



วรรณคดีที่เกี่ยวข้อง

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบคะแนนสอบ และค่าความเที่ยงของแบบสอบ ระหว่างการให้น้ำหนักคะแนนรายข้อเท่ากัน (Raw Scoring Method) ตามทฤษฎีมาตรฐานเดิม (Classical Test Theory) กับการให้น้ำหนักคะแนนรายข้อต่างกัน (Optimal Scoring Weight Method) ตามทฤษฎีการสนองตอบข้อกระทง (Item Response Theory : IRT) ซึ่งมีการศึกษาวิจัยที่เกี่ยวข้องทั้งในต่างประเทศและในประเทศดังต่อไปนี้

เร็นซ์และบาชอร์ (Renz and Bashaw 1975 อ้างถึงใน อวยพร วิบูลย์กาญจน์ 2526) ได้วิจัยพบว่ากลุ่มตัวอย่างระหว่าง 500 - 1,000 คน จะทำให้ค่าประมาณค่าพารามิเตอร์มีความคงที่เพิ่มขึ้นอย่างรวดเร็วและจะค่อย ๆ เพิ่มขึ้น เมื่อกลุ่มตัวอย่างที่ใช้ประมาณค่าพารามิเตอร์มีจำนวน 2,000 และ 4,000 คน

เจมส์ รี (James Ree 1981 : 11 - 19) ได้ศึกษาวิจัยผลของการใช้ขนาดกลุ่มตัวอย่างและขนาดข้อกระทงทั้งหมด (item pools) ต่าง ๆ กันในการกำหนดขนาดของการสอบ (adaptive testing) โดยใช้โมเดลพารามิเตอร์ 3 ตัว (Three-parameter Logistic Model : 3 PL) ขนาดของกลุ่มตัวอย่างเป็น 500, 1,000 และ 2,000 และ item pools เป็น 100, 200 หรือ 300 กำหนดความยาวของ adaptive test ให้มีข้อกระทง 10, 15, 20, 25, 30 และ 35 ศึกษาด้วยวิธี simulation 500 ครั้ง สำหรับการวัดขนาดของ item pool และกลุ่มตัวอย่างแต่ละครั้ง ผลการวิจัยพบว่า ความสัมพันธ์ระหว่างค่าความลำาถและค่าความลำาถจากการประมาณจะสูงมากถ้าการสอบมีจำนวนข้อกระทงที่เพียงพอ (sufficient number of items) ค่าความคลาดเคลื่อนมาตรฐานของการทำนายค่าพารามิเตอร์ของข้อกระทงและของผู้สอบสามารถทำให้น้อยลงได้ เมื่อมีข้อกระทงอย่างน้อย 20 ข้อ และกลุ่มตัวอย่าง 2,000 คน

เรคเคิล (Reckase 1979 : 207 - 230) ได้ศึกษาประเมินโมเดลที่มีพารามิเตอร์ 1 ตัว (1 PL) และโมเดลที่มีพารามิเตอร์ 3 ตัว (3 PL) กับข้อมูลที่เป็น multivariate โดยการ simulated ข้อมูล 5 ชุดและใช้ข้อมูลจริง 5 ชุด และวิเคราะห์โดยใช้เทคนิคโมเดลที่มีพารามิเตอร์ 1 ตัว (1 PL) โมเดลที่มีพารามิเตอร์ 3 ตัว (3 PL) การวิเคราะห์องค์ประกอบ (factor analysis) และการวิเคราะห์ข้อกระทงแบบเดิม (traditional item analysis) พบว่าโมเดลที่มีพารามิเตอร์ 3 ตัว (3 PL) จะวัดเพียงองค์ประกอบเดียวและสามารถจำแนกระดับความสามารถในองค์ประกอบนี้ได้ โดยไม่สนใจองค์ประกอบอื่น : ขณะที่โมเดลที่มีพารามิเตอร์ 1 ตัว (1 PL) เหมือนจะประมาณผลรวมขององค์ประกอบ แต่ก็แสดงให้เห็นว่าคะแนนรวมเป็นสถิติที่พอเพียงสำหรับประมาณระดับความสามารถ และเมื่อองค์ประกอบที่หนึ่ง (first factors) มีความสัมพันธ์กับองค์ประกอบอื่น ๆ มาก ทั้งโมเดลที่มีพารามิเตอร์ 1 ตัว และมีพารามิเตอร์ 3 ตัว จะวัดองค์ประกอบแรก โดยความสัมพันธ์กับคะแนนองค์ประกอบ (factor scores) เกือบจะเป็นทางเดียวกัน (identical) ค่าระดับความสามารถที่ประมาณได้จะเป็นค่าที่ดีแม้ว่าองค์ประกอบแรกมีความแปรปรวนน้อยกว่า 10 เปอร์เซ็นต์ แต่ถ้าจะให้ดีมากกว่าควรมีความแปรปรวนอย่างน้อย 20 เปอร์เซ็นต์ของความแปรปรวนของแบบล่อ

เต คี ชู และ ชูศักดิ์ ชัมภลสิทธิ์ (Tse-chi Hsu and Choosak Kamplikit 1982 : 10 - 16) ทำการศึกษาเปรียบเทียบผลของการให้คะแนนแบบให้น้ำหนักคะแนนรายข้อต่างกัน (Optimal Scoring Weight Method) ตามทฤษฎีการสนองตอบข้อกระทง (IRT) โมเดลที่มีพารามิเตอร์ 3 ตัว (3 PL) กับการให้คะแนนแบบให้น้ำหนักคะแนนรายข้อเท่ากัน (Raw Scoring Method) คะแนนมาตรฐาน (Z) และคะแนนความสามารถ ( $\hat{\theta}$ ) โดยศึกษาข้อมูลจากสภาพการสอบคัดเลือกจริงและได้ใช้แบบล่อฉบับที่มีลักษณะต่างกันคือ แบบล่อภาษาจีนและแบบล่อประวัติคำศัพท์ ซึ่งเป็นแบบล่อที่เขียนขึ้นใหม่ ยังไม่มีการทดลองหรือใช้มาก่อน มีสมมติฐานของการวิจัยว่าการให้คะแนนแบบให้น้ำหนักคะแนนรายข้อต่างกัน (Optimal Scoring Weight Method) น่าจะมีส่วนได้เปรียบกว่าการให้คะแนนแบบอื่นในด้านที่ว่าผลของการให้คะแนนแบบให้น้ำหนักคะแนนรายข้อต่างกัน (Optimal Scoring Weight Method) ขึ้นอยู่กับคุณลักษณะของแบบล่อน้อยกว่าผลของการให้คะแนน

อื่น วิธีการเปรียบเทียบคือ สุ่มตัวอย่างที่จะรับผู้สมัครสอบเข้า เรียง 10% ของจำนวนผู้สมัคร แล้ว การเรียงลำดับตามแบบการให้คะแนนชนิดต่าง ๆ - จะให้จำนวนผู้สอบได้ทั้งสองวิธีได้ก็เปอร์เซ็นต์ โดยคิดจากคะแนนรวมของแบบสอบทั้งสองฉบับ ผลการวิจัยพบว่าในการสอบวัดดังกล่าวสามารถนำข้อมูลจากการสอบไปใช้กับทฤษฎีการสนองตอบข้อกระทง (IRT) โมเดลที่มีพารามิเตอร์ 3 ตัว (3 PL) ได้ คุณภาพของข้อกระทงที่เขียนขึ้นมาใหม่นั้นดูเหมือนจะไม่เป็นปัญหาสำหรับการให้คะแนน โดยวิธีนี้คือ การให้คะแนนแบบให้น้ำหนักคะแนนรายข้อต่างกัน (Optimal Scoring Weight Method) ทำให้ค่าความเที่ยงของแบบสอบ (Reliability) และค่าอำนาจจำแนกของแบบสอบ (discriminating power) สูงกว่าการให้คะแนนแบบอื่น ในด้านความสัมพันธ์ของคะแนนพบว่าคะแนนจากการให้น้ำหนักคะแนนรายข้อเท่ากัน (Raw Scoring Method) กับแบบให้น้ำหนักคะแนนรายข้อต่างกัน (Optimal Scoring Weight Method) มีความสอดคล้องกันสูงถึง 87% ของจำนวนผู้สมัครสอบ 10% ที่ถูกคัดเลือก โดยคะแนนทั้งสองแบบ อาจสรุปได้ว่าการให้คะแนนแบบให้น้ำหนักคะแนน รายข้อเท่ากัน (Optimal Scoring Weight) ตามทฤษฎีการสนองตอบข้อกระทง (IRT) จะช่วยขจัดปัญหาของแบบสอบที่ขาดคุณภาพ และในขณะที่เดียวกันจะช่วยให้คะแนนจากการสอบวัดมีประสิทธิภาพสูงขึ้นอีกด้วย

๙ วัลลภ กันทรพย และ เฮคส์เตียน (Wanlop Kansup and Hakstian 1975 : 219 - 230) ได้ศึกษาเปรียบเทียบค่าความเที่ยง ความตรงของแบบสอบเหตุผลเชิงคณิตศาสตร์ (Mathematical Reasoning Test) และแบบสอบความสามารถทางภาษา (Verbal Ability Test) ระหว่างการให้คะแนนแบบปกติ (Conventional Scoring) คือตอบถูกให้ 1 ตอบผิดหรือไม่ตอบให้ 0 กับแบบให้น้ำหนักตามเหตุผล (Logically Weighted Scoring) โดยที่ข้อกระทงแต่ละข้อจะมีคะแนนน้อยกว่าหรือเท่ากับ 5 ( $\leq 5$ ) สำหรับผู้สอบแต่ละคน และเปรียบเทียบระหว่างการให้คะแนนแบบทดสอบความมั่นใจ (Confidence - Weighted Scoring) ซึ่งมีสูตรการให้คะแนนที่ต่างกัน 5 วิธี คือ

วิธีให้คะแนน	ช่วงคะแนน $Y_{ij}$
1. CW1 : $Y_{ij} = C_{ij}$	0 - 10
2. CW2 : $Y_{ij} = \log_{10} C_{ij}$	0 - 10
3. CW3 : $Y_{ij} = 1 - \log_{10} (10 - C_{ij})$	0 - 10
4. CW4 : $Y_{ij} = C_{ij}^2 / 10$	0 - 10
5. CW5 : $Y_{ij} = Z_{ij}$	- 1.69 - + 1.69

เมื่อ  $Y_{ij}$  แทน คะแนนสำหรับผู้สอบ  $i$  จากข้อกระทง  $j$   
 $X_i$  แทน  $\sum Y_{ij}$   
 $C_{ij}$  แทน คะแนนความมั่นใจ (0 ถึง 10) ที่ตัดสินโดยผู้สอบ  $i$  จากข้อกระทง  $j$   
 $Z_{ij}$  แทน คะแนนปกติมาตรฐานของ  $C_{ij}$

ผลการวิจัยพบว่า การให้คะแนนแบบให้น้ำหนักตามเหตุผล (Logically Weighted Scoring) ทำให้ค่าความเที่ยงของแบบสอบเหตุผลเชิงคณิตศาสตร์เพิ่มจากวิธีให้คะแนนแบบปกติ (Conventional Scoring) เล็กน้อยแต่ไม่มีนัยสำคัญ สำหรับวิธีให้คะแนนแบบทดสอบความมั่นใจ (Confidence Weighted Scoring) พบว่าวิธี CW5 กับ CW2 ให้ค่าความเที่ยงต่างกัน ส่วนแบบสอบความสามารรถทางภาษาพบว่าค่าความเที่ยงต่างกันเฉพาะวิธี CW3 กับ CW4 ในด้านความตรงพบว่า การให้คะแนนแบบปกติ (0 - 1) กับแบบให้น้ำหนักตามเหตุผล ให้ค่าความตรงของแบบสอบต่างกันอย่างมีนัยสำคัญ

ในปี 1975 เช่นกัน เอกส์เตียม และวัลลภ กันทรพิทย์ (Hakstian and Wanlop Kansup 1975 : 231 - 239) ได้ศึกษาเปรียบเทียบค่าความเที่ยง ความตรงของแบบสอบเหตุผลเชิงคณิตศาสตร์ (Mathematical Reasoning Test) และแบบสอบทางภาษา (Verbal Test) ระหว่างวิธีสอบและการตรวจให้คะแนนที่ต่างกัน 3 วิธีคือ วิธีปกติ (Conventionally - Tested) วิธีทดสอบความมั่นใจ (Confidence - Tested) และวิธีตัดข้อผิดออก (Elimination Tested) พบว่าระหว่างวิธีให้คะแนนตัดข้อผิดออกเปรียบเทียบกับวิธีปกติ ทำให้ค่าความเที่ยงของแบบสอบต่างกันบ้างเล็กน้อยอย่างไม่มีนัยสำคัญ ส่วนการให้คะแนนแบบทดสอบความมั่นใจปรากฏชัดเจนว่าค่าความเที่ยง

ของแบบล่อสูงกว่า วิธีให้คะแนนแบบปกติและแบบตัดข้อผิดพลาด ส่วนด้านความตรงของแบบ  
 ล่อไม่ปรากฏรูปแบบที่ชัดเจนระหว่าง การให้คะแนนแบบปกติและแบบตัดข้อผิดพลาด ส่วนการให้  
 คะแนนแบบปกติกับแบบทดสอบความมั่นใจพบว่า วิธีให้คะแนนแบบปกติทำให้ค่าความตรงของแบบล่อ  
 ทางภาษาล่งกว่า : อย่างไรก็ตามพบว่าวิธีให้คะแนนแบบตัดข้อผิดพลาด และแบบทดสอบความ  
 มั่นใจทำให้ค่าความตรงของแบบล่อสูงขึ้น เมื่อเทียบกับวิธีให้คะแนนแบบปกติอยู่หลายกรณี  
 สำหรับกรณีที่ค่าความตรงลดลงนั้น เป็นข้อสังเกตที่ควรศึกษา

บายุก (Bayuk 1973 : 4546 - 4) ศึกษาผลของการให้น้ำหนักตามแบบ  
 การตอบ (response - category weighting) และการให้น้ำหนักข้อกระทง (item-  
 weighting) ที่มีต่อค่าความเที่ยง และค่าความตรงเชิงพยากรณ์ของแบบล่อ ผลปรากฏว่า  
 คะแนนจากการให้น้ำหนักตามแบบการตอบ (response - category weighting) สูง  
 กว่าคะแนนจากจำนวนข้อของการตอบถูก (number - right scores) หรือมากกว่า  
 คะแนนที่ตอบถูกจากโอกาส (score corrected for chance success) อย่างมีนัยสำคัญ  
 การให้คะแนนแต่ละข้อกระทงเมื่อนำมาใช้ในแบบให้น้ำหนักตามแบบการตอบมีความเที่ยงต่ำ  
 กว่าคะแนนที่ตอบถูกจากโอกาสอย่างมีนัยสำคัญ ซึ่งเหตุผลนี้ไม่ชัดเจน ในด้านความตรงพบว่า  
 ไม่มีความแตกต่างระหว่าง การให้คะแนนแบบให้น้ำหนักตามแบบการตอบกับคะแนนที่ตอบถูก  
 จากโอกาส แต่การให้น้ำหนักข้อกระทงมีความตรงมากกว่าแบบคะแนนที่ตอบถูกจากโอกาส  
 อย่างมีนัยสำคัญ

โอเลจเบต (Olagbade 1971 : 1922 - A) ศึกษาผลของการให้คะแนนกับ  
 แบบล่อที่ครูสร้างขึ้น 5 วิธีคือ I. วิธีความน่าจะเป็น (probabilistic) II. วิธี  
 ตัดข้อผิดพลาด โดยมีการตัดคะแนน (elimination - with penalty) III. วิธีตัด  
 ข้อผิดพลาดโดยไม่มีการตัดคะแนน (elimination - without penalty) IV. วิธีให้  
 น้ำหนักตัวเลือก (weighted - choice) และ V. วิธีปกติ (conventional) ผลปรากฏ  
 ดังนี้

1. ค่าความเที่ยงของแบบล่อจากวิธีให้คะแนนที่ต่างกันคือ .862 (II) .854  
 (III) .834 (V) .833 (I) .828 (IV) ซึ่งไม่ต่างกันอย่างมีนัยสำคัญ

2. มัชฌิมเลขคณิตจากวิธีให้คะแนนที่ต่างกันคือ 127.16 (IV) 109.50 (V) 106.40 (I) 108.82 (III) 92.78 (II) ยกเว้นวิธี II และ III แล้วพบว่ามีความแตกต่างกันอย่างมีนัยสำคัญ

3. วิธีตัดข้อผิดออก (elimination model) แสดงให้เห็นว่าช่วยลดการเดามากกว่าวิธีความน่าจะเป็น (probabilistic) แต่ความแตกต่างไม่มีนัยสำคัญ วิธีให้น้ำหนักตัวเลือก (weighted - choice technique) ไม่มีผลต่อการลดการเดา

4. สหสัมพันธ์ระหว่างวิธีให้คะแนนแต่ละแบบสูงมาก (.90 - .98) โดยวิธี I, II, III มีสหสัมพันธ์สูงสุด แบบ III และ IV มีสหสัมพันธ์ต่ำสุด

5. ตารางจัดลำดับ (table of rank) แสดงให้เห็นว่า วิธีความน่าจะเป็น (probabilistic) มีสหสัมพันธ์สูงสุด ส่วนวิธีให้น้ำหนักตัวเลือก (weighted - choice) มีสหสัมพันธ์ต่ำสุด กับลำดับของทั้งหมด

นอกจากนั้น โอเลจเบต ให้ข้อสังเกตว่า วิธีให้น้ำหนักข้อกระทงโดยทั่วไปมีประโยชน์น้อยที่สุดในการประเมินความรู้บางส่วน (partial knowledge) และพอสรุปโดยไม่มีพิจารณาเหตุผลอื่น ๆ ได้ว่า วิธีง่าย ๆ ไม่ซับซ้อนคือวิธีปกติ (conventional technique) (without penalty) ควรใช้ต่อไปขณะที่การศึกษาทดลองยังคงต้องสืบสาวต่อไปในอนาคต

คูมบส์และโวเมอร์ (Coombs and Womer 1956 : 13 - 17) ได้ทำการศึกษาเปรียบเทียบวิธีที่เสนอนี้ คือการเลือกคำตอบที่ผิดแทนการเลือกคำตอบที่ถูก กับวิธีธรรมดา โดยใช้แบบสอบ 3 ฉบับ ฉบับละ 40 ข้อกระทง โดยตรวจให้คะแนน 3 วิธีคือ วิธี 0 - 1 วิธีของคูมบส์ที่เสนอนี้ และวิธี 0 - 1 กับวิธีของคูมบส์รวมกันผลปรากฏดังนี้

1. ค่าความเที่ยงของแบบสอบจากวิธีการตอบและตรวจตามที่คูมบส์ เสนอแนะ สูงกว่าวิธี 0 - 1

2. ค่าความตรงของแบบสอบแตกต่างกันอย่างไม่มีนัยสำคัญ

3. มาตรฐานความมั่นใจในการทำแบบสอบ วิธีที่คูมบส์ เสนอแนะมีมากกว่าวิธี 0 - 1 และสามารถป้องกันการเดาได้มากกว่า

4. ทศน คคิของกลุ่มตัวอย่่าง ฆอบวิธีที่คุมบัล้เล่เสนอแนะมากรักว่า โดยให้เหตุผลว่า เป็นวิธีที่ยุติธรรมและง่า ยกว่าวิธี 0 - 1

\* เดวิส และฟิฟเฟอร์ (Davis and Fiffer 1959 : 159 - 169) ทำการ ศึกษาทดลองเกี่ ยวกับค่าความเที่ยงและค่าความตรงของแบบล่อบวัดความถนัดและแบบล่อบวัด ผลลัสมฤทธิ์ทางการ เรียบวิชาคณิตค่าลัศร์ โดยกำหนดให้หน้าหน้กคะแนนของแต่ละตัวเลือกไม่ เท่ากันเปรียบเทียบกับวิธี 0 - 1 เดวิสและฟิฟเฟอร์มีความคิดว่่า การคิดจ้อกห้ยคคิคณิตค่าลัศร์ นั้นตัวเลือกแต่ละตัวมีความยากง่า ยไม่เท่ากัน และใช้ลัสมรรถภาพทางล่อมองในการคิดไม่เท่ากัน ด้วย จึงกำหนดคะแนนของแต่ละตัวเลือกไม่เท่ากัน ผลการศึกษพบว่า

1. การตรวจให้คะแนนโดยกำหนดคะแนนของแต่ละตัวเลือกไม่เท่ากันนั้น ให้ค่าความเที่ยงสูงกว่าการตรวจโดยวิธี 0 - 1 อย่างมีนัยสำคัญ ✓
2. การตรวจโดยกำหนดคะแนนของแต่ละตัวเลือกไม่เท่ากันนั้น ให้ค่าความตรงสูงกว่าการตรวจโดยวิธี 0 - 1 แต่ไม่มีนัยสำคัญ

▷ แพทเนอิค และทรูบ (Patnaik and Traub 1973 : 281 - 285) ทำ การทดลองศึกษาวิธีตอบและตรวจให้คะแนนที่คล้ายกับวิธีของเดวิส และฟิฟเฟอร์ โดยกำหนด คะแนนของตัวเลือกถูกในแต่ละข้อให้มีคะแนนต่างกัน เปรียบเทียบกับวิธีตอบแบบ 0 - 1 โดยการกำหนดน้ำหนักของตัวเลือกที่ถูกในแต่ละข้อ ใช้การตัดสินของครูและผู้เชี่ยวชาญรวม 61 คน จัดอันดับน้ำหนักคะแนนของตัวเลือกที่ถูกในแต่ละข้อ แล้วหาค่าเฉลี่ยออกมาเป็นคะแนนในแต่ละข้อ ผลการศึกษพบว่าล่อดคล่องกับผลการวิจัยของเดวิสและฟิฟเฟอร์ กล่าวคือ ค่าความเที่ยงของ แบบล่อบที่มีวิธีให้คะแนน วิธี 0 - 1 และแบบให้คะแนน ข้อถูกแต่ละข้อไม่เท่ากันเมื่อคำนวณโดย วิธีแบ่งครึ่งแบบล่อบมีลัสมประลัทธิความเที่ยงเป็น .881 และ .915 ตามล่่าดับ ส่วนค่าความตรงของแบบล่อบซึ่งหาความตรงเชิงพยากรลัสมพบว่า วิธีตรวจแบบให้ค่าน้หนักคะแนนแต่ละข้อไม่ เท่ากันมีความตรงน้อยกว่่าวิธีตอบและตรวจให้คะแนนแบบ 0 - 1

ฟิฟ และบรันซ่า (Pugh and Brunza 1975 : 73 - 78) ได้เสนอให้มี วิธีตอบโดยบอกความม่ั่นใจในการตอบ และน้หนักความม่ั่นใจในการตอบมาเป็นลัสม ในการให้คะแนนด้วย ฟิฟและบรันซ่าจึงได้ศึกษาหาค่าความเที่ยงของแบบล่อบที่มีวิธีตอบและตรวจให้ คะแนน 2 วิธี คือ วิธีตอบแบบธรรมด่า และวิธีตอบโดยให้ผู้ตอบบอกระดับความม่ั่นใจในการ เลือกตอบข้อค่าถามเหล่าน้ัน ผลการศึกษพบว่า ค่าความเที่ยงของแบบล่อบซึ่งหาโดยวิธี

วิเคราะห์ความแปรปรวนของแบบลอบที่มีวิธีตอบโดยบอกความมั่นใจในการตอบสูงกว่าแบบลอบที่ตอบแบบธรรมดา คือ .85 และ .57 ตามลำดับ

อาบู ซาอีฟ และไดมอนด์ (Abu - Sayf and Diamond 1976 : 62 - 65) ทำการศึกษาเกี่ยวกับค่าความเที่ยงและค่าความตรงของแบบลอบเมื่อตรวจให้คะแนน 3 วิธี คือให้คะแนนเฉพาะข้อถูก และตอบอย่างมั่นใจ ให้คะแนนข้อที่ถูกและไม่ได้เท่า และให้คะแนนทุกข้อที่ตอบถูกโดยไม่ว่ามีค่าความมั่นใจในการตอบหรือไม่เพียงใด แล้วหาค่าความเที่ยงของแบบลอบโดยใช้สูตร ครอสมัค แอลฟา ( $\alpha$  - Coefficient) ผลปรากฏว่า การให้คะแนนเฉพาะข้อที่ตอบถูกและมีความมั่นใจได้ค่าความเที่ยงสูงสุดคือ .898 ส่วนการให้คะแนนข้อที่ตอบถูกโดยไม่ได้เท่าได้ค่าความเที่ยงสูงกว่าการให้คะแนนทุกข้อที่ตอบถูก คือได้ค่าความเที่ยง .871 และ .724 ตามลำดับ ส่วนค่าความตรงนั้นวิธีการให้คะแนนข้อที่ตอบถูกและไม่ได้เท่ามีความตรงสูงสุดคือ .260 ส่วนการให้คะแนนเฉพาะข้อที่ตอบถูกและมีความมั่นใจให้ค่าความตรงสูงกว่าการให้คะแนนทุกข้อที่ตอบถูกคือ .236 และ .219 ตามลำดับ

ชิน ซิง เวิน (Shih - Sung Wen 1975 : 197 - 199) ได้ศึกษาลหสัมพันธ์ระหว่างคะแนนที่นับจากผลรวมจำนวนข้อที่ตอบถูก กับคะแนนการทดสอบความมั่นใจพบว่ามีความสัมพันธ์กันสูงถึง .97 ซึ่งในทางปฏิบัติแล้วคะแนนที่ได้จากทั้งสองวิธีนี้สามารถนำมาจัดอันดับ (Rank order) ได้เหมือนกันแต่เขาก็ให้ข้อคิดว่า ถึงแม้ลหสัมพันธ์ระหว่างคะแนนที่ได้จากทั้งสองวิธีนี้จะสูง แต่ก็ไม่ได้เป็นข้ออ้างที่จะปฏิเสธวิธีการทดสอบความมั่นใจ ถึงอย่างไรการทดสอบความมั่นใจก็สามารถให้รายละเอียดหรือข้อมูลในการตอบข้อกระทงที่มากกว่าสำหรับผลการศึกษาวิจัยในประเทศไทยที่เกี่ยวข้องมีดังนี้

อรรวรรณ ศัลย์เจริญรัตน์ (2517 : 23 - 26) ได้ศึกษาเปรียบเทียบค่าความเที่ยงและเปอร์เซ็นต์การเดาของแบบลอบที่มีวิธีการตอบและตรวจให้คะแนนตามเงื่อนไขวิธี 0 - 1 วิธีของคูมบัส และวิธีใหม่ตามแนวคิดของอนันต์ ศิริโสภา คือให้ตอบทุกตัวเลือกว่าตัวเลือกใดถูกตัวเลือกใดผิด ผลการวิจัยพบว่าค่าความเที่ยงของแบบลอบที่มีวิธีตอบและตรวจตามวิธีของอนันต์ ศิริโสภา สูงกว่าค่าความเที่ยงของแบบลอบที่มีวิธีตอบและตรวจตามวิธีของคูมบัส และวิธีตอบแบบธรรมดาให้ค่าความเที่ยงต่ำสุด เปอร์เซ็นต์การเดาของการตอบตามเงื่อนไขของอนันต์ ศิริโสภา ต่ำกว่าการเดาในการตอบตามเงื่อนไขของคูมบัส แต่สูงกว่าการเดาในการตอบตามวิธี 0 - 1



กาญจนา ศิริวัฒนพงษ์ (2520) ได้ศึกษาเปรียบเทียบค่าความเที่ยง ค่าความตรง ค่าอำนาจจำแนก และเปอร์เซ็นต์การเดาของแบบลอบปรนัยชนิดเลือกตอบที่มีวิธีการตอบและตรวจต่างกัน 3 วิธี คือ ตอบแบบธรรมดา วิธีของคูมบ์ลี และวิธีของอนันต์ ศิริโสภา โดยใช้แบบลอบวิชาวิทยาศาสตร์ ทดลอบกับกลุ่มตัวอย่างที่เป็นนักเรียนมัธยมศึกษาปีที่ 3 ในจังหวัดปทุมธานี รวม 810 คน ผลปรากฏว่า การตอบและตรวจโดยวิธีของอนันต์ ศิริโสภา ให้ค่าความเที่ยงและค่าความตรงของแบบลอบสูงกว่าวิธีธรรมดา และวิธีของคูมบ์ลี และวิธีตรวจแบบธรรมดาให้ค่าความเที่ยงสูงกว่าวิธีของคูมบ์ลี แต่ค่าความตรงไม่แตกต่างกัน ส่วนค่าอำนาจจำแนกของแบบลอบจากการตรวจแต่ละวิธีไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ ในด้านกาเดาปรากฏว่า วิธีตอบและตรวจให้คะแนนด้วยวิธีของอนันต์ ศิริโสภามีการเดาน้อยกว่าวิธีตอบแบบธรรมดา และวิธีของคูมบ์ลีโดยที่วิธีของคูมบ์ลีมีการเดาลงกว่าวิธีอื่น ๆ

ธีรศักดิ์ อินทรมาตย์ (2520) ได้ศึกษาเปรียบเทียบค่าความเที่ยง ค่าความตรง และปริมาณการเดาของแบบลอบปรนัยชนิดเลือกตอบที่มีวิธีการตอบแตกต่างกัน 3 วิธี คือวิธีตอบแบบธรรมดา วิธีตอบโดยบอกความมั่นใจในการตอบ และวิธีตอบทุกตัวเลือก โดยใช้แบบลอบวิชาวิทยาศาสตร์ทดสอบกลุ่มตัวอย่างที่เป็นนักเรียนชั้นประถมศึกษาปีที่ 7 ที่มีระดับความสามารถในวิชาวิทยาศาสตร์แตกต่างกัน 3 กลุ่ม ผลการศึกษากาปรากฏว่าค่าความเที่ยงแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ในกลุ่มนักเรียนที่มีความสามารถในวิชาวิทยาศาสตร์ระดับสูง และปานกลาง และแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ในกลุ่มนักเรียนที่มีความสามารถในวิชาวิทยาศาสตร์ระดับต่ำ สำหรับค่าความตรงของแบบลอบที่มีวิธีการตอบแตกต่างกันพบว่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติในทุกระดับความสามารถในวิชาวิทยาศาสตร์ของนักเรียน ในด้านปริมาณการเดาพบว่า แบบลอบที่ตอบแบบธรรมดามีปริมาณการเดาลงสุด รองลงมาคือแบบลอบที่บอกระดับความมั่นใจในการตอบ และแบบลอบที่ตอบทุกตัวเลือกมีปริมาณการเดาต่ำสุด และจากกาการศึกษาเรื่องปริมาณการเดานี้พบว่านักเรียนที่มีความสามารถในวิชาวิทยาศาสตร์ระดับต่ำจะตอบแบบลอบโดยกาเดามากกว่านักเรียนที่มีความสามารถในวิชาวิทยาศาสตร์ระดับสูงทุกวิธีการตอบ

เพ็ญศรี ล่วงเนตร (2520) ศึกษาความเที่ยงของแบบลอบชนิดเลือกตอบเมื่อกำหนดเทคนิคการให้คะแนนที่ต่างกัน 4 วิธี คือ

1. วิธีไม่แกการเตาโดยกำหนดน้ำหนักคะแนน 1 ให้ข้อที่ถูกต้องและ 0 ให้ข้อที่ตอบผิดหรือเว้น
2. วิธีแกการเตาโดยลดคะแนน ของข้อที่ตอบผิด
3. วิธีแกการเตาโดยเพิ่มคะแนนให้ข้อที่เว้น
4. วิธีแกการเตาโดยกา รทลสอบความมั่นใจ

#### ผลการวิจัยพบว่า

1. เทคนิคให้คะแนนแกการเตาโดยวิธีทลสอบความมั่นใจให้ค่าความเที่ยงสูงขึ้มาจากวิธีไม่แกการเตาอย่างมีนัยสำคัญทางสถิติที่ระดับ .01
2. เทคนิคให้คะแนนแกการเตาโดยกา รลดคะแนนของข้อที่ตอบผิด ทำให้ค่าความเที่ยงเพิ่มขึ้นจากวิธีไม่แกการเตาเพียงเล็กน้อย ไม่มีนัยสำคัญทางสถิติ
3. สหสัมพันธ์ระหว่างเทคนิคการให้คะแนนทุกคู่มีค่าสูงตั้งแต่ .92 ขึ้นไป ยกเว้นเทคนิคแกการเตาโดยวิธีเพิ่มคะแนนให้ข้อที่เว้น กับวิธีทลสอบความมั่นใจมีสัมประสิทธิ์สหสัมพันธ์ .89

นอกจากนั้น เพ็ญศรี ล้วางเนตร ยังให้ข้อสรุปว่า เทคนิคการให้คะแนนความมั่นใจมีประสิทธิภาพที่สุด ในการเพิ่มค่าความเที่ยงของการวัดผลสัมฤทธิ์ โดยแบบลสอบชนิดเลือกตอบ

× สาราญ มีแจ้จ (2525) ศึกษาผลของคำสั่งและการให้คะแนนที่ต่างกัน 4 วิธี คือ วิธี 0 - 1 วิธีของคุมบ์ส์ วิธีของอนันต์ ศรีโสภา และวิธีของสาราญ มีแจ้จ คือ เลือกตามจำนวนตัวเลือกที่แน่ใจว่ามีตัวเลือกถูกที่สุดรวมอยู่ด้วย นั่นคืออาจจะเลือก 1 หรือ 2 หรือ 3 ตัวเลือก ที่มีต่อค่าความเที่ยง ค่าความตรง และค่าอำนาจจำแนกของแบบลสอบชนิดเลือกตอบวิชาคณิตศาสตร์ ทลสอบกับนักเรียนชั้นมัธยมศึกษาปีที่ 4 จำนวน 240 คน

#### ผลการวิจัยพบว่า

1. ค่าความเที่ยงของแบบลสอบที่ใช้คำสั่งและการให้คะแนนตามวิธีของสาราญ มีแจ้จ สูงกว่าค่าความเที่ยงของแบบลสอบที่ใช้คำสั่งและการให้คะแนนตามวิธี 0 - 1 ( $P < .01$ ) และค่าความเที่ยงของแบบลสอบที่ใช้คำสั่งและการให้คะแนนตามวิธีของคุมบ์ส์สูงกว่าค่าความเที่ยงของแบบลสอบที่ใช้คำสั่งและการให้คะแนนตามวิธี 0 - 1 ( $P < .05$ )

2. ค่าความตรงของแบบล่อที่ใช้คำสั่งและการให้คะแนน ตามวิธีของส้าราญ มีแล้ง  
 ลุงกว่า ค่าความตรงของแบบล่อที่ใช้คำสั่งและการให้คะแนนตามวิธี 0 - 1 ( $P < .01$ )  
 และค่าความตรงของแบบล่อที่ใช้คำสั่งและการให้คะแนนตามวิธีของอนันต์ ลุงกว่าค่าความตรง  
 ของแบบล่อที่ใช้คำสั่งและการให้คะแนนตามวิธี 0 - 1 ( $P < .05$ )

3. คำสั่งและการให้คะแนนทั้ง 4 วิธี ไม่ทำให้ค่าอำนาจจำแนกของแบบล่อแตก  
 ต่างกัน ( $P < .05$ )

จากรายงานการวิจัยที่กล่าวมาทั้งหมด ผลการวิจัยและแนวคิดยังแยกออกเป็น 2  
 กลุ่มคือ กลุ่มหนึ่งพบว่า การให้คะแนนแบบอื่น ๆ ซึ่งต่างจากการให้คะแนนตามปกติคือการให้  
 น้ำหนักคะแนนรายข้อเท่ากัน (Raw Scoring Method) ซึ่งนิยมใช้กันทั่วไปทำให้ค่าความ  
 เทียบของแบบล่อสูงขึ้น แต่อีกกลุ่มหนึ่งมีความเห็นว่า การให้คะแนนแบบอื่น ๆ มีความยุ่ง  
 ยาก ซ้ำซ้อน และให้ประโยชน์น้อยควรจะยังคงใช้วิธีให้น้ำหนักคะแนนรายข้อเท่ากัน แบบ  
 เดิมต่อไปขณะที่ยังไม่มีการค้นพบที่ดีกว่า แต่จะสังเกตได้ว่างานวิจัยดังกล่าวยังไม่มีการนำ  
 วิธีให้คะแนนตามทฤษฎีการสนองตอบข้อกระทง (IRT) มาศึกษาเปรียบเทียบกับ การให้น้ำหนัก  
 คะแนนรายข้อเท่ากัน (Raw Scoring Method) ตามทฤษฎีมาตรฐานเดิม (Classical  
 Test Theory) ยกเว้นงานวิจัยของเต.ศิ ชู และชูศักดิ์ ชัมภลสิทธิ์ ซึ่งก็เป็นการศึกษาวิจัย  
 ในต่างประเทศและมีได้เปรียบเทียบถึงความแตกต่างของมัชฌิมเลขคณิต ส่วนเบี่ยงเบน  
 มาตรฐาน จากการตรวจให้คะแนนที่แตกต่างกัน ผู้วิจัยจึงสนใจที่จะศึกษาถึงผลของการให้  
 คะแนนที่ต่างกันคือ การให้น้ำหนักคะแนนรายข้อเท่ากัน (Raw Scoring Method) ตาม  
 ทฤษฎีมาตรฐานเดิม (Classical Test Theory) กับการให้น้ำหนักคะแนนรายข้อต่างกัน  
 (Optimal Scoring Weight Method) ตามทฤษฎีการสนองตอบข้อกระทง (Item  
 Response Theory : IRT) ต่อค่ามัชฌิมเลขคณิต ส่วนเบี่ยงเบนมาตรฐาน และคุณภาพ  
 ของแบบล่อด้านความเที่ยง (Reliability) ตลอดจนความสัมพันธ์ของคะแนนล่อระหว่าง  
 การตรวจให้คะแนนที่ต่างกัน โดยอาศัยเอกสารและงานวิจัยที่กล่าวมา เป็นแนวทางในการ  
 ศึกษาเปรียบเทียบครั้งนี้