บทที่ 2

วรรณคดีที่เกี่ยวข้อง

จากการศึกษาผู้วิจัยพบว่ามีเอกสารและรายงานการวิจัยที่เกี่ยวข้องกับการศึกษาค้นคว้าหลายเรื่องด้วยกัน จึงจะนำเสนอตามลำดับดังนี้

1. การสร้างแบบสอบคำศัพท์ภาษาอังกฤษ
2. ทฤษฎีการตอบสนองข้อสอบ
3. วิธีการประมาณค่าพารามิเตอร์
4. การทดสอบแบบ ซี เอ ที
   - 4.1 กลยุทธ์ที่ใช้ในการทดสอบแบบ ซี เอ ที
   - 4.2 องค์ประกอบของกระบวนการทดสอบแบบ ซี เอ ที
   - 4.3 การดำเนินการสอบ
   - 4.4 ค่าความเที่ยงและความคลาดเคลื่อนมาตรฐานในการวัด
   - 4.5 ความตรงของการทดสอบแบบ ซี เอ ที
   - 4.6 งานวิจัยที่เกี่ยวข้องกับการทดสอบแบบ ซี เอ ที

1. การสร้างแบบสอบคำศัพท์ภาษาอังกฤษ

แฮริส (Harris, D.P. 1969:48) กล่าวว่า การสร้างข้อสอบในแบบสอบสมิทธิภาพ (Proficiency Test) นั้นปัญหาอยู่ที่การเลือกคำถามให้เหมาะสม ซึ่งเป็นเรื่องที่ค่อนข้างยากกว่าการสร้างข้อสอบในแบบสอบสัมฤทธิ์ผล และได้ให้ข้อสังเกตว่าการสร้างข้อสอบในแบบสอบสมิทธิภาพนั้นสิ่งที่จำเป็นที่จะต้องคำนึงถึงประการแรกคือ การตัดสินใจว่าจะทดสอบคำศัพท์ซึ่งเป็นคำที่ใช้บ่อย (Active Vocabulary) หรือคำศัพท์ที่ไม่ค่อยใช้ (Passive Vocabulary) และใช้คำที่มีความถี่สูงเพราะสามารถนำมาเป็นดัชนีบ่งชี้ที่น่าเชื่อถือในเรื่องค่าความยาก

แมดเซน (Madsen, H.S 1983:17) ให้ความเห็นว่า การทดสอบคำศัพท์ควรหลีกเลี่ยงการเสนอคำศัพท์เดี่ยว ๆ แต่ใช้คำแนะในบริบท (context clue) และทดสอบคำศัพท์ที่มีความหมายในตัวเอง (Content word) ซึ่งได้แก่ คำนาม คำกริยา คำคุณศัพท์และคำวิเศษณ์

วิธีการเขียนข้อสอบคำศัพท์

แฮริส (Harris, D.P. 1969:54-57) ได้เสนอแนะวิธีการเขียนข้อสอบคำศัพท์ไว้ดังนี้

1. ถ้าคำถามที่ให้เลือกคำศัพท์ที่มีความหมายตรงกับคำจำกัดความนั้น จะต้องใช้คำจำกัดความที่ง่ายและผู้สอบเข้าใจได้โดยง่าย

2. ตัวเลือกทุกตัวจะต้องมีระดับความยากพอ ๆ กันและมีความยาวพอ ๆ กัน

3. ตัวเลือกทุกตัวควรจะเป็นคำศัพท์จำพวกเดียวกันหรือเป็นกิจกรรมชนิดเดียวกัน

4. ตัวลวงจะต้องไม่มีตัวสะกดคล้ายหรือออกเสียงคล้ายกับคำที่ถูกต้อง

ฮีตัน (Heaton, J.B. 1975:44-46) เสนอแนวทางในการเขียนข้อสอบคำศัพท์ไว้ดังนี้

1. ถ้าขอบเขตของปัญหาของสิ่งที่จะทดสอบอยู่ในตัวเลือกแล้ว ตัวคำถามควรจะเขียนให้ง่าย แต่ถ้าขอบเขตของปัญหาอยู่ในตัวคำถามแล้ว ควรเขียนตัวเลือกให้ง่ายเพื่อให้ผู้สอบทุกคนอ่านแล้วเข้าใจ

2. ตัวเลือกแต่ละตัวควรจะเป็นคำในชุดเดียวกันกับคำในตัวคำถาม โดยเฉพาะอย่างยิ่งเมื่อคำ ๆ นั้นปรากฏในเนื้อหาของประโยค

3. ตัวเลือกถูกและตัวลวงควรจะมีระดับความยากพอ ๆ กัน และมีความยาว เท่า ๆ กัน

4. ความสัมพันธ์ของตัวเลือกกับขอบเขตของปัญหาที่จะนำมาทดสอบควรจะจำกัดตัวเลือกโดยใช้ขอบเขตของกิจกรรมในกลุ่มเดียวกัน

5. ควรหลีกเลี่ยงการใช้คำที่มีความหมายเหมือนกัน 2 คำ เป็นตัวลวงหรือใช้คำที่มีความหมายตรงข้ามเป็นตัวลวง

แมดเซน (Madsen, H.S. 1983:19-27) ได้ให้ข้อเสนอแนะในการเขียนตัวลวงพอสรุปได้ดังนี้

1. ตัวลวงควรจะอยู่ในรูปแบบของคำชนิดเดียวกับคำตอบที่ถูกต้อง และเกี่ยวข้องกับสิ่งที่จะทดสอบ ไม่ใช่ตัวลวงที่มีความหมายเหมือนกัน

2. อย่าแสดงคำตอบถูกผ่านตัวแนะทางไวยากรณ์ (grammartical clues)

3. ตัวลวงควรจะมีความยาวพอ ๆ กันและถ้าตัวลวงเป็นวลีพยายามไม่ให้คำตอบที่ถูกต้องนั้นยาวที่สุด

4. หลีกเลี่ยงคำถามหลอกล่อโดยการใช้คำสะกดที่ใกล้เคียงกันหรือคำที่มีเสียงตรงข้ามและหลีกเลี่ยงการจับคู่คำที่มีความหมายตรงข้ามกับคำตอบถูก และไม่มีคำตอบถูกมากกว่า 1 คำตอบ

รูปแบบการสร้างแบบสอบคำศัพท์

ในด้านรูปแบบการสร้างข้อสอบคำศัพท์นั้น ฮีตัน (Heaton, J.B. 1975:42-48) ได้แบ่งรูปแบบการสร้างแบบสอบแบบเลือกตอบออกเป็น 2 กลุ่มใหญ่ ๆ คือ

กลุ่มที่ 1 เลือกตัวอักษรหน้าคำที่มีความหมายใกล้เคียงกับคำที่ขีดเส้นใต้ และยังแบ่งย่อยได้เป็น

ชนิดที่ 1 เป็นคำถามที่เกี่ยวกับการระลึกได้ ตัวคำถาม (stem) เป็นรูปภาพให้ผู้สอบเลือกคำที่เหมาะสมที่สุดจากตัวเลือกที่กำหนดให้ คำถามประเภทนี้เหมาะสำหรับเด็กในระดับต้น ๆ

ชนิดที่ 2 เป็นคำถามที่ประกอบด้วยคำจำกัดความ โดยให้ผู้สอบเลือกตัวเลือกที่ถูกต้องตามคำจำกัดความที่กำหนดให้

ชนิดที่ 3 ตัวคำถามประกอบด้วยคำถามที่มีความหมายตามพจนานุกรม (lexical meaning) โดยให้ผู้สอบเลือกคำที่มีความหมายคล้ายคลึงหรือคำจำกัดความที่ดีที่สุด

ชนิดที่ 4 ตัวคำถามประกอบด้วยประโยค 1 ประโยค คำที่จะนำมาทดสอบปรากฏอยู่ในบริบท การทดสอบชนิดที่ 4 นี้มีประโยชน์กว่า 3 ชนิดแรก

กลุ่มที่ 2 เลือกลำอักษรของคำที่ถูกที่สุดเติมลงในประโยคให้ได้ใจความสมบูรณ์ การทดสอบในกลุ่มที่ 2 นี้สร้างยากกว่าการสร้างแบบสอบในกลุ่มที่ 1 ปัญหาหลักก็คือ เนื้อหาวิชามีน้อยเกินไปไม่เพียงพอที่จะสร้างสถานการณ์ที่มีความหมาย ในขณะเดียวกันเนื้อหามีมากเกินไปก็จะทำให้ผู้สอบได้แนวทางในการตอบมากเกินไป

แมดเซน (Madsen, H.S. 1983:12-33) แบ่งรูปแบบคำศัพท์ออกเป็น 4 ชนิดคือ

ชนิดที่ 1 เป็นแบบจำกัดคำตอบ เหมาะสำหรับผู้เริ่มเรียนโดยให้นักเรียนตอบเพียงใช่ หรือ ไม่ใช่

ชนิดที่ 2 แบบเติมคำให้สมบูรณ์ ซึ่งตัวคำถามจะมีคำที่ขาดหายไปแล้วให้นักเรียนเลือกคำตอบจากตัวเลือกที่กำหนดไว้ให้ 4 ตัวเลือก

ชนิดที่ 3 แบบถ่ายโยงข้อความ ในตัวคำถามจะเป็นประโยคและขีดเส้นใต้คำ ๆ หนึ่งแล้วให้นักเรียนเลือกคำที่มีความหมายใกล้เคียงกับคำที่ขีดเส้นใต้มากที่สุดจากตัวเลือกที่กำหนดไว้ให้ 4 ตัวเลือก

ชนิดที่ 4 เติมคำง่าย ๆ ให้สมบูรณ์โดยให้นักเรียนเขียนส่วนของคำที่ขาดหายไป

ฟินอคคิอาโรและซาโด (Finocchiaro, M. and Sako, S.1983:69-70) กล่าวว่าการทดสอบความรู้ด้านคำศัพท์อาจใช้รูปแบบได้มากมาย เช่น

1. บอกว่าประโยคที่ให้มานั้นถูกหรือผิด
2. เลือกคำที่เหมาะสมเติมลงในประโยคให้สมบูรณ์
3. เลือกคำที่ไม่สัมพันธ์จากกลุ่มคำที่กำหนดให้
4. เลือกภาพที่เหมาะสมให้สอดคล้องกับคำที่กำหนดให้
5. บอกชื่อสิ่งของที่พบในห้องเรียน
6. ให้หาคำที่มีความหมายเหมือนกันหรือคำที่มีความหมายตรงข้าม
7. เติมคำวิภัติ (prefix) ปัจจัย (suffix)
8. เปลี่ยนคำกริยาให้เป็นคำนาม คำนามให้เป็นคำคุณศัพท์ คำคุณศัพท์ให้เป็นคำกริยา
9. เลือกคำที่มีความหมายเหมือนกับคำที่ขีดเส้นใต้
10. ถอดความจากประโยคโดยใช้คำที่มีโครงสร้างต่างกันแต่ให้มีความหมายเหมือนกัน
11. เขียนประโยคหลาย ๆ ประโยคโดยใช้คำ ๆ เดียวกัน แต่ให้ความหมายต่างกัน

12. ใช้แบบสอบไดลช

13. เขียนตามคำบอก

ในการทดสอบคำศัพท์นั้นมักจะทดสอบคำศัพท์ที่มีความหมายในตัวเอง (content word) ซึ่งได้แก่ คำนาม คำกริยา คำคุณศัพท์และคำวิเศษณ์ (Madsen 1983:17) ผู้วิจัยจึงเลือกใช้คำศัพท์ที่มีความหมายในตัวเองในการสร้างแบบสอบคำศัพท์และจากการศึกษางานวิจัยของกัลยา ไกวัลวณิช (2527:48-119) เกี่ยวกับการหาความถี่ของคำศัพท์ซึ่งเป็นคำที่มีความหมายในตัวเอง (Content Word) ได้แก่ คำนาม คำกริยา คำคุณศัพท์และคำวิเศษณ์ โดยไม่รวมถึงคำที่เป็นชื่อเฉพาะชื่อเดือน วัน คำบอกจำนวน คำแสดงความเป็นเจ้าของและคำกริยานุเคราะห์ จากหนังสือแบบเรียนของนักเรียนชั้นมัธยมศึกษาปีที่ 1 ได้แก่ Get Set, Magic Sounds and Words Book One, Magic Sounds and Words Book Two, Active English for Thailand Book One, Junior Active Context English Book One, และ Yes! Book D ได้คำที่มีความหมายในตัวเองจำนวน 1,416 คำ แต่งานวิจัยนี้ไม่ได้แจกแจงความถี่ของคำเหล่านี้ไว้ว่ามีความถี่คิดเป็นร้อยละเท่าใดผู้วิจัยจึงนำคำที่ปรากฏในตารางที่ 4 ในงานวิจัยชิ้นนี้มาแจกแจง พบว่า เป็นคำนามร้อยละ 59.61 คำกริยาร้อยละ 20.34 คำคุณศัพท์ร้อยละ 12.71 และคำวิเศษณ์ร้อยละ 7.35 ผลการแจกแจงนี้ใกล้เคียงกับงานวิจัยของทิพาวรรณ จันทรพิบูลย์ (2519:62) ได้ประเมินศัพท์สกรณ์ภาษาอังกฤษสำหรับนักเรียนระดับชั้นมัธยมศึกษาตอนต้นจากหนังสือ A Direct Method English Course Book I, II, III, English for Thai Students Book I, II, III, และ Intermediate English Course for Thailand Book I, II, III ผลวิจัยพบว่าได้คำศัพท์ทั้งสิ้น 2,750 คำ เป็นคำศัพท์สกรณ์ (Active Vocabulary) 1,270 คำ คิดเป็นร้อยละ 46.18 ของคำศัพท์ทั้งหมดและจากคำศัพท์สกรณ์ 1,270 คำนี้แยกประเภทตามหน้าที่ของคำได้ผลดังนี้ คำนามร้อยละ 51.18 คำกริยาร้อยละ 21.18 คำคุณศัพท์ร้อยละ 15.90 คำวิเศษณ์ร้อยละ 9.21 คำสรรพนามร้อยละ 2.98 คำบุพบทร้อยละ 2.67 และอื่น ๆ คิดเป็นร้อยละ 1.57

จากงานวิจัยดังกล่าวพบว่า ในหนังสือแบบเรียนระดับมัธยมศึกษาตอนต้นนั้นมีคำที่มีความหมายในตัวเองที่เป็นคำนามมีน้ำหนักมากที่สุด คำกริยา คำคุณศัพท์และคำวิเศษณ์มีน้ำหนักรองลงมาตามลำดับ ผู้วิจัยจึงนำผลวิจัยดังกล่าวมากำหนดน้ำหนักของคำศัพท์ที่มีความหมายในตัวเอง ในการสร้างแบบสอบคำศัพท์ภาษาอังกฤษเป็น 5.5 : 2.0 : 1.5 : 1.0 และเลือกรูปแบบที่ใช้ในการสร้างข้อสอบเป็น 4 รูปแบบ คือ

1. แบบเลือกคำจากัดความที่ดีที่สุดให้เหมาะกับคำที่กำหนดให้

2. แบบเลือกคำที่มีความหมายเหมือนกับคำที่ขีดเส้นใต้

3. แบบเลือกคำที่มีความหมายตรงข้ามกับคำที่ขีดเส้นใต้

4. แบบเติมคำให้สมบูรณ์โดยให้เลือกคำที่มีความหมายเหมาะกับประโยคที่กำหนดให้

แมดเซน (Madsen) กล่าวถึงการทดสอบคำศัพท์ว่าควรจะหลีกเลี่ยงการเสนอคำศัพท์เดี่ยว ๆ แต่ควรใช้คำแนะในบริบท (context clues) จากรูปแบบดังกล่าวพบว่ารูปแบบเติมคำให้สมบูรณ์โดยให้เลือกคำที่มีความหมายเหมาะกับประโยคที่กำหนดให้นั้นเป็นรูปแบบที่มีประโยชน์มากที่สุด เพราะเป็นรูปแบบที่มีบริบทเหมาะกับการใช้ภาษาจริงมากกว่ารูปแบบอื่น ๆ ซึ่งสอดคล้องกับความคิดเห็นและข้อเสนอแนะของผู้เชี่ยวชาญว่าควรจะกำหนดน้ำหนักของรูปแบบเป็น 1 : 1 : 1 : 7

อนึ่ง ในการทดสอบแบบ ซี เอ ที (Computerized Adaptive Testing) จำเป็นต้องใช้กลุ่มข้อสอบ (Item Pool) ที่ประกอบด้วยข้อสอบจำนวนมาก ๆ ที่ครอบคลุมความสามารถของผู้สอบและวัดความสามารถมิติเดียวกัน ผู้วิจัยจึงสร้างแบบสอบคำศัพท์ประกอบด้วยข้อสอบจำนวน 500 ข้อ โดยใช้คำศัพท์ที่มีความหมายในตัวเองโดยการเลือกคำศัพท์ที่ปรากฏในหนังสือแบบเรียนภาษาอังกฤษชั้นมัธยมศึกษาตอนต้นซึ่งโรงเรียนในสังกัดกรมสามัญศึกษา จังหวัดนนทบุรี ใช้เป็นส่วนใหญ่คือ Junior Active Context English Book I, II, III, Learning Kit Book I, II, III, IV, V, VI, English for a Changing World Book I, II, III, และ ODYSSEY Book I, II, III ได้คำศัพท์ที่เป็นคำนาม 275 ข้อคำกริยา 100 ข้อ คำคุณศัพท์ 75 ข้อและคำวิเศษณ์ 50 ข้อ จากนั้นผู้วิจัยเขียนข้อสอบตามวิธีการเขียนข้อสอบที่ได้ศึกษามา

2. ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory)

ทฤษฎีการทดสอบแบบประเพณีนิยม (Classical Test Theory) เป็นทฤษฎีที่ไม่เหมาะสมกับการทดสอบแบบ ซี เอ ที เนื่องจากดัชนีความเที่ยง ความตรงและคุณภาพของข้อสอบแบบประเพณีนิยมนั้นจะสอดคล้องกันเมื่อผู้สอบทุกคนได้ทำข้อสอบชุดเดียวกัน แต่การทดสอบแบบ ซี เอ ที นั้นผู้สอบแต่ละคนจะได้รับข้อสอบต่างกัน จึงจำเป็นต้องใช้ทฤษฎีที่เหมาะสม เช่น ทฤษฎีลาเท้นท์เทรต (Latent Traits Theory) (Birnbaum 1958 อ้างถึงใน Green, et al. 1984:347) ลอร์ด (Lord) นำมาปรับปรุงใหม่โดยใช้ชื่อว่า ทฤษฎีการตอบสนองข้อสอบ (IRT) ทฤษฎีนี้อธิบายถึงความสัมพันธ์ระหว่างการตอบข้อสอบกับระดับความสามารถที่มีอยู่ภายในตัวผู้สอบ โดยใช้รูปแบบทางคณิตศาสตร์ที่กำหนดด้วยฟังก์ชันทางคณิตศาสตร์ เรียกว่า ฟังก์ชันลักษณะข้อสอบ (Item Characteristic Function) หรือโค้งลักษณะข้อสอบ (Item Characteristic Curve) เพื่อที่จะกำหนดความสัมพันธ์ระหว่างคุณลักษณะแฝง (Traits) ข้อสอบและการตอบข้อสอบจากฟังก์ชันทางคณิตศาสตร์และความสัมพันธ์นี้สามารถดัดแปลงได้หลายรูปแบบตามลักษณะข้อตกลงเบื้องต้นเกี่ยวกับทฤษฎีการตอบข้อสอบ

015846

คุณสมบัติของรูปแบบการตอบข้อสอบ

แฮมเบิลตันและสวามินาแธน (Hambleton and Swaminathan 1985:9) ได้สรุปไว้ดังนี้

1. เป็นรูปแบบที่อธิบายเกี่ยวกับปฏิบัติการของผู้สอบในการทำแบบสอบว่าสามารถอธิบายได้ในลักษณะของคุณลักษณะหนึ่งหรือหลายคุณลักษณะ ซึ่งเรียกว่า คุณลักษณะแฝง

2. รูปแบบการตอบข้อสอบกำหนดเป็นความสัมพันธ์ระหว่างปฏิบัติการของผู้สอบในการตอบข้อสอบที่สังเกตได้และคุณลักษณะแฝงหรือความสามารถที่ถือว่าเป็นปฏิบัติการหลักของพฤติกรรมการทำแบบสอบ

3. รูปแบบการตอบข้อสอบที่ดีจะต้องให้วิถีทางในการประมาณคะแนนของผู้สอบได้ตรงกับคุณลักษณะแฝงนั้น

4. คุณลักษณะแฝงอธิบายได้จากปฏิบัติการที่สังเกตได้ของผู้สอบในการตอบข้อสอบ

ข้อตกลงเบื้องต้นของทฤษฎีการตอบข้อสอบ

1. **มิติของลาเท้นท์สเปส** (Dimensionality of the Latent Space) ลาเท้นท์สเปส (Latent Space) หมายถึง คุณลักษณะภายในหรือความสามารถที่เป็นตัวกำหนดพฤติกรรมการตอบข้อสอบข้อใดข้อหนึ่งอาจมีได้หลายคุณลักษณะ จำนวนคุณลักษณะทั้งหมดในลาเท้นท์สเปสก็คือ มิติของลาเท้นท์สเปส โดยทั่วไปวิธีการทดสอบแบบ ซี เอ ที และทฤษฎีการตอบสนองข้อสอบมักจะถือว่าข้อสอบแต่ละข้อในแบบสอบนั้นวัดความสามารถหรือคุณลักษณะเดียวกัน (Unidimensionality) หรือข้อสอบเหล่านี้มีลักษณะเป็นเอกพันธ์ (Homogeneous) การกำหนดเช่นนี้ก็เพื่อช่วยให้รูปแบบของทฤษฎีนี้มีความซับซ้อนน้อยลงและง่ายแก่การแปลความหมายของคะแนนที่ได้จากแบบสอบ

วิธีการที่จะทำให้แน่ใจได้ว่าข้อสอบชุดนี้วัดความสามารถหรือคุณลักษณะเดียวกันมีหลายวิธีได้แก่ การวิเคราะห์องค์ประกอบ (Factor Analysis) การเลือกข้อสอบที่มีค่าอำนาจจำแนกสูง เนื่องจากข้อสอบที่มีค่าอำนาจจำแนกสูงจะมีสหสัมพันธ์กับคะแนนแบบสอบทั้งหมดสูง ยูริ (Urry 1981) เสนอไว้ว่าควรเลือกข้อสอบที่มีค่าอำนาจจำแนกตั้งแต่ .8 ขึ้นไป (Green, et al. 1984:347)

2. **ความเป็นอิสระในการตอบข้อสอบ** (Local Independence) หมายถึง การตอบข้อสอบข้อต่าง ๆ ในแบบสอบของผู้สอบนั้นมีความเป็นอิสระในเชิงสถิติ นั่นคือ การตอบข้อสอบ

ข้อหนึ่ง ๆ จะไม่มีผลกระทบต่อการตอบข้อสอบข้ออื่น ๆ ในแบบสอบ แฮมเบิลตันและสแวมมินาแธน (Hambleton and Swaminathan 1985:25) มีความเห็นตรงกับแมคโดแนล (McDonald) ว่าคำจำกัดความของความเป็นมิติเดียวกันของข้อสอบจะมีความหมายควรจะอิงกับคำจำกัดความของความเป็นอิสระในการตอบข้อสอบ

3. **โค้งลักษณะข้อสอบ** (Item Characteristic Curve) เป็นฟังก์ชันทางคณิตศาสตร์ที่แสดงถึงความสัมพันธ์ระหว่างโอกาสในการตอบข้อสอบนั้นได้ถูกต้องกับระดับความสามารถที่วัดได้โดยชุดของข้อสอบหรือแบบสอบนั้น โค้งลักษณะข้อสอบมีได้หลายรูปแบบตามความเชื่อเกี่ยวกับความสัมพันธ์ระหว่างระดับความสามารถกับโอกาสที่จะตอบถูก จึงทำให้รูปแบบที่ใช้อธิบายความสัมพันธ์ดังกล่าวมีหลายรูปแบบด้วยกัน เช่น Guttman Perfect Scale, Latent Distance Model, Linear Model, Normal Ogive Model, Logistic Model, Nominal Response Model และ Grade Response Model (จักรกฤษณ์ สาราญใจ 2531:30-38) แต่รูปแบบที่นิยมใช้กันแพร่หลายคือ รูปแบบโลจิสติก (Logistic Model) รูปแบบนี้แบ่งออกเป็น 4 รูปแบบตามจำนวนพารามิเตอร์ของโค้งลักษณะข้อสอบดังนี้ (Hambleton and Swaminathan 1985:37-49)

1. รูปแบบโลจิสติกที่มี 2 พารามิเตอร์ (Two-Parameter Logistic Model) เป็นรูปแบบที่เบิร์นบอม (Birnbaum 1957,1958a,1958b,1968) เสนอไว้แทนรูปแบบนอร์มอล ไอโจ้ เนื่องจาก รูปแบบโลจิสติกเป็นรูปแบบเชิงคณิตศาสตร์ที่นำมาใช้ในทางปฏิบัติได้สะดวกกว่า รูปแบบนี้ถือว่าไม่มีการเดา ดังสมการ

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

โดยที่ $P_i(\Theta)$ คือ ความน่าจะเป็นที่ผู้สอบที่มีระดับความสามารถ Θ ตอบข้อสอบข้อนั้นถูกต้อง

Θ คือ ระดับความสามารถของผู้สอบที่ประมาณได้จากคะแนนรวมของการตอบแบบสอบ โดยการปรับให้เป็นคะแนนที่มีค่าเฉลี่ยเท่ากับ 0 และมีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1 โดยมีพิสัยการกระจายเท่ากับ $-\infty$ ถึง $+\infty$ แต่ในทางปฏิบัติระดับความสามารถจะมีค่าอยู่ระหว่าง -3 ถึง +3

a_i คือ ค่าอำนาจจำแนกของข้อสอบข้อ i ที่เป็นสัดส่วนโดยตรงกับความชันของโค้งลักษณะข้อสอบ ณ จุดเปลี่ยนโค้งหรือที่จุด O ในเชิงทฤษฎีค่า a_i มีค่าอยู่ในช่วงระหว่าง $-\infty$ ถึง $+\infty$ แต่ในทางปฏิบัติจะใช้ค่าที่อยู่ระหว่าง 0 ถึง 2

b_i คือ ค่าความยากของข้อสอบข้อ i ที่แสดงถึงระดับความสามารถที่จุดโค้งลักษณะข้อสอบมีความชันมากที่สุด ในทางปฏิบัติค่าความยากจะมีค่าอยู่ระหว่าง -2 ถึง +2

D คือ ค่า scaling factor มีค่าเท่ากับ 1.7

e คือ ค่าคงที่มีค่าเท่ากับ 2.71828

2. รูปแบบโลจิสติกที่มี 3 พารามิเตอร์ (Three - Parameter Logistic Model) เป็นรูปแบบที่ได้จากรูปแบบที่ใช้ 2 พารามิเตอร์ โดยเพิ่มค่าพารามิเตอร์ c_i คือ ค่าการเดา รูปแบบดังสมการ

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

โดยที่ c_i คือ ค่าการเดาของข้อสอบข้อ i ซึ่งได้จากเส้นกำกับโค้งลักษณะข้อสอบและแทนความน่าจะเป็นของผู้สอบที่มีความสามารถต่ำตอบข้อสอบได้ถูกต้อง ในทางปฏิบัติ ค่า c_i มีค่าระหว่าง 0 ถึง 0.3 เมื่อข้อสอบมี 4 ตัวเลือก

3. รูปแบบโลจิสติกที่มี 1 พารามิเตอร์ (One-Parameter Logistic Model) หรือรูปแบบราส์ช (Rasch Model) เป็นรูปแบบที่มีข้อตกลงเบื้องต้นว่า ค่าอำนาจจำแนกของข้อสอบทุกข้อเท่ากันและค่าการเดาของข้อสอบเป็น 0 ดังสมการ

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

4. รูปแบบโลจิสติกที่มี 4 พารามิเตอร์ (Four-Parameter Logistic Model) รูปแบบนี้คำนึงถึงผู้สอบที่มีความสามารถสูง เนื่องจากผู้สอบที่มีความสามารถสูงบางครั้งอาจมีความสะเพร่าในการทำข้อสอบ จึงทำให้ตอบข้อสอบข้อนั้นผิดพลาดไป แมคโดแนล (McDonald อ้างถึงใน Hambleton and Swaminathan 1985:48) จึงเสนอรูปแบบนี้ขึ้นดังสมการ

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

รูปแบบนี้ต่างจากรูปแบบ 3 พารามิเตอร์คือค่า γ ให้มีค่าต่ำกว่า 1 เล็กน้อย แต่อย่างไรก็ตาม รูปแบบนี้ยังไม่สามารถนำมาใช้ในทางปฏิบัติได้

อนึ่ง ค่าพารามิเตอร์ต่าง ๆ ที่ได้จากการวิเคราะห์ข้อสอบรายข้อจากรูปแบบต่าง ๆ ข้างต้นนี้ เป็นค่าที่ไม่เปลี่ยนแปลงไปตามกลุ่มตัวอย่างและไม่เปลี่ยนแปลงไปตามตำแหน่งของข้อสอบ คุณลักษณะดังกล่าวนี้คือความเป็นปรนัยของการวัดผล (Objectivity of Measurement) ที่มีความสำคัญมากในทางวัดผล (Wright and Stones อ้างถึงใน สุพัฒน์ สุกมลสันต์ 2528:49)

ค่าฟังก์ชันสารสนเทศแบบสอบ

สารสนเทศนิยามได้จากลักษณะของความไม่แน่นอนของเหตุการณ์บางประการ ถ้าความไม่แน่นอนในการเกิดเหตุการณ์หนึ่งมีน้อย แสดงว่ามีสารสนเทศเกี่ยวกับเหตุการณ์นั้นมาก จากนิยามนี้จึงนำไปใช้เป็นมโนทัศน์ทั่วไปในทฤษฎีการตอบสนองข้อสอบ โดยพิจารณาจากกระบวนการอ้างอิงในเชิงสถิติ ความไม่แน่นอนในการประมาณค่าพารามิเตอร์ของประชากรในเชิงทฤษฎีเป็นค่าที่ไม่จำกัดมีค่าอยู่ระหว่าง $-\infty$ ถึง $+\infty$

ความไม่แน่นอนเกี่ยวกับประชากรหรือพารามิเตอร์ในเชิงทฤษฎีสามารถแสดงได้ในลักษณะของความแปรปรวนจากการสุ่มของการประมาณค่าพารามิเตอร์จากกลุ่มตัวอย่าง ถ้าเราประมาณค่าเฉลี่ยของประชากรจะได้ ความคลาดเคลื่อนมาตรฐานกำลังสองของค่าเฉลี่ยเป็น $\sigma^2_{\bar{x}} = \sigma^2_x/N$ ถ้าความคลาดเคลื่อนมาตรฐานกำลังสองมีค่าน้อยเท่าไร จะได้สารสนเทศมากยิ่งขึ้น ดังนั้น จึงกำหนดสารสนเทศเกี่ยวกับพารามิเตอร์เป็นสัดส่วนกลับของความคลาดเคลื่อนมาตรฐานกำลังสอง ดังนี้

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

สารสนเทศเป็นเครื่องมือที่มีประโยชน์ในการประเมินปฏิบัติการของแบบสอบ ซึ่งเป็นสิ่งสำคัญมากในแนวทฤษฎีการตอบสนองข้อสอบ เนื่องจากแบบสอบและสเกลต่าง ๆ พัฒนาขึ้นโดยใช้กระบวนการตอบข้อสอบที่ประเมินได้จากความแน่นอนในการประมาณค่าความสามารถ (θ) จากแบบสอบและสเกลต่าง ๆ โดยทั่วไปแล้วจะใช้ค่าความคลาดเคลื่อนมาตรฐานกำลังสองของ θ

หรือสารสนเทศเป็นการวัดความแน่นอนในการประมาณค่า สารสนเทศไม่จำเป็นต้องคงที่ตลอดพิสัยความสามารถ ค่า สารสนเทศ ณ ระดับความสามารถใดขึ้นกับจำนวนและคุณลักษณะของข้อสอบที่ใช้ในการประมาณค่าความสามารถ (θ)

ความคลาดเคลื่อนมาตรฐานในการประมาณความสามารถอธิบายได้ว่าเป็นส่วนเบี่ยงเบนมาตรฐานของการกระจายความสามารถโดยใช้ความสัมพันธ์ระหว่างสารสนเทศและความคลาดเคลื่อนมาตรฐานกำลังสอง นั่นคือ $6\hat{\theta}/\theta = 1 / \sqrt{I(\theta)}$

เบิร์นบอม (Birnbaum 1968) และลอร์ด (Lord 1980) (อ้างถึงใน Hulin, et al. 1983:55) ได้นิยามฟังก์ชันสารสนเทศแบบสอบเป็นสัดส่วนกลับของความแปรปรวนในการสุ่มของการประมาณค่าความสามารถแบบความเป็นไปได้สูงสุด (Maximum Likelihood Estimation) แสดงว่าเมื่อจำนวนข้อสอบเพิ่มขึ้น $1 / \sqrt{I(\theta)}$ เป็นความคลาดเคลื่อนมาตรฐานของ θ

เมื่อใช้วิธีการประมาณค่าแบบความเป็นไปได้สูงสุด นั่นคือ

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

เมื่อ $P_i(\theta)'$ เป็น ค่าอนุพันธ์ของโค้งลักษณะข้อสอบ (ICC) ของข้อสอบข้อ i ณ ระดับความสามารถ θ

และค่าสารสนเทศของข้อสอบได้จากสูตร

$$I(\theta,U_i) = \frac{[P_i(\theta)']^2}{P_i(\theta)\ Q_i(\theta)}$$

ในการสอบทุกครั้งย่อมมีความคลาดเคลื่อนเกิดขึ้นความคลาดเคลื่อนมาตรฐานในการประมาณค่าเป็นส่วนเบี่ยงเบนมาตรฐานที่คาดของความคลาดเคลื่อนในการประมาณค่าความสามารถ นั่นคือ ถ้าเรานำแบบสอบชุดหนึ่งไปสอบกับผู้สอบกลุ่มหนึ่งที่มีความสามารถคล้ายคลึงกันและประมาณความสามารถของเขาโดยใช้แบบสอบฉบับนั้น ส่วนเบี่ยงเบนมาตรฐานของการประมาณค่าจะเป็นความคลาดเคลื่อนมาตรฐานของการประมาณค่า (Standard Error of Estimate) ซึ่งคำนวณได้จากสูตร

$$SEE = \frac{1}{\sqrt{I(\theta)}}$$

นั่นคือ SEE เท่ากับรากที่สองของส่วนกลับของฟังก์ชันสารสนเทศของแบบสอบ

เนื่องจากค่าฟังก์ชันสารสนเทศแบบสอบมีค่าแปรเปลี่ยนไปตามสเกลความสามารถ เช่นเดียวกับความคลาดเคลื่อนมาตรฐานของการประมาณค่า นั่นคือ ถ้าค่าฟังก์ชันสารสนเทศยิ่งมากค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถยิ่งน้อยลง ในการสอบวัดทุกครั้งเราต้องการให้ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถมีน้อย นอกจากนี้ค่าเฉลี่ยของ SEE ($\overline{SEE}$) ของผู้สอบทั้งหมดมีความสัมพันธ์กับค่าความเที่ยงในเชิงทฤษฎีแบบประเพณีนิยมด้วยเมื่อคะแนนเป็นมาตรฐานและมีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1 ดังสูตร (Warm 1978:77)

$$r_{tt} = 1 - (\overline{SEE})^2$$

3. วิธีการประมาณค่าพารามิเตอร์

แฮมเบิลตันและคณะ (Hambleton, et al.1977 อ้างถึงใน จักรกฤษณ์ สาราญใจ 2531:38) สรุปวิธีการประมาณค่าพารามิเตอร์ของรูปแบบต่าง ๆ ในทฤษฎีการตอบสนองข้อสอบว่ามี 3 วิธีการใหญ่ ๆ คือ

1. วิธีการประมาณค่าแบบความเป็นไปได้สูงสุด (Maximum Likelihood Estimation)

2. วิธีการประมาณค่าแบบฮิวริสติค (Heuristic Estimation)

3. วิธีการประมาณค่าแบบเบย์ (Bayesian Estimation)

การประมาณค่าแบบความเป็นไปได้สูงสุด

การประมาณค่าแบบความเป็นไปได้สูงสุดเป็นวิธีการประมาณค่าที่นิยมใช้กันมากเพราะมีประสิทธิภาพค่อนข้างสูงได้มีการนำวิธีการนี้มาพัฒนาเป็นโปรแกรมคำสั่งสำเร็จรูปโดยให้คอมพิวเตอร์ประมาณค่าพารามิเตอร์หลายโปรแกรม เช่น LOGIST, BICAL, DATAGEN (Hambleton and Cook 1977:89-90) NORMOG (Hulin, Drasgow and Parson 1983:53), PML และ BILOG (Hambleton and Swaminathan 1985:148) และวิธีการประมาณค่าแบบนี้มีคุณสมบัติที่สำคัญและมีประโยชน์ภายใต้สถานการณ์ทั่ว ๆ ไป ดังนี้คือ (Hambleton and Swaminathan 1985: 88-89)

1. มีความคงที่ (consistent) ได้แก่ เมื่อขนาดของกลุ่มตัวอย่างและจำนวนข้อสอบเพิ่มขึ้น ตัวประมาณค่าจะเป็นค่าที่ถูกต้อง

หอสมุดกลาง สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

2. ทำหน้าที่ของค่าสถิติที่เพียงพอเมื่อมีค่าสถิติที่เพียงพอ (Functions of sufficient statistics when sufficient statistics exist) ได้แก่ ค่าสถิติที่เพียงพอจะมีสารสนเทศทั้งหมดเกี่ยวกับค่าพารามิเตอร์ ไม่จำเป็นต้องใช้ข้อมูลใด ๆ เพิ่มเติม

3. มีประสิทธิภาพ (Efficient) นั่นคือ ตัวประมาณค่าแบบความเป็นไปได้สูงสุดมีความแปรปรวนน้อยที่สุด

4. มีลักษณะของการกระจายเข้าใกล้ปกติ (Asymptotically normally distributed)

กลุ่มข้อสอบที่ใช้ในการทดสอบแบบ ซี เอ ที ประกอบด้วยข้อสอบที่หาบค่าพารามิเตอร์ของข้อสอบ ได้แก่ ค่าอำนาจจำแนกของข้อสอบ ค่าความยากของข้อสอบและค่าการเดาของข้อสอบ จึงใช้การประมาณค่าความสามารถอย่างมีเงื่อนไข (Conditional Maximum Likelihood Eastimation of ability) (Hambleton and Swaminathan 1985:81) การประมาณค่าความสามารถแบบความเป็นไปได้สูงสุดได้จากฟังก์ชันความเป็นไปได้ (Likelihood Function) มีลักษณะดังนี้

ความน่าจะเป็นของผู้สอบที่มีความสามารถ Θ ตอบข้อสอบข้อ i เป็น

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

แสดงได้ดังสมการ

$$P(U_i/\Theta) = P(U_i = 1/\Theta)\, P(U_i = 0/\Theta)$$

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

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

เมื่อ $Q_i = 1-P_i$

ถ้าผู้สอบที่มีระดับความสามารถ Θ ตอบข้อสอบ n ข้อ ความน่าจะเป็นร่วมของการตอบ $U_1, U_2, \ldots, U_n$ สามารถแทนด้วย $P(U_1, U_2, \ldots, U_n/\Theta)$ ถ้ากลุ่มข้อสอบเป็นมิติเดียวกันแล้ว ความเป็นอิสระในการตอบข้อสอบ n ข้อเป็นอิสระจากกัน นั่นคือ

$$P(U_1, U_2, \ldots, U_n/\Theta) = P(U_1/\Theta)\, P(U_2/\Theta), \ldots, P(U_n/\Theta)$$

$$= \sum_{i=1}^{n} P(U_i/\Theta) \qquad (2)$$

$$= \sum_{i=1}^{n} P_i^{u_i} Q_i^{(1-u_i)} \qquad (3)$$

สมการข้างต้นเป็นสมการความน่าจะเป็นร่วมของการตอบข้อสอบ n ข้อ แต่อย่างไรก็ตาม เมื่อสามารถสังเกตการตอบได้ เช่น $U_1, U_2, \ldots, U_n$ เป็นตัวแปรสุ่มที่มีค่าเฉพาะเป็น $u_1, u_2, u_3, \ldots, u_n$ โดยที่ u_i เป็น 1 หรือ 0 แล้วสมการดังกล่าวก็ไม่ใช่สมการความน่าจะเป็นอีกต่อไปหรือกล่าวได้ว่า สมการนี้เป็นฟังก์ชันคณิตศาสตร์ของ Θ ซึ่งเรียกว่า ฟังก์ชันความเป็นไปได้ (likelihood function) แสดงได้ดังนี้

$$L(u_1, u_2, \ldots, u_n/\Theta) = \prod_{i=1}^{n} P_i^{u_i} Q_i^{(1-u_i)} \qquad (4)$$

เมื่อ $U_i = 1$ ค่าของ Q_i ก็ตัดทิ้งไปและเมื่อ $U_i = 0$ ค่าของ P_i ก็ตัดทิ้ง

การประมาณค่าจากสมการข้างต้นนี้ไม่สามารถเป็นไปได้ในทางปฏิบัติ ฟังก์ชันความเป็นไปได้สูงสุด $L(U/\Theta)$ หรือ $\ln L(U/\Theta)$ เมื่อ

$$\ln L(U/\Theta) = \sum_{i=1}^{n} [U_i \ln P_i + (1 - u_i) \ln Q_i] \qquad (5)$$

แล้ว $\frac{d}{d\Theta} \ln L(u/\Theta) = 0 \qquad (6)$

แต่สมการนี้ไม่เป็นเส้นตรงและไม่สามารถนำไปแก้ปัญหาได้อย่างชัดเจน จึงต้องใช้กระบวนการตัวเลข (numerical) ซึ่งได้แก่ กระบวนการนิวตัน ราฟสัน (Newton Raphson) ดังสมการต่อไปนี้

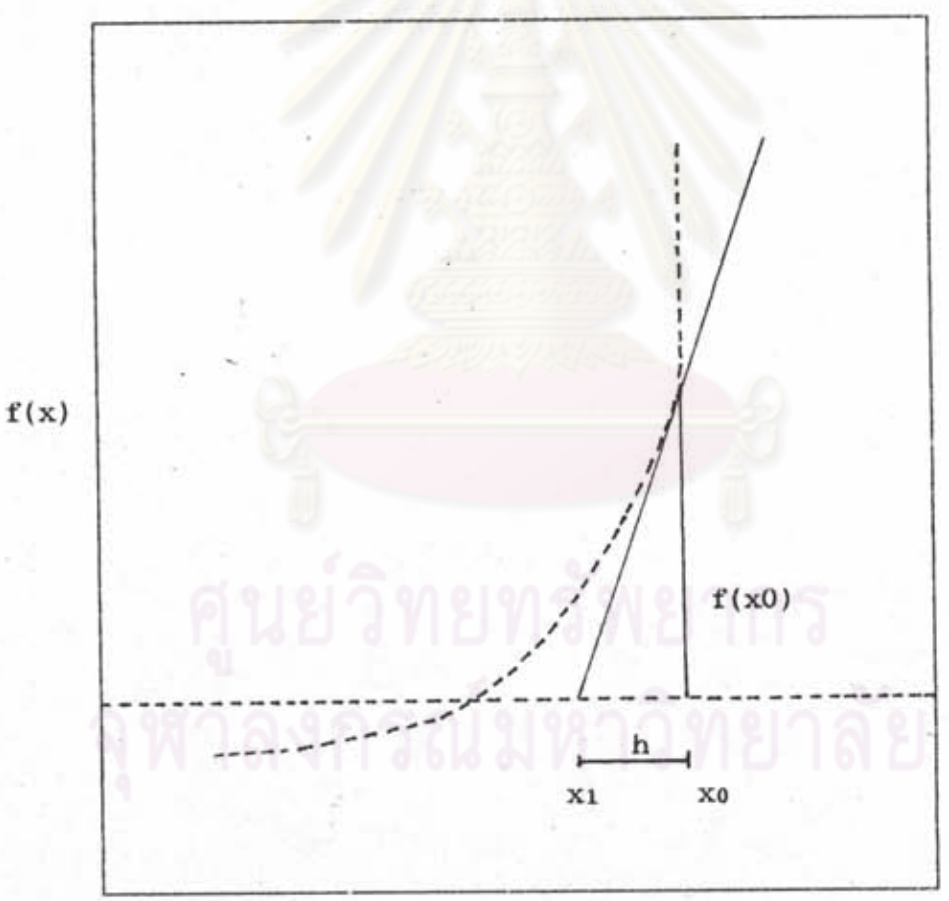
$$f(x) = 0$$

$$x_1 = x_0 - h$$

โดยที่ x_0 ผลลัพธ์เชิงประมาณของสมการ

x_1 ผลลัพธ์ที่คงที่มากกว่า

ภาพที่ 1 แสดงวิธีการนิวตัน ราฟสัน



จากภาพที่ 1 จะได้ว่า

$$h = \frac{f(x_0)}{\tan \alpha}$$

เนื่องจาก $\tan \alpha$ เป็นความชันของฟังก์ชัน $f(x)$ ที่ x_0 ดังนั้น $\tan \alpha = f'(x_0)$ ซึ่ง $f'(x_0)$ เป็นอนุพันธ์ของฟังก์ชัน x_0 ดังนั้น

$$h = \frac{f(x_0)}{f'(x_0)}$$

จะได้

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)} \qquad (7)$$

กระบวนการนี้เกิดขึ้นซ้ำ ๆ ถ้า x_m เป็นผลลัพธ์เชิงประมาณในขั้นที่ m แล้ว ผลลัพธ์ที่ปรับใหม่จะได้

$$x_{m+1} = x_m - \frac{f(x_m)}{f'(x_m)} \qquad (8)$$

กระบวนการนี้ทำซ้ำ ๆ จนกระทั่งความแตกต่างระหว่างการประมาณค่าขั้นที่ m และขั้นที่ m+1 แตกต่างกันเพียงเล็กน้อย กล่าวได้ว่าสิ้นสุดกระบวนการทำซ้ำ ผลลัพธ์จากสมการ $f(x) = 0$ จะเป็น x_m แสดงว่า กระบวนการนิวตัน ราฟสัน สิ้นสุดลงอย่างรวดเร็วตรงที่ $f'(x)$ ไม่เท่ากับ 0 นั่นคือ

$$f(x) = \frac{d}{d\theta} \ln L(U/\theta)$$

และ

$$f'(x) = \frac{d^2}{d\theta^2} \ln L(U/\theta)$$

ถ้าให้ θ_m เป็นการประมาณค่าขั้นที่ m จากสมการที่ 8 จะได้

$$\theta_{m+1} = \theta_m - \left[\frac{d}{d\theta} \ln L(U/\theta)\right]_m \Bigg/ \left[\frac{d^2}{d\theta^2} \ln L(U/\theta)\right]_m \qquad (9)$$

จากกระบวนการนิวตัว ราฟสัน ข้างต้นนี้นำมาใช้ในการประมาณค่าความสามารถแบบความเป็นไปได้สูงสุด เมื่อหาค่าพารามิเตอร์ของข้อสอบโดยกำหนดค่าความสามารถเริ่มต้นสำหรับผู้สอบคนที่ a เป็น

$$\theta_{oa} = \ln [\, r_a / (n - r_a) \,] \qquad (10)$$

เมื่อ r_a = จำนวนข้อสอบที่ผู้สอบคนที่ a ตอบถูก

n = จำนวนข้อสอบที่ผู้สอบคนที่ a ทำทั้งหมด

ln = natural logarithm

จะได้ค่า θ ของการทำซ้ำครั้งที่ m+1 นั่นคือ

$$\theta_{m+1} = \theta_m - h_m$$

โดยที่

$$h_m = \left[\frac{\partial \ln L}{\partial \theta_a}\right] \Bigg/ \left[\frac{\partial^2 \ln L}{\partial \theta_a{}^2}\right] \qquad (11)$$

เมื่อ $h_m = .001$ จึงสิ้นสุดกระบวนการทำซ้ำ

ค่าอนุพันธ์สำหรับรูปแบบการตอบข้อสอบที่มีจำนวนพารามิเตอร์ต่างกันสามารถแสดงได้ดังตารางที่ 1

ตารางที่ 1 ค่าอนุพันธ์ครั้งที่ 1,2 ของ Log-Likelihood Function สำหรับรูปแบบการตอบข้อสอบที่มีจำนวนพารามิเตอร์ต่างกัน (Hambleton and Swaminathan 1985:84)

ค่าอนุพันธ์	รูปแบบ 1 พารามิเตอร์	รูปแบบ 2 พารามิเตอร์	รูปแบบ 3 พารามิเตอร์
อันดับที่ 1 $\left(\frac{\partial \ln L}{\partial \theta a}\right)$	$D\sum_{i=1}^{n}(uia - Pia)$	$D\sum_{i=1}^{n} ai(uia-Pia)$	$D\sum_{i=1}^{n} a(uia-Pia)(Pia-ci)/Pia(1-ci)$
อันดับที่ 2 $\left(\frac{\partial^2 \ln L}{\partial \theta a^2}\right)$	$-D^2\sum_{i=1}^{n} Pia(1-Pia)$	$-D^2\sum_{i=1}^{n} ai^2(Pia(1-Pia)$	$D^2\sum_{i=1}^{n} ai^2(Pia-ci)(uiaci-Pia^2)$ $Qia/Pia^2(1-ci)^2$

จากตารางที่ 1 จะได้

$$\frac{\partial \ln L}{\partial \theta_a} = \sum_{i=1}^{n} k_i u_{ia} - \sum_{i=1}^{n} k_i P_{ia} = 0 \quad (a=1,\ldots,N) \tag{12}$$

ค่าของ k_i สำหรับรูปแบบไลจิสติกที่มี 1,2 และ 3 พารามิเตอร์ตามลำดับจะได้

$$k_i = D \tag{13}$$

$$k_i = D a_i \tag{14}$$

และ

$$k_i = D a_i (P_{ia}-c_i)/P_{ia}(1-c_i) \tag{15}$$

เนื่องจากค่าอำนาจจำแนกของข้อสอบ (a_i) มีค่าเป็นบวก และเนื่องจากในรูปแบบ 3 พารามิเตอร์ $P_{ia}-c_i$ ไม่เป็นลบ (เพราะ c_i เป็นขีดจำกัดล่าง) k_i เป็นจำนวนเชิงบวกสำหรับทั้ง 3 รูปแบบ ปรากฏว่า เมื่อผู้สอบตอบข้อสอบผิดทุกข้อ, $u_i = 0$ สำหรับ $i = 1,n$ ดังนั้น สมการความเป็นไปได้ในสมการความเป็นไปได้ในสมการที่ 12 จึงลดลงเป็น

$$\sum_{i=1}^{n} k_i P_{ia} = 0 \tag{16}$$

เนื่องจาก k_i มีค่าเป็นบวกและ P_{ia} คือ ความน่าจะเป็นของการตอบถูก สมการนี้จึงใช้ได้เมื่อ $\theta_a = -\infty$ เท่านั้น ในทำนองเดียวกัน เมื่อผู้สอบตอบข้อสอบถูกทุกข้อ, $U_{ia} = 1$ สำหรับข้อสอบทุกข้อและสมการความเป็นไปได้จึงลดลงเหลือ

$$\sum_{i=1}^{n} k_i = \sum_{i=1}^{n} k_i P_{ia} \tag{17}$$

และใช้ได้เมื่อ $\theta_a = +\infty$ จึงปรากฏว่า ตัวประมาณค่าแบบความเป็นไปได้สูงสุดจะใช้ไม่ได้ในกรณีเหล่านี้ ผู้สอบที่ได้คะแนนเต็มหรือศูนย์จะถูกกำจัดออกจากกระบวนการประมาณค่าแบบนี้ ซึ่งก็ยังไม่ทราบว่ามีผลกระทบของการกำจัดผู้สอบโดยใช้ฟังก์ชันความเป็นไปได้อย่างไร แต่ก็อาจเดาว่าคุณสมบัติของตัวประมาณค่าความเป็นไปได้สูงสุดใช้ไม่ได้ในสถานการณ์เหล่านี้

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ปัญหาต่อไปที่เกิดขึ้นในการใช้ตัวประมาณค่าความเป็นไปได้สูงสุดก็คือ ฟังก์ชันความเป็นไปได้อาจจะเป็นขีดสูงสุดในช่วง $-\infty < \theta < \infty$ ในกรณีนี้ค่าที่กำหนดโดยการแก้ปัญหาสมการ $\partial \ln L/\partial\theta = 0$ อาจจะไม่สอดคล้องกับค่าสูงสุดอย่างแท้จริง สิ่งนี้อาจจะเกิดขึ้นถ้าค่าของฟังก์ชันความเป็นไปได้ ณ $\theta = \pm\theta$ ใหญ่กว่าค่าสูงสุดที่พบในช่วง $-\infty < \theta < \infty$ ตัวประมาณค่าความเป็นไปได้สูงสุดก็ใช้ไม่ได้ในกรณีนี้และปัญหานี้จะเพิ่มขึ้นในกรณีที่ใช้รูปแบบ 3 พารามิเตอร์ ซึ่งซาเมจิมา (Samejima 1973a) กล่าวไว้เป็นคนแรก

ต่อไปนี้เป็นตัวอย่างของรูปแบบ 3 พารามิเตอร์ และเวคเตอร์ของการตอบของผู้สอบเป็น

$$u = [1\ 0\ 1\ 0\ 0]$$

และให้ค่าพารามิเตอร์ของผู้สอบเป็น

$$a = [2.0\ \ 1.0\ \ 2.5\ \ 1.5\ \ 2.5]$$

$$b = [0.0\ \ -.5\ \ 0.0\ \ -.5\ \ 0.5]$$

และ

$$b = [.25\ \ 0.0\ \ .25\ \ 0.0\ \ 0.1]$$

เมื่อ $\theta \longrightarrow \infty$, $P_{ia} \longrightarrow c_i$ และ $Q_{ia} \longrightarrow (1-c_i)$ ดังนั้น

$$\lim_{\theta \to -\infty} \ln L(u_1, u_2, \ldots, u_n/\theta) = \sum_{i=1}^{n} [u_i \ln c_i + (1-u_i)\ln(1-c_i)]$$

$$= \ln .25 + \ln 1.0 + \ln .25 + \ln 1.0 + \ln .9$$

$$= -2.88$$

ปัญหาของขีดสูงสุดเป็นปัญหาที่สำคัญมากที่จำเป็นจะต้องแก้ไขอย่างจริงจัง อย่างไรก็ตามก็ควรจะระบุว่า รูปแบบการตอบผิดปกติ และมีจำนวนข้อน้อยเกินไป ลอร์ด (Lord 1980 a;p.51 cited in Hambleton and Swaminathan 1985:87) ให้ข้อสังเกตว่า ปัญหานี้จะลดลงเมื่อใช้ข้อสอบมากกว่า 20 ข้อ

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

4. การทดสอบแบบ ซี เอ ที

การทดสอบแบบ ซี เอ ที (Computerized Adaptive Testing) หรือเรียกว่า การทดสอบแบบเทเลอร์ (Tailored Testing) เป็นการทดสอบที่พยายามคัดเลือกข้อสอบที่มีความยากให้เหมาะสมกับความสามารถของผู้สอบเป็นรายบุคคล (Hambleton and Swaminathan 1985:296-297) โดยใช้กลุ่มข้อสอบที่มีค่าพารามิเตอร์ ซึ่งได้แก่ ค่าความยาก ค่าอำนาจจำแนก และค่าการเดา ที่เหมาะสมจำนวนมาก เพื่อจะได้คัดเลือกข้อสอบตามกฎหรือกลยุทธ์ที่กำหนดไว้ก่อนการทดสอบ (Weiss 1974:1) ความจริงแล้ว ความคิดในเรื่องการทดสอบที่ปรับข้อสอบให้เหมาะสมกับความสามารถของผู้สอบ (tailoring) นั้นมีมานานแล้ว พบได้ในแบบสอบเชาว์ปัญญาของบิเนท์ และการทดสอบเป็นรายบุคคล เช่น มาตรวัดเชาว์ปัญญาของเวคสเลอร์ (Wechsler,1985) และแบบสอบ Peabody Picture Vocabulary Test (Dunn, 1965 cited in McCormick, D.J. 1978) แบบสอบเหล่านี้ให้ผู้สอบเริ่มทำข้อสอบตามระดับความยากที่ประมาณค่าจากช่วงอายุหรือตัวบ่งชี้อื่น ๆ ซึ่งมีข้อสอบจำนวนหนึ่งที่เหมาะสมกับช่วงอายุนั้น ๆ ถ้าผู้สอบตอบข้อสอบถูกต้องก็จะเริ่มทำข้อสอบข้อที่ยากขึ้น จนกระทั่งผู้สอบทำข้อสอบผิดหลาย ๆ ข้อ ก็สรุปว่าข้อสอบที่มีความยากน้อยกว่าข้อสอบที่ผู้สอบได้ทดสอบแล้ว ผู้สอบก็อาจจะสามารถทำได้ถูกต้อง

ความพยายามในการปรับข้อสอบให้เหมาะสมกับความสามารถของผู้สอบเริ่มเป็นระบบขึ้นครั้งแรกในปี ค.ศ. 1946 ในงานเขียนของคาวเด็น (Cowden อ้างถึงใน McCormick,D.J. 1978) คาวเด็นประยุกต์เทคนิคการวิเคราะห์ตามลำดับขั้น (sequential) ของวาลด์ (Wald 1947) ซึ่งเป็นเทคนิคที่นิยมใช้กันมากในการทดสอบแบบอิงเกณฑ์ที่ใช้ทดสอบในโปรแกรมช่วยสอนโดยใช้คอมพิวเตอร์ ต่อมาจึงมีการพัฒนารูปแบบการทดสอบแบบนี้ขึ้นอย่างกว้างขวางดังกลยุทธ์ที่จะกล่าวต่อไป

4.1 กลยุทธ์ที่ใช้ในการทดสอบแบบ ซี เอ ที

กลยุทธ์ที่ใช้ในการทดสอบแบบนี้มีหลายกลยุทธ์ด้วยกันตั้งแต่ง่ายที่สุดจนถึงวิธีการที่ยุ่งยากซับซ้อน โดยจะนำเสนอพอเป็นสังเขปดังนี้

1. กลยุทธ์สองขั้นตอน (Two-Stage Strategies) เป็นแบบที่ง่ายที่สุดที่จัดให้เหมาะสมกับผู้สอบรายบุคคล กลยุทธ์นี้ประกอบด้วยแบบสอบเพื่อแยกทาง (Routing Test) และแบบสอบวัด แบบสอบเพื่อแยกทางประกอบด้วยข้อสอบที่มีระดับความยากกระจายในช่วงกว้างหรืออาจจะประกอบด้วยข้อสอบที่มีระดับความยากปานกลาง เมื่อทำแบบสอบเพื่อแยกทางซึ่งมีจำนวนข้อไม่มากนักก็จะนำผลคะแนนที่ได้นั้นมาใช้เป็นเกณฑ์เพื่อแยกให้ผู้สอบไปทำแบบสอบวัดชุดใดชุดหนึ่ง โดยทั่วไปแบบสอบวัดจะประกอบด้วยแบบสอบย่อย จำนวน 3-5 ชุด ในแต่ละชุดจะมีค่าความยากเฉลี่ยเท่า ๆ กัน เรียงจากข้อสอบที่ง่าย ๆ ไปจนถึงชุดที่ยากที่สุด

กลยุทธการทดสอบสองขั้นตอนนี้ บางครั้งก็ใช้กลยุทธแยกทางสองครั้ง (Double Routing Two-Stage Strategies) บางครั้งเรียกว่า แบบสอบสามขั้นตอน (Three Stage Test)

2. กลยุทธแบบหลายขั้นตอน (Multistage Strategies) กลยุทธนี้ยังแบ่งออกเป็น

2.1 รูปแบบแยกทางคงที่ (Fixed Branching Model) กลยุทธนี้แตกต่างจากกลยุทธสองขั้นตอนคือ กลยุทธสองขั้นตอนใช้การตัดสินใจแยกทางเพียงครั้งเดียว นั่นคือ จากแบบสอบเพื่อแยกทางไปยังแบบสอบวัดเลย ส่วนกลยุทธหลายขั้นตอน การตัดสินใจแยกทางเกิดขึ้นหลังจากที่ผู้สอบตอบข้อสอบแต่ละข้อ ซึ่งข้อสอบนี้ใช้กลุ่มข้อสอบที่คำนวณค่าพารามิเตอร์ข้อสอบไว้แล้วและจัดโครงสร้างได้ก่อนแล้ว ซึ่งสามารถจัดเป็นโครงสร้างต่าง ๆ ได้หลายวิธีกลุ่มข้อสอบที่จัดโครงสร้างแตกต่างกันก็จะกำหนดเป็นกลยุทธการทดสอบที่ต่างกัน กลยุทธแยกทางคงที่นี้ใช้โครงสร้างกลุ่มข้อสอบเดียวกัน แต่ผู้สอบแต่ละคนจะได้รับข้อสอบต่างข้อกัน การทดสอบแบบนี้จะต้องมีกฎการแยกทางที่กำหนดไว้ล่วงหน้า เพื่อกำหนดให้ผู้สอบแต่ละคนทำข้อสอบจากขั้นหนึ่งไปยังอีกขั้นหนึ่ง ทั้งนี้กฎการแยกทางนี้ใช้สารสนเทศของผู้สอบในการตอบข้อสอบว่าถูกหรือผิดเพื่อกำหนดว่า ผู้สอบจะต้องทำข้อสอบข้อใดบ้างในโครงสร้าง กลยุทธหลายขั้นตอนนี้แบ่งออกเป็นรูปแบบปิระมิด (Pyramidal Model) รูปแบบเฟลกซิเลเวล (Flexilevel Model) และแบบสอบปรับระดับขั้น (Stradaptive Test)

2.1.1 รูปแบบปิระมิดหรือรูปแบบโครงสร้างต้นไม้ เป็นรูปแบบการทดสอบแบบ ซี เอ ที แบบแรกที่มีการวิจัยกันมากที่สุดและมีการศึกษาในหลาย ๆ ลักษณะ โดยการใช้ขนาดขั้นคงที่ (constant step size) ขนาดขั้นแปรผัน (variable step size) ปิระมิดข้างตัด (truncated pyramid) ปิระมิดที่มีหลายข้อในแต่ละขั้น (multiple pyramids) และรูปแบบปิระมิดที่ใช้ค่าน้ำหนักแก่ตัวเลือกที่ต่างกันเพื่อการแยกทาง (pyramidal using differential response option branching)

2.1.2 รูปแบบเฟลกซิเลเวล เป็นรูปแบบที่ลอร์ด (Lord อ้างถึงใน Weiss 1974:36) และแมคคอร์มิค (McCormick,D.J.,1978) เสนอขึ้นเพื่อใช้เป็นเทคนิคการทดสอบโดยใช้กระดาษในการปรับข้อสอบให้เหมาะสมกับระดับความสามารถของผู้สอบ แบบสอบเฟลกซิเลเวลต้องการกระดาษคำตอบที่ให้สารสนเทศแก่ผู้สอบว่าข้อสอบที่ทำผ่านมานั้นถูกหรือผิด จากสารสนเทศนี้ร่วมกับคำแนะนำในการทำแบบสอบ ผู้สอบก็จะทราบว่าจะทำข้อสอบข้อใดในแบบสอบ รูปแบบนี้แตกต่างจากรูปแบบปิระมิดคือ รูปแบบปิระมิดมีข้อสอบมากกว่า 1 ข้อ ในแต่ละช่วงความยาก ในขณะที่แบบสอบเฟลกซิเวลจะมีข้อสอบเพียงข้อเดียวในแต่ละช่วงความยาก เนื่องจากมีข้อสอบเพียงข้อเดียวในแต่ละช่วงความยาก กฎในการแยกทางจึงเน้นว่าถ้าผู้สอบตอบข้อสอบนั้น ๆ ถูกต้องก็จะได้ทำข้อสอบที่มีความยากเพิ่มขึ้น ถ้าตอบผิดก็จะได้ทำข้อสอบข้อที่ง่ายลงและข้อสอบในทั้งสอบกรณีเป็นข้อสอบที่ยังไม่ได้นำออกมาใช้ทดสอบกับผู้สอบจนนั้น ๆ ผู้สอบจะต้องปฏิบัติตามกฎเกณฑ์นี้จนกระทั่งการทดสอบจะจะเสร็จสิ้นลง เมื่อผู้สอบตอบข้อสอบไปจำนวนครั้งหนึ่งของข้อสอบทั้งหมดในแบบทดสอบ

2.1.3 รูปแบบปรับระดับชั้น เป็นรูปแบบที่ทดสอบโดยให้คอมพิวเตอร์ดำเนินการสอบ โดยเลือกข้อสอบจากกลุ่มข้อสอบที่จัดออกเป็นระดับชั้น (strata) ตามค่าความยากของข้อสอบ กลุ่มข้อสอบในรูปแบบนี้แบ่งออกเป็น 9 ระดับชั้น แต่ละระดับชั้นจะเพิ่มค่าความยากขึ้นตามลำดับ รูปแบบนี้แตกต่างจากรูปแบบแยกทางคงที่อื่น ๆ เนื่องจากใช้การประมาณระดับความสามารถของผู้สอบจากสารสนเทศของผู้สอบที่ได้มาจากการรายงานของตนเอง สารสนเทศกำหนดระดับในการเริ่มต้นทำข้อสอบ ซึ่งเป็นไปตามระดับของชั้นค่าความยากที่เพิ่มขึ้นจากสารสนเทศที่ได้ล่วงหน้านี้ทำให้ผู้สอบที่มีความสามารถต่ำก็จะเริ่มทำข้อสอบที่ง่าย ๆ ถ้าผู้สอบรายงานว่าตนมีความสามารถสูงก็จะได้เริ่มต้นทำข้อสอบที่มีความยากเพิ่มขึ้น

ลักษณะของรูปแบบปรับระดับชั้นนี้เด่นกว่ารูปแบบแยกทางคงที่อื่น ๆ ก็คือ ไม่จำกัดจำนวนข้อสอบที่ดำเนินการสอบกับผู้สอบแต่ละคนแต่จะขึ้นอยู่กับว่าผู้สอบบรรลุเกณฑ์การยุติการสอบได้เร็วเท่าใด จากการดำเนินการสอบโดยใช้ข้อสอบที่มีประสิทธิภาพก็จะบรรลุถึงระดับชั้นสูงสุด (ceiling strata) การทดสอบก็สิ้นสุดลง

2.3 รูปแบบแยกทางแปรผัน เป็นรูปแบบที่ดำเนินการสอบจากกลุ่มข้อสอบที่ไม่ได้จัดโครงสร้างไว้ก่อนและไม่ได้กำหนดข้อสอบไว้ล่วงหน้าว่าถ้าผู้สอบตอบถูกจะต้องไปทำข้อใดหรือถ้าตอบผิดจะต้องไปทำข้อสอบข้อใด ในรูปแบบแยกทางแปรผันนี้จะไม่มีขนาดของชั้นแต่จะดำเนินการสอบจากกลุ่มข้อสอบที่คำนวณค่าความยากและค่าอำนาจจำแนกของข้อสอบไว้แล้ว กฎการคัดเลือกข้อสอบข้อถัดไปเพื่อให้เหมาะกับความสามารถของผู้สอบนั้นจะใช้รูปแบบทางคณิตศาสตร์เข้ามาช่วย ได้แก่ รูปแบบเบย์และรูปแบบความเป็นไปได้สูงสุด

2.3.1 กลยุทธเบย์ (Bayesian Strategies) กลยุทธนี้ได้ประยุกต์หฤษฎีของเบย์ในกระบวนการตอบข้อสอบโดยใช้วิธีการทดสอบแบบ ซี เอ ที มีขั้นตอนดังต่อไปนี้

ก ประมาณค่าความสามารถของผู้สอบและความคลาดเคลื่อนมาตรฐานในการประมาณค่าไว้ก่อน ในแต่ละขั้นของการทดสอบจากสารสนเทศที่มีอยู่เกี่ยวกับผู้สอบ

ข คัดเลือกข้อสอบจากกลุ่มข้อสอบที่ได้คำนวณค่าพารามิเตอร์ของข้อสอบไว้แล้วข้อสอบในกลุ่มข้อสอบทุกข้อที่ยังไม่ได้นำมาให้ผู้สอบคนนั้น ๆ สอบก็จะถือว่าเป็นข้อสอบที่มีโอกาสที่จะนำมาใช้สอบได้ กระบวนการนี้แสดงให้เห็นว่าข้อสอบข้อใด ๆ ในกลุ่มข้อสอบที่จะนำมาใช้ในการทดสอบกับผู้สอบคนใด ๆ ก็ตาม จะเป็นข้อสอบที่ลดความไม่แน่นอนในการประมาณความสามารถของผู้สอบมากที่สุด ข้อสอบที่จะนำมาใช้ทดสอบกับผู้สอบจะเป็นข้อสอบที่มีระดับความยากใกล้กับระดับความสามารถของผู้สอบหลังจากดำเนินการสอบโดยใช้ข้อสอบที่คัดเลือกไว้ก็จะประมาณความสามารถก่อนและสารสนเทศที่ได้จากการตอบข้อสอบนั้น (เช่น การตอบถูกหรือการตอบผิด) นำมารวมกันโดยวิธีหฤษฎีเบย์ก็จะได้การประมาณความสามารถภายหลัง การประมาณค่าครั้งหลังนี้เป็นการประมาณค่าที่ปรับจากสิ่งที่รู้เกี่ยวกับข้อมูลที่อธิบายประชากรผู้สอบก็สามารถนำมาใช้เป็นสารสนเทศก่อนได้ กระบวนการนี้จะสิ้นสุดลงเมื่อความแปรปรวนภายหลังมีค่าน้อยกว่าที่กำหนดไว้ล่วงหน้า

2.3.2 กลยุทธความเป็นไปได้สูงสุด (Maximum Likelihood Estimation) ยูริ (Urry 1970 cited in Weiss 1975:62) ได้พัฒนาการทดสอบแบบ ซี เอ ที (Adaptive Testing) โดยใช้วิธีการความเป็นไปได้สูงสุดตามแนวทฤษฎีการตอบสนองข้อสอบ (IRT) กลยุทธความเป็นไปได้สูงสุดดำเนินการในแนวทางที่คล้ายคลึงกับกระบวนการเบย์ แม้ว่าเหตุผลทางคณิตศาสตร์จะต่างกัน หลังจากที่ผู้สอบตอบข้อสอบข้อหนึ่ง ๆ ถูกต้องและอีกข้อหนึ่งผิดก็จะสามารถนำไปแก้สมการความเป็นไปได้สูงสุดและจะได้ค่าการประมาณความสามารถและความคลาดเคลื่อนมาตรฐาน ข้อสอบข้อถัดไปที่เลือกมาใช้สอบเป็นข้อสอบจากกลุ่มข้อสอบเป็นข้อสอบที่มีค่าความยากเหมาะกับความสามารถของผู้สอบมากที่สุด เมื่อผู้สอบตอบข้อสอบข้อนั้นแล้วก็จะมีการประเมินคำตอบทันที จากบันทึกการตอบข้อสอบทั้งหมดของผู้สอบซึ่งรวมถึงข้อสอบข้อสุดท้ายที่ผู้สอบตอบ จากนั้นก็ประมาณค่าความสามารถใหม่และความคลาดเคลื่อนมาตรฐานใหม่โดยใช้สมการความเป็นไปได้สูงสุด ถ้าผู้สอบตอบข้อสอบข้อสุดท้ายผิด การประมาณความสามารถก็จะลดต่ำลง ในขั้นสุดท้ายของกระบวนการความคลาดเคลื่อนมาตรฐาน ในการประมาณค่าความสามารถลดลงเป็นผลจากสารสนเทศใหม่ที่ได้จากข้อสอบข้อสุดท้ายร่วมกับรูปแบบการตอบข้อสอบที่ทำผ่านมาของผู้สอบ

วิธีการนี้เริ่มจากการใช้กลุ่มข้อสอบเพียงบางส่วนก็สามารถประมาณค่าความสามารถเบื้องต้นแล้ว ก็มีแนวทางให้ผู้สอบแต่ละคนทำตามจนกระทั่งผู้สอบตอบข้อสอบข้อหนึ่งถูกและอีกข้อหนึ่งผิด เมื่อได้สารสนเทศนี้แล้วกระบวนการความเป็นไปได้สูงสุดก็จะประมาณค่าและกระบวนการแยกทางคงที่ก็จะไม่มีอีกต่อไป กระบวนการความเป็นไปได้สูงสุดนี้จะประมาณค่าความสามารถและความคลาดเคลื่อนตามการตอบข้อสอบแต่ละข้อ

4.2 องค์ประกอบของกระบวนการทดสอบแบบ ซี เอ ที

การทดสอบแบบ ซี เอ ที เป็นวิธีการทดสอบที่มีรูปแบบแตกต่างไปจากวิธีการทดสอบแบบประเพณีนิยม เนื่องจากการทดสอบแบบ ซี เอ ที จะต้องกำหนดองค์ประกอบต่าง ๆ ที่ใช้ในการดำเนินการสอบไว้ล่วงหน้าทั้งสิ้น ซึ่งองค์ประกอบที่กล่าวนี้ประกอบด้วย (Weiss 1984:362-364)

1. **รูปแบบการตอบข้อสอบ** (Item Response Models) ตามกรอบทฤษฎีการตอบสนองข้อสอบ (IRT) มีข้อตกลงเบื้องต้นว่า คุณลักษณะที่สอบวัดจะต้องเป็นมิติเดียวกันค่าพารามิเตอร์ของข้อสอบมีความคงที่ข้ามกลุ่มผู้สอบและมีความเป็นอิสระในการตอบข้อสอบ (Lord and Novick cited in Koch, B.R. and Reckase, M.D 1978:1) ดังนั้น การทดสอบแบบ ซี เอ ที จึงสามารถประยุกต์ใช้รูปแบบที่มี 1,2 หรือ 3 พารามิเตอร์ และอาจใช้รูปแบบนอร์มอล โอไจว์ (Normal Ogive Models) ก็ได้ การเลือกใช้รูปแบบใดก็ควรพิจารณาธรรมชาติของข้อสอบและต้องเหมาะสมกับข้อมูลการตอบข้อสอบด้วย แต่จากงานวิจัยของ แมคดินเลย์ และเรคเคส

(McKinley,R.L and Reckase, M.D. 1980) และคอคและเรคเคส (Koch, B.R. and Reckase, M.D. 1978) เปรียบเทียบการทดสอบแบบ ซี เอ ที โดยใช้รูปแบบการตอบข้อสอบที่มี 1 พารามิเตอร์และ 3 พารามิเตอร์ พบว่ารูปแบบที่มี 3 พารามิเตอร์มีความเหมาะสมมากกว่ารูปแบบที่มี 1 พารามิเตอร์

2. **กลุ่มข้อสอบ** (Item Pool) ในกลุ่มข้อสอบจะต้องประกอบด้วยข้อสอบที่ประมาณค่าตามหฤษฎีการตอบข้อสอบ พารามิเตอร์ของข้อสอบทุกข้อต้องได้รับการจัดให้อยู่ในมาตรวัดเดียวกัน โดยใช้กระบวนการที่เหมาะสม ข้อสอบในกลุ่มข้อสอบจะต้องครอบคลุมช่วงของระดับความสามารถในประชากรและวัดความสามารถมิติเดียวกัน ยูริ (Urry 1977:184) กล่าวว่ากลุ่มข้อสอบควรจะประกอบด้วยข้อสอบอย่างน้อย 100 ข้อ ค่าอำนาจจำแนกของข้อสอบเกิน .8 ความความยากกระจายอยู่ในช่วง -2.0 ถึง +2.0 และค่าการเดาต้องน้อยกว่า .3

3. **ระดับในการเริ่มต้น** (Entry Level) ในการทดสอบแบบ ซี เอ ที ผู้สอบแต่ละคนจะได้รับข้อสอบต่างกัน ดังนั้นจึงสามารถที่จะให้ผู้สอบเริ่มต้นทำข้อสอบที่มีระดับความยากที่ต่างกันได้ การเริ่มทำข้อสอบที่คลาดเคลื่อนไปจากความสามารถนั้น ไม่มีผลกระทบต่อการประมาณความสามารถเท่าใดนัก แต่ถ้าให้ผู้สอบเริ่มทำข้อสอบที่ตรงกับความสามารถของเขาก็จะทำให้ลดจำนวนข้อที่จะใช้ในการทดสอบลง

4. **กฎในการคัดเลือกข้อสอบ** กลยุทธของเบย์ซึ่งเป็นรูปแบบหนึ่งของการทดสอบแบบ ซี เอ ที นั้นจะใช้วิธีการประมาณค่าความสามารถแบบเบย์ และการคัดเลือกข้อสอบแบบเบย์ ซึ่งเป็นกระบวนการที่ลดความแปรปรวนภายหลังที่คาดของการประมาณค่าความสามารถเพราะสารสนเทศและความแปรปรวนภายหลังแบบเบย์เป็นฟังก์ชันที่เกี่ยวข้องกัน ส่วนกลยุทธแบบความเป็นไปได้สูงนั้นนิยมใช้เทคนิคการคัดเลือกข้อสอบดังนี้

4.1 วิธีการจับคู่ค่าความยาก (b_i) กับค่าความสามารถ (θ)

4.2 วิธีการจับคู่ค่า m_i กับค่าความสามารถ (θ) โดยใช้สูตร

$$m_i = b_i + \frac{1}{Da_i} \ln \left[\frac{1+\sqrt{1+8c_i}}{2} \right]$$ (Birnbuam, 1968)

$\ln(x)$ = natural logarithm of x

4.3 วิธีการคัดเลือกแบบสารสนเทศสูงสุด (Maximum information item selection)

5. **วิธีการให้คะแนนหรือการประมาณค่าความสามารถของแต่ละบุคคล** โดยใช้กลยุทธของเบย์และกลยุทธความเป็นไปได้สูงสุดนั้นในการประมาณค่าความสามารถแบบเบย์ (Bayesian Estimation) และการประมาณค่าความสามารถแบบความเป็นไปได้สูงสุด (Maximum Likelihood Estimation) วิธีเบย์ ให้การประมาณค่าความสามารถที่พิเศษกว่าการประมาณค่าความสามารถแบบความเป็นไปได้สูงสุด นั่นคือ สามารถประมาณค่าความสามารถได้จากการตอบข้อสอบเพียงข้อเดียวหรือชุดของข้อสอบที่ตอบถูกต้องทั้งหมดหรือผิดทั้งหมด ซึ่งวิธีการประมาณค่าแบบความเป็นไปได้สูงสุดไม่สามารถประมาณค่าได้ในสถานการณ์เช่นนั้น แต่อย่างไรก็ตามการประมาณค่าแบบความเป็นไปได้สูงสุดให้การประมาณค่าที่ไม่อคติ

6. **เกณฑ์การยุติการสอบ** (termainal criterion) ลักษณะสำคัญของการสอบแบบ ซี เอ ที ก็คือ การทดสอบจะดำเนินต่อไปตราบที่จำเป็นสำหรับผู้สอบแต่ละคน โดยทั่วไปแล้วจะใช้ค่าความคลาดเคลื่อนมาตรฐานเป็นเกณฑ์ในการยุติการสอบ และสามารถนำค่าความคลาดเคลื่อนมาตรฐานนี้มาคำนวณค่าความเที่ยงได้ด้วย ดังนั้น ในการวัดความสามารถของผู้สอบจะถือเกณฑ์ที่ตั้งไว้

4.3 การดำเนินการสอบ

ในด้านการดำเนินการสอบโดยวิธีการทดสอบแบบ ซี เอ ที เริ่มต้นโดยการให้ผู้สอบนั่งหน้าจอคอมพิวเตอร์ซึ่งอาจเป็นเครื่องไมโครคอมพิวเตอร์หรือเครื่องมินิคอมพิวเตอร์ ที่มีจานขับแม่เหล็ก (disk drive) ที่สามารถนำแผ่นบันทึกข้อมูล (diskette) ใส่เข้าหรือถอดออกมาได้ ในแผ่นบันทึกข้อมูลจะมีโปรแกรมการทดสอบที่สร้างขึ้นตามองค์ประกอบต่าง ๆ ของการทดสอบแบบ ซี เอ ที ที่กล่าวมาข้างต้นนั้ ผู้ดำเนินการสอบจัดการให้เครื่องอยู่ในสภาพที่พร้อมที่จะทดสอบ คอมพิวเตอร์จะให้คำแนะนำในการทำและรับข้อมูลส่วนตัวของนักเรียน เช่น ชื่อ - ชื่อสกุล เลขประจำตัว ชั้น/ห้อง เกรดเฉลี่ย เป็นต้น หลังจากนั้น คอมพิวเตอร์จะแสดงข้อสอบข้อแรกบนจอภาพเพื่อให้ผู้สอบได้ตอบ ข้อสอบข้อแรกนี้เป็นข้อสอบที่มีค่าความยากปานกลางหรือเป็นข้อสอบที่คัดเลือกตามกฎในการคัดเลือกข้อสอบก็ได้ ถ้าผู้สอบตอบข้อสอบข้อแรกถูกต้องข้อสอบข้อถัดไปก็จะเป็นข้อสอบที่มีความยากเพิ่มขึ้น ในทำนองเดียวกัน ถ้าผู้สอบตอบข้อสอบข้อแรกผิด ข้อสอบข้อถัดไปก็จะเป็นข้อสอบที่มีความยากลดลง จากนั้นก็จะประมาณค่าความสามารถการทดสอบจะดำเนินต่อไปตามแนวทางต่อไปนี้ คือ

1. หลังจากประมาณค่าความสามารถของผู้สอบแล้ว คอมพิวเตอร์ก็จะคัดเลือกข้อสอบข้อใหม่ที่มีประโยชน์ในการกำหนดความสามารถจริง

2. แสดงข้อสอบข้อใหม่บนจอภาพและผู้สอบตอบข้อสอบโดยการเลือกตัวเลือกที่ต้องการแล้วพิมพ์ตัวเลือกที่ต้องการ คอมพิวเตอร์จะรับคำตอบแล้วตรวจข้อสอบ

3. ประมาณค่าความสามารถของผู้สอบอีกครั้งหนึ่งตามการตอบข้อสอบข้อที่ผ่านมา

4. พิจารณาว่าผู้สอบควรจะทำข้อสอบข้อต่อไปหรือไม่ตามเกณฑ์การยุติการสอบที่กำหนดไว้

5. ถ้าผู้สอบจะต้องทำข้อสอบข้อต่อไปก็ย้อนกลับไปที่ข้อ 1 อีกครั้ง

เมื่อสิ้นสุดการทดสอบสารสนเทศเกี่ยวกับการตอบข้อสอบของผู้สอบก็จะถูกบันทึกไว้ในแผ่นบันทึกข้อมูลและแสดงผลการทดสอบให้ผู้สอบทราบ

4.4 ค่าความเที่ยงและความคลาดเคลื่อนของการวัด

ในเชิงทฤษฎีการทดสอบแบบประเพณีนิยมนั้น ความเที่ยงนิยามได้จากอัตราส่วนของความแปรปรวนของคะแนนที่สังเกตได้ในกลุ่มประชากรคน ซึ่งได้จากผู้สอบที่ถือว่าเป็นกลุ่มตัวอย่างที่สุ่มขึ้นมา ซึ่งสามารถแสดงได้ในรูปสหสัมพันธ์ร่วมดังนี้

$$\rho = \frac{\sigma^2 x - \sigma^2 e}{\sigma^2 x}$$

ซึ่ง $\sigma^2 x$ เป็นความแปรปรวนของคะแนนที่สังเกตได้
$\sigma^2 e$ เป็นความแปรปรวนคลาดเคลื่อนของการวัด
ρ เป็นค่าสหสัมพันธ์ที่ประมาณโดยตรงโดยใช้ดัชนีความเที่ยง เช่น แบบสอบคู่ขนานหรือการทดสอบซ้ำหรือโดยอ้อมโดยใช้ความเที่ยงแบบแบ่งครึ่ง

ความแปรปรวนคลาดเคลื่อนของการวัดกำหนดเป็น

$$\sigma e^2 = \sigma x^2 (1 - \rho)$$

ดังนั้น ความคลาดเคลื่อนมาตรฐานของการวัดจึงเท่ากับ

$$\sigma e = \sigma x \sqrt{1 - \rho}$$

แม้ว่า ความเที่ยงจะเป็นจำนวนที่ไม่มีหน่วยและสะดวกในการแปลความหมายเพราะมีค่าอยู่ระหว่าง 1 กับ 0 แต่ค่าความคลาดเคลื่อนมาตรฐานของการวัดจะมีประโยชน์ในการแปลผลคะแนนมากกว่า

สูตรดังกล่าวใช้กันอย่างแพร่หลายในทางปฏิบัติแต่สูตรเหล่านี้มีข้อตกลงเบื้องต้นที่ไม่ถูกต้อง กล่าวคือ ความแปรปรวนคลาดเคลื่อนของผู้สอบทุกคนเท่ากันซึ่งที่ถูกต้องแล้วความแปรปรวนคลาดเคลื่อนควรจะแปรเปลี่ยนไปตามความสามารถของแต่ละบุคคล ที่เป็นเช่นนี้เพราะสูตรในเชิง

I17611454

ทฤษฎีการทดสอบแบบประเพณีนิยมใช้ค่าเฉลี่ยสำหรับความแปรปรวนคลาดเคลื่อนเพียงค่าเดียว ดังนั้น สัมประสิทธิ์ความเที่ยงแบบประเพณีนิยมจึงให้คำอธิบายของสถานะที่แท้จริงของสิ่งใดสิ่งหนึ่งได้อย่างหยาบ ๆ เท่านั้น

ซาเมจิมา (Samejima 1977 cited in Green, et al.1984:352) ระบุว่า คำจำกัดความของค่าความเที่ยงมีความเกี่ยวข้องกับการวัดในแนวทฤษฎีการตอบสนองข้อสอบเพียงเล็กน้อย ซึ่งความแปรปรวนคลาดเคลื่อนกำหนดเป็นค่าฟังก์ชันของความสามารถ โดยใช้วิธีการประมาณค่าแบบความเป็นไปได้สูงสุด ความแปรปรวนของความคลาดเคลื่อนในการวัดถูกกำหนดเป็นสัดส่วนกลับของฟังก์ชันสารสนเทศ ในทฤษฎีของเบย์ ความแปรปรวนคลาดเคลื่อนเป็นค่าความแปรปรวนภายหลังของค่าความสามารถ

ความแปรปรวนคลาดเคลื่อนของการประมาณความสามารถเกี่ยวข้องโดยตรงกับค่าอำนาจจำแนกของข้อสอบที่ให้ผู้สอบแต่ละคนได้ทดสอบ การทดสอบแบบประเพณีนิยมสร้างขึ้นเพื่อที่จะจำแนกผู้สอบในกลุ่มความสามารถปานกลาง ดังนั้น จะให้ค่าความแปรปรวนคลาดเคลื่อนค่อนข้างต่ำในช่วงความสามารถปานกลางและจะมีความแปรปรวนคลาดเคลื่อนค่อนข้างสูงในช่วงความสามารถสูงมากหรือต่ำมาก ผลที่ได้จากทฤษฎีการตอบสนองข้อสอบนี้จะไม่สับสนกับค่าความคลาดเคลื่อนในการวัดสำหรับคะแนนจริง ในแนวทฤษฎีการทดสอบแบบประเพณีนิยม กล่าวคือ ในแนวทฤษฎีการทดสอบแบบประเพณีนิยมความคลาดเคลื่อนของคะแนนข้อที่ตอบถูกมีแนวโน้มต่ำมาก เมื่อผู้สอบที่ได้คะแนนเต็มในแบบสอบฉบับหนึ่งก็จะได้คะแนนเต็มในแบบสอบฉบับคู่ขนาน ซึ่งไม่ได้คำนึงถึงความเหมาะสมของข้อสอบกับความสามารถของผู้สอบ เช่น ให้ผู้ใหญ่มาทดสอบแบบสอบคำศัพท์ของนักเรียนชั้นประถมปีที่ 3 ก็จะได้คะแนนเต็ม คะแนนเหล่านั้นมาทดสอบซ้ำได้จะถือว่ามีความถูกต้องแม่นยำ ซึ่งในทางทฤษฎีการตอบสนองข้อสอบแล้วจะแปลผลในทางตรงข้ามคือวัดความสามารถได้ต่ำมาก ความคลาดเคลื่อนในการวัดในแนวทฤษฎีการตอบสนองข้อสอบเกี่ยวข้องกับความสามารถไม่ใช่จำนวนข้อที่ตอบถูก

ในการทดสอบแบบ ซี เอ ที ความแปรปรวนคลาดเคลื่อนในการวัดขึ้นอยู่กับกฎในการยุติการสอบ ซึ่งมีกฎการยุติการสอบอยู่ข้อหนึ่งที่เกี่ยวข้องโดยตรงกับความแปรปรวนคลาดเคลื่อน คือ ผู้สอบทุกคนจะรับการทดสอบโดยใช้ค่าความแปรปรวนคลาดเคลื่อนเดียวกัน จากกฎนี้จึงทำให้ค่าความคลาดเคลื่อนมาตรฐานของการวัดโดยประมาณมีค่าคงที่ ถ้ากลุ่มข้อสอบมีจำนวนข้อสอบไม่เพียงพอที่จะให้ผู้สอบได้ทดสอบจนถึงเกณฑ์ความแปรปรวนคลาดเคลื่อนที่กำหนด หรือถ้าใช้กฎในการยุติการสอบแบบอื่น ๆ เช่น กำหนดให้จำนวนข้อคงที่แล้วความแปรปรวนคลาดเคลื่อนก็จะขึ้นกับระดับความสามารถ (Θ) ดังนั้น ความคลาดเคลื่อนมาตรฐานของการวัดของคะแนนแบบสอบของผู้สอบแต่ละคนควรจะประเมินเป็นฟังก์ชันของคะแนนแบบสอบตามมาตรฐานของคะแนนที่รายงานไว้

ความเที่ยง บางครั้งอาจจะต้องใช้ดัชนีความเที่ยง แม้ว่า จะเป็นดัชนีความเที่ยงนี้เป็นสิ่งที่ประดิษฐ์ขึ้น นักวัดทางจิตจำนวนมากมีความเห็นว่า สัมประสิทธิ์ความเที่ยงในการทดสอบแบบ ซี เอ ที ไม่เหมาะสมและทำให้เข้าใจผิดได้ แต่ถ้าต้องการจะใช้ก็อาจกำหนดเป็นความเที่ยงอย่างมีเงื่อนไข ดังสูตร

$$\rho(\theta) = \frac{\sigma^2\theta - \sigma^2 e(\theta)}{\sigma^2\theta}$$

ซึ่ง $\sigma^2\theta$ เป็นความแปรปรวนของความสามารถที่สังเกตได้ วิธีการนี้จะเป็นความเที่ยงได้ถ้าผู้สอบทุก ๆ คนวัดโดยใช้ความมั่นใจในการวัดเดียวกัน ฟังก์ชันนี้เป็นสิ่งที่คล้ายกับฟังก์ชันสารสนเทศ

อีกวิธีหนึ่งที่เป็นไปได้ที่จะกำหนดความคลาดเคลื่อนของการวัดเฉลี่ย เมื่อประชากรซึ่งมีการกระจายความสามารถ $g(\theta)$ แล้ว

$$\sigma em^2 = \frac{\int_{-\infty}^{\infty} \sigma e^2(\theta)\, g(\theta) d\theta}{g(\theta) d\theta}$$

ซึ่งถ้าความสามารถกระจายเป็นปกติแล้วความเที่ยงก็กำหนดเป็น

$$\rho = \frac{\sigma^2\theta - \sigma^2 em}{\sigma^2\theta}$$

ยูริ (Urry 1977:183) ได้อธิบายว่า ความคลาดเคลื่อนมาตรฐานสัมพันธ์กับความเที่ยงของการทดสอบแบบ ซี เอ ที โดยตรง นั่นคือ สหสัมพันธ์กำลังสองระหว่างค่าประมาณความสามารถ, $\hat{\theta}$, และความสามารถจริง, θ, ความเที่ยงจึงกำหนดโดย

$$\rho^2 \hat{\theta}\theta = 1 - \sigma^2 e$$

ภายใต้เงื่อนไขที่ว่า การทดสอบแบบ ซี เอ ที สิ้นสุดสำหรับผู้สอบทุกคนเมื่อค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าที่กำหนดไว้ล่วงหน้า (σe หรือ SEE) ดังนั้น ค่าความเที่ยงนี้จึงสามารถควบคุมได้โดยการทดสอบสิ้นสุดลงก็ต่อเมื่อค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าของผู้สอบมีค่าเท่ากับค่าที่กำหนดไว้ล่วงหน้า ซึ่งค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถนั้นคำนวณได้จากสูตร

$$SEE = \frac{1}{\sqrt{I(\theta)}}$$

ซึ่ง $I(\theta)$ = ฟังก์ชันสารสนเทศแบบสอบ

4.5 ความตรง

กรีนและคณะ (Green,et al.1984:35) กล่าวถึงความตรงของการทดสอบแบบ ซี เอ ที ว่าเมื่อการทดสอบแบบ ซี เอ ที ได้รับการเสนอขึ้นมาให้เป็นแบบสอบคู่ขนานกันแบบสอบที่ใช้ทดสอบแบบประเพณีนิยมแล้วความตรงของแบบสอบปัจจุบันก็สามารถนำมาใช้เป็นความตรงของการทดสอบแบบ ซี เอ ที ได้เมื่อสร้างฟอร์มคู่ขนานขึ้นจริง ๆ แล้วคะแนนของทั้งสองฟอร์มควรจะสัมพันธ์กันได้ดีพอ ๆ กับค่าความเที่ยง กรีนและคณะยังกล่าวอีกว่าการเพิ่มความตรงอาจทำได้ใน 2 แนวทางคือ ใช้ข้อสอบที่มีค่าอำนาจจำแนกสูง ๆ จะทำให้การวัดแบบ ซี เอ ที บริสุทธิ์ขึ้นและถูกต้องขึ้นและอีกแนวทางหนึ่งก็คือการสร้างข้อสอบให้ประณีตละเอียดลออ

ในแนวทางของยูริ (Urry 1977:183) กล่าวถึงความตรงในเชิงการทดสอบแบบ ซี เอ ที ว่าสามารถควบคุมได้เช่นเดียวกับความเที่ยงโดยใช้ค่าความคลาดเคลื่อนมาตรฐานซึ่งเรียกว่าเป็นความตรงเชิงโครงสร้างซึ่งกำหนดได้จากค่าสหสัมพันธ์ระหว่างการประมาณค่าของคุณลักษณะแฝง, $\hat{\theta}$, และคุณลักษณะแฝงจริง, θ, ซึ่งความเที่ยงของการทดสอบแบบ ซี เอ ที มีค่าเท่ากับความตรงเชิงโครงสร้างยกกำลังสอง โดยใช้สูตร

$$\rho_{\hat{\theta}\theta} = \sqrt{1 - \sigma_e^2}$$

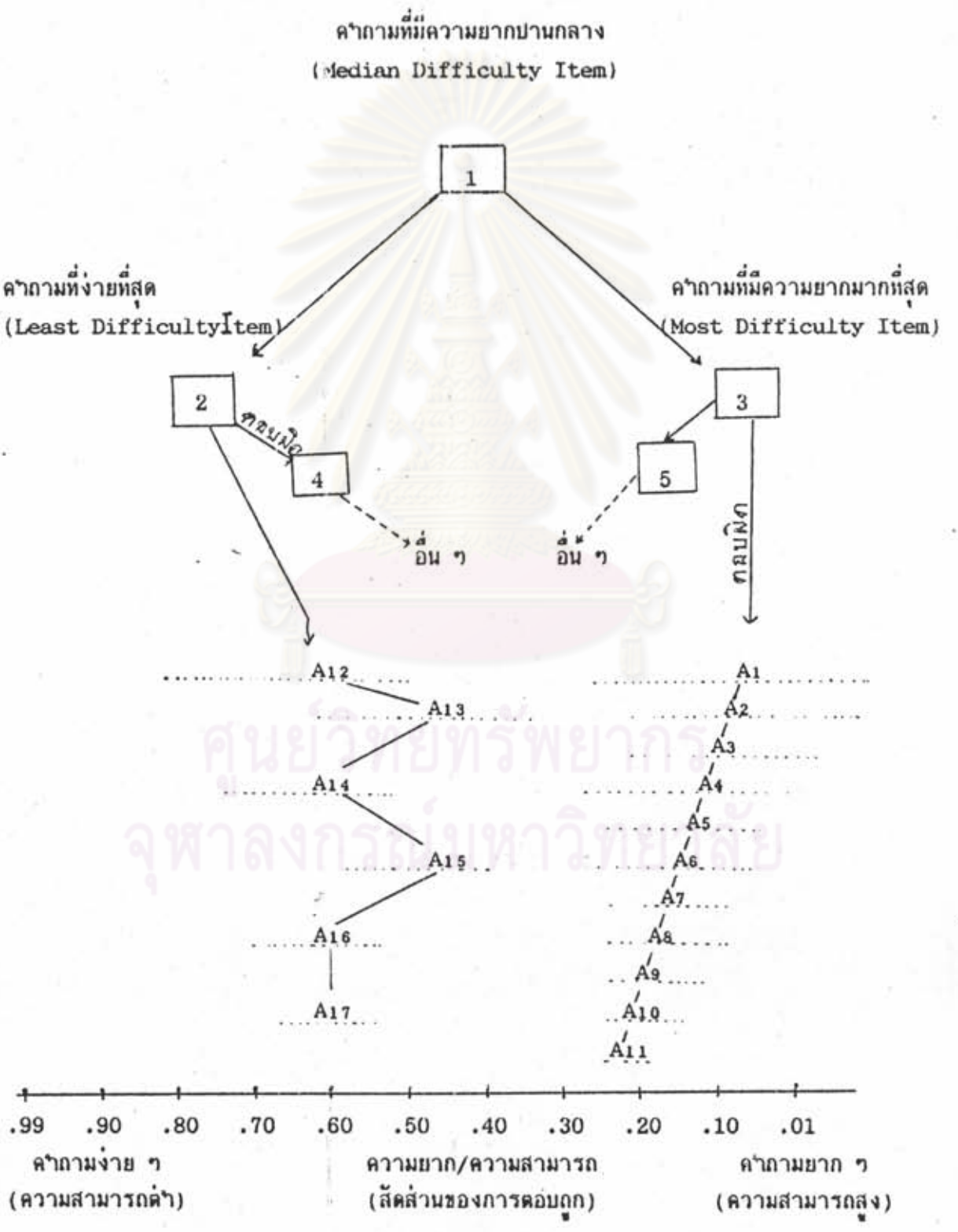
4.6 งานวิจัยที่เกี่ยวข้องกับการทดสอบแบบ ซี เอ ที

ในสหรัฐอเมริกาได้มีการศึกษาเกี่ยวกับการทดสอบแบบ ซี เอ ที ไว้มากพอควร แต่งานวิจัยส่วนมากจะเป็นงานวิจัยที่ศึกษาเกี่ยวกับการทดสอบโดยใช้รูปแบบแยกทางคงที่ เช่น การศึกษาคุณภาพของการทดสอบโดยใช้รูปแบบบิระมิดเปรียบเทียบกับการทดสอบแบบประเพณีนิยม (Larkin, K.C. and Weiss, D.J.1974) การศึกษาเปรียบเทียบเชิงประจักษ์ในการทดสอบโดยใช้รูปแบบสองขั้นตอนและรูปแบบบิระมิด (Larkin, K.C. and Wesiss, D.J. 1975) การศึกษาเชิงประจักษ์ของการทดสอบวัดความสามารถแบบสองขั้นตอน (ERIC Jan-Dec, 1974) การศึกษาในสถานการณ์จำลองในการทดสอบความสามารถโดยใช้แบบสอบปรับระดับขั้น (ERIC Jul-Dec,1976) การศึกษาในสถานการณ์จำลองและการศึกษาเชิงประจักษ์ในการทดสอบความสามารถโดยใช้รูปแบบเฟล็กซิเลเวล (ERIC Jul-Dec,1976) การศึกษาเชิงประจักษ์ในการทดสอบความสามารถโดยใช้รูปแบบบิระมิดในการดำเนินการสอบโดยใช้คอมพิวเตอร์ (ERIC Jan-June,1975) เป็นต้น การทดสอบโดยทดสอบกับคอมพิวเตอร์โดยใช้รูปแบบแยกทางแปรผันซึ่งได้แก่ กลยุทธเบย์และกลยุทธความเป็นไปได้สูงสุดมักจะเป็นการศึกษาในสถานการณ์จำลองและที่จะกล่าวต่อไปนี้เป็นการศึกษาการทดสอบแบบ ซี เอ ที โดยใช้กลยุทธความเป็นไปได้สูงสุดในสถานการณ์จำลอง

ยูริ (Urry 1970 cited in Weiss 1975:62) ได้พัฒนาการทดสอบแบบ ซี เอ ที โดยให้ข้อสอบข้อแรกแก่ผู้สอบทุกคนได้ทดสอบเป็นข้อสอบที่มีความยากปานกลาง ถ้าผู้สอบตอบข้อสอบข้อแรกถูกก็จะได้รับข้อสอบที่มีความยากมากที่สุดในกลุ่มข้อสอบ แต่ถ้าผู้สอบตอบข้อสอบข้อแรกผิดก็จะได้รับข้อสอบที่ง่ายที่สุดในกลุ่มข้อสอบ การตอบถูกหรือผิดนี้จำเป็นต้องใช้เพื่อการแก้สมการความเป็นไปได้สูงสุด (Maximum Likelihood Estimation) การตอบข้อสอบข้อแรกถูกต้องแสดงว่าความสามารถของผู้สอบสูงกว่าความสามารถปานกลาง แต่ยังไม่สามารถบอกได้อย่างแน่นอนว่าความสามารถของเขาเท่ากับเท่าไรจากการตอบข้อสอบเพียงข้อเดียว ในทำนองเดียวกัน ถ้าผู้สอบตอบข้อสอบข้อแรกผิดก็แสดงว่าความสามารถของผู้สอบต่ำกว่าความสามารถปานกลางแต่ยังบอกไม่ได้แน่นอนเช่นกันว่าความสามารถของเขาเท่ากับเท่าไรจากการตอบข้อสอบเพียงข้อเดียว วิธีการทดสอบของยูริแสดงในภาพที่ 2

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ภาพที่ 2
บันทึกการตอบในเชิงทฤษฎีจากยุทธวิธีการทดสอบแบบ ซี เอ ที โดยใช้การประมาณค่าแบบความเป็นไปได้สูงสุดของยูริ
(Hypothetical Response Records from Urry's Maximum Likelihood Adaptive Testing Strategy)



จากภาพ 2 ถ้าผู้สอบตอบข้อสอบข้อที่ 3 ผิด ก็สามารถคำนวณค่าความสามารถเบื้องต้นของผู้สอบโดยใช้กระบวนการความเป็นไปได้สูงสุดได้เท่ากับ A1 ค่าความสามารถโดยประมาณ (A1) อยู่ระหว่างข้อสอบข้อที่ยากที่สุดและข้อสอบที่มีความยากปานกลางยังมีความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถอยู่มาก ในทำนองเดียวกัน ถ้าผู้สอบตอบข้อสอบข้อที่ 1 ผิด ก็จะได้รับข้อสอบข้อที่ง่ายที่สุดในกลุ่มข้อสอบ ถ้าผู้สอบตอบข้อสอบข้อที่ 2 ได้ถูกต้องก็จะประมาณค่าความสามารถเท่ากับ (A12) และความคลาดเคลื่อนมาตรฐานในการประมาณค่าโดยใช้กระบวนการสารสนเทศสูงสุด

จากภาพแสดงให้เห็นว่าข้อสอบที่นำไปใช้สอบกับผู้สอบคนที่ A1 ไม่ใช่ข้อสอบข้อเดียวกับที่ทดสอบกับผู้สอบที่มีความสามารถเท่ากับ A12 ข้อสอบที่ผู้สอบได้รับจะเป็นข้อสอบที่มีความยากใกล้เคียงกับระดับความสามารถ การตอบข้อสอบถูกจะทำให้ค่าความสามารถที่ประมาณใหม่สูงขึ้น(เช่น A13) และความคลาดเคลื่อนที่เล็กลง ในขณะที่การตอบผิดจะให้การประมาณความสามารถที่ต่ำลง(เช่น A14) และความคลาดเคลื่อนก็ลดลงด้วย กระบวนการนี้จะสิ้นสุดลงเมื่อระดับความสามารถของผู้สอบมีค่าใกล้เคียงกับความสามารถจริงของผู้สอบและค่าความคลาดเคลื่อนมาตรฐานโดยที่มีความคลาดเคลื่อนมาตรฐานในการประมาณค่าเท่ากับค่าที่กำหนดไว้ล่วงหน้า

เรคเคส (Reckase cited in Weiss 1975:64) ได้พัฒนากลยุทธความเป็นไปได้สูงสุดโดยวิธีการที่ต่างจากของยูริคือ ยูริให้ผู้สอบทุกคนเริ่มทำข้อสอบข้อแรกเป็นข้อสอบที่มีความยากปานกลาง แต่เรคเคสให้ผู้สอบเริ่มทำข้อสอบข้อแรกเป็นข้อสอบที่มีความยากแตกต่างกันตามสารสนเทศเกี่ยวกับความสามารถของผู้สอบ เช่น ถ้าทราบว่าผู้สอบคนนี้เก่งก็อาจให้ผู้สอบทำข้อสอบที่มีความยากค่อนข้างสูงให้ใกล้เคียงกับความสามารถของเขา

เรคเคสยังใช้กระบวนการที่ต่างจากของยูริ เพื่อให้ได้สารสนเทศที่จำเป็นในการเริ่มกระบวนการประมาณค่าความสามารถแบบความเป็นไปได้สูงสุด เนื่องจากจะไม่สามารถประมาณค่าความสามารถได้ถ้าผู้สอบตอบข้อสอบถูกหมดหรือผิดทั้งหมด ดังนั้น เรคเคสจึงใช้วิธีการดังนี้ คือ ถ้าผู้สอบตอบข้อสอบข้อที่ 1 ถูกก็จะได้รับข้อสอบข้อที่มีความยากเป็น 2 เท่าของข้อแรกในขั้นที่ 2 ถ้าตอบข้อสอบข้อที่ 2 ถูกต้องก็จะได้รับข้อสอบข้อที่ 3 ที่มีความยากเป็น 2 เท่าของข้อที่ 2 จนกระทั่งผู้สอบตอบข้อสอบผิด ในทำนองเดียวกัน ถ้าผู้สอบตอบข้อสอบข้อแรกผิดก็จะได้รับข้อสอบข้อที่ง่ายลงครึ่งหนึ่งของข้อเดิม การทดสอบจะดำเนินต่อไปจนกระทั่งผู้สอบตอบข้อสอบได้ถูกต้อง จึงมีการประมาณค่าความสามารถโดยใช้กระบวนการความเป็นไปได้สูงสุด

เนื่องจากในการประมาณค่าแบบความเป็นไปได้สูงสุดนี้ได้จากการตอบของผู้สอบ แนวทางของเรคเคสจึงต่างจากของยูริเพียงเล็กน้อย แต่จุดประสงค์ของทั้งสองแนวทางก็เพื่อที่จะค้นหาข้อสอบที่ยังคงเหลืออยู่ในกลุ่มข้อสอบที่มีความยากใกล้เคียงกับระดับความสามารถของผู้สอบมาให้เขาได้ทดสอบ อย่างไรก็ตาม ในกลุ่มข้อสอบที่จำกัดโดยใช้ข้อสอบจริง ๆ มักจะไม่มีข้อสอบที่มีค่าความยากเท่า ๆ กับ

ระดับความสามารถของผู้สอบที่ประมาณได้ ยูริแก้ปัญหานี้โดยการใช้ข้อสอบที่มีค่าความยากใกล้เคียงกับค่าความยากในจินตนาการ นั่นคือ ข้อสอบที่ได้รับเลือกอาจเป็นข้อสอบที่ง่ายกว่าที่ต้องการเล็กน้อยหรือยากกว่าที่ต้องการเล็กน้อย แต่เรคเคสมักจะเลือกข้อสอบที่ง่ายกว่าข้อสอบในจินตนาการเล็กน้อยแม้ว่าอาจจะมีข้อสอบที่ยากกว่าข้อสอบในจินตนาการ กระบวนการนี้อาจจะใช้ได้เพื่อที่จะลดผลการเดา

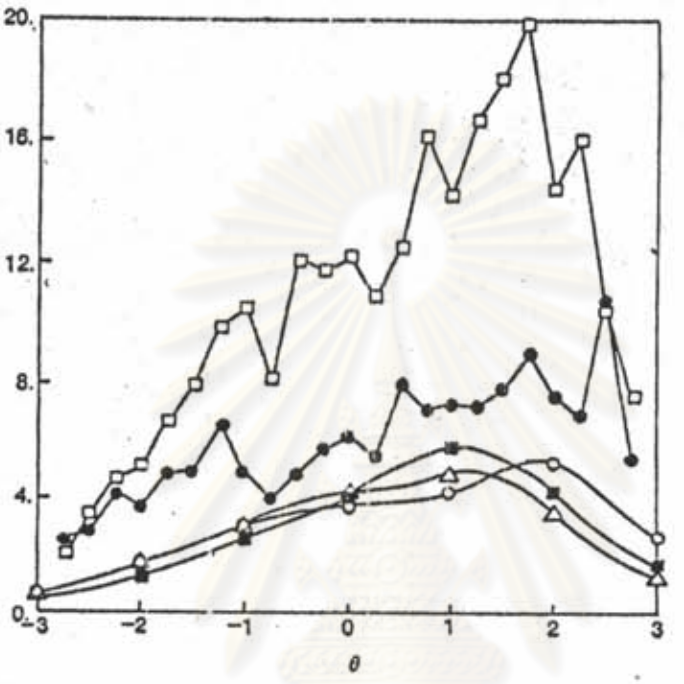
เนื่องจากกลยุทธของเรคเคสสามารถใช้ประโยชน์จากสารสนเทศเบื้องต้น ซึ่งอาจจะทำให้มีผลต่อการประหยัดข้อสอบได้บางข้อเมื่อเปรียบเทียบกับกลยุทธของยูริ กล่าวอีกนัยหนึ่งกระบวนการของยูริมักจะเริ่มกระบวนการประมาณค่าได้เร็วกว่าของเรคเคส ซึ่งเป็นผลให้ประหยัดเวลาในการทดสอบขณะที่ กระบวนการของยูริถ้าใช้ข้อสอบจริง ๆ อาจจะเป็นผลให้เกิดการเดาข้อสอบที่ยากมากกว่าที่ต้องการ กระบวนการของเรคเคสอาจส่งผลในการดำเนินการสอบโดยใช้ชุดข้อสอบที่ให้สารสนเทศเฉลี่ยต่อข้อสอบต่ำกว่าของยูริ ผลก็คืออาจทำให้ใช้ข้อสอบที่ดำเนินการสอบในแนวทางของเรคเคสมากกว่ากระบวนการของยูริ

ฮูลิน ดราสโกว์และพาร์สัน (Hulin, C.L., Drasgow, F. and Parsons, C.K. 1983:222-232) ได้ทำการประเมินกระบวนการทดสอบแบบ ซี เอ ที ภายใต้สถานการณ์ การทดสอบจำลองโดยใช้เครื่องคอมพิวเตอร์ดำเนินการสอบโดยใช้ข้อสอบที่มีสารสนเทศสูงสุดที่ระดับความสามารถ (θ) เท่ากับ 0 ถ้าผู้สอบตอบข้อสอบข้อแรกถูกก็จะแสดงข้อสอบที่มีสารสนเทศสูงสุดที่ระดับความสามารถ (θ) เท่ากับ 1 และถ้าตอบถูกอีกก็จะแสดงข้อสอบที่มีสารสนเทศสูงสุดที่ระดับความสามารถ (θ) เท่ากับ 2 ถ้าตอบถูกอีกก็จะแสดงข้อสอบที่มีสารสนเทศที่ระดับความสามารถ (θ) เท่ากับ 3 ในทำนองเดียวกัน ถ้าผู้สอบตอบข้อสอบข้อแรกที่มีสารสนเทศสูงสุดที่ระดับความสามารถ (θ) เท่ากับ 0 ผิด ก็จะได้รับข้อสอบที่มีสารสนเทศสูงสุดที่ระดับความสามารถ (θ) เท่ากับ -1 ถ้าตอบผิดอีกก็จะได้รับข้อสอบที่มีสารสนเทศสูงสุดที่ระดับความสามารถ (θ) เท่ากับ -2 ถ้าตอบผิดอีกก็จะได้รับข้อสอบที่มีสารสนเทศสูงสุดที่ระดับความสามารถ (θ) เท่ากับ -3 ตามลำดับ

เมื่อผู้สอบตอบข้อสอบถูกหรือผิดอย่างน้อย 1 ข้อ ก็จะประมาณค่าความสามารถโดยใช้กระบวนการความเป็นไปได้สูงสุด ดังนั้นจะต้องมีการตอบข้อสอบ 2 ข้อขึ้นไปก่อน จึงจะประมาณค่าความสามารถได้และใช้วิธีการเลือกข้อสอบโดยการใช้สารสนเทศสูงสุด การแสดงข้อสอบและการประมาณค่าความสามารถใหม่จะดำเนินไปจนกระทั่งครบ 25 ข้อ

จากสถานการณ์จำลองโดยใช้ผู้สอบ 200 คน ที่มีระดับความสามารถ (θ) เท่ากับ 0 พบว่าฟังก์ชันสารสนเทศข้อสอบที่ประมาณได้นั้นเพิ่มขึ้นเกือบเป็นเส้นตรงเช่นเดียวกับจำนวนของข้อสอบที่เพิ่มเข้าไปและพบว่า การลดความคลาดเคลื่อนของ $1/\sqrt{I(\theta)}$ ของความสามารถที่ประมาณได้ค่อนข้างเร็วในตอนเริ่มการทดสอบแต่จะลดลงค่อนข้างช้าในตอนท้ายของการทดสอบดังภาพที่ 3

ภาพที่ 3 ค่าฟังก์ชันสารสนเทศโดยประมาณของการทดสอบแบบ ซี เอ ที และการทดสอบแบบประเพณีนิยม เมื่อปรับความยากของข้อสอบให้เท่ากับ 25 ข้อ



การคัดเลือกข้อสอบโดยวิธีการสารสนเทศสูงสุด

การคัดเลือกข้อสอบโดยวิธีการจับคู่ค่า m_i กับ Θ

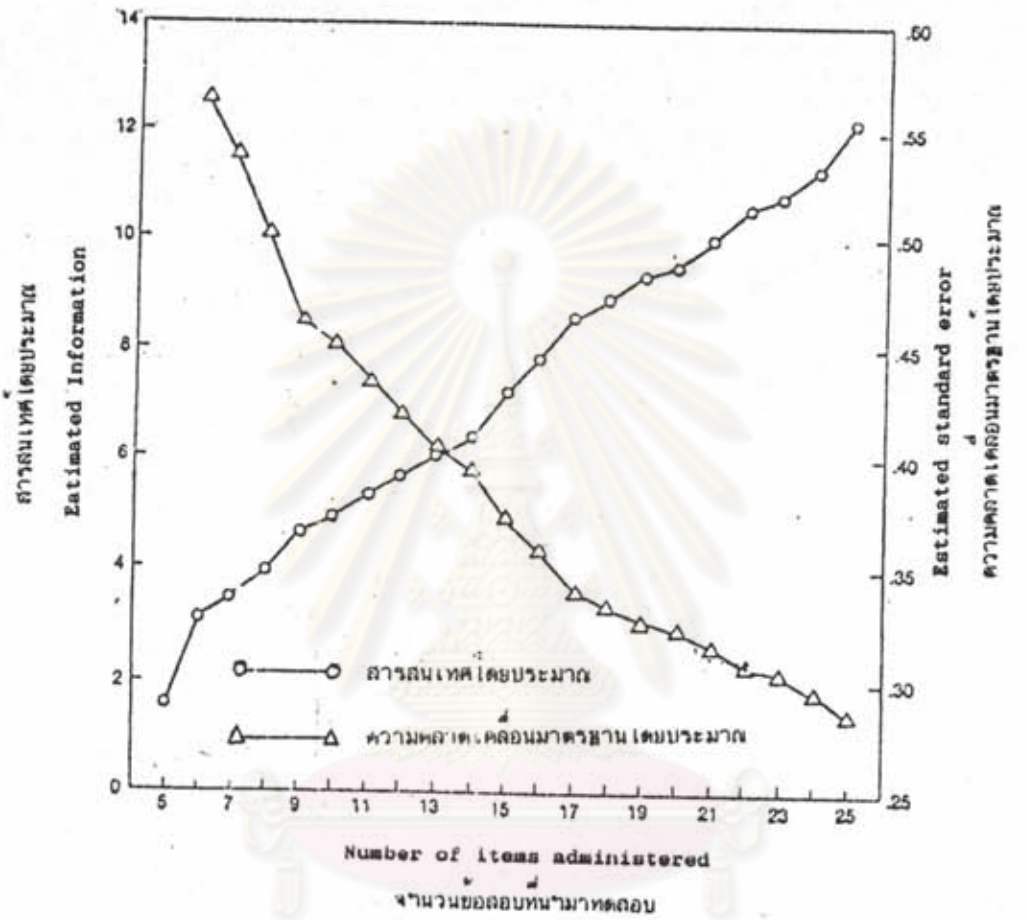
แบบสอบ SAT - V ของเลวินและดราสโกว์ (Levine and Drasgow)

แบบสอบ SAT - V ของลอร์ด (Lord)

แบบสอบ GRE - V ของดราสโกว์ (Drasgow)

จากภาพที่ 3 ค่าฟังก์ชันสารสนเทศแบบสอบที่ประมาณได้ของการทดสอบแบบ ซี เอ ที จำนวน 25 ข้อ ฟังก์ชันสารสนเทศของแบบสอบ Gre-V และ SAT-V จำนวน 2 ชุด หลังจากปรับความยาวร่วมกันจำนวน 25 ข้อ ดังภาพนั้น พบว่า การทดสอบแบบ ซี เอ ที ให้การวัดที่เด่นกว่าการทดสอบแบบประเพณีนิยมจำนวน 3 ฉบับและพบว่าผู้สอบที่มีความสามารถต่ำเท่ากับ -1.75 ในการทดสอบแบบ ซี เอ ที จะมีค่าความสามารถที่ประมาณได้ถูกต้องมากกว่าค่าความสามารถในช่วงใด ๆ ในการทดสอบแบบประเพณีนิยม

ตารางที่ 2 สารสนเทศโดยประมาณและความคลาดเคลื่อนมาตรฐานของ θ ของการทดสอบแบบ ซี เอ ที โดยใช้การคัดเลือกข้อสอบแบบสารสนเทศสูงสุด เมื่อ $\theta = 0$



จากตารางที่ 2 พบปัญหาเกี่ยวกับวิธีการใช้สารสนเทศสูงสุดของข้อสอบเป็นเกณฑ์ในการเลือกข้อสอบ จากการเปรียบเทียบการกระจายความถี่ของจำนวนครั้งที่ข้อสอบแต่ละข้อที่นำออกมาใช้ดำเนินการสอบระหว่างการทดสอบแบบ ซี เอ ที โดยใช้สารสนเทศสูงสุดของข้อสอบเป็นเกณฑ์ในการคัดเลือกและการทดสอบแบบประเพณีนิยม พบว่า วิธีการคัดเลือกข้อสอบโดยใช้สารสนเทศสูงสุดนั้นจะมีข้อสอบจำนวน 141 ข้อที่ไม่ได้ถูกนำออกมาใช้ทดสอบกับผู้สอบคนใดเลย ในทางปฏิบัติอาจทำให้เกิดปัญหาคือ การทุจริตในการสอบ สาเหตุก็คือ วิธีการคัดเลือกข้อสอบโดยใช้สารสนเทศสูงสุดของข้อสอบนั้นจะใช้ข้อสอบที่มีสารสนเทศสูงสุดจำนวนไม่กี่ข้อจากกลุ่มข้อสอบ จึงทำให้เกิดปัญหาการจดจำข้อสอบเพียงไม่กี่ข้อก็จะได้คะแนนสูง

ผู้วิจัยจึงดำเนินการวิจัยในสถานการณ์จำลองโดยใช้เกณฑ์การคัดเลือกข้อสอบโดยการจับคู่ค่า m_i กับค่าความสามารถ (Θ) เพื่อแก้ปัญหาการใช้ข้อสอบบางข้อมากเกินไปซึ่งอาจช่วยแก้ปัญหาการทุจริต

ในสถานการณ์จำลองนี้ข้อสอบข้อเริ่มต้นนั้นได้มาจากการสุ่มข้อสอบ 30 ข้อจากกลุ่มข้อสอบที่มีค่า m_i ในช่วง (-.20,.21) ถ้าตอบถูกก็จะสุ่มข้อสอบจากกลุ่มข้อสอบ 30 ข้อที่มีค่า m_i อยู่ในช่วง (.86,1.23) ถ้าตอบผิดก็จะสุ่มข้อสอบจากข้อสอบ 30 ข้อที่มีค่า m_i อยู่ในช่วง (-1.22,-.76) ถ้าตอบถูกอีกครั้งก็จะสุ่มข้อสอบที่อยู่ในช่วงที่มีค่า m_i เพิ่มขึ้น ถ้าตอบผิดอีกครั้งก็จะสุ่มข้อสอบจากข้อสอบที่อยู่ในช่วงที่มีค่า m_i ลดลง จนกระทั่งมีการตอบถูก 1 ข้อหรือผิด 1 ข้อก็จะประมาณค่าความสามารถโดยใช้วิธีความเป็นไปได้สูงสุดและใช้เกณฑ์ในการเลือกข้อสอบโดยการจับคู่ m_i กับ Θ การนำข้อสอบมาให้ผู้สอบสอบ การประมาณค่าความสามารถการเลือกข้อสอบจะดำเนินต่อไปจนกระทั่งครบ 25 ข้อ

ฟังก์ชันสารสนเทศแบบสอบโดยใช้วิธีการจับคู่ m_i กับ Θ นี้พบว่า มีการสูญเสียสารสนเทศไปเมื่อใช้วิธีการจับคู่ค่า m_i กับ Θ แทนวิธีการสารสนเทศสูงสุด อย่างไรก็ตามวิธีการนี้ก็ให้สารสนเทศเพิ่มขึ้นมากกว่าวิธีการทดสอบแบบประเพณีนิยมในระดับความสามารถตั้งแต่ -1.25 ลงมา พบว่าแบบสอบ ซี เอ ที ให้สารสนเทศมากเป็น 2 เท่าของแบบสอบที่ทดสอบแบบประเพณีนิยม

ผลการวิจัยที่นำเสนอไปแล้วข้างต้นนี้เป็นการศึกษาในสถานการณ์จำลองโดยใช้เครื่องคอมพิวเตอร์ ซึ่งการศึกษาในลักษณะนี้สามารถกำหนดค่าพารามิเตอร์ของข้อสอบได้แน่นอนและคะแนนผลการสอบก็ไม่มีผลกระทบจากผู้ดำเนินการสอบ ผู้สอบ วิธีดำเนินการสอบ ตลอดจนสภาพแวดล้อมในขณะที่ทำการทดสอบ ดังนั้น การศึกษาในลักษณะดังกล่าวจึงมีข้อจำกัดในการสรุปครอบคลุมไปถึงกรณีทั่ว ๆ ไปนอกเหนือจากข้อจำกัดในสถานการณ์จำลองดังกล่าวแล้ว การวิจัยเหล่านี้ยังมีความผันแปรในเรื่องจำนวนข้อสอบ ขนาดของช่วงห่างระหว่างค่าความยากของข้อสอบและเกณฑ์ที่ใช้ในการประเมินประสิทธิภาพของการทดสอบ จากข้อจำกัดเหล่านี้ทำให้ได้ข้อสรุปว่า ยังมีความต้องการการศึกษาประสิทธิภาพของการทดสอบแบบ ซี เอ ที อยู่มาก โดยเฉพาะอย่างยิ่งในสถานการณ์จริงและการศึกษาเกี่ยวกับการทดสอบแบบ ซี เอ ที ในประเทศไทยมีการศึกษาเฉพาะรูปแบบแยกทางคงที่ยังไม่มีงานวิจัยใด ๆ ที่ศึกษาเกี่ยวกับรูปแบบแยกทางแปรผัน เนื่องจากเป็นรูปแบบที่จำเป็นต้องใช้เครื่องคอมพิวเตอร์ช่วยในด้านการดำเนินการสอบและการคำนวณค่าความสามารถขณะทำการทดสอบ ซึ่งการคำนวณค่าความสามารถนี้จำเป็นต้องใช้การคำนวณทางคณิตศาสตร์ที่ซับซ้อนเข้ามาช่วย จึงเป็นเรื่องใหม่ที่น่าสนใจที่จะศึกษาประสิทธิภาพของการทดสอบแบบ ซี เอ ที โดยใช้รูปแบบแยกทางแปรผัน ซึ่งผลการวิจัยครั้งนี้อาจทำให้เกิดแนวทางในการนำรูปแบบการทดสอบ ซี เอ ที มาใช้ในการทดสอบในวงการศึกษาของไทยได้กว้างขวางยิ่งขึ้น