

บทที่ 2

วรรณคดีที่เกี่ยวข้อง

ในบทนี้จะได้กล่าวถึงเอกสารและงานวิจัยที่เกี่ยวข้อง โดยผู้วิจัยได้ศึกษาเนื้อหาต่างๆดังจะเสนอตามลำดับ ดังต่อไปนี้

1. การทดสอบด้วยข้อสอบเฉพาะตัว
 - 1.1 หลักการทดสอบด้วยข้อสอบเฉพาะตัว
 - 1.2 ประเภทของการทดสอบด้วยข้อสอบเฉพาะตัว
 - 1.3 การประมาณค่าความสามารถของการทดสอบด้วยข้อสอบเฉพาะตัว
2. แบบสอบตามระดับความสามารถ
 - 2.1 หลักการเกี่ยวกับแบบสอบตามระดับความสามารถ
 - 2.2 การดำเนินการสอบแบบสอบตามระดับความสามารถ
 - 2.3 การให้คะแนนแบบสอบตามระดับความสามารถ
3. ทฤษฎีการตอบสนองข้อสอบ
 - 3.1 หลักการเกี่ยวกับทฤษฎีการตอบสนองข้อสอบ
 - 3.2 ข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบ
 - 3.3 การตรวจสอบคุณภาพของแบบสอบ
 - 3.4 การนำทฤษฎีการตอบสนองข้อสอบไปประยุกต์ใช้
4. งานวิจัยที่เกี่ยวข้อง
 - 4.1 งานวิจัยต่างประเทศ
 - 4.2 งานวิจัยในประเทศ

1. การทดสอบด้วยข้อสอบเฉพาะตัว (Tailored testing)

1.1 หลักการทดสอบด้วยข้อสอบเฉพาะตัว

การทดสอบด้วยข้อสอบเฉพาะตัว เป็นการทดสอบที่พยายามจะปรับระดับความยากของข้อสอบให้เหมาะสมกับระดับความสามารถของผู้สอบแต่ละคน (Lord 1971) โดยที่ผู้สอบแต่ละคนไม่จำเป็นต้องตอบข้อสอบเหมือนกันทุกข้อ ทั้งนี้ขึ้นอยู่กับผลการตอบข้อสอบในแบบสอบของแต่ละคน (Weiss and Kingbury 1984) การทดสอบด้วยข้อสอบเฉพาะตัว จะประมาณระดับความสามารถของผู้สอบตลอดระยะเวลาการสอบ ซึ่งการประมาณความสามารถของผู้สอบนี้จะพิจารณาจากคำตอบของผู้สอบที่เพิ่งตอบผ่านไป โดยทั่วไปจะเริ่มการทดสอบด้วยข้อสอบที่มีความยากปานกลาง แล้วพิจารณาผลการตอบข้อสอบ เมื่อผู้สอบตอบข้อสอบผิดจะได้ไปตอบข้อสอบที่ง่ายขึ้นเป็นข้อต่อไป และเมื่อผู้สอบตอบข้อสอบถูก จะได้ไปตอบข้อสอบที่ยากขึ้นเป็นข้อต่อไป การทดสอบจะกระทำเช่นนี้ไปจนสิ้นสุดการทดสอบ (Green and Others 1984) ซึ่งวิธีการจัดข้อสอบให้เหมาะสมกับผู้สอบ จะแตกต่างกันตามประเภทของการทดสอบด้วยข้อสอบเฉพาะตัว

1.2 ประเภทของการทดสอบด้วยข้อสอบเฉพาะตัว

การทดสอบด้วยข้อสอบเฉพาะตัว ประกอบด้วยวิธีการทดสอบ 2 ประเภทใหญ่ๆ ดังนี้

1. ยุทธวิธีสองขั้นตอน (Two-Stage Strategies)
2. ยุทธวิธีหลายขั้นตอน (Multi-Stage Strategies)
 - 2.1 แบบจำลองแยกทางคงที่ (Fixed-Branching Model)
 - 2.1.1 แบบจำลองรูปปิรามิด (Pyramidal Model)
 - 2.1.2 แบบสอบตามระดับความสามารถ (Flexilevel Test)
 - 2.1.3 แบบสอบปรับระดับขึ้น (Stradaptive Test)
 - 2.2 แบบจำลองแยกทางแปรผัน (Variable-Branching Model)
 - 2.2.1 ยุทธวิธีของเบย์ (Bayesian Strategies)
 - 2.2.2 ยุทธวิธีความเป็นไปได้สูงสุด (Maximum Likelihood

Strategies)

ซึ่งจะนำเสนอพลเป็นสังเขปดังนี้

1. **ยุทธวิธีสองขั้นตอน (Two-Stage Strategies)** เป็นแบบที่ง่ายที่สุดที่จัดให้เหมาะสมกับผู้สอบเป็นรายบุคคล ยุทธวิธีนี้ประกอบด้วยแบบสอบ 2 ฉบับ แบบสอบฉบับแรก เรียกว่า แบบสอบเพื่อแยกทาง (Routing Test) จะประกอบด้วยข้อสอบที่มีระดับความยากกระจายในช่วงกว้าง หรืออาจจะประกอบด้วยข้อสอบที่มีระดับความยากปานกลาง โดยปกติจะมีประมาณ 10 ข้อ ซึ่งผู้สอบทุกคนต้องสอบเหมือนกัน เพื่อนำผลการสอบมาเป็นเกณฑ์ในการแยกผู้สอบไปสอบแบบสอบฉบับที่สอง ที่เรียกว่า แบบสอบวัด (Main or Measurement Test) ซึ่งแบบสอบวัดจะประกอบด้วยแบบสอบย่อย จำนวน 3-5 ฉบับ แยกตามระดับความยาก โดยที่แบบสอบย่อยแต่ละฉบับจะมีค่าความยากเฉลี่ยเท่าๆกัน และมีจำนวนข้อสอบประมาณฉบับละ 20-30 ข้อ ผู้สอบแต่ละคนจะได้สอบแบบสอบชุดที่สองต่างกันตามความสามารถที่ประเมินได้จากแบบสอบฉบับแรก ถ้าผู้สอบได้รับการประเมินว่ามีความสามารถสูงจะได้สอบแบบสอบย่อยฉบับที่ยาก ถ้าได้รับการประเมินว่ามีความสามารถปานกลางจะได้สอบแบบสอบย่อยฉบับที่ยากปานกลาง และถ้าได้รับการประเมินว่ามีความสามารถต่ำจะได้สอบแบบสอบย่อยฉบับที่ง่าย

ยุทธวิธีสองขั้นตอนนี้ บางครั้งใช้วิธีการแยกทางสองครั้ง (Double Routing Two-Stage Strategies) บางครั้งเรียกว่าแบบสอบสามขั้นตอน (Three-Stage Test)

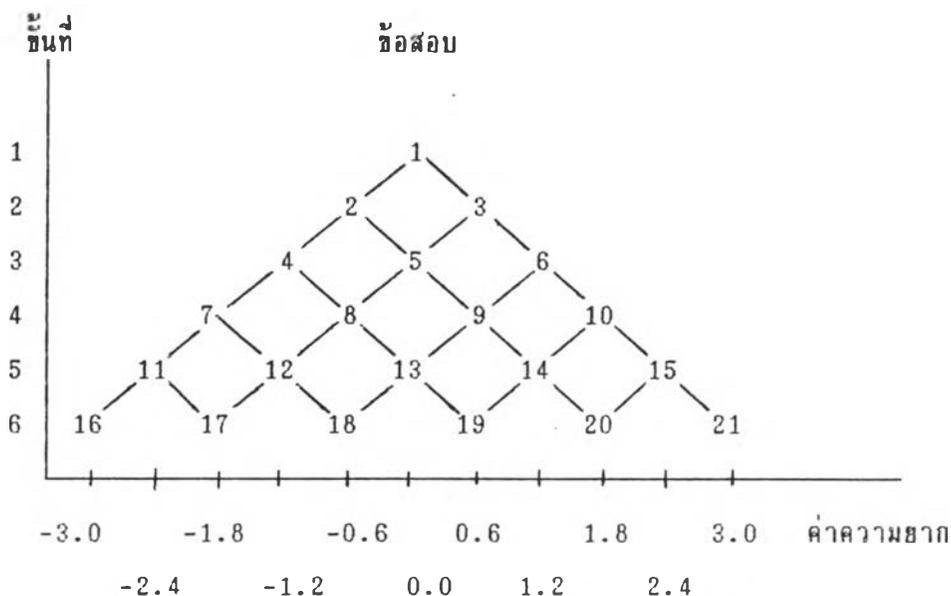
2. ยุทธวิธีหลายขั้นตอน (Multi-Stage Strategies)

2.1 ยุทธวิธีหลายขั้นตอน (Multi-Stage Strategies) แบบแยกทางคงที่ (Fixed-Branching Model) หมายถึงการทดสอบด้วยข้อสอบเฉพาะตัวที่มีหลายขั้นตอน ในแต่ละขั้นตอนอาจจะมีข้อสอบเพียงข้อเดียวหรือหลายข้อก็ได้ โดยมีรูปแบบหรือเส้นทางในการตอบข้อสอบที่แน่นอน ประกอบด้วย 3 รูปแบบ ดังนี้

2.1.1 รูปแบบปิรามิด (Pyramidal Model) เป็นรูปแบบการทดสอบด้วยข้อสอบเฉพาะตัว ที่จัดเป็นรูปสามเหลี่ยม ซึ่งการทดสอบรูปแบบนี้มีการศึกษาในหลายๆลักษณะ โดยการใช้ขนาดขั้นคงที่ (Constant Step Size) ขนาดขั้นแปรผัน (Variable Step Size) ปิรามิดข้างตัด (Truncated Pyramid) ปิรามิดที่มีหลายข้อในแต่ละขั้น (Multiple Pyramids) และรูปแบบปิรามิดที่ใช้ค่าน้ำหนักแก่ตัวเลือกที่ต่างกันเพื่อการแยกทาง (Pyramidal Using Differential Response Option Branching) ซึ่งในที่นี้จะขอกล่าวถึงการ
ใช้ขนาดขั้นคงที่

รูปแบบปิรามิดขนาดขั้นคงที่ (Constant Step Size Pyramidal) เป็นการจัดข้อสอบเป็นรูปสามเหลี่ยม โดยมีจำนวนข้อสอบในแต่ละขั้นเท่ากับลำดับที่ของขั้นนั้น กล่าว

คือชั้นที่ 1 จะมีข้อสอบหนึ่งข้อ ชั้นที่ 2 จะมีข้อสอบสองข้อ ในชั้นที่ 6 จะมีข้อสอบหกข้อ ดังนั้น การทดสอบรูปปรางค์ที่มีหกชั้น จะมีข้อสอบทั้งหมด 21 ข้อ ดังแสดงในภาพที่ 1



ภาพที่ 1 ตัวอย่างการจัดโครงสร้างข้อสอบรูปแบบปรางค์ขนาดชั้นคงที่ 6 ชั้น

จากภาพที่ 1 แกนนอนแสดงถึงค่าความยากของข้อสอบ มีค่าอยู่ระหว่าง -3.0 ถึง 3.0 โดยแบ่งออกเป็น 10 ช่วงเท่าๆกัน ฉะนั้นค่าความยากจะห่างกันช่วงละ 0.6 ข้อสอบที่อยู่ในแนวตั้งแนวเดียวกัน จะมีค่าความยากเท่ากันและช่วงห่างค่าความยากของข้อที่อยู่ติดกันภายในชั้นที่มีค่าเท่ากันตลอดคือ 1.2

ในการทำแบบสอบผู้สอบจะต้องทำข้อสอบชั้นละหนึ่งข้อ โดยเริ่มทำข้อที่อยู่บนยอดสามเหลี่ยมซึ่งมีความยากปานกลางก่อน ถ้าทำถูกขั้นต่อไปจะแยกไปทำข้อที่ยากขึ้น แต่ถ้าทำผิดขั้นต่อไปจะไปทำข้อที่ง่ายขึ้น กระบวนการทดสอบจะเป็นเช่นนี้จนถึงชั้นสุดท้าย

2.1.2 แบบสอบตามระดับความสามารถ (Flexilevel Test) เป็นรูปแบบของการทดสอบด้วยข้อสอบเฉพาะตัว ที่ลอร์ด คิดขึ้นเพื่อใช้กับการทดสอบด้วยการเขียนตอบ แบบสอบชนิดนี้คัดแปลงมาจากรูปแบบปรางค์ โดยที่โครงสร้างของรูปแบบปรางค์ มีข้อสอบในแต่ละระดับความยากมากกว่า 1 ข้อ ทั้งนี้ยกเว้นข้อที่มีค่าความยากสูงสุดและต่ำสุด แต่โครงสร้างแบบสอบตามระดับความสามารถจะมีข้อสอบเพียงข้อเดียวในแต่ละระดับความยาก ดังนั้นการใช้กฎการแยกทางจึงแตกต่างกัน วิธีการของแบบสอบชนิดนี้ คือ แบ่งกลุ่มข้อสอบออกเป็น 2 กลุ่ม คือข้อสอบกลุ่มง่าย กับข้อสอบกลุ่มยาก และคัดเลือกข้อสอบที่มีความ

ยากอยู่กึ่งกลางระหว่างข้อสอบกลุ่มง่ายกับกลุ่มยาก ไว้ต่างหากอีกข้อหนึ่ง เพื่อเป็นข้อเริ่มต้นสำหรับผู้สอบทุกคน ภายในข้อสอบกลุ่มง่ายจะเรียงลำดับข้อสอบจากข้อที่ยากที่สุดในกลุ่มนี้ไปหาข้อที่ง่ายที่สุด ข้อสอบกลุ่มนี้จะพิมพ์ด้วยหมึกสีแดง ส่วนข้อสอบในกลุ่มยากจะเรียงลำดับจากข้อที่ง่ายที่สุดในกลุ่มนี้ไปหาข้อที่ยากที่สุด และจะพิมพ์ด้วยหมึกสีน้ำเงิน แสดงดังภาพที่ 2

ข้อ 0 (มีความยากอยู่กึ่งกลางระหว่างข้อสอบกลุ่มง่ายกับกลุ่มยาก) (ข้อสอบกลุ่มง่าย)	(ข้อสอบกลุ่มยาก)
ข้อ 1. (ข้อที่ยากที่สุดในกลุ่มนี้)	ข้อ 1. (ข้อที่ง่ายที่สุดในกลุ่มนี้)
.	.
ข้อ 37. (ข้อที่ง่ายที่สุดในกลุ่มนี้)	ข้อ 37. (ข้อที่ยากที่สุดในกลุ่มนี้)

ภาพที่ 2 การจัดข้อสอบในแบบสอบตามระดับความสามารถ (Lord 1980 : 114)

สำหรับกระดาษคำตอบก็ได้รับการออกแบบมาพิเศษ คือ เมื่อผู้สอบตอบถูกจุดสีน้ำเงินก็จะปรากฏขึ้น ข้อต่อไปผู้สอบจะต้องไปเลือกทำข้อสอบในกลุ่มยาก ในข้อที่มีหมายเลขค่าสถิติยังไม่ได้ทำ และถ้าตอบผิดจุดสีแดงก็จะปรากฏขึ้น ผู้สอบจะต้องไปเลือกทำข้อสอบในกลุ่มง่าย ในข้อที่มีหมายเลขค่าสถิติยังไม่ได้ทำ และจะทำตามกฎไปจนสิ้นสุดการทดสอบ การทดสอบจะสิ้นสุดลงเมื่อผู้สอบทำข้อสอบไปเป็นจำนวน $(n+1)/2$ ข้อ เมื่อ n คือจำนวนข้อสอบทั้งหมดในแบบสอบ (Lord 1980 : 114)

2.1.3 แบบสอบปรับระดับขึ้น (Stradaptive Test) เป็นรูปแบบของการทดสอบด้วยข้อสอบเฉพาะตัวที่ใช้คอมพิวเตอร์ในการดำเนินการสอบ แบบสอบชนิดนี้จะคัดเลือกข้อสอบจากกลุ่มข้อสอบออกเป็นกลุ่มย่อยๆตามระดับความยากของข้อสอบ โดยกลุ่มย่อยแต่ละกลุ่มเรียกว่า ระดับชั้น (Stratum) ในแต่ละระดับชั้นจะประกอบด้วยข้อสอบที่มีระดับความยากใกล้เคียงกัน และความยากของแต่ละระดับชั้นจะเพิ่มขึ้นตามลำดับ รูปแบบนี้แตกต่างจากรูปแบบแยกทางคงที่อื่นๆ เนื่องจากใช้การประมาณระดับความสามารถของผู้สอบจากรายงานของผู้สอบเอง ซึ่งจะเป็นตัวกำหนดระดับเริ่มต้นในการทำข้อสอบ ถ้าผู้สอบรายงานระดับความสามารถต่ำก็จะเริ่มทำข้อสอบในระดับชั้นที่มีข้อสอบง่าย ถ้าผู้สอบรายงานระดับความสามารถสูง ก็จะเริ่มต้นทำข้อสอบที่ระดับชั้นที่มีข้อสอบยาก

2.2 แบบจำลองแยกทางแปรผัน (Variable-Branching Model) เป็นรูปแบบที่ไม่ได้กำหนดโครงสร้างและข้อสอบไว้ล่วงหน้า ว่าถ้าผู้สอบตอบถูกหรือผิดจะต้องไปทำข้อใด ในรูปแบบแยกทางแปรผันนี้ จะไม่มีขนาดของชั้น แต่จะดำเนินการสอบจากกลุ่มข้อสอบที่คำนวณค่าความยากและค่าอำนาจจำแนกของข้อสอบไว้แล้ว กฎการคัดเลือกข้อสอบข้อถัดไปเพื่อให้เหมาะสมกับความสามารถของผู้สอบนั้นจะใช้รูปแบบทางคณิตศาสตร์เข้ามาช่วย ได้แก่ รูปแบบเบย์ และรูปแบบความเป็นไปได้สูงสุด

2.2.1 ยุทธวิธีของเบย์ (Bayesian Strategies) เป็นยุทธวิธีที่ได้จากการประยุกต์ทฤษฎีของเบย์ในกระบวนการตอบข้อสอบ และใช้คอมพิวเตอร์ในการทดสอบ โดยมีขั้นตอนดังนี้

ก. ประมวลค่าความสามารถของผู้สอบและความคลาดเคลื่อนมาตรฐานในการประมวลค่าไว้ก่อน ในแต่ละชั้นของการทดสอบจากสารสนเทศที่มีอยู่เกี่ยวกับตัวผู้สอบ

ข. คัดเลือกข้อสอบจากกลุ่มข้อสอบที่ได้คำนวณค่าพารามิเตอร์ของข้อสอบไว้แล้ว ข้อสอบทุกข้อที่ยังไม่ได้นำมาให้ผู้สอบคนนั้นๆ สอบ ก็จะได้ว่าเป็นข้อสอบที่มีโอกาสนำมาใช้ได้ กระบวนการนี้แสดงให้เห็นว่าข้อสอบข้อใดๆ ในกลุ่มข้อสอบที่นำมาใช้ในการทดสอบกับผู้สอบคนใดก็ตาม จะเป็นข้อสอบที่ลดความไม่แน่นอนในการประมวลความสามารถของผู้สอบมากที่สุด ข้อสอบที่นำมาใช้กับผู้สอบจะเป็นข้อสอบที่มีระดับความยากใกล้เคียงกับระดับความสามารถของผู้สอบ หลังจากดำเนินการสอบโดยใช้ข้อสอบที่คัดเลือกไว้ก็จะประมวลความสามารถก่อน และสารสนเทศที่ได้จากการตอบข้อสอบนั้น เช่น การตอบถูก หรือการตอบผิด นำมารวมกัน โดยวิธีทฤษฎีเบย์ ก็จะได้การประมวลค่าที่ปรับจากสิ่งที่เกี่ยวข้องกับข้อมูลที่อธิบายประชากรผู้สอบก็สามารถนำมาใช้เป็นสารสนเทศได้ กระบวนการนี้จะสิ้นสุดลงเมื่อความแปรปรวนภายหลังมีค่าน้อยกว่ามากกำหนดไว้ล่วงหน้า

2.2.2 ยุทธวิธีความเป็นไปได้สูงสุด (Maximum Likelihood Strategies) เป็นยุทธวิธีที่ดำเนินการในแนวทางคล้ายคลึงกับกระบวนการของเบย์ แม้ว่าเหตุผลจะต่างกัน หลังจากที่ยุทธวิธีของเบย์ตอบข้อสอบข้อหนึ่งๆ ถูกต้อง และอีกข้อหนึ่งผิดก็จะสามารถนำไปแก่สมการความเป็นไปได้สูงสุด และจะได้ค่าการประมวลความสามารถและความคลาดเคลื่อนมาตรฐาน ข้อสอบข้อถัดไป ที่เลือกมาใช้สอบเป็นข้อสอบจากกลุ่มข้อสอบที่มีความยากเหมาะสมกับความสามารถของผู้สอบมากที่สุด เมื่อผู้สอบตอบข้อนั้นแล้วก็จะมีการประเมินค่าตอบทันที จากนั้นก็ทำการตอบข้อสอบทั้งหมดของผู้สอบ ซึ่งรวมถึงข้อสอบข้อสุดท้ายที่ผู้สอบตอบ จากนั้นก็ประมวลค่าความสามารถใหม่และความคลาดเคลื่อนมาตรฐานใหม่ โดยใช้สมการความเป็นไปได้สูงสุด ถ้าผู้สอบตอบข้อสุดท้ายผิด การประมวลค่าความสามารถจะลดต่ำลง ในขั้นสุดท้ายของกระบวนการ

การความคลาดเคลื่อนมาตรฐาน ในการประมาณค่าความสามารถลดลงเป็นผลจากสารสนเทศใหม่ที่ได้จากข้อสอบข้อสุดท้ายร่วมกับรูปแบบการตอบข้อสอบที่ผ่านมาของผู้สอบ

1.3 การประมาณค่าความสามารถของการทดสอบด้วยข้อสอบเฉพาะตัว

การประมาณค่าความสามารถของผู้สอบด้วยการทดสอบด้วยข้อสอบเฉพาะตัว มีผู้เสนอวิธีไว้หลายวิธี พอสรุปได้ดังนี้

(1) ประมาณค่าความสามารถจากค่าความยากของข้อที่ยากที่สุดที่ทำได้ (The most difficulty of item answered correctly) เป็นการประมาณความสามารถโดยพิจารณาจากข้อสอบที่ผู้สอบทำทั้งหมด ข้อใดที่มีค่าความยากสูงสุดที่ผู้สอบทำถูกต้องว่าเป็นค่าความสามารถของผู้สอบคนนั้น ค่าความยากที่วานี้ควรจะเป็นค่าความยากที่วิเคราะห์มาจากทฤษฎีการตอบสนองข้อสอบ

(2) ประมาณค่าความสามารถตามค่าเฉลี่ยความยากของข้อที่ทำได้ (Average difficulty of all item answered correctly) โดยการเอาค่าความยากของข้อสอบทุกข้อที่ผู้สอบทำถูกมาหาค่าเฉลี่ย ค่าเฉลี่ยที่ได้นี้จะถือเป็นค่าความสามารถของผู้สอบ

(3) ประมาณค่าความสามารถตามค่าเฉลี่ยความยากของทุกข้อที่ทำ (Average difficulty) วิธีการนี้เสนอโดย ลอร์ด (Lord 1970, cited by Weiss 1974) โดยการนำค่าความยากของข้อสอบทุกข้อที่ผู้สอบได้ทำ โดยไม่คำนึงว่าจะตอบถูกหรือไม่ มาหาค่าเฉลี่ย ค่าเฉลี่ยนี้จะถือเป็นความสามารถของผู้สอบ

(4) ประมาณค่าความสามารถตามค่าความยากของข้อสุดท้ายที่ทำ (Difficulty of the final term) วิธีการนี้ไม่คำนึงว่าข้อสุดท้ายจะตอบถูกหรือไม่ก็ตาม จะถือว่าค่าความยากของข้อสุดท้ายที่ได้ทำเป็นค่าความสามารถของผู้สอบ ซึ่งข้อสอบถูกจัดตามระดับความยาก

(5) ประมาณความสามารถตามค่าความยากของข้อสอบที่ต่อจากข้อสุดท้ายที่ทำ (Difficulty of (N+1)th Item) ซึ่งมีลักษณะคล้ายกับวิธีที่ 4 แต่เป็นการตรวจสอบอีกชั้นหนึ่ง กล่าวคือ ถ้าชั้นสุดท้ายที่ทำได้ ค่าความยากในชั้นต่อไปจะเพิ่มขึ้น ดังนั้นค่าความสามารถก็จะเท่ากับค่าความยากในชั้นต่อไปที่เพิ่มขึ้น หรือถ้าชั้นสุดท้ายทำผิด ค่าความยากในชั้นต่อไปก็จะลดลง ดังนั้นค่าความสามารถจะเท่ากับค่าความยากในชั้นต่อไปที่ลดลง

(6) ประมาณค่าความสามารถโดยใช้หลักการของเบย์ (Bayesian Strategies) การใช้หลักการของเบย์ยังมีวิธีการย่อยหลายวิธี วิธีที่นิยมใช้กันมาก คือ วิธี

ที่นักสถิติโรเจอร์ เจ โอเวน (Owen, 1975) ผู้เสนอขึ้นเพื่อใช้ในการประมาณค่าความสามารถของผู้สอบในการทดสอบด้วยข้อสอบเฉพาะตัวที่ให้ ชื่อว่า เบย์เซียน อัปเดตติง (Bayesian Updating) ซึ่งสามารถประมาณค่าความสามารถของผู้สอบได้ค่อนข้างคงที่ โดยมีสูตร ดังนี้

ในการแก้ที่ตอบข้อสอบถูก

$$e_m + 1 = e_m + (1 - c_i) \left[\frac{\alpha_m^k}{\sqrt{\frac{1}{a_i^2} + \alpha_m^k}} \right] \left[\frac{\phi(D)}{c_i + (1 - c_i)\phi(-D)} \right]$$

และ

$$\alpha_m^k + 1 = \alpha_m^k \left[1 - \frac{1 - c_i}{1 + \frac{1}{a_i^2 \alpha_m^k}} \right] \left[\frac{\phi(D)}{A} \right] \left[\frac{(1 - c_i)\phi(D) - D}{A} \right]$$

ในการแก้ที่ตอบข้อสอบผิด

$$e_m + 1 = e_m - \left[\frac{\alpha_m^k}{\sqrt{\frac{1}{a_i^2} + \alpha_m^k}} \right] \left[\frac{\phi(D)}{\phi(D)} \right]$$

และ

$$\alpha_m^k + 1 = \alpha_m^k \left[1 - \frac{\phi(D)}{1 + \frac{1}{a_i^2 \alpha_m^k}} \right] \left[\frac{\phi(D) + D}{\phi(D)} \right] \left[\phi(D) \right]$$

- เมื่อ $\phi(D)$ คือ ค่าออร์ดิเนต ของโค้งปกติ จุด D
- $\phi(D)$ คือ พื้นที่ใต้โค้งปกติจากค่า D ต่ำสุดจนถึงจุด D
- $$D = \frac{b_1 - e_m}{\sqrt{\frac{1}{a_1} + \alpha_m^2}}$$
- A = $c_1 + (1-c_1)\phi(-D)$
- e_m คือ ความสามารถของผู้สอบที่ประมาณค่าได้ก่อนตอบข้อสอบข้อที่ m+1 โดยปกติถ้าเป็นการประมาณค่าครั้งแรก ที่ยังไม่ทราบค่าความสามารถของผู้สอบ จะให้ $e_m = 0.000000$
- α_m คือ ความแปรปรวนในการประมาณค่าความสามารถของผู้สอบก่อนตอบข้อสอบข้อที่ m+1 โดยปกติถ้าเป็นการประมาณค่าครั้งแรกที่ยังไม่ทราบค่าความแปรปรวนจะให้ $e_m = 1.0000$
- e_{m+1} คือ ความสามารถของผู้สอบที่ประมาณค่าได้เมื่อตอบข้อสอบข้อที่ m+1
- α_{m+1} คือ ความแปรปรวนในการประมาณค่าความสามารถของผู้สอบเมื่อตอบข้อสอบข้อที่ m+1 แล้ว
- a_1 คือ ค่าอำนาจจำแนกของข้อสอบข้อที่ m+1
- b_1 คือ ค่าความยากของข้อสอบข้อที่ m+1
- c_1 คือ ค่าการคาดของข้อสอบข้อที่ m+1

(7) ประมาณค่าความสามารถโดยใช้หลักการความเป็นไปได้สูงสุด (Maximum Likelihood) การประมาณค่าโดยวิธีนี้เป็นที่นิยมใช้กันมากเพราะมีประสิทธิภาพค่อนข้างสูง ซึ่งวิธีการประมาณค่าแบบนี้ยังมีวิธีย่อยหลายวิธี แต่เป็นที่นิยมใช้กันมาก คือ วิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข (Conditional Maximum Likelihood) การประมาณค่าโดยวิธีนี้มีข้อจำกัดในกรณีที่ผู้สอบตอบข้อสอบทุกหมัดหรือผิดหมด จะไม่สามารถประมาณค่าได้ นอกจากนี้ยังมีจิมา (Samejima 1973, cited by Hambleton and Swaminathan 1985) ยังพบว่าในกรณีที่ใช้ 3 พารามิเตอร์ บางครั้งอาจจะมีค่าสูงสุดหลายค่า ในกรณีนี้ ลอร์ดได้ให้ข้อสังเกตว่าถ้าใช้ข้อสอบมากข้อ ($k > 20$) ในการสอบ ปัญหาจะไม่เกิดขึ้น ซึ่งมีขั้นตอนในการประมาณค่า ดังนี้ (Hambleton & Swaminathan 1985)

ขั้นที่ 1 ประเมินค่าความสามารถเริ่มต้น (e_m ; $m=0$) โดยใช้สูตร ดังนี้

$$e_0 = \ln [r_u / (k-r_u)]$$

เมื่อ r_u = ผลรวมของค่าอำนาจจำแนกของข้อที่ตอบถูก

$$= \sum a_{ij} u_{ij}; [u_{ij} = 1 \text{ เมื่อตอบข้อสอบถูก,}$$

$$u_{ij} = 0 \text{ เมื่อตอบข้อสอบผิด,}$$

a_{ij} คือ ค่าอำนาจจำแนกของข้อสอบข้อที่ j]

k คือ จำนวนข้อสอบทั้งหมดที่ผู้สอบได้ทำ

ขั้นที่ 2 หาค่า $P_{ij}(e_m)$ และ $Q_{ij}(e_m)$ โดยใช้สูตร ดังนี้

$$Da_{ij}(e_m - b_{ij})$$

$$P_{ij}(e_m) = c_{ij} + (1-c_{ij}) \frac{e^{Da_{ij}(e_m - b_{ij})}}{1 + e^{Da_{ij}(e_m - b_{ij})}}$$

$$Q_{ij}(e_m) = 1 - P_{ij}(e_m) = \frac{(1-c_{ij})}{1 + e^{Da_{ij}(e_m - b_{ij})}}$$

เมื่อ e_m คือ ความสามารถของผู้สอบที่ประเมินค่าได้ในครั้งที่ m

a_{ij} คือ ค่าอำนาจจำแนกของข้อสอบข้อที่ j

b_{ij} คือ ค่าความยากของข้อสอบข้อที่ j

c_{ij} คือ ค่าการเดาของข้อสอบข้อที่ j

D คือ ค่าคงที่ ซึ่งมีค่าเท่ากับ 1.7

ขั้นที่ 3 หาค่าตัวปรับแก้ (h_m) โดยใช้สูตร

$$h_m = \frac{D[r_u - \sum P_{ij}(e_m)]}{-D^2 \sum P_{ij}(e_m) Q_{ij}(e)}$$

ขั้นที่ 4 ประเมินค่าความสามารถของผู้สอบใหม่ โดยใช้สูตร

$$e_{m+1} = e_m - h_m$$

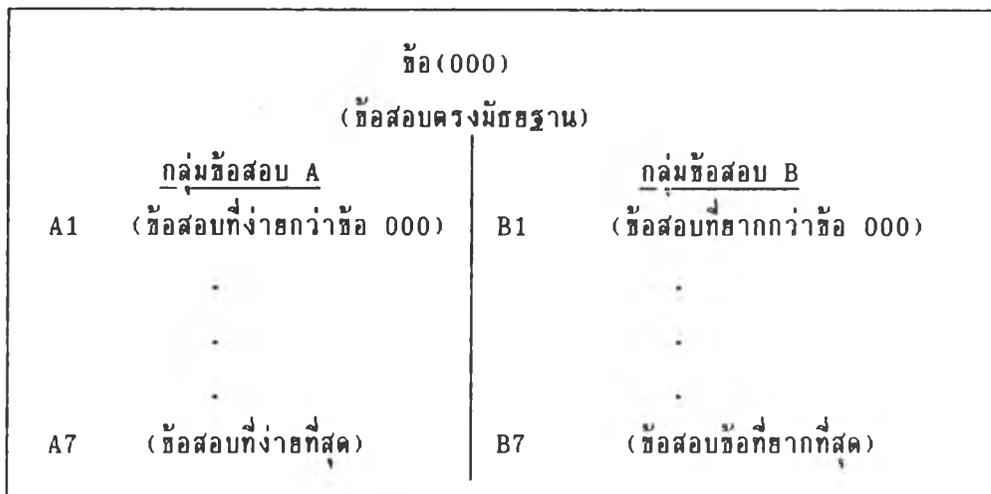
ทำการคำนวณซ้ำในขั้นที่ 2, 3, และ 4 จนกระทั่ง h_m เข้าใกล้ศูนย์ ($h_m < 0.001$)

2. แบบสอบตามระดับความสามารถ (Flexilevel Test)

2.1 หลักการเกี่ยวกับแบบสอบตามระดับความสามารถ

แบบสอบตามระดับความสามารถ เป็นแบบสอบชนิดหนึ่งของการทดสอบด้วยข้อสอบเฉพาะตัว (Tailored Testing) ซึ่ง ลอร์ด เป็นผู้เสนอขึ้น เพื่อใช้วัดค่าในการสอบด้วยการเขียนตอบ แบบสอบตามระดับความสามารถจัดเป็นแบบจำลองแยกทางคงที่ โดยปรับเปลี่ยนมาจากแบบสอบรูปปิรามิด เพื่อให้มีจำนวนข้อสอบในกลุ่มข้อสอบน้อยลงโดยที่แต่ละระดับความยากจะมีข้อสอบเพียงข้อเดียว แบบสอบชนิดนี้มุ่งหมายเช่นเดียวกับแบบสอบชนิดอื่นๆของการทดสอบด้วยข้อสอบเฉพาะตัว

แบบสอบตามระดับความสามารถ สามารถดัดแปลงมาจากแบบสอบประเพณีนิยมได้ โดยการปรับปรุง ค่าชี้แจง สมุดข้อสอบ (Test booklet) และกระดาษคำตอบ วิธีการก็คือ นำแบบสอบประเพณีนิยมที่มีข้อสอบจำนวน n ข้อ ซึ่งข้อสอบทั้งหมดผ่านการทดลองใช้ และคำนวณค่าพารามิเตอร์ของข้อสอบมาแล้ว นำข้อสอบทั้ง n ข้อ มาจัดเรียงลำดับตามค่าความยาก โดยเรียงจากง่ายไปหายาก จากนั้นคัดเลือกข้อสอบที่อยู่ตรงกลางหรือตรงมีมาตรฐานของข้อสอบทั้งหมดมา 1 ข้อ เพื่อแยกไว้สำหรับเป็นข้อเริ่มต้นในการสอบ ส่วนข้อสอบที่เหลือทั้งหมดที่มีความยากน้อยกว่าข้อสอบตรงมีมาตรฐาน เราเรียกว่า กลุ่มข้อสอบที่ง่าย และข้อสอบที่เหลือทั้งหมดที่มีความยากมากกว่าข้อสอบตรงมีมาตรฐาน เราเรียกว่า กลุ่มข้อสอบที่ยาก ซึ่งจะขอยกตัวอย่างให้เห็นได้ชัดเจนยิ่งขึ้นดังนี้ คือ ถ้ามีแบบสอบประเพณีนิยมฉบับหนึ่งมีข้อสอบจำนวน 15 ข้อ ซึ่งข้อสอบทั้ง 15 ข้อนี้ผ่านการทดลองใช้และหาค่าพารามิเตอร์ของข้อสอบมาแล้ว นำข้อสอบทั้ง 15 ข้อมาจัดเรียงลำดับตามค่าความยาก โดยเรียงจากข้อที่มีค่าความยากน้อยไปหาข้อที่มีค่าความยากมาก จากนั้นคัดเลือกข้อสอบที่อยู่ในลำดับที่ 8 ซึ่งเป็นข้อสอบที่อยู่ตรงมีมาตรฐาน กำหนดให้เป็นข้อสอบข้อที่ 000 ส่วนข้อสอบที่อยู่ในลำดับที่ 1 ถึง 7 จัดเป็นข้อสอบกลุ่มที่ง่าย ซึ่งเราอาจจะตั้งชื่อว่าเป็น กลุ่มข้อสอบ A และข้อสอบในลำดับที่ 7, 6, 5, ..., 1 เราจะให้เป็นข้อสอบข้อที่ A1, A2, A3, ..., A7 ตามลำดับ ส่วนข้อสอบที่อยู่ในลำดับที่ 9 ถึง 15 จัดเป็นข้อสอบกลุ่มที่ยาก ซึ่งเราอาจจะตั้งชื่อว่าเป็น กลุ่มข้อสอบ B และข้อสอบในลำดับที่ 9, 10, 11, ..., 15 เราจะให้เป็นข้อสอบข้อที่ B1, B2, B3, ..., B7 ตามลำดับ ดังแสดงในภาพที่ 3



ภาพที่ 3 ตัวอย่างการจัดโครงสร้างข้อสอบของแบบสอบตามระดับความสามารถ
(ดัดแปลงมาจากของ ลอร์ด (Lord ,1980)

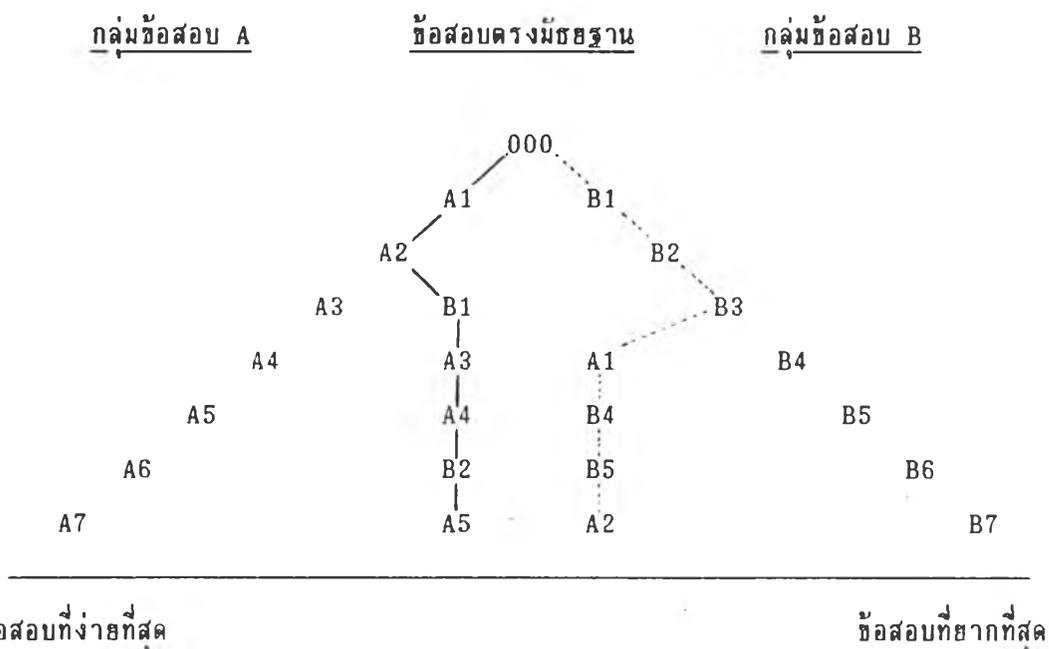
ส่วนกระดาษคำตอบที่ใช้กับแบบสอบตามระดับความสามารถ เป็นรูปแบบที่ผู้สอบสามารถทราบผลการตอบข้อสอบแต่ละข้อโดยทันทีว่าถูกหรือผิด ในกรณีที่ผู้สอบเลือกตัวเลือกที่เป็นคำตอบถูกจะปรากฏอักษร B ตรงตัวเลือกที่ผู้สอบแคะกระดาษคำตอบที่ปิดนั้นออก และถ้าผู้สอบเลือกตัวเลือกที่เป็นคำตอบผิด จะปรากฏอักษร A ตรงตัวเลือกนั้น ซึ่งดัดแปลงมาจากของ ลอร์ด โดยที่ลอร์ดใช้จุดสีแดง และจุดสีน้ำเงิน แทนอักษร A และ B ตามลำดับ ส่วนรูปแบบการจัดตำแหน่งของข้อสอบในกระดาษคำตอบจะคล้ายกับการจัดตำแหน่งของข้อสอบในแบบสอบ แต่จะต่างกันในกลุ่มข้อสอบ B ที่เรียงลำดับข้อสอบสลับกัน คือ จะเรียงจากบนลงล่างเป็น B7, B6, B5, ..., B1 เป็นต้น

2.2 การดำเนินการสอบแบบสอบตามระดับความสามารถ

หลักในการดำเนินการสอบของแบบสอบตามระดับความสามารถจะเริ่มต้นให้ผู้สอบทำข้อสอบที่มีระดับความยากตรงมีพื้นฐานของข้อสอบ (Median Difficulty Item) ทั้งหมดในแบบสอบ แล้วพิจารณาจากผลการตอบของผู้สอบ ถ้าผู้สอบตอบข้อสอบถูกจะได้ไปทำข้อสอบที่มีระดับความยากเพิ่มขึ้นเป็นข้อต่อไป ถ้าผู้สอบตอบข้อสอบผิดจะได้ไปทำข้อสอบที่มีระดับความยากลดลงเป็นข้อต่อไป (Weiss 1974) สำหรับการทำข้อสอบข้อต่อไปจะให้ผู้สอบปฏิบัติตามกฎนี้ตลอดไปจนสิ้นสุดการทดสอบ เกณฑ์ในการสิ้นสุดการทดสอบ คือ ผู้สอบต้องทำข้อสอบ

ทั้งหมด $(n+1)/2$ ข้อ เมื่อ n คือ จำนวนข้อสอบทั้งหมดในแบบสอบ

การตอบแบบสอบถามระดับความสามารถ ผู้สอบจะปฏิบัติตามกฎ ดังนี้ คือ เมื่อตอบข้อใด ถูก จะต้องไปตอบข้อสอบข้อที่มีหมายเลขน้อยที่สุดที่ยังไม่ได้ตอบ ในกลุ่มข้อสอบที่ยาก หรือกลุ่มข้อสอบ B และเมื่อตอบข้อสอบข้อใด ผิด จะต้องไปตอบข้อสอบที่มีหมายเลขน้อยที่สุดที่ยังไม่ได้ตอบ ในกลุ่มข้อสอบที่ง่าย หรือกลุ่มข้อสอบ A เป็นข้อต่อไป ดังจะยกตัวอย่างแนวทางในการตอบในภาพที่ 4



ภาพที่ 4 ตัวอย่างแนวทางในการตอบแบบสอบถามระดับความสามารถ

จากภาพที่ 4 แนวทางในการตอบแบบสอบถามระดับความสามารถ ผู้สอบทุกคนจะเริ่มสอบข้อสอบหมายเลข 000 เป็นข้อแรก โดยให้เส้นทึบแสดงแนวทางในการตอบแบบสอบถามของผู้สอบคนที่ 1 และเส้นปะแสดงแนวทางในการตอบแบบสอบถามของผู้สอบคนที่ 2

ผู้สอบคนที่ 1 เริ่มทำข้อสอบหมายเลข 000 ตอบผิดไปทำข้อสอบหมายเลข A1 ตอบผิดไปทำข้อสอบหมายเลข A2 ตอบถูกไปทำข้อสอบหมายเลข B1 ตอบผิดไปทำข้อสอบหมายเลข A3 ตอบผิดไปทำข้อสอบหมายเลข A4 ตอบถูกไปทำข้อสอบหมายเลข B2 ตอบผิดไปทำข้อสอบหมายเลข A5 คิดเป็นข้อสอบที่ทำทั้งหมดจำนวน 8 ข้อ และถ้าข้อสุดท้ายตอบถูกจำนวนข้อที่ตอบถูกเท่ากับ 3 ข้อ และถ้าข้อสุดท้ายตอบผิด จำนวนข้อที่ตอบถูกเท่ากับ 2 ข้อ

ผู้สอบคนที่ 2 ทำข้อสอบหมายเลข 000, B1, B2, B3, A1, B4, B5, A2 คิดเป็นข้อสอบที่ทำทั้งหมดจำนวน 8 ข้อ และถ้าทำถูกในข้อสุดท้าย จำนวนข้อที่ตอบถูกเท่ากับ 6 ข้อ และถ้าข้อสุดท้ายตอบผิด จำนวนข้อที่ตอบถูกเท่ากับ 5 ข้อ

จากลักษณะของแบบสอบตามระดับความสามารถที่กล่าวข้างต้น จะเห็นว่าถ้าผู้สอบที่มีระดับความสามารถสูงซึ่งตอบข้อสอบในช่วงแรกๆถูกต้อง จะได้ตอบกลุ่มข้อสอบที่ยากกว่า ผู้สอบที่มีระดับความสามารถต่ำกว่า ซึ่งตอบข้อสอบในช่วงแรกๆไม่ถูกต้อง นั้นแสดงให้เห็นว่าแบบสอบตามระดับความสามารถจะมีการจัดระดับความยากของข้อสอบในระหว่างดำเนินการสอบโดยอัตโนมัติ เพื่อให้พอเหมาะกับความสามารถของผู้สอบแต่ละคน

2.3 การให้คะแนนแบบสอบตามระดับความสามารถ

การให้คะแนนของแบบสอบตามระดับความสามารถ คือ การนับจำนวนข้อสอบที่ตอบได้ถูกต้องของผู้สอบแต่ละคน ยกเว้นผู้ที่ตอบผิดในข้อสุดท้ายจะได้รับคะแนนเพิ่มอีกครั้ง คะแนนเป็นโบนัส ซึ่งจากการให้คะแนนแบบนี้จะพบว่าคะแนนของผู้สอบบางคนจะอยู่ในรูปเศษส่วน ซึ่งอาจจะทำให้เป็นจำนวนเต็มได้โดยการคูณคะแนนของผู้สอบแต่ละคนด้วย 2 และจากการพิจารณาคะแนนของผู้สอบที่ได้คะแนนเท่ากัน ถ้าทำข้อสอบในชุดเดียวกันถือว่ามีความสามารถเท่ากัน แต่ถ้าทำข้อสอบต่างชุดกัน ผู้ที่ทำข้อสอบในชุดที่ยากกว่าจะเป็นผู้ที่มีความสามารถสูงกว่า ซึ่งวิธีการพิจารณาแบบนี้เป็นการประมาณความสามารถอย่างหยาบๆ แต่ถ้าต้องการที่จะทราบค่าความสามารถที่แน่นอน ควรจะนำผลการสอบที่ได้ไปประมาณค่าความสามารถตามวิธีการของการประมาณค่าความสามารถของการทดสอบด้วยข้อสอบเฉพาะตัว

3. ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory)

ทฤษฎีการทดสอบแบบประเพณีนิยม (Classical Test Theory) เป็นทฤษฎีที่มีจุดอ่อนหลายประการ อาทิเช่น ค่าพารามิเตอร์ของข้อสอบไม่คงที่ แปรเปลี่ยนไปตามกลุ่มตัวอย่างของผู้สอบ (Hambleton and Swaminathan 1985) ซึ่งไม่เหมาะสมกับการทดสอบด้วยข้อสอบเฉพาะตัว เนื่องจากผู้สอบแต่ละคนจะได้รับข้อสอบต่างชุดกัน และค่าความยากง่ายของข้อสอบต่างกัน ดังนั้นทฤษฎีที่เหมาะสมและสามารถนำมาใช้กับการทดสอบด้วยข้อสอบเฉพาะตัวได้อย่างมีประสิทธิภาพ คือ ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory) (Urry 1977) โดยมีหลักการที่จะเสนอในรายละเอียดต่อไป



3.1 หลักการของทฤษฎีการตอบสนองข้อสอบ

ทฤษฎีการตอบสนองข้อสอบ เป็นทฤษฎีที่อธิบายถึงความสัมพันธ์ระหว่างลักษณะหรือความสามารถที่มีอยู่ภายในตัวบุคคล กับพฤติกรรมการตอบสนองต่อข้อสอบของบุคคลนั้น (Lord and Novick 1968) โดยทฤษฎีนี้มีความเชื่อว่าพฤติกรรมการตอบสนองข้อสอบของผู้เข้าสอบ ซึ่งเป็นสิ่งที่สังเกตได้โดยตรง จะถูกกำหนดโดยคุณลักษณะ (Trait) หรือความสามารถ (Ability) ที่มีอยู่ในตัวบุคคลซึ่งไม่สามารถสังเกตได้โดยแสดงความสัมพันธ์ดังกล่าวในลักษณะโมเดลทางคณิตศาสตร์ที่กำหนดด้วยฟังก์ชันทางคณิตศาสตร์ ซึ่งเรียกว่า ฟังก์ชันลักษณะข้อสอบ (Item Characteristic Function) หรือเรียกว่าโค้งคุณลักษณะข้อสอบ (Item Characteristic Curve) หรือฟังก์ชันการตอบสนองข้อสอบ (Item Response Function) (Lord and Novick 1968) ซึ่งกำหนดได้หลายรูปแบบ แต่ทั้งนี้ถ้าจะใช้รูปแบบของทฤษฎีนี้ได้ต้องขึ้นอยู่กับว่าข้อมูลที่ได้จากการทดสอบเป็นไปตามข้อตกลงเบื้องต้นหรือไม่

3.2 ข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบ

ในการพิจารณาทฤษฎีการตอบสนองข้อสอบไปใช้ จะต้องคำนึงถึงว่าแบบสอบที่จะนำมาใช้เป็นไปตามข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบดังนี้

3.2.1 แบบสอบมิติเดียว (Unidimension Tests) หมายถึงว่า ข้อสอบแต่ละข้อในแบบสอบจะต้องวัดความสามารถหรือคุณลักษณะเดียวกัน หรือมีความเป็นเอกพันธ์กัน การตรวจสอบความเป็นมิติเดียวของแบบสอบส่วนใหญ่ใช้วิธีการวิเคราะห์ตัวประกอบ

3.2.2 ความเป็นอิสระในการตอบข้อสอบ (Local Independence) หมายถึงว่า การตอบสนองต่อข้อสอบต่างๆกัน ในแบบสอบของผู้สอบคนใดคนหนึ่งมีความเป็นอิสระในทางสถิติ นั่นคือ การตอบข้อสอบข้อใดข้อหนึ่งก็จะมีผลต่อการตอบข้อสอบข้ออื่น ๆ ซึ่งเป็นข้อตกลงเบื้องต้นที่สัมพันธ์กับข้อตกลงเบื้องต้นของคุณสมบัติความเป็นแบบสอบมิติเดียว นั่นคือถ้ามีคุณสมบัติความเป็นแบบสอบมิติเดียว ก็จะมีความเป็นอิสระในการตอบข้อสอบ และในทางกลับกัน ถ้ามีความเป็นอิสระในการตอบข้อสอบก็จะเป็นแบบสอบมิติเดียว

3.2.3 โค้งลักษณะข้อสอบ (Item Characteristic Curve) เป็นฟังก์ชันทางคณิตศาสตร์ที่แสดงความสัมพันธ์ระหว่างความน่าจะเป็นในการตอบข้อสอบข้อหนึ่งได้ถูกต้อง กับระดับความสามารถที่วัดโดยข้อสอบข้อนั้น ซึ่งเป็นฟังก์ชันการถดถอยชนิดไม่เป็นเชิงเส้นตรง (Non-Linear Regression) ของคะแนนข้อสอบบนระดับความสามารถที่วัดโดยแบบสอบ

โคงลักษณะข้อสอบมีหลายรูปแบบแตกต่างกันขึ้นอยู่กับโมเดลการตอบสนองข้อสอบ ในรูปแบบทางคณิตศาสตร์ที่ใช้อธิบายความสัมพันธ์ดังกล่าว

3.2.4 ทฤษฎีการตอบสนองข้อสอบ อธิบายว่าการที่ผู้สอบจะมีโอกาสที่จะตอบถูกหรือผิดขึ้นอยู่กับความสามารถของผู้สอบเท่านั้น ดังนั้นข้อสอบที่มีลักษณะเป็นข้อสอบวัดความเร็ว จะไม่สามารถประยุกต์ใช้กับทฤษฎีการตอบสนองข้อสอบนี้ได้ เพราะข้อสอบที่ผู้สอบไม่ได้ทำแต่ได้คะแนนเป็นศูนย์ อาจเป็นเพราะไม่มีเวลาทำไม่ใช่เพราะทำแล้วผิด

3.3 โมเดลในทฤษฎีการตอบสนองข้อสอบ

โมเดลในทฤษฎีการตอบสนองข้อสอบ เป็นโมเดลทางคณิตศาสตร์ที่อธิบายโคงลักษณะข้อสอบซึ่งมีหลายโมเดลด้วยกัน แต่ที่ได้รับการพัฒนาและมีการนำมาใช้มากที่สุดในการดำเนินการทดสอบ คือ โมเดลโลจิสติก (Logistic Model) เนื่องจากการนำไปใช้และการคำนวณสะดวกกว่าโมเดลอื่น โมเดลโลจิสติกมีพัฒนาขึ้น 4 รูปแบบ คือ โมเดลโลจิสติกหนึ่งพารามิเตอร์ สองพารามิเตอร์ สามพารามิเตอร์ และสี่พารามิเตอร์ดังรายละเอียดต่อไปนี้

3.3.1 โมเดลโลจิสติกหนึ่งพารามิเตอร์ (One-Parameter Logistic Model)

โมเดล เบิร์นบอม (Birnbaum) ได้พัฒนาขึ้นในปี 1968 เป็นโมเดลที่อธิบายข้อสอบด้วยพารามิเตอร์ค่าความยากเพียงค่าเดียว โดยมีข้อตกลงเบื้องต้นว่าค่าอำนาจจำแนกของข้อสอบทุกข้อมีค่าเท่ากัน และค่าการเดาของข้อสอบเท่ากับ 0 มีฟังก์ชัน ดังนี้

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

- เมื่อ $P_i(\theta)$ คือ โอกาสที่ผู้มีความสามารถ θ จะทำข้อสอบข้อที่ i ได้ถูกต้อง
 θ คือ ระดับความสามารถที่แท้จริงของผู้สอบ
 b_i คือ ค่าความยากของข้อสอบข้อที่ i
 e คือ ค่าคงที่มีค่าเท่ากับ 2.7182818

3.3.2 โมเดลโลจิสต์สองพารามิเตอร์ (Two-Parameter Logistic Model)

โมเดลนี้อธิบายข้อสอบด้วยสองพารามิเตอร์ คือ ค่าอำนาจจำแนกของข้อสอบ และค่าความยากของข้อสอบ โดยกำหนดให้ค่าการเดาของข้อสอบเป็น 0 มีฟังก์ชันดังนี้

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

- เมื่อ $P_i(\theta)$ คือ โอกาสที่ผู้มีความสามารถ θ จะทำข้อสอบข้อที่ i ได้ถูกต้อง
 θ คือ ระดับความสามารถที่แท้จริงของผู้สอบ
 a_i คือ ค่าอำนาจจำแนกของข้อสอบข้อที่ i
 b_i คือ ค่าความยากของข้อสอบข้อที่ i
 e คือ ค่าคงที่มีค่าเท่ากับ 2.7182818
 D คือ ตัวประกอบของสเกล มีค่าเท่ากับ 1.7

3.3.3 โมเดลโลจิสต์สามพารามิเตอร์ (Three-Parameter Logistic Model)

โมเดลนี้พัฒนาต่อจากโมเดลโลจิสต์สองพารามิเตอร์ เพื่อใช้กับแบบสอบที่มีอิทธิพลจากการเดาแฝงอยู่ด้วย มีฟังก์ชันดังนี้

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

- เมื่อ $P_i(\theta)$ คือ โอกาสที่ผู้มีความสามารถ θ จะทำข้อสอบข้อที่ i ได้ถูกต้อง
 θ คือ ระดับความสามารถที่แท้จริงของผู้สอบ
 a_i คือ ค่าอำนาจจำแนกของข้อสอบข้อที่ i
 b_i คือ ค่าความยากของข้อสอบข้อที่ i
 c_i คือ ค่าการเดาของข้อสอบข้อที่ i
 e คือ ค่าคงที่มีค่าเท่ากับ 2.7182818
 D คือ ตัวประกอบของสเกล มีค่าเท่ากับ 1.7

3.3.4 โมเดลโลจิสต์สี่พารามิเตอร์ (Four-Parameter Logistic Model)

โมเดลนี้แตกต่างจากโมเดลโลจิสต์สามพารามิเตอร์ โดยการเพิ่มพารามิเตอร์ y_i ซึ่งมีค่าต่ำกว่า 1 เล็กน้อย หมายถึง ค่าที่ผู้สอบที่มีความสามารถสูงมีความสะเพร่าตอบข้อสอบไม่ถูก ในทางปฏิบัติโมเดลนี้ยังไม่สามารถนำมาใช้ได้ มีฟังก์ชันดังนี้

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

3.4 การตรวจสอบคุณภาพแบบสอบ

การตรวจสอบคุณภาพของแบบสอบตามแนวคิดของทฤษฎีการตอบสนองข้อสอบ จะพิจารณาจากค่าฟังก์ชันสารสนเทศแบบสอบ (Test Information Function) ซึ่งจะเป็นดัชนีที่บอกความแม่นยำในการประมาณค่าความสามารถที่แท้จริง (Lord 1980) ซึ่งได้จากผลรวมของฟังก์ชันสารสนเทศข้อสอบ (Item Information Function) โดยที่ค่าฟังก์ชันสารสนเทศแบบสอบเป็นดัชนีผสม (Composite Index) ที่สร้างจากดัชนีบอกคุณลักษณะของข้อสอบหลายลักษณะ เช่น ค่าอำนาจจำแนก ค่าความยาก เป็นต้น รวมเป็นดัชนีเพียงตัวเดียวเพื่อชี้ถึงคุณภาพของแบบสอบได้ จะเห็นว่าในทฤษฎีการตอบสนองข้อสอบ จะใช้ค่าฟังก์ชันสารสนเทศแบบสอบแทนการหาความเที่ยงของคะแนนและความคลาดเคลื่อนมาตรฐานในการวัดตามแนวคิดของทฤษฎีการทดสอบแบบประเพณีนิยม (Hambleton, 1977) สำหรับโมเดลโลจิสต์ 3 พารามิเตอร์จะสามารถหาค่าฟังก์ชันสารสนเทศข้อสอบในแต่ละระดับความสามารถ (θ) ได้จากสูตร (Lord, 1980)

$$I(\theta, U_i) = \frac{(1.7 a_i)^2 (1 - c_i)}{[c_i + e^{1.7a(\theta - b_i)}] [1 + e^{-1.7a(\theta - b_i)}]}$$

และสามารถกำหนดคโด่งสารสนเทศข้อสอบได้จากสมการ

$$I(\theta, U_i) = (p'_i)^2 / P_i Q_i$$

เมื่อ $I(\theta, U_i)$ คือ ค่าฟังก์ชันสารสนเทศข้อสอบ
 P_i คือ ความชันของ ICC ที่ระดับความสามารถ θ
 P'_i คือ ความน่าจะเป็นที่ผู้สอบมีความสามารถ θ จะตอบ
 ข้อสอบข้อ i ได้ถูกต้อง
 Q_i คือ $1 - P_i$

และจะสามารถหาค่าฟังก์ชันสารสนเทศแบบสอบได้จากผลรวมของฟังก์ชันสารสนเทศข้อสอบทั้งหมดในแบบสอบเขียนสมการได้ดังนี้

$$I(\theta) = \sum I(\theta, U_i)$$

เมื่อ $I(\theta)$ คือ ค่าฟังก์ชันสารสนเทศแบบสอบ

ในการสอบทุกครั้งย่อมมีความคลาดเคลื่อนเกิดขึ้น ความคลาดเคลื่อนมาตรฐานในการประมาณค่า (Standard Error of Estimate : SEE) คือส่วนเบี่ยงเบนมาตรฐานในการประมาณค่าความสามารถ (θ) ซึ่งมีความสัมพันธ์กับค่าฟังก์ชันสารสนเทศแบบสอบ ความคลาดเคลื่อนมาตรฐานของการประมาณค่าความสามารถ คำนวณได้จากสมการ

$$SEE = \frac{1}{\sqrt{I(\theta)}} \quad (\text{Warm 1978})$$

จากสมการจะพบว่าค่า SEE เท่ากับรากที่สองของส่วนกลับของฟังก์ชันสารสนเทศของแบบสอบ ดังนั้น ถ้าค่าฟังก์ชันสารสนเทศสูงมาก ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความรู้อิ่งน้อยลง ในการสอบวัดทุกครั้งเราต้องการให้ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถมีค่าน้อยที่สุด นอกจากนี้ค่าเฉลี่ยของ SEE (SEE) ของผู้เข้าสอบทั้งหมดมีความสัมพันธ์กับค่าความเที่ยงในทฤษฎีการทดสอบแบบประเพณีนิยมด้วย เมื่อคะแนนเป็นมาตรฐาน และมีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1 ดังสมการ

$$r_{xx} = 1 - (\text{see})^2 \quad (\text{Warm 1978})$$

3.5 การนำเอาทฤษฎีการตอบสนองข้อสอบมาประยุกต์ใช้

จากคุณสมบัติที่สำคัญของทฤษฎีการตอบสนองข้อสอบในเรื่องค่าพารามิเตอร์ของข้อสอบไม่ขึ้นอยู่กับความสามารถของผู้สอบ และความสามารถของผู้สอบก็ไม่ขึ้นอยู่กับความยากหรือง่ายของข้อสอบ และค่าพารามิเตอร์ไม่เปลี่ยนแปลงไปตามตำแหน่งของข้อสอบ คุณลักษณะดังกล่าวคือความเป็นปรนัยของการวัดผล (Objectivity of Measurement) ที่มีความสำคัญมากในทางวัดผล (Write and Stones อ้างถึงใน สุวัฒน์ สุขมลสันต์ : 2528) จึงสามารถนำเอาคุณสมบัติเด่นเหล่านี้มาประยุกต์ใช้ในการวัดผลการศึกษาได้อย่างมีประสิทธิภาพในหลายๆเรื่อง ดังนี้

1. การทดสอบด้วยข้อสอบเฉพาะตัว (tailored testing) หมายถึง การทดสอบที่มีการจัดข้อสอบให้มีความยากที่เหมาะสมกับระดับความสามารถของผู้สอบ นับเป็นการนำเอาทฤษฎีการตอบสนองข้อสอบมาประยุกต์ใช้ได้อย่างมีประสิทธิภาพ (Urry 1977) โดยการนำเอาคุณสมบัติความไม่แปรเปลี่ยนของค่าพารามิเตอร์ของข้อสอบและความสามารถของผู้สอบมาใช้ การทดสอบด้วยข้อสอบเฉพาะตัวนั้นผู้สอบแต่ละคนจะได้รับชุดของข้อสอบที่แตกต่างกันตามระดับความสามารถของผู้สอบ แต่ก็ยังสามารถนำเอาความสามารถที่ประมาณค่าได้จากการทดสอบนี้มาเปรียบเทียบกันได้ ทั้งนี้เพราะความสามารถที่ประมาณค่าได้ยังอยู่ในมาตรฐานวัดเดียวกัน (Common Ability)

2. การสร้างคลังข้อสอบ (Item Bank) โดยการนำเอาทฤษฎีการตอบสนองข้อสอบมาใช้ในการวิเคราะห์และสร้างคลังข้อสอบ ทำให้ข้อสอบแต่ละข้อมีค่าสถิติต่างๆเป็นของประจำข้อไม่ขึ้นอยู่กับข้ออื่นและกลุ่มตัวอย่าง จึงทำให้เกิดความสะดวกในการคัดเลือกข้อสอบมาใช้ และจัดชุดข้อสอบให้ได้คุณภาพตามเกณฑ์ที่ต้องการ

3. การหาความลำเอียงของข้อสอบ (Item Bias) สามารถใช้โด่งลักษณะข้อสอบ มาตรวจสอบความลำเอียงของข้อคำถาม โดยปกติถ้าข้อคำถามนั้นไม่มีความลำเอียง โด่งลักษณะข้อสอบ ไม่ว่าจะได้จากผู้สอบกลุ่มใดก็ตามจะมีรูปร่างเหมือนกัน แต่ถ้าข้อคำถามนั้นมีความลำเอียง โด่งลักษณะข้อสอบจะมีรูปร่างแตกต่างกัน

4. การเทียบมาตรา (Equating) หมายถึง กระบวนการทางสถิติเพื่อปรับคะแนนที่ได้จากการทดสอบด้วยข้อสอบที่ต่างชุดกัน ที่มีโครงสร้างเดียวกันให้สามารถเทียบกันได้ แต่ถ้าการทดสอบนั้นวัดความสามารถต่างกันแล้วจะนำมาเทียบกันไม่ได้ เนื่องจากความสามารถของผู้สอบที่ประมาณค่ามาจากทฤษฎีการตอบสนองข้อสอบจะมีหน่วยการวัดเดียวกัน และไม่ขึ้นอยู่กับชุดของแบบสอบ จึงสามารถนำมาเทียบกันได้

4. งานวิจัยที่เกี่ยวข้อง

4.1 งานวิจัยจากต่างประเทศ

ลอร์ด (Lord, 1971) ได้ทำการประเมินกระบวนการทดสอบด้วยแบบสอบตามระดับความสามารถ โดยใช้คอมพิวเตอร์จำลองสถานการณ์ ด้วยวิธีแบบจำลองนอร์มอลไอโวจ์ ในการประมาณความสามารถของผู้สอบ (๑) และคำนวณค่าพารามิเตอร์ของข้อสอบ โดยใช้เกณฑ์ในการคัดเลือกข้อสอบ คือ ค่าอำนาจจำแนกจำแนกเท่ากับ .50 ศึกษาเปรียบเทียบกับแบบสอบประเพณีนิยมที่คัดเลือกข้อสอบด้วยเกณฑ์เดียวกัน โดยใช้แบบสอบประเพณีนิยมจำนวน 60 ข้อ และแบบสอบตามระดับความสามารถจำนวน 119 ข้อ แต่ผู้สอบตอบเพียง 60 ข้อ ผลการศึกษาพบว่า เมื่อขนาดของช่วงห่างระหว่างความยากของข้อสอบใหญ่ขึ้น จะทำให้สารสนเทศของผู้สอบลดลงทุกระดับความสามารถ และจะมีประสิทธิภาพสูงสุด เมื่อขนาดของช่วงห่างระหว่างความยากมีค่าอยู่ระหว่าง 0.033 และ 0.067 และจะไม่มีประสิทธิภาพเมื่อมีค่าใกล้เคียง 0.1 นอกจากนี้แบบสอบตามระดับความสามารถจะให้สารสนเทศสูงกว่าแบบสอบประเพณีนิยม ในกลุ่มผู้สอบที่มีระดับความสามารถทั้งสูงและต่ำ แต่เมื่อผู้สอบมีระดับความสามารถใกล้เคียงกับ 0 หรือมีระดับความสามารถปานกลาง จะให้สารสนเทศต่ำกว่าแบบสอบแบบประเพณีนิยม และประสิทธิภาพของการวัดของแบบสอบประเพณีนิยมจะเท่ากับแบบสอบตามระดับความสามารถ จำนวน 60 ข้อ เมื่อไม่มีการเดา ต้องใช้แบบสอบประเพณีนิยม จำนวน 58, 60, 69, และ 86 ข้อ ที่ระดับความสามารถเท่ากับ 0, +1, +2, และ +3 และเมื่อการเดาเป็น 0.2 จะต้องใช้แบบสอบประเพณีนิยมจำนวน 58, 60, 70, 83 และ 114 ข้อ ที่ระดับความสามารถ 0, +1, +2 หรือ +2.25, 3 และ -3

ผลการศึกษาของลอร์ดพบว่า แบบสอบตามระดับความสามารถจะมีประสิทธิภาพสูงเมื่อผู้สอบมีความสามารถสูงหรือต่ำ แต่ที่ระดับความสามารถปานกลาง แบบสอบตามระดับความสามารถจะให้สารสนเทศต่ำกว่าแบบสอบประเพณีนิยมและเมื่อใช้ข้อสอบ จำนวน 60 ข้อเท่านั้น แบบสอบตามระดับความสามารถจะมีประสิทธิภาพสูงกว่าแบบสอบประเพณีนิยม และใช้เวลาน้อยกว่าแบบสอบประเพณีนิยม

เวลล์ (Vale 1975 อ้างถึงใน เสรี ปรมชวลิตโรจน์ 2531) ได้รายงานผลการประเมินประสิทธิภาพของแบบสอบตามระดับความสามารถ แบบสอบประเพณีนิยม แบบสอบสองขั้นตอน แบบสอบสามขั้นตอน แบบสอบปิรามิด แบบสอบปรับระดับชั้น และสุทธวิธีของเบย์ โดยสุกัลฉบับให้ผู้สอบตอบเพียง 8 ข้อ กลุ่มข้อสอบที่นำมาสร้างแบบสอบ จำนวน 68 ข้อแบ่งออก

เป็น 17 กลุ่มศึกษา โดยใช้สถานการณ์จำลองด้วยคอมพิวเตอร์โดยมีข้อตกลงเบื้องต้น 4 ประการ คือ ข้อสอบทั้งหมดอยู่ในเกณฑ์ที่ดี มีค่าอำนาจจำแนกเท่ากัน ไม่มีการเดา ให้คะแนนด้วยวิธีของเบย์ (Bayesian Scoring) และประมาณค่าความสามารถเริ่มต้นของผู้สอบเท่ากับ 0.5 โดยใช้โด่งสารสนเทศในการประเมินแบบสอบ ผลการศึกษาพบว่าแบบสอบตามระดับความสามารถที่ประกอบด้วยข้อสอบที่มีค่าความยากกระจาย ตั้งแต่ง่ายสุดไปจนถึงยากที่สุด ให้สารสนเทศสูงกว่าแบบสอบประเพณีนิยมทุกระดับความสามารถ แบบสอบตามระดับความสามารถที่ประกอบด้วยข้อสอบที่มีค่าระดับความสามารถใกล้ๆ 0 จะให้สารสนเทศต่ำกว่าแบบสอบประเพณีนิยม ส่วนแบบสอบชนิดอื่นของการทดสอบด้วยข้อสอบเฉพาะตัว ให้สารสนเทศสูงกว่าแบบสอบตามระดับความสามารถในทุกระดับความสามารถ

ซีควิน (Sequin 1978 อ้างถึงใน งานนิศย์ ธาตุทอง 2527) ได้ศึกษาเปรียบเทียบแบบสอบตามระดับความสามารถ จำนวน 6 ฉบับ วิเคราะห์หาค่าพารามิเตอร์ตามทฤษฎีการทดสอบแบบประเพณีนิยม 3 ฉบับ วิเคราะห์ตามนอร์มอลโอโจฟฟ์ ชนิด 3 พารามิเตอร์ และแบบสอบแบบประเพณีนิยม จำนวน 6 ฉบับ การจัดโครงสร้างแบบสอบทั้ง 12 ฉบับ ใช้ 3 วิธี คือ จัดตามข้อสอบที่มีค่าอำนาจจำแนกสูงสุด จัดตามความยากเฉลี่ยใกล้เคียงกับความสามารถของกลุ่มตัวอย่างและจัดตามความยากกระจายกว้าง (rectangular distribution) ใช้กลุ่มตัวอย่าง 1,317 คน สร้างจากกลุ่มข้อสอบสองกลุ่ม คือ ใช้วิชาภาษาฝรั่งเศส จำนวน 35 ข้อ วิชาคณิตศาสตร์ 27 ข้อ และยังนำข้อสอบในกลุ่มมาสร้างเป็นแบบสอบต้นฉบับ N ข้อ กับแบบสอบที่ได้จากการลดข้อสอบในแบบสอบต้นฉบับให้เหลือ n ข้อ แบบสอบทุกฉบับใช้สอบในสถานการณ์จำลอง

ผลการศึกษาพบว่า แบบสอบตามระดับความสามารถที่เรียงค่าความยากจากนอร์มอลโอโจฟฟ์ พารามิเตอร์ 3 ตัว ดีกว่าแบบสอบตามระดับความสามารถที่เรียงตามค่าสัดส่วนการตอบถูกอย่างมีนัยสำคัญ ไม่แตกต่างกัน แบบสอบชนิดเดียวกันที่เรียงตามความสามารถของผู้สอบที่ให้สารสนเทศสูงสุดของข้อสอบแต่ละข้อและแบบสอบประเพณีนิยม ที่เรียงตามค่าอำนาจจำแนกสูงสุด จะมีสหสัมพันธ์บิสerial-ไบซีเรียล (Point-biserial correlation coefficient) กับ แบบสอบต้นฉบับสูง

ต่อมาได้มีผู้นำเอาแบบสอบตามระดับความสามารถมาบริหารการสอบโดยใช้คอมพิวเตอร์แทนการเขียนตอบ โดยมีผู้วิจัย ดังนี้

ลี (Lee 1987) ได้พัฒนาการทดสอบแบบเอ็มเอเอที (Microcomputerized Adaptive Achievement Test: MAAT) โดยใช้ทฤษฎี 2 แบบ คือแบบเอ็สพีอาร์ทีของวอลท์ (Wald's Sequential Probability Ratio Test: SPRT) และทฤษฎีของแบบสอบ

ตามระดับความสามารถของลอร์ด (Lord's Flexilevel Test) โดยใช้ข้อสอบที่มีอยู่ในคลังข้อสอบพีอีทีเอ (The Pitt Educational Testing Aids: PETA) ได้ทดลองศึกษากับนักศึกษาระดับบัณฑิตศึกษา จำนวน 3 ห้องเรียน ในรายวิชาการศึกษาเบื้องต้น ผลการศึกษาพบว่า

(1) มีความเป็นไปได้สูงในอนาคตที่จะนำเอาการทดสอบแบบเอ็มเอเอที (MAAT) มาใช้ในการทดสอบผลสัมฤทธิ์ทางการเรียนในระดับบัณฑิตศึกษา ผู้ดำเนินการสอบไม่จำเป็นต้องมีความรู้ทางด้านจิตวิทยาหรือคอมพิวเตอร์มากนักก็สามารถดำเนินการสอบได้ เพียงแต่ดำเนินการตามขั้นตอนต่างๆตามคู่มือครุ

(2) คะแนนจากการทดสอบแบบ เอ็มเอเอที (MAAT) มีความสัมพันธ์ทางบวกกับคะแนนผลสอบกลางเทอม

(3) นักศึกษามีทัศนคติที่ดีต่อการทดสอบแบบเอ็มเอเอที (MAAT) ในทางบวก

4.2 งานวิจัยในประเทศ

งามนิศย์ ธาตทอง และปรีชา เครือวรรณ (2527) สร้างและประเมินแบบสอบตามระดับความสามารถในวิชาคณิตศาสตร์ ระดับชั้นมัธยมศึกษาปีที่ 2 โดยใช้แบบสอบตามระดับความสามารถ 1 ฉบับ จำนวน 35 ข้อ ผู้สอบตอบเพียง 18 ข้อ แบบสอบประเพณีนิยม 1 ฉบับจำนวน 18 ข้อ จากการศึกษาพบว่า แบบสอบตามระดับความสามารถมีความตรงตามเกณฑ์เท่ากับ .7238 และแบบสอบประเพณีนิยมมีความตรงตามเกณฑ์เท่ากับ .6109 และเมื่อทดสอบนัยสำคัญทางสถิติ พบว่าแตกต่างกันอย่างมีนัยสำคัญที่ระดับ .01 การประเมินประสิทธิภาพเชิงปฏิบัติพบว่านักเรียนส่วนใหญ่มีทัศนคติที่ดีต่อการสอบด้วยแบบสอบตามระดับความสามารถ จากผลการวิจัยจะเห็นว่าแบบสอบตามระดับความสามารถมีความตรงตามเกณฑ์สูงกว่าแบบสอบประเพณีนิยม

เสรี ปรมชวลิตโรจน์ (2531) ได้ศึกษาความตรงในการจำแนกความรอบรู้และการประมาณค่าความสามารถโดยใช้แบบสอบตามระดับความสามารถจำนวน 2 ฉบับ และแบบสอบรูปปิรามิดขนาดชั้นแปดจำนวน 2 ฉบับ โดยทำการศึกษากับนักเรียนชั้นมัธยมศึกษาปีที่ 3 จำนวน 378 คน ผลการศึกษาพบว่าความตรงในการจำแนกความรอบรู้ เมื่อใช้แบบสอบอิงเกณฑ์ที่มีความแม่นยำในการวัดผู้สอบที่มีระดับความสามารถ 1.0 ที่สร้างขึ้นเป็นเกณฑ์ ปรากฏว่าแบบสอบตามระดับความสามารถและแบบสอบรูปปิรามิด แบบขนาดชั้นแปดทุกฉบับ มีความตรงใน

การจำแนกความรอบรู้ได้อย่างมีนัยสำคัญทางสถิติ และเมื่อเปรียบเทียบค่าความตรงแต่ละแบบ พบว่าแบบสอบตามระดับความสามารถทั้งสองฉบับมีความตรง ในการจำแนกความรอบรู้ได้สูงกว่า แบบสอบรูปปรีามิคออยู่หนึ่งฉบับ อย่างมีนัยสำคัญทางสถิติที่ระดับ .01

จากการศึกษา เอกสาร และงานวิจัยที่เกี่ยวข้อง พบว่า แบบสอบตามระดับความสามารถใช้กลุ่มข้อสอบขนาดเล็กที่สุดเมื่อเปรียบเทียบกับแบบสอบชนิดอื่นๆ ของการทดสอบด้วยข้อสอบเฉพาะตัว และสามารถบริหารการสอบด้วยการเขียนตอบได้ ซึ่งจะสามารถนำไปใช้โดยทั่วไปได้ และจากงานวิจัยที่เกี่ยวข้องจากต่างประเทศ จะพบว่า แบบสอบตามระดับความสามารถให้สารสนเทศดีกว่าแบบสอบประเพณีนิยม ดังที่ ลอร์ด (Lord 1971) และ เวลล์ (Vale 1975) ได้ศึกษาไว้ แต่งานวิจัยส่วนใหญ่ศึกษาจากสถานการณ์จำลอง ซึ่งสามารถกำหนดค่าความยากต่างๆได้ ซึ่งในสถานการณ์จริงค่าความยากไม้อาจจะกำหนดได้ตามที่ต้องการ ส่วนงานวิจัยภายในประเทศยังมีเป็นจำนวนน้อย และการวิจัยส่วนใหญ่ใช้เฉพาะในการสอบย่อย ซึ่งผลการวิจัยที่ได้ พบว่า แบบสอบตามระดับความสามารถมีความตรงตามเกณฑ์สูงกว่าแบบสอบประเพณีนิยมและสร้างทัศนคติที่ดีให้กับผู้สอบ อีกทั้งมีความตรงในการจำแนกความรอบรู้ ดังนั้นผู้วิจัยจึงสนใจที่จะพัฒนาแบบสอบตามระดับความสามารถในวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ซึ่งเป็นการวัดผลสัมฤทธิ์ปลายภาคเรียนในสถานการณ์การทดสอบจริง เพื่อศึกษาถึงคุณภาพของแบบสอบตามระดับความสามารถที่สร้างขึ้นว่าจะวัดได้เหมาะสมกับระดับความสามารถของผู้สอบจริงหรือไม่ โดยในการวิจัยครั้งนี้ ใช้โมเดลโลจิสติก ชนิด 3 พารามิเตอร์ในการวิเคราะห์หาค่าพารามิเตอร์ของข้อสอบ เนื่องจากโมเดลรูปแบบนี้ให้สารสนเทศเกี่ยวกับแบบสอบเลือกตอบได้ดีที่สุด และใช้วิธีการประมาณค่าความสามารถของผู้สอบด้วยวิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข (Conditional Maximum Likelihood) เนื่องจากเป็นวิธีที่นิยมใช้กันมาก และมีประสิทธิภาพค่อนข้างสูง โดยนำค่าความสามารถที่ได้มาเทียบกับเกณฑ์ คือ เกณฑ์เฉลี่ยวิชาคณิตศาสตร์ (ค 101- ค 311) เพื่อศึกษาคุณภาพของแบบสอบในด้านความตรงตามเกณฑ์