาตรฐานทางการศึกษาและจิตวิทยา, 1967)

ความตรง หมายถึง ความสามารถที่แบบสอบนั้นสามารถวัดได้ในขอบเขตที่ต้องการภายใต้สถานการณ์หนึ่งกับประชากรกลุ่มหนึ่ง (เยาวดี วิบูลย์ศรี, 2526)

จากความหมาย สรุปได้ว่า ความตรงของแบบสอบ หมายถึง คุณสมบัติของแบบสอบที่สามารถในสิ่งที่ต้องการวัดได้อย่างถูกต้องและแม่นยำ

ประเภทของความตรง

ความตรงมีหลายชนิด สามารถแบ่งเป็นประเภทใหญ่ ๆ ได้ 3 ประเภท คือ (สมาคมจิตวิทยาอเมริกา, 1974 อ้างถึงใน Mehren and Lehmah, 1984)

1. ความตรงตามเนื้อหา (Content Validity)
2. ความตรงตามเกณฑ์สัมพันธ์ (Criterion Related Validity) ซึ่งได้แก่
   2.1 ความตรงเชิงทำนาย (Predictive Validity)
   2.2 ความตรงตามสภาพ (Concurrent Validity)
3. ความตรงตามทฤษฎี (Construct Validity)

ในที่นี้ผู้วิจัยจะกล่าวถึงค่าความตรงที่เกี่ยวข้องกับงานวิจัย คือ ค่าความตรงตามสภาพ ซึ่งเป็นชนิดหนึ่งของความตรงตามเกณฑ์สัมพันธ์เท่านั้น

ความตรงตามเกณฑ์สัมพันธ์ (Criterion Related Validity)

ความตรงตามเกณฑ์สัมพันธ์ของแบบสอบ หมายถึง ความสามารถของแบบสอบในการทำนายพฤติกรรมของบุคคลในสถานการณ์เฉพาะ หรือแสดงถึงความสัมพันธ์ระหว่างคะแนนที่ผู้สอบกระทำได้จากแบบสอบกับเกณฑ์ภายนอก (External Criterion) แบ่งเป็น

1. ความตรงเชิงทำนาย (Predictive Validity) โดยจะบ่งชี้ถึงความสัมพันธ์ระหว่างคะแนนจากการสอบกับผลิตผลที่คาดว่าจะเกิดขึ้นในอนาคต การใช้ความตรงในแบบสอบประเภทนี้ มีวัตถุประสงค์เพื่อใช้ทำนายอนาคตของบุคคลจากความสามารถที่ผู้สอบนั้นทำได้ในปัจจุบัน

2. ความตรงตามสภาพ (Concurrent Validity) ของแบบสอบ หมายถึง ความสามารถของแบบสอบที่จะบ่งชี้ว่า ผู้สอบมีความสามารถ หรือคุณสมบัติตามเกณฑ์ที่ตั้งขึ้นจากสภาพปัจจุบันเพียงใด หรือความสัมพันธ์ระหว่างคะแนนที่ได้จากการทำแบบสอบ กับคะแนนเกณฑ์ที่ได้มาจากการกำหนดในขณะนั้น (เยาวดี วิบูลย์ศรี, 2526) เช่น แบบสอบที่ใช้วัดผลสัมฤทธิ์ทางการเรียนวิชาวิทยาศาสตร์ ถ้าต้องการทราบว่าแบบสอบนั้นมีความตรงตามสภาพของแบบสอบหรือไม่ ให้นำคะแนนที่ได้จากการสอบไปหาความสัมพันธ์กับคะแนนเกณฑ์ที่กำหนดขึ้น อาจจะเป็นคะแนนผลสัมพันธ์ทางการเรียนวิชาวิทยาศาสตร์ในภาคเรียนก่อน ถ้ามีความสัมพันธ์กันสูง ก็แสดงว่าแบบสอบมีความตรงตามสภาพ นอกจากนั้นยังสามารถหาความตรงตามสภาพของแบบสอบได้โดยการเปรียบเทียบกลุ่มที่แบ่งขึ้นตามเกณฑ์ (Brown, 1976 อ้างถึงใน สำราญ มีแจ้ง, 2525) โดยดูค่าคะแนนเฉลี่ยที่ได้จากแบบสอบของกลุ่ม

<u>การวัดเกณฑ์</u> (Measuring the Criterion)

Mehren and Lehman (1984) ได้กล่าวถึง หลักการในการพิจารณาเกณฑ์ที่ใช้ในการวัด เพื่อหาค่าความตรงตามเกณฑ์สัมพันธ์ คือ

1. ความเกี่ยวข้อง (Relevance) เกณฑ์ที่ใช้วัดจะต้องมีความสัมพันธ์เกี่ยวข้องกับเรื่องที่จะศึกษา และจะต้องสะท้อนให้เห็นคุณลักษณะที่สำคัญของสิ่งที่ศึกษาอย่างเพียงพอ ซึ่งโดยปกติแล้วการพิจารณาความเกี่ยวข้องจะเป็นการตัดสินคุณค่า

2. ความเที่ยง (Reliability) นอกจากความเที่ยงของแบบสอบจะมีผลต่อความตรงตามเกณฑ์สัมพันธ์แล้ว ความเที่ยงของเกณฑ์ก็จะมีผลต่อค่าความตรงตามเกณฑ์สัมพันธ์ด้วย กล่าวคือ โดยทฤษฎีความสัมพันธ์ทั่วไป พบว่า ความสัมพันธ์ที่มีค่าสูงสุดระหว่างตัวแปร 2 ตัว จะมีค่าเท่ากับหรือน้อยกว่า รากที่สองของผลคูณของค่าความเที่ยง ดังนั้น ค่าความตรงตามเกณฑ์สัมพันธ์ จึงเขียนสมการได้เป็น

$$r_{xy} \leq \sqrt{(r_{xx})(r_{yy})}$$

เมื่อ r_{xy} คือ สหสัมพันธ์ระหว่างตัวทำนาย x กับตัวเกณฑ์ Y

r_{xx} คือ ความเที่ยงของแบบสอบหรือตัวทำนาย

r_{yy} คือ ความเที่ยงของเกณฑ์

แสดงว่า ค่าความตรงตามเกณฑ์สัมพันธ์ จะมีค่ามากหรือน้อยก็ขึ้นอยู่กับค่าความเที่ยงของตัวทำนาย และความเที่ยงของตัวเกณฑ์

3. ความเป็นอิสระปราศจากอคติ (Free from bias) นั่นคือ เกณฑ์ที่ใช้วัดจะต้องไม่เป็นเกณฑ์มลทิน (Criterion Contamination) กล่าวคือ เกณฑ์มลทินจะเกิดขึ้นโดยได้รับอิทธิพลจากการรู้คะแนนทำนายตัวอย่างเช่น ตัวอย่างเช่น การนำผลการสอบของนักเรียนในตอนต้นปีไปทำนายผลการเรียนในปลายปี ถ้าครูผู้ประเมินได้นำความรู้ที่ได้จากคะแนนทำนายไปมีส่วนประเมินผลการเรียนของนักเรียนในปลายปี โดยอาจไม่มีผลทำให้การประเมินนั้นสูงขึ้นหรือต่ำลง แสดงว่า คะแนนที่ใช้เป็นเกณฑ์นั้นจะไม่มีความเป็นอิสระ และเป็นเกณฑ์มลทิน ซึ่งทำให้ค่าความตรงตามเกณฑ์สัมพันธ์ของแบบสอบสูงขึ้น วิธีที่จะป้องกันไม่ให้เกิดเกณฑ์มลทิน ก็คือ ไม่ให้ผู้ประเมินได้ทราบคะแนนที่ใช้ในการทำนายนั่นเอง

<u>องค์ประกอบที่มีอิทธิพลต่อค่าความตรงของแบบสอบ</u>

กรอนลันด์ (Gronlund, 1981) ได้กล่าวถึงองค์ประกอบที่มีผลต่อความตรงของแบบสอบดังนี้

1. องค์ประกอบภายในแบบสอบ

1.1 คำสั่งไม่ชัดเจน เช่น คำสั่งที่ไม่ชัดเจนเกี่ยวกับวิธีตอบข้อสอบ การบันทึกคำตอบ ซึ่งทำให้ค่าความตรงลดลง

1.2 โครงสร้างของประโยคและคำศัพท์ที่ยากเกินไป ทำให้เกิดความสับสนแก่ผู้สอบ ผู้ที่ตอบแบบสอบได้ถูกต้องกลับเป็นการวัดความสามารถทางภาษา

1.3 ระดับความยากของข้อสอบที่ไม่เหมาะสม ข้อสอบอิงกลุ่มที่ง่ายหรือยากเกินไป ทำให้ไม่สามารถจำแนกนักเรียนได้ ทำให้ค่าความตรงลดลง ส่วนข้อสอบอิงเกณฑ์ที่มีความยากไม่เหมาะสมกับสภาพการเรียนรู้ มีผลทำให้ค่าความตรงลดลง

1.4 การสร้างข้อสอบที่ไม่ดีและไม่ระมัดระวัง อาจเป็นการชี้แนะคำตอบได้

1.5 ข้อสอบที่ใช้ภาษากำกวม ทำให้นักเรียนเกิดความสับสนและแปลความผิดพลาด

1.6 การใช้ชนิดของแบบสอบที่ไม่เหมาะสมกับพฤติกรรมที่ต้องการวัด

1.7 ความยาวของแบบสอบ แบบสอบที่สั้นเกินไปจะไม่สามารถวัดได้ครอบคลุมวัตถุประสงค์ของการวัด

1.8 การจัดเรียงข้อกระทงไม่เหมาะสม โดยปกติข้อกระทง ควรเรียงจากง่ายไปหายาก แบบสอบที่เรียงข้อกระทงจากยากไปหาง่าย ทำให้นักเรียนใช้เวลาในการคิดหาคำตอบนาน จนอาจไม่มีเวลาในการทำข้อสอบข้อหลัง และอาจมีผลไปลดแรงจูงใจในการทำข้อสอบให้ลดลง จึงทำให้วัดไม่ตรงกับความสามารถที่แท้จริง

1.9 การจัดเรียงข้อถูกอย่างเป็นระบบ ทำให้นักเรียนสามารถเดาคำตอบถูกในข้อหลัง ๆ ได้

2. องค์ประกอบด้านเนื้อหาและวิธีดำเนินการสอน การสอนในเนื้อหาที่ตรงกับเนื้อหาที่ต้องการวัด ทำให้แบบสอบนั้นขาดความตรงในการวัดขบวนการทางสมองที่ซับซ้อน

3. องค์ประกอบเกี่ยวกับการบริหารแบบสอบ และการให้คะแนน เช่น เวลาในการสอบที่น้อยเกินไป การชี้แนะนักเรียนบางคนในการตอบข้อสอบ การให้คะแนนที่ผิด การไม่ปฏิบัติตามคู่มือในการสอบ ทำให้ความตรงของแบบสอบลดลง

4. องค์ประกอบภายในตัวผู้สอบ เช่น ความวิตกกังวล การขาดแรงจูงใจ การเปลี่ยนแปลงทางอารมณ์ สิ่งเหล่านี้มีผลทำให้ความตรงลดลง

5. องค์ประกอบเกี่ยวกับธรรมชาติของกลุ่มและเกณฑ์ โดยปกติความตรงมักจะมีค่าสูงเฉพาะกลุ่ม แต่เนื่องจากในการสอบวัด อาจมีอิทธิพลมาจากองค์ประกอบอื่น ๆ เช่น อายุ เพศ ระดับความสามารถ ภูมิหลังทางการศึกษาและวัฒนธรรม ดังนั้นในการพิจารณาความตรงของแบบสอบ ควรคำนึงถึงธรรมชาติของกลุ่มและเกณฑ์ที่ใช้หาความตรงของแบบสอบด้วย

ตอนที่ 2 มโนทัศน์เกี่ยวกับวิธีการให้น้ำหนักคะแนนตามวิธีของกิบบอนส์และคณะกับวิธีของลอร์ด

ในขณะที่การให้น้ำหนักคะแนนโดยทั่วไป ที่ปฏิบัติกันอยู่ในปัจจุบันในแบบสอบเลือกตอบ (Multiple Choice) ซึ่งเป็นแบบสอบที่นิยมใช้กันอย่างแพร่หลายมากที่สุด คือ วิธีการให้คะแนนแบบธรรมดา คือการให้คะแนนข้อที่ถูกเป็น 1 กับข้อที่ผู้ตอบตอบถูก และให้คะแนนเป็น 0 กับข้อที่ผู้ตอบตอบผิดหรือเว้นว่างเอาไว้ การคิดคะแนนรวมสำหรับผู้เข้าสอบแต่ละคน จะนับจากผลรวมของข้อกระทงที่ตอบถูก ซึ่งวิธีการนี้พบว่ามีความสะดวกในการตรวจให้คะแนน แต่ในสภาพตามเป็นจริงพบว่า พฤติกรรมส่วนหนึ่งของผู้สอบ คือ การเดา จึงอาจไม่เป็นการยุติธรรมในกรณีที่ผู้ไม่รู้แล้วเดาคำตอบได้ถูกต้อง จะได้ค่าคะแนนเท่า ๆ กับผู้ที่มีความรู้จริงที่ตอบถูก และจะมีผลทำให้ค่าความเที่ยงและความตรงของแบบสอบมีค่าลดลง นักวัดผลการศึกษาหลายท่านได้พยายามทำการศึกษา และนำเทคนิคการให้คะแนนด้วยวิธีต่าง ๆ กันมาใช้ ซึ่งในที่นี้จะกล่าวถึงเทคนิควิธีการที่เกี่ยวข้องกับการวิจัยครั้งนี้เท่านั้น

2.1 วิธีการให้น้ำหนักคะแนนตามวิธีของกิบบอนส์และคณะ

กิบบอนส์ โอลคิน และโซเบล (Gibbons Olkins and Sobel, 1979) ได้เสนอวิธีการตอบและวิธีการตรวจให้คะแนนที่เรียกว่า Subset Selection Technique (SST) ในแบบสอบเลือกตอบที่มีตัวเลือกถูกเพียงตัวเลือกเดียว โดยผู้สอบสามารถเลือกตัวเลือกได้ตั้งแต่ 1 ตัว จนกระทั่งถึง k -1 ตัวเลือก ซึ่งผู้สอบคาดว่าจะมีคำตอบถูก หรือตัวเลือกที่ถูกรวมอยู่ด้วย เมื่อ k คือ จำนวนตัวเลือกทั้งหมดในแต่ละข้อ การจะได้คะแนนของผู้ตอบจะขึ้นกับจำนวนตัวเลือกที่ผู้สอบเลือกยิ่งจำนวนตัวเลือกเพิ่มขึ้น ค่าของคะแนนจะลดลง ส่วนวิธีการตรวจให้คะแนน ยึดถือหลักว่า คะแนนคาดหวัง (Expected Score) ที่ได้จากการเดาสุ่มในแต่ละข้อคำถามเป็นศูนย์ เช่น ในกรณีที่แบบสอบเลือกตอบเป็นแบบสอบ 5 ตัวเลือก และผู้ตอบตอบอย่างสุ่มมา 2 ตัวเลือก ความน่าจะเป็นที่ตัวเลือกทั้ง 2 จะตอบถูก คือ 2 ใน 5 และความน่าจะเป็นที่เขาจะตอบผิดคือ 3 ใน 5 ดังนั้นคะแนนความคาดหวัง คือ

$$E(S) = 3(215) - 2(315) = 0$$

ในกรณีนี้ถ้าตอบถูกหมายถึงมีคำตอบที่ถูกรวมอยู่จะได้ 3 คะแนน แต่ถ้าตอบผิด คือ ไม่มีคำตอบที่ถูกรวมอยู่ด้วยจะได้ -2 คะแนน โดยรูปแบบของการให้คะแนน สามารถแสดงให้เห็นดังตารางต่อไปนี้

ตารางแสดงการให้คะแนนตามวิธีของกิบบอนส์และคณะในแบบสอบเลือกตอบ 5 ตัวตัวเลือก

จำนวนตัวเลือกที่ตอบ	คะแนนเมื่อมีตัวเลือกที่ถูก	คะแนนเมื่อไม่มีตัวเลือกที่ถูก
0	-	0
1	4	-1
2	3	-2
3	2	-3
4	1	-4
5	0	0

จะเห็นได้ว่า ในกรณีที่ผู้ตอบเลือกตอบในเมื่อไม่มีตัวเลือกที่ถูกรวมอยู่ด้วย คะแนนที่ได้จะมีค่าติดลบ วิธีการนี้ กิบบอนส์ (Gibbons 1979, 262) กล่าวว่า เป็นวิธีที่ลดการเดาสุ่มลง และยังสามารถวัดส่วนของความรู้ (Partial Information) ของผู้สอบในเรื่องนั้น ๆ ได้อีกด้วย

นอกจากนี้ กิบบอนส์ และคณะ ยังได้นำวิธีการ Subset Selection Technique นี้ มาใช้ในแบบสอบเลือกตอบที่มีคำตอบที่ถูกหลายตัว (Multiple Answer) โดยพบว่า วิธีการนี้มีประสิทธิภาพสูงกว่าวิธีการให้คะแนนแบบธรรมดาตามประเพณีนิยม ไม่ว่าผู้ตอบจะเลือกตอบด้วยจำนวนตัวเลือกหลายตัวเลือกก็ตาม

2.2 วิธีการให้น้ำหนักคะแนนของลอร์ด

ในขณะที่ทฤษฎีการวัดแบบมาตรฐานดั้งเดิม (Classical Test Theory) ถูกพบว่ามีจุดอ่อนในบางประการ เช่น ค่าพารามิเตอร์ที่แปรเปลี่ยนไปตามกลุ่มผู้สอบ การเปรียบเทียบความสามารถของผู้สอบต้องใช้ข้อกระทงชุดเดียวกันหรือแบบสอบที่เป็นคู่ขนานกัน และค่าความคลาดเคลื่อนมาตรฐานในการวัดแต่ละบุคคลมีค่าเท่ากัน ด้วยเหตุเหล่านี้จึงเกิดทฤษฎีวัดผลแนวใหม่ที่กล่าวถึงความสัมพันธ์ระหว่างความสามารถที่แท้จริงกับลักษณะการตอบข้อสอบ ซึ่งเรียกว่า ทฤษฎีสนองตอบข้อสอบ (Item Response Theory IRT) ผู้ที่ริเริ่มคือ Ferguson และ Lawley ในปี ค.ศ. 1942 และ 1943 โดยอาศัยหลักการว่า ผลการสอบของผู้สอบแต่ละคนในการทำแบบสอบใด ๆ จะขึ้นอยู่กับคุณลักษณะแฝง (Latent Trait) หรือความสามารถ (Ability) หรือ ทักษะ (Skill) ของผู้สอบ ต่อมาในปี ค.ศ. 1952 Lord ได้เสนอทฤษฎีนี้ โดยเรียกทฤษฎีดังกล่าวว่า ทฤษฎีโค้งคุณลักษณะของข้อสอบ (Item Characteristic Curve Theory) ต่อมาเรียกว่า Normal Ogive Model ซึ่งเขาไม่ได้พัฒนาต่อเพราะโมเดลนี้มีการคำนวณค่อนข้างยุ่งยาก

ต่อมาในปี ค.ศ. 1960 Rasch ได้เสนอโมเดลหนึ่งของทฤษฎี IRT เรียกว่า Rasch Model เป็นโมเดลที่มีพารามิเตอร์ 1 ตัว คือ ค่าความยากต่อมาในปี ค.ศ. 1965 Lord ได้พัฒนาทฤษฎี IRT อีกครั้ง เนื่องจากเขาพบว่า ปัญหาที่พบจากการศึกษาในครั้งก่อนไม่เป็นปัญหาที่แท้จริง ในปี ค.ศ. 1968 Birnbaum ได้เสนอแนวคิดเกี่ยวกับ Logistic Model และ Normal Ogive Model ที่มี 2 และ 3 พารามิเตอร์ และในปี ค.ศ. 1977 Lord ได้เปลี่ยนชื่อ Item Characleristic Curve Theory มาเป็น Item Response Theory และใช้ Logistic Function แทน Normal Ogive Function เนื่องจากวิธีการคำนวณทางคณิตศาสตร์สะดวกกว่า (Warm, 1978) ปัจจุบันที่นิยมใช้กันมี 3 Model คือ

1. One-Parameter Logistic Model เป็นโมเดลที่มีพารามิเตอร์ 1 ตัว คือ ค่าความยากของแบบสอบ (b_i) โดยมีความเชื่อว่า โอกาสที่ผู้สอบจะทำข้อสอบได้ถูกหรือไม่นั้น ขึ้นอยู่กับระดับความสามารถ (θ) ของตนเองกับระดับความยากของข้อสอบ โดยถือว่า ค่าการเดาของข้อสอบ (c_i) เท่ากับ 0 และค่าอำนาจจำแนกของข้อสอบ (a_i) ทุกข้อมีค่าเท่ากัน สูตรในการคำนวณเป็น

$$P_i(\theta) = \frac{e^{(\theta - b_i)}}{1 + e^{(\theta - b_i)}} \quad ; i=1,2,3,\ldots,n$$

เมื่อ $P_i(\theta)$ คือ โอกาสที่ผู้สอบซึ่งมีความสามารถ (θ) จะทำข้อสอบข้อที่ i ได้ถูกต้อง

θ คือ ระดับความสามารถของผู้สอบ

b_i คือ ค่าความยากข้อสอบ

e คือ ค่าคงที่ มีค่าเท่ากับ 2.71888

2. Two - Parameter Logistic Model เป็นโมเดลที่กำหนดให้ค่าการเดาของข้อสอบ (c_i) เป็น 0 รูปของสมการจะแสดงค่าพารามิเตอร์ 2 ตัว คือ ค่าอำนาจจำแนกของข้อสอบ (a_i) และค่าความยากของข้อสอบ (b_i)

$$P_i(\theta) = \frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}} \quad ; i=1,2,3,\ldots n$$

เมื่อ a_i คือ ค่าอำนาจจำแนกข้อสอบข้อที่ i

D คือ Scaling Factor มีค่า 1.7 เป็นค่าที่มากที่สุดในการปรับ Logistic Model และ Normal Ogive Model ให้สอดคล้องกัน

3. Three - Parameter Logistic Model เป็นโมเดลที่พัฒนามาจาก Two - Parameter Logistic Model เพื่อให้เหมาะกับสถานการณ์ของการสอบที่มีอิทธิพลของการเดาเข้ามาเกี่ยวข้อง รูปสมการ คือ

$$P_i(\theta) = c_i + (1-c_i)\frac{e^{Da_i(\theta - b_i)}}{1+e^{Da_i(\theta - b_i)}} \quad ; i = 1,2,3,\ldots,n$$

เมื่อ C_i คือ ค่าการเดาของข้อสอบข้อที่ i

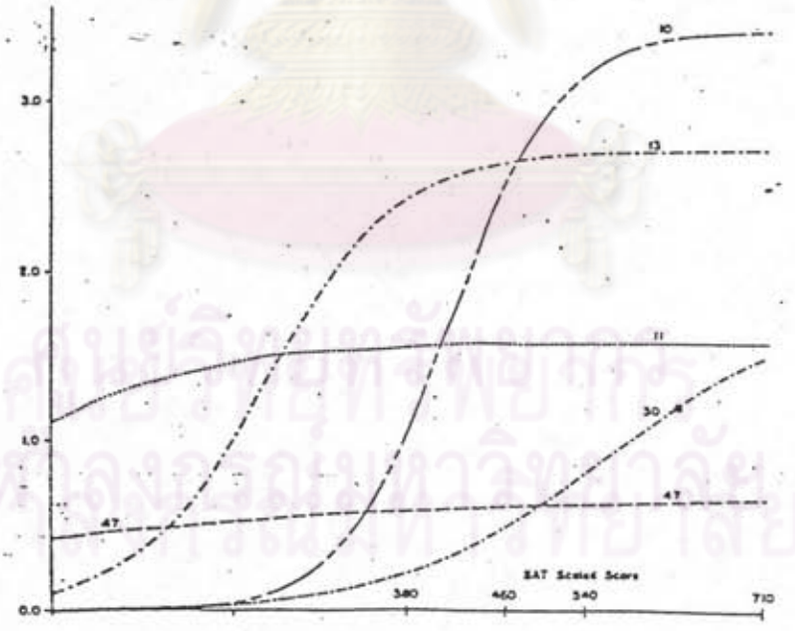
ภายใต้ทฤษฎีสนองตอบข้อสอบ (IRT) โมเดลที่มี 3 พารามิเตอร์ (Three - Parameter Logistic Model) ได้กำหนดสูตรซึ่งใช้คำนวณค่าน้ำหนักคะแนนรายข้อที่ต่างกัน โดยอาศัยค่าพารามิเตอร์ทั้ง 3 (Optimal Scoring Weight) คือ

$$W_i(\theta) = \frac{Da_i}{1+c_i e^{-DL_i}}$$

เมื่อ $L_i = a_i(\theta - b_i)$

<u>แผนภาพที่ 1</u> โค้งแสดงค่าน้ำหนักคะแนนรายข้อ โมเดลที่มี 3 พารามิเตอร์

ค่าน้ำหนักคะแนนที่เหมาะสม



ความสามารถ

จากโค้งแสดงค่าน้ำหนักคะแนนที่เหมาะสมรายข้อ พบว่าเมื่อความสามารถมีค่าเพิ่มขึ้น เส้นโค้งจะขนานกับแกนนอน นั่นคือ ที่ความสามารถสูง ค่าน้ำหนักคะแนนจะเป็นอิสระจากค่าความสามารถแต่จะเป็นสัดส่วนกับค่าอำนาจจำแนก และเมื่อค่าความสามารถลดลง ค่าน้ำหนักคะแนนจะลดลงด้วย ที่เป็นเช่นนี้เพราะที่ความสามารถต่ำจะมีค่าการเดาเกิดขึ้นและค่าการเดาจะทำให้เส้นโค้งน้ำหนักลดต่ำลง ยิ่งความสามารถมีค่าต่ำลงมากขึ้น ค่าน้ำหนักคะแนนจะมีค่าใกล้ศูนย์ เนื่องจากค่าการเดามีค่ามากขึ้น

อย่างไรก็ดี การให้คะแนนแก่ผู้สอบจะไม่ถูกต้อง ถ้าไม่ทราบค่าความสามารถ (θ) ของผู้สอบ Lord (1980) จึงได้เสนอสูตรในการคำนวณค่าน้ำหนักคะแนนที่เหมาะสมเป็นรายข้อ โดยใช้ค่าความยาก (Difficulty) จากการคำนวณโดยวิธีประเพณีนิยมขึ้นมาแทน และค่า D ซึ่งเป็นค่าคงที่ 1.7 นั้นตัดทิ้งไป สูตรในการคำนวณน้ำหนักคะแนนที่เหมาะสมของลอร์ด คือ

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

เมื่อ W_i แทนค่าน้ำหนักคะแนนรายข้อที่เหมาะสม

a_i แทนค่าอำนาจจำแนกของข้อสอบข้อที่ i

p_i แทนค่าความยากของข้อสอบข้อที่วิเคราะห์โดยวิธีประเพณีนิยมตามทฤษฎีมาตรฐานดั้งเดิม

c_i แทนค่า การเดาของข้อสอบข้อที่ i

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ตอนที่ 3 งานวิจัยที่เกี่ยวกับวิธีการให้คะแนนที่ต่างกัน

3.1 งานวิจัยในต่างประเทศ

คูมบส์ (Coombs) เสนอให้เลือกคำตอบที่ผิดแทนการเลือกคำตอบที่ถูก โดยเชื่อว่าการเลือกคำตอบผิดที่มีหลายตัวเลือกในข้อสอบข้อหนึ่ง ผู้สอบจะต้องใช้ส่วนของความรู้ย่อยมาพิจารณา การให้คะแนน ถ้านักเรียนทำเครื่องหมายตรงตัวเลือกที่ผิดจริงจะได้

1 คะแนน แต่ถ้าทำเครื่องหมายตรงตัวเลือกที่ถูก ตัวเลือกนั้นจะได้ 1 - k คะแนน (เมื่อ k คือ ตัวเลือกทั้งหมด) คูมบส์ มิลฮอลแลนด์ และ โวเมอร์ (Commbs, Milholland and Womer, 1956) ได้นำวิธีการตอบและให้คะแนนของคูมบส์มาศึกษาเปรียบเทียบกับวิธีธรรมดา ในแบบสอบ 3 ฉบับ ๆ ละ 40 ข้อ คือ แบบสอบเกี่ยวกับคำศัพท์แบบสอบเกี่ยวกับความรู้นักขับรถ และแบบสอบเกี่ยวกับการใช้สายตาด้านมิติสัมพันธ์ วิธีการตรวจมี 3 วิธี คือ วิธี 0-1 วิธีของคูมบส์ และวิธี 0-1 กับของคูมบส์รวมกัน กลุ่มตัวอย่างเป็นนักศึกษาปีที่ 3 และ 4 ของแจคสันไอสคูล จำนวน 855 คน แบ่งเป็น 3 กลุ่ม ผลการศึกษาปรากฏว่า

1. ค่าความเที่ยงจากวิธีการตรวจของคูมบส์สูงกว่าวิธี 0-1
2. ค่าความตรงของแบบสอบแตกต่างกันอย่างไม่มีนัยสำคัญ
3. วิธีที่คูมบส์เสนอแนะสามารถป้องกันการเดาได้ดีกว่า
4. ทัศนคติของผู้สอบ ชอบวิธีที่คูมบส์เสนอแนะมากกว่า

เดวิส และ ฟิฟเฟอร์ (Davis and Fiffer, 1959) ศึกษาเกี่ยวกับค่าความเที่ยงและความตรงของแบบสอบวัดความถนัด และแบบสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ โดยเปรียบเทียบวิธีการให้คะแนนแบบธรรมดา 0-1 กับวิธีกำหนดน้ำหนักคะแนนแต่ละตัวเลือกไม่เท่ากัน โดยเขามีความคิดว่าโจทย์คณิตศาสตร์นั้น ตัวเลือกแต่ละตัวมีความยากง่ายไม่เท่ากัน ซึ่งจะใช้สมรรถภาพทางสมองต่างกัน กลุ่มตัวอย่างเป็นนักเรียนการบินของ Lackland Air Force Base จำนวน 65 คน ผลการศึกษาพบว่า

1. การให้คะแนนโดยมีน้ำหนักคะแนนแต่ละตัวเลือกไม่เท่ากัน มีค่าความเที่ยงสูงกว่า วิธีการให้คะแนนแบบ 0-1 อย่างมีนัยสำคัญ

2. การให้คะแนนโดยมีน้ำหนักคะแนนแต่ะตัวเลือกไม่เท่ากัน มีค่าความตรงสูงกว่าวิธีการให้คะแนนแบบ 0-1 แต่ไม่มีนัยสำคัญ

แพทเนอิค และทรูป (Patnaik and Traub, 1973) ทดลองศึกษาวิธีตอบและวิธีตรวจที่คล้ายกับวิธีของเดวิสและทิฟเฟอร์ นั่นคือ วิธีกำหนดน้ำหนักคะแนนตัวเลือกต่างกัน กับวิธีธรรมดาแบบ 0-1 โดยในการกำหนดน้ำหนักคะแนนของตัวเลือก ใช้การตัด

สินใจของครูผู้สอนกับผู้เชี่ยวชาญ จำนวน 61 คน จัดอันดับน้ำหนักคะแนนแต่ละตัวเลือกที่ถูกในแต่ละข้อแล้วหาค่าเฉลี่ยออกมาเป็นคะแนนในแต่ละข้อ ผลการศึกษาพบว่า สอดคล้องกับการวิจัยของเดวิส และ ทิฟเฟอร์ นั่นคือ ค่าความเที่ยงที่คำนวณโดยวิธีแบบแบ่งครึ่งของวิธีการให้คะแนนแต่ละตัวเลือกไม่เท่ากันกับวิธีให้คะแนนแบบธรรมดา Ø-1 มีค่า .915 และ .881 ตามลำดับ ส่วนค่าความตรงเชิงพยากรณ์พบว่า วิธีการให้คะแนนแต่ละตัวเลือกไม่เท่ากัน มีค่าความตรงน้อยกว่า วิธีธรรมดาแบบ Ø-1

พัฟ และ ปรันซา (Pugh and Brunza, 1975) ศึกษาวิธีการตอบและการตรวจ โดยบอกความมั่นใจในการตอบ โดยนำเอาน้ำหนักความมั่นใจในการเลือกข้อที่ถูกมาเป็นส่วนในการให้คะแนน เปรียบเทียบกับวิธีการตอบและการตรวจแบบธรรมดา ผลการศึกษาพบว่า ค่าความเที่ยงซึ่งคำนวณโดยวิธีวิเคราะห์ความแปรปรวนของแบบสอบที่มีวิธีตอบโดยระบุความมั่นใจในการสอบสูงกว่า แบบสอบที่ตอบโดยวิธีธรรมดา คือ มีค่า .85 และ .57 ตามลำดับ

อาบู ซาอีฟ และ ไดมอน (Abu - Sayf & Daimond, 1976) ศึกษาเกี่ยวกับค่าความเที่ยงและความตรงของแบบสอบ เมื่อตรวจให้คะแนน 3 วิธี คือ วิธีที่ 1 ให้คะแนนเฉพาะข้อที่ถูกและตอบอย่างมั่นใจ วิธีที่ 2 ให้คะแนนข้อที่ตอบถูกโดยไม่ได้เดา และวิธีที่ 3 ให้คะแนนทุกข้อที่ตอบถูก โดยไม่คำนึงว่ามีความมั่นใจในการตอบถูกหรือไม่เพียงใด แล้วหาค่าความเที่ยงโดยใช้สูตรครอบบัค แอลฟา (Cronbach's Alpha) ผลการศึกษาปรากฏว่า การให้คะแนนเฉพาะข้อที่ถูก และมีความมั่นใจมีค่าความเที่ยงสูงสุด คือ Ø.898 และการให้คะแนนเฉพาะข้อที่ตอบถูกและไม่ได้เดา มีค่าความเที่ยงสูงกว่าการให้คะแนนทุกข้อที่ตอบถูก โดยมีค่าความเที่ยง Ø.871 และ Ø.774 ตามลำดับ สำหรับค่าความตรง ปรากฏว่าวิธีการให้คะแนนเฉพาะข้อที่ตอบถูกและไม่ได้เดามีค่าสูงสุดคือ .260 รองลงมาคือ วิธีการให้คะแนนเฉพาะข้อที่ตอบถูกและมีความมั่นใจ และสุดท้ายคือ วิธีการให้คะแนนทุกข้อที่ตอบถูก คือมีค่าความตรงเป็น Ø.236 และ Ø.219 ตามลำดับ

ชิ ซัง เวน (Shih - Sung Wen, 1975) ศึกษาสหสัมพันธ์ระหว่างคะแนนที่นับจากผลรวมจำนวนข้อที่ตอบถูก กับคะแนนทดสอบความมั่นใจพบว่ามีสหสัมพันธ์สูงถึง .97 โดยในทางปฏิบัติแล้ว คะแนนที่ได้จาก 2 วิธีนี้ สามารถนำมาจัดอันดับ (Rank Order) ได้เช่นกัน และพบว่า การทดสอบความมั่นใจสามารถให้รายละเอียดของข้อมูลในการตอบข้อกระทงได้ดีกว่า

ลอร์ด (Lord, 1980) ได้วิเคราะห์แบบสอบทางภาษาที่ชื่อ SCAT II Form 2A ซึ่งเป็นแบบสอบเลือกตอบ 4 ตัวเลือก จำนวน 50 ข้อ เป็นเครื่องมือ กลุ่มตัวอย่างเป็นนักศึกษาระดับไฮสคูล โดยศึกษาเปรียบเทียบผลของการให้คะแนนข้อสอบแต่ละข้อแตกต่างกัน ลอร์ด พบว่า การให้คะแนนข้อสอบแต่ละข้อ โดยวิธีใช้ค่าน้ำหนักคะแนนที่เหมาะสมโดยสูตรของลอร์ดจะให้สาระของการสอบมากที่สุด และการให้ค่าน้ำหนักคะแนนรายข้อเท่ากันตามวิธีประเพณีนิยม จะให้สาระของการสอบน้อยที่สุด

เต คิ ซู และ ชูศักดิ์ ขัมภลิขิต (Tse - Chi Hsu and Choosak Khampalikit, 1982) ได้ศึกษาเปรียบเทียบผลของการให้คะแนน โดยวิธีให้น้ำหนักคะแนนรายข้อที่เหมาะสม ตามทฤษฎีสนองตอบข้อสอบ (IRT) โมเดลที่มีพารามิเตอร์ 3 ตัว กับวิธีการให้น้ำหนักคะแนนรายข้อเท่ากัน ตามทฤษฎีมาตรฐานเดิม โดยศึกษาข้อมูลกับสภาพการสอบคัดเลือกจริงในแบบสอบภาษาจีน และแบบสอบประวัติศาสตร์ซึ่งเป็นแบบสอบที่เขียนขึ้นมาใหม่และยังไม่ได้มีการทดลองใช้มาก่อน โดยมีสมมติฐานของการวิจัยว่า การให้น้ำหนักคะแนนรายข้อที่เหมาะสมตามค่าพารามิเตอร์ของข้อกระทง น่าจะมีส่วนได้เปรียบกว่าวิธีการให้คะแนนแบบธรรมดาในเรื่องที่ว่า ผลของการให้คะแนนรายข้อที่เหมาะสมนั้น ค่าพารามิเตอร์ของข้อสอบจะขึ้นอยู่กับคุณลักษณะของแบบสอบน้อยกว่าการให้คะแนนแบบอื่น ผลการวิจัย พบว่าข้อมูลจากการสอบวัดในครั้งนี้ สามารถนำมาใช้กับทฤษฎีตอบสนองข้อสอบ (IRT) ได้ และยังเป็นการขจัดปัญหาแบบสอบที่ขาดคุณภาพได้ด้วย นั่นคือ การให้น้ำหนักคะแนนรายข้อที่เหมาะสม จะให้ค่าความเที่ยง (Reliability) ของแบบสอบ และค่าอำนาจจำแนก (Discriminating Power) สูงกว่า การให้น้ำหนักคะแนนรายข้อเท่ากันตามวิธีประเพณีนิยม

ในด้านความสัมพันธ์ของคะแนน พบว่า คะแนนที่ได้จากการให้น้ำหนักคะแนนรายข้อเท่ากับตามวิธีประเพณีนิยม กับคะแนนที่ได้จากการให้น้ำหนักคะแนนที่เหมาะสม (Optimal Scoring Weight Method) โมเดลที่มีพารามิเตอร์ 3 ตัว มีความสอดคล้องกันสูงถึง 87 % ของจำนวนผู้สมัครสอบ 10 % ที่ถูกคัดเลือกโดยคะแนนทั้ง 2 แบบ

จาราดัท และซาเวจ (Jaradat and Savage, 1986) ได้เปรียบเทียบวิธีการให้คะแนน 3 วิธี คือ วิธีธรรมดาแบบ 0-1 วิธีใช้สูตรแก้การเดา และวิธีของกิบบอนส์และคณะ ที่เรียกว่า Subbset Selection Technique โดยศึกษาค่าความเที่ยงและความตรงของแบบสอบ กลุ่มตัวอย่างเป็นนักศึกษาระดับ 9 ของโรงเรียนในประเทศจอร์แดน จำนวน 160 คน เครื่องมือที่ใช้เป็นแบบสอบวิชาเคมี เรื่องทฤษฎีอะตอมและปฏิกิริยาเคมี โดยแบบสอบที่สร้างขึ้นได้สร้างตามกระบวนการสร้างแบบสอบที่มีคุณภาพ มีความตรงเชิงเนื้อหา จากผู้เชี่ยวชาญเนื้อหาวิชา และผู้เชี่ยวชาญด้านการสร้างแบบสอบ (นักวัดผล) จำนวน 6 ท่าน มีคุณภาพรายข้อตามเกณฑ์ที่กำหนด ผลการวิจัย โดยการคำนวณจากกลุ่มตัวอย่างทั้งหมด พบว่า ค่าความเที่ยงตามวิธีการให้คะแนนแบบธรรมดา (0 - 1) อย่างมีนัยสำคัญ แต่สูงกว่าการให้คะแนนโดยใช้สูตรแก้การเดาอย่างไม่มีนัยสำคัญ ในด้านความตรงของแบบสอบ พบว่า วิธีการให้คะแนนวิธีของกิบบอนส์และคณะสูงกว่า วิธีการให้คะแนนแบบธรรมดา (0 - 1) และวิธีใช้สูตรแก้การเดาอย่างมีนัยสำคัญ ทั้ง 2 วิธี

3.2 งานวิจัยในประเทศ

อรวรรณ ตัณฑ์เจริญรัตน์ (2517) ศึกษาเปรียบเทียบค่าความเที่ยงและเปอร์เซนต์การเดาของแบบสอบเลือกตอบ ที่มีวิธีการตอบและตรวจให้คะแนน 3 วิธี คือ วิธี 0 - 1 วิธีของคูมบส์ และวิธีของอนันต์ ศรีโสภา ที่ให้ตอบทุกตัวเลือกว่าตัวเลือกใดถูกหรือตัวเลือกใดผิด เครื่องมือที่ใช้เป็นแบบสอบวิชาภาษาไทยระดับประกาศนียบัตรวิชาการศึกษา กลุ่มตัวอย่างเป็นนักศึกษาระดับประกาศนียบัตรวิชาการศึกษาจำนวน 214 คน แบ่งออกเป็น 3 กลุ่ม แต่ละกลุ่มได้รับแบบทดสอบที่มีวิธีการตอบและตรวจกลุ่มละวิธี ผลการศึกษาพบว่าค่าความเที่ยงของแบบสอบที่มีวิธีตอบและวิธีตรวจตามวิธีของอนันต์ ศรีโสภา สูงกว่าค่าความ

เที่ยงของแบบสอบที่ตรวจให้คะแนน โดยวิธีของคูมบส์ และวิธีธรรมดาแบบ ๐- 1 ให้ค่าความเที่ยงที่สุด เปอร์เซ็นต์การเดาตามวิธีของคูมบส์มีค่าสูงสุด รองลงมาคือ เปอร์เซนต์การเดาตามวิธีของอนันต์ และเปอร์เซนต์การเดาตามวิธีธรรมดา แบบ ๐ - 1 มีค่าต่ำสุด

ธีรศักดิ์ อินทรมาตย์ (2520) เปรียบเทียบค่าความเที่ยงและความตรงและปริมาณการเดา เมื่อมีวิธีการตอบต่างกัน 3 วิธี คือ วิธีตอบแบบธรรมดา วิธีตอบโดยบอกความมั่นใจในการตอบและวิธีตอบทุกตัวเลือก แบบสอบที่ใช้เป็นแบบสอบวิชาวิทยาศาสตร์ กลุ่มตัวอย่างเป็นนักเรียนชั้นประถมศึกษาปีที่ 7 ของโรงเรียนเมืองสมุทรสงคราม จำนวน 216 คน โดยแบ่งกลุ่มตัวอย่างเป็น 3 กลุ่ม ที่มีระดับความสามารถต่างกัน ผลการศึกษาพบว่า ค่าความเที่ยงแตกต่างกันอย่างมีนัยสำคัญที่ระดับ .๐1 ในกลุ่มนักเรียนที่มีความสามารถในวิชาวิทยาศาสตร์ ในระดับสูงและปานกลาง และแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .๐5 ในกลุ่มที่มีความสามารถทางวิชาวิทยาศาสตร์ในระดับต่ำ ส่วนค่าความตรงของแบบสอบพบว่าแตกต่างกันอย่างไม่มีนัยสำคัญในทุกระดับความสามารถของนักเรียน ในด้านปริมาณการเดา พบว่า แบบสอบที่ตอบแบบธรรมดามีปริมาณการเดาสูงสุด รองลงมาคือ แบบสอบที่ตอบโดยบอกความมั่นใจ และแบบสอบที่ตอบทุกตัวเลือกจะมีปริมาณการเดาต่ำสุด และจากการศึกษาเรื่องปริมาณการเดา พบว่า นักเรียนที่มีความสามารถในวิชาวิทยาศาสตร์ในระดับต่ำจะตอบแบบสอบโดยการเดามากกว่า นักเรียนที่มีความสามารถในวิชาวิทยาศาสตร์ระดับสูงในทุกวิธีการตอบ

เพ็ญศรี สว่างเนตร (2520) ศึกษาเปรียบเทียบค่าความเที่ยงของแบบสอบเลือกตอบ เมื่อมีวิธีการให้คะแนนต่างกัน 4 วิธี คือ วิธีที่ 1 แบบธรรมดา (๐ - 1) วิธีที่ 2 วิธีลดคะแนนข้อที่ตอบผิด โดยมีสมการในการให้คะแนน คือ

$$\text{คะแนนรวม} = \text{จำนวนข้อที่ตอบถูก} - \frac{\text{จำนวนข้อที่ตอบผิด}}{\text{จำนวนตัวเลือก} - 1}$$

วิธีที่ 3 วิธีเพิ่มคะแนนข้อที่เว้นไว้ โดยมีสมการในการให้คะแนน คือ

$$\text{คะแนนรวม} = \text{จำนวนข้อที่ตอบถูก} + \frac{\text{จำนวนข้อที่เว้น}}{\text{จำนวนตัวเลือก}}$$

และวิธีที่ 4 เป็นวิธีทดสอบความมั่นใจ เครื่องมือที่ใช้เป็นแบบสอบเลือกตอบ 5 ตัวเลือก วิชาภาษาอังกฤษ กลุ่มตัวอย่างเป็นนักศึกษาประกาศนียบัตรการศึกษาชั้นสูง จำนวน 218 คน ผลการวิจัยพบว่า ค่าความเที่ยงเมื่อให้คะแนนทดสอบค่าความมั่นใจสูงกว่า วิธีให้คะแนนแบบธรรมดา (0 - 1) อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ส่วนค่าความเที่ยงของวิธีให้คะแนนแบบธรรมดา (0 - 1) วิธีลดคะแนนข้อที่ตอบผิด กับวิธีการเพิ่มคะแนนข้อที่เว้นไว้มีค่าไม่แตกต่างกันอย่างมีนัยสำคัญ และพบว่าวิธีการทดสอบความมั่นใจให้ค่าความเที่ยงสูงที่สุด

กาญจนา ศิริวัฒนพงศ์ (2520) ศึกษาเปรียบเทียบค่าความเที่ยง ความตรง อำนาจจำแนก และเปอร์เซนต์การเดาของแบบสอบเลือกตอบ 3 แบบ คือ ตัวเลือกถูกตัวเดียว ตัวเลือกถูกที่สุดเพียงตัวเดียว และตัวเลือกถูกหลายตัว เมื่อมีวิธีตอบและวิธีตรวจให้คะแนนต่างกัน 3 วิธี คือ วิธี 0 - 1 วิธีของคูมบส์ และวิธีของอนันต์ เครื่องมือที่ใช้เป็นแบบสอบเลือกตอบวิชาวิทยาศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 3 จำนวน 810 คน โดยแบ่งเป็น 9 กลุ่ม ๆ ละ 90 คน เพื่อทำแบบสอบทั้ง 3 แบบด้วยวิธีการตอบและการตรวจทั้ง 3 วิธี ผลการศึกษาพบว่า

1. ค่าความตรง และค่าอำนาจจำแนกของแบบสอบ เมื่อตรวจให้คะแนนโดยวิธีธรรมดาแบบ 0 - 1 และวิธีของคูมบส์แตกต่างกันอย่างไม่มีนัยสำคัญของแบบสอบทั้ง 3 แบบ

2. การตรวจให้คะแนนด้วยวิธีของอนันต์ ค่าความเที่ยงของแบบสอบมีตัวเลือกถูกตัวเดียวสูงกว่า แบบสอบที่มีตัวเลือกที่ถูกที่สุดเพียงตัวเดียว และแบบสอบตัวเลือกถูกหลายตัว ส่วนค่าความตรงและค่าอำนาจจำแนกไม่พบว่าแตกต่างกัน

3. การตรวจให้คะแนนโดยวิธีของอนันต์ ในแบบสอบที่มีตัวเลือกถูกตัวเดียว และแบบสอบตัวเลือกที่ถูกที่สุดเพียงตัวเดียว ให้ค่าความเที่ยงและความตรงสูงกว่า วิธี ๐ - 1 และวิธีของคูมบส์ ส่วนค่าอำนาจจำแนกไม่แตกต่างกัน

4. แบบสอบที่มีตัวเลือกถูกหลายตัว เมื่อตรวจให้คะแนนทั้ง 3 วิธี ให้ค่าความเที่ยง ความตรง และค่าอำนาจจำแนกไม่แตแต่างกัน

สำราญ มีแจ้ง (2525) ได้เสนอวิธีการตอบและการตรวจให้คะแนนในแบบสอบเลือกตอบ 5 ตัวเลือก โดยผู้ตอบสามารถเลือกตอบได้ตั้งแต่ 1 ถึง 3 ตัวเลือกตามความมั่นใจ ถ้ามั่นใจตัวเลือกใดก็เลือกตอบตัวเลือกนั้น ถ้าไม่มั่นใจ 2 หรือ 3 ตัวเลือก ก็ให้ตอบทั้ง 2 หรือ 3 ตัวเลือกนั้น ถ้าตอบตัวเลือกเดียวแล้วถูกจะได้ 4 คะแนน ถ้าตอบ 2 ตัวเลือกแล้วถูกจะได้ 2 คะแนน ถ้าตอบ 3 ตัวเลือกแล้วถูกจะได้ 1 คะแนน ในทุกกรณีถ้าตอบผิดจะได้ ๐ คะแนน จากนั้น สำราญได้นำวิธีการของเขามาศึกษาเปรียบเทียบกับวิธีการให้คะแนนแบบอื่น ๆ อีก 3 วิธี คือ วิธี ๐ - 1 วิธีของคูมบส์ และวิธีของอนันต์ ศรีโสภา โดยศึกษาเปรียบเทียบในเรื่องความเที่ยง ความตรง และอำนาจจำแนกของแบบสอบ เครื่องมือที่ใช้เป็นแบบสอบเลือกตอบ 5 ตัวเลือกวิชาคณิตศาสตร์ จำนวน 60 ข้อ กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 4 โรงเรียนสุรศักดิ์มนตรี และโรงเรียนวัดน้อยใน จำนวน 4 กลุ่ม ๆ ละ 60 คน แต่ละกลุ่มได้รับการทดสอบจากแบบสอบที่มีวิธีการตอบและตรวจกลุ่มละวิธี ผลการวิจัยปรากฏว่า ค่าความเที่ยงของแบบสอบที่ใช้คำสั่ง และวิธีการให้คะแนนตามวิธีของสำราญ มีแจ้ง สูงกว่าวิธี ๐ - 1 ค่าความเที่ยงของแบบสอบที่ใช้คำสั่งและวิธีการให้คะแนนของคูมบส์ สูงกว่าวิธี ๐ - 1 ส่วนค่าความตรง พบว่า ความตรงที่ตรวจและให้คะแนนโดยวิธีของอนันต์ ศรีโสภา สูงกว่าความตรงที่ตรวจให้คะแนนโดยวิธี ๐ - 1 ส่วนค่าอำนาจจำแนกของแบบสอบที่ใช้คำสั่ง และวิธีการตรวจให้คะแนนทั้ง 4 วิธี ไม่แตกต่างกันอย่างมีนัยสำคัญ

ทวี ทองคำ (2526) ศึกษาเปรียบเทียบคำสั่งและวิธีการให้คะแนนต่างกัน 3 วิธี วิธีที่ 1 คือ วิธี ๐ - 1 วิธีที่ 2 คือวิธีของสำราญ และวิธีที่ 3 คือวิธีของกิบบอนส์และคณะ ต่อค่าความเที่ยง ความตรง ค่าอำนาจแนกของแบบสอบเลือกตอบ

เครื่องมือที่ใช้เป็นแบบสอบเลือกตอบ 5 ตัวเลือก จำนวน 60 ข้อวิชา คณิตศาสตร์พาณิชยกรรม 2 กลุ่มตัวอย่างเป็นนักศึกษาระดับประกาศนียบัตรวิชาชีพชั้นปีที่ 1 ของวิทยาลัยเทคนิคยะลา จำนวน 180 คน โดยแบ่งกลุ่มตัวอย่างเป็น 3 กลุ่ม แต่ละกลุ่มได้รับแบบสอบที่มีวิธีการตอบและการตรวจในแต่ละวิธี ผลการวิจัยพบว่า

1. ค่าความเที่ยงของแบบสอบ เมื่อมีวิธีการตรวจให้คะแนนที่ต่างกัน 3 วิธี ไม่แตกต่างกันที่ระดับความนัยสำคัญ .05 โดยวิธีการให้คะแนนตามวิธีของสำราญ ให้ค่าความเที่ยงสูงสุด ส่วนวิธีการให้คะแนนของกิบบอนส์และคณะมีค่าต่ำที่สุด

2. ค่าความตรงตามสภาพของแบบสอบ โดยใช้คะแนนผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ 1 เป็นเกณฑ์ พบว่า ไม่แตกต่างกันอย่างมีนัยสำคัญที่ระดับ .05 ทั้ง 3 วิธี อย่างไรก็ดี พบว่าวิธีการให้คะแนนตามวิธีของกิบบอนส์และคณะ ให้ค่าความตรงตามสภาพที่สูงที่สุด

3. ค่าความตรงตามสภาพเมื่อให้ครูผู้สอนเป็นผู้จัดกลุ่มความสามารถเป็น 3 กลุ่ม พบว่า วิธีการให้คะแนนทั้ง 3 วิธี ให้ค่าความตรงตามสภาพที่ระดับ .05 ทั้ง 3 วิธี และพบว่า แบบสอบที่ใช้คำสั่งและวิธีการตรวจตามวิธีของสำราญ และวิธีการวิธีของกิบบอนส์และคณะ สามารถจำแนกนักเรียนระหว่างกลุ่มได้ทุกกลุ่ม

4. ค่าอำนาจจำแนกของแบบสอบ เมื่อใช้คำสั่งและวิธีการตรวจให้คะแนนต่างกัน 3 วิธี แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

นภา กาญจนกิจโสภณ (2528) เปรียบเทียบคะแนนสอบระหว่างวิธีการให้น้ำหนักคะแนนรายข้อเท่ากันตามทฤษฎีมาตรฐานเดิม กับ 1 วิธีการให้น้ำหนักคะแนนรายข้อต่างกันตามทฤษฎีสนองตอบข้อสอบ โมเดลที่มีพารามิเตอร์ 3 ตัว โดยศึกษาข้อมูลจากคะแนนสอบในวิชาคณิตศาสตร์ วิทยาศาสตร์ และพลานามัย ซึ่งเป็นแบบสอบของโครงการตรวจสอบคุณภาพทางการศึกษา กรมวิชาการ กระทรวงศึกษาธิการ ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2526 ได้จำนวนกลุ่มตัวอย่างทั้งหมดวิชาละ 3,907 คน โดยได้นำข้อมูลเดิม ซึ่งตรวจให้คะแนนแบบให้น้ำหนักรายข้อเท่ากัน มาตรวจให้คะแนน เมื่อให้น้ำหนักรายข้อต่างกัน โดยใช้โมเดลที่มีพารามิเตอร์ 3 ตัว และคำนวณค่าน้ำหนักคะแนนในแต่ละระดับความสามารถจากสูตร

$$W_i(\theta) = \frac{Da_i}{1+c_i e^{-b\theta_i}} \quad , i = 1,2,3,\ldots n$$

แล้วนำค่าคะแนนที่ได้ไปเปรียบเทียบกับคะแนนที่ได้จากการตรวจให้คะแนนน้ำหนักรายข้อเท่ากัน ผลการวิจัยพบว่า

1. ค่ามัชฌิมเลขคณิต และส่วนเบี่ยงเบนมาตรฐานของคะแนนสอบ โดยวิธีให้น้ำหนักคะแนนรายข้อเท่ากัน แตกต่างกับคะแนนสอบที่ได้จากวิธีให้น้ำหนักคะแนนรายข้อต่างกันอย่างมีนัยสำคัญ ($p < .05$) ทั้ง 3 วิชา

2. ความสัมพันธ์ของคะแนนสอบระหว่างการตรวจให้คะแนนทั้ง 2 วิธี ในวิชาคณิตศาสตร์ เป็น .937 วิชาวิทยาศาสตร์มีค่า .929 และวิชาพลานามัยมีค่า .897 พบว่ามีนัยสำคัญ ($p < .01$) ทั้ง 3 วิชา

3. ค่าความเที่ยงของแบบสอบวิชาคณิตศาสตร์ วิทยาศาสตร์ และพลานามัยของวิธีการให้น้ำหนักคะแนนรายข้อต่างกันตามทฤษฎีสนองตอบข้อสอบสูงกว่า ค่าความเที่ยงที่ให้คะแนนโดยวิธีให้น้ำหนักคะแนนรายข้อเท่ากัน อย่างมีนัยสำคัญทางสถิติที่ 3 วิชา

สุพัฒน์ สุกมลสันต์ (2528) เปรียบเทียบวิธีการให้นำหนักคะแนนที่ต่างกัน 5 วิธี คือ วิธีที่ 1 วิธีใช้ค่าเดลต้า (Δ) วิธีที่ 2 วิธีใช้ค่าเบต้าแปลง (β_t) วิธีที่ 3 วิธีใช้ค่าน้ำหนักคะแนนที่เหมาะสม (w) จากสูตรของลอร์ด วิธีที่ 4 วิธีใช้ค่าผลบวกของความแปรปรวนร่วม (h^2) และวิธีที่ 5 วิธีประเพณีนิยม เครื่องมือที่ใช้เป็นแบบสอบการอ่านเข้าใจความภาษาอังกฤษ ที่มีโครงสร้างต่างกัน 2 ฉบับ คือ ฉบับที่ 1 เป็นฉบับที่ไม่มีชื่อเรื่องและภาพประกอบ ฉบับที่ 2 เป็นฉบับที่มีชื่อเรื่องและภาพ กลุ่มตัวอย่างเป็นนิสิตชั้นปีที่ 1 ที่เรียนวิชาภาษาอังกฤษพื้นฐานเฉพาะการอ่าน 2 ของจุฬาลงกรณ์มหาวิทยาลัย จำนวน 1,872 คน ผลการวิจัยพบว่า

1. ค่าความตรงเชิงพยากรณ์ของแบบสอบอิงกลุ่ม

1.1 ค่าความตรงเชิงพยากรณ์ของแบบสอบที่ไม่มีชื่อเรื่องและภาพประกอบ พบว่าวิธีการให้คะแนนโดยใช้ผลบวกความแปรปรวนร่วม (h^2) จะมีค่าสูงสุด

และสูงกว่าค่าความตรงเชิงพยากรณ์ของแบบสอบที่ให้คะแนนโดยวิธีใช้ค่าน้ำหนักคะแนนที่เหมาะสม (w) ซึ่งมีค่าต่ำสุดอย่างมีนัยสำคัญ

1.2 ในแบบสอบที่มีชื่อเรื่องและภาพประกอบ พบว่า ค่าความตรงเชิงพยากรณ์ที่ให้ค่าน้ำหนักคะแนนที่เหมาะสม (w) มีค่าสูงสุด และสูงกว่าวิธีเบต้าแปลง ซึ่งมีค่าต่ำสุดอย่างมีนัยสำคัญ

1.3 ค่าความตรงเชิงพยากรณ์ของแบบสอบทั้ง 2 แบบ ที่ให้คะแนนโดยวิธีต่าง ๆ ที่ลำดับที่ต่างกัน

2. ค่าความเที่ยงของแบบสอบอิงกลุ่ม

2.1 ในแบบสอบที่ไม่มีชื่อเรื่องและภาพประกอบ ค่าความเที่ยงที่ให้คะแนนโดยวิธีให้น้ำหนักคะแนนที่เหมาะสม (w) มีค่าสูงสุด และสูงกว่าค่าความเที่ยงที่ให้คะแนนโดยวิธีอื่น ๆ อย่างมีนัยสำคัญ โดยวิธีใช้ค่าเบต้าแปลง (β_t) จะมีค่าต่ำสุด

2.2 ในแบบสอบที่มีชื่อเรื่องและภาพประกอบ ค่าความเที่ยงที่ให้คะแนนโดยวิธีให้น้ำหนักคะแนนที่เหมาะสม (w) มีค่าสูงสุด โดยสูงกว่าค่าความเที่ยงที่ให้คะแนนโดยวิธีอื่น ๆ อย่างมีนัยสำคัญ โดยค่าความเที่ยงที่ใช้วิธีเบต้าแปลง ($\beta_{t.}$) จะมีค่าต่ำสุด

2.3 ค่าลำดับที่ของค่าความเที่ยงของแบบสอบทั้ง 2 ที่ให้คะแนนโดยวิธีต่าง ๆ เหมือนกันทั้ง 2 แบบสอบ

3. ค่าความเที่ยงของแบบสอบอิงเกณฑ์ เมื่อจุดตัดอยู่ระหว่างร้อยละ 0 ถึง 100 พบว่า วิธีการให้คะแนนตามวิธีประเพณีนิยม มีค่าสูงสุด และวิธีให้คะแนนโดยใช้ค่าน้ำหนักคะแนนที่เหมาะสมมีค่าต่ำสุด ทั้งใน 2 แบบสอบ แต่ถ้าจุดตัดอยู่ระหว่างร้อยละ 50 ถึง 100 วิธีการให้คะแนนโดยใช้ค่าเบต้าแปลง จะมีค่าความเที่ยงสูงสุด และวิธีให้น้ำหนักคะแนนที่เหมาะสมให้ค่าต่ำสุด ทั้งใน 2 แบบสอบเช่นกัน และค่าลำดับที่ของแบบสอบทั้ง 2 เหมือนกันทั้ง 2 กรณี

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

I19003454

4. ความคงที่ของอันดับที่ของคะแนน พบว่า อันดับที่ของคะแนนของแบบสอบที่ให้คะแนนโดยวิธีใช้ค่าผลบวกความแปรปรวนร่วม (h^2) มีความคงที่มากที่สุดทั้งในแบบสอบ 2 ฉบับ และสำหรับแบบสอบที่ไม่มีชื่อเรื่องและภาพประกอบวิธีการให้น้ำหนักคะแนนที่เหมาะสมมีค่าความคงที่ต่ำสุด ส่วนแบบสอบที่มีชื่อเรื่องและภาพประกอบวิธีใช้ค่าเบต้าแปลง มีค่าต่ำสุด

5. ค่าความสัมพันธ์ของคะแนน พบว่าวิธีการให้คะแนนด้วยวิธีต่าง ๆ มีความสัมพันธ์กันอย่างมีนัยสำคัญ ทั้งใน 2 แบบสอบ

พนิชา สวัสดีมงคล (2530) ศึกษาเปรียบเทียบค่าความเที่ยง ความตรงของแบบสอบเลือกตอบที่วัดระดับความรู้ขั้นต่ำ และขั้นสูง โดยมีวิธีการให้คะแนนต่างกัน 2 วิธี คือ วิธีธรรมดา และวิธีการให้คะแนนความมั่นใจ ในกลุ่มผู้สอบที่มีผลสัมฤทธิ์ทางการเรียนต่างกัน คือ สูง ปานกลาง และต่ำ เครื่องมือที่ใช้คือ แบบสอบวิชาการพยาบาลผู้ป่วยอายุรกรรมและศัลยกรรม จำนวน 60 ข้อ กลุ่มตัวอย่างคือนักเรียนพยาบาลที่เรียนวิชาดังกล่าว

ผลการวิจัยพบว่า ด้านความเที่ยง ความตรงของแบบสอบในกลุ่มผู้สอบที่มีผลสัมฤทธิ์ทางการเรียนสูง ปานกลางและต่ำ 1) แบบสอบที่วัดระดับความรู้ขั้นต่ำและมีวิธีให้คะแนนวิธีธรรมดา ค่าความเที่ยงของแบบสอบระหว่างกลุ่มต่ำกับกลุ่มปานกลางและสูง มีค่าแตกต่างกันอย่างมีนัยสำคัญที่ระดับ .01 ระหว่างกลุ่มสูงกับปานกลางไม่แตกต่างกัน ค่าความตรงของแบบสอบที่ได้จากทั้ง 3 กลุ่ม มีค่าไม่แตกต่างกัน 2) แบบสอบที่วัดระดับความรู้ขั้นสูงและมีวิธีให้คะแนนวิธีธรรมดา ค่าความเที่ยง ความตรงของแบบสอบที่ได้จากทั้ง 3 กลุ่มมีค่าไม่แตกต่างกันอย่างมีนัยสำคัญ 3) แบบสอบที่วัดระดับความรู้ขั้นต่ำและมีวิธีให้คะแนนความมั่นใจ ค่าความเที่ยงของแบบสอบที่ได้จากทั้ง 3 กลุ่ม มีค่าไม่แตกต่างกันอย่างมีนัยสำคัญ ค่าความตรงของแบบสอบระหว่างกลุ่มสูงกับต่ำมีค่าแตกต่างกันอย่างมีนัยสำคัญที่ระดับ .05 ระหว่างกลุ่มปานกลางกับกลุ่มสูงและต่ำมีค่าไม่แตกต่างกัน 4) แบบสอบที่วัดระดับความรู้ขั้นสูงและมีวิธีให้คะแนนความมั่นใจ ค่าความเที่ยงของแบบสอบระหว่างกลุ่มสูงกับกลุ่มต่ำ มีค่าแตกต่างกันอย่างมีนัยสำคัญ .05 ระหว่างกลุ่มปานกลางกับสูงและต่ำมีค่าไม่แตกต่างกัน ค่าความตรงของแบบสอบที่ได้จากทั้ง 3 กลุ่มมีค่าไม่แตกต่างกัน

ผลการเปรียบเทียบค่าความเที่ยง ความตรงของแบบสอบระหว่างวิธีให้คะแนนวิธีธรรมดา และวิธีให้คะแนนความมั่นใจ พบว่า 1) แบบสอบที่วัดระดับความรู้ขั้นต่ำ ค่าความเที่ยงของแบบสอบที่ได้จากวิธีให้คะแนนทั้ง 2 วิธีมีค่าไม่แตกต่างกันอย่างมีนัยสำคัญที่ระดับ .01 ในกลุ่มสูง ส่วนกลุ่มปานกลางและต่ำ มีค่าไม่แตกต่างกัน 2) แบบสอบที่วัดระดับความรู้ชั้นสูง ค่าความเที่ยงของแบบสอบที่ได้จากวิธีให้คะแนนทั้ง 2 วิธี มีค่าไม่แตกต่างกันอย่างมีนัยสำคัญทั้งในกลุ่มสูง ปานกลาง และต่ำ ค่าความตรงของแบบสอบมีค่าแตกต่างกันอย่างมีนัยสำคัญที่ระดับ .01 ในกลุ่มต่ำ ส่วนกลุ่มปานกลางและสูงมีค่าไม่แตกต่างกัน

พินิจ อุไรรักษ์ (2534) ศึกษาผลของวิธีการให้คะแนนที่มีต่อคะแนนสอบ ค่าสัมประสิทธิ์ความเที่ยง และค่าสัมประสิทธิ์ความตรง จากวิธีการให้คะแนน 4 วิธี คือ วิธีการให้น้ำหนักคะแนนรายข้อเท่ากัน วิธีการให้น้ำหนักคะแนนรายข้อต่างกันตามระดับความมั่นใจในการตอบ วิธีการให้น้ำหนักคะแนนรายข้อต่างกันค่าพารามิเตอร์ของข้อกระทงโดยสูตรของลอร์ด และวิธีการให้น้ำหนักคะแนนรายข้อต่างกันตามค่าความสามารถของผู้ตอบ กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 1 ปีการศึกษา 2534 เขตจังหวัดบุรีรัมย์ จำนวน 1,011 คน เครื่องมือที่ใช้เป็นแบบสอบวัดผลสัมฤทธิ์วิชาวิทยาศาสตร์ ผลการวิจัยพบว่า

1. ความสัมพันธ์ของคะแนนสอบระหว่างวิธีการให้คะแนนทั้ง 4 วิธี มีค่าตั้งแต่ .941 ถึง .983

2. ค่าสัมประสิทธิ์ความเที่ยงของแบบสอบ พบว่า ค่าความเที่ยงของแบบสอบจากวิธีให้น้ำหนักคะแนนรายข้อเท่ากัน วิธีการให้น้ำหนักคะแนนรายข้อต่างกันตามระดับความมั่นใจ วิธีการให้น้ำหนักคะแนนรายข้อต่างกันค่าพารามิเตอร์ของข้อกระทงโดยสูตรของลอร์ด และวิธีการให้น้ำหนักคะแนนรายข้อต่างกันตามค่าความสามารถของผู้ตอบ มีค่าเท่ากับ .705 , .781 , .886 และ .894 ตามลำดับ

3. วิธีการให้น้ำหนักคะแนนรายข้อต่างกันตามระดับความมั่นใจในการตอบ วิธีการให้น้ำหนักคะแนนรายข้อต่างกันตามค่าพารามิเตอร์ของข้อกระทงโดยสูตรของลอร์ด และวิธีการให้น้ำหนักคะแนนรายข้อต่างกันตามค่าความสามารถของผู้ตอบ โดยเฉลี่ยแล้วให้ค่าสัมประสิทธิ์ของความตรงของแบบสอบสูงกว่า วิธีการให้น้ำหนักคะแนนรายข้อเท่ากัน