บทที่ 2

วรรณคดีที่เกี่ยวข้อง

ในการวิจัยครั้งนี้ผู้วิจัยได้นำเสนอวรรณคดีที่เกี่ยวข้องในการวิจัย โดยแบ่งออกเป็น 2 ตอน ดังนี้

ตอนที่ 1 ทฤษฎีการทดสอบ

1.1 ทฤษฎีคลาสสิคอล

1.2 ทฤษฎีการตอบสนองข้อสอบ

1.3 ความแตกต่างระหว่างทฤษฎีคลาสสิคอลกับทฤษฎีการตอบสนองข้อสอบ

ตอนที่ 2 งานวิจัยที่เกี่ยวข้อง

2.1 ตามกรอบของทฤษฎีคลาสสิคอล

2.2 ตามกรอบของทฤษฎีการตอบสนองข้อสอบ

ตอนที่ 1 ทฤษฎีการทดสอบ

1.1 ทฤษฎีคลาสสิคอล

การวัดผลการศึกษาโดยเฉพาะการทดสอบ นักวัดผลได้ใช้แบบสอบไปเร้าให้ผู้สอบได้แสดงพฤติกรรมของความสามารถที่แท้จริงภายในตัวผู้สอบซึ่งไม่สามารถวัดได้โดยตรง แล้วนำเอาพฤติกรรมที่ได้จากจำนวนข้อที่ตอบถูกซึ่งได้กำหนดในรูปของคะแนนเป็นตัวแทนของความสามารถที่แท้จริงของผู้สอบ ซึ่งทฤษฎีการวัดความสามารถต่าง ๆ ก็พยายามให้ผลการทดสอบแต่ละครั้งตรงกับความสามารถที่แท้จริงของผู้สอบที่เรียกว่าคะแนนจริง (True Score) มากที่สุด ในปัจจุบันทฤษฎีการวัดและประเมินผลการศึกษาที่ได้พัฒนาและนิยมใช้กันอย่างกว้างขวาง คือ ทฤษฎีคลาสสิคอล ซึ่ง กูลลิคเซ็น (Gulliksen)

(1950) ได้เป็นผู้ที่เสนอทฤษฎีนี้ (Warm 1978: 15) ทฤษฎีคลาสสิคอลนี้เป็นโมเดลทางคณิตศาสตร์ที่แสดงถึงความคลาดเคลื่อนของการวัดที่มีผลต่อคะแนนที่วัดได้หรือสังเกตได้

ข้อตกลงเบื้องต้นของทฤษฎีคลาสสิคอล

1. คะแนนที่วัดได้ เท่ากับ ผลรวมของคะแนนจริงกับคะแนนความคลาดเคลื่อน

2. ค่าความคาดหวังของคะแนนที่วัดได้ เท่ากับ คะแนนจริง

3. คะแนนความคลาดเคลื่อนและคะแนนจริงไม่มีสหสัมพันธ์กัน

4. คะแนนความคลาดเคลื่อนของแบบสอบที่แตกต่างกัน 2 ฉบับ ไม่มีสหสัมพันธ์

5. คะแนนความคลาดเคลื่อนของแบบสอบฉบับหนึ่งไม่มีสหสัมพันธ์กันกับคะแนนจริงของแบบสอบอีกฉบับหนึ่ง

6. ถ้าแบบสอบ 2 ฉบับ มีคะแนนที่วัดได้เป็น X และ X' และเป็นไปตามข้อตกลงเบื้องต้น 1-5 และถ้าสำหรับทุก ๆ ประชากรผู้สอบซึ่งมีคะแนนจริงและคะแนนความคลาดเคลื่อน โดยที่ $T = T'$ และ $\sigma_E^2 = \sigma_{E'}^2$ แล้วแบบสอบ 2 ฉบับนี้เรียกว่า แบบสอบคู่ขนานกัน

7. ถ้าแบบสอบ 2 ฉบับ มีคะแนนที่วัดได้เป็น X_1 และ X_2 และเป็นไปตามข้อตกลงเบื้องต้น 1-5 และถ้าสำหรับทุก ๆ ประชากรผู้สอบซึ่งมีคะแนนจริง โดยที่ $T_1 = T_2 + c_{12}$ เมื่อ c_{12} เป็นค่าคงที่ แล้วแบบสอบ 2 ฉบับนี้เรียกว่า แบบสอบที่สมมูลกัน

การตรวจสอบคุณภาพของแบบสอบ

ตามกรอบของทฤษฎีคลาสสิคอล มีกระบวนการตรวจสอบคุณภาพของแบบสอบ 2 กระบวนการ คือ

1. การวิเคราะห์ข้อสอบรายข้อ โดยพิจารณาคุณลักษณะที่สำคัญของข้อสอบ 2 ประการ คือ

1.1 ค่าความยากของข้อสอบ

ค่าความยากของข้อสอบ หมายถึง สัดส่วนของผู้สอบทั้งหมดที่ตอบข้อสอบแต่ละข้อถูกต้อง ซึ่งมีค่าตั้งแต่ 0 ถึง 1 ถ้าค่าความยากของข้อสอบใกล้ 0 ข้อสอบข้อนั้นเป็นข้อสอบที่ยาก ถ้าค่าความยากของข้อสอบใกล้ 1 ข้อสอบข้อนั้นเป็นข้อสอบที่ง่าย ซึ่งทั้ง 2 ลักษณะนี้เป็นข้อสอบที่ไม่สามารถแยกความสามารถของผู้สอบออกจากกันได้จึงควรตัดออกไป ถ้าค่าความยากของข้อสอบใกล้ .5 ข้อสอบข้อนั้นมีความยากพอเหมาะสามารถแยกผู้สอบออกจากกันได้ ในทางปฏิบัติโดยทั่วไปค่าความยากของข้อสอบที่ใช้มีค่าตั้งแต่ .3 ถึง .7

1.2 ค่าอำนาจจำแนกของข้อสอบ

ค่าอำนาจจำแนกของข้อสอบ หมายถึง ประสิทธิภาพของข้อสอบที่จะสามารถแยกความแตกต่างระหว่างกลุ่มที่ได้คะแนนสูงกับกลุ่มที่ได้คะแนนต่ำ ซึ่งมีค่าตั้งแต่ -1.00 ถึง 1.00 ข้อสอบที่มีค่าอำนาจจำแนกใกล้ 1.00 แสดงถึงกลุ่มที่ได้คะแนนสูงตอบข้อสอบข้อนั้นได้ถูกต้องมากกว่ากลุ่มที่ได้คะแนนต่ำ ข้อสอบที่มีค่าอำนาจจำแนกใกล้ -1.00 แสดงถึงกลุ่มที่ได้คะแนนสูงตอบข้อสอบข้อนั้นได้ถูกต้องน้อยกว่ากลุ่มที่ได้คะแนนต่ำ ถ้าข้อสอบที่มีค่าอำนาจจำแนกเป็น 0 แสดงถึงกลุ่มที่ได้คะแนนสูงตอบข้อสอบข้อนั้นได้ถูกต้องเท่ากับกลุ่มที่ได้คะแนนต่ำ ซึ่งไม่สามารถแยกความแตกต่างของทั้ง 2 กลุ่มได้เลย ในทางปฏิบัติโดยทั่วไปค่าอำนาจจำแนกของข้อสอบที่ใช้มีค่าตั้งแต่ .20 ขึ้นไป

2. การวิเคราะห์แบบสอบทั้งฉบับ โดยพิจารณาคุณลักษณะที่สำคัญ 2 ประการ คือ

2.1 ความตรง

ความตรง หมายถึง แบบสอบนั้น ๆ สามารถวัดในสิ่งที่เราต้องการจะวัดได้อย่างถูกต้อง ซึ่งความตรงสามารถพิจารณาได้หลายวิธีขึ้นอยู่กับแบบสอบและความต้องการที่จะใช้วัด ความตรงที่สำคัญมี 3 ชนิด คือ

1. ความตรงตามเนื้อหา เป็นความตรงที่เกี่ยวกับการวิเคราะห์ตรวจสอบอย่างมีระบบในเนื้อหาของแบบสอบ ซึ่งความตรงนี้การพิจารณาขึ้นอยู่กับการ

ตัดสินของแต่ละบุคคลและการตัดสินแบบปรนัย ความตรงตามเนื้อหาที่สำคัญมี 2 ชนิด คือ

1.1 <u>ความตรงตามปรากฏ</u> เป็นความตรงที่ผู้ตรวจแบบสอบซึ่งเป็นผู้เชี่ยวชาญได้สรุปว่า แบบสอบนั้นวัดได้ตรงตามความต้องการที่จะวัด

1.2 <u>ความตรงตามตรรกหรือการสุ่ม</u> เป็นความตรงที่เกี่ยวกับการสุ่มเนื้อหามาใช้ โดยมีการนิยามโดเมนของพฤติกรรมที่จะวัดในการสอบ แล้วสุ่มเนื้อหาในการสร้างข้อสอบโดยให้ครอบคลุมเนื้อหาที่สำคัญของโดเมนที่นิยามไว้ ซึ่งมีประโยชน์มากในการพัฒนาแบบสอบวัดผลสัมฤทธิ์.

2. <u>ความตรงตามเกณฑ์สัมพันธ์</u> เป็นความสัมพันธ์ระหว่างคะแนนจากการสอบกับเกณฑ์ โดยที่เกณฑ์เป็นพฤติกรรมที่ได้จากคะแนนการสอบที่ใช้ทำนาย ความตรงตามเกณฑ์สัมพันธ์แสดงในรูปของสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนการสอบหรือตัวทำนายกับคะแนนเกณฑ์มี 2 ชนิด คือ

2.1 <u>ความตรงตามทำนาย</u> คือ สัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนการสอบกับคะแนนเกณฑ์ ซึ่งคะแนนเกณฑ์เป็นคะแนนที่ได้รวบรวมภายหลัง ความตรงตามทำนายนี้เป็นการใช้คะแนนการสอบทำนายพฤติกรรมในอนาคต

2.2 <u>ความตรงร่วมสมัย</u> คือ สัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนการสอบกับคะแนนเกณฑ์ ซึ่งคะแนนทั้งสองนี้ได้วัดในระยะเวลาเดียวกัน ความตรงร่วมสมัยนี้เป็นการใช้คะแนนการสอบประมาณเกณฑ์ร่วมสมัยมากกว่าใช้ทำนายเกณฑ์ในอนาคต

3. <u>ความตรงตามทฤษฎี</u> เป็นขอบเขตความหมายหรือคุณลักษณะประจำตามทฤษฎีที่แบบสอบนั้นสามารถวัดได้ วิธีการหาความตรงตามทฤษฎีมีหลายวิธี เช่น การหาสหสัมพันธ์ การหาความสอดคล้องภายใน เป็นต้น โดยเฉพาะการหาความตรงตามทฤษฎีด้วยวิธีการหาสหสัมพันธ์มี 2 ชนิด คือ

3.1 <u>ความตรงหลายคุณลักษณะและหลายวิธี</u> เป็นวิธีที่ใช้หาความตรงเมื่อมีคุณลักษณะ 2 คุณลักษณะขึ้นไป ที่ถูกวัดโดยวิธีตั้งแต่ 2 วิธีขึ้นไป โดยการตรวจสอบในรูปของสหสัมพันธ์มี 2 ชนิด คือ

3.1.1 <u>ความตรงลู่เข้า</u> เป็นความตรงที่มีค่าสหสัมพันธ์สูงระหว่างคะแนนการสอบที่วัดคุณลักษณะเดียวกันด้วยวิธีที่แตกต่างกัน

3.1.2 <u>ความตรงจำแนก</u> เป็นความตรงที่มีค่าสหสัมพันธ์ต่ำระหว่างคะแนนการสอบที่วัดคุณลักษณะที่แตกต่างกันด้วยการใช้วิธีที่เหมือนกัน

3.2 <u>ความตรงตามตัวประกอบ</u> เป็นการพิจารณาโดยใช้กระบวนการวิเคราะห์ตัวประกอบของสหสัมพันธ์ของคะแนนจากแบบสอบ

2.2 <u>ความเที่ยง</u>

ความเที่ยง หมายถึง ความคงที่ของคะแนนที่ได้จากการทดสอบคนกลุ่มเดียวกันสองครั้ง ด้วยแบบสอบเดิมในเวลาที่ต่างกัน หรือทดสอบคนกลุ่มเดียวกันด้วยข้อสอบต่างชุดที่มีข้อสอบเทียบเท่ากัน หรือภายใต้สภาพการสอบที่แตกต่างกัน (อนาสตาซี 2519: 73)

จากนิยามของความเที่ยงสามารถแปลความหมายได้หลายอย่าง เช่น ถ้าเราได้คะแนนที่วัดได้และคะแนนจริงจากการสอบแบบสอบฉบับหนึ่ง โดยที่คะแนนที่วัดได้มีสหสัมพันธ์สูงกับคะแนนจริงก็ถือว่าแบบสอบนั้นมีความเที่ยง สหสัมพันธ์กำลังสองระหว่างคะแนนที่วัดได้และคะแนนจริง เรียกว่า สัมประสิทธิ์ความเที่ยงของแบบสอบ หรือความเที่ยงสามารถแสดงในรูปของสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนที่วัดได้จากแบบสอบคู่ขนานกัน ถ้า X และ X' เป็นคะแนนที่วัดได้จากแบบสอบคู่ขนานกัน สหสัมพันธ์ของคะแนนคู่นี้ก็เป็นสัมประสิทธิ์ความเที่ยง แต่โดยทั่วไปเราไม่สามารถรู้คะแนนจริงและไม่สามารถบ่งชี้ได้ชัดเจนว่าแบบสอบจะคู่ขนานกัน ดังนั้นจึงต้องใช้การประมาณความเที่ยงโดยวิธีอื่น ๆ ในการแปลความหมายของสัมประสิทธิ์ความเที่ยงสามารถแปลความหมายได้ 6 วิธี ดังนี้

1. สัมประสิทธิ์ความเที่ยง คือ สหสัมพันธ์ระหว่างคะแนนที่วัดได้จากแบบสอบคู่ขนานกัน

2. สัมประสิทธิ์ความเที่ยง คือ สัดส่วนของความแปรปรวนใน X ที่อธิบายได้ด้วยความสัมพันธ์เชิงเส้นตรงกับ X'

3. สัมประสิทธิ์ความเที่ยง คือ อัตราส่วนของความแปรปรวนของคะแนนจริงกับความแปรปรวนของคะแนนที่วัดได้

4. สัมประสิทธิ์ความเที่ยง คือ สหสัมพันธ์กำลังสองระหว่างคะแนนที่วัดได้กับคะแนนจริง

5. สัมประสิทธิ์ความเที่ยง คือ ค่าความแตกต่างระหว่าง 1 กับสหสัมพันธ์กำลังสองระหว่างคะแนนที่วัดได้กับคะแนนความคลาดเคลื่อน

6. สัมประสิทธิ์ความเที่ยง คือ ค่าความแตกต่างระหว่าง 1 กับอัตราส่วนของความแปรปรวนของคะแนนความคลาดเคลื่อนกับความแปรปรวนของคะแนนที่วัดได้

จากนิยามของความเที่ยงนั้น โดยทั่วไปเราไม่สามารถรู้คะแนนจริงจากการสอบได้ และไม่สามารถที่จะบ่งชี้ได้ชัดเจนว่าแบบสอบคู่ขนานกัน จึงใช้วิธีการประมาณความเที่ยงโดยวิธีอื่น ๆ ซึ่งมี 5 วิธี ดังนี้ (อนาสตาซี 2519: 79)

1. ความเที่ยงแบบสอบซ้ำ
2. ความเที่ยงแบบสลับฟอร์ม
3. ความเที่ยงชนิดแบ่งครึ่ง
4. ความเที่ยงแบบคูเดอร์-ริชาร์ดสัน
5. ความเที่ยงของผู้ให้คะแนน

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

014007

1.2 <u>ทฤษฎีการตอบสนองข้อสอบ</u>

ทฤษฎีการตอบสนองข้อสอบได้เข้ามามีบทบาทในวงการวัดและประเมินผลการศึกษาตั้งแต่ปี ค.ศ. 1936 และมีบทบาทมากขึ้นเรื่อย ๆ ในปัจจุบัน เนื่องจากพบว่าทฤษฎีคลาสสิคอลมีจุดอ่อนหลายประการดังนี้ (Hambleton and Swaminathan 1985: 1-3)

1. ค่าสถิติของข้อสอบ ได้แก่ ค่าความยากของข้อสอบ และค่าอำนาจจำแนกของข้อสอบ ขึ้นอยู่กับกลุ่มตัวอย่างผู้สอบ หมายความว่า ค่าสถิติของข้อสอบดังกล่าวมีประโยชน์ในการสร้างแบบสอบเฉพาะแต่กลุ่มประชากรผู้สอบที่มีลักษณะคล้ายคลึงกับกลุ่มตัวอย่างผู้สอบที่ใช้ในการหาค่าสถิติเหล่านั้น

2. การเปรียบเทียบผู้สอบในเรื่องของความสามารถที่ระดับหนึ่ง ซึ่งวัดจากแบบสอบถูกจำกัดอยู่กับสถานการณ์ที่ผู้สอบถูกทดสอบด้วย นั่นคือใช้ข้อสอบชุดเดียวกันหรือคู่ขนานกัน โดยส่วนใหญ่เป็นแบบสอบวัดผลสัมฤทธิ์และความถนัดเฉพาะนักเรียนที่มีความสามารถปานกลาง ซึ่งแบบสอบเหล่านี้ไม่ให้การประมาณค่าความสามารถที่ถูกต้องแม่นยำสำหรับผู้สอบที่มีความสามารถระดับสูงหรือต่ำ

3. ความเที่ยงของแบบสอบที่ได้นิยามในลักษณะของแบบสอบคู่ขนานซึ่งมโนทัศน์ของการวัดแบบนี้ยากที่จะสัมฤทธิ์ผลได้ในการปฏิบัติ เพราะผู้สอบแต่ละคนไม่มีผลการสอบเหมือนเดิมของการสอบครั้งที่สอง ซึ่งอาจเกิดจากการลืม การพัฒนาทักษะใหม่ ๆ การเปลี่ยนแปลงของระดับแรงจูงใจหรือความวิตกกังวล เป็นต้น

4. เทคนิคการสอบไม่ได้ให้พื้นฐานเพื่อกำหนดว่า ผู้สอบคนหนึ่งจะสามารถทำข้อสอบได้ดีเพียงใด เมื่อได้เผชิญกับข้อสอบข้อหนึ่ง ซึ่งควรจะมีการประมาณความน่าจะเป็นที่ผู้สอบจะตอบข้อสอบได้ถูกต้อง ด้วยค่าที่เหมาะสมตรงกับระดับความสามารถของผู้สอบ ซึ่งเป็นสารสนเทศที่จำเป็น

5. ความแปรปรวนของความคลาดเคลื่อนของการวัดของผู้สอบแต่ละคนมีค่าเท่ากัน ซึ่งความแปรปรวนของความคลาดเคลื่อนของการวัดแปรเปลี่ยนไปตามความสามารถกล่าวคือ ค่าความแปรปรวนของความคลาดเคลื่อนของการวัดสำหรับผู้สอบที่มีความสามารถสูง จะมีค่าน้อยกว่าผู้สอบที่มีความสามารถปานกลาง

ทฤษฎีการตอบสนองข้อสอบ เป็นทฤษฎีที่อธิบายถึงความสัมพันธ์ระหว่างการตอบข้อสอบของผู้สอบกับระดับความสามารถที่มีอยู่ภายในตัวผู้สอบ ซึ่งความสัมพันธ์ดังกล่าวนี้อธิบายได้โดยแสดงในลักษณะโมเดลทางคณิตศาสตร์ ที่กำหนดด้วยฟังก์ชั่นทางคณิตศาสตร์ซึ่งเรียกว่า ฟังก์ชั่นลักษณะข้อสอบ (Item Characteristic Function) หรือเรียกว่า โค้งลักษณะข้อสอบ (Item Characteristic Curve) และสำหรับโมเดลทางคณิตศาสตร์ที่นำมาใช้อธิบายในทฤษฎีการตอบสนองข้อสอบมีหลายโมเดลด้วยกัน แต่ละโมเดลแตกต่างกันด้วยรูปแบบของฟังก์ชั่นทางคณิตศาสตร์ที่ใช้ และด้วยจำนวนของพารามิเตอร์ที่ใช้อธิบายโค้งลักษณะข้อสอบ

<u>โมเดลทางคณิตศาสตร์ในทฤษฎีการตอบสนองข้อสอบ</u>

โมเดลทางคณิตศาสตร์ที่สำคัญที่อธิบายโค้งลักษณะข้อสอบในทฤษฎีการตอบสนองข้อสอบ ที่ได้รับการพัฒนาและมีการนำไปใช้มากที่สุดในสถานการณ์ของการสอบ คือ โมเดลโลจิสติก (Logistic Model) เนื่องจากมีความสะดวกในการใช้ ซึ่งแบ่งออกเป็น 4 โมเดลตามจำนวนพารามิเตอร์ของโค้งลักษณะข้อสอบดังนี้

1. <u>โมเดลโลจิสติกที่มีพารามิเตอร์ 3 ตัว</u> (Three-Parameter Logist Model)

เป็นโมเดลที่มีพารามิเตอร์อธิบายโค้งลักษณะข้อสอบ 3 ตัว ซึ่งมีรูปสมการดังนี้

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

เมื่อ $P_i(\theta)$ คือ ความน่าจะเป็นที่ผู้สอบที่มีระดับความสามารถ θ ตอบข้อสอบข้อ i ถูกต้อง

θ คือ ระดับความสามารถของผู้สอบที่ประมาณได้จากคะแนนรวมของการตอบแบบสอบ โดยการปรับให้เป็นคะแนนที่มีค่าเฉลี่ยเท่ากับ 0 และมีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1 โดยมีพิสัยการกระจายอยู่ระหว่าง $-\infty$ ถึง $+\infty$ แต่ในทางปฏิบัติระดับความสามารถจะมีค่าอยู่ระหว่าง -3

ถึง +3 ซึ่งระดับความสามารถ -3 หมายถึง ผู้สอบมีระดับความสามารถต่ำมาก และระดับความสามารถ +3 หมายถึง ผู้สอบที่มีระดับความสามารถสูงมาก

a_i คือ ค่าอำนาจจำแนกของข้อสอบข้อ i ที่เป็นสัดส่วนโดยตรงกับความชันของโค้งลักษณะข้อสอบ ณ จุดเปลี่ยนโค้ง หรือที่จุด $\theta = b_i$ โดยทั่วไปพิสัยของค่าอำนาจจำแนกของข้อสอบอยู่ระหว่าง 0 ถึง 2

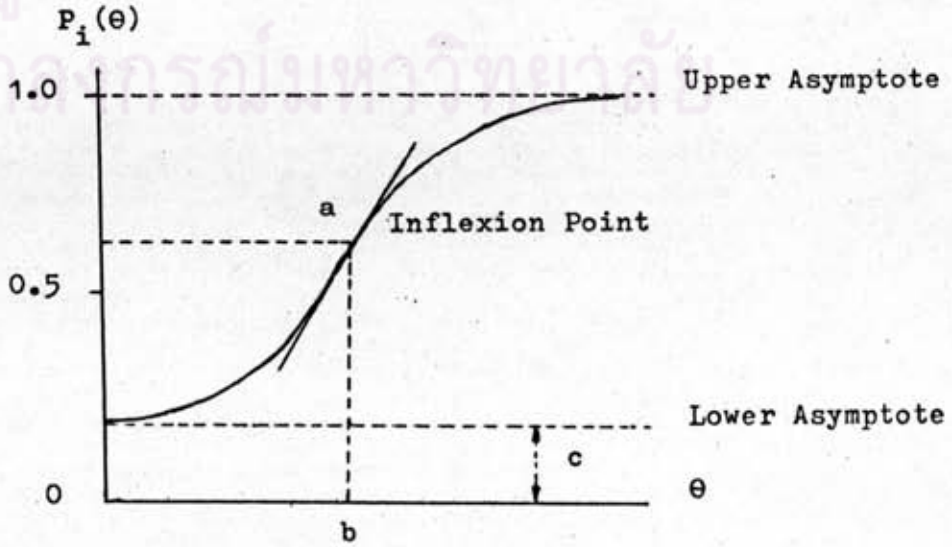
b_i คือ ค่าความยากของข้อสอบข้อ i ที่แสดงถึงระดับความสามารถ ที่จุดโค้งลักษณะข้อสอบมีความชันมากที่สุด โดยทั่วไปมีพิสัยของค่าความยากของข้อสอบอยู่ระหว่าง -2 ถึง +2 ซึ่งค่าที่ใกล้ -2 หมายถึง ข้อสอบที่ง่ายมาก และค่าที่ใกล้ +2 หมายถึง ข้อสอบที่ยากมาก

c_i คือ ค่าการเดาของข้อสอบข้อ i ซึ่งเป็นค่าความน่าจะเป็นที่ผู้สอบมีความสามารถต่ำมาก จะมีโอกาสตอบข้อสอบข้อ i ถูก หรือเป็นค่ากำกับโค้งที่ต่ำสุด (Lower asymptote) ของโค้งลักษณะข้อสอบมีค่า 0 ถึง 1

D คือ a scaling factor มีค่าเท่ากับ 1.7

e คือ ค่าคงที่ มีค่าเท่ากับ 2.71828...

ภาพที่ 1 ความหมายของค่าพารามิเตอร์ของข้อสอบ 3 พารามิเตอร์

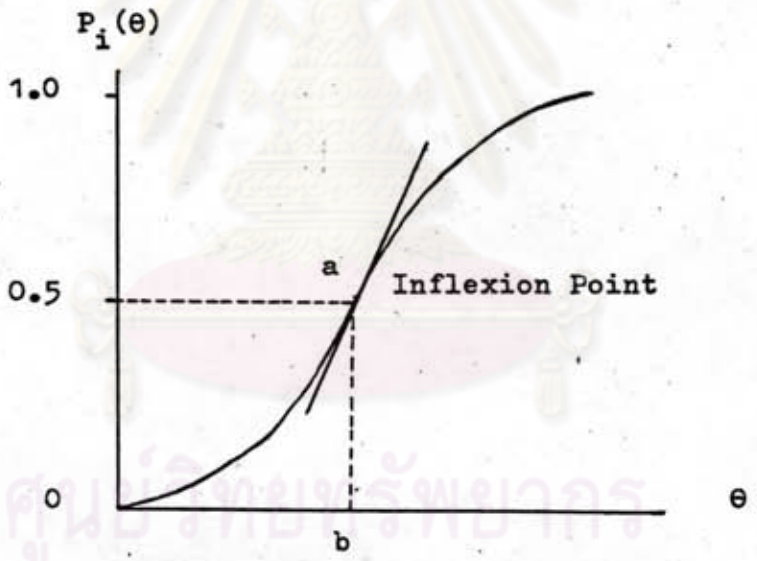


2. <u>โมเดลโลจิสติกที่มีพารามิเตอร์ 2 ตัว</u> (Two-Parameter Logistic Model)

เป็นโมเดลที่กำหนดให้ค่าการเดาของข้อสอบเป็น 0 รูปสมการก็จะเหลือค่าพารามิเตอร์เพียง 2 ตัว คือ ค่าอำนาจจำแนกของข้อสอบ และค่าความยากของข้อสอบ ดังนี้

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

ภาพที่ 2 ความหมายของค่าพารามิเตอร์ของข้อสอบ 2 พารามิเตอร์

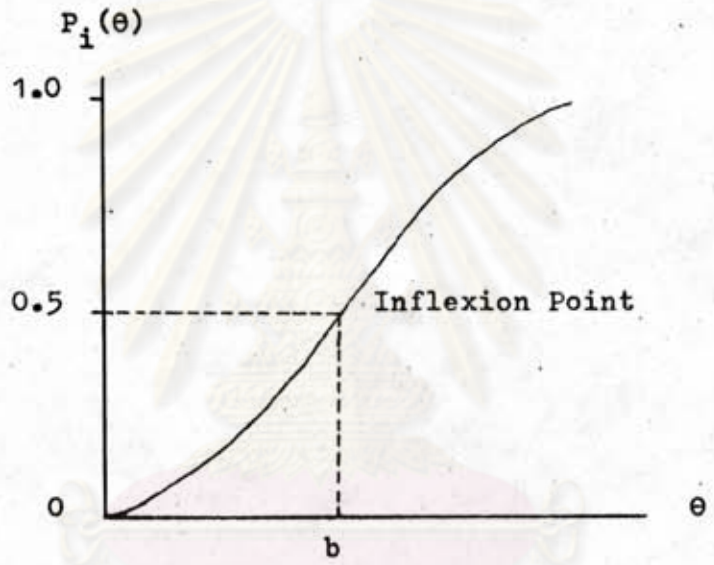


3. <u>โมเดลโลจิสติกที่มีพารามิเตอร์ 1 ตัว</u> (One-Parameter Logistic Model)

เป็นโมเดลที่มีข้อตกลงเบื้องต้น คือ ค่าอำนาจจำแนกของข้อสอบทุกข้อมีค่าเท่ากัน และค่าการเดาของข้อสอบเท่ากับ 0 รูปสมการก็จะเหลือค่าพารามิเตอร์เพียง 1 ตัว ที่มีค่าแปรเปลี่ยน ดังนี้

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

ภาพที่ 3 ความหมายของค่าพารามิเตอร์ของข้อสอบ 1 พารามิเตอร์



4. <u>โมเดลโลจิสติกที่มีพารามิเตอร์ 4 ตัว</u> (Four-Parameter Logistic Model)

เป็นโมเดลที่แตกต่างจากโมเดลโลจิสติกที่มีพารามิเตอร์ 3 ตัว โดยเพิ่มพารามิเตอร์อีก 1 ตัว คือ γ_i ซึ่งมีค่าต่ำกว่า 1 เล็กน้อย ซึ่งหมายถึง ค่าที่ผู้สอบที่มีความสามารถสูงมีความสะเพร่าตอบข้อสอบไม่ถูก แต่เป็นโมเดลที่ยังไม่สามารถหาวิธีปฏิบัติได้ โดยมีรูปสมการดังนี้

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

หอสมุดกลาง สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบ

1. มิติของลาเทนท์สเปส (Dimensionality of the Latent Space)

ลาเทนท์สเปส หมายถึง คุณลักษณะภายในหรือความสามารถที่เป็นลักษณะกำหนดพฤติกรรมการตอบข้อสอบของแบบสอบทั้งหมด จำนวนคุณลักษณะภายในทั้งหมดของลาเทนท์สเปส หมายถึง มิติของลาเทนท์สเปส โดยทั่วไปทฤษฎีการตอบสนองข้อสอบมีข้อตกลงเบื้องต้นว่า ลาเทนท์สเปสมีเพียงมิติเดียว (Unidimensionality) นั่นคือมีข้อตกลงเบื้องต้นว่า ข้อสอบในแบบสอบมีลักษณะเป็นเอกพันธ์ (Homogeneous) ที่ได้วัดคุณลักษณะภายในหรือความสามารถเพียงอย่างเดียว หรือมีคุณสมบัติการวัดเพียงมิติเดียว การตรวจสอบคุณสมบัติการวัดเพียงมิติเดียวนี้ สามารถใช้การวิเคราะห์ตัวประกอบตรวจสอบ

2. ความเป็นอิสระในการตอบข้อสอบ (Local Independence)

ข้อตกลงเบื้องต้นข้อนี้กำหนดว่า การตอบสนองต่อข้อสอบต่าง ๆ กันในแบบสอบของผู้สอบคนใดคนหนึ่งมีความเป็นอิสระในทางสถิติ นั่นคือการตอบข้อสอบข้อใดข้อหนึ่งของผู้สอบคนหนึ่งต้องไม่มีผลต่อการตอบข้อสอบข้ออื่น ๆ ในแบบสอบ และเป็นข้อตกลงเบื้องต้นที่สมมูลกับข้อตกลงเบื้องต้นของคุณสมบัติการวัดเพียงมิติเดียว นั่นคือถ้ามีคุณสมบัติการวัดเพียงมิติเดียวก็จะมีความเป็นอิสระในการตอบข้อสอบ และถ้ามีความเป็นอิสระในการตอบข้อสอบก็จะมีคุณสมบัติการวัดเพียงมิติเดียว ดังนั้นการตรวจสอบความเป็นอิสระในการตอบข้อสอบก็ใช้การวิเคราะห์ตัวประกอบด้วยเหมือนกัน

3. โค้งลักษณะข้อสอบ (Item Characteristic Curve)

โค้งลักษณะข้อสอบเป็นฟังก์ชั่นทางคณิตศาสตร์ ที่แสดงถึงความสัมพันธ์ระหว่างความน่าจะเป็นของการตอบข้อสอบข้อหนึ่งได้ถูกต้องกับระดับความสามารถที่วัดโดยข้อสอบข้อนั้น ซึ่งเป็นฟังก์ชั่นการถดถอยชนิดไม่เป็นเชิงเส้นตรง (Non-Linear Regression Function) ของคะแนนข้อสอบบนระดับความสามารถที่วัดโดยแบบสอบ โค้งลักษณะข้อสอบมีหลายรูปแบบแตกต่างกันขึ้นอยู่กับโมเดลการตอบสนองข้อสอบ ในรูปแบบทางคณิตศาสตร์ที่ใช้อธิบายความสัมพันธ์ดังกล่าว

ดัชนีที่กำหนดคุณภาพและเปรียบเทียบคุณภาพของแบบสอบ

ทฤษฎีการตอบสนองข้อสอบพิจารณาคุณภาพของแบบสอบจากดัชนีที่กำหนดคุณภาพ คือ ค่าฟังก์ชั่นสารสนเทศแบบสอบ ซึ่งเป็นฟังก์ชั่นของค่าพารามิเตอร์ของข้อสอบที่เป็นดัชนีผสมโดยสร้างมาจากดัชนีที่กำหนดคุณภาพของข้อสอบหลาย ๆ ดัชนีด้วยกันของทุก ๆ ข้อในแบบสอบ โดยกำหนดว่า ณ ระดับความสามารถใดที่ค่าฟังก์ชั่นสารสนเทศแบบสอบมีค่าสูง แสดงว่าแบบสอบมีคุณภาพดีมากในการประมาณค่าความสามารถที่ระดับนั้น และดัชนีที่ใช้เปรียบเทียบคุณภาพของแบบสอบ คือ ค่าประสิทธิภาพสัมพัทธ์ของแบบสอบ โดยนำค่าฟังก์ชั่นสารสนเทศแบบสอบมาหาค่าอัตราส่วนกัน ณ ระดับความสามารถเดียวกัน ซึ่งดัชนีทั้งสองนี้มีรายละเอียดดังนี้

ค่าฟังก์ชั่นสารสนเทศแบบสอบ

สารสนเทศสามารถนิยามได้จากลักษณะของความไม่แน่นอนของเหตุการณ์ที่เกิดขึ้น ถ้าทราบว่าจะมีเหตุการณ์เกิดขึ้นค่อนข้างแน่นอน ก็ย่อมมีสารสนเทศเกี่ยวกับเหตุการณ์นั้นมาก ซึ่งเป็นมโนทัศน์ทั่วไปที่ได้นำมาใช้ในทฤษฎีการตอบสนองข้อสอบ โดยได้พิจารณาในกระบวนการอ้างอิงเชิงสถิติที่เกี่ยวกับค่าพารามิเตอร์ ความไม่แน่นอนของการประมาณค่าพารามิเตอร์ของกลุ่มประชากรพิจารณาในช่วง $-\infty$ ถึง $+\infty$ ถ้าเรามีสารสนเทศเกี่ยวกับกลุ่มประชากรบ้างแล้ว ช่วงของการประมาณค่าพารามิเตอร์จะลดลงหรือแคบเข้า หรือการประมาณค่ามีความแน่นอนมากขึ้น ความไม่แน่นอนของการประมาณค่าพารามิเตอร์สามารถแสดงได้ด้วยความแปรปรวนของกลุ่มตัวอย่างที่ใช้ประมาณค่าพารามิเตอร์หรือความกว้างของช่วงความเชื่อมั่นของค่าพารามิเตอร์ ถ้าเราประมาณค่าเฉลี่ยของกลุ่มประชากรแล้วความแปรปรวนของค่าเฉลี่ยหรือความคลาดเคลื่อนมาตรฐานกำลังสองของค่าเฉลี่ยเป็น $\sigma^2_{\bar{x}} = \sigma^2_x/N$ ถ้าความแปรปรวนนี้มีค่าน้อยก็มีผลทำให้สารสนเทศมีมากขึ้น ดังนั้นโดยทั่วไปจึงได้นิยามสารสนเทศเกี่ยวกับค่าพารามิเตอร์ว่าเป็นสัดส่วนกลับกับความแปรปรวนหรือความคลาดเคลื่อนมาตรฐานกำลังสองดังนี้

$$I_{\bar{x}} = \frac{1}{\sigma^2_{\bar{x}}}$$

สารสนเทศมีความสำคัญมากในทฤษฎีการตอบสนองข้อสอบ เพราะแบบสอบและสเกลที่พัฒนาขึ้นโดยการใช้ทฤษฎีนี้สามารถประเมินได้จากความคงที่แน่นอนของการประมาณ θ (ค่าความสามารถ) ที่กำหนดไว้ ดังนั้นสารสนเทศจึงเป็นสัดส่วนกลับกับความคลาดเคลื่อนมาตรฐานกำลังสองของ θ สารสนเทศที่ θ ใด ๆ ขึ้นอยู่กับจำนวนและลักษณะของข้อสอบที่ใช้ประมาณ θ เมื่อสารสนเทศที่ระดับความสามารถใด ๆ มีค่าสูง เราก็มีช่วงการประมาณค่าความสามารถนั้นแคบขึ้น ถ้าเรามีสารสนเทศน้อยเราก็มีช่วงการประมาณกว้างขึ้น เราไม่มีสารสนเทศเกี่ยวกับความสามารถของแต่ละบุคคลก่อนที่จะมีการตอบสนองข้อสอบ แต่หลังจากการตอบสนองข้อสอบแล้วเราสามารถประมาณ θ ได้ และสามารถหาค่าสารสนเทศได้ สารสนเทศเกี่ยวกับ θ นั้นกำหนดได้โดย

$$I(\theta) = \frac{1}{\sigma^2_{\hat{\theta}/\theta}}$$

หรือสารสนเทศเกี่ยวกับ θ เป็นสัดส่วนกลับกับความคลาดเคลื่อนมาตรฐานกำลังสองของ $\hat{\theta}$ จาก θ ที่กำหนด ความคลาดเคลื่อนมาตรฐานของ $\hat{\theta}$ จาก θ ที่กำหนด หมายถึง ส่วนเบี่ยงเบนมาตรฐานของการกระจายของ $\hat{\theta}$ สำหรับแต่ละบุคคลทั้งหมดกับ θ ดังนั้นจึงสามารถแสดงความสัมพันธ์ระหว่างสารสนเทศและความคลาดเคลื่อนมาตรฐานของ $\hat{\theta}$ จาก θ ที่กำหนดได้ดังนี้

$$\sigma_{\hat{\theta}/\theta} = \frac{1}{\sqrt{I(\theta)}}$$

เบิร์นบอม (Birnbaum) (1968) ได้ให้นิยามฟังก์ชั่นสารสนเทศ $I(\theta,y)$ สำหรับคะแนน y ใด ๆ ว่าเป็นสัดส่วนกลับกับกำลังสองของความกว้างของช่วงความเชื่อมั่น สำหรับการประมาณค่าความสามารถ θ จากคะแนน y (Lord 1980: 65)

กำหนดให้ $Z = x/n$ เป็นคะแนนสัดส่วนการตอบถูกที่วัดได้ (สัดส่วนของการตอบข้อสอบถูก n ข้อ) การถดถอย (Regression) ของ Z บนความสามารถ θ มีสมการเป็น

$$\mu_{Z/\theta} = \frac{1}{n} \sum_{i=1}^{n} P_i(\theta) = \zeta$$

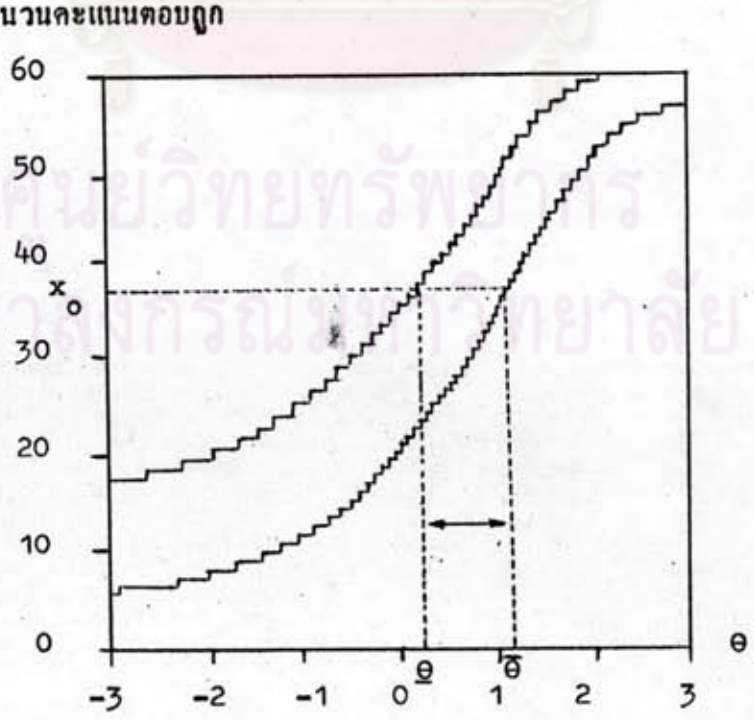
การถดถอยนี้ไม่เปลี่ยนแปลงไปตามความยาวของแบบสอบ และความแปรปรวนของ Z สำหรับ θ ที่คงที่ใด ๆ มีสมการเป็น

$$\sigma^2_{Z/\theta} = \frac{1}{n^2} \sum_{i=1}^{n} P_i Q_i = \frac{1}{n}(\bar{P}\bar{Q} - \sigma^2_{P/\theta})$$

ความแปรปรวนนี้มีค่าเข้าใกล้ 0 เมื่อ n มีขนาดใหญ่

พิจารณาจากภาพดังต่อไปนี้

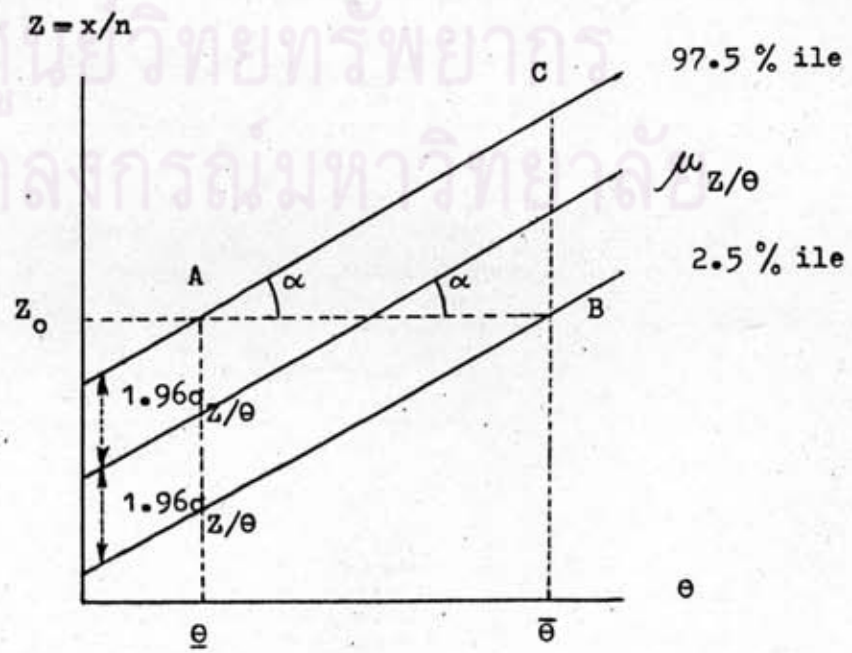
ภาพที่ 4 ช่วงความเชื่อมั่น $(\underline{\theta}, \bar{\theta})$ สำหรับการประมาณค่าความสามารถ (แบบสอบคณิตศาสตร์ SAT, January 1971)



จากภาพที่ 4 แกนตั้งแทนด้วยสเกลจำนวนคะแนนตอบถูก x โดยสูตร $Z = x/n$ ซึ่งมีพิสัยจาก Z=0 ถึง Z=1 การถดถอย $\mu_{Z/\theta}$ ของ Z บน θ อาจจะมองเห็นเป็นโค้งรูป Ogive ซึ่งอยู่ตรงกลางระหว่างฟังก์ชั่นแบบชั้นบันได 2 ฟังก์ชั่น ฟังก์ชั่นทั้งสองนี้ประมาณค่าเป็นเปอร์เซ็นต์ไทล์ที่ $2\frac{1}{2}$ และ $97\frac{1}{2}$ ของการแจกแจงของ Z จาก θ ที่กำหนด ถ้า n มีขนาดมากขึ้นการถดถอยนี้ไม่เปลี่ยนแปลง แต่ $\sigma^2_{Z/\theta}$ ลดน้อยลงเข้าสู่ 0 ดังนั้นฟังก์ชั่นแบบชั้นบันไดทั้งสองจะเลื่อนเข้าหาโค้งการถดถอย ในขณะเดียวกันจำนวนชั้นบันไดเพิ่มมากขึ้นจนกระทั่งฟังก์ชั่นแบบชั้นบันไดมีลักษณะการประมาณเป็นโค้งเรียบยิ่งขึ้น เมื่อการแจกแจงแบบมีเงื่อนไขของ Z จาก θ ที่กำหนดมีลักษณะการประมาณเป็นแบบปกติ ระยะห่างจากโค้งการถดถอยถึงตำแหน่งเปอร์เซ็นต์ไทล์ที่ $2\frac{1}{2}$ และถึงตำแหน่งเปอร์เซ็นต์ไทล์ที่ $97\frac{1}{2}$ จะมีค่าเท่ากับ 1.96 ของส่วนเบี่ยงเบนมาตรฐานหรือเท่ากับ $(1.96/n)\sqrt{\Sigma_i^n P_i Q_i}$

ช่วงความเชื่อมั่นที่ตรงกับ Z_o คือ $(\underline{\theta}, \bar{\theta})$ ความกว้างของช่วงความเชื่อมั่นคือ ระยะทาง AB เป็นการอธิบายถึงประสิทธิภาพของการวัดความสามารถของแบบสอบ พิจารณารูปสามเหลี่ยม ABC จากภาพดังนี้

ภาพที่ 5 โครงสร้างของช่วงความเชื่อมั่น 95 % $(\underline{\theta}, \bar{\theta})$ สำหรับความสามารถ θ



เราจะได้ว่า

$$\tan\alpha = \frac{\overline{CB}}{\overline{AB}} = \frac{2(1.96\sigma_{Z/\theta})}{\overline{AB}}$$

หรือ

$$\overline{AB} = \frac{3.92\sigma_{Z/\theta}}{\tan\alpha}$$

เนื่องจาก $\tan\alpha$ เป็นความชันของเส้นถดถอย $\mu_{Z/\theta}$ จากนิยามของฟังก์ชั่นสารสนเทศที่ว่า เป็นสัดส่วนกลับกับกำลังสองของความกว้างของช่วงความเชื่อมั่นสำหรับการประมาณค่าความสามารถ θ จากคะแนน y ใด ๆ ดังนั้นฟังก์ชั่นสารสนเทศสำหรับคะแนน Z คือ สัดส่วนกับ

$$\frac{1}{\overline{AB}^2} \equiv \frac{(\frac{d}{d\theta}\mu_{Z/\theta})^2}{(3.92)^2 Var(Z/\theta)}$$

จากภาพที่ 5 ได้แสดงการประมาณความสามารถจากคะแนนสัดส่วนการตอบถูก Z สำหรับแบบสอบที่มีลักษณะเป็นมิติเดียวก็สามารถประยุกต์ใช้ได้กับคะแนนแบบสอบ ดังนั้น เบิร์นบอม จึงได้นิยามฟังก์ชั่นสารสนเทศสำหรับคะแนน y ใด ๆ ดังนี้

$$I(\theta,y) \equiv \frac{(\frac{d}{d\theta}\mu_{y/\theta})^2}{Var(y/\theta)}$$

นั่นคือฟังก์ชั่นสารสนเทศสำหรับคะแนน y ใด ๆ นิยามว่า เป็นกำลังสองของอัตราส่วนของความชันของการถดถอยของ y บน θ กับความคลาดเคลื่อนมาตรฐานของการวัดคะแนน y สำหรับ θ ที่คงที่

และเมื่อไรก็ตามที่ θ คงที่ คะแนนจริง η ที่สัมพันธ์กับคะแนนสอบ y ก็มีค่าคงที่ด้วย และ $Var(y/\theta)$ ก็มีค่าเท่ากับกำลังสองของความคลาดเคลื่อนมาตรฐานของการวัด $\sigma^2_{y/n}$ ที่ได้ใช้กันอยู่

สารสนเทศที่ได้จากคะแนน y สำหรับการประมาณ θ มีค่าแปรเปลี่ยนตามระดับความสามารถ θ ที่แตกต่างกัน การแปรเปลี่ยนนี้มาจากสาเหตุ 2 ประการ คือ

1. เมื่อความคลาดเคลื่อนมาตรฐานของการวัด $\sigma_{y/n}$ มีค่าลดน้อยลงจะทำให้ได้สารสนเทศ y เกี่ยวกับ θ มากขึ้น

2. เมื่อความชันของการถดถอย $\mu_{y/\theta}$ มากขึ้น (y จะแปรเปลี่ยนกับ θ อย่างมาก) จะทำให้ได้สารสนเทศ y เกี่ยวกับ θ มากขึ้น

ตัวประมาณความน่าจะเป็นสูงสุด $\hat{\theta}$ (The Maximum Likelihood Estimator) เป็นคะแนนการสอบชนิดหนึ่ง ดังนั้นเราสามารถใช้สมการ

$$I(\theta,y) \equiv \frac{(\frac{d}{d\theta}\mu_{y/\theta})^2}{Var(y/\theta)}$$

เพื่อใช้ในการหาฟังก์ชั่นสารสนเทศของตัวประมาณความน่าจะเป็นสูงสุด ซึ่งเราต้องมีสูตรสำหรับการถดถอย $\mu_{\hat{\theta}/\theta}$ และสูตรสำหรับความแปรปรวน $\sigma^2_{\hat{\theta}/\theta}$

ตัวประมาณความน่าจะเป็นสูงสุด $\hat{\theta}$ ของพารามิเตอร์ θ มีลักษณะการแจกแจงแบบปกติ ที่มีค่าเฉลี่ย θ_o (เป็นค่าพารามิเตอร์จริงที่ไม่ทราบค่า) และมีความแปรปรวนเป็น

$$Var(\hat{\theta}/\theta_o) = \frac{1}{\mathcal{E}\left[(\frac{d \ln L}{d\theta})^2_{\theta_o}\right]}$$

เมื่อ L เป็น ฟังก์ชั่นความน่าจะเป็น (Likelihood Function)

และเมื่อหราบค่าพารามิเตอร์ของข้อสอบ เราได้ว่า

$$\frac{1}{\mathrm{Var}(\hat{\theta}/\theta_o)} = \mathcal{E}\left\{\left[\sum_{i=1}^{n}(u_i-P_i)P_i/P_iQ_i\right]^2/\theta_o\right\}$$

$$= \mathcal{E}\left\{\left[\sum_{i=1}^{n}(u_i-P_i)P_i/P_iQ_i\right]\cdot\left[\sum_{i=1}^{n}(u_j-P_j)P_j/P_jQ_j\right]/\theta_o\right\}$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n}\frac{P'_{io}P'_{jo}}{P_{io}P_{jo}Q_{io}Q_{jo}}\mathcal{E}\left[(u_i-P_i)(u_j-P_j)/\theta_o\right]$$

และเนื่องจาก $\mathcal{E}(u_i/\theta_o) = P_{io}$ ดังนั้นค่าความคาดหวังภายใต้เครื่องหมายรวมเป็นความแปรปรวนร่วม และเพราะว่ามีความเป็นอิสระในการตอบข้อสอบ u_i จึงมีการแจกแจงที่เป็นอิสระกับ u_j สำหรับ θ ที่คงที่ มีผลทำให้ความแปรปรวนร่วมเป็น 0 ยกเว้นเมื่อ $i = j$ ลักษณะนี้ก็เป็นความแปรปรวนนั่นเอง ดังนั้น

$$\frac{1}{\mathrm{Var}(\hat{\theta}/\theta_o)} = \sum_{i=1}^{n}\frac{P_i'^2}{P_i^2Q_i^2}\mathrm{Var}(u_{io}/\theta_o)$$

$$= \sum_{i=1}^{n}\frac{P_{io}'^2}{P_{io}^2Q_{io}^2}P_{io}Q_{io}$$

เมื่อตัดตัวห้อยท้าย o ออก สูตรสำหรับตัวประมาณความน่าจะเป็นสูงสุดของความแปรปรวน คือ

$$\mathrm{Var}(\hat{\theta}/\theta) = \frac{1}{\sum_{i=1}^{n}\frac{P_i'^2}{P_iQ_i}}$$

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

และ $\hat{\theta}$ เป็นตัวประมาณที่ลงที่ ดังนั้น $\mu_{\hat{\theta}/\theta} \longrightarrow \theta$ ซึ่งทำให้จำนวนของฟังก์ชั่นสารสนเทศสำหรับคะแนน $\hat{\theta}$ คือ

$$\left(d\mu_{\hat{\theta}/\theta}/d\theta \right)^2 = 1$$

ดังนั้นฟังก์ชั่นสารสนเทศของตัวประมาณความน่าจะเป็นสูงสุดของ θ ก็คือ สัดส่วนกลับกับความแปรปรวน ซึ่งเรียกว่า ฟังก์ชั่นสารสนเทศแบบสอบ มีสมการเป็น

$$I(\theta) = I(\theta,\hat{\theta}) = \sum_{i=1}^{n} \frac{P_i'^2}{P_i Q_i}$$

เมื่อ P_i คือ ความน่าจะเป็นที่ผู้สอบที่มีระดับความสามารถ θ ตอบข้อสอบข้อ i ถูกต้อง

Q_i คือ $1 - P_i$

P_i' คือ ความชันของโค้งลักษณะข้อสอบที่ระดับความสามารถ θ

ลอร์ด (Lord) ได้สรุปฟังก์ชั่นสารสนเทศแบบสอบเป็นทฤษฎี 2 ทฤษฎี ดังนี้

<u>ทฤษฎีที่ 1</u> ฟังก์ชั่นสารสนเทศสำหรับตัวประมาณที่ไม่ลำเอียง (คงที่) ของความสามารถเป็นสัดส่วนกลับกับความแปรปรวนของตัวประมาณนั้น

<u>ทฤษฎีที่ 2</u> ฟังก์ชั่นสารสนเทศแบบสอบ เป็นขอบเขตบนของสารสนเทศซึ่งสามารถได้รับโดยวิธีการให้คะแนนแบบสอบวิธีต่าง ๆ

จากลักษณะที่สำคัญของสมการฟังก์ชั่นสารสนเทศแบบสอบ คือ สารสนเทศแบบสอบมีลักษณะเป็นอิสระและเป็นผลรวมที่สนับสนุน (Contribution) จากข้อสอบต่าง ๆ ในแบบสอบ การสนับสนุนของข้อสอบแต่ละข้อไม่ขึ้นกับข้อสอบข้ออื่น ๆ ที่รวมอยู่ในแบบสอบ การสนับสนุนของข้อสอบแต่ละข้อคือ $P_i'^2/P_i Q_i$ ซึ่งเรียกว่า ฟังก์ชั่นสารสนเทศข้อสอบ มีสมการดังนี้

$$I(\theta,u_i) = \frac{P_i'^2}{P_iQ_i}$$

จึงกล่าวได้ว่า ฟังก์ชั่นสารสนเทศแบบสอบเป็นผลรวมของฟังก์ชั่นสารสนเทศข้อสอบทั้งหมดในแบบสอบ (Warm 1978: 73) เขียนเป็นสมการได้ดังนี้

$$I(\theta) = \sum_{i=1}^{n} I(\theta,u_i)$$

ซึ่งค่าต่าง ๆ ของฟังก์ชั่นสารสนเทศแบบสอบ ในโมเดลโลจิสติกที่มีพารามิเตอร์ 3 ตัว สามารถคำนวณได้จากสูตรดังนี้ (Lord 1980: 60-61)

$$P_i' = \frac{1.7a_i(1 - c_i)}{e^{1.7a_i(\theta-b_i)} + 2 + e^{-1.7a_i(\theta-b_i)}}$$

$$P_i = \frac{c_i + e^{1.7a_i(\theta-b_i)}}{1 + e^{1.7a_i(\theta-b_i)}}$$

$$Q_i = \frac{1 - c_i}{1 + e^{1.7a_i(\theta-b_i)}}$$

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ในการสอบทุก ๆ แบบสอบที่ใช้ในการวัดย่อมมีความคลาดเคลื่อนเกิดขึ้น และจากนิยามของฟังก์ชั่นสารสนเทศแบบสอบซึ่งเป็นสัดส่วนกลับกับความแปรปรวนหรือความคลาดเคลื่อนมาตรฐานกำลังสองของการประมาณค่าความสามารถ โดยที่ความคลาดเคลื่อนมาตรฐานของการประมาณค่าความสามารถ (The Standard Error of Estimate Ability (S.E.E.))หมายถึง ส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนการประมาณค่าความสามารถ ถ้าเราสอบกับกลุ่มผู้สอบที่มีความสามารถต่างกัน การประมาณค่าความสามารถของผู้สอบแต่ละแบบสอบ จะมีความคลาดเคลื่อนมาตรฐานของการประมาณค่าความสามารถ ซึ่งมีค่าเท่ากับ รากที่สองของส่วนกลับของฟังก์ชั่นสารสนเทศแบบสอบ เขียนในรูปสมการดังนี้ (Warm 1978: 76-77)

$$S.E.E. = \frac{1}{\sqrt{I(\theta)}}$$

ฟังก์ชั่นสารสนเทศแบบสอบจะมีค่าแปรเปลี่ยนไปตามสเกลของความสามารถ นั่นคือ ถ้าฟังก์ชั่นสารสนเทศแบบสอบมีค่ามาก จะทำให้ค่า S.E.E. มีค่าน้อย ซึ่งเป็นค่าที่เราต้องการมากในการสอบ นอกจากนี้ค่าเฉลี่ยของ S.E.E. ($\overline{S.E.E.}$) ของผู้สอบทั้งหมดมีความสัมพันธ์กับค่าความเที่ยงตามกรอบของทฤษฎีคลาสสิคอลด้วย เมื่อคะแนนเป็นมาตรฐานที่มีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1 ซึ่งเราสามารถคำนวณค่าความเที่ยงได้จากสูตรดังนี้ (Warm 1978: 77)

$$r_{tt} = 1 - (\overline{S.E.E.})^2$$

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

<u>ค่าประสิทธิภาพสัมพัทธ์ของแบบสอบ</u>

จากแนวคิดของทฤษฎีการตอบสนองข้อสอบที่มีจุดมุ่งหมายในการสอบ คือ เมื่อใช้แบบสอบวัดความสามารถของแต่ละบุคคล แล้วต้องการให้ผลการสอบนั้นไปประมาณค่าความสามารถที่มีอยู่ภายในตัวบุคคลที่ไม่สามารถสังเกตได้หรือวัดได้โดยตรง ดังนั้น

แบบสอบที่มีคุณภาพที่ดีควรเป็นแบบสอบที่ให้ผลการสอบสามารถนำไปใช้ประมาณค่าความสามารถที่ต้องการวัดได้ถูกต้องแม่นยำมากที่สุด ดัชนีที่ใช้ประมาณค่าความสามารถที่ต้องการวัดได้ถูกต้องแม่นยำมากที่สุดซึ่งเป็นการพิจารณาคุณภาพของแบบสอบ คือ ค่าฟังก์ชั่นสารสนเทศแบบสอบ โดยกำหนดว่า ณ ระดับความสามารถใดที่ค่าฟังก์ชั่นสารสนเทศแบบสอบมีค่าสูง แสดงว่าแบบสอบมีคุณภาพที่ดีในการประมาณค่าความสามารถที่ระดับนั้น และนอกจากใช้พิจารณาคุณภาพของแบบสอบแล้ว ยังสามารถนำมาใช้เปรียบเทียบคุณภาพของแบบสอบได้อีกด้วยดังนี้ (Hambleton and Swaminathan 1985: 101)

ถ้าเรามีแบบสอบหลายฉบับที่วัดความสามารถอย่างเดียวกันแล้ว เราสามารถเปรียบเทียบคุณภาพของแบบสอบเหล่านั้นได้ โดยการนำเอาค่าฟังก์ชั่นสารสนเทศแบบสอบของแต่ละแบบสอบที่ต้องการเปรียบเทียบ 2 ฉบับ ณ ระดับความสามารถเดียวกัน มาหาค่าอัตราส่วนกัน ค่าอัตราส่วนของค่าฟังก์ชั่นสารสนเทศแบบสอบ 2 ฉบับที่นำมาเปรียบเทียบกัน ณ ระดับความสามารถเดียวกัน คือ ค่าประสิทธิภาพสัมพัทธ์ของแบบสอบ ซึ่งนิยามดังนี้

ค่าประสิทธิภาพสัมพัทธ์ของคะแนนแบบสอบ y กับคะแนนแบบสอบ x คือ อัตราส่วนของค่าฟังก์ชั่นสารสนเทศแบบสอบของแบบสอบทั้ง 2 ฉบับ ดังสมการ

$$RE(y,x) = \frac{I(\theta,y)}{I(\theta,x)}$$

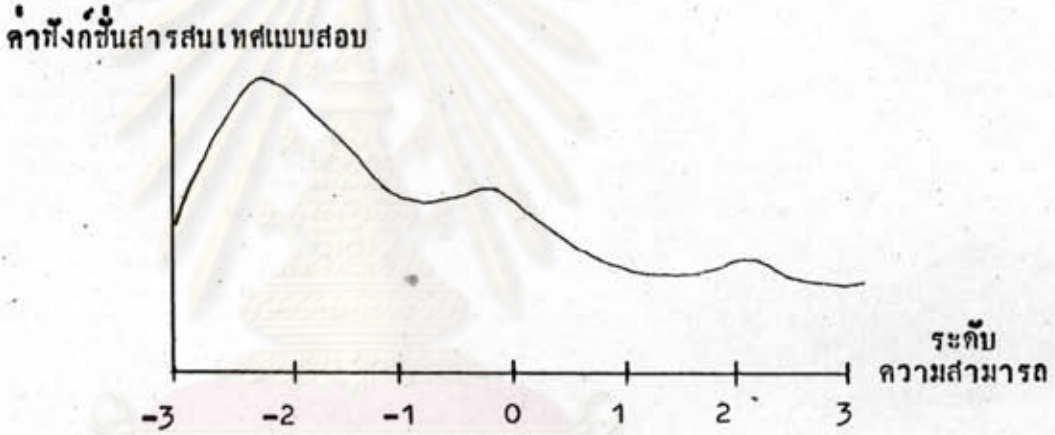
โดยที่คะแนน x และ y เป็นคะแนนจากแบบสอบที่แตกต่างกัน 2 ฉบับ ณ ระดับความสามารถเดียวกัน หรือ x และ y เป็นผลจากวิธีการให้คะแนนที่แตกต่างกัน 2 วิธี ของแบบสอบฉบับเดียวกัน ตามนิยามของค่าประสิทธิภาพสัมพัทธ์ของแบบสอบนั้น ค่าความสามารถใน $I(\theta,y)$ ต้องเป็นค่าความสามารถเดียวกันใน $I(\theta,x)$ และค่าประสิทธิภาพสัมพัทธ์ของคะแนนแบบสอบ 2 ฉบับ มีค่าแปรเปลี่ยนไปตามระดับความสามารถด้วย (Lord 1980: 83)

การแปลความหมายของค่าประสิทธิภาพสัมพัทธ์ของแบบสอบพิจารณาดังนี้ ณ ระดับความสามารถใด ถ้าค่าประสิทธิภาพสัมพัทธ์ของแบบสอบมีค่ามากกว่า 1 แสดงว่า

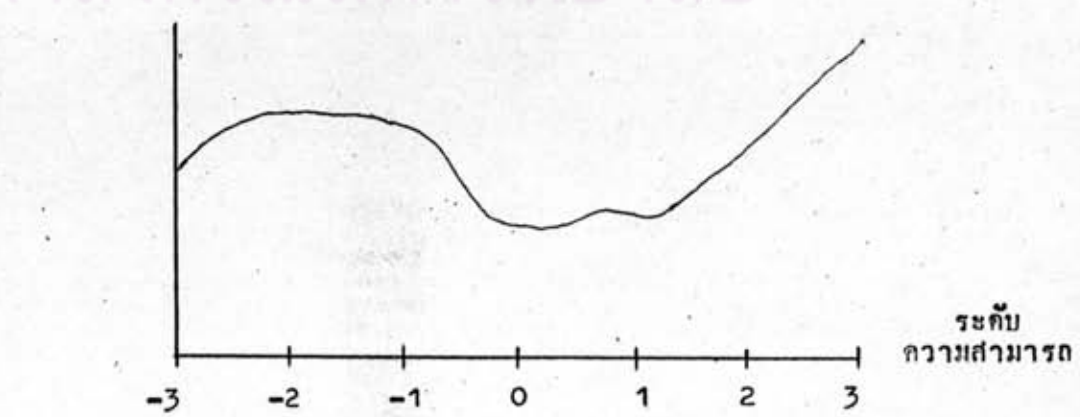
แบบสอบ y มีคุณภาพสูงกว่าแบบสอบ x ที่ระดับความสามารถนั้น ถ้าค่าประสิทธิภาพสัมพัทธ์ของแบบสอบมีค่าน้อยกว่า 1 แสดงว่าแบบสอบ y มีคุณภาพต่ำกว่าแบบสอบ x ที่ระดับความสามารถนั้น และถ้าค่าประสิทธิภาพสัมพัทธ์ของแบบสอบมีค่าเท่ากับ 1 แสดงว่าแบบสอบ y และแบบสอบ x มีคุณภาพเท่ากัน ที่ระดับความสามารถนั้น (Warm 1978: 76)

พิจารณาจากภาพดังนี้

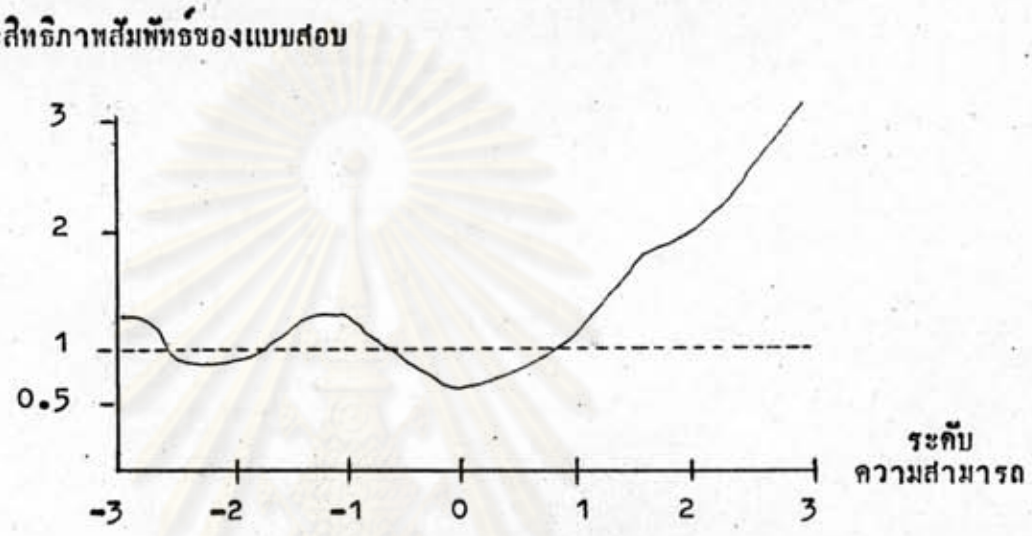
ภาพที่ 6 ค่าฟังก์ชั่นสารสนเทศแบบสอบ ที่มีประสิทธิภาพสำหรับการวัดระดับความสามารถต่ำ ($\theta = -2.3$)



ภาพที่ 7 ค่าฟังก์ชั่นสารสนเทศแบบสอบ ที่มีประสิทธิภาพสำหรับการวัดทั้งระดับความสามารถสูงและต่ำ



ภาพที่ 8 ค่าประสิทธิภาพสัมพัทธ์ของแบบสอบ ของค่าฟังก์ชั่นสารสนเทศแบบสอบภาพที่ 7 เมื่อเปรียบเทียบกับค่าฟังก์ชั่นสารสนเทศแบบสอบภาพที่ 6



จากภาพที่ 8 เป็นการแสดงถึงการเปรียบเทียบคุณภาพของแบบสอบ 2 ฉบับ โดยการนำเอาค่าฟังก์ชั่นสารสนเทศแบบสอบภาพที่ 7 และภาพที่ 6 มาหาค่าอัตราส่วนกัน ณ ระดับความสามารถเดียวกัน ของทุก ๆ ระดับความสามารถตั้งแต่ -3 ถึง 3 ซึ่งสรุปได้ว่า ในกลุ่มนักเรียนระดับความสามารถสูง (ระดับความสามารถมากกว่า 1 ถึง 3) แบบสอบของภาพที่ 7 มีคุณภาพสูงกว่าแบบสอบของภาพที่ 6 ในกลุ่มนักเรียนระดับความสามารถปานกลาง (ระดับความสามารถตั้งแต่ -1 ถึง 1) แบบสอบของภาพที่ 7 มีคุณภาพต่ำกว่าแบบสอบของภาพที่ 6 และในกลุ่มนักเรียนระดับความสามารถต่ำ (ระดับความสามารถน้อยกว่า -1 ถึง -3) มีทั้งแบบสอบของภาพที่ 7 มีคุณภาพสูงกว่าแบบสอบของภาพที่ 6 และแบบสอบของภาพที่ 7 มีคุณภาพต่ำกว่าแบบสอบของภาพที่ 6

โปรแกรมโลจิสต์

มาริลีน เอส. วินเจอร์สกี (Marilyn S. Wingersky) และคณะ ได้ร่วมกันพัฒนาโปรแกรมโลจิสต์ ซึ่งเป็นโปรแกรมคอมพิวเตอร์ที่ใช้ในการประมาณค่าพารามิเตอร์ของข้อสอบและประมาณค่าพารามิเตอร์ความสามารถของผู้สอบ ในโมเดลโลจิสติกที่มีพารามิเตอร์ 3 ตัว จากคำตอบของกลุ่มผู้สอบในกลุ่มข้อสอบกลุ่มใดกลุ่มหนึ่ง วิธีการประมาณค่าพารามิเตอร์ของข้อสอบและความสามารถของผู้สอบนี้ ได้ใช้วิธีการประมาณความน่าจะเป็นสูงสุด (Maximum Likelihood Estimation) ซึ่งวิธีนี้มีคุณสมบัติที่สำคัญดังนี้ (Hambleton and Swaminathan 1985: 88-89)

1. มีความคงที่ เมื่อจำนวนของกลุ่มผู้สอบและข้อสอบเพิ่มขึ้น การประมาณก็จะมีความคงที่ไปสู่ค่าที่แท้จริง

2. ฟังก์ชั่นของค่าสถิติมีความเที่ยงพอของสารสนเทศทั้งหมดเกี่ยวกับพารามิเตอร์

3. มีประสิทธิภาพ ซึ่งวิธีการประมาณความน่าจะเป็นสูงสุด มีความแปรปรวนน้อยที่สุด

4. มีลักษณะของการกระจายเข้าใกล้แบบปกติ

ลักษณะที่สำคัญของโปรแกรมโลจิสต์ คือ สามารถประมาณค่าพารามิเตอร์ของข้อสอบและความสามารถของผู้สอบ โดยการคำนวณกลับไปกลับมาหลาย ๆ ครั้ง จนกว่าจะได้ค่าใกล้เคียงที่มีความคงที่ ถ้าผู้สอบ N คน ได้ตอบแบบสอบชุดหนึ่ง n ข้อ ค่าที่ได้จากการคำนวณจะมีจำนวน N+3n ค่า คือ ค่าความสามารถของผู้สอบ (θ) ของแต่ละคน N ค่า และค่าพารามิเตอร์ของข้อสอบ (a_i, b_i, c_i) อีกข้อละ 3 ค่าเป็นจำนวน 3n ค่า แต่เนื่องจากค่าพารามิเตอร์ของข้อสอบและความสามารถของผู้สอบไม่สามารถวัดหรือสังเกตได้ จึงทำให้มีระดับของความไม่แน่นอนในโมเดลเกิดขึ้น วิธีการแก้ไขจึงจำเป็นกำหนดสเกลค่าความสามารถ (หรือค่าความยาก) ให้มีค่าคงที่ โดยทำให้ค่าเฉลี่ยเท่ากับ 0 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1 ซึ่งมีผลทำให้จำนวนของค่าพารามิเตอร์น้อยลง 2 ค่า ดังนั้นจำนวนพารามิเตอร์ทั้งหมดในโมเดลที่ได้ประมาณค่ามี

ที่ 74111610

จำนวน N+3n-2 ค่า (Hambleton and Swaminathan 1985: 126-127)

วิธีการประมาณค่าพารามิเตอร์ของข้อสอบและความสามารถของผู้สอบด้วยโปรแกรมโลจิสต์ มีลักษณะดังนี้

จากการนิยามความน่าจะเป็นของผู้สอบในการตอบข้อสอบข้อ i เมื่อ

$$U_i = \begin{cases} 1 & \text{สำหรับการตอบถูก} \\ 0 & \text{สำหรับการตอบผิด} \end{cases}$$

แสดงได้ดังสมการดังนี้

$$P(U_i/\theta,b,a,c) = P(U_i=1/\theta,b,a,c)P(U_i=0/\theta,b,a,c)$$

$$= P_i^{U_i}(1-P_i)^{1-U_i}$$

$$= P_i^{U_i} Q_i^{1-U_i} \qquad , Q_i = 1-P_i$$

ถ้าผู้สอบตอบข้อสอบ n ข้อ และแบบสอบที่สอบมีลักษณะการวัดเพียงมิติเดียว ความน่าจะเป็นของการตอบ แสดงได้ดังสมการต่อไปนี้

$$P(U_1,U_2,\ldots U_n/\theta,b,a,c) = P(U_1/\theta,b,a,c)P(U_2/\theta,b,a,c),\ldots,P(U_n/\theta,b,a,c)$$

$$= \prod_{i=1}^{n} P(U_i/\theta,b,a,c)$$

$$= \prod_{i=1}^{n} P_i^{U_i} Q_i^{1-U_i}$$

จากสมการดังกล่าวข้างต้นซึ่งเป็นความน่าจะเป็นการตอบข้อสอบ n ข้อ ที่สามารถวัดหรือสังเกตได้ โดยที่ $U_1,U_2,\ldots,U_n$ เป็นตัวแปรสุ่มที่มีค่าเฉพาะเป็น $u_1,u_2,\ldots,u_n$ เมื่อ u_i มีค่าเท่ากับ 1 หรือ 0 หรืออีกลักษณะหนึ่งสมการนี้เป็น

ฟังก์ชั่นคณิตศาสตร์ของค่า θ, b, a, c ซึ่งเรียกว่า ฟังก์ชั่นความน่าจะเป็น (Likelihood Function) แสดงได้ดังนี้

$$L(u_1,u_2,\ldots,u_n/\theta,b,a,c) = \prod_{i=1}^{n} P_i^{U_i} Q_i^{1-U_i}$$

เมื่อ $u_i=1$ ค่าของ Q_i ก็หมดไป และเมื่อ $u_i=0$ ค่าของ P_i ก็หมดไป และฟังก์ชั่นความน่าจะเป็นที่มีผู้สอบ N คน ตอบข้อสอบ n ข้อ มีสมการดังนี้

$$L(u/\theta,b,a,c) = L(u_1,u_2,\ldots,u_N/\theta_1,\theta_2,\ldots,\theta_N,b_1,b_2,\ldots,b_n,$$
$$a_1,a_2,\ldots,a_n,c_1,c_2,\ldots,c_n)$$

$$= \prod_{j=1}^{N} \prod_{i=1}^{n} L(U_{ij}/\theta,b,a,c)$$

$$= \prod_{j=1}^{N} \prod_{i=1}^{n} P_{ij}^{U_{ij}} Q_{ij}^{1-U_{ij}}$$

เมื่อ $P_{ij} \equiv P_i(\theta_j,b_i,a_i,c_i)$

ลอการิทึม (Logarithm) ของฟังก์ชั่นความน่าจะเป็นนี้ คือ

$$\ln L(u/\theta,b,a,c) = \sum_{j=1}^{N} \sum_{i=1}^{n} (u_{ij} \ln P_{ij} + (1-u_i) \ln Q_{ij})$$

การพิจารณาการประมาณความน่าจะเป็นสูงสุดของค่าพารามิเตอร์ θ, a, b และ c โดยการทำให้ค่าพารามิเตอร์เหล่านี้มีค่าสูงสุดในลอการิทึมของฟังก์ชั่นความน่าจะเป็น ซึ่งใช้วิธีการแก้สมการความน่าจะเป็น (Likelihood Equation) ดังนี้

$$\partial \ln L / \partial t_k = 0 \qquad (k = 1,2,\ldots,N+3n-2)$$

เมื่อ t_k เป็นสมาชิกของพารามิเตอร์เวคเตอร์ t นิยามดังนี้

$$t' = \left[\theta' \; a' \; b' \; c'\right]$$

แต่สมการความน่าจะเป็นนี้โดยทั่วไปไม่เป็นสมการเส้นตรง และไม่สามารถหาค่าพารามิเตอร์ได้ชัดเจน ดังนั้นจึงได้ใช้วิธีด้านจำนวน (Numerical) เข้ามาช่วยในการหาค่าพารามิเตอร์ วิธีที่นิยมใช้กัน คือ วิธีของนิวตัน (Newton-Raphson Procedure) โดยมีการประมาณค่าซ้ำ ๆ กันจนมีค่าคงที่ ดังสมการต่อไปนี้

$$t^{(j+1)} = t^{j} - \left[f''(t^{(j)})\right]^{-1} f'(t^{(j)})$$

เมื่อ t คือ $\left[\theta \; a \; b \; c\right]$

f(t) คือ ลอการิทึมของฟังก์ชั่นความน่าจะเป็น ($\ln L(u/\theta,b,a,c)$)

$f'(t)$ คือ เวคเตอร์ของอนุพันธ์อันดับที่ 1 ที่ประมาณค่า $t^{(j)}$

$f''(t)$ คือ เมตริกซ์ของอนุพันธ์อันดับที่ 2 ของ f(t)

j,j+1 คือ จำนวนครั้งที่ได้ประมาณค่า

ซึ่งได้ทำการประมาณค่าจนความแตกต่างระหว่างการประมาณครั้งที่ j+1 และครั้งที่ j มีเพียงเล็กน้อยก็หยุดการประมาณถือว่ามีค่าคงที่ โดยเริ่มจากกำหนดค่าพารามิเตอร์ a,b และ c แล้วนำไปประมาณค่าพารามิเตอร์ความสามารถของผู้สอบ และนำค่าพารามิเตอร์ความสามารถที่ได้ไปประมาณค่าพารามิเตอร์ a,b และ c ซึ่งทั้ง 2 ขั้นตอนนี้ได้ถูกทำซ้ำ ๆ จนมีค่าคงที่ และในการทำซ้ำแต่ละครั้งค่าความสามารถถูกปรับให้มีค่าเฉลี่ยเท่ากับ 0 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1 ค่าพารามิเตอร์ a,b และ c ได้ถูกปรับไปด้วย

โดยสรุปการประมาณค่าพารามิเตอร์ของข้อสอบและความสามารถของผู้สอบ มี 2 ขั้นตอน คือ การประมาณค่าความสามารถของผู้สอบ N คน โดยการปรับค่าความสามารถในการทำซ้ำแต่ละครั้ง จนมีค่าเฉลี่ยเท่ากับ 0 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1 และการนำค่าความสามารถนี้ไปประมาณค่าพารามิเตอร์ของข้อสอบ n ข้อ ซึ่งในโปรแกรมโลจิสต์มีการทำงาน 4 ขั้นตอน ดังนี้

1. กำหนดให้ค่าอำนาจจำแนกของข้อสอบทุกข้อเท่ากับ 1 และค่าการเดาของข้อสอบเท่ากับ 1/A - 0.05 เมื่อ A แทน จำนวนตัวเลือกของข้อสอบ แล้วทำการประมาณค่าความสามารถของผู้สอบแต่ละคน และค่าความยากของข้อสอบแต่ละข้อ

2. นำค่าความสามารถของผู้สอบแต่ละคนจากขั้นที่ 1 ไปประมาณค่าอำนาจจำแนกของข้อสอบ ค่าความยากของข้อสอบ และค่าการเดาของข้อสอบแต่ละข้อ

3. นำค่าอำนาจจำแนกของข้อสอบและค่าการเดาของข้อสอบจากขั้นที่ 2 ไปประมาณค่าความสามารถของผู้สอบแต่ละคน และค่าความยากของข้อสอบแต่ละข้อ

4. นำค่าความสามารถของผู้สอบแต่ละคนจากขั้นที่ 3 ไปประมาณค่าอำนาจจำแนกของข้อสอบ ค่าความยากของข้อสอบ และค่าการเดาของข้อสอบแต่ละข้อ

ซึ่งแสดงได้ดังตารางที่ 1 ต่อไปนี้

ตารางที่ 1 ขั้นตอนการประมาณค่าพารามิเตอร์ในโปรแกรมโลจิสต์

ขั้นตอนที่	ค่าพารามิเตอร์			
	ความสามารถ	ค่าอำนาจจำแนก	ค่าความยาก	ค่าการเดา
1	ประมาณ	กำหนด	ประมาณ	กำหนด
2	กำหนด	ประมาณ	ประมาณ	ประมาณ
3	ประมาณ	กำหนด	ประมาณ	กำหนด
4	กำหนด	ประมาณ	ประมาณ	ประมาณ

ในแต่ละขั้นตอนของการประมาณค่าพารามิเตอร์ของข้อสอบและความสามารถของผู้สอบจะมีหลาย ๆ ระยะ ซึ่งโปรแกรมได้กำหนดไว้ 25 ระยะด้วยกัน และในแต่ละระยะในขั้นตอนที่ 1 และ 3 ประกอบด้วย 3 ส่วนด้วยกัน คือ

1. ค่าความสามารถได้รับการประมาณค่า ขณะที่ค่าความยากของข้อสอบถูกกำหนดเป็นค่าคงที่ไว้

2. ถ้ามีการเปลี่ยนแปลงสเกล ก็จะมีการปรับค่าความสามารถ ค่าอำนาจจำแนกของข้อสอบ และค่าความยากของข้อสอบ ถึงแม้ว่าค่าอำนาจจำแนกของข้อสอบถูกกำหนดเป็นค่าคงที่ไว้ ก็จะมีการเปลี่ยนแปลงไปด้วย

3. ค่าความยากของข้อสอบได้รับการประมาณค่าแต่ละข้อ ขณะที่ค่าความสามารถเป็นค่าที่ถูกกำหนดเป็นค่าคงที่ ที่ได้รับการประมาณไว้ในตอนเริ่มต้นของระยะนี้

ส่วนในแต่ละระยะของขั้นตอนที่ 2 และ 4 ก็จะประกอบด้วยการประมาณค่าพารามิเตอร์ของข้อสอบแต่ละข้อจนครบทั้งหมดทุกข้อ นอกจากในระยะแรกของขั้นตอนที่ 2 ค่าอำนาจจำแนกของข้อสอบ และค่าความยากของข้อสอบ ได้รับการประมาณค่าแต่ละข้อ

ในแต่ละระยะค่าความสามารถได้ถูกประมาณค่าทีละค่าจำนวนหลายครั้ง เพื่อให้ค่าที่คงที่ และค่าพารามิเตอร์ของข้อสอบ a, b และ c ในแต่ละข้อก็ถูกประมาณค่าหลาย ๆ ครั้ง เช่นเดียวกัน ในระยะที่ 1 และ 3 จำนวนครั้งของการประมาณถูกกำหนดโดยขนาดของการเปลี่ยนแปลงของค่าพารามิเตอร์ที่ถูกประมาณ ในระยะที่ 2 และ 4 ค่าพารามิเตอร์ของข้อสอบทั้ง 3 ค่า ได้ถูกประมาณค่าพร้อม ๆ กัน และสิ้นสุดเมื่อค่าอำนาจจำแนกของข้อสอบมีการเปลี่ยนแปลงเพียงเล็กน้อย ซึ่งน้อยกว่าเกณฑ์ที่กำหนดไว้

ในแต่ละขั้นตอนจะสิ้นสุดหรือมีความคงที่ เมื่อค่าพารามิเตอร์ของระยะที่ n และ n+1 มีค่าเพิ่มขึ้นน้อยกว่าเกณฑ์ที่กำหนดไว้ในขั้นตอนนั้น ๆ ซึ่งโปรแกรมได้กำหนดเกณฑ์ในขั้นตอนที่ 1 ถึง 4 คือ 200 % , 20 % , 2 % และ 0.2 % ตามลำดับ

1.3 <u>ความแตกต่างระหว่างทฤษฎีคลาสสิคอลกับทฤษฎีการตอบสนองข้อสอบ</u>

การเปรียบเทียบความแตกต่างระหว่าง 2 ทฤษฎีนี้ ได้สรุปในประเด็นที่สำคัญ ซึ่งแสดงได้ด้วยตารางที่ 2 ดังนี้ (อวยพร วิบูลย์กาญจน์ 2526: 24)

ตารางที่ 2 ความแตกต่างระหว่างทฤษฎีคลาสสิคอลกับทฤษฎีการตอบสนองข้อสอบ

	ทฤษฎีคลาสสิคอล	ทฤษฎีการตอบสนองข้อสอบ
ลักษณะ	เป็นโมเดลทางคณิตศาสตร์ที่แสดงถึงความคลาดเคลื่อนของการวัดที่มีผลต่อคะแนนที่วัดได้หรือสังเกตได้ เป็นการเน้นการประมาณค่าคะแนนจริงหรือความสามารถที่แท้จริง โดยใช้คะแนนของกลุ่มผู้สอบเป็นสำคัญ	เป็นโมเดลทางคณิตศาสตร์ที่อธิบายถึงความสัมพันธ์ระหว่างการตอบข้อสอบของผู้สอบกับระดับความสามารถที่มีอยู่ภายในตัวผู้สอบ เป็นการเน้นการประมาณค่าพารามิเตอร์ของข้อสอบและค่าพารามิเตอร์ของความสามารถของผู้สอบ
ค่าความยาก	สัดส่วนของผู้สอบทั้งหมดที่ตอบข้อสอบแต่ละข้อถูกต้อง มีค่าตั้งแต่ 0 ถึง 1	ค่าที่แสดงถึงระดับความสามารถที่จุดโค้งลักษณะข้อสอบมีความชันมากที่สุด โดยทั่วไปมีค่าอยู่ระหว่าง -2 ถึง 2
ค่าอำนาจจำแนก	ประสิทธิภาพของข้อสอบที่จะสามารถแยกความแตกต่างระหว่างกลุ่มที่ได้คะแนนสูงกับกลุ่มที่ได้คะแนนต่ำ มีค่าตั้งแต่ -1 ถึง 1	ค่าที่เป็นสัดส่วนโดยตรงกับความชันของโค้งลักษณะข้อสอบ ณ จุดเปลี่ยนโค้ง หรือที่จุด $\theta=b_i$ โดยทั่วไปมีค่าอยู่ระหว่าง 0 ถึง 2

ตารางที่ 2 (ต่อ)

	ทฤษฎีคลาสสิคอล	ทฤษฎีการตอบสนองข้อสอบ
ค่าการเดา	อัตราส่วนของ 1 กับจำนวนตัวเลือกในข้อสอบแต่ละข้อของแบบสอบ	ค่าความน่าจะเป็นที่ผู้สอบมีความสามารถต่ำมาก มีโอกาสตอบข้อสอบข้อนั้นถูก หรือเป็นค่ากำกับโค้งที่ต่ำสุดของโค้งลักษณะข้อสอบมีค่า 0 ถึง 1
ความตรง	ความสามารถในการวัดของสิ่งที่เราต้องการจะวัดได้อย่างถูกต้อง มีค่าตั้งแต่ 0 ถึง 1	เป็นการใช้แบบสอบให้เป็นไปตามวัตถุประสงค์ที่ต้องการของการสอบ โดยพิจารณาจากค่าฟังก์ชั่นสารสนเทศแบบสอบ ซึ่งเป็นค่าฟังก์ชั่นของค่าพารามิเตอร์ของข้อสอบตามโมเดลที่ใช้ มีค่าตั้งแต่ 0 ขึ้นไป
ความเที่ยง	ความคงที่ของคะแนนที่ได้จากการทดสอบคนกลุ่มเดียวกันสองครั้งด้วยแบบสอบเดิมในเวลาที่ต่างกัน หรือทดสอบคนกลุ่มเดียวกันด้วยข้อสอบต่างชุดที่มีข้อสอบเทียบเท่ากัน หรือภายใต้สภาพการสอบที่แตกต่างกัน มีค่าตั้งแต่ 0 ถึง 1	ความถูกต้องแม่นยำของการประมาณค่าความสามารถที่แท้จริงของผู้สอบ โดยพิจารณาจากค่าฟังก์ชั่นสารสนเทศแบบสอบ

ตอนที่ 2 งานวิจัยที่เกี่ยวข้อง

2.1 ตามกรอบของทฤษฎีคลาสสิคอล

การเปรียบเทียบคุณภาพของแบบสอบแบบเลือกตอบที่มีจำนวนตัวเลือกไม่เท่ากัน ตามกรอบของทฤษฎีคลาสสิคอล มีผลการศึกษาวิจัยของนักวัดและประเมินผลการศึกษาทั้งของต่างประเทศและไทยดังนี้

วิลเลียมส์ และ อีเบล (Williams and Ebel 1957: 64 อ้างถึงใน Lord 1980: 107) ได้วิจัยพบว่า การตัดตัวลวงที่มีค่าอำนาจจำแนกน้อยที่สุดของตัวเลือกของแบบสอบคำศัพท์ที่มีจำนวนตัวเลือก 4 ตัวเลือก ให้เหลือเพียง 3 ตัวเลือก และ 2 ตัวเลือก ทำให้แบบสอบแบบเลือกตอบที่มี 2 ตัวเลือก มีค่าความเที่ยงสูงกว่าแบบสอบแบบเลือกตอบที่มี 3 ตัวเลือก และ 4 ตัวเลือก แต่ไม่มีความแตกต่างกันอย่างมีนัยสำคัญที่ระดับ .10

คอสติน (Costin 1970: 353-358) ได้วิจัยพบว่า การสุ่มตัดตัวลวงจากแบบสอบแบบเลือกตอบวัดผลสัมฤทธิ์ทางการเรียนที่มี 4 ตัวเลือก เป็น 3 ตัวเลือก ทำให้ค่าเฉลี่ยของค่าอำนาจจำแนกและค่าความยากของแบบสอบแบบเลือกตอบที่มี 3 ตัวเลือก มีค่าสูงกว่าแบบสอบแบบเลือกตอบที่มี 4 ตัวเลือก และค่าความเที่ยงของแบบสอบแบบเลือกตอบที่มี 3 ตัวเลือก ยังมีค่าสูงกว่าแบบสอบแบบเลือกตอบที่มี 4 ตัวเลือกด้วย

อีเบล (Ebel 1969: 565-570) ได้ศึกษาเกี่ยวกับค่าความเที่ยงของแบบสอบแบบเลือกตอบที่มีจำนวนตัวเลือกไม่เท่ากัน พบว่า ค่าความเที่ยงของแบบสอบจะสูงมากขึ้นหรือสมควร ถ้าเพิ่มจำนวนตัวเลือกจาก 2 ตัวเลือก เป็น 3 ตัวเลือก และค่าความเที่ยงของแบบสอบจะเพิ่มขึ้นเล็กน้อย เมื่อเพิ่มจำนวนตัวเลือกเป็น 4 ตัวเลือก และจะยังเพิ่มขึ้นเล็กน้อยเรื่อย ๆ ไป ตามจำนวนตัวเลือกที่เพิ่มขึ้น

วารุณี ปิตธวัชชัย (2513: 29-30) ได้วิจัยพบว่า การสุ่มตัดตัวลวงออกทีละข้อของแบบสอบความถนัดทางการเรียนประเภทอุปมาอุปไมยที่มีจำนวนตัวเลือก 5 ตัวเลือก เป็น 4 ตัวเลือก และ 3 ตัวเลือก ทำให้ค่าความเที่ยงของแบบสอบทั้ง 3 ฉบับแตกต่างกัน โดยค่าความเที่ยงของแบบสอบแบบเลือกตอบที่มี 5 ตัวเลือก มีค่าสูงสุดและ

ลดต่ำลงมาเป็นแบบสอบแบบเลือกตอบที่มี 4 ตัวเลือก และ 3 ตัวเลือก ตามลำดับ และค่าความเที่ยงของแบบสอบแบบเลือกตอบที่มี 5 ตัวเลือก กับ 3 ตัวเลือก และแบบสอบแบบเลือกตอบที่มี 4 ตัวเลือก กับ 3 ตัวเลือก มีค่าแตกต่างกันอย่างมีนัยสำคัญที่ระดับ .01 และ .05 ตามลำดับ แต่แบบสอบแบบเลือกตอบที่มี 5 ตัวเลือก กับ 4 ตัวเลือก มีค่าความเที่ยงไม่แตกต่างกันอย่างมีนัยสำคัญ และค่าอำนาจจำแนกและค่าความยากของข้อสอบแต่ละข้อของแบบสอบทั้ง 3 ฉบับ ไม่แตกต่างกันอย่างมีนัยสำคัญ

มาลี ชิดสวน (2514: 37-39) ได้วิจัยพบว่า การสุ่มตัดตัวลวงของแบบสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาวิทยาศาสตร์แบบเลือกตอบที่มี 5 ตัวเลือก เป็น 4 ตัวเลือก และ 3 ตัวเลือก ทำให้ค่าความเที่ยงของแบบสอบทั้ง 3 ฉบับแตกต่างกัน โดยค่าความเที่ยงของแบบสอบแบบเลือกตอบที่มี 5 ตัวเลือก มีค่าสูงสุด และลดต่ำลงมาเป็นแบบสอบแบบเลือกตอบที่มี 4 ตัวเลือก และ 3 ตัวเลือก ตามลำดับ และค่าความเที่ยงของแบบสอบแบบเลือกตอบที่มี 5 ตัวเลือก กับ 3 ตัวเลือก มีค่าแตกต่างกันอย่างมีนัยสำคัญ แต่แบบสอบแบบเลือกตอบที่มี 5 ตัวเลือก กับ 4 ตัวเลือก และ แบบสอบแบบเลือกตอบที่มี 4 ตัวเลือก กับ 3 ตัวเลือก มีค่าความเที่ยงไม่แตกต่างกันอย่างมีนัยสำคัญ ค่าความตรงของแบบสอบทั้ง 3 ฉบับ ไม่แตกต่างกันอย่างมีนัยสำคัญ ค่าอำนาจจำแนกของข้อสอบแต่ละข้อของแบบสอบทั้ง 3 ฉบับ ไม่แตกต่างกันอย่างมีนัยสำคัญ ค่าความยากของข้อสอบแต่ละข้อของแบบสอบทั้ง 3 ฉบับ แตกต่างกันอย่างมีนัยสำคัญ โดยแบบสอบแบบเลือกตอบที่มี 5 ตัวเลือกมีค่าสูงสุด และลดต่ำลงมาเป็นแบบสอบแบบเลือกตอบที่มี 4 ตัวเลือก และ 3 ตัวเลือก ตามลำดับ และคะแนนเฉลี่ยของแบบสอบทั้ง 3 ฉบับ แตกต่างกันอย่างมีนัยสำคัญ โดยคะแนนเฉลี่ยของแบบสอบแบบเลือกตอบที่มี 3 ตัวเลือก มีค่าสูงสุด และลดต่ำลงมาเป็นแบบสอบแบบเลือกตอบที่มี 4 ตัวเลือก และ 5 ตัวเลือก ตามลำดับ

2.2 ตามกรอบของทฤษฎีการตอบสนองข้อสอบ

การพิจารณาค่าประสิทธิภาพสัมพัทธ์ของแบบสอบ ที่ใช้เปรียบเทียบคุณภาพของแบบสอบแบบเลือกตอบที่มีจำนวนตัวเลือกไม่เท่ากันตามกรอบของทฤษฎีการตอบสนองข้อสอบ ยังไม่เป็นที่แพร่หลายมากนัก จึงยังไม่มีรายงานผลการวิจัยเกี่ยวกับเรื่องนี้ แต่มีผลงานของ ลอร์ด ที่ใช้แนวคิดนี้เปรียบเทียบคุณภาพของแบบสอบต่าง ๆ โดยศึกษาค่าประสิทธิภาพสัมพัทธ์ของแบบสอบแบบคำศัพท์ ระดับ 6 จำนวน 7 ฉบับด้วยกัน (Lord 1980: 96-100) และใช้แนวคิดนี้ศึกษาจำนวนตัวเลือกที่เหมาะสมของข้อสอบ โดยเปรียบเทียบคุณภาพของแบบสอบที่มีความยาวของแบบสอบและค่าการเดาของข้อสอบที่แตกต่างกัน (Lord 1980: 110-112) ดังนี้

การเปรียบเทียบคุณภาพของแบบสอบต่าง ๆ โดยศึกษาค่าประสิทธิภาพสัมพัทธ์ของแบบสอบ โดยได้เปรียบเทียบ Metropolitan Reading Tests (1970), Intermediate Level, Form F, Word Analysis subtest (MAT) กับแบบสอบอื่น ๆ อีก 6 ฉบับดังนี้

1. Sequential Tests of Education Progress Series II (1969), Level 4, Form A, Reading subtest (STEP)

2. California Achievement Tests (1970), Level 4, Form A, Reading Vocabulary (CAT)

3. Iowa Test of Basic Skills (1970), Level 12, Form 5, Vocabulary (ITBS)

4. Standford Reading Tests (1964), Intermediate II, Form W, Word Meaning (Standford)

5. Comprehensive Tests of Basic Skills (1968), Level 3, Form Q, Reading Vocabulary (CTBS)

6. SRA Achievement Series (1971), Green Edition, Form E, Vocabulary (SRA)

ผลการศึกษาสรุปได้ดังนี้

1. STEP มีคุณภาพสูงกว่า MAT ในกลุ่มนักเรียนระดับความสามารถต่ำ (ต่ำกว่าเปอร์เซ็นต์ไทล์ที่ 20) ส่วนระดับความสามารถอื่น ๆ STEP มีคุณภาพต่ำกว่า

2. CAT มีคุณภาพสูงกว่า MAT ในกลุ่มนักเรียนระดับความสามารถสูงมาก (มากกว่าเปอร์เซ็นต์ไทล์ที่ 95) ส่วนระดับความสามารถอื่น ๆ CAT มีคุณภาพต่ำกว่า

3. ITBS มีคุณภาพสูงกว่า MAT ในกลุ่มนักเรียนระดับความสามารถสูงมาก (มากกว่าเปอร์เซ็นต์ไทล์ที่ 90) ส่วนระดับความสามารถอื่น ๆ ITBS มีคุณภาพต่ำกว่า

4. Standford มีคุณภาพสูงกว่า MAT ในกลุ่มนักเรียนระดับความสามารถปานกลางและสูง (มากกว่าเปอร์เซ็นต์ไทล์ที่ 50) และมีคุณภาพต่ำกว่า MAT ในกลุ่มนักเรียนระดับความสามารถปานกลางและต่ำ (น้อยกว่าเปอร์เซ็นต์ไทล์ที่ 50)

5. CTBS มีคุณภาพสูงกว่า MAT ในกลุ่มนักเรียนระดับความสามารถสูง (มากกว่าเปอร์เซ็นต์ไทล์ที่ 80) มีคุณภาพใกล้เคียงกันในกลุ่มนักเรียนระดับความสามารถปานกลางและค่อนข้างสูง (มากกว่าเปอร์เซ็นต์ไทล์ที่ 50 ถึงเปอร์เซ็นต์ไทล์ที่ 80) และมีคุณภาพต่ำกว่า MAT ในกลุ่มนักเรียนระดับความสามารถปานกลางและต่ำ (น้อยกว่าเปอร์เซ็นต์ไทล์ที่ 50)

6. SRA มีคุณภาพต่ำกว่า MAT ในกลุ่มนักเรียนระดับความสามารถต่ำ (น้อยกว่าเปอร์เซ็นต์ไทล์ที่ 30) ส่วนระดับความสามารถอื่น ๆ SRA มีคุณภาพสูงกว่า

การศึกษาจำนวนตัวเลือกที่เหมาะสมของข้อสอบโดยเปรียบเทียบคุณภาพของแบบสอบที่มีความยาวของแบบสอบและค่าการเดาของข้อสอบที่แตกต่างกัน ซึ่งได้ศึกษาแบบสอบความถนัดทางการเรียนด้านภาษา (TSA13) คือ The College Board Scholastic Aptitude Test, Form TSA13 เป็นแบบสอบแบบเลือกตอบที่มี 5 ตัวเลือก จำนวน 90 ข้อ ค่าพารามิเตอร์ของข้อสอบได้ประมาณค่าตามโมเดลโลจิสติกที่มีพารามิเตอร์ 3 ตัว ค่ามัธยฐานของค่าการเดาของข้อสอบมีค่าประมาณ .15 และได้สร้างแบบสอบสมมุติ (Hypothetical Test) ขึ้นจากแบบสอบนี้อีก 4 ฉบับ โดยการแทนค่าการเดาของข้อสอบใหม่ในแต่ละข้อของแบบสอบสมมุติแต่ละฉบับตามลำดับดังนี้

.20, .25, .333 และ .50 ส่วนค่าอำนาจจำแนกของข้อสอบและค่าความยากของข้อสอบยังใช้ค่าเดิมใน TSA13 ซึ่งค่าการเดาของข้อสอบใหม่ในแบบสอบสมมุติแต่ละฉบับมีความสัมพันธ์กับจำนวนตัวเลือกของข้อสอบตามลำดับดังนี้ 5 ตัวเลือก 4 ตัวเลือก 3 ตัวเลือก และ 2 ตัวเลือก และเนื่องจากจำนวนของข้อสอบในแบบสอบสมมุติแต่ละฉบับเป็นสัดส่วนโดยตรงกับค่าการเดาของข้อสอบ ดังนั้นจำนวนของข้อสอบในแบบสอบสมมุติ 4 ฉบับ มีจำนวนแตกต่างกันตามลำดับดังนี้ 120 ข้อ 150 ข้อ 200 ข้อ และ 300 ข้อ แล้วเปรียบเทียบคุณภาพของแบบสอบโดยการหาค่าประสิทธิภาพสัมพัทธ์ของแบบสอบของ TSA13 ที่มีค่ามัธยฐานของค่าการเดาของข้อสอบประมาณ .15 แบบสอบสมมุติที่มีค่าการเดาของข้อสอบเป็น .20, .25 และ .50 เปรียบเทียบกับแบบสอบสมมุติที่มีค่าการเดาของข้อสอบเป็น .333 ผลการศึกษาพบว่า ที่ระดับความสามารถสูงแบบสอบที่มีคุณภาพสูงที่สุดคือ แบบสอบที่มีค่าการเดาของข้อสอบเป็น .50 รองลงมาเป็น .333, .25, .20 และ TSA13 ตามลำดับ และที่ระดับความสามารถต่ำแบบสอบที่มีคุณภาพสูงที่สุดคือ TSA13 รองลงมาเป็นแบบสอบที่มีค่าการเดาของข้อสอบเป็น .20, .25, .333 และ .50 นั่นคือ ผลกระทบจากการลดจำนวนตัวเลือกของข้อสอบโดยที่เป็นสัดส่วนโดยตรงกับความยาวของแบบสอบ ทำให้คุณภาพของแบบสอบเพิ่มขึ้นสำหรับกลุ่มผู้สอบที่มีระดับความสามารถสูง และลดลงสำหรับกลุ่มผู้สอบที่มีระดับความสามารถต่ำ

จากแบบสอบสมมุติ 4 ฉบับนี้ ลอร์ด ได้ศึกษาเพิ่มเติมโดยพิจารณาถึงค่าความยากของข้อสอบในแบบสอบด้วย ซึ่งจากค่ามัธยฐานของค่าความยากของข้อสอบของ TSA13มีค่าประมาณ .5 ดังนั้นค่าความยากของข้อสอบของแบบสอบสมมุติทั้ง 4 ฉบับก็มีค่าประมาณ .5 ด้วย โดยเปรียบเทียบคุณภาพของแบบสอบสมมุติทั้ง 4 ฉบับ ซึ่งหาค่าประสิทธิภาพสัมพัทธ์ของแบบสอบของแบบสอบที่มีค่าการเดาของข้อสอบ .50 (โดยลดค่าความยากของข้อสอบทุกข้อเท่ากับ 0.1 เพื่อให้ง่ายขึ้นเล็กน้อย) แบบสอบที่มีค่าการเดาของข้อสอบ .25 (โดยการเพิ่มค่าความยากของข้อสอบทุกข้อเท่ากับ 0.1 เพื่อให้ยากขึ้นเล็กน้อย) และแบบสอบที่มีค่าการเดาของข้อสอบ .20 (โดยการเพิ่มค่าความยากของข้อสอบทุกข้อเท่ากับ 0.2 เพื่อให้ยากขึ้นมาก) เปรียบเทียบกับแบบสอบที่มีค่าการเดาของข้อสอบ .333 (โดยมีค่าความยากของข้อสอบทุกข้อเท่ากับ 0.5) ผลการศึกษาพบว่า

แบบสอบที่มีค่าการเดาของข้อสอบ .333 และมีค่าความยากของข้อสอบทุกข้อเท่ากับ 0.5 มีคุณภาพสูงกว่าแบบสอบที่มีค่าการเดาของข้อสอบ .50 และมีความยากของข้อสอบทุกข้อเท่ากับ 0.4 แบบสอบที่มีค่าการเดาของข้อสอบ .25 และมีค่าความยากของข้อสอบทุกข้อเท่ากับ 0.6 และแบบสอบที่มีค่าการเดาของข้อสอบ .20 และมีค่าความยากของข้อสอบทุกข้อเท่ากับ 0.7

จากงานวิจัยที่เกี่ยวข้องตามกรอบของทฤษฎีคลาสสิคอลนั้น ในการเปรียบเทียบคุณภาพของแบบสอบแบบเลือกตอบที่มีจำนวนตัวเลือกไม่เท่ากัน ซึ่งเป็นแบบสอบแบบเลือกตอบชุดเดียวกัน แล้วลดจำนวนตัวเลือกจาก 5 ตัวเลือก เป็น 4 ตัวเลือก และ 3 ตัวเลือก พบว่า ค่าความเที่ยงของแบบสอบทั้ง 3 ฉบับมีค่าแตกต่างกันอย่างมีนัยสำคัญ โดยค่าความเที่ยงของแบบสอบแบบเลือกตอบที่มี 5 ตัวเลือก มีค่าสูงสุด และลดต่ำลงมาเป็นแบบสอบแบบเลือกตอบที่มี 4 ตัวเลือก และ 3 ตัวเลือก ตามลำดับ ค่าอำนาจจำแนกของข้อสอบแต่ละข้อของแบบสอบทั้ง 3 ฉบับ ไม่แตกต่างกันอย่างมีนัยสำคัญ ส่วนค่าความยากของข้อสอบไม่สามารถสรุปได้ เพราะมีผลการวิจัยที่ขัดแย้งกัน ดังนั้นจึงไม่สามารถสรุปได้อย่างชัดเจนว่าแบบสอบแบบเลือกตอบฉบับใดที่มีคุณภาพสูงกว่าเมื่อนำมาเปรียบเทียบกัน เพราะเป็นการเปรียบเทียบคุณลักษณะที่สำคัญของการตรวจสอบคุณภาพของแบบสอบทีละประเด็น ส่วนงานวิจัยที่เกี่ยวข้องตามกรอบของทฤษฎีการตอบสนองข้อสอบในเรื่องนั้นยังไม่มีผู้ที่ทำการวิจัยมาก่อน

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย