หอสมุดกลาง สถาบันวิทยบริการ จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 2

วรรณคดีที่เกี่ยวข้อง

การทดสอบในประเทศไทยส่วนใหญ่ใช้เพื่อการเรียนการสอนในสถานศึกษาและการคัดเลือกบุคคลเข้าทำงาน โดยมีการใช้แบบสอบประเภทต่าง ๆมากมาย รูปแบบในการวิเคราะห์ข้อสอบใช้ทฤษฎีมาตรฐานดั้งเดิม (Classical Test Theory) เป็นส่วนใหญ่ ซึ่งในปัจจุบันนี้ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory) เป็นทฤษฎีหนึ่งที่กำลังเป็นที่นิยม โดยเฉพาะรูปแบบของราสช์ (Rasch Model) นั้นได้มีการนำมาประยุกต์ใช้อย่างกว้างขวางในต่างประเทศ จากรายงานการวิจัยของต่างประเทศปรากฏว่ารูปแบบของราสช์สามารถใช้ได้กับแบบสอบประเภทต่าง ๆ เช่น แบบสอบผลสัมฤทธิ์ แบบสอบวัดความถนัดและอื่น ๆ สำหรับในการวิจัยครั้งนี้ผู้วิจัยได้นำรูปแบบของราสช์มาประยุกต์ใช้กับแบบสอบผลสัมฤทธิ์ ซึ่งจะได้นำเสนอมโนทัศน์ที่เกี่ยวข้องในการวิจัย โดยแบ่งเป็นหัวข้อที่สำคัญ ดังนี้

ตอนที่ 1 มโนทัศน์เกี่ยวกับแบบสอบผลสัมฤทธิ์

ตอนที่ 2 มโนทัศน์เกี่ยวกับรูปแบบของราสช์ กับงานวิจัยที่เกี่ยวข้อง

ตอนที่ 3 มโนทัศน์ในการออกแบบแบบสอบ

<u>ตอนที่ 1 มโนทัศน์เกี่ยวกับแบบสอบผลสัมฤทธิ์</u>

แบบสอบผลสัมฤทธิ์ส่วนมากจะเป็นเครื่องมือสำหรับช่วยครูให้สามารถตัดสินผลสัมฤทธิ์ของนักเรียนได้อย่างมีความเป็นปรนัย (objectively) (เยาวดี วิบูลย์ศรี 2528 : 10) เพราะเป็นวิธีประเมินพฤติกรรมของนักเรียนที่มีความเป็นอิสระได้มากกว่าวิธีอื่น ๆในกระบวนการเรียนการสอน แบบสอบผลสัมฤทธิ์ที่ใช้ในโรงเรียนมุ่งวัดความสำเร็จในวิชาเฉพาะและทักษะต่าง ๆโดยมีวัตถุประสงค์พื้นฐานของการใช้แบบสอบผลสัมฤทธิ์ 2 ประการคือ ประการแรกก็เพื่อเป็นเครื่องมือในการวัดประเมินผลสัมฤทธิ์ของนักเรียนอันเป็นข้อมูลสำหรับการวัดการเรียนการสอนเป็นรายบุคคล (Individual Instruction) ประการที่สองก็เพื่อตรวจสอบความสามารถของบุคคลที่แตกต่างกันโดยธรรมชาติ

ความหมายของแบบสอบผลสัมฤทธิ์

ความหมายโดยทั่วไปนั้น Encyclopedia World Dictionary ให้ความหมายว่า คือ แบบสอบที่สร้างขึ้นเพื่อใช้ในการวัดผลการเรียนการสอน Webster 's New International Dictionary of The English Language ได้ให้ความหมายว่า เป็นแบบสอบที่จัดเป็นมาตรฐานที่ใช้สำหรับวัดทักษะหรือความรู้ที่เรียนมา (เยาวดี วิบูลย์ศรี 2528: 11) สำหรับ Dictionary of Education นั้น Carter V. Good ได้ให้ความหมายไว้ว่า เป็นแบบสอบที่ออกแบบสำหรับวัดความรู้ของบุคคล ทักษะความเข้าใจในสิ่งต่าง ๆเป็นต้น หรือการวัดอย่างอื่นที่ประกอบด้วยวิชาต่าง ๆและมีคะแนนให้สำหรับผู้สอบแต่ละคน แบบสอบที่ใช้กันอยู่โดยทั่วไปเรียกว่า ชุดของแบบสอบผลสัมฤทธิ์(Achievement battery) จากความหมายที่กล่าวไว้แล้วข้างต้นพอจะสรุปได้ว่า แบบสอบผลสัมฤทธิ์เป็นแบบสอบที่สร้างขึ้นเพื่อใช้ในการเรียนการสอนสำหรับวัดความรู้ที่ได้เรียนไปแล้วของนักเรียน ซึ่งอาจจะเป็นแบบสอบที่ครูสร้างขึ้นหรือแบบสอบมาตรฐานก็ได้

ข้อตกลงเบื้องต้นของแบบสอบผลสัมฤทธิ์

ในการสร้างแบบสอบผลสัมฤทธิ์ควรคำนึงถึงข้อตกลงเบื้องต้น 3 ข้อดังนี้ (เยาวดี วิบูลย์ศรี 2528 : 14)

1. เนื้อหา (content) หรือ ทักษะ (skill) ของโดเมน (domain) ที่ครอบคลุมในแบบสอบผลสัมฤทธิ์นั้นจะต้องสามารถจำกัดในรูปของพฤติกรรมที่มีความเฉพาะเจาะจงในลักษณะที่จะสื่อสารไปยังบุคคลอื่นได้ ถ้าเป้าหมายทางการศึกษาที่ไม่สามารถจำกัดในรูปพฤติกรรมแล้วย่อมไม่สามารถที่จะวัดในลักษณะของผลสัมฤทธิ์ได้

2. ผลิตผลที่แบบสอบผลสัมฤทธิ์วัดนั้นจะต้องเป็นผลิตผลเฉพาะที่ต้องการเท่านั้น ตัวอย่าง เช่น ถ้าต้องการวัดว่านักเรียนได้เรียนรู้อะไรในชั้นเรียนโดยเฉพาะ แต่แบบสอบผลสัมฤทธิ์ที่สร้างขึ้นมานั้นกลับไปวัดเพียงประสบการณ์ทางอ้อมที่ได้มาจากนอกชั้นเรียนเท่านั้น ย่อมแสดงว่าแบบสอบนั้นมิได้วัดสิ่งที่เป็นจริงตามสิ่งที่ได้เรียนรู้ไป ดังนั้นผลิตผลที่วัดนั้นจะต้องมีความสำคัญต่อการเรียนรู้เท่านั้น

3. ผลสัมฤทธิ์ที่แบบสอบผลสัมฤทธิ์วัดได้นั้น ถ้าจะนำไปเปรียบเทียบกันแล้ว ผู้สอบทุกคนจะต้องมีโอกาสเรียนรู้สิ่งต่าง ๆที่แบบสอบครอบคลุมอย่างเท่าเทียมกัน โดยทั่วไปแล้วข้อตกลงเบื้องต้นข้อนี้จะบรรลุผลได้ดีสำหรับแบบสอบผลสัมฤทธิ์ที่ครูสร้างขึ้นเพื่อใช้ในชั้นเรียนเนื่องจากนักเรียนต่างก็เรียนในวิชาเดียวกันและได้ประสบการณ์จากการทำแบบฝึกหัด

เฉพาะเจาะจงเช่นเดียวกัน

ประเภทของแบบสอบผลสัมฤทธิ์

เราสามารถจำแนกประเภทของแบบสอบผลสัมฤทธิ์ได้หลายประเภทดังต่อไปนี้

1. จำแนกตามขอบข่ายของเนื้อหาวิชาที่วัด (content area) เช่น แบบสอบผลสัมฤทธิ์บางประเภทจะวัดเนื้อหาวิชาคณิตศาสตร์ ประวัติศาสตร์ ภาษาไทย เป็นต้น ในต่างประเทศมักจะรวมแบบสอบให้เป็นชุด ๆ (series of achievement batteries) ซึ่งส่วนใหญ่จะเป็นแบบสอบมาตรฐานในทุกระดับชั้น

2. จำแนกตามลักษณะหน้าที่ทั่วไปของแบบสอบ โดยแบ่งแบบสอบผลสัมฤทธิ์ออกเป็น 3 ประเภท ดังนี้

2.1 แบบสอบเพื่อการสำรวจผลสัมฤทธิ์ (survey tests)

2.2 แบบสอบเพื่อวินิจฉัยผลสัมฤทธิ์ (diagnostic tests)

2.3 แบบสอบเพื่อวัดความพร้อม (readiness tests)

สำหรับ Mehrens and Lehman (Mehrens and Lehman 1978: 467) ได้เสนอไว้ดังนี้

1) เนื้อหาวิชาเดียว (single-subject matter) ใช้วัดเนื้อหาวิชาเดี่ยว ๆ

2) ชุดสำรวจวิชา (survey or survey battery) ประกอบด้วยแบบสอบหลายชุดวิชาใช้กับตัวอย่างประชากรกลุ่มเดียวกันใช้วัดว่าแต่ละบุคคลมีความรู้อยู่มากน้อยเท่าใด มีความรู้ในวิชาต่าง ๆนั้นหรือไม่

3) วินิจฉัย (diagnostic) ใช้ในการพิจารณาว่าทำไมบุคคลประสบผลสัมฤทธิ์เพียงพอหรือไม่เพียงพอแยกให้เห็นจุดอ่อนและจุดเด่นเฉพาะของแต่ละบุคคลเฉพาะความรู้แต่ละวิชา ใช้ในการแนะแนว การสอนซ่อมเสริม เป็นต้น

3. จำแนกตามคำตอบที่ใช้ โดยทั่วไปแล้วแบบสอบผลสัมฤทธิ์ส่วนใหญ่ที่ใช้กันจะเป็นประเภทข้อเขียน (paper and pencil test) ซึ่งมีการแยกออกเป็น 2 ระดับ คือ ระดับ Recognition และ ระดับ Recall (เยาวดี วิบูลย์ศรี 2528: 18) ซึ่งระดับ Recogniton นั้นจะประกอบด้วยคำตอบหลายคำตอบที่เป็นไปได้รวมอยู่ในคำถามผู้ตอบจะต้องเลือกคำตอบระหว่างตัวเลือกต่าง ๆตัวอย่างของข้อสอบประเภทนี้ได้แก่แบบสอบประเภทหลายตัวเลือก (multiple-choice test) แบบสอบแบบถูก-ผิด

(true-false test) และแบบสอบแบบจับคู่(matching) สำหรับระดับRecall นั้น ผู้สอบจะต้องให้คำตอบเองและต้องเป็นคำตอบที่ถูกต้องเหมาะสม ตัวอย่างของข้อสอบประเภทนี้ได้แก่ แบบสอบจำพวกเติมคำและตอบสั้น

<u>ประโยชน์ของแบบสอบผลสัมฤทธิ์</u> (สุภาพ วาดเขียน 2525: 144)

1. สำรวจทั่ว ๆไปเกี่ยวกับตำแหน่งทางการเรียนในโรงเรียนเมื่อเทียบกับเกณฑ์ปกติทำให้เข้าใจนักเรียนได้ดีขึ้น

2. การแนะแนวและประเมินค่าเกี่ยวกับการสอบได้ตกของแต่ละบุคคล จุดอ่อนและจุดเด่นของแต่ละบุคคล การสอนซ่อมเสริมให้กับนักเรียนฉลาดและนักเรียนที่ต้องการความช่วยเหลือ การปรับปรุงการสอนเมื่อทราบถึงความผิดพลาดซ้ำซากที่เกิดแก่นักเรียนเนื่องมาจากการสอนของครู

3. การจัดกลุ่มนักเรียน ความสามารถสูง ปานกลางและต่ำ หรือแยกประเภทเรียนเร็ว เรียนช้า เป็นการช่วยจัดโปรแกรมการเรียนการสอนทั้งโรงเรียน

4. ช่วยวิจัยการศึกษา เปรียบเทียบผลการเรียนในวิชาที่ทำการสอนแตกต่างกัน โดยใช้แบบสอบมาตรฐานเป็นเครื่องมือวัด

ในการวิจัยครั้งนี้ผู้วิจัยได้เลือกใช้แบบสอบชนิดเลือกตอบ (multiple-choice test) โดยที่แบบสอบชนิดหลายตัวเลือกเป็นแบบสอบที่นิยมใช้กันมากในปัจจุบัน เพราะเป็นแบบสอบที่สามารถจำแนกระดับความรู้ต่าง ๆได้ดีกว่าแบบสอบประเภทถูกผิด หรือประเภทอื่น ๆ (กานดา พูนลาภทวี 2528 : 63) โดยเฉพาะถ้าสร้างให้มีมาตรฐาน แบบสอบชนิดหลายตัวเลือกจะสามารถวัดระดับการเรียนรู้ขั้นสูงได้ เช่น ระดับการวิเคราะห์ การสังเคราะห์ ตลอดจนการประเมินผล แต่ทั้งนี้การสร้างแบบสอบแบบหลายตัวเลือกให้มีคุณภาพดี ผู้สร้างจำเป็นต้องมีความเข้าใจในหลักการออกข้อสอบและเข้าใจกระบวนการคิดของสมอง ตลอดจนเนื้อหาวิชาที่จะเขียนข้อสอบเป็นอย่างดีประกอบกัน

<u>โครงสร้างของแบบสอบชนิดหลายตัวเลือก</u>

ข้อสอบชนิดหลายตัวเลือกโดยทั่วไปจะประกอบด้วยสองส่วน คือ คำถาม (stem) และอีกส่วนหนึ่งเป็นตัวเลือก (option) ซึ่งประกอบด้วยคำตอบถูก 1 คำตอบ (key) และที่เหลือเป็นตัวลวง (distracters) ลักษณะของคำถามอาจจะเป็นคำถามโดยตรง (direct question) หรือเป็นข้อความที่ไม่สมบูรณ์ (incomplete statement) สำหรับรูปแบบที่เป็นคำถามโดยตรงเหมาะที่จะใช้กับผู้สอบที่มีระดับความรู้ต่ำ โดยทั่วไปแล้ว

แบบข้อความที่ไม่สมบูรณ์เป็นข้อสอบที่สั้นกระทัดรัดได้ใจความ ภาษาที่ใช้ไม่ฟุ่มเฟือย จึงเป็นที่นิยมใช้กันมาก

รูปแบบของข้อสอบชนิดหลายตัวเลือกมีอยู่หลายรูปแบบซึ่งอาจกล่าวโดยแยกเป็นประเภทได้ดังนี้ (กานดา พูนลาภทวี 2528 : 64-65)

1. แบบคำตอบถูกต้องข้อเดียว (one correct answer) เป็นข้อสอบที่มีคำตอบที่ถูกต้องตามหลักวิชาเพียงข้อเดียว ตัวลวงอื่น ๆเป็นคำตอบที่ผิด

2. แบบคำตอบดีที่สุด (best answer) ข้อสอบแบบนี้จะมีคำตอบถูกหลายตัวแต่ระดับของความถูกต้องต่างกัน ผู้ตอบจะต้องพิจารณาว่าตัวเลือกใดถูกต้องมากที่สุด

3. แบบให้หาที่ผิดหรือตรงกันข้าม (reverse) ข้อสอบแบบนี้จะให้ผู้ตอบหาที่ผิดในส่วนที่เป็นปัญหาหรือในตัวลวง หรือให้ผู้ตอบหาสิ่งที่ตรงกันข้ามกับสิ่งที่กำหนดให้

4. อุปมาอุปไมย (analogy) เป็นข้อสอบที่ให้ผู้ตอบค้นหาความสัมพันธ์ระหว่างสิ่ง 2 สิ่งที่กำหนดให้ว่ามีความสัมพันธ์กันอย่างไร แล้วใช้ความสัมพันธ์นี้หาส่วนที่สี่เมื่อกำหนดส่วนที่สามมาให้

ข้อเสนอแนะในการเขียนข้อสอบชนิดเลือกตอบหลายตัวเลือก

นักวัดผลทางการศึกษาได้ให้ข้อเสนอแนะในการเขียนข้อสอบชนิดเลือกตอบหลายตัวเลือกไว้หลายท่าน จะขอนำมาเสนอไว้ ณ ที่นี้เพียง 2 ท่าน คือ Hambleton และ Eignor กับของ Berk ซึ่งมีรายละเอียดดังต่อไปนี้

Hambleton และ Eignor (Hambleton and Eignor : 1978) ได้ให้ข้อเสนอแนะในการเขียนข้อสอบชนิดเลือกตอบหลายตัวเลือก ดังนี้

1. ข้อสอบแต่ละข้อออกแบบเพื่อวัดจุดประสงค์ที่สำคัญใช่หรือไม่
2. ตัวคำถามของข้อสอบมีความชัดเจนในการกำหนดปัญหาใช่หรือไม่
3. ข้อสอบมีลักษณะที่เป็นคำถามที่สามารถตอบได้ใช่หรือไม่
4. ข้อสอบแต่ละข้อมีเนื้อหาที่มีความสัมพันธ์กัน
5. ข้อสอบแต่ละข้อเขียนถูกต้องตามหลักไวยากรณ์ใช่หรือไม่
6. ข้อสอบมีจำนวนของข้อความเป็นปฏิเสธอยู่น้อยมากใช่หรือไม่
7. ถ้าข้อสอบเป็นประโยคปฏิเสธ มีความชัดเจนในการเน้นคำปฏิเสธใช่หรือไม่
8. มีคำตอบที่ถูกที่สุดเพียงข้อเดียว
9. หลีกเลี่ยงการใช้คำตอบที่ว่า " ถูกทุกข้อ " และ " ไม่มีข้อใดถูก "
10. ตัวเลือกต่าง ๆมีตัวลวงที่มีเหตุผลชัดเจนที่ใช้ลวงผู้สอบที่ไม่มีทักษะความ

สามารถให้เลือกตอบได้

11. หลีกเลี่ยงการให้แนวทางหรือรายละเอียดเกี่ยวกับตัวเลือกที่ถูกต้อง

12. คำตอบทุกคำตอบมีหลักไวยากรณ์ที่แน่นอนตลอดทั้งแบบสอบ

13. หลีกเลี่ยงคำตอบ 2 คำตอบที่มีความหมายคล้ายกัน(เหมือนกัน) จนทำให้ตัดสินไม่ได้ว่าข้อใดถูกต้อง

14. ในแต่ละตัวเลือกควรหลีกเลี่ยงคำขยายที่ว่า "บางครั้ง " และ" บ่อย ๆ "

15. ลักษณะของตัวลวงต้องพิจารณาแล้วว่ามีความสำคัญพอ ๆกับตัวเลือกที่เป็นคำตอบที่ถูกต้อง

16. คำตอบทุกคำตอบควรมีความยาวหรือความซับซ้อนพอ ๆกัน

17. คำตอบทุก ๆคำตอบควรมีลักษณะเป็นเรื่องราวเดียวกัน

18. คำตอบที่ถูกมีความยาวพอเหมาะไม่ยาวจนเป็นเหตุให้แยกแยะได้ว่าเป็นคำตอบที่ถูกต้อง

19. คำตอบแต่ละคำตอบเขียนแยกกันให้เห็นอย่างชัดเจน

20. มีตัวอักษรกำกับหน้าคำตอบทุก ๆข้อ

21. มีการใช้ความรู้ใหม่ ๆในการตั้งปัญหาที่จะวัดความเข้าใจหรือความสามารถในการประยุกต์กฎเกณฑ์ต่าง ๆ

สำหรับ Berk (Berk 1980 : 119-120)ได้ให้ข้อเสนอแนะในการเขียนข้อสอบชนิดเลือกตอบหลายตัวเลือกดังนี้

1. ข้อคำถามมีการเขียนที่ชัดเจนสำหรับที่มุ่งจะวัดในกลุ่มผู้สอบนั้น ๆ

2. ข้อคำถามมีความสำคัญที่เป็นอิสระจากกันในแต่ละข้อ

3. ข้อคำถามกำหนดไว้อย่างชัดเจนในแต่ละข้อ

4. ตัวเลือกแต่ละข้อมีการเขียนที่ชัดเจนที่มุ่งจะวัดในกลุ่มผู้สอบนั้น ๆ (เหมาะสมกับวัยของผู้ตอบ)

5. ตัวเลือกแต่ละข้อมีความเป็นอิสระจากกัน

6. มีคำตอบที่ถูกที่สุดเพียงข้อเดียว

7. ไม่มีการใช้คำว่า " บ่อย ๆ " " ไม่ทั้งหมด " หรือ " ทั้งหมด "

8. ผู้สอบมีโอกาสที่จะทำผิดได้เท่า ๆกัน

9. หลีกเลี่ยงตัวเลือกที่ว่า " ถูกหมดทุกข้อ "

10. ตัวเลือกถูกแต่ละข้อมีการสลับที่ข้อโดยการสุ่ม

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย
013072

11. ตัวเลือกแต่ละข้อมีการจัดตามลำดับเหตุผล

12. หลีกเลี่ยงคำซ้ำซ้อน หรือที่จะเป็นแนวทางในการตอบคำถามข้อนั้น ๆ

13. ตัวเลือกทุกตัวเขียนโดยใช้เครื่องหมายวรรคตอน และไวยากรณ์ที่ถูกต้องตามหลักวิชา

14. ตัวเลือกทุกตัวควรมีความยาวพอ ๆกัน

15. ประโยคปฏิเสธมีการขีดเส้นใต้ไว้

16. ข้อคำถามและตัวเลือกเขียนได้ถูกต้องเหมาะสมกับโอกาสและกาลเทศะ

17. รูปแบบของข้อคำถามมีความเหมาะสมสำหรับที่จะวัดตามจุดประสงค์ที่ตั้งไว้

ข้อดีและข้อจำกัดของข้อสอบชนิดเลือกตอบหลายตัวเลือก มีดังนี้

<u>ข้อดี</u>

1. สามารถวัดการเรียนรู้ตั้งแต่ระดับต่ำสุดจนถึงระดับสูงสุดได้

2. สามารถออกข้อสอบได้มากข้อจนสามารถวัดสิ่งที่ต้องการได้ครอบคลุมเนื้อหาวิชาที่ต้องการ

3. ตรวจคะแนนได้ง่าย ถูกต้อง และรวดเร็ว สามารถใช้เครื่องจักรกลหรือบุคคลอื่นตรวจแทนผู้ออกข้อสอบได้

4. สามารถวินิจฉัยสาเหตุที่ทำข้อสอบผิด โดยพิจารณาจากการเลือกตัวเลือกต่าง ๆได้

5. สามารถนำมาวิเคราะห์ข้อสอบ เก็บไว้ใช้ในโอกาสต่อไปตลอดจนจัดทำเป็นแบบสอบมาตรฐาน หรือธนาคารข้อสอบ(Item Bank) ได้

<u>ข้อจำกัด</u>

1. ไม่เหมาะในการวัดสิ่งที่ต้องรวบรวมและเสนอความคิดริเริ่มสร้างสรรค์

2. ผู้ไม่มีความรู้สามารถตอบถูกโดยการเดาได้ง่าย

3. ในการสร้างข้อสอบให้ดีจะต้องใช้เวลานานมาก

<u>งานวิจัยที่เกี่ยวข้องของแบบสอบผลสัมฤทธิ์</u>

การศึกษาวิจัยในเรื่องแบบสอบผลสัมฤทธิ์ทั้งในประเทศและต่างประเทศส่วนใหญ่จะพิจารณาในเรื่องประเภทของแบบสอบชนิดต่าง ๆว่ามีคุณภาพเท่าเทียมกันหรือแตกต่างกันเพียงใด โดยการศึกษาในแต่ละเรื่องจะมีเทคนิคต่าง ๆมากมายผลที่ได้นั้นบางครั้งก็มีความตรงกันหรือขัดแย้งกันบ้าง ซึ่งก็แล้วแต่เหตุผลของผู้ที่ดำเนินการวิจัยและขึ้นอยู่กับลักษณะของ

กลุ่มตัวอย่างที่นำมาใช้ในการวิจัยแต่ละเรื่อง ซึ่งสามารถหาอ่านได้มากมาย ผู้วิจัยขอสรุปเฉพาะงานวิจัยที่เกี่ยวข้องกับงานวิจัยครั้งนี้คือ

Hambleton Mills และ Simon (Hambleton Mills and Simon 1983) ได้ศึกษาข้อมูลที่สร้างขึ้นจากเครื่องคอมพิวเตอร์ และหาคำตอบจากประเด็นต่อไปนี้

1. ผลกระทบของการใช้ข้อสอบที่มีค่าความยากต่างกันต่อการกำหนดความยาวของแบบสอบ

2. วิธีการต่าง ๆในการสุ่มข้อสอบขึ้นมาทดสอบที่มีผลต่อความยาวของข้อสอบที่ใช้ในความหมายของความคงที่ในการจัดประเภทบุคคล

3. ผลกระทบของค่าอำนาจจำแนกที่มีต่อความยาวของแบบสอบอิงเกณฑ์

ผลการศึกษาพบว่า

1. แบบสอบที่มีความยาวมากขึ้น(จำนวนข้อมากขึ้น) จะมีค่าความเที่ยงเพิ่มขึ้น

2. แบบสอบที่มีความยากใกล้เคียงกัน จะมีค่าความเที่ยงเพิ่มขึ้น ณ ที่ระดับความยาวเท่ากัน

3. แบบสอบที่มีค่าอำนาจจำแนกสูงขึ้น จะมีค่าความเที่ยงสูงกว่าแบบสอบที่เป็นคู่ขนานแบบสุ่ม ซึ่งมีการกระจายของค่าความยากเท่ากันและมีความยาวเท่ากัน

Hambleton และ Gruijter (Hambleton and Gruijter 1983) ได้ศึกษามาตรการการกำหนดความยาวของแบบสอบอิงเกณฑ์ที่สร้างข้อสอบที่มีการกระจายของค่าความยากและค่าอำนาจจำแนกต่างกัน โดยใช้ทฤษฎีการตอบสนองข้อสอบซึ่งผลการศึกษาโดยใช้ข้อมูลจากเครื่องคอมพิวเตอร์พบว่า

1. การใช้ข้อสอบที่มีระดับความยากพอเหมาะกับระดับความสามารถของผู้สอบจะลดความคลาดเคลื่อนลงได้มากกว่าการใช้แบบสอบที่สร้างขึ้นจากการสุ่มข้อสอบขึ้นมาใช้

2. แบบสอบที่มีความยาวเท่ากัน ข้อสอบที่มีค่าความยากเท่ากันและตรงกับระดับความสามารถของผู้สอบ แบบสอบที่ประกอบด้วยข้อสอบที่มีค่าอำนาจจำแนกสูงกว่าจะได้ค่าความคลาดเคลื่อนต่ำกว่า

Haladyna และ Roid (Haladyna and Roid 1983) ได้ศึกษาเปรียบเทียบผลของความคลาดเคลื่อนในการประมาณค่าความสามารถของบุคคลจากการใช้แบบสอบต่างกัน 2 แบบ คือ แบบแรกเป็นแบบสอบที่สุ่มข้อสอบขึ้นมาจากประชากรข้อสอบที่มีอยู่ทั้งหมด และแบบที่สองเป็นการเลือกข้อสอบให้เหมาะกับระดับความสามารถของผู้สอบ โดยศึกษาจากรายวิชาต่าง ๆ 4 วิชา ซึ่งประกอบด้วยข้อสอบที่ผ่านการวิเคราะห์ด้วย

รูปแบบของราสช์มาแล้ว ข้อสอบเหล่านี้เป็นประชากรข้อสอบทั้งหมดในโดเมน จากนั้นทดลองสุ่มสร้างแบบสอบให้มีจำนวนเป็น 10,20,30 และ 40 ข้อตามลำดับ แล้วศึกษาเปรียบเทียบอัตราความแตกต่างเฉลี่ยระหว่างคะแนนที่สอบได้กับคะแนนโดเมนที่ประมาณได้ (The Average Absolute Difference between Observed and Domain Scores) ตัวย่อ AAD จากแบบสอบที่มีความยาวระดับต่าง ๆที่กำหนดหารด้วยส่วนเบี่ยงเบนมาตรฐานเรียกค่าที่ได้ว่าอัตราส่วน AAD/SD. โดยในการสร้างแบบสอบย่อยจะมีวิธีการสร้าง 4 วิธี คือ

1. สุ่มข้อสอบขึ้นมาจากโดเมน
2. เลือกข้อสอบให้มีความยากตรงกับระดับความสามารถของผู้สอบ
3. เลือกข้อสอบให้มีค่าความยากกระจายมากกว่าระดับความสามารถของผู้สอบเล็กน้อย
4. เลือกข้อสอบให้มีค่าความยากกระจายมากกว่าระดับความสามารถของผู้สอบ

จากการศึกษาพบว่า การสอบที่จัดให้ข้อสอบมีค่าความยากตรงกับระดับความสามารถของผู้สอบจะให้ค่าอัตราส่วน AAD/SD. ต่ำสุด แบบสอบที่ได้รับการสุ่มข้อสอบขึ้นมาดัชนีดังกล่าวต่ำเป็นลำดับที่สอง และแบบสอบที่มีค่าความยากกระจายมากจะยิ่งมีค่าดัชนีดังกล่าวสูงมากยิ่งขึ้น ผู้เขียนทั้งสองอภิปรายไว้ว่า ในการดำเนินการสอบไม่ว่าจะใช้วิธีการใดจัดระดับความยากให้เหมาะสมกับระดับความสามารถของผู้สอบจะใช้ได้ดีในกรณีที่ต้องการวัดคุณลักษณะความสามารถที่ต่อเนื่องกัน การทดสอบจะเป็นการพยายามหาตำแหน่งที่ของระดับความสามารถของบุคคลในช่วงความต่อเนื่องนั้น แต่อย่างไรก็ดีเนื่องจากวิธีการสุ่มข้อสอบก็ยังคงให้ผลดีเป็นลำดับที่สองและมีความสะดวกในการปฏิบัติจึงน่าจะใช้เทคนิคทั้งสองในการสร้างแบบสร้างแบบสอบอิงเกณฑ์ การที่จะตัดสินใจว่าวิธีการใดจะดีกว่าจะต้องรอดูผลจากประโยชน์ที่ได้จากการสอบจริง ๆ

ตอนที่ 2 มโนทัศน์เกี่ยวกับรูปแบบของราสช์ และงานวิจัยที่เกี่ยวข้อง

ทฤษฎีการทดสอบที่ได้เข้ามามีบทบาทในการวัดคุณลักษณะ(trait) ที่แฝงอยู่ในของบุคคลในปัจจุบันก็คือ ทฤษฎีการตอบสนองข้อสอบ(Item Response Theory) ซึ่งทฤษฎีนี้แสดงถึงความสัมพันธ์ของคุณลักษณะของข้อสอบด้วยรูปแบบทางคณิตศาสตร์(Mathematical Model) โดยมีจุดมุ่งหมายที่จะอธิบายคุณลักษณะของข้อสอบและคุณลักษณะของสิ่งที่จะวัดในตัวบุคคลด้วยค่าพารามิเตอร์(parameter) ของข้อสอบและของบุคคลตามลำดับสำหรับรูปแบบทางคณิตศาสตร์ที่นำมาใช้อธิบายในทฤษฎีการตอบสนองข้อสอบมีหลายรูปแบบแต่ที่นิยมใช้มี 2 รูปแบบ คือ Normal Ogive Model และ Logistic Model แต่ในปัจจุบันนักทดสอบนิยมใช้รูปแบบโลจีสติกมากกว่าเนื่องจากรูปแบบ Normal Ogive มีปัญหายุ่งยากในการคำนวน ซึ่ง Birnbuam ได้พัฒนารูปแบบโลจีสติกขึ้นมาใช้เพื่อลดความยุ่งยากในการคำนวนและสามารถใช้เครื่องคอมพิวเตอร์ช่วยคำนวนได้ผลดีในการประมาณค่าพารามิเตอร์ของข้อสอบและของบุคคล Wood , Wingersky และ Lord ได้ใช้โปรแกรมโลจีสช่วยในการคำนวนการประมาณค่าพารามิเตอร์โดยใช้ Maximum Likelihood Method ในการคำนวน รูปแบบโลจีสติกสามารถแบ่งออกเป็น 3 รูปแบบดังนี้ (Hambleton and Cook 1977: 80-82)

1. รูปแบบที่ใช้พารามิเตอร์ 2 ตัว(Two-parameters Logistic Model) Birnbuam ได้พัฒนาขึ้นมาในปี ค.ศ.1968 โดยเสนอรูปแบบของคุณลักษณะแฝง(Latent Trait) ที่ใช้กับโค้งลักษณะของข้อสอบ(Item Characteristic Curve) จากรูปแบบโลจีสติกไว้ดังนี้คือ

$$P_g(\theta) = \frac{e^{Da_g(\theta - b_g)}}{1 + e^{Da_g(\theta - b_g)}} \quad ; g = 1,2,\ldots\ldots n$$

เมื่อ $P_g(\theta)$ คือ ความน่าจะเป็นในการตอบข้อสอบของผู้สอบด้วยความสามารถที่จะตอบข้อสอบข้อที่ g ถูก

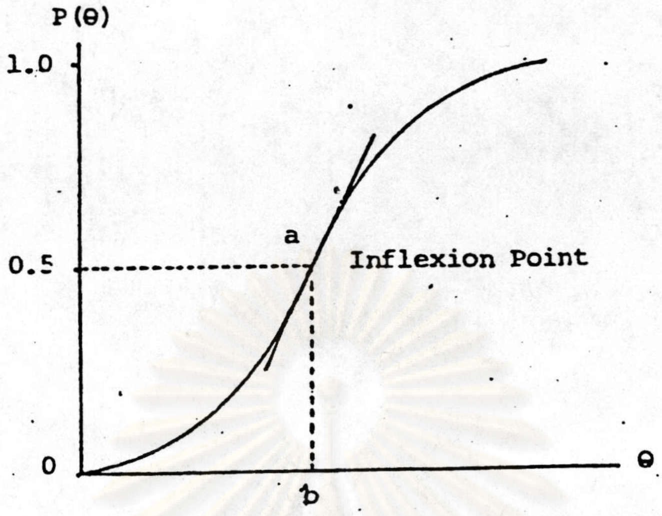
a_g คือ ค่าอำนาจจำแนกของข้อสอบข้อที่ g

b_g คือ ค่าความยากของข้อสอบข้อที่ g

D คือ scaling factor มีค่า 1.702

e คือ ค่าคงที่มีค่าเท่ากับ 2.7182818...

แผนภาพที่ 1 แสดงความหมายของค่า Item parameter (2 พารามิเตอร์)ของข้อสอบ

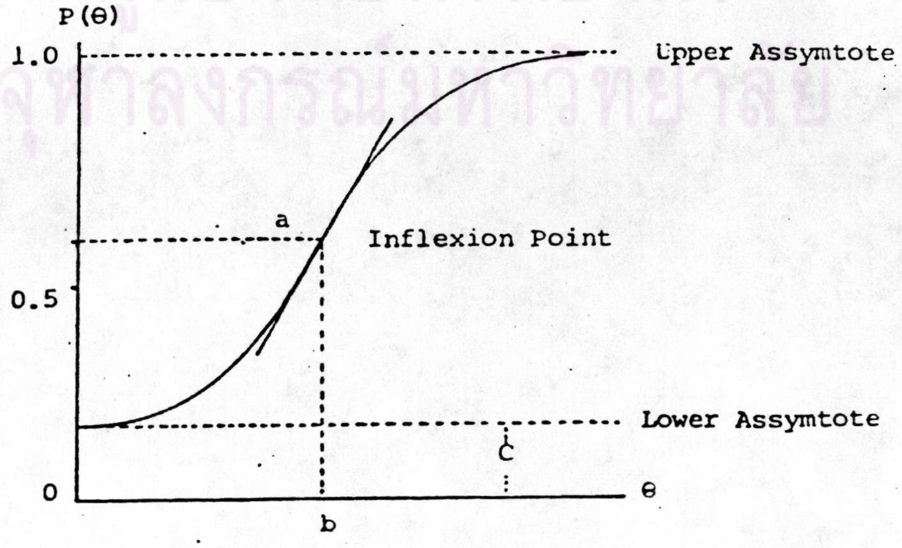


2. รูปแบบที่ใช้พารามิเตอร์ 3 ตัว (Three-parameters Logistic Model) รูปแบบ 3 พารามิเตอร์สามารถเขียนได้จากรูปแบบ 2 พารามิเตอร์ โดยเพิ่มพารามิเตอร์ตัวที่ 3 คือ ค่าการเดาเข้าไป ใช้สัญลักษณ์ c_g มีสูตรการคำนวนดังนี้

$$P_g(\Theta) = c_g + (1-c_g)\frac{e^{Da_g(\Theta - b_g)}}{1 + e^{Da_g(\Theta - b_g)}}$$

เมื่อ c_g คือ ค่าการเดาของข้อสอบข้อที่ g

แผนภาพที่ 2 แสดงความหมายของค่า Item parameter (3 พารามิเตอร์)ของข้อสอบ

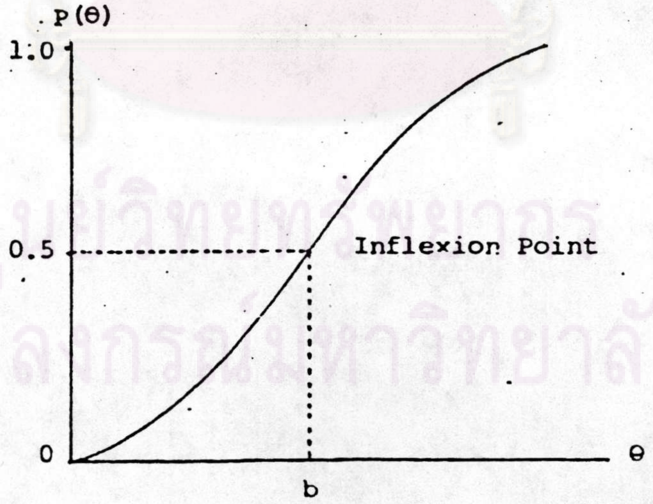


ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

3. รูปแบบที่ใช้พารามิเตอร์ 1 ตัว (One-parameter Logistic Model) รูปแบบพารามิเตอร์ 1 พารามิเตอร์นี้ได้มีนักคณิตศาสตร์ชาวเดนมาร์กชื่อ Georg Rasch ได้พัฒนารูปแบบในแนวทางของเขาขึ้นในปี ค.ศ.1960 ซึ่งใช้ Poisson Model มาแก้ปัญหาแต่มีความยุ่งยากในการคำนวนเมื่อจำนวนประชากรที่ใช้เพิ่มมากขึ้น Rasch ได้แก้ปัญหาโดยพัฒนารูปแบบโลจีสติกขึ้นมาใช้แทนผลงานของ Rasch ได้ถูกพิสูจน์โดยนักวัดผลหลายท่านแล้วมีความเห็นว่าเป็นกรณีพิเศษของรูปแบบ 2 พารามิเตอร์ในรูปแบบโลจีสติกที่ Birnbuam ได้พัฒนาขึ้นซึ่งมีข้อตกลงเบื้องต้นที่พิจารณาคือ ค่าอำนาจจำแนกของข้อสอบ (a) เป็นค่าคงที่สำหรับข้อสอบแต่ละข้อและค่าความยากของข้อสอบแปรเปลี่ยนไป ซึ่งสามารถเขียนในรูปฟังก์ชั่นของลักษณะข้อสอบ(Item Characteristic Function) ได้ดังนี้

$$P_g(\Theta) = \frac{e^{D\bar{a}(\Theta - b_g)}}{1 + e^{D\bar{a}(\Theta - b_g)}} \quad ; g = 1,2,....n$$

แผนภาพที่ 3 แสดงความหมายของค่า Item parameter (1 พารามิเตอร์)ของข้อสอบ หรือรูปแบบของราสช์



ซึ่งWright ได้กล่าวว่า การเขียนรูปแบบที่ใช้Dā นั้น Wright ได้รวมเข้าไว้ในสเกลของความสามารถ (Θ) แล้วเขียนขึ้นใหม่ดังนี้

$$P_g(\Theta) = \frac{e^{(\Theta' - b'_g)}}{1 + e^{(\Theta' - b'_g)}} \quad ; g = 1,2,...n$$

ต่อมา Rasch Wright และคนอื่น ๆมีความเห็นว่าควรจะเขียนให้เป็นแนวเดียวกันโดยมีรูปแบบที่ใช้พารามิเตอร์ 1 ตัวมีสูตรดังนี้

$$P_g(\theta) = \frac{\theta^*}{\theta^* + b_g^*} \quad ; g = 1,2,\ldots n$$

เมื่อ $\theta^* = e^{D\bar{a}\theta}$ และ $b_g^* = e^{D\bar{a}b_g}$

ค่า $e^{D\bar{a}}$ เป็น scaling factor ในสเกลของความสามารถโดยที่ θ^* และ b_g^* กำหนดค่าขึ้นในสเกลจาก 0 ถึง ถ้าใส่ log ความสามารถและ log ค่าความยาก ประมาณค่าได้ ($\log \theta^*$ and $\log b_g^*$) เมื่อนั้น θ และ b และ $\log \theta^*$ และ $\log b_g^*$ สามารถวัดได้โดยสเกลร่วมกันตั้งแต่ $-\infty$ ถึง ∞ ซึ่งทำให้เราสามารถอธิบายความหมายในสเกลความสามารถต่าง ๆได้ดีขึ้น

จากข้อความที่กล่าวมาแล้วนั้นทำให้เราทราบว่ารูปแบบของราสช์สามารถใช้แทนได้กับรูปแบบพารามิเตอร์ 1 ตัวในรูปแบบโลจีสติก ซึ่งแนวคิดนี้ Wright ได้เผยแพร่ตั้งแต่ปีค.ศ.1967 เป็นต้นมา

แนวคิดพื้นฐานของรูปแบบของราสช์

สำหรับแนวคิดพื้นฐานของราสช์ในระยะแรกนั้น ราสช์ได้พิจารณาจากรูปแบบปัวซอง (poisson model) โดยที่รูปแบบปัวซองเป็นการประมาณค่าคะแนนของแบบสอบบนพื้นฐานของจำนวนข้อสอบหลาย ๆข้อสามารถที่จะนำมาประยุกต์ในเรื่องจำนวนข้อที่ตอบถูกหรือผิดในแบบสอบ ถ้าพิจารณาจำนวนคำตอบที่ถูกในแบบสอบ ความน่าจะเป็นของคำตอบถูกจะต้องน้อย ถ้าพิจารณาจำนวนคำตอบผิดในการสอบ ความน่าจะเป็นของคำตอบผิดจะต้องน้อยจึงจะเหมาะสมกับรูปแบบนี้ ซึ่งการพิจารณาการกระจายในรูปแบบปัวซองเป็นกรณีเฉพาะของการกระจายแบบไบโนเมียน (binomial) ราสช์ได้เสนอการพิจารณาจากรูปแบบปัวซองดังนี้ ราสช์พิจารณาจากอัตราส่วนของความสามารถของผู้สอบและความยากของข้อสอบ โดยอธิบายได้ด้วยข้อตกลงเบื้องต้นดังนี้

1. แบบสอบทุก ๆฉบับที่มีผู้สอบจำนวนมาก ข้อสอบจะต้องมีอิสระจากกัน ไม่มีข้อจำกัดของจำนวนข้อสอบที่จะใช้ในแต่ละแบบสอบ แต่เท่าที่ใช้ประมาณ 75 ข้อหรือมากกว่า

2. ความน่าจะเป็นเฉลี่ยในการตอบผิดมีจำนวนน้อยในแต่ละแบบสอบที่ผู้สอบทำซึ่งไม่ได้จำกัดว่าจะน้อยเพียงใด แต่ความน่าจะเป็นของการตอบผิดมักจะน้อยกว่า .10

3. ความน่าจะเป็นเฉลี่ยของผู้สอบคนที่ j ที่ตอบคำตอบผิดในแต่ละข้อของ

แบบสอบที่ i ให้ P_{ij} เป็นอัตราส่วนของ 2 พารามิเตอร์นี้

σ_i คือ ความยากของแบบสอบ

ζ_j คือ ความสามารถของผู้สอบ

$$P_{ij} = \sigma_i / \zeta_j$$

ค่า σ_i และ ζ_j ในแต่ละ P_{ij} จะมีค่าอยู่ระหว่าง 0 ถึง 1 ความน่าจะเป็นในการตอบผิดจะเพิ่มขึ้นเมื่อค่าความยากของแบบสอบเพิ่มขึ้น และความสามารถของผู้สอบลดลง ดังนั้นความน่าจะเป็นของการตอบผิดจะสูงสำหรับแบบสอบที่ยากและผู้สอบไม่มีความสามารถที่จะทำแบบสอบ P_{ij} มีความคล้ายคลึงกับคะแนนจริงใน Binomial-error model

4. ถ้าให้ X_{ij} เป็นตัวแปรหนึ่งสำหรับจำนวนผู้สอบที่ตอบผิดที่ได้จากผู้สอบคนที่ j ในแบบสอบที่ i (X_{ij} เป็นค่าเฉพาะของค่าที่วัดได้ของ X_{ij} สำหรับแต่ละ i และ j)

$$X_{ij} = N_i P_{ij} + E_{ij}$$

เมื่อ N_i คือ จำนวนข้อสอบในแบบสอบที่ i และ E_{ij} คือ ความคลาดเคลื่อนมาตรฐานสำหรับแบบสอบที่ i และผู้สอบคนที่ j ซึ่งมีข้อตกลงเบื้องต้นคล้ายกับข้อตกลงเบื้องต้นของคะแนนจริงและค่าความคลาดเคลื่อนมาตรฐานในการหาคะแนนจริงของทฤษฎีมาตรฐานดั้งเดิม (Classical Test Theory)

5. $\mathcal{E}(X_{ij}/N_i P_{ij}) = N_i P_{ij}$ เป็นค่าคาดหวังของจำนวนคำตอบผิดสำหรับผู้สอบคนที่ j ในแบบสอบที่ i $N_i P_{ij}$ หรือ $N_i \sigma_i \zeta_i$ ตัวอย่างเช่นถ้าแบบสอบมีค่า $\sigma_i = .03$ ซึ่งได้มาจากผู้สอบที่มีความสามารถ $\zeta_i = 10$ ผู้สอบจะมีค่าเฉลี่ยความน่าจะเป็น $= (.3/10) = .03$ ในการทำข้อสอบผิด ถ้าแบบสอบมีข้อสอบ 100 ข้อ เราก็สามารถคาดหวังได้ว่า ผู้สอบจะทำผิดได้ (100)(.03) หรือ 3 ข้อ ในจำนวนที่มีข้อถูก 97 ข้อ

ถ้าเราให้ λ_{ij} เป็นค่าคาดหวังของจำนวนข้อสอบผิด ที่ผู้สอบคนที่ j และแบบสอบที่ i จะได้ว่า

$$\lambda_{ij} = N_i \sigma_i / \zeta_j$$

จากข้อตกลงเบื้องต้น5 ข้อของรูปแบบปัวซองที่ราสช์เสนอสามารถทำนายได้จาก

$$P(X_{ij} = x_{ij}/ \lambda_{ij}) = \frac{e^{-\lambda_{ij}} \lambda_{ij}^{x_{ij}}}{x_{ij}!}, x_{ij} = 0,1,2,..n$$

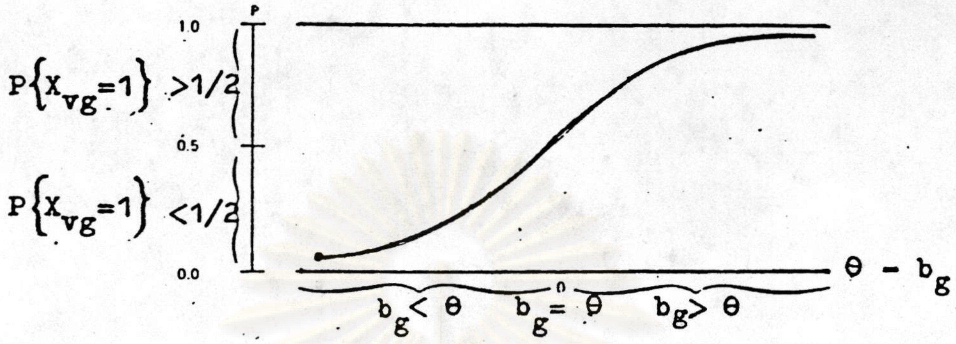
ในการพิจารณาการกระจายแบบปัวซองให้พารามิเตอร์ λ_{ij} เป็นฟังก์ชั่นของการกระจาย e เป็นค่าคงที่มีค่า 2.718 ถ้าให้ $\lambda_{ij}=1$ ความน่าจะเป็นของความไม่คลาดเคลื่อนเป็น $e^{-1}(1^0)/0!=.37$ นั่นคือ ความน่าจะเป็นในการตอบผิด 1 ข้อ คือ $e^{-1}(1^1)/1! = .37$ ถ้าความน่าจะเป็นในการตอบผิด 2 ข้อ คือ $e^{-1}(1^2)/2!=.18$

ราสช์ได้อธิบายการประมาณค่าความคลาดเคลื่อนมาตรฐานของความสามารถและความยากว่าสามารถทำได้โดยการประมาณค่าจะต้องพิจารณาผู้สอบแต่ละคนและแบบสอบแต่ละชุดในทฤษฎีมาตรฐานดั้งเดิม ค่าความคลาดเคลื่อนมาตรฐานในการวัด (Standard error of measurment) จะประมาณค่าได้จากการประมาณค่าคะแนนจริงและเป็นค่าที่เท่ากัน(เหมือนกัน) สำหรับการวัดซึ่งใช้คะแนนจริงทั้งหมด ซึ่งราสช์กล่าวว่าค่าความคลาดเคลื่อนมาตรฐานของการประมาณค่าแต่ละครั้งจะแตกต่างกัน เมื่อความสามารถต่างกัน และแบบสอบมีความยากต่างกัน

รูปแบบปัวซองของราสช์เหมาะสมสำหรับชุดของแบบสอบที่มีความน่าจะเป็นในการตอบผิดน้อย ๆและจำนวนข้อสอบมาก ๆ ราสช์จึงได้เสนอรูปแบบอื่นคือ รูปแบบโลจีสติกที่ใช้ได้เหมาะสมกว่าซึ่งไม่ต้องจำกัดความน่าจะเป็นของความคลาดเคลื่อนให้น้อยและจำนวนข้อมาก ๆสามารถประยุกต์กับการสอบได้หลายแบบซึ่งดีกว่ารูปแบบปัวซองของราสช์

ต่อไปนี้จะเสนอแนวคิดของราสช์ที่ใช้กับรูปแบบโลจีสติก โดยที่ราสช์ได้เสนอว่าโอกาสที่คนจะทำข้อสอบได้ถูกหรือไม่นั้นขึ้นอยู่กับความสามารถของตนเอง(ability parameter: θ) และระดับความยากของข้อสอบ (difficulty parameter: b) (Wright 1977: 97-99) เช่น ถ้า b = .5 และ θ = .5 โอกาสที่บุคคลจะสามารถทำข้อสอบข้อนั้นได้ถูกต้องประมาณ 50 % และถ้าหากความสามารถของบุคคล(θ) น้อยกว่าความยาก(b) แล้ว โอกาสที่จะทำข้อสอบข้อนั้นน้อยกว่าความยาก(b) แล้ว โอกาสที่จะทำข้อสอบข้อนั้นก็ย่อมจะน้อยกว่า 50 % และในทำนองเดียวกัน ถ้าค่าความสามารถ(θ) มากกว่าค่าความยากของข้อสอบ(b) แล้ว โอกาสที่บุคคลนั้นจะทำข้อสอบได้ถูกก็มีมากกว่า 50 % ดังแผนภาพที่ 4

แผนภาพที่ 4 แสดงความสัมพันธ์ระหว่างโอกาสที่จะทำข้อสอบได้ถูกต้อง(P) กับความสัมพันธ์ระหว่าง Θ และ b_g บนตัวแปร



ซึ่งอาจจะกล่าวได้ว่าโอกาสที่บุคคลจะทำข้อสอบได้ถูกหรือไม่นั้นขึ้นอยู่กับผลต่างระหว่างค่าความสามารถ (Θ) และค่าความยากของข้อสอบ (b) ซึ่งสามารถเขียนเป็นสูตรการคำนวนได้ดังนี้

$$P_g(\Theta) = \frac{e^{(\Theta - b_g)}}{1 + e^{(\Theta - b_g)}} \quad ; g = 1,2,\ldots.n$$

จะเห็นได้ว่า ถ้า $\Theta > b_g$ มากค่า $P_g(\Theta)$ จะมีค่าสูง

จากแนวคิดของรูปแบบของราสช์นี้ ค่าพารามิเตอร์ที่เกี่ยวข้องมีเพียงค่าเดียวเท่านั้นคือ ค่าความยากของข้อสอบ ค่าอำนาจจำแนกมีค่าเท่ากับ 1 โอกาสในการเดาเป็น 0 เพราะเป็นข้อตกลงเบื้องต้นว่า ข้อสอบที่จะนำมาใช้กับรูปแบบของราสช์นั้นจะต้องมีค่าอำนาจจำแนกเท่า ๆกัน หรือมีลักษณะที่จะทำให้เกิดค่าการเดาน้อยที่สุด แต่ในทางปฏิบัติลักษณะทั้งสองนี้อาศัยความแกร่ง (robustness) ของรูปแบบของราสช์แก้ไขได้

แนวความคิดที่สำคัญของรูปแบบของราสช์คือ เป็นแนวความคิดที่เชื่อว่าความเป็นปรนัยของการวัด (Objectivity of measurement) ที่ไม่สามารถหาได้จากทฤษฎีมาตรฐานดั้งเดิม ซึ่ง Whitely และ Dawis แห่งมหาวิทยาลัยมินนิโซต้าได้ทำการวิจัยมีผลสอดคล้องกับแนวความคิดนี้

1. ความเป็นอิสระจากกลุ่มตัวอย่าง (Sample-free test calibration) กล่าวคือ การคำนวนค่าพารามิเตอร์ของข้อสอบเป็นอิสระจากกลุ่มตัวอย่าง ค่าต่าง ๆของข้อสอบ เช่นค่าความยากของข้อสอบ (b) จะไม่แปรเปลี่ยนไปตามลักษณะของกลุ่มตัวอย่าง

2. ความเป็นอิสระจากข้อสอบ (Item-free person measurement)

กล่าวคือ การคำนวนค่าพารามิเตอร์ของบุคคลเป็นอิสระจากข้อสอบ เช่น ความสามารถของบุคคล(Θ) จะไม่แปรเปลี่ยนไปตามลักษณะข้อสอบ ความสามารถของบุคคลดังกล่าวจะคงที่ไม่ว่าจะวัดด้วยข้อสอบใด ข้อสอบหนึ่ง ๆที่บุคคลนั้นเคยทำแล้วจะมีค่าคงที่เมื่อวัดกับบุคคลเดิมไม่ว่าข้อสอบนั้นจะไปปรากฏอยู่ที่ส่วนใดของแบบสอบ

แนวความคิดนี้ได้รับการสนับสนุนและมีการพิสูจน์ยืนยันว่าเป็นจริงได้ โดยนักวัดผลหลายท่าน เช่น Wright และ Panchapakesan ในปี ค.ศ.1969 Wright และ Douglass ในปี ค.ศ.1975 เป็นต้น การทดสอบได้ยืนยันลักษณะดังกล่าวนี้ว่าเป็นจริงได้แม้แต่กลุ่มตัวอย่างที่ไม่ได้มาจากการสุ่มก็ตามและมีความแกร่ง (robustness) มากคือ สามารถทดสอบเพื่อคำนวนหาความสามารถของบุคคล (Θ) และความยากของข้อสอบ (b) ได้ทั้ง ๆที่กลุ่มตัวอย่างบุคคลและข้อสอบแตกต่างไปจากข้อตกลงเบื้องต้นของการใช้รูปแบบของราสช์นี้ก็ตาม เช่น ขนาดของกลุ่มตัวอย่างมีเพียง 100 คนก็ใช้ได้และไม่ต้องมาจากการสุ่ม (Wright 1977: 106) อนึ่งข้อแตกต่างของรูปแบบของราสช์กับรูปแบบโลจีสติกที่ใช้ค่าพารามิเตอร์ 1 ตัว คือรูปแบบของราสช์ประมาณพารามิเตอร์ความสามารถที่สอดคล้องกับคะแนนสอบ ส่วนรูปแบบพารามิเตอร์ 1 ตัว ประมาณค่าพารามิเตอร์ของผู้สอบแต่ละคน ดังนั้นรูปแบบของราสช์จึงนิยมใช้ในการสร้างแบบสอบมากกว่ารูปแบบพารามิเตอร์ 1 ตัวที่ใช้ในการวิเคราะห์ปัญหาการวัด(อุทุมพร จามรมาน 2529: 98)

<u>สูตรการคำนวนค่าพารามิเตอร์จากรูปแบบของราสช์</u>

ดังได้กล่าวแล้วในตอนต้นว่า ราสช์ได้พัฒนารูปแบบของเขาในรูปแบบปัวซองแต่มีข้อจำกัด ราสช์จึงหันมาพิจารณารูปแบบโลจีสติกซึ่งผลงานของเขาได้มีความสัมพันธ์และตรงกับรูปแบบพารามิเตอร์ 2 ตัวของรูปแบบโลจีสติกที่พัฒนาโดย Birnbuam และได้เป็นที่นิยมใช้กันอย่างแพร่หลายจนกระทั่งปัจจุบันนี้ จากแนวคิดของรูปแบบของราสช์ที่เกี่ยวข้องกับความสัมพันธ์ของความสามารถของบุคคล (Θ) และค่าความยากของข้อสอบเท่านั้น ความสัมพันธ์นี้เป็นความสัมพันธ์กันในเชิงคณิตศาสตร์ กล่าวคือ โอกาสของบุคคล v ที่มีระดับความสามารถ Θ (หรือ Θ_v) จะทำข้อสอบข้อที่ g ที่มีระดับความยาก b (หรือ b_g) จะทำข้อสอบได้ถูกต้องมากน้อยเพียงใดขึ้นอยู่กับระดับความแตกต่างของ $\Theta_v - b_g$ นั้นคือ

โอกาสของความสำเร็จ (odd of success) = $(\Theta - b_g)$

ค่าของ $(\Theta - b_g)$ นั้นมีค่าระหว่าง $-\infty$ ถึง ∞ แต่ว่าโอกาสของความสำเร็จ

มีได้ระหว่าง 0 ถึง 1 เท่านั้น เพื่อให้ค่าของ $(\Theta - b_g)$ เป็นค่าที่มีหน่วยเล็กลงและต่อเนื่องหลาย ๆค่าและคงที่เหมาะแก่การคำนวนจึงใส่ค่า exponential เข้าไปเพื่อจะได้ค่าตั้งแต่ 0 ถึง $+\infty$ เช่น

โอกาสของความสำเร็จ $e^{(\Theta_v - b_g)} = \exp(\Theta - b_g)$ .... (1)

ซึ่งรูปแบบนี้เรียกว่า โลจิท(logit) ค่า $e = 2.71828..$ ค่า Θ_v จะเริ่มที่ 0 และ b_g จะเริ่มที่ 0 เช่นกัน ค่าของ $\exp(\Theta - b_g)$ เป็นค่าในมาตราอันตรภาค(interval scale) ต้องการให้มีค่าจาก 0ถึง1 สมการที่ 1 อาจเขียนใหม่ได้เป็น

$$\exp(\Theta_v - b_g)/\left[1 + \exp(\Theta_v - b_g)\right] \text{ ....(2)}$$

ถ้าเราให้ $X_{vg} = 1$ เมื่อบุคคล v ตอบถูก

$x_{vg} = 0$ เมื่อบุคคล v ตอบผิด

ดังนั้นสูตรที่แสดงวาโอกาสที่บุคคล v ที่มีความสามารถ Θ_v จะทำข้อสอบข้อที่ g ที่มีความยาก b_g ได้ถูกต้องก็คือ

$$P\left\{X_{vg} = 1/\Theta_v\ ,\ b_g\right\} = \exp(\Theta_v - b_g)/\left[1 + \exp(\Theta_v - b_g)\right] \text{.... (3)}$$

และถ้าบุคคลนั้นตอบข้อสอบข้อที่ g ผิด ก็จะได้สูตร

$$P\left\{X_{vg} = 0/\Theta_v\ ,\ b_g\right\} = 1/\left[1+\exp(\Theta_v - b_g)\right] \text{ .... (4)}$$

สมการที่ 3 ก็คือรูปแบบของราสช์ หรือรูปแบบพารามิเตอร์ 1 ตัว ตัวอย่างเช่น ถ้าความแตกต่าง $(\Theta_v - b_g) = 2$ ในระบบ log-odd unit โอกาสที่บุคคลจะทำข้อสอบข้อที่ g ได้ถูกต้องคือ

$$P\left\{X_{vg} = 1/\Theta_v\ ,\ b_g\right\} = \frac{e^2}{1 + e^2} = \frac{(2.71828)^2}{1+(2.71828)^2} = .881 = 88.1\%$$

ที่กล่าวมานี้เป็นการคำนวนในกรณีที่กลุ่มตัวอย่างมีขนาดเล็ก ถ้ากลุ่มตัวอย่างใหญ่และมีการสอบหลาย ๆคนและข้อสอบหลาย ๆข้อ โดยกำหนดให้ข้อสอบทั้งหมดมี L ข้อและกลุ่มตัวอย่างมี N คน คะแนนรวมของคนที่มีความสามารถ Θ_v โดยกำหนดให้เท่ากับ

$$R_v = \sum_{g}^{L} X_{vg} \quad , v = 1,2,\ldots N \quad \text{(person score)}$$

และจำนวนคนที่สอบข้อสอบแต่ละข้อเท่ากับ

$$S_g = \sum_{g}^{N} X_{vg} \quad , g = 1,2,\ldots L \quad \text{(item score)}$$

ดังนั้นถ้าหากกำหนดให้ $((X_{vg}))$ แทนเมตริกซ์ทั้งหมดของผลการสอบแล้วโอกาสที่เกิดผลการสอบดังกล่าวก็คือ

$$P\left\{((X_{vg}))/(\Theta_v),(b_g)\right\} = \prod_{v}^{N}\prod_{g}^{L}\left[\frac{\exp(X_{vg}(\Theta_v - b_g)}{1+\exp(\Theta_v - b_g)}\right] \ldots\ldots\ldots (5)$$

เราสามารถที่จะทำผลรวมของผลคูณ $\prod_{v}^{N}$ และ $\prod_{g}^{L}$ ในสมการที่ 5 ซึ่งอยู่ในรูปของ $\exp(X_{vg}(\Theta_v - b_g))$ ให้อยู่ในรูปของ $\sum_{v}^{N}$ และ $\sum_{g}^{L}$ ได้ดังนั้น

$$\prod_{v}^{N}\prod_{g}^{L}\exp(X_{vg}(\Theta_v - b_g) = \exp\sum_{v}^{N}\sum_{g}^{L}X_{vg}(\Theta_v - b_g)$$

โดยที่ $$\sum_{v}^{N}\sum_{g}^{L}X_{vg}\Theta_v = \sum_{v}^{N}R_v\Theta_v$$

และ $$\sum_{v}^{N}\sum_{g}^{L}X_{vg}b_g = \sum_{g}^{L}S_g\Theta_v$$

สมการที่ 5 อาจเขียนได้ใหม่ดังนี้

$$P\left\{((X_{vg}))/(\Theta_v),(b_g)\right\} = \frac{\exp(\sum_{v}^{N}R_v\Theta_v)\ \exp(-\sum_{g}^{L}S_g b_g)}{\prod_{v}^{N}\prod_{g}^{L}\left[1+\exp(\Theta_v - b_g)\right]} \ldots\ldots\ldots (6)$$

สมการที่ 6 แสดงให้เห็นถึงลำดับของการประมาณค่าพารามิเตอร์ Θ_v และ b_g ซึ่งสามารถเขียนใหม่ได้ดังนี้

$$P\left\{((X_{vg}))/(\Theta_v),(b_g)\right\} = \frac{\left[\exp(\sum_{v}^{N}R_v\Theta_v)\right]\left[\exp(-\sum_{g}^{L}S_g b_g)\right]}{\prod_{v}^{N}\prod_{g}^{L}\left[1+\exp(\Theta_v - b_g)\right]} \ldots\ldots (7)$$

ศูนย์วิทยทรัพยากร จุฬาลงกรณ์มหาวิทยาลัย

สมการที่ 7 มีความสำคัญมากเนื่องจากแสดงให้ทราบว่า

1. ในการคำนวนหาค่าความสามารถของข้อสอบแต่ละข้อ (คือความยากหรือ b_g เราอาศัยแต่เพียงคะแนนรวมของแต่ละบุคคล (person score : R_v) เท่านั้นก็เป็นการเพียงพอในการกำจัด Θ_v ออกจากสมการที่ 7 การคำนวนหาค่า b_g จึงเป็นอิสระจากกลุ่มตัวอย่าง (sample-free item calibation)

2. ในการคำนวนหาค่าความสามารถของแต่ละบุคคล (Θ_v)ก็ใช้วิธีเดียวกันคือ เราเพียงแต่อาศัยคะแนนรวมของแต่ละข้อ (item score : S_g) เท่านั้นก็เพียงพอในการกำจัดค่า b_g ออกจากสมการที่ 7 ดังนั้นการคำนวนหาค่าความสามารถของแต่ละบุคคล (Θ_v)จึงเป็นอิสระจากกลุ่มข้อสอบ (item-free person measurement)

ซึ่ง การคำนวนในข้อที่ 1 และข้อที่ 2 ค่าของ Θ_v และ b_g จึงเป็นอิสระจากกันซึ่งเป็นแนวคิดสำคัญของรูปแบบของราสช์ที่ได้กล่าวมาแล้วข้างต้น

วิธีการคำนวนค่า Θ_v และ b_g โดยการกำจัดค่าใดค่าหนึ่งออกไปจากสมการ เรียกว่าวิธี Conditional Maximum Likelihood Procedure เป็นวิธีการที่ยุ่งยากในการคำนวน ซึ่ง Wright และ Douglass (Wright 1978: 17) ได้ทำการศึกษาพบว่าจะใช้วิธีกับการวิเคราะห์ข้อสอบที่ยาวเกินกว่า 25 ข้อไม่ได้ ดังนั้น Wright และ Panchapakesan ได้พัฒนาวิธีซึ่งคำนวนค่า Θ_v และ b_g โดยใช้วิธี Unconditional Maximum Likelihood Procedure: UCON ขึ้น ซึ่งสามารถใช้ได้กับการวิเคราะห์ข้อสอบที่มีจำนวนมากและน้อยได้ด้วย แม้ว่าจะมีอคติ(bias) ในการคำนวนเพียงเล็กน้อยแต่ก็แก้ไขได้ วิธีดังกล่าวนี้ในโปรถกรมไบคาล (BICAL) แนะนำให้ใช้เมื่อข้อสอบมีจำนวนน้อย(ประมาณ 25 ข้อ) และมีกลุ่มตัวอย่างขนาดใหญ่ (มากกว่า 400 คน) และเชื่อว่าการกระจายของคะแนนมีแนวโน้มว่าจะเป็นโค้งปกติ การคำนวนค่า Θ_v และ b_g ควรจะใช้วิธี Cohen's Approximation (PROX) จะได้ผลดีกว่า ในทางปฏิบัตินั้นเมื่อใช้กลุ่มตัวอย่างมากหรือขนาดใหญ่ค่าประมาณที่ได้จะมีค่าใกล้เคียงกัน

นอกจากนี้ Whitely และ Dawis ได้สนับสนุนให้เริ่มต้นประมาณค่าพารามิเตอร์ด้วยวิธี Least squares ซึ่งขึ้นอยู่กับข้อสอบโดยใช้คะแนนความถี่ของกลุ่มนั้น วิธีนี้ราสช์ได้ใช้ครั้งแรกเมื่อประมาณปี ค.ศ.1950 และได้เสนอไว้ในหนังสือของเขาเมื่อปี ค.ศ. 1960 ว่าเป็นวิธีที่ไม่ค่อยดีนักเพราะว่าต้องอาศัยกลุ่มตัวอย่างขนาดใหญ่มากและให้ค่าไม่ถูกต้องนัก

หอสมุดกลาง สถาบันวิทยบริการ จุฬาลงกรณ์มหาวิทยาลัย

ข้อตกลงเบื้องต้นในการใช้รูปแบบของราสช์

ข้อตกลงเบื้องต้นของราสช์พอจะสรุปได้ดังนี้ (Hambleton 1977: 77-78)

1. แบบสอบทั้งฉบับต้องวัดคุณลักษณะเดียวกัน (Unidimension Latent space) หรือความสามารถเดียวกัน (single ability) กล่าวคือ ข้อสอบจะต้องเป็นเอกพันธ์ (homogeneous) ในการที่จะตัดสินใจว่าข้อสอบมีลักษณะดังกล่าวหรือไม่นั้นอาจทำได้โดยอาศัยการวิเคราะห์ตัวประกอบ (Factor Analysis)

2. ข้อสอบจะต้องมีความเป็นอิสระของตำแหน่ง(Local independent) ใน 2 ลักษณะ คือ

2.1 มีความเป็นอิสระทางสถิติ (Statistically independent) กล่าวคือ ข้อสอบแต่ละข้อเป็นอิสระไม่เกี่ยวข้องกัน แต่ละข้อวัดความสามารถไม่ซ้ำกันเลย ดังนั้นคำตอบของแต่ละข้อของแต่ละคนเป็นอิสระต่อกันแต่เมื่อรวมกันแล้วจะวัดคุณลักษณะเดียวกันเท่านั้น

2.2 มีความเป็นอิสระของตำแหน่ง(Uncorrelated independent) กล่าวคือ ข้อสอบแต่ละข้อจะปรากฏ อยู่ในตำแหน่งใดของแบบสอบก็ได้ไม่มีผลต่อการสอบ

3. คะแนนเป็น dichotomous เท่านั้น คือตอบถูกเป็น 1 และผิดเป็น 0

4. ความเร็วในการทำข้อสอบไม่มีผลต่อโอกาสในการตอบถูก

5. โอกาสในการตอบถูกขึ้นอยู่กับความสัมพันธ์ของความสามารถของผู้สอบกับความยากของข้อสอบเท่านั้น

ความแกร่ง (robustness) ของรูปแบบของราสช์

การใช้รูปแบบของราสช์เพื่อการทดสอบนั้นได้มีการทดลองใช้ในรูปแบบต่าง ๆกัน ปรากฏว่ารูปแบบของราสช์มีข้อได้เปรียบกว่ารูปแบบอื่น ๆและเป็นที่นิยมใช้กันมากดังนี้

1. ถ้าข้อมูลเหมาะสมกับรูปแบบของราสช์อาจจะใช้กลุ่มตัวอย่างขนาดเล็กประมาณ 100 คน ก็ได้(Wright 1977: 219)

2. การกระจายของความสามารถของกลุ่มตัวอย่างหรือความยาวของแบบสอบไม่จำเป็นต้องเป็นโค้งปกติ เพียงแต่มีแนวโน้มเป็นโค้งปกติเท่านั้นและกลุ่มตัวอย่างไม่จำเป็นต้องมาจากการสุ่มก็ได้ (Wright and Stone 1979: 20)

3. ข้อสอบไม่จำเป็นต้องมีค่าอำนาจจำแนกเท่ากันและไม่จำเป็นต้องคำนึงถึงค่าการเดา(Wright and Panchapakesan 1967: 25)

ศูนย์วิทยทรัพยากร จุฬาลงกรณ์มหาวิทยาลัย

4. ข้อสอบแบบ multiple point สามารถใช้กับรูปแบบของราสช์ได้ และไม่จำเป็นต้องเป็นแบบถูกได้ 1 และผิดได้ 0 ก็สามารถใช้ได้(Willmott 1980: 195)

5. รูปแบบของราสช์มีความซับซ้อนน้อยที่สุด เมื่อเทียบกับรูปแบบอื่น ๆของทฤษฎีการตอบสนองข้อสอบจึงมีผู้นำมาใช้มากที่สุด

6. สามารถใช้กับการทดสอบทั่วไป การรวมคะแนนสามารถรวมได้เลยเพราะการกำหนดน้ำหนักไม่ซับซ้อนเหมือนกับรูปแบบอื่น ๆ (Wright 1977: 102)

วิธีการประมาณค่าพารามิเตอร์ในรูปแบบของราสช์

การประมาณค่าพารามิเตอร์ในรูปแบบของราสช์เพื่อการวิเคราะห์ข้อสอบสามารถประมาณค่าได้หลายวิธี คือ (Wright afterword in Rasch 1980: 188-193)

1. The Log Method ราสช์ได้พัฒนาขึ้นในปี ค.ศ.1953 เป็นวิธีที่ง่ายใช้ได้โดยประมาณค่าพารามิเตอร์ในรูปlog ของความสำเร็จ ซึ่งเป็นวิธีที่เหมาะสมกับแบบสอบที่มีค่าอำนาจจำแนกเท่ากับ 1 และใช้จำนวนคนในการประมาณค่าพารามิเตอร์เป็นจำนวนมาก

2. The PAIR Method เป็นวิธีการประมาณค่าพารามิเตอร์โดยใช้ข้อสอบจับคู่เป็นคู่ ๆ เช่น 1-2 ,1-3,1-4,2-4 เป็นต้น โดยเอาจำนวนคนที่พยายามตอบ 2 ข้อแต่ทำถูกเพียงข้อเดียวมาวิเคราะห์ เพื่อหา sample free ดังนั้นจึงเหมาะที่จะใช้ประมาณค่าแบบสอบที่มีจำนวนข้อสอบไม่มากนัก

3. The FCON Method เป็นวิธีประมาณค่าพารามิเตอร์โดยใช้ข้อสอบทุกข้อและผู้สอบทุกคนมาวิเคราะห์ เหมาะสำหรับแบบสอบที่มีจำนวนข้อสอบน้อยกว่า 30 ข้อ ถ้ามากกว่า 30 ข้อ จะทำให้การประมาณคาพารามิเตอร์มีความคลาดเคลื่อน แต่ภายหลังได้มีการปรับปรุงใหม่สามารถใช้กับแบบสอบที่มีความยาว 60-70 ข้อได้

4. The UCON Method เป็นวิธีประมาณคาพารามิเตอร์ด้วยวิธี Unconditional Maximum Likelihood Method ใช้กับแบบสอบที่การกระจายของความสามารถของบุคคลและความยากของแบบสอบเป็นการกระจายแบบโค้งปกติและใช้กับแบบสอบที่มีข้อสอบตั้งแต่ 25 ข้อขึ้นไป โดยมากจะใช้กับแบบสอบที่มี 1,000 ข้อขึ้นไป

5. The PROX Method เป็นวิธีประมาณค่าพารามิเตอร์ด้วยวิธีการของ Cohen's Approximation ใช้กับแบบสอบที่มีการกระจายของความสามารถของบุคคลและความยากของแบบสอบเป็นลักษณะเบ้ หรือมีแนวโน้มที่จะเป็นโค้งปกติ โดยมากการประมาณค่าด้วยวิธี PROX และ UCON จะให้ผลเหมือนกันต่างกันที่มีค่าความคลาด

เคลื่อนมาตรฐานเพียงเล็กน้อย

6. The UFORM Method ใช้กับแบบสอบที่มีการกระจายของค่าความยาก ของข้อสอบเป็นแบบ Uniform

<u>การวิเคราะห์ข้อสอบด้วยรูปแบบของราสช์</u>

ในการวิเคราะห์ข้อสอบด้วยรูปแบบของราสช์มีทั้งการวิเคราะห์ด้วยมือและการวิเคราะห์ด้วยเครื่องคอมพิวเตอร์ ซึ่งขั้นตอนของทั้งสองวิธีจะมีความแตกต่างกัน ในตอนแรกจะขอกล่าวถึงการวิเคราะห์ด้วยมือ ซึ่งมีขั้นตอนการวิเคราะห์ดังต่อไปนี้

<u>การวิเคราะห์ด้วยมือ</u> (Item Calibration by Hand)

วิธีนี้ใช้วิธีประมาณค่าพารามิเตอร์ด้วยวิธี PROX ซึ่ง Wright (Wright 1977 : 100- 101) เสนอขั้นตอนไว้โดยสรุปดังนี้

1. หลังจากตรวจแบบสอบแล้ว ให้ตัดข้อที่คนตอบถูกและผิดหมดออกและตัดคนที่ได้คะแนนเต็มและคะแนนศูนย์ออก

2. กำหนดค่า S_i คือ จำนวนคนที่ทำข้อสอบข้อที่ i ถูก เมื่อข้อสอบมีจำนวนข้อจาก i ถึง L และกำหนดให้ n_r คือ ความถี่หรือจำนวนคนที่ได้คะแนน r เมื่อค่า r=1 ถึง L-1 (เพราะคนที่ได้คะแนนเต็มและคะแนนศูนย์ถูกตัดออก)

3. คำนวนค่า

$x_i = \ln\left[(N-S_i)/S_i\right]$ คือ ค่าอัตราส่วนลอกของจำนวนคนทำผิดต่อจำนวนคนทำถูกในแต่ละข้อ

$\bar{x} = \sum_{i}^{L} x_i/L$ คือ ค่าเฉลี่ยของ x_i จากข้อสอบจำนวน L ข้อ

$U = \sum_{i}^{L} (x_i - \bar{x})^2/L$ คือ ค่าความแปรปรวนของ x_i จากทุกๆข้อ

$y_r = \ln\left[r/(L-r)\right]$ คือ ค่าอัตราส่วนลอกของจำนวนคนทำถูกต่อคนทำผิดที่ L คะแนน

$\bar{y} = \sum_{r}^{L-1} n_r y_r/N$ คือ ค่าเฉลี่ยของ y_r จากจำนวน N คน

$V = \sum_{r}^{L-1} n_r (y_r - y) /L-1$ คือ ค่าความแปรปรวนของ y_r จำนวน N คน

โดยที่ $X = \left[\dfrac{1 + U/2.89}{1 - UV/8.35}\right]^{\frac{1}{2}}$ คือ ค่าปรับขยายของความยากของข้อสอบเมื่อคำนึงถึงความยาวของข้อสอบแล้ว

$Y = \left[\dfrac{1 + V/2.89}{1 - UV/8.35}\right]^{\frac{1}{2}}$ คือ ค่าปรับขยายของความสามารถของบุคคลเมื่อคำนึงถึงกลุ่มคนทั้งหมดแล้ว

$d = Y(x_i - \bar{x})$ คือ ค่าความยากของข้อสอบแต่ละข้อเมื่อปรับแล้ว

$SE(d_i) = Y\left[N/S_i(N-S_i)\right]^{\frac{1}{2}}$ คือ ค่าความคลาดเคลื่อนมาตรฐานของความยากของข้อสอบ

$b = Xy_r$ คือ ค่าความสามารถของบุคคลแต่ละระดับคะแนนที่ปรับแล้ว

$SE(b_r) = X\left[L/r(L-r)\right]^{\frac{1}{2}}$ คือ ค่าความคลาดเคลื่อนมาตรฐานของความสามารถของบุคคล

<u>การทดสอบความเหมาะสม(fit)ของบุคคลและข้อสอบกับรูปแบบของราสช์</u>

หลังจากหาค่าพารามิเตอร์ d และ b แล้วจำเป็นต้องมีการทดสอบว่าข้อสอบหรือบุคคลนั้น ๆ มีความเหมาะสมกับรูปแบบเพียงใด โดยพิจารณาจากเหตุการณ์ที่แต่ละบุคคลตอบข้อสอบข้อหนึ่ง ๆเป็นไปตามความคาดหมายเพียงใดหรือไม่ กล่าวคือ ถ้าบุคคลหนึ่ง ๆ ตอบข้อสอบข้อนั้นผิดทั้ง ๆที่เขามีความสามารถมากกว่าความยากของข้อสอบหรือตอบข้อสอบนั้น ๆที่ความสามารถของเขามีน้อยกว่าความยากของข้อสอบ ซึ่งทั้งสองเหตุการณ์นี้เป็นลักษณะของการตอบที่เป็นไปไม่ได้ตามความคาดหวัง (Unexpected Responses) แสดงว่าต้องมีสิ่งใดสิ่งหนึ่งไม่เหมาะสมกับรูปแบบ คือ อาจจะเป็นข้อสอบหรือบุคคลที่ทำข้อสอบซึ่งต้องมีการทดสอบต่อไปและ Wright ได้เสนอวิธีการทดสอบความเหมาะสมซึ่งสามารถสรุปได้ดังนี้ (Wright and Stone 1979: 66-80)

1. เมื่อบุคคล v ทำข้อสอบข้อที่ i จะได้

$X_{vg} = 0$ เมื่อตอบผิด และ $X_{vg} = 1$ เมื่อตอบถูก

2. ทดสอบความเหมาะสมของบุคคลที่ v (person fit)

2.1 $V_v = \sum_{i}^{L} Z_{vi}/(L-1) \sim F_{L-1,\infty}$ L คือ จำนวนข้อ

2.2 $Z_{vi}^2 = \exp(b-d)$ เมื่อ $X_{vi} = 0$ หรือตอบผิด

$Z_{vi}^2 = \exp(d-b)$ เมื่อ $X_{vi} = 1$ หรือตอบถูก

เมื่อ Z_{vi} คือ ค่าความคลาดเคลื่อนมาตรฐานกำลังสองของเหตุการณ์ที่บุคคลที่ v ทำข้อสอบข้อที่ i

2.3 $t_v = \left[\ln(V_v) + (V_v - 1)\right]\left[(L-1)/8\right]^{\frac{1}{2}} \sim N(0,1)$

2.4 เปรียบเทียบค่า t_v กับค่า t จากตารางทางเดียว t = t-test

3. ทดสอบความเหมาะสมของข้อสอบข้อที่ i (item fit)

3.1 $V_i = \sum_{v}^{N} Z_{vi}^2/(N-1) \sim F_{N-1}$ N คือ จำนวนคน

3.2 $t_i = \left[\ln(V_i) + (V_i-1)\right]\left[(N-1)/8\right]^{\frac{1}{2}} \sim N(0,1)$

3.3 เปรียบเทียบค่า t_i กับค่า t จากตารางทางเดียว (t-test)

สำหรับการวิเคราะห์ด้วยเครื่องคอมพิวเตอร์นั้น เนื่องจากมีโปรแกรมสำเร็จรูปเพื่อวิเคราะห์ข้อสอบ คือโปรแกรม MICROSCALE ซึ่งพัฒนาโดย Wright กับบริษัท MEDIAX เป็นเทปเล็ก ๆซึ่งเมื่อใส่ข้อมูลของคนและข้อสอบเป็นรายข้อ รายคน จะช่วยให้สามารถวิเคราะห์ข้อสอบได้ (อุทุมพร จามรมาน 2529: 100) ต่อไปนี้จะขอเสนอการวิเคราะห์ด้วยเครื่องคอมพิวเตอร์ที่ใช้โปรแกรมไบคาล (BICAL) ซึ่งเป็นโปรแกรมสำเร็จรูปเช่นกันแต่มีความสามารถในการคำนวนที่ดีกว่า

<u>การวิเคราะห์ด้วยเครื่องคอมพิวเตอร์</u>

การวิเคราะห์ด้วยเครื่องคอมพิวเตอร์ใช้โปรแกรมไบคาล (BICAL) ซึ่งเขียนโดย Wright และ Mead (Wright and Mead 1979) ในปี ค.ศ.1978 ซึ่งมีอยู่ 2 สับโปรแกรม คือ PROX และ UCON โดยที่ข้อตกลงเบื้องต้นในการเลือกใช้โปรแกรมย่อยทั้ง 2 นี้ได้กล่าวมาแล้วในตอนต้น ขั้นตอนการวิเคราะห์โดยสังเขปมีดังนี้

<u>ขั้นที่ 1</u> ตรวจข้อสอบ

โปรแกรมจะทำการตรวจสอบเพื่อหาว่าข้อสอบแต่ละข้อมีคนทำได้ถูกต้องเท่าใด และจะนับจำนวนความถี่ของคะแนนรวมในแต่ละจำนวน

<u>ขั้นที่ 2</u> กำจัดข้อสอบบางข้อออก

ข้อสอบที่ไม่มีคนทำได้เลย หรือข้อสอบที่คนทำได้ทุกคนและคนที่ทำคะแนนไม่ได้เลยหรือทำได้เต็มจะถูกกำจัดออกจากการวิเคราะห์เพราะถือว่าข้อสอบหรือคะแนนดังกล่าวไม่มีลักษณะที่จำเป็นมาพิจารณาวิเคราะห์เพราะว่าเป็นข้อสอบที่ยากที่สุดหรือง่ายที่สุดและแสดงว่ามีบางคนที่มีความสามารถต่ำที่สุดและสูงที่สุด

ในทางปฏิบัติเกณฑ์ในการกำจัดข้อสอบและคนบางคนออกจากการวิเคราะห์อาจใช้อัตราส่วนของการเดาได้ถูกต้องเป็นเกณฑ์ เช่น ถ้าข้อสอบชนิดเลือกตอบมีจำนวน 100 ข้อ คนที่ได้คะแนนต่ำกว่าหรือเท่ากับ 20 คะแนนและสูงกว่า 80 คะแนน จะถูกกำจัดออกจากการวิเคราะห์ถ้าเป็นความต้องการของผู้วิเคราะห์ทั้งนี้โปรแกรมจะมีคำสั่งให้เลือกด้วย

ในขั้นนี้โปรแกรมจะแบ่งกลุ่มตัวอย่างออกเป็นกลุ่ม ๆตามจำนวนที่ผู้วิเคราะห์ข้อสอบต้องการ โดยอาศัยคะแนนรวมในการสอบเป็นเกณฑ์แต่ไม่เกิน 6 กลุ่ม

<u>ขั้นที่ 3</u> กำหนดค่าเริ่มแรกของ Θ_v

<u>ขั้นที่ 4</u> กำหนดค่าเริ่มแรกของ b_g

<u>ขั้นที่ 5</u> กำหนดค่าของชุดของ b_g

<u>ขั้นที่ 6</u> คำนวนค่าของ b_g ของแต่ละข้อโดยวิธีการทำซ้ำ (iteration) ค่า b_g ในสมการที่หาค่ามาจากสมการ

$$P\left\{X_{vg}/\Theta_v, b_g\right\} = \exp\left[X_{vg}(\Theta_v - b_g)/\left\{1 + \exp(\Theta_v - b_g)\right\}\right]$$

<u>ขั้นที่ 7</u> คำนวนค่าของ Θ_v ของคะแนนรวมที่แตกต่างกันโดยการทำซ้ำค่าในสมการอีกสมการหนึ่งที่ derived มาจากสมการในขั้นที่ 6 โดยอาศัยค่าของ b_g ด้วย

<u>ขั้นที่ 8</u> เปรียบเทียบค่า b_g ในขั้นที่ 6 และค่าที่เปลี่ยนไปในขั้นที่ 7 ว่าแตกต่างกันที่ $\alpha = 0.0001$ หรือไม่ ถ้าแตกต่างกันก็กระทำในขั้นที่ 5 6 7 และ 8 อีกหลาย ๆครั้ง จนกระทั่งค่า b_g ของข้อต่าง ๆคงที่

<u>ขั้นที่ 9</u> ปรับปรุงค่า b_g ของแต่ละข้อให้ถูกต้อง

<u>ขั้นที่ 10</u> คำนวนค่า Θ_v ของผู้สอบที่ได้คะแนนรวมแตกต่างกันโดยอาศัยค่าในขั้นที่ 9 ที่ปรับปรุงแล้ว

<u>ขั้นที่ 11</u> ปรับปรุงค่า Θ_v ของผู้ที่ได้คะแนนรวมแตกต่างกันให้ถูกต้อง

<u>ขั้นที่ 12</u> ทำการทดสอบโค้งลักษณะของข้อสอบ (Item Characteristic Curve: ICC) ของข้อสอบแต่ละข้อว่าเหมาะสม (fit) กับ ICC ของรูปแบบหรือไม่ โดยการทดสอบ 2 ลักษณะคือ

1. ทดสอบความเหมาะสมของ ICC โดยยึดเอาจุดต่าง ๆที่แสดงค่าร้อยละของผู้ตอบข้อสอบแต่ละข้อได้ถูกต้อง เมื่อจุดเหล่านั้นแทนระดับความสามารถของกลุ่มตัวอย่างที่แตกต่างกัน(ซึ่งเกิดจากการแบ่งโดยอาศัยคะแนนรวมเป็นเกณฑ์ในขั้นที่ 2) ว่าเหมาะสมกับ ICC ของรูปแบบหรือไม่ หากว่าค่า meansquare =1 แสดงว่าโค้งทั้งสองเหมาะสมกันพอดี

2. ทดสอบความเหมาะสมของ ICC โดยยึดเอาระดับความสามารถต่าง ๆของผู้สอบที่แตกต่างกันแต่ละคนเป็นเกณฑ์ว่าสามารถทำข้อสอบนั้นได้ร้อยละที่แตกต่างกันกระจาย

I 10292032

แล้วเป็นโค้งที่เหมาะสมกับ ICC ของรูปแบบหรือไม่ หากว่าค่า meansquare =1 แสดงว่าโค้งทั้งสองเหมาะสมกับ ICC พอดี

<u>ขั้นที่ 13</u> โปรแกรมคำนวนหาค่าอำนาจจำแนก

สืบเนื่องจากรูปแบบของราสช์ไม่ได้คำนึงถึงเรื่องค่าอำนาจจำแนก เพราะว่าเป็นข้อตกลงเบื้องต้นแล้วว่าข้อสอบแต่ละข้อมีค่าอำนาจจำแนกเท่า ๆกันและไม่ได้คำนวนค่าพารามิเตอร์อำนาจจำแนกดังกล่าวด้วย แต่โปรแกรมไบคาลให้ค่าอำนาจจำแนกด้วยซึ่งเป็นการคำนวนภายหลังที่คำนวนด้วยวิธีของรูปแบบของราสช์แล้ว ค่าอำนาจจำแนกดังกล่าวนี้บอกถึงระดับความเหมาะสมระหว่างICC ของข้อสอบแต่ละข้อกับ ICCของรูปแบบกล่าวคือ ถ้าค่าอำนาจจำแนกยิ่งใกล้ 1.00 มากเท่าใดก็แสดงว่าโค้งทั้งสองมีความใกล้ชิดกันมาก ถ้าค่าอำนาจจำแนกน้อยกว่า 1.00 แสดงว่าICC ของข้อสอบแบนกว่า(flat)กว่าของรูปแบบแสดงว่าข้อสอบนั้นไม่มีอำนาจจำแนกดีเท่ากับข้ออื่น ๆที่มีอำนาจจำแนกสูงกว่า

แต่ถ้าค่าอำนาจจำแนกมากกว่า 1.00 แสดงว่าข้อสอบนั้นโค้งกว่าICC ของรูปแบบและมีค่าอำนาจจำแนกมากเกินไป จึงควรนำมาพิจารณาว่าเหตุใดจึงเป็นเช่นนั้น

ค่าสหสัมพันธ์แบบไบซีเรียล (biserial correlation) นั้นมีความหมายเหมือนกับการวิเคราะห์ข้อสอบโดยทฤษฎีมาตรฐานดั้งเดิม

<u>ขั้นที่ 14</u> การเรียงลำดับข้อสอบ

การเรียงลำดับข้อสอบที่วิเคราะห์จากโปรแกรมมี 3 ลักษณะ คือ

1. เรียงตามลำดับที่ของข้อสอบ
2. เรียงตามลำดับที่ความยากจากข้อที่ง่ายที่สุดไปยังข้อที่ยากที่สุด
3. เรียงตามลำดับระดับความเหมาะสมกันระหว่าง(fit)ของข้อสอบกับของรูปแบบ จากข้อที่เหมาะสมมากที่สุดไปถึงข้อที่เหมาะสมน้อยที่สุด

อนึ่งโปรแกรมวิเคราะห์ข้อสอบด้วยรูปแบบของราสช์โปรแกรมไบคาลนี้สามารถติดต่อขอใช้บริการได้ที่สถาบันบริการคอมพิวเตอร์ จุฬาลงกรณ์มหาวิทยาลัย โดยผู้วิเคราะห์ต้องทำการศึกษาโปรแกรมจากคู่มือการวิเคราะห์ BICAL 3ที่ Wright และ Mead (Wright and Mead 1979) ได้เรียบเรียงไว้ก่อนเพื่อความสะดวกในการวิเคราะห์ข้อมูล

<u>การนำรูปแบบของราสช์มาประยุกต์ใช้ในการทดสอบและวัดผลทางการศึกษา</u>

1. <u>ใช้ประโยชน์ในการวิเคราะห์ข้อสอบ</u> (Item Analysis) เนื่องจากการวิเคราะห์ข้อสอบตามแนวเดิมเป็นวิธี 27 % นั้นมีข้อบกพร่องจำนวนมากเช่น ประการแรกค่าสถิติต่าง ๆขึ้นอยู่กับสภาพของตัวอย่างที่ทำการสอบ ประการที่สอง ค่าอำนาจจำแนกเป็นค่าที่ไม่ถูกต้อง และประการที่สาม แก้ปัญหาเรื่องการเดาไม่ได้ แต่ปัญหาต่าง ๆเหล่านี้อาจจะแก้ไขได้โดยวิธีวิเคราะห์ด้วยรูปแบบของราสช์ เพราะค่าความยากเป็นค่าคงที่และแก้ปัญหาเรื่องการเดาหรือค่าอำนาจจำแนกได้

2. <u>ใช้ในการสร้างธนาคารข้อสอบ</u> (Item Bank) เนื่องจากข้อสอบที่วิเคราะห์แล้วค่าพารามิเตอร์มีลักษณะไม่แปรเปลี่ยน(invariant) ดังนั้นข้อสอบเหล่านี้จึงนำมาใช้สร้างข้อสอบชุดใหม่ ๆตามเกณฑ์ที่ต้องการวัด

3. <u>ใช้สร้างแบบสอบที่มีน้ำหนักคะแนนต่างกัน</u> (multiple point test) รูปแบบของราสช์ทำการวิเคราะห์ข้อสอบที่เป็น multiple point ได้ตามระดับของความถูกต้องของข้อสอบและสามารถจะสร้างข้อสอบลักษณะดังกล่าวจากธนาคารข้อสอบที่วิเคราะห์ไว้แล้ว

4. <u>ใช้ในการกำหนดเกณฑ์ของระดับการรอบรู้</u> (master level) ของแบบสอบอิงเกณฑ์ ผลการวิเคราะห์ข้อสอบคะแนนของผู้สอบจะถูกแปลงให้เป็นคะแนนซึ่งสามารถเปรียบเทียบกับคะแนนความสามารถที่เป็นเกณฑ์คงที่ที่ทำให้เราทราบว่าระดับของ minimum master level ของข้อสอบแต่ละข้อควรเป็นเท่าใด

5. <u>ใช้วินิจฉัยความสามารถของผู้สอบ</u> ในกรณีที่มี Item Characteristic Curve: ICC ของข้อสอบไม่เหมาะสมกับโค้งของรูปแบบแสดงว่ามีบางสิ่งบางอย่างผิดปกติในตัวผู้สอบที่เราสนใจควรแก้ไข

6. <u>ใช้ในการค้นคว้าหาความเป็นอคติของข้อสอบ</u> (item Bias) เมื่อมีความสามารถอื่นแฝงเข้ามาในความสามารถที่ต้องการวัด เราสามารถทำการตรวจอคติของข้อสอบได้จากการศึกษาความเหมาะสมของ ICC ของกลุ่มบุคคลที่มีความสามารถเท่า ๆกัน

7. <u>ใช้ในการวัดระดับความสามารถของแต่ละบุคคล</u> (Self-Tailoring Test) เราอาจสุ่มข้อสอบที่วิเคราะห์แล้วและมีระดับความยากเรียงกันตามลำดับเพียงจำนวนข้อไม่มากมาทดสอบความสามารถของแต่ละบุคคลได้

8.. <u>ใช้ในการสร้างแบบสอบที่ดีที่สุด</u> (Best Test Design) ผลจากการ

วิเคราะห์ด้วยรูปแบบของราสช์สามารถนำมาใช้ในการออกแบบแบบสอบที่มีลักษณะต่าง ๆ ตามที่ต้องการได้ เช่น ให้ความยากของข้อสอบที่มีลักษณะตามโค้งลักษณะแบบสอบที่ต้องการ และถ้าแบบสอบมีข้อสอบที่มีความยากอยู่ระหว่าง $M = \pm 2$ SD.(M = mean ability) และมีความยาวของข้อสอบเป็น $L = 6/SEM^2$ (SEM = Standard error of measurement) แล้วจะได้แบบสอบที่ดีที่สุด

<u>งานวิจัยที่เกี่ยวข้องกับรูปแบบของราสช์</u>

Moses Adedoyin Soriyan (1971) ได้ศึกษาการวัดสารรูปสนิทสุด (Goodness of fit) ของรูปแบบของราสช์ในการวิเคราะห์ผลสัมฤทธิ์ตามจุดประสงค์ของแบบสอบในโรงเรียนเขตแอฟริกาตะวันตก โดยมีจุดมุ่งหมายดังนี้

1. พิจารณาความเหมาะสมของรูปแบบของราสช์กับข้อสอบแบบสอบผลสัมฤทธิ์ในระดับมัธยมศึกษาตอนปลาย

2. ค้นหาขอบเขตของข้อสอบที่เหมาะสมกับรูปแบบของราสช์ เมื่อมีค่าอำนาจจำแนกเท่า ๆกัน

3. ศึกษาความไม่แปรเปลี่ยนของค่าพารามิเตอร์ที่ได้อ้างไว้

4. ค้นหาขอบเขตของการวัดคุณลักษณะมิติเดียวของแบบสอบผลสัมฤทธิ์แต่ละฉบับ

ผลการศึกษาพบว่า ข้อสอบในวิชาคณิตศาสตร์ ภูมิศาสตร์ชีวภาพ และภาษาอังกฤษมีความเหมาะสมกับรูปแบบของราสช์ดีพอใช้ที่ระดับ .01 ค่าอำนาจจำแนกของข้อสอบได้ค่าที่ไม่เท่ากันจริง ๆ จะอยู่ในช่วง 0.5 - 1.25 ผู้วิจัยได้สรุปว่า

1. รูปแบบของราสช์เหมาะสมกับแบบสอบผลสัมฤทธิ์ทางวิชาการที่ใช้วัดในระดับมัธยมศึกษาตอนปลาย ทั้งนี้ควรจะเลือกข้อสอบที่มีความคงที่ภายในที่เหมาะสมด้วย

2. รูปแบบของราสช์แม้มีข้อตกลงเบื้องต้นว่าค่าอำนาจจำแนกรายข้อเท่ากัน เมื่อทำการทดสอบการคัดเลือกข้อสอบค่าอำนาจจำแนกอาจไม่เท่ากันก็ได้

3. การพิจารณาความเป็นอิสระของบุคคลในการคัดเลือกข้อสอบ(เมื่อปราศจากพารามิเตอร์ของข้อสอบ) สามารถใช้ได้กับรูปแบบของราสช์

4. ในการวิเคราะห์ค่า พารามิเตอร์ของข้อสอบ (calibration) ความเป็นอิสระของบุคคลไม่มีผลด้วย

5. ความเป็นอิสระของข้อสอบต่อบุคคลในการวัดครั้งนี้ไม่เกิดผล

Whitely และ Dawis (1974) วิเคราะห์แบบสอบอุปมาอุปไมยทางภาษาจำนวน 60 ข้อ โดยใช้รูปแบบของราสช์มีข้อสอบที่ไม่เหมาะสมกับรูปแบบประมาณ 30-40 %

Rentz และ Bashaw (1975) ได้ศึกษาและพบว่าขนาดของกลุ่มตัวอย่างระหว่าง 500-1,000 คน จะทำให้การประมาณค่าพารามิเตอร์ของรูปแบบของราสช์มีความคงที่เพิ่มขึ้นอย่างรวดเร็ว และจะค่อย ๆเพิ่มขึ้นเมื่อกลุ่มตัวอย่างที่ใช้ประมาณค่า พารามิเตอร์มีจำนวน 2,000 คน และ 4,000 คน

Rentz และ Rinderour (1978) ได้ใช้รูปแบบของราสช์วิเคราะห์แบบสอบผลสัมฤทธิ์ แบบสอบ Atlanta Assessment Project และแบบสอบผลสัมฤทธิ์แสตนด์ฟอร์ด พบว่าข้อสอบมีความเหมาะสมกับรูปแบบอยู่ระหว่าง 85-97%,92-97 % และ 80 % ตามลำดับ

Beres (Beres 1980: 4685 - A.) ใช้รูปแบบของราสช์วิเคราะห์แบบสอบประเมินตนเอง (SAS : Self Assessment Scalesได้ข้อสอบที่เหมาะสมกับรูปแบบ 41 % มีค่าความเป็นอิสระของข้อสอบเท่ากับ 87 %และความอิสระของความสามารถของผู้สอบ 84 %

James Bruce Douglass (1980) ได้ศึกษาเปรียบเทียบการใช้รูปแบบ Item Characteristic Curve กับการสอบในระบบห้องเรียน โดยใช้รูปแบบทั้ง 3 รูปแบบของทฤษฎีการตอบสนองข้อสอบประมาณค่าพารามิเตอร์ของข้อสอบ 100 ข้อ ที่เป็นข้อสอบปลายปีของ 4 ปีการศึกษา จากกลุ่มตัวอย่างระดับวิทยาลัยจำนวน 594 - 1,082 คน ข้อสอบแบ่งเป็น 4 ฉบับแล้วสุ่มข้อสอบฉบับละ 43-53 ข้อแล้วทำการสุ่มผู้สอบมา 200,600 และ 800 คนมาวิเคราะห์ปรากฏว่า การวิเคราะห์ข้อสอบด้วยรูปแบบของราสช์มีความคงที่ของค่าพารามิเตอร์ดีกว่ารูปแบบ 2 พารามิเตอร์และเมื่อมีการเทียบมาตรา (equating) ของแบบสอบข้ามกลุ่มตัวอย่าง รูปแบบของราสช์ก็มีความคงที่ดีกว่าด้วย

Thavachai VOrpongthorn (1980) ได้ศึกษาการประยุกต์รูปแบบของราสช์ในการศึกษาความลำเอียงของข้อสอบของแบบสอบที่ใช้แบบสอบเข้ามหาวิทยาลัยในประเทศไทย โดยมีจุดมุ่งหมายของการศึกษาดังนี้

1. การวัดของแบบสอบแต่ละฉบับวัดในสิ่งที่เหมือนกันสำหรับผู้ชายและผู้หญิงหรือไม่

2. ข้อสอบภายในแบบสอบแต่ละฉบับมีศักยภาพของความเป็นอคติเพียงใด เมื่อเทียบกันระหว่างผู้ชายกับผู้หญิง

3. ถ้ามีศักยภาพในด้านอคติระหว่างเพศต่อข้อสอบ ลักษณะเฉพาะของเนื้อหาของข้อสอบเป็นอย่างไรเมื่อใช้รูปแบบของราสช์พิจารณา

ผู้วิจัยได้ใช้ผลการสอบของนักเรียนชายและนักเรียนหญิงที่สุ่มมาจากประชากรผู้สอบเข้ามหาวิทยาลัยสาขาวิทยาศาสตร์ ปีค.ศ.1979 แล้วนำมาวิเคราะห์โดยใช้โปรแกรมไบคาล

ผลการศึกษาพบว่า

1. แบบสอบแต่ละฉบับวัดในสิ่งเดียวกันสำหรับเพศชายและหญิง โดยมีหลักฐานยืนยันจากภาพ scattergram และมีค่าสัมประสิทธิ์สหสัมพันธ์สูงในข้อสอบที่ยาก

2. มีข้อสอบบางข้อในแต่ละฉบับที่มีศักยภาพของความเป็นอคติเมื่อเทียบกันระหว่างเพศชายและหญิงโดยที่แบบสอบภาษาอังกฤษมีจำนวนข้อสอบที่มีศักยภาพของความเป็นอคติมากที่สุด รองลงมาได้แก่ แบบสอบชีววิทยาและเคมี

3. การวิเคราะห์เนื้อหาของศักยภาพความเป็นอคติของข้อสอบไม่มีผลสรุปที่สมบูรณ์ แต่มีผลสอดคล้องกันระหว่างลักษณะเฉพาะของข้อสอบที่บ่งชี้ถึงศักยภาพของความเป็นอคติโดยเกณฑ์ความยากของข้อสอบกับความสัมพันธ์ของอัตราการตัดสินความเป็นอคติของเพศ

Hsing Sheng Chen (1981) ได้ทำการประเมินคุณลักษณะมิติเดียวของการสอบจากโปรแกรมการประเมินผลทางการศึกษาของมิชิแกน (Michigan Educational Assessment Program : MEAP) โดยใช้กลุ่มตัวอย่างนักเรียนเกรด 4 จำนวน 1,224 คนจากโรงเรียนประถมศึกษา 26 โรงเรียนที่สอบ MEAP ในฤดูใบไม้ร่วงปี ค.ศ.1979 จุดมุ่งหมายของการศึกษาได้ทำการเลือกศึกษาข้อมูลในการอ่านของนักเรียนเกรด 4 โดยวัดจุดประสงค์ 19 จุดประสงค์และทำการวัดคุณลักษณะมิติเดียวจากหัวข้อต่อไปนี้

1. ค่าสถิติของความเหมาะสมของรูปแบบของราสช์
2. การวิเคราะห์ด้วยสเกลของ Guttman
3. การวิเคราะห์ตัวประกอบด้วยการหมุนแกนแบบแวริแมกซ์
4. Biserial Test
5. การวิเคราะห์ตัวประกอบของความคลาดเคลื่อน

ค่าความสัมพันธ์ระหว่างดัชนีของคุณลักษณะมิติเดียวกับค่าสถิติของความเหมาะสมของข้อมูลในแต่ละวิธี คำนวนจากค่าสหสัมพันธ์ของเพียร์สัน(Pearson Product Moment Correlation) ผลการศึกษาพบว่า

1. ข้อสอบของแบบสอบอิงเกณฑ์วัดจุดประสงค์เดี่ยว ๆเป็นคุณลักษณะของมิติเดียว

2. ผลของความสัมพันธ์ในคุณลักษณะมิติเดียวของแต่ละจุดประสงค์จากการวิเคราะห์ตัวประกอบด้วยการหมุนแกนแบบแวริแมกซ์ ,Biserial Test และการวิเคราะห์ตัวประกอบของความคลาดเคลื่อน converge กัน และการวิเคราะห์ด้วยสเกลของ Guttman ไม่ converge กับการวิเคราะห์รูปแบบอื่น ๆ

3. ถ้าข้อสอบไม่เป็นคุณลักษณะมิติเดียวอย่างแท้จริงจะใช้ได้ไม่เหมาะสมกับรูปแบบของราสช์

4. การวิเคราะห์ตัวประกอบด้วยการหมุนแกนแบบแวริแมกซ์ เมื่อกลุ่มข้อสอบต่างกันแต่ใช้จุดประสงค์เดียวกันสามารถลดจำนวนของความคลาดเคลื่อนลงได้

5. การวิเคราะห์ตัวประกอบด้วยการหมุนแกนแบบแวริแมกซ์โดยใช้แบบสอบทั้งฉบับและเฉพาะบางส่วน และการวิเคราะห์ตัวประกอบของความคลาดเคลื่อน converge กันกับการแบ่งตัวแปรและโครงสร้างของคุณลักษณะมิติเดียว

Leah R. Hutten (1981) ได้ศึกษาความเหมาะสมของข้อมูลจริงกับรูปแบบของทฤษฎีการตอบสนองข้อสอบ 2 รูปแบบ คือ รูปแบบของราสช์ กับรูปแบบ 3 พารามิเตอร์ในการประมาณค่าพารามิเตอร์ความสามารถและความยากของข้อสอบ ใช้กลุ่มตัวอย่างนักเรียน 1,000 คน จุดมุ่งหมายเพื่อพิจารณาความแม่นยำ(precision) สำหรับกลุ่มตัวอย่างขนาดเล็ก(250 คน) และแบบสอบสั้น(20 ข้อ) ผลการศึกษาพบว่า รูปแบบของราสช์และรูปแบบ 3 พารามิเตอร์มีความเหมาะสมกับข้อมูล 80 % ทั้งสองรูปแบบ เมื่อพิจารณาคะแนนที่ทำนายกับคะแนนที่วัดได้จากค่าสถิติ Kolmogorov - Smirnov แล้ว รูปแบบของราสช์เหมาะสมกับข้อมูลทั้งหมดดีกว่ารูปแบบ 3 พารามิเตอร์(รูปแบบของราสช์ ค่า K-S=1.304 รูปแบบ 3 พารามิเตอร์ค่า K-S =1.289) ผลยังยืนยันว่าการใช้ข้อมูลเชิงประจักษ์กับรูปแบบของราสช์และรูปแบบ 3 พารามิเตอร์ควรจะคำนึงถึงคุณลักษณะมิติเดียวเป็นอันดับแรก(ค่าความสัมพันธ์ระหว่างค่าสถิติในความเหมาะสมกับดัชนีของคุณลักษณะมิติเดียวมีนัยสำคัญที่ระดับ .05)

ในกรณีการประมาณค่าความสามารถจากแบบสอบสั้น 20 ข้อ และใช้กลุ่มตัวอย่างขนาดเล็ก 250 คน รูปแบบของราสช์เหมาะสมกว่ารูปแบบ 3 พารามิเตอร์ ผลการศึกษาได้สนับสนุนให้ใช้รูปแบบ 3 พารามิเตอร์กับกลุ่มตัวอย่างที่มีขนาดตั้งแต่ 1,000 คนขึ้นไป เพื่อจะได้ความแม่นยำในการประมาณค่าพารามิเตอร์ต่าง ๆ ทั้งค่าใช้จ่ายในการวิเคราะห์ด้วยรูปแบบของราสช์ถูกกว่ารูปแบบ 3 พารามิเตอร์เมื่อใช้กลุ่มตัวอย่าง 1,000 คน(รูปแบบของราสช์เสียค่าใช้จ่าย 12.50 เหรียญสหรัฐ ส่วนรูปแบบ 3 พารามิเตอร์เสียค่าใช้จ่าย 35.12 เหรียญสหรัฐ)

Lash (Lash 1983: 1429-A.) ได้ศึกษาเกี่ยวกับการใช้รูปแบบของราสช์ในแบบสอบที่วัดหลาย ๆคุณลักษณะ (Multidimensional Item) โดยแปลงให้เป็นการวัดคุณลักษณะเดียวหลาย ๆกลุ่ม วิธีการที่ใช้ประมาณค่านั้นใช้คะแนนจริงจากการสอบ จากการศึกษาพบว่า ค่าพารามิเตอร์ที่ได้ของการวัดแต่ละกลุ่มมีการกระจายที่ไม่เปลี่ยนแปลง

ไปตามกลุ่มผู้สอบและแบบสอบ และจากการเปรียบเทียบความแตกต่างของพารามิเตอร์ข้อสอบระหว่างการวัดแต่ละด้านพบว่า ระดับความสามารถของกลุ่มมีความเป็นอิสระแต่ไม่พบในการกระจายของข้อสอบ

Donald John McPheson (1983) ได้ศึกษาการใช้รูปแบบของราสช์ในการประเมินข้อสอบสำหรับนักเรียนเกรด 4 และเกรด 7 ที่สอบในโปแกรมการประเมินทางการศึกษาของมิชิแกน (Michigan Educational Assessment Program: MEAP) ซึ่งเป็นแบบสอบอิงเกณฑ์ในมิติของการอ่านที่ใช้สอบในปี ค.ศ.1973-1979 ผู้วิจัยต้องการทดสอบว่า รูปแบบของราสช์จะเหมาะสมหรือไม่กับแบบสอบที่ใช้เวลามากในการที่จะให้เกิดผลสัมฤทธิ์ ผลปรากฏว่า แบบสอบทั้ง 14 ฉบับเหมาะสมกับรูปแบบของราสช์เพียง 1 ฉบับ ผู้วิจัยจึงได้เสนอแนะว่า การนำรูปแบบของราสช์มาใช้นั้นการวัดควรจะวัดในจุดประสงค์ที่แท้จริงทั้งยังต้องพิจารณาถึงข้อตกลงเบื้องต้นของรูปแบบของราสช์ในการวิเคราะห์ข้อสอบด้วย

Jeffrey Oescher (1983) ได้ศึกษาเปรียบเทียบการวิเคราะห์ข้อสอบด้วยวิธีการคัดเลือก 3 วิธี โดยการสร้างแบบสอบขึ้นมา 3 ฉบับ ๆละ 30 ข้อ ซึ่งคัดเลือกข้อสอบมาจากประชากรข้อสอบ 110 ข้อ แบบสอบฉบับแรกใช้เกณฑ์การคัดเลือกแบบคลาสสิคอล ประกอบด้วยข้อสอบที่มีค่าดัชนีอำนาจจำแนกสูงสุด แบบสอบฉบับที่สองคัดเลือกจากการพิจารณาค่าความยากโดยใช้รูปแบบของราสช์ และแบบสอบฉบับที่สามใช้การสุ่มตัวอย่างแบบแบ่งชั้น ผลปรากฏว่า ไม่มีความแตกต่างกันอย่างมีนัยสำคัญทางค่าสถิติของการคัดเลือกตามเกณฑ์จากการวิเคราะห์ด้วย 3 วิธีดังกล่าว เมื่อใช้จุดมุ่งหมายของการสอบเพื่อพิจารณาตำแหน่งของนักเรียนบนสเกลของความรู้ (Continuum Knowledge Scale) แต่มีความแตกต่างกันในการวัดที่มีจุดมุ่งหมายแตกต่างกันเมื่อนักเรียนมีความแตกต่างกันคือ เมื่อพิจารณาความแตกต่างกันของนักเรียน แบบสอบที่พัฒนาโดยใช้เกณฑ์คัดเลือกแบบคลาสสิคอลประมาณค่าเกินความเป็นจริง(overestimate) ในเรื่องความแปรปรวนจริงของการกระจายของคะแนนในแบบสอบซึ่งเป็นสาเหตุสำคัญทำให้ประเมินการปฏิบัติของนักเรียนอยู่ในพิสัยควอไทล์ต่ำ ในทางตรงกันข้ามแบบสอบที่พัฒนาโดยใช้รูปแบบของราสช์ประเมินค่าต่ำกว่าความเป็นจริง (underestimate)ในเรื่องความแปรปรวนเมื่อพิจารณาผลการ generalization ของแบบสอบที่พัฒนาโดยใช้รูปแบบที่พัฒนาด้วยเกณฑ์ของคลาสสิคอลประเมินค่าต่ำกว่าความเป็นจริง(underestimate) ในเรื่องสัดส่วนของเด็กที่ตอบถูก แต่แบบสอบที่พัฒนาด้วยเกณฑ์ของรูปแบบของราสช์ประเมินค่าเกินความเป็นจริง (overestimate) ในเรื่องนี้

Innocent G.Akar(1984) ได้ศึกษาผลของการเรียงลำดับข้อสอบตามความยาก

และความสามารถในแบบสอบผลสัมฤทธิ์ชนิดเลือกตอบ โดยใช้การวิเคราะห์ด้วยรูปแบบของราสช์ ซึ่งใช้กลุ่มตัวอย่างจากการสอบเด็กนักเรียน 3 ห้องที่สอบแบบสอบผลสัมฤทธิ์ในฤดูใบไม้ร่วง ปีค.ศ.1982 ที่มหาวิทยาลัยนิวยอร์ค ลักษณะของแบบสอบมี 3 ชุดคือ แบบสอบ 2 ชุดแรกเป็นแบบสอบที่มีการเรียงลำดับข้อสอบต่างกันแต่ลำดับของตัวเลือกเหมือนกัน แบบสอบชุดที่ 3 เป็นแบบสอบที่มีฟอร์มเหมือน 2 ชุดแรก แต่มีการเรียงลำดับตำแหน่งตัวถูกแตกต่างกันไป ผลการศึกษาพบว่า ข้อสอบหรือผลการตอบที่สลับลำดับที่ข้อสอบจะเปลี่ยนความยากของข้อสอบแต่ละข้อ แต่อย่างไรก็ตามแนวโน้มว่าความยากของข้อสอบทั้งหมดจะแปรเปลี่ยนไปเมื่อข้อสอบหรือผลการตอบสลับที่ลำดับกัน เมื่อทำการเปรียบเทียบข้ามแบบสอบในแต่ละชุดไม่มีความแตกต่างกัน ผู้วิจัยสรุปว่า ผลของการเรียงลำดับข้อมีความสำคัญในทางปฏิบัติเมื่อใช้กับข้อสอบเดี่ยว ๆมากกว่าที่จะใช้กับข้อสอบเป็นชุด ๆ

Bloomquist (1984) ได้ศึกษาความเหมาะสมของรูปแบบของราสช์กับการใช้ข้อมูลที่ได้จากแบบสอบอิงเกณฑ์ในการสอบด้วยแบบสอบทักษะพื้นฐาน (Basic Skills Tests) โดยใช้การวิเคราะห์ตัวประกอบและรูปแบบสมการวิเคราะห์โครงสร้าง (Structural equation modeling method) ของการตอบแบบถูกเป็น 1 ผิดเป็น 0 ปรากฏว่ารูปแบบของราสช์สามารถใช้ได้เหมาะสมกับข้อมูลที่ได้จากแบบสอบอิงเกณฑ์ที่วัดหลาย ๆจุดประสงค์ ค่าพารามิเตอร์ของข้อสอบที่ประมาณได้ไม่เปลี่ยนแปลง ซึ่งสามารถยืนยันได้จากทั้งข้อมูลจริงและการจำลองสถานการณ์ ทั้งยังมีความแกร่ง (robustness) ในเรื่องของการทดสอบคุณลักษณะมิติเดียว (unidimensionality)

Tompkins (1984) ได้ใช้รูปแบบของราสช์ศึกษาเกี่ยวกับผลของความยาวของข้อสอบต่อการประมาณค่าความสามารถในการอ่านและจำแนกความรอบรู้ของนักเรียนเกรด 9 ในรัฐแมรี่แลนด์ จากการศึกษาโดยการกำหนดความยาวของข้อสอบเป็น 10 20 และ 30 ข้อ และคำนวนความสามารถของบุคคลพบว่า ความยาวของข้อสอบมีความสัมพันธ์กับการประมาณค่าความสามารถของบุคคลและการจำแนกความรอบรู้ของผู้เรียน

อวยพร วิบูลย์กาญจน์(อวยพร วิบูลย์กาญจน์ 2526) ได้เปรียบเทียบผลการวิเคราะห์แบบสอบอุปมาอุปไมยด้วยคลาสสิคอลกับราสช์โมเดล โดยใช้กลุ่มตัวอย่าง 1,884 คน ซึ่งเป็นนักเรียนที่จบชั้นประถมศึกษาปีที่ 6 ในปีการศึกษา 2525 ที่สมัครสอบคัดเลือกเข้าเรียนต่อในชั้นมัธยมศึกษาปีที่ 1 เครื่องมือที่ใช้เป็นแบบสอบมาตรฐานวิชาวิทยาศาสตร์ และแบบสอบอุปมาอุปไมยของสำนักทดสอบทางการศึกษาและจิตวิทยา มหาวิทยาลัยศรีนครินทรวิโรฒประสานมิตร วิเคราะห์ข้อสอบด้วยคอมพิวเตอร์โปรแกรมไบคาลและโปรแกรมวิเคราะห์ข้อสอบ

ทดสอบความแตกต่างระหว่างจำนวนข้อสอบโดยใช้ค่าสถิติทดสอบ ซี (Z -test) แล้วพบว่า วิธีคลาสสิคอลได้ 34 ข้อ และราสช์โมเดลได้ 26 ข้อหาค่าความเที่ยงโดยใช้สูตร K-R 20 ได้ค่าความเที่ยง .85 และแบบราสช์โมเดลได้ .739 หาความตรงเชิงสภาพแบบคลาสสิคอล ได้ .49 และแบบราสช์โมเดลได้ .46

ชวลิต โพธิ์นคร(ชวลิต โพธิ์นคร 2528) ได้เปรียบเทียบผลของการกำหนดจุดตัด แบบทดสอบอิงเกณฑ์ชนิดเลือกตอบโดยวิธีของเบอร์ก วิธีประยุกต์ราสช์โมเดล และวิธีกำหนด เกณฑ์ผ่านระดับต่ำสุด โดยใช้นักเรียนชั้นประถมศึกษาปีที่ 4 ปีการศึกษา 2527 ในโรงเรียน สังกัดสำนักงานการประถมศึกษาจังหวัดนครสวรรค์ จำนวน 325 คน เครื่องมือที่ใช้ในการ ศึกษาคือ แบบสอบวัดความถนัดทางการเรียนด้านตัวเลข 2 ฉบับ เพื่อจัดระดับความสามารถ ของกลุ่มตัวอย่าง แบบสอบอิงเกณฑ์กลุ่มทักษะวิชาคณิตศาสตร์ชนิดเลือกตอบ 3 ฉบับเพื่อจัด กลุ่มรอบรู้และแบบสอบวัดผลสัมฤทธิ์ทางการเรียนกลุ่มทักษะวิชาคณิตศาสตร์เพื่อวัดผลสัมฤทธิ์ ของกลุ่มรอบรู้ ผลการศึกษาพบว่า

1. คะแนนผลสัมฤทธิ์ของกลุ่มรอบรู้ที่ผ่านจุดตัดที่กำหนดโดยวิธีของเบอร์ก วิธี ประยุกต์ราสช์โมเดล และวิธีกำหนดเกณฑ์ผ่านระดับต่ำสุด แตกต่างกันอย่างมีนัยสำคัญทาง สถิติที่ระดับ .05 และพบว่าค่าเฉลี่ยของกลุ่มรอบรู้ที่ผ่านจุดตัดโดยวิธีประยุกต์ราสช์โมเดล ต่ำกว่ากลุ่มที่ผ่านจุดตัดโดยวิธีกำหนดเกณฑ์ผ่านระดับต่ำสุดอย่างมีนัยสำคัญทางสถิติที่ระดับ.05

2. ค่าความตรงเชิงโครงสร้างของแบบสอบอิงเกณฑ์ที่ได้จากการกำหนดจุดตัด โดยวิธีของเบอร์กมีแนวโน้มให้ค่าสูงสุด รองลงมาได้แก่ วิธีประยุกต์ราสช์โมเดลและวิธี กำหนดเกณฑ์ผ่านระดับต่ำสุดตามลำดับ

3. จุดตัดของแบบสอบอิงเกณฑ์ที่คำนวนโดยวิธีของเบอร์กในกลุ่มตัวอย่างที่มี ความสามารถสูง ปานกลางและต่ำ พบว่าได้คะแนนจุดตัดที่แตกต่างกันและมีแนวโน้มว่า กลุ่มสูงมีจุดตัดสูงสุด รองลงมาได้แก่ กลุ่มปานกลางและกลุ่มต่ำตามลำดับ

4. จุดตัดแบบสอบอิงเกณฑ์ที่คำนวนโดยวิธีประยุกต์ราสช์โมเดลในกลุ่มตัวอย่าง ที่มีความสามารถสูง ปานกลาง และต่ำ พบว่าในกลุ่มสูงได้คะแนนจุดตัดแตกต่างไปจากกลุ่ม ปานกลาง และกลุ่มต่ำ ส่วนระหว่างกลุ่มปานกลางและกลุ่มต่ำได้คะแนนจุดตัดเท่ากัน แต่เมื่อ คำนวนร้อยละของคะแนนจุดตัดพบว่าเท่ากับร้อยละ 50 ของคะแนนทั้งหมด เมื่อคิดเฉพาะ ข้อที่เป็นไปตามข้อตกลงของราสช์โมเดลในแต่ละกลุ่ม

สมพร บุญอิ่ม(สมพร บุญอิ่ม 2529) ได้ศึกษาความคงที่ของค่าพารามิเตอร์ความยาก ในการวิเคราะห์ข้อกระทงด้วยราสช์โมเดลกับประชากรที่มีความสามารถต่างกัน 3 กลุ่ม

มีความโด่ง และความแปรปรวนเท่ากัน การแจกแจงของแต่ละกลุ่ม คือ 1) แจกแจงเบ้บวก สำหรับกลุ่มที่มีความสามารถต่ำ 2) แจกแจงแบบปกติสำหรับกลุ่มที่มีความสามารถปานกลาง 3) แจกแจงแบบเบ้ลบสำหรับกลุ่มที่มีความสามารถสูง โดยใช้เทคนิคมอนติคาร์โลซิมูเลชั่นจำลองการทดลองในเครื่องคอมพิวเตอร์ศึกษากับกลุ่มตัวอย่างกลุ่มละ 500 ข้อสอบจำนวน 30 ข้อ ทดลองซ้ำในแต่ละสถานการณ์ 100 ครั้ง ผลการศึกษาพบว่า

1. พารามิเตอร์ความยากของข้อสอบมีความคงที่ในข้อที่มีความยากไม่สูงหรือต่ำมาก กลุ่มที่มีความสามารถต่ำและกลุ่มที่มีความสามารสูง พารามิเตอร์ความยากมีความคงที่ 19 ข้อ กลุ่มที่มีความสามารถปานกลางมีความคงที่ 22 ข้อ

2. การกระจายของค่าความคลาดเคลื่อนมาตรฐานของค่าความยากของข้อสอบ เมื่อผู้สอบมีความสามารถต่ำแล้ผู้สอบที่มีความสามารถสูงมีแนวโน้มไม่เป็นสมมาตร เมื่อผู้สอบมีความสามารถปานกลางมีการกระจายเป็นสมมาตร

3. ค่าความคลาดเคลื่อนมาตรฐานของค่าความยากของข้อสอบทั้ง 3 กลุ่มไม่แตกต่างกันอย่างมีนัยสำคัญ

ตอนที่ 3 มโนทัศน์ในการออกแบบแบบสอบ

หลังจากที่ราสช์ได้เสนอรูปแบบในการวิเคราะห์ข้อสอบของเขาแล้ว ได้มีนักวัดผลจำนวนมากได้นำรูปแบบของราสช์มาศึกษาวิจัยในด้านต่าง ๆมากมาย รวมทั้งมีการเปรียบเทียบกับรูปแบบอื่น ๆของทฤษฎีการตอบสนองข้อสอบ เช่น รูปแบบ 2 พารามิเตอร์และรูปแบบ 3 พารามิเตอร์ในการวิเคราะห์ข้อสอบซึ่งทำให้ทฤษฎีการตอบสนองข้อสอบได้แพร่หลายมากขึ้น

ในปีค.ศ.1975 Graham A. Douglass (Douglass 1976: 4427-A.) ได้เสนอแนวคิดในการออกแบบแบบสอบซึ่งเป็นการกำหนดลักษณะเฉพาะของแบบสอบขึ้นจากโค้งลักษณะเฉพาะของแบบสอบ (Item Characteristic Curve) ที่มีส่วนสูง ความกว้าง และความยาวของข้อสอบ แล้วใช้รูปแบบของราสช์มาวิเคราะห์ มโนทัศน์เกี่ยวกับเรื่องนี้ได้มีส่วนที่พัฒนาประสิทธิภาพและกลวิธีในการออกแบบแบบสอบเพื่อให้เหมาะสมกับแบบสอบที่นักทดสอบต้องการนำไปทดสอบซึ่งผลลัพธ์ที่ได้จะทำให้ได้แบบสอบที่ประมาณค่าความสามารถของกลุ่มตัวอย่างที่ทำการทดสอบได้ใกล้เคียงมากที่สุด

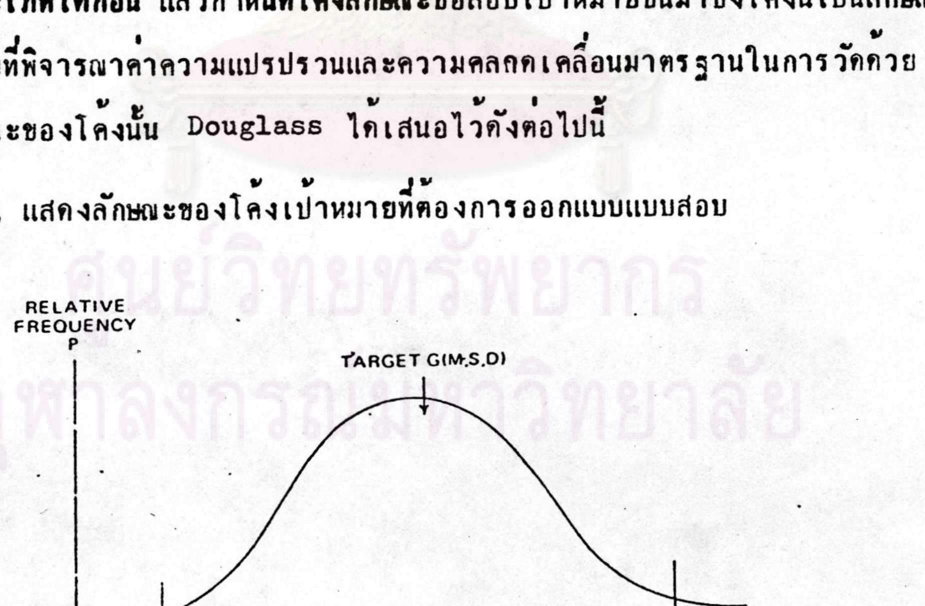
ปีค.ศ.1977 Frederic M. Lord (Lord 1977: 117-138) ได้เสนอการคัดเลือกข้อสอบและสร้างแบบสอบโดยพิจารณาค่าพารามิเตอร์ของข้อสอบแล้วคำนวนค่าฟังก์ชั่นแสดงสารสนเทศของข้อสอบ (Item information function)แต่ละข้อ

แล้วนำมาหาผลรวมเป็นค่าฟังก์ชั่นแสดงสารสนเทศของแบบสอบ (Test information function) แล้วพิจารณารูปแบบของโค้งเป้าหมายที่ตั้งไว้กับความสัมพันธ์ของข้อสอบแต่ละข้อที่บรรจุลงในแบบสอบว่ามีความสัมพันธ์กับโค้งเป้าหมายเพียงใด ซึ่งเป็นการพัฒนาการออกแบบแบบสอบเพื่อใช้สอบตามกลุ่มตัวอย่างที่ต้องการ

ปีค.ศ.1977-1978 Charles L. Hulin และคณะ (Hulin 1983: 83-93) ได้ใช้รูปแบบการพิจารณาข้อสอบจากโค้งเป้าหมายที่กำหนดไว้คล้ายกับที่ Lord ได้เสนอไว้ แต่ Hulin ได้ใช้ข้อมูลจาก Job Descriptive Index (JDI) ที่มีอยู่เป็นข้อมูลในการทดลอง โดยใช้รูปแบบ 2 พารามิเตอร์พิจารณา Attitude Scale และรูปแบบ 3 พารามิเตอร์พิจารณาแบบสอบความสามารถทางภาษา (Test of Verbal Ability) ทำให้การออกแบบแบบสอบได้มีการพัฒนาขึ้นโดยใช้ข้อมูลจริง

Wright และ Stone (Wright and Stone 1979) ได้เสนอแนวคิดในการออกแบบแบบสอบไว้ในหนังสือ Best Test Design ซึ่งมี Graham A. Douglass ให้ความร่วมมือในการนำเสนอแนวความคิด โดยสรุปแล้วแนวคิดในการออกแบบแบบสอบคือ ผู้สร้างแบบสอบควรมีหลักและเป้าหมายในการออกแบบโดยกำหนดเป้าหมายว่าต้องการแบบสอบประเภทใดก่อน แล้วกำหนดโค้งลักษณะข้อสอบเป้าหมายขึ้นมาซึ่งโค้งนี้เป็นลักษณะการกระจายที่พิจารณาค่าความแปรปรวนและความคลาดเคลื่อนมาตรฐานในการวัดด้วยรูปร่างลักษณะของโค้งนั้น Douglass ได้เสนอไว้ดังต่อไปนี้

แผนภาพที่ 5 แสดงลักษณะของโค้งเป้าหมายที่ต้องการออกแบบแบบสอบ



SHAPE D	M ± 2S	M ± 3S
Interval	.75+	.89+
Normal	.95	.99

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

พารามิเตอร์ต่าง ๆที่นำมาพิจารณาคือ

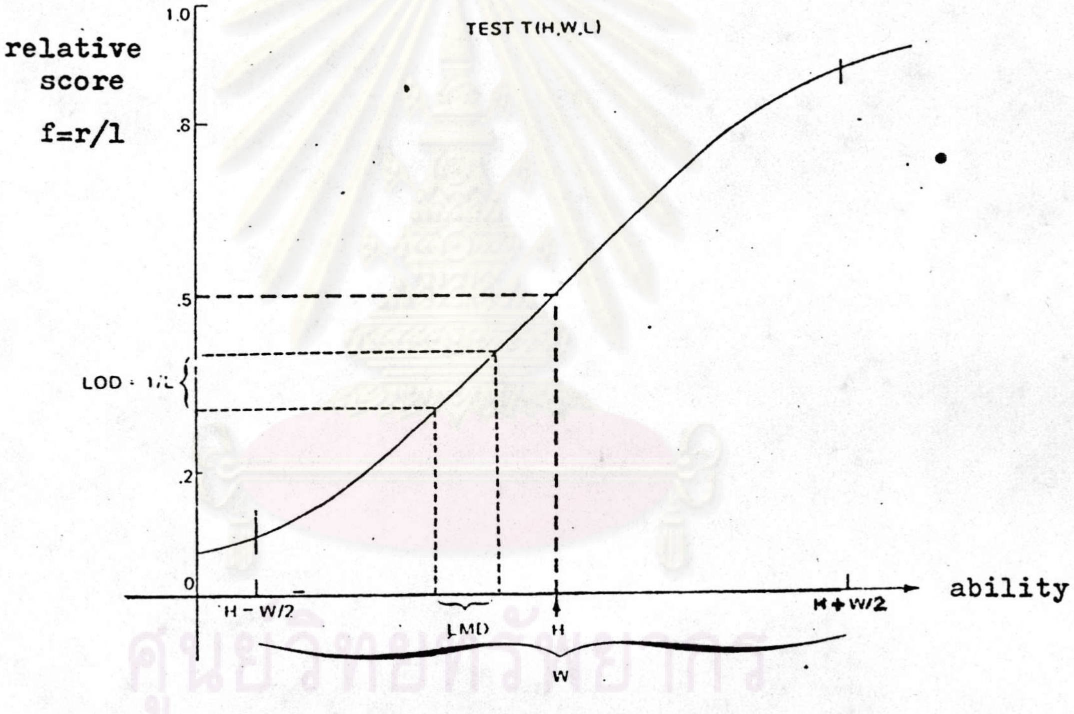
M หมายถึง ค่าเฉลี่ยของความสามารถที่เราต้องการตั้งเป็นเป้าหมาย

S หมายถึง ค่าเบี่ยงเบนมาตรฐานของโค้งที่ต้องการ

D หมายถึง ลักษณะการกระจายของโค้งที่ต้องการ

จากโค้งลักษณะเป้าหมายที่ต้องการนำมาหาความสัมพันธ์กับพารามิเตอร์ของแบบสอบที่ต้องการ โดยพารามิเตอร์ของแบบสอบแต่ละตัวต่อไปนี้มีความสัมพันธ์กันทำให้เกิดลักษณะของโค้งที่ต้องการดังแผนภาพที่ 6

แผนภาพที่ 6 แสดงความสัมพันธ์ของพารามิเตอร์ของข้อสอบที่ต้องการออกแบบแบบสอบ



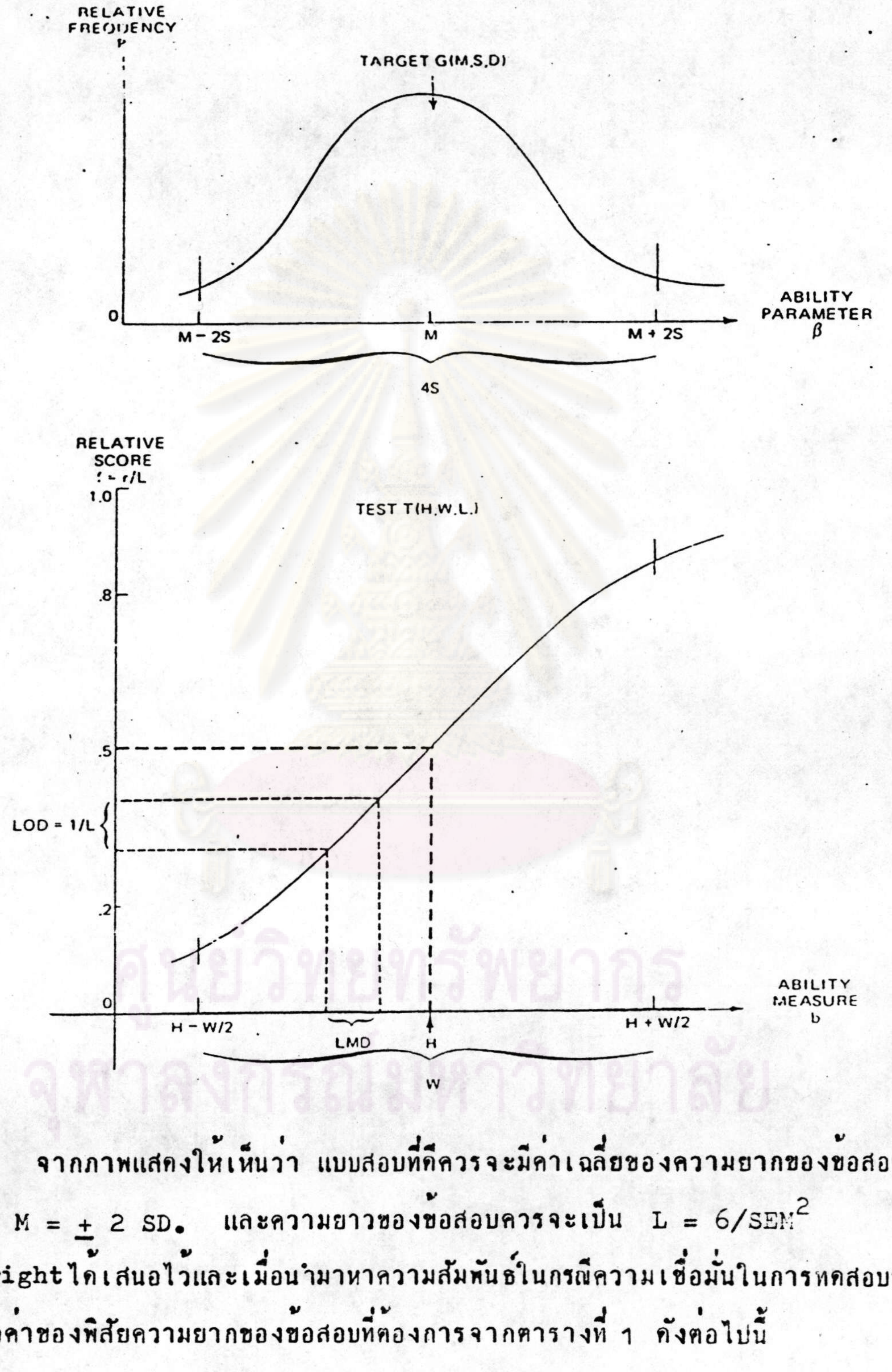
H (Height) หมายถึง ค่าเฉลี่ยของความยากของข้อสอบในแบบสอบที่เราเลือกมา

W (Width) หมายถึง พิสัยของความยากของข้อสอบที่ครอบคลุมแบบสอบฉบับนี้

L (Length) หมายถึง จำนวนข้อสอบทุกข้อที่ประกอบกันขึ้นเป็นแบบสอบ

Wright และ Douglass ได้พิจารณาโค้งที่เป็นลักษณะการกระจายแบบเดียวกันหรือเป็นโค้งปกติในการอธิบายรูปร่างของแบบสอบที่ดีและเขาได้นำลักษณะรูปร่างของแบบสอบ (Operation of a test)และโค้งลักษณะของแบบสอบเป้าหมาย (Target Test) มาพิจารณาความสัมพันธ์ร่วมกันในการออกแบบแบบสอบแต่ละฉบับ ซึ่งแสดงไว้ดังแผนภาพที่ 7

แผนภาพที่ 7 แสดงลักษณะการกระจายของโค้งเป้าหมายกับลักษณะของข้อสอบที่ต้องการ



จากภาพแสดงให้เห็นว่า แบบสอบที่ดีควรจะมีค่าเฉลี่ยของความยากของข้อสอบเท่ากับ $M = \pm 2$ SD. และความยาวของข้อสอบควรจะเป็น $L = 6/SEM^2$ ซึ่ง Wright ได้เสนอไว้และเมื่อนำมาหาความสัมพันธ์ในกรณีความเชื่อมั่นในการทดสอบนั้นได้เสนอค่าของพิสัยความยากของข้อสอบที่ต้องการจากตารางที่ 1 ดังต่อไปนี้

ตารางที่ 1 แสดงช่วงพิสัยของความเชื่อมั่นในการทดสอบของแบบสอบที่มีลักษณะของโค้งเป้าหมายที่ต่างกัน

TARGET STD. DEV. S	NORMAL TARGET error minimized over N(M,S²)	INTERVAL TARGET error minimized at (M ± 2S)	SIMPLE RULE W = 4S	EFFICIENCY	
				Normal	Interval
.5	0	0	2	94	97
.6	0	0	2		
.7	0	2	3	90	100
.8	0	3	3		
.9	0	4	4		
1.0	0	5	4	89	98
1.1	0	6	4		
1.2	1	6	5	92	96
1.3	2	7	5		
1.4	3	7	6		
1.5	4	8	6	96	91
1.6	5	9	6		
1.8	6	10	7	98	87
2.0	8	11	8	99	84

Wright ได้สรุปขั้นตอนในการออกแบบแบบสอบไว้ดังนี้

1. คาดคะแนนค่าเฉลี่ยของความสามารถ M (M = mean ability) และการกระจายของความเบี่ยงเบนมาตรฐานของความสามารถ S (Standard Deviation of Ability) ที่เราต้องการวัดซึ่ง Wright ได้เสนอไว้ให้ M = $\pm$ 2 SD. จะได้แบบสอบที่ดีที่สุด

2. ออกแบบแบบสอบด้วยค่าพารามิเตอร์ความยากของข้อสอบซึ่งมีค่าเฉลี่ยที่เท่ากับ M โดยให้ค่าความยากกระจายอยู่ในช่วง M - 2S ถึง M + 2S ด้วยแบบสอบที่สร้างขึ้นจำนวนข้อสอบที่น่าจะดีที่สุดคือ จำนวนข้อสอบ $L = 6/SEM^2$ เมื่อ SEM คือ ค่าความคลาดเคลื่อนมาตรฐานในการวัดที่พอใจในรูปแบบโลจิท(logit model)

3. เลือกข้อสอบจากคลังข้อสอบ (Item Bank) มาบรรจุให้เต็มโค้งที่ได้ออกแบบไว้และใช้ช่วงค่าความยากของข้อสอบซึ่งกำหนดไว้จาก h (height) และช่วงความกว้างของความยากที่ครอบคลุมโค้ง w (width) ค่าความคลาดเคลื่อนมาตรฐานในการวัดซึ่งหาได้จากสูตร $SEM = 2.5/L^{1/2}$

จากการออกแบบแบบสอบที่ Wright ได้เสนอไว้ข้างต้นนี้ ผู้สร้างแบบสอบจำเป็นต้องมีความรู้ในเรื่องฟังก์ชั่นแสดงสารสนเทศของข้อสอบและแบบสอบ (Item and Test

หอสมุดกลาง สถาบันวิทยบริการ จุฬาลงกรณ์มหาวิทยาลัย

information function) โค้งแสดงสารสนเทศของแบบสอบ (Test information curve)

Wright ได้เสนอเป็นรูปแบบโค้งของแบบสอบ แต่สำหรับโค้งแสดงสารสนเทศของแบบสอบคือ โค้งแสดงสารสนเทศของแบบสอบที่เราได้จากข้อสอบแต่ละข้อที่เรานำมาใช้จริง ๆแล้วกำหนดขึ้นให้คล้ายกับเป้าหมายของแบบสอบที่เราต้องการ ดังนั้นจึงควรที่จะต้องทำความรู้จักกับฟังก์ชั่นแสดงสารสนเทศของข้อสอบและของแบบสอบเสียก่อน เพื่อที่จะทำให้ผู้สร้างแบบสอบสามารถนำข้อสอบที่ต้องการมาใช้ได้ถูกต้อง

<u>ฟังก์ชั่นแสดงสารสนเทศของข้อสอบ</u> (Item information function)

ฟังก์ชั่นแสดงสารสนเทศของข้อสอบเป็นความสัมพันธ์ของอัตราส่วนระหว่างกำลังสองของค่าอนุพันธ์ ของโอกาสในการตอบถูกของผู้สอบในระดับความสามารถนั้น ๆ กับผลคูณของโอกาสในการตอบถูกและผิดของผู้สอบในระดับความสามารถนั้น ๆ(ความแปรปรวนรายข้อ) ดังสูตร

$$I(\theta,u_i) = \frac{(P'(\theta))^2}{P_i(\theta)Q_i(\theta)} \qquad \text{(Hulin 1983: 155)}$$

เมื่อ $P_i(\theta)$ คือ โอกาสในการตอบข้อสอบข้อที่ i ถูก ณ ความสามารถนั้น ๆ

$Q_i(\theta)$ คือ โอกาสในการตอบข้อสอบข้อที่ i ผิด ณ ความสามารถนั้น ๆ

$P_i'(\theta)$ คือ ค่าอนุพันธ์ (derivative) ของโอกาสในการตอบข้อสอบข้อที่ i ถูก

ซึ่งในรูปแบบโลจีสติก (Logistic Model) จะให้ค่าฟังก์ชั่นแสดงสารสนเทศของข้อสอบแตกต่างกันไปตามค่าพารามิเตอร์ความยาก (b) ค่าอำนาจจำแนก (a) และค่าการเดา (c) ที่พิจารณาได้ดังตารางที่ 2

ตารางที่ 2 แสดงค่าอนุพันธ์ (devatives) และค่าฟังก์ชั่นแสดงสารสนเทศของข้อสอบในรูปแบบโลจีสติกทั้งสามรูปแบบ

	Derivative of $P_i(\theta)$ with respect to θ: $P_i(\theta)'$	*Item Information Function: $I(\theta, u_i)$*
Three-parameter model	$\frac{(1-c_i)Da_i\exp\{-Da_i(\theta-b_i)\}}{[1+\exp\{-Da_i(\theta-b_i)\}]^2}$	$\frac{(1-c_i)D^2a_i^2\exp\{-Da_i(\theta-b_i)\}}{[1+c_i\exp\{-Da_i(\theta-b_i)\}][1+\exp\{-Da_i(\theta-b_i)\}]^2}$
Two-parameter model	$\frac{Da_i\exp\{-Da_i(\theta-b_i)\}}{[1+\exp\{-Da_i(\theta-b_i)\}]^2}$	$\frac{D^2a_i^2\exp\{-Da_i(\theta-b_i)\}}{[1+\exp\{-Da_i(\theta-b_i)\}]^2}$
One-parameter model	$\frac{D\exp\{-D(\theta-b_i)\}}{[1+\exp\{-D(\theta-b_i)\}]^2}$	$\frac{D^2\exp\{-D(\theta-b_i)\}}{[1+\exp\{-D(\theta-b_i)\}]^2}$

<u>ฟังก์ชั่นแสดงสารสนเทศของแบบสอบ</u> (Test information function)

ฟังก์ชั่นแสดงสารสนเทศของแบบสอบสามารถหาได้จากผลรวมของค่าฟังก์ชั่นแสดงสารสนเทศของข้อสอบแต่ละข้อที่ต้องการนำมาสร้างเป็นแบบสอบซึ่งเราสามารถเสนอในรูปความสัมพันธ์เชิงคณิตศาสตร์ได้ดังสูตร

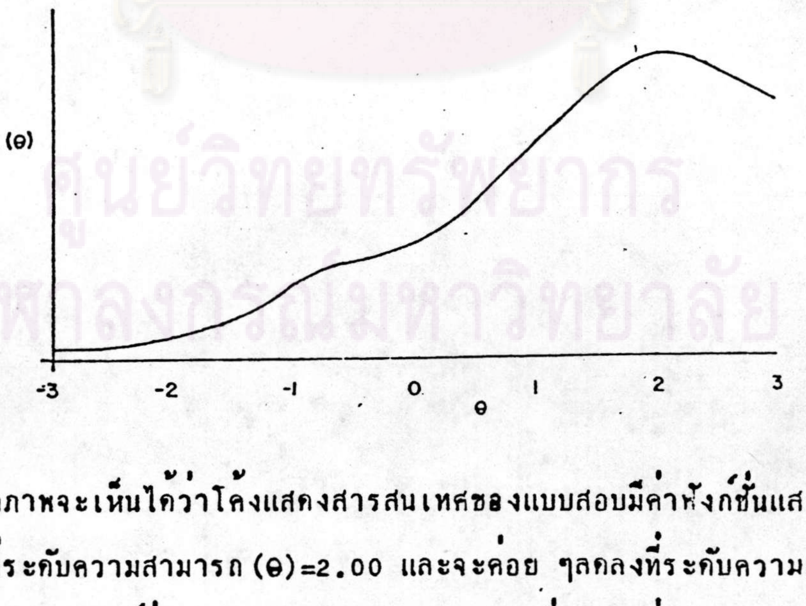
$$I(\theta) = \sum_{i=1}^{n} I(\theta, u_i) \qquad ; i = 1,2,\ldots n$$

(Hulin 1983: 55)

จากสูตรจะเห็นได้ว่าถ้าเรานำค่าฟังก์ชั่นแสดงสารสนเทศของแบบสอบมาเขียนเป็นภาพจะได้โค้งซึ่งเราเรียกว่าโค้งแสดงสารสนเทศของแบบสอบ (Test information function)

ค่าฟังก์ชั่นแสดงสารสนเทศของแบบสอบแสดงถึงความสัมพันธ์ของฟังก์ชั่นแสดงสารสนเทศของข้อสอบในแต่ละจุดของความสามารถในสเกลความสามารถที่เราพิจารณา ดังนั้นเราสามารถที่จะออกแบบโค้งแสดงสารสนเทศของแบบสอบเพื่อให้มีประโยชน์ในการวัดความสามารถที่เราต้องการได้ดังแผนภาพที่ 8 (Warm 1978: 74-75)

แผนภาพที่ 8 แสดงโค้งแสดงสารสนเทศของแบบสอบที่ตั้งไว้โดยมีประสิทธิภาพสำหรับการวัดในกลุ่มความสามารถสูง

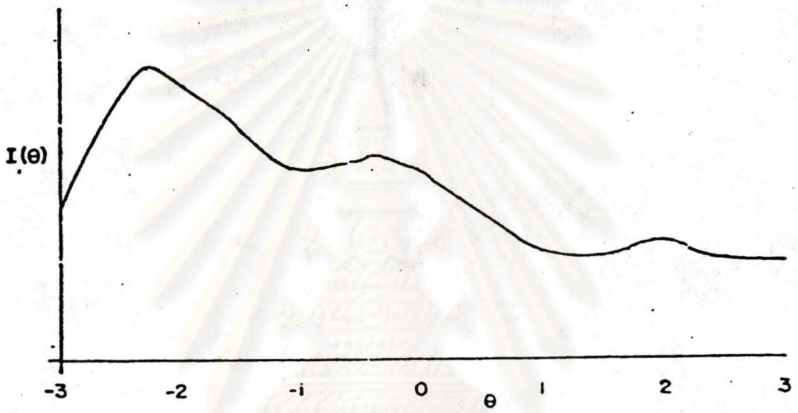


ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

จากภาพจะเห็นได้ว่าโค้งแสดงสารสนเทศของแบบสอบมีค่าฟังก์ชั่นแสดงสารสนเทศแบบสอบสูงสุดที่ระดับความสามารถ (θ)=2.00 และจะค่อย ๆลดลงที่ระดับความสามารถต่ำ เนื่องจากความสูงของฟังก์ชั่นแสดงสารสนเทศของแบบสอบที่ระดับหนึ่งของความสามารถเป็นเครื่องวัดโดยตรงถึงประโยชน์ของแบบสอบในการวัดความสามารถที่ระดับนั้น ๆได้อย่างแน่นอนและเป็นเครื่องแสดงถึงความถูกต้องแม่นยำ (precision) ของค่าความสามารถ

ที่ประมาณได้ การตอบข้อสอบแต่ละข้อ การส่งผลของข้อสอบข้อหนึ่ง ๆต่อประสิทธิผลของการวัดของข้อสอบทั้งฉบับไม่ขึ้นอยู่กับข้อสอบข้ออื่น ๆที่อยู่ในแบบสอบเดียวกัน ซึ่งสถานการณ์เช่นนี้ ตรงกันข้ามกับเรื่องความเชื่อถือได้ (reliabilityในรูปแบบการวัดที่ใช้กันอยู่ในปัจจุบัน (ผจงจิต อินทสุวรรณ 2525 : 65)

แผนภาพที่ 9 แสดงโค้งแสดงสารสนเทศของแบบสอบที่ตั้งไว้โดยมีประสิทธิภาพสำหรับการวัดในกลุ่มความสามารถต่ำ



ประโยชน์โดยตรงของความรู้เรื่องโค้งแสดงสารสนเทศของแบบสอบก็คือ ถ้าเรามีกลุ่มข้อสอบที่เราทราบสารสนเทศของข้อสอบ(Item information) เราก็สามารถสร้างแบบสอบฉบับหนึ่งให้มีโค้งแสดงสารสนเทศของแบบสอบ ณ ระดับหนึ่งของความสามารถตามที่เราต้องการได้เพื่อจุดประสงค์ในการใช้ข้อสอบนั้นโดยเฉพาะ เช่น เราต้องการสร้างแบบสอบเพื่อคัดเลือกนักเรียนให้ได้รับทุนเราก็ต้องการข้อสอบที่มีประสิทธิภาพสูงสุดที่ระดับความสามารถสูง ๆนั่นคือ ต้องการสร้างโค้งแสดงสารสนเทศของแบบสอบที่มีค่าฟังก์ชั่นแสดงสารสนเทศของแบบสอบสูงสุดที่ระดับความสามารถสูง ๆ เป็นต้น

ทุก ๆแบบสอบเมื่อใช้ในการสอบวัดย่อมมีความคลาดเคลื่อนในการวัดทุกครั้ง ความคลาดเคลื่อนมาตรฐานในการประมาณค่า(Standard Error of Estimate: S.E.E.) คือ ความเบี่ยงเบนมาตรฐานในการวัดของการประมาณค่าความสามารถ (θ) ถ้าในกลุ่มผู้สอบที่วัดมีความสามารถแตกต่างกันไป การประมาณค่าความสามารถของผู้สอบแต่ละแบบสอบจะมีค่าความคลาดเคลื่อนในการประมาณค่าซึ่งเราสามารถคำนวนได้จากสูตรดังนี้

$$S.E.E. = \frac{1}{\sqrt{I(\theta)}}$$ (Warm 1978: 77)

ค่าฟังก์ชั่นแสดงสารสนเทศของแบบสอบจะเปลี่ยนแปลงไปตามสเกลของความสามารถ ดังนั้นค่า S.E.Eจะเป็นดังนี้ ถ้าค่าของฟังก์ชั่นแสดงสารสนเทศของแบบสอบมากค่า S.E.E. จะน้อยลง ค่า S.E.E. ที่น้อย ๆจะเป็นค่าที่เราพอใจอย่างยิ่งในการวัดแต่ละครั้ง นอกจากนี้ ค่าเฉลี่ยของ S.E.E. คือ $\overline{S.E.E.}$ ของผู้สอบทั้งหมดยังมีความสัมพันธ์กับค่าความเที่ยงของทฤษฎีมาตรฐานดั้งเดิมเมื่อคะแนนเป็น Standardized ที่มีค่าความเบี่ยงเบนในการวัดเท่ากับ 1 ซึ่งเราสามารถเขียนเป็นสูตรได้ดังนี้

$$r_{tt} = 1 - \overline{S.E.E.}^2$$ (Warm 1978: 77)

เมื่อเราทราบความหมายของค่าฟังก์ชั่นแสดงสารสนเทศของข้อสอบและแบบสอบแล้วเราก็สามารถคัดเลือกข้อสอบที่เราต้องการบรรจุลงในโค้งลักษณะเป้าหมายที่ต้องการได้ ซึ่ง Lord ได้เสนอขั้นตอนไว้ดังนี้

1. ออกแบบลักษณะโค้งแสดงสารสนเทศของแบบสอบที่เราต้องการโดยพิจารณาว่าจะวัดผู้สอบที่มีความสามารถในระดับใด ซึ่งจะเป็นโค้งแสดงสารสนเทศของแบบสอบที่เราต้องการสอบวัด

2. คัดเลือกข้อสอบโดยพิจารณาค่าฟังก์ชั่นแสดงสารสนเทศของข้อสอบแต่ละข้อที่เราต้องการมาใส่ไว้ในโค้งแสดงสารสนเทศของแบบสอบจนเต็มพื้นที่โค้งเป้าหมายที่ตั้งไว้

3. คำนวนค่าฟังก์ชั่นแสงดสารสนเทศของข้อสอบแต่ละข้อที่นำมานั้นแล้วหาผลรวมเพื่อให้ได้ค่าฟังก์ชั่นแสดงสารสนเทศของแบบสอบที่สามารถพิจารณาได้ทุก ๆค่าความสามารถที่เราต้องการ

4. ปรับแก้โค้งแสดงสารสนเทศของแบบสอบเป้าหมายให้มีความเหมาะสมตามที่เราต้องการ เพื่อจะได้นำข้อสอบแต่ละข้อที่คัดเลือกมาไปใช้ในการสอบวัดนักเรียนตามจุดประสงค์ที่ตั้งไว้ต่อไป.

จากมโนทัศน์ของแบบสอบผลสัมฤทธิ์ มโนทัศน์ของรูปแบบของราสช์ และมโนทัศน์ในการออกแบบแบบสอบ พร้อมทั้งจากรายงานการวิจัยที่กล่าวมาทั้งหมดแสดงให้เห็นว่าการนำเอามโนทัศน์ของการออกแบบแบบสอบมาใช้ในการสอบวัดแบบสอบผลสัมฤทธิ์ยังไม่มีผู้ที่ทำการวิจัย

โดยศึกษาจากข้อมูลของการสอบจริงเลย ที่ทำการวิจัยก็มีเพียง Hulin และคณะ(1983) ที่ใช้มโนทัศน์ในการออกแบบแบบสอบจากโค้งแสดงสารสนเทศเป้าหมายมาวิเคราะห์กับแบบสอบวัดทัศนคติและแบบสอบความสามารถทางภาษาที่ใช้ข้อมูลของ Job Descriptive Index (JDI) โดยที่ Hulin ใช้รูปแบบ 2 พารามิเตอร์และรูปแบบ 3 พารามิเตอร์ของทฤษฎีการตอบสนองข้อสอบทำการวิเคราะห์ข้อมูลแล้วศึกษาผลของการเพิ่มจำนวนข้อสอบที่มีผลต่อโค้งแสดงสารสนเทศของแบบสอบว่าจะมีลักษณะรูปร่างของโค้งเป็นอย่างไรและตรงกับโค้งแสดงสารสนเทศเป้าหมายที่ตั้งไว้หรือไม่เพียงใด จากผลดังกล่าวนี้ผู้วิจัยจึงสนใจที่จะศึกษาถึงการนำเอามโนทัศน์ของการออกแบบแบบสอบมาใช้ในการออกแบบโค้งแสดงสารสนเทศของแบบสอบผลสัมฤทธิ์วิชาคณิตศาสตร์ ด้วยการจัดข้อสอบที่มีความยากง่ายเหมาะสมกับระดับความสามารถของผู้สอบที่แตกต่างกัน 3 ระดับ คือ กลุ่มความสามารถสูง ความสามารถปานกลาง และและความสามารถต่ำ โดยอาศัยเอกสารและงานวิจัยที่กล่าวมาเป็นแนวทางในการศึกษาครั้งนี้ ซึ่งผู้วิจัยจะใช้รูปแบบของราสช์ในการวิเคราะห์ค่าพารามิเตอร์ต่าง ๆทั้งหมด

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย