บทที่ 2

วรรณคดีที่เกี่ยวข้อง

วรรณคดีที่เกี่ยวข้องซึ่งจะเสนอในบทนี้แบ่งออกเป็น 5 ส่วนคือ ทฤษฎีการตอบสนองต่อข้อสอบ การประมาณค่าความสามารถและค่าพารามิเตอร์ของข้อสอบ วิธีการวิเคราะห์ความลำเอียงของข้อสอบ ดัชนีความลำเอียงของข้อสอบและรายงานการวิจัยที่เกี่ยวข้องกับการเปรียบเทียบวิธีการวิเคราะห์ความลำเอียงของข้อสอบ

ทฤษฎีการตอบสนองต่อข้อสอบ

ความเป็นมาของทฤษฎีการตอบสนองต่อข้อสอบ

Ferguson (1942) และ Lawley (1943) เป็นผู้ริเริ่มทฤษฎีการตอบสนองต่อข้อสอบ โดยมีหลักการว่า ผลการสอบของผู้สอบจากแบบสอบใด ๆ ขึ้นอยู่กับความสามารถ (Ability or Skill) ของผู้สอบ ต่อมาในปี ค.ศ. 1952 Lord ได้เสนอทฤษฎีนี้ใหม่ในรูปโค้งลักษณะข้อสอบ (Item Characteristic Curve : ICC) โดย Lord ได้เสนอว่าโค้งลักษณะข้อสอบมีลักษณะเป็นโค้งปกติสะสม ต่อมาเรียกว่า Normal Ogive Model ซึ่งโมเดลนี้จะกล่าวถึงพารามิเตอร์ 2 ตัวคือค่าความยากและค่าอำนาจจำแนก แต่เนื่องจากโมเดลนี้มีการคำนวณยุ่งยากมาก และขาดแคลนโปรแกรมคอมพิวเตอร์ที่จำเป็นต้องใช้วิเคราะห์ข้อมูลตามทฤษฎี จึงทำให้ Lord หยุดความสนใจในทฤษฎีนี้ไประยะหนึ่ง

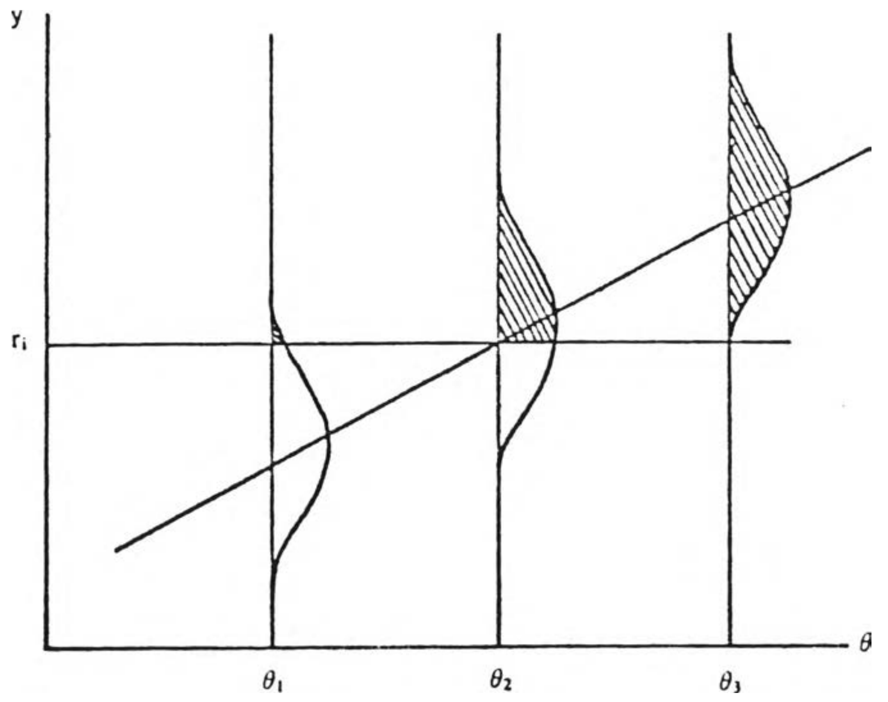
ปี ค.ศ. 1960 Rasch ได้ศึกษาเกี่ยวกับเรื่องนี้ โดยไม่ทราบแนวคิดของ Lord มาก่อน ซึ่ง Rasch ได้เสนอแนวคิดในรูปของพารามิเตอร์ตัวเดียว คือ ค่าความยาก บางครั้ง

จึงมีผู้เรียกแบบจำลองนี้ว่า Rasch Model และในปี ค.ศ. 1965 Lord ก็หันมาสนใจและพัฒนาทฤษฎีการตอบสนองต่อข้อสอบใหม่ (Warm 1979 : 19)

ปี ค.ศ. 1968 Birnbaum ได้เสนอ Logistic Model ที่ใช้พารามิเตอร์ 2 ตัวคือ ค่าความยากและค่าอำนาจจำแนก ซึ่งเป็นแบบจำลองที่ง่ายกว่าของ Lord จึงทำให้ Logistic Model เป็นที่นิยมแพร่หลายและมีการพัฒนาขึ้นเรื่อย ๆ จนกระทั่งสามารถใช้ได้กับพารามิเตอร์ตัวเดียว และพารามิเตอร์สามตัว (Warm 1979 : 19-21)

หลักการของทฤษฎีการตอบสนองต่อข้อสอบ

ทฤษฎีการตอบสนองต่อข้อสอบเป็นทฤษฎีที่อธิบายถึงความสัมพันธ์ระหว่างลักษณะ หรือความสามารถที่มีอยู่ในตัวบุคคลกับพฤติกรรมการตอบสนองต่อข้อสอบของบุคคลนั้น โดยทฤษฎีนี้มีความเชื่อว่าพฤติกรรมการตอบสนองต่อข้อสอบของบุคคล จะถูกกำหนดโดยลักษณะหรือความสามารถที่มีอยู่ในตัวบุคคล (Lord and Novick 1968 : 358) ซึ่งไม่สามารถจะสังเกตได้ ทฤษฎีนี้พยายามที่จะอธิบายความสัมพันธ์ระหว่างลักษณะภายในตัวบุคคลกับพฤติกรรมที่บุคคลตอบสนองต่อข้อสอบ การอธิบายความสัมพันธ์ดังกล่าวแสดงออกมาในรูปแบบจำลองทางคณิตศาสตร์ โดยให้คะแนนที่ได้รับจากการตอบข้อสอบ (y) แทนพฤติกรรมการตอบสนองต่อข้อสอบ ให้ θ แทนลักษณะหรือความสามารถในตัวบุคคล และ r_i เป็นเกณฑ์ที่บอกว่า y แค่ไหนจึงจะทำข้อสอบข้อ i ได้ถูก ดังนั้นถ้า $y > r_i$ แสดงว่าทำข้อสอบข้อ i ได้ถูก และถ้า $y < r_i$ แสดงว่าทำข้อสอบข้อ i ผิด ความสัมพันธ์ระหว่างความสามารถ (θ) กับพฤติกรรมการตอบสนอง (y) แสดงได้ดังภาพที่ 4



ภาพที่ 4 แสดงความสัมพันธ์ระหว่างความสามารถกับพฤติกรรมการตอบสนอง

จากภาพจะเห็นได้ว่า ถ้านำโอกาสที่จะตอบถูก (พื้นที่ส่วนที่แรเงา) ในระดับความสามารถต่าง ๆ มาเขียนกราฟใหม่ จะได้โค้งลักษณะข้อสอบ (ICC) เป็นรูปต่าง ๆ ทั้งขึ้นอยู่กับฟังก์ชันทางคณิตศาสตร์และจำนวนพารามิเตอร์ที่จะใช้อธิบาย ฟังก์ชันทางคณิตศาสตร์ที่แสดงความสัมพันธ์ระหว่างระดับความสามารถกับพฤติกรรมการตอบสนองต่อข้อสอบเรียกว่า ฟังก์ชันการตอบสนองต่อข้อสอบ (Item Characteristic Function) (Lord and Novick 1968:360) ซึ่งสามารถเขียนฟังก์ชันได้ดังนี้

$$P_i(\theta) = \text{Prob}(U_i = 1|\theta) \quad \text{เมื่อ} \quad U_i = 0,1$$

จากฟังก์ชันข้างต้นนี้ หมายถึงโอกาสที่ผู้สอบซึ่งมีความสามารถ θ จะตอบคำถามข้อ i ได้ถูกต้อง

<u>ข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองต่อข้อสอบ</u>

ทฤษฎีการตอบสนองต่อข้อสอบนั้น มีข้อตกลงเบื้องต้นดังนี้

1. Unidimensionality เป็นการสมมุติว่าข้อสอบในแบบสอบมีลักษณะเป็นเอกพันธ์ นั่นคือแบบสอบนั้นจะต้องมุ่งวัดความสามารถอย่างใดอย่างหนึ่งเพียงความสามารถเดียว หากไม่กำหนดข้อตกลงเบื้องต้นเช่นนี้ จะทำให้แบบจำลองของทฤษฎีมีความยุ่งยากมาก (Hambleton and Cook 1977 : 77) ส่วนการตรวจสอบว่าข้อมูลจากการสอบเป็นไปตามข้อตกลงเบื้องต้นหรือไม่ อาจทำได้โดยการวิเคราะห์องค์ประกอบ (Factor Analysis)

2. Local Independence เป็นการกำหนดข้อตกลงเบื้องต้นเกี่ยวกับความเป็นอิสระในการตอบสนองต่อข้อสอบ กล่าวคือการตอบสนองต่อข้อสอบข้อใดข้อหนึ่งของผู้สอบไม่มีผลต่อการตอบสนองต่อข้อสอบข้ออื่น ๆ ในแบบสอบ ในการตรวจสอบว่าข้อสอบแต่ละข้อเป็นไปตามข้อตกลงเกี่ยวกับ Local Independence หรือไม่ ทำได้โดยอาศัยเทคนิคการวิเคราะห์องค์ประกอบเช่นกัน (Hambleton and Cook 1977 : 78)

3. Item Characteristic Curve เป็นข้อตกลงเกี่ยวกับโค้งลักษณะข้อสอบ กล่าวคือโค้งลักษณะข้อสอบเป็นกราฟของฟังก์ชันทางคณิตศาสตร์ที่แสดงถึงความสัมพันธ์ระหว่างโอกาสในการตอบข้อสอบข้อนั้นถูกกับความสามารถหรือลักษณะที่วัดโดยข้อสอบข้อนั้น โค้งลักษณะข้อสอบมีหลายรูปแบบทั้งนี้ขึ้นอยู่กับแบบจำลองที่ใช้อธิบายความสัมพันธ์ดังกล่าว

<u>ลักษณะเด่นของทฤษฎีการตอบสนองต่อข้อสอบ</u>

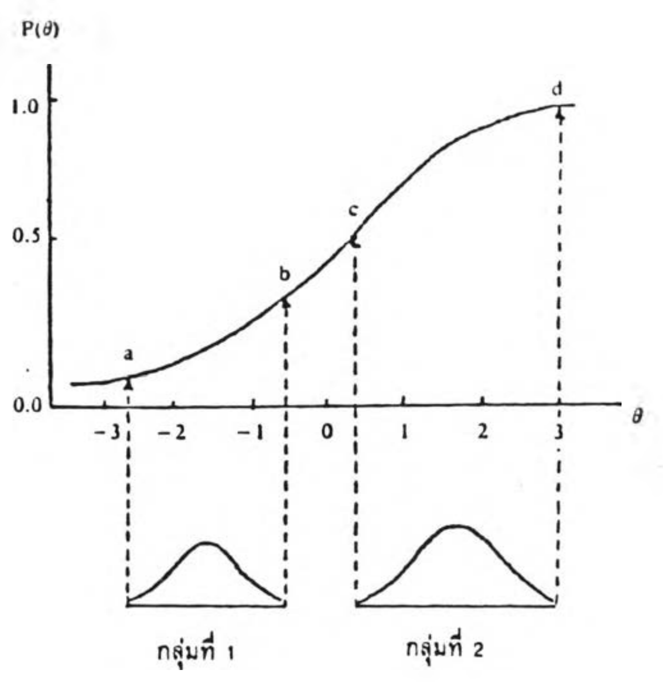
ทฤษฎีการตอบสนองต่อข้อสอบมีความเหนือกว่า (Superiority) ทฤษฎีการสอบแบบคลาสสิค ดังนี้

1. ความไม่แปรเปลี่ยนของค่าพารามิเตอร์ (Invariant of Item Parameter) กล่าวคือ ไม่ว่าจะประมาณค่าจากกลุ่มตัวอย่างที่มีความสามารถระดับใดก็ตาม ค่าพารามิเตอร์จะไม่แปรเปลี่ยนไปตามกลุ่มของผู้สอบ นั่นคือค่าความยาก ค่าอำนาจจำแนกและค่าโอกาสในการตอบถูกโดยการเดาจะไม่แปรเปลี่ยนไม่ว่าจะประมาณค่าจากผู้สอบกลุ่มใด

2. **จะใช้ข้อสอบกับใครก็ได้** (Person - free) **นั่นคือไม่ว่าจะนำข้อสอบไปใช้**สอบกับบุคคลกลุ่มใด โค้งลักษณะข้อสอบก็จะคงเดิม

3. **จะใช้ข้อสอบข้อใดก็ได้** (Item - free) **ในการประมาณค่าความสามารถแท้** (θ) **ของผู้สอบ จะใช้ข้อสอบชุดใด จำนวนเท่าใดก็ได้** บางครั้งอาจใช้ข้อสอบเพียง 3 - 5 ข้อ ก็สามารถประมาณค่าความสามารถแท้ของผู้สอบได้แล้ว

เกี่ยวกับความไม่แปรเปลี่ยนของค่าพารามิเตอร์นั้นสามารถอธิบายได้ด้วยภาพที่ 5



ภาพที่ 5 ความไม่แปรเปลี่ยนของค่าพารามิเตอร์ของข้อสอบที่ประมาณค่าจากกลุ่มที่แตกต่างกัน

จากภาพแสดงให้เห็นว่าในการประมาณค่าจากกลุ่มผู้สอบที่มีความสามารถต่ำ (กลุ่มที่ 1) เส้นโค้งลักษณะข้อสอบที่ได้จะอยู่ในช่วง a - b เส้นโค้งลักษณะข้อสอบในช่วงนี้จะอธิบายถึงโอกาสในการตอบข้อสอบข้อนี้ถูกของผู้สอบที่มีความสามารถต่ำ และทำนองเดียวกันเส้นโค้งลักษณะข้อสอบในช่วง c - d ก็จะแสดงถึงโอกาสที่จะตอบข้อสอบข้อนี้ถูกของผู้สอบที่มีความสามารถสูง (กลุ่มที่ 2) ถ้าข้อสอบข้อนี้วัดความสามารถอันเดียวกันของกลุ่มตัวอย่างทั้งสองกลุ่มแล้วเส้นโค้งลักษณะข้อสอบย่อมมีเส้นเดียว เส้นโค้งในช่วง a - b และช่วง c - d ย่อมเป็นส่วนหนึ่งของเส้นโค้งลักษณะข้อสอบข้อนี้ เมื่อเส้นโค้งลักษณะข้อสอบมีได้เส้นเดียว ตัวพารามิเตอร์ซึ่งเป็นตัวกำหนดรูปร่างลักษณะของเส้นโค้งก็ย่อมมีได้ค่าเดียวเช่นกัน หรืออาจกล่าวได้ว่าถ้าพารามิเตอร์ของข้อสอบไม่แปรเปลี่ยนไปตามกลุ่มผู้สอบแล้วข้อสอบที่นำไปสอบกับผู้สอบที่มีความสามารถแตกต่างกัน เมื่อนำข้อมูลที่ได้จากแต่ละกลุ่มมาวิเคราะห์จะมีค่าพารามิเตอร์ของข้อสอบเป็นค่าเดียวกัน

Lord (1980 : 36) ได้อธิบายให้เห็นว่าการไม่แปรเปลี่ยนของค่าพารามิเตอร์นั้นไม่ได้หมายความว่าค่าพารามิเตอร์ที่ได้จากการประมาณค่าโดยใช้กลุ่มตัวอย่างที่แตกต่างกันจะมีค่าเท่ากันเสมอ ค่าพารามิเตอร์ของข้อสอบจะเท่ากันหรือไม่ ขึ้นอยู่กับเงื่อนไขในการทดสอบบางประการเช่น ถ้าเลือกมาตรวัดที่มีจุดเริ่มต้นเดียวกันและหน่วยในการวัดหน่วยเดียวกันแล้วค่าพารามิเตอร์ที่ประมาณค่าจากกลุ่มตัวอย่างที่แตกต่างกันจะมีค่าเท่ากัน ในทำนองกลับกันหากเลือกมาตรวัดที่มีจุดเริ่มต้นและมีหน่วยในการวัดแตกต่างกันแล้ว การไม่แปรเปลี่ยนของค่าพารามิเตอร์หมายความว่าค่าพารามิเตอร์ที่ประมาณได้จากกลุ่มตัวอย่างที่ต่างกันของข้อสอบชุดหนึ่งจะมีความสัมพันธ์กันในเชิงเส้นตรง

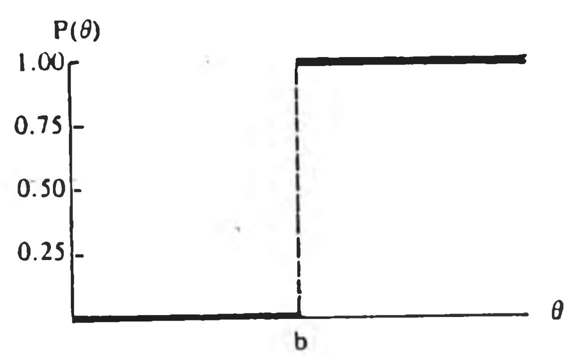
นอกจากนี้ Baker (1977 : 170) ยังได้อธิบายเพิ่มเติมเกี่ยวกับการไม่แปรเปลี่ยนของค่าพารามิเตอร์นี้ว่าขึ้นอยู่กับเงื่อนไข 2 ประการ คือ ความสามารถที่กล่าวถึงต้องสามารถนิยามได้ชัดเจนและวัดได้ด้วยข้อสอบ กับความสามารถที่วัดนั้นจะต้องมีความคงที่ภายในช่วงระยะเวลาหนึ่ง

<u>แบบจำลองที่ใช้ในทฤษฎีการตอบสนองต่อข้อสอบ</u>

ทฤษฎีการตอบสนองต่อข้อสอบได้มีการพัฒนาแบบจำลองขึ้นมาหลายแบบด้วยกัน โดยแบบจำลองแต่ละแบบนั้นจะต่างกันในรูปของฟังก์ชันทางคณิตศาสตร์และจำนวนพารามิเตอร์ที่จะอธิบาย ซึ่งในที่นี้จะขอเสนอแบบจำลองทางคณิตศาสตร์ในทฤษฎีการตอบสนองต่อข้อสอบดังนี้

<u>Guttman Perfect Scale</u>

ลักษณะของฟังก์ชันเป็นแบบขั้นบันได (Step Function) กล่าวคือโอกาสในการตอบข้อสอบข้อใดข้อหนึ่งถูกมีความสัมพันธ์กับความสามารถของบุคคล โดยโอกาสในการตอบข้อสอบถูกจะเป็น 0 หรือ 1 เท่านั้น เส้นโค้งลักษณะข้อสอบจะมีลักษณะดังภาพที่ 6



ภาพที่ 6 เส้นโค้งลักษณะข้อสอบที่เป็น Guttman Perfect Scale

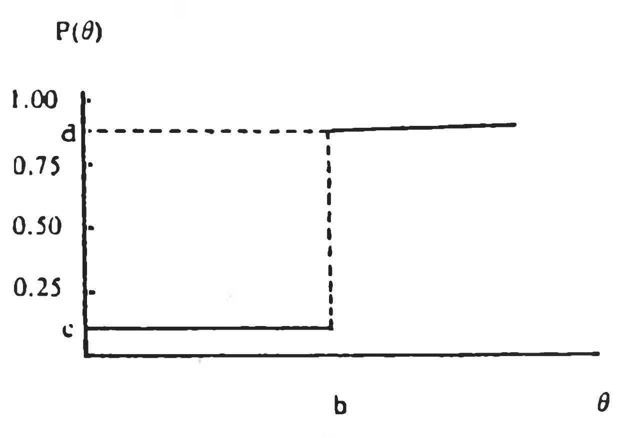
จากภาพผู้ที่มีความสามารถต่ำกว่า b ($\theta < b$) มีโอกาสที่จะตอบข้อสอบถูกเท่ากับ 0 ส่วนผู้ที่มีความสามารถมากกว่า b ($\theta > b$) มีโอกาสที่จะตอบข้อสอบถูกเป็น 1 สมการที่แสดงความสัมพันธ์ระหว่าง θ กับโอกาสที่จะตอบข้อสอบถูก ($P(\theta)$) เป็นดังนี้

$$P(\theta) = \begin{cases} 1 & \text{ถ้า } \theta > b \\ 0 & \text{ถ้า } \theta < b \end{cases} \qquad (1)$$

แบบจำลอง Guttman Perfect Scale มีลักษณะเป็น Deterministic Model และมีการนำไปประยุกต์ใช้น้อยมาก ทั้งนี้เพราะเป็นไปได้ยากที่ข้อมูลจากการสอบจะสอดคล้องกับแบบจำลองนี้

Latent Distance Model

ลักษณะของฟังก์ชันยังคงเป็นแบบขั้นบันไดเช่นเดียวกันกับ Guttman Perfect Scale แต่โอกาสในการตอบข้อสอบถูกจะแปรเปลี่ยนระหว่าง 0 ถึง 1 โค้งลักษณะข้อสอบมีลักษณะดังภาพที่ 7



ภาพที่ 7 เส้นโค้งลักษณะข้อสอบใน Latent Distance Model

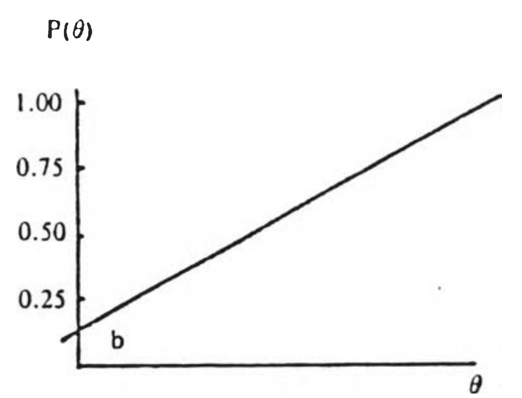
จากภาพแสดงว่าผู้ที่มีความสามารถต่ำกว่า b มีโอกาสที่จะตอบข้อสอบข้อนี้ถูกเท่ากับ c ส่วนผู้ที่มีความสามารถมากกว่า b มีโอกาสที่จะตอบข้อสอบข้อนี้ถูกเท่ากับ d โดยที่ $0 < c < d < 1.00$ ซึ่งสามารถเขียนสมการแสดงความสัมพันธ์ระหว่าง θ กับโอกาสที่จะตอบข้อสอบถูก (P (θ)) ได้ดังนี้

$$P(\theta) = \begin{cases} c & \text{ถ้า } \theta < b \\ d & \text{ถ้า } \theta > b \end{cases} \qquad (2)$$

แบบจำลองนี้มีลักษณะเป็น Stochastic Model อย่างไรก็ตามโอกาสที่ข้อมูลจะสอดคล้องกันแบบจำลองนี้ก็เป็นไปได้ยาก เพราะบุคคลที่มีความสามารถมากกว่า หรือน้อยกว่า b จะแสดงพฤติกรรมการตอบที่คงเส้นคงวาเช่นนี้คงเป็นไปได้ยาก

<u>Linear Model</u>

โค้งลักษณะข้อสอบเป็นเส้นตรง โดยโอกาสในการตอบข้อสอบถูกจะเป็นสัดส่วนโดยตรงกับความสามารถของบุคคล เส้นโค้งลักษณะข้อสอบจะมีลักษณะดังภาพที่ 8



ภาพที่ 8 เส้นโค้งลักษณะข้อสอบตาม Linear Model

สมการที่แสดงความสัมพันธ์ระหว่าง θ กับโอกาสที่จะตอบข้อสอบถูก (P (θ)) เป็นดังนี้

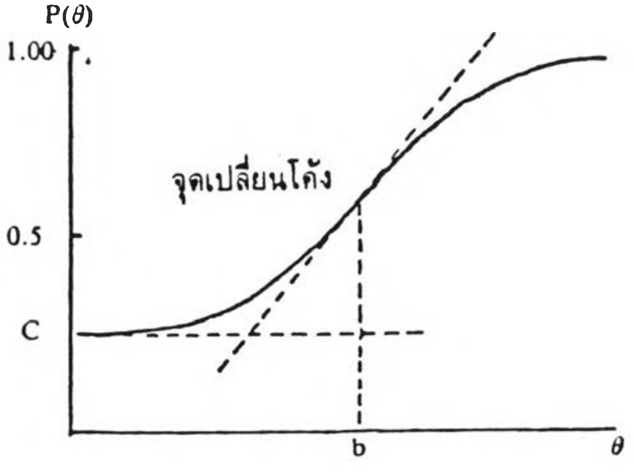
$$P(\theta) = b + a\theta \qquad (3)$$

จากสมการ (3) ถ้า a ≠ 0 แล้วผู้มีความสามารถในระดับต่ำมาก ๆ มีโอกาสที่จะตอบข้อสอบถูกเป็นลบ ส่วนผู้ที่มีความสามารถสูงมาก ๆ ก็มีโอกาสที่จะตอบข้อสอบถูกมีค่าเกิน 1 ได้ Lazarsfeld ผู้เสนอแบบจำลองนี้ได้กำหนดข้อตกลงเบื้องต้นเพิ่มเติมว่าการแจกแจงของความสามารถ (θ) จะต้องอยู่ในช่วงที่ทำให้ P (θ) มีค่าอยู่ระหว่าง 0 ถึง 1 เท่านั้น (Torgerson 1958 : 368) แบบจำลองดังกล่าวนี้มีข้อจำกัดในการนำไปใช้มาก ทั้งนี้เนื่องจากข้อตกลงเบื้องต้นที่กำหนดเพิ่มเติมนั่นเอง

แบบจำลองทั้งหมดที่กล่าวมานี้ เป็นแบบจำลองที่ไม่ค่อยจะมีความสอดคล้องกับข้อมูลจากการสอบในสถานการณ์จริงมากนัก แต่อย่างไรก็ตามแบบจำลองดังกล่าวก็เป็นแบบจำลองที่ให้แนวคิดเชิงทฤษฎีที่มีคุณค่าต่อการพัฒนาแบบจำลองที่เหมาะสมในระยะเวลาต่อมา

Normal Ogive Model

โค้งลักษณะข้อสอบตามแบบจำลองนี้จะเป็นรูปตัว S ดังภาพที่ 9



ภาพที่ 9 เส้นโค้งลักษณะข้อสอบของ Normal Ogive Model

จากภาพแสดงให้เห็นว่าโค้งลักษณะข้อสอบเป็นกราฟที่แสดงความสัมพันธ์ระหว่างโอกาสที่จะตอบข้อสอบข้อใดข้อหนึ่งถูกกับระดับความสามารถของผู้สอบ โดยผู้สอบที่มีระดับความสามารถสูงมีโอกาสที่จะตอบข้อสอบถูกมากกว่าผู้สอบที่มีความสามารถต่ำ ค่าความสามารถและค่าความยากของข้อสอบจะอยู่ในสเกลเดียวกัน โดยมีค่าอยู่ระหว่าง -∞ ถึง +∞ แต่ในทางปฏิบัติแล้วระดับความสามารถจะมีค่าอยู่ระหว่าง - 3 ถึง + 3 ค่า - 3 หมายถึงผู้สอบมีความสามารถต่ำมาก ค่า + 3 หมายถึงผู้สอบมีความสามารถสูงมาก ค่าความยากจะมีค่าอยู่ระหว่าง -∞ ถึง +∞ เช่นเดียวกันกับค่าอำนาจจำแนกของข้อสอบ (Hambleton and Cook 1977) ส่วนค่าการเดานั้นหมายถึงค่าที่ปลายต่ำสุด (Lower Tail) ของโค้งลักษณะข้อสอบ ซึ่งมีค่าอยู่ระหว่าง 0 ถึง 1

ความสัมพันธ์ดังกล่าวนี้จะอยู่ในรูปฟังก์ชันการแจกแจงสะสมแบบปกติ (Cumulative Normal Ogive) ดังนี้

$$P_i(\theta) = c_i + (1 - c_i) \int_{-\infty}^{a_i(\theta - b_i)} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt \qquad (4)$$

เมื่อ $P_i(\theta)$ แทนโอกาสที่ผู้สอบซึ่งมีความสามารถ θ จะตอบคำถามข้อ i ได้ถูกต้อง

a_i แทนค่าอำนาจจำแนกของข้อสอบข้อ i

b_i แทนค่าความยากของข้อสอบข้อ i

c_i แทนค่าการเดา

ค่าพารามิเตอร์ของข้อสอบ (a_i, b_i และ c_i) จะเป็นตัวที่ทำให้รูปร่างของโค้งลักษณะข้อสอบแต่ละข้อแตกต่างกันไป ซึ่งสามารถอธิบายรายละเอียดของค่าพารามิเตอร์ต่าง ๆ เหล่านี้ได้ดังนี้

a_i เป็นสัดส่วนโดยตรงกับความชัน (Slope) ของโค้งลักษณะข้อสอบ ณ จุดเปลี่ยนโค้ง

$$\text{ความชัน ณ จุดเปลี่ยนโค้ง} = \begin{cases} \dfrac{a(1-c)}{\sqrt{2\pi}} & \text{กรณีที่มีการเดา} \\ \dfrac{a}{\sqrt{2\pi}} & \text{กรณีที่ไม่มีการเดา} \end{cases}$$

ค่า a_i นี้จะมีค่าอยู่ระหว่าง $-\infty$ ถึง $+\infty$

b_i เป็นค่าของระดับ θ ที่ทำให้โค้งลักษณะข้อสอบมีการเปลี่ยนโค้ง (จุดเปลี่ยนโค้งจะอยู่ที่ $b, P(b)$) ค่าพารามิเตอร์ b_i นี้จะบอกตำแหน่งของโค้งลักษณะข้อสอบ ณ $\theta = b$ ซึ่งถ้าไม่มีการเดา ($c = 0$) แล้วโอกาสที่จะตอบข้อสอบข้อนั้นถูกจะมีค่าเท่ากับ $(1 - c)/2$ หรือมีค่าเท่ากับ 0.5

ค่า b_i นี้มีค่าอยู่ระหว่าง $-\infty$ ถึง $+\infty$

c_i เป็นโอกาสของผู้สอบที่มีความสามารถในระดับต่ำจะตอบคำถามข้อ i ได้ถูกต้อง ซึ่งก็คือค่าโอกาสที่จะตอบถูกโดยการเดานั่นเอง c_i เป็น Lower Assymtote ของโค้งลักษณะข้อสอบ

ค่า c_i นี้จะมีค่าอยู่ระหว่าง 0 ถึง 1

สมการ (4) เป็นสมการที่มีพารามิเตอร์ 3 ตัว จึงอาจเรียกว่า Three - parameter Normal Ogive Model แต่หากในสถานการณ์การสอบที่ผู้สอบมีโอกาสตอบถูกโดยการเดามีน้อยมากหรือไม่มีเลย สมการ (4) ก็จะเหลือค่าพารามิเตอร์เพียง 2 ตัว ดังนี้

$$P_i(\theta) = \int_{-\infty}^{a_i(\theta - b_i)} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt \qquad (5)$$

Normal Ogive Model นั้นมีปัญหาบางประการทั้งในด้านทฤษฎี และคณิตศาสตร์ จึงได้มีผู้พยายามคิดหาแบบจำลองที่มีรูปร่างคล้ายคลึงกันกับ Normal Ogive Model แต่มีความสะดวกกว่าในทางคณิตศาสตร์ แบบจำลองดังกล่าวนี้ก็คือ Logistic Model

Logistic Model

เป็นแบบจำลองที่มีลักษณะใกล้เคียงกับ Normal Ogive Model มากที่สุด โค้งลักษณะข้อสอบมีลักษณะเป็นรูปตัว S เช่นเดียวกับ Normal Ogive Model ส่วนความสัมพันธ์ระหว่างโอกาสในการตอบถูกกับระดับความสามารถอยู่ในรูปฟังก์ชันการแจกแจงสะสมแบบโลจิสต์ (Logistic Cumulative Distribution Function) ดังนี้

$$P_i(\theta) = c_i + (1 - c_i) \left[1 + e^{-1.7 a_i (\theta - b_i)}\right]^{-1} \qquad (6)$$

โดยที่ a_i, b_i และ c_i เป็นค่าพารามิเตอร์ของข้อสอบ และมีความหมายเช่นเดียวกับ Normal Ogive Model ส่วน e เป็นค่าคงที่มีค่าประมาณ 2.71828... และเมื่อมีการปรับค่าด้วย Scaling Factor (มีค่า 1.7) แล้วจะทำให้ค่าฟังก์ชัน Normal Ogive กับค่าฟังก์ชัน Logist แตกต่างกันเพียงเล็กน้อย แต่ในแง่ของการคำนวณแล้ว Logistic Model มีค่าความง่ายและสะดวกกว่ามาก นอกจากนี้ในการสอบจริงอาจมีผู้ที่มีความสามารถสูงตอบผิดด้วยความเลินเล่อ กรณีเช่นนี้ Logistic Model มีความแกร่ง (Robustness) ต่อข้อมูลมากกว่า Normal Ogive Model จึงทำให้ Logistic Model เป็นที่นิยมกันมากในทางปฏิบัติ (Lord 1980 : 14)

สมการ (6) นี้ อาจเรียกว่า Three - parameter Logistic Model เพราะมีตัวพารามิเตอร์ในสมการ 3 ตัว

Haebara (1979) ได้อธิบายโค้งลักษณะข้อสอบที่มีพารามิเตอร์ 3 ตัว ไว้ดังนี้

$$\lim_{\theta \to -\infty} P_i(\theta) = c_i \qquad (7)$$

$$P_i(\theta \mid \theta = b_i) = \frac{1}{2} + \frac{c_i}{2} \qquad (8)$$

$$\left(\frac{\partial P_i(\theta)}{\partial \theta}\right)_{\theta = b_i} = \frac{D}{4} a_i (1 - c_i) \qquad (9)$$

$$\frac{\partial^2 P_i(\theta)}{\partial \theta^2} \begin{cases} = 0, & \text{ถ้า } \theta = b_i \\ > 0, & \text{ถ้า } \theta < b_i \\ < 0, & \text{ถ้า } \theta > b_i \end{cases} \qquad (10)$$

จากสมการ (7) จะเห็นได้ว่าผู้สอบที่มีความสามารถต่ำสุด ($-\infty$) มีโอกาสที่จะตอบข้อสอบข้อ i ถูกด้วยความน่าจะเป็นเท่ากับ c_i และในกรณีที่ผู้สอบมีระดับความสามารถเท่ากับความยากของข้อสอบ ($\theta = b_i$) แล้วโอกาสที่จะตอบข้อสอบข้อนั้นถูกจะเท่ากับ $1/2 + c_i/2$ ดังสมการ (8) และที่ระดับความสามารถของผู้สอบ ณ จุดนี้ข้อสอบจะมีอำนาจจำแนกสูงโดยโค้งลักษณะข้อสอบจะมีความชันเท่ากับ $(D/4)\ a_i\ (1 - c_i)$ ดังสมการ (9) ค่าความชันนี้จะค่อย ๆ เพิ่มขึ้นเมื่อระดับความสามารถของผู้สอบมีค่าน้อยกว่าความยาก ($\theta < b_i$) และค่าความชันนี้จะค่อย ๆ ลดลง เมื่อระดับความสามารถมีค่ามากกว่าค่าความยากของข้อสอบ ($\theta > b_i$)

ในกรณีที่ผู้สอบมีโอกาสตอบถูกโดยการเดาน้อยมากหรือไม่มีเลย อาจกำหนดให้ค่าการเดา (c_i) เป็น 0 สมการก็จะเหลือค่าพารามิเตอร์เพียง 2 ตัวคือ a_i และ b_i ดังนี้

$$P_i(\theta) = \frac{1}{1 + e^{-1.7 a_i (\theta - b_i)}} \tag{11}$$

สมการ (11) นี้ก็คือ Two - parameter Logistic Model

ในกรณีที่กำหนดให้ $c_i = 0$ และถือว่าข้อสอบทุกข้อมีอำนาจจำแนกเท่ากัน ($a_i = \bar{a}$) สมการก็จะเหลือค่าพารามิเตอร์เพียงตัวเดียวคือ b_i ดังนี้

$$P_i(\theta) = \frac{1}{1 + e^{-1.7 \bar{a} (\theta - b_i)}} \tag{12}$$

สมการ (12) นี้ก็คือ One - parameter Logistic Model

ในขณะที่ Lord พัฒนา Normal Ogive Model และ Birnbaum ได้เสนอ Logistic Model นั้น Rasch นักคณิตศาสตร์ชาวเดนมาร์กได้เสนอแบบจำลองการตอบสนองต่อข้อสอบ ดังนี้

$$P_i(\theta^*) = \frac{\theta^*}{\theta^* + b_i^*} \tag{13}$$

เมื่อ $\theta^* = e^{1.7 \bar{a} \theta}$

$b_i^* = e^{1.7 \bar{a} b_i}$

เมื่อพิจารณาแล้วจะเห็นว่าสมการ (13) ของ Rasch ก็คือสมการ (12) ของ One - parameter Logistic Model นั่นเอง บางครั้งจึงมีผู้เรียกแบบจำลองนี้ว่า Rasch Model

นอกจาก Logistic Model แล้ว Hambleton และคณะ (1978 : 476) ได้เสนอแบบจำลองอื่น ๆ อีก ดังนี้

Nominal Response Model

จากแบบจำลองที่กล่าวมาแล้วข้างต้นนั้นใช้ได้กับการให้คะแนนแบบ 2 ค่า (Binary Item) ส่วน Nominal Response Model นั้นเป็นแบบจำลองที่สามารถประยุกต์ใช้ได้กับข้อสอบที่มีการตรวจให้คะแนนมากกว่า 2 ค่า (Multichotomusly Scored) ผู้เสนอแบบจำลองนี้คือ Bock จุดมุ่งหมายของแบบจำลองนี้ก็เพื่อความแม่นยำในการประมาณค่าความสามารถของผู้สอบ แบบจำลองนี้จะมีโค้งลักษณะของตัวเลือกแต่ละตัว (Item - option Characteristic Curve) โค้งของตัวเลือกที่เป็นคำตอบถูกจะมีลักษณะเป็นฟังก์ชันที่มีค่าเพิ่มขึ้นโดยตลอด (Monotonic Increasing Function) ส่วนโค้งของตัวเลือกที่ผิดจะมีลักษณะอย่างไรขึ้นอยู่กับการเลือก รูปแบบทางคณิตศาสตร์ของโค้งลักษณะของตัวเลือกมีได้หลายแบบ ตัวอย่างเช่น รูปแบบของ Bock ที่สมมุติว่าโอกาสที่ผู้มีความสามารถเท่ากับ θ จะเลือกตอบตัวเลือก k (จากตัวเลือกทั้งหมด m ตัว) ของข้อสอบข้อที่ i จะเป็นดังนี้

$$P_{ik}(\theta) = \frac{e^{b^*_{ik} + a^*_{ik}\theta}}{\sum_{k=1}^{m} e^{b^*_{ik} + a^*_{ik}}} \qquad (14)$$

เมื่อ b^*_{ik} และ a^*_{ik} คือค่าพารามิเตอร์ของตัวเลือกที่ k สำหรับระดับความสามารถ θ ใด ๆ ผลรวมของโอกาสในการเลือกตัวเลือกทุกตัวจะเท่ากับ 1 ในกรณีที่

ข้อสอบมี 2 ตัวเลือก ($m = 2$) แบบจำลองนี้ก็คือ Two - parameter Logistic Model นั่นเอง

Grade Response Model

เป็นกรณีพิเศษของ Nominal Response Model แบบจำลองนี้จะใช้ในกรณีที่พฤติกรรมการตอบสนองสามารถจัดเรียงลำดับ (Order) ได้ เช่นกรณีคะแนนจากแบบสอบวัดทัศนคติ เป็นต้น Samejima ได้เสนอแบบจำลองนี้ โดยสมมุติว่าพฤติกรรมการตอบสนองต่อข้อสอบสามารถจัดแบ่งออกได้เป็น $m_i + 1$ ประเภท และให้คะแนน $x_i = 0, 1, \ldots, m_i$ ตามลำดับ โอกาสที่ผู้มีความสามารถระดับ θ จะได้คะแนนข้อนี้เป็น x_i คือ

$$P_{x_i}(\theta) = P^{*}_{x_i}(\theta) - P^{*}_{(x_i+1)}(\theta) \qquad (15)$$

เมื่อ $P^{*}_{x_i}(\theta)$ คือ Item Response Function ของการให้คะแนนแบบ 2 ค่า โดยผู้ที่ได้คะแนนต่ำกว่า x_i ถือเป็น 0 และผู้ที่ได้คะแนนเท่ากับ x_i หรือมากกว่าถือเป็น 1

การประมาณค่าความสามารถและค่าพารามิเตอร์ของข้อสอบ

ในการประมาณค่าความสามารถและค่าพารามิเตอร์ของข้อสอบนั้นจะใช้คอมพิวเตอร์โปรแกรม LOGIST ซึ่งพัฒนาโดย Wood, Wingersky และ Lord (1976) โดยโปรแกรม LOGIST นี้จะทำการประมาณค่าต่าง ๆ ด้วยวิธี Maximum Likelihood

สำหรับแบบจำลองโลจิสติ (Logistic Model) ที่มีพารามิเตอร์ 3 ตัว จะทำการประมาณค่าพารามิเตอร์ของข้อสอบแต่ละข้อ 3 ตัวคือ a_i, b_i และ c_i ($i = 1, 2, \ldots, n$)

และค่าความสามารถของผู้สอบแต่ละคน (θ_a) เมื่อ $a = 1,2, \ldots, N$ ดังนั้นถ้ามีผู้สอบ N คน และข้อสอบ n ข้อ ค่าพารามิเตอร์ที่จะต้องประมาณค่าจะมีทั้งหมดเท่ากับ N + 3n ตัว ซึ่งมีฟังก์ชันในการประมาณค่าพารามิเตอร์ดังกล่าว ดังสมการ (16)

$$L^* = \prod_{a=1}^{n} \prod_{i=1}^{n_a} P_{ia}^{V_{ia}} Q_{ia}^{1-V_{ia}} \qquad (16)$$

เมื่อ P_{ia} คือโอกาสที่ผู้สอบคนที่ a จะตอบข้อสอบข้อที่ i ได้ถูก หรือ คือ $P_i(\theta)$ นั่นเอง

Q_{ia} คือ $1 - P_{ia}$

n คือจำนวนข้อสอบ

N คือจำนวนผู้สอบ

n_a คือจำนวนข้อสอบที่ผู้สอบคนที่ a ทำเสร็จ

V_{ia} คือคะแนนในข้อที่ i ของผู้สอบคนที่ a โดยถ้าตอบถูกจะมีค่าเท่ากับ 1 ตอบผิดมีค่าเท่ากับ 0 และถ้าเว้นไว้ไม่ตอบจะมีค่าเท่ากับ 1/A เมื่อ A คือจำนวนตัวเลือกของข้อสอบข้อที่ i

ค่า L^* นี้จะถูกปรับใหม่ให้มีค่าต่ำสุดเท่ากับ F ดังสมการ (17)

$$F = -\log L^* = -\sum_{a=1}^{N} \sum_{i=1}^{n_a} [V_{ia} \log P_{ia} + (1 - V_{ia}) \log Q_{ia}] \qquad (17)$$

โดยกำหนดให้ Z แทนค่าอนุพันธ์ที่ 1 ของ F ซึ่งมีค่าเท่ากับศูนย์และเป็นเวคเตอร์ของพารามิเตอร์ซึ่งประกอบด้วยเวคเตอร์ของความสามารถและเวคเตอร์ของพารามิเตอร์ของข้อสอบทุกข้อ แต่เนื่องจาก F ไม่เป็นสมการเส้นตรง ในการคำนวณค่าพารามิเตอร์ดังกล่าวจึงต้องใช้วิธีการประมาณค่าซ้ำ ๆ กันหลายครั้งจนกระทั่งค่าประมาณจะเป็นอันหนึ่งอันเดียวกัน ในการประมาณค่าซ้ำ ๆ กันนี้จะใช้วิธีการของ Newton ซึ่งมีวิธีการดังนี้

$$\underset{\sim}{Z}^{r+1} = \underset{\sim}{Z}^{r} - (H^{-1} \frac{\partial F}{\partial \underset{\sim}{Z}}) \Big|_{\underset{\sim}{Z}^{r}} \qquad (18)$$

เมื่อ r และ r + 1 คือจำนวนครั้งที่ประมาณค่า $\underset{\sim}{Z}$

H คือแมทริกซ์ของอนุพันธ์ที่ 2 ของ F

กล่าวโดยสรุปการประมาณค่าพารามิเตอร์ดังกล่าวสามารถแบ่งได้เป็น 2 ตอนคือ ตอนแรกจะประมาณค่าความสามารถของผู้สอบ N คน ด้วยสมการเส้นตรง N สมการ หลังจากนี้จึงทำการปรับค่าความสามารถของผู้สอบให้มีค่าเฉลี่ยเป็น 0 และมีส่วนเบี่ยงเบนมาตรฐานเป็น 1 ในตอนที่ 2 จะเป็นการประมาณค่าพารามิเตอร์ของข้อสอบ n ข้อ ด้วยสมการเส้นตรง 3n สมการ การคำนวณค่าพารามิเตอร์ทั้ง 2 ตอนที่กล่าวมานี้ สามารถอธิบายขั้นตอน (Wood, Wingersky and Lord 1976 : 11) ได้ดังนี้

ขั้นที่ 1 กำหนดให้ค่าอำนาจจำแนกและค่าการเดาของข้อสอบเป็น 1 และ 0.15 ตามลำดับ หลังจากนั้นจึงประมาณค่าความยากของข้อสอบแต่ละข้อ และความสามารถของผู้สอบแต่ละคน

ขั้นที่ 2 นำค่าความสามารถของผู้สอบทุกคนที่คำนวณได้ในขั้นที่ 1 ไปประมาณค่าความยาก ค่าอำนาจจำแนก และค่าการเดาของข้อสอบทุกข้อ

ขั้นที่ 3 นำค่าอำนาจจำแนกและค่าการเดาที่ประมาณได้ในขั้นที่ 2 ไปประมาณค่าความสามารถของผู้สอบทุกคน และค่าความยากของข้อสอบแต่ละข้อ

ขั้นที่ 4 นำค่าความสามารถของผู้สอบแต่ละคนและค่าการเดาของข้อสอบแต่ละข้อไปประมาณค่าอำนาจจำแนกและค่าความยากของข้อสอบแต่ละข้อ

การประมาณค่าพารามิเตอร์ของข้อสอบนี้คอมพิวเตอร์โปรแกรม LOGIST จะประมาณค่าเริ่มต้นครั้งแรกของพารามิเตอร์ทุกตัวในขั้นที่ 1 ก่อน แล้วจึงประมาณค่าความ

สามารถของผู้สอบให้เป็นค่ามาตรฐาน โดยมีค่าเฉลี่ยเป็น 0 ส่วนเบี่ยงเบนมาตรฐานเป็น 1 แล้วทำการประมาณค่าพารามิเตอร์ของข้อสอบ และเมื่อประมาณค่าพารามิเตอร์ของข้อสอบได้แล้วจะกลับไปประมาณค่าความสามารถของผู้สอบอีก การประมาณค่าพารามิเตอร์ทุกครั้งจะประกอบด้วยการคำนวณ 4 ขั้นตอนดังกล่าว โดยในครั้งต่อไปต้องคำนวณค่าเริ่มต้นเหมือนครั้งที่ 1 แต่จะใช้ค่าที่คำนวณได้ในครั้งก่อนหน้านั้นเป็นค่าเริ่มต้นทุกครั้ง และในการประมาณค่าทุกครั้งค่าความสามารถของผู้สอบและค่าพารามิเตอร์ของข้อสอบจะถูกประมาณค่าใหม่เสมอ และหยุดการประมาณค่าหลังจากประมาณค่าพารามิเตอร์ทั้งหมดใน 4 ขั้น เมื่อได้ค่าพารามิเตอร์ของข้อสอบแต่ละข้อแล้วก็จะนำไปเขียนกราฟแสดงโค้งลักษณะข้อสอบ ซึ่งโค้งลักษณะข้อสอบนี้จะแสดงความสัมพันธ์ระหว่างระดับความสามารถกับโอกาสในการตอบข้อสอบถูก

ข้อจำกัดของโปรแกรมดังกล่าวก็คือ จะไม่มีการประมาณค่าความสามารถของผู้สอบที่ตอบข้อสอบถูกทุกข้อหรือตอบผิดทุกข้อ หรือผู้ที่ตอบข้อสอบเพียง 1/3 ของข้อสอบทั้งหมดหรือน้อยกว่า (Laksana 1979 : 20)

<u>วิธีการวิเคราะห์ความลำเอียงของข้อสอบ</u>

วิธีการวิเคราะห์ความลำเอียงของข้อสอบอาจแบ่งเป็น 2 วิธีใหญ่ ๆ คือ การวิเคราะห์ความลำเอียงของข้อสอบโดยใช้เกณฑ์ภายนอกและเกณฑ์ภายใน (Rudner, Getson, and Knight 1980 b) รายละเอียดของแต่ละวิธีเป็นดังนี้

<u>การวิเคราะห์ความลำเอียงของข้อสอบโดยใช้เกณฑ์ภายนอก</u>

การวิเคราะห์ความลำเอียงของข้อสอบโดยวิธีการนี้สามารถวิเคราะห์ได้ทั้งรายข้อและรายฉบับ วิธีการวิเคราะห์จะเขียนกราฟแสดงความสัมพันธ์ระหว่างคะแนนจากตัวแปรเกณฑ์

ภายนอกกับตัวแปรทำนาย วิธีนี้มีจุดมุ่งหมายที่จะวิเคราะห์ความถดถอยของตัวแปรทั้งสอง แล้วทำการเปรียบเทียบค่าความชัน (Slope) ค่าการตัดแกน (Intercept) ของเส้นกราฟ ในการวิเคราะห์ความลำเอียงของข้อสอบทั้งฉบับจะใช้คะแนนรวมเป็นตัวแปรทำนาย แต่ถ้าเป็นการวิเคราะห์รายข้อจะใช้ค่าความยาก (p) ของข้อสอบแต่ละข้อเป็นตัวแปรทำนาย ส่วนตัวแปรเกณฑ์ภายนอกจะใช้คะแนนรวมหรือเกรดเกี่ยวกับผลสัมฤทธิ์ของงานบางอย่างที่ให้ทำ (Cronbach 1970 ; Anastasi 1976)

การวิเคราะห์ความลำเอียงของข้อสอบด้วยวิธีการนี้จึงเป็นการเปรียบเทียบเส้นกราฟระหว่างกลุ่มผู้สอบ โดยถ้าเส้นกราฟดังกล่าวมีค่าความชันและค่าการตัดแกนแตกต่างกันในแต่ละกลุ่มแล้ว ข้อสอบข้อนั้นหรือแบบสอบฉบับนั้นก็จะมีความลำเอียงต่อกลุ่มผู้สอบที่มีค่าการตัดแกน และค่าความชันมากกว่า (Haebara 1979)

การวิเคราะห์ความลำเอียงของข้อสอบตามวิธีการที่กล่าวมานี้มีจุดอ่อนตรงที่ในทางปฏิบัติแล้วเป็นการยากมากที่จะหาตัวแปรเกณฑ์ภายนอกที่มีความตรงเชิงพยากรณ์และมีความยุติ-ธรรม ซึ่งหากตัวแปรเกณฑ์ภายนอกขาดคุณสมบัติดังกล่าว จะทำให้ผลการวิเคราะห์ความลำเอียงของข้อสอบขาดความสมบูรณ์และถูกต้องเท่าที่ควร จากความยุ่งยากในการสร้างเกณฑ์ภายนอกที่ไม่ลำเอียง นักวัดทั้งหลายจึงหันมาตรวจสอบโครงสร้างภายในของแบบสอบ โดยดูว่าคะแนนจากแบบสอบมีความหมายเชิงจิตวิทยาอย่างเดียวกันในแต่ละกลุ่มย่อย ๆ หรือไม่ วิธีการดังกล่าวนี้จะใช้ข้อมูลผลการตอบคำถามในแบบสอบเท่านั้น จึงเรียกวิธีการนี้ว่าการวิเคราะห์ความลำเอียงของข้อสอบโดยใช้เกณฑ์ภายใน ซึ่งจะได้กล่าวถึงรายละเอียดต่อไป

<u>การวิเคราะห์ความลำเอียงของข้อสอบโดยใช้เกณฑ์ภายใน</u>

การวิเคราะห์ความลำเอียงของข้อสอบโดยวิธีนี้ ยึดหลักการที่ว่าข้อสอบที่ผู้สอบที่มีระดับความสามารถเท่ากันแต่อยู่ต่างกลุ่มกันมีคะแนนจริงผลการสอบแตกต่างกันเป็นข้อสอบที่มีความลำเอียง การวิเคราะห์ความลำเอียงโดยวิธีนี้อาจแบ่งได้เป็น 3 วิธีใหญ่ ๆ (Laksana and Coffman 1980) ดังนี้

1. การวัดความเบี่ยงเบนสัมพัทธ์ (Relative Deviation) ของข้อสอบแต่ละข้อจากแนวโน้มเข้าสู่ส่วนกลาง วิธีนี้มีข้อตกลงเบื้องต้นว่าข้อสอบส่วนใหญ่ในแบบสอบมีความเป็นเอกพันธ์ (Homogeneous) ในการวัดความสามารถใด ความสามารถหนึ่งในผู้สอบต่างกลุ่มกัน ข้อสอบข้อใดที่เบี่ยงเบนไปจากส่วนกลางมากกว่าที่คาดหวังไว้ก็อาจตั้งข้อสงสัยได้ว่าเป็นข้อสอบที่มีความลำเอียง วิธีการหนึ่งที่ใช้กันบ่อย ๆ ก็คือการวิเคราะห์ความแปรปรวน (ANOVA) ข้อสอบที่ลำเอียงคือข้อสอบที่ Interaction Effect ระหว่างข้อกับกลุ่มผู้สอบมีนัยสำคัญทางสถิติ ส่วนอีกวิธีหนึ่งได้แก่ Bivariate Plot วิธีการนี้ค่าพารามิเตอร์เกี่ยวกับความยาก เช่นค่า p ค่า Arcsin ค่า Delta ค่า b จากกลุ่มหนึ่ง จะ Plot คู่กับค่าพารามิเตอร์จากอีกกลุ่มหนึ่ง ข้อสอบข้อใดที่เบี่ยงเบนไปจากแกนหลัก (Principal Axis) มาก ถือว่าเป็นข้อสอบที่มีความลำเอียง

2. การประเมินความตรงเชิงโครงสร้าง (Construct Validity) ของแบบสอบ วิธีการหนึ่งที่ใช้กันมากคือการวิเคราะห์องค์ประกอบ ดัชนีความลำเอียงของข้อสอบตามวิธีการนี้ก็คือความแตกต่าง (Discrepancy) ของ Factor Loading จากผู้สอบต่างกลุ่มกัน หรือความแตกต่างอย่างมีนัยสำคัญระหว่างค่าเฉลี่ยของ Factor Scores จากแต่ละกลุ่มที่นำมาเปรียบเทียบกัน ดัชนีความลำเอียงทั้งสองที่กล่าวมาเป็นตัวชี้ว่าข้อสอบไม่ได้วัดสิ่งเดียวกันในผู้สอบต่างกลุ่มกัน

3. การประมาณค่าโอกาสในการตอบข้อสอบแต่ละข้อถูก วิธีการที่ใช้กันทั่ว ๆ ไปคือ วิธีไคสแควร์ (Chi - square) และวิธีโค้งลักษณะข้อสอบ (Item Characteristic Curve : ICC) วิธีการทั้งสองนี้มีความคล้ายคลึงกันในแง่การใช้โอกาสในการตอบถูกที่แตกต่างกันจากผู้สอบ 2 กลุ่ม หรือมากกว่าเป็นดัชนีความลำเอียงของข้อสอบ ส่วนข้อแตกต่างระหว่างวิธีการทั้งสองก็คือวิธีไคสแควร์จะประมาณความสามารถของผู้สอบโดยใช้คะแนนดิบ ส่วนวิธีโค้งลักษณะข้อสอบจะประมาณความสามารถของผู้สอบจากคุณลักษณะแฝง (Latent Trait)

ในการวิจัยครั้งนี้ผู้วิจัยจะเลือกใช้วิธีการวิเคราะห์ความลำเอียงของข้อสอบโดยไม่ใช้เกณฑ์ภายนอก ทั้งนี้เพราะเห็นว่าจะมีปัญหาน้อยกว่าวิธีการวิเคราะห์ความลำเอียงโดยใช้เกณฑ์ภายนอก กล่าวคือในการหาตัวแปรเกณฑ์ภายนอกให้มีความตรงเชิงพยากรณ์ และยุติธรรมนั้นเป็นสิ่งที่ทำได้ลำบากมาก และหากตัวแปรภายนอกขาดคุณสมบัติดังกล่าวจะทำให้ผลการวิเคราะห์ความลำเอียงมีความคลาดเคลื่อนเกิดขึ้นได้ ด้วยเหตุผลดังที่ได้กล่าวมาผู้วิจัยจึงเลือกใช้วิธีวิเคราะห์ความลำเอียงโดยไม่ใช้เกณฑ์ภายนอก 4 วิธี ซึ่งได้แก่ การแปลงค่าความยากของข้อสอบ (TID) การวิเคราะห์ความแปรปรวน (ANOVA) การวิเคราะห์ด้วยโค้งลักษณะข้อสอบที่มีพารามิเตอร์ 1 ตัว (ICC 1) และการวิเคราะห์ด้วยโค้งลักษณะข้อสอบที่มีพารามิเตอร์ 3 ตัว (ICC 3) เป็นกรอบทฤษฎีในการวิเคราะห์ความลำเอียงของข้อสอบ ดังรายละเอียดของแต่ละวิธีวิเคราะห์ ดังนี้

<u>การแปลงค่าความยากของข้อสอบ</u> (TID)

การหาความลำเอียงของข้อสอบตามวิธีการนี้จะต้องทำการคำนวณหาค่าความยาก (p) ของข้อสอบแต่ละข้อจากผู้สอบแต่ละกลุ่ม ค่าความยากของข้อสอบแต่ละข้อจะมี 2 ค่า ตามจำนวนกลุ่มที่แบ่ง ในการเปรียบเทียบค่าความยากนี้จะใช้วิธีการแปลงค่าความยาก (p - values) ไปเป็นค่าซี (z - values) แล้วทำการแปลงค่าซีเป็นค่าเดลต้า ($\triangle$ - values) อีกต่อหนึ่งโดยการใช้สมการ $\triangle = 4z + 13$ นำค่าเดลต้าแต่ละคู่มา Plot ลงบน Bivariate

Graph โดยให้แนวนอนแทนค่าเดลต้าสำหรับกลุ่มหนึ่ง และให้แนวตั้งแทนค่าเดลต้าของอีกกลุ่มหนึ่ง จุดต่าง ๆ ที่ปรากฏนี้จะอยู่ในรูปลักษณะที่เป็น Ellipse แล้วทำการคำนวณหาระยะทางที่จุดต่าง ๆ เหล่านี้ห่างจากเส้นแกนหลัก ข้อสอบข้อใดที่ห่างจากเส้นแกนหลัก $\geq \pm 3\ S_d$ ถือว่าเป็นข้อสอบที่มีความลำเอียง (Angoff cited in Berk 1982 : 107)

<u>การวิเคราะห์ความแปรปรวน</u> (ANOVA)

การวิเคราะห์ความแปรปรวนสามารถนำมาประยุกต์ใช้ในการหาความลำเอียงของข้อสอบได้ โดยถือว่าข้อสอบที่ไม่วัดความสามารถเดียวกันจากผู้สอบต่างกลุ่มกันเป็นข้อสอบที่มีความลำเอียง ซึ่งหมายความว่าถ้ามีข้อสอบที่ง่ายสำหรับกลุ่มหนึ่ง แต่ยากสำหรับอีกกลุ่มหนึ่ง ก็อาจตั้งข้อสงสัยได้ว่าข้อสอบข้อนั้นเป็นข้อสอบที่มีความลำเอียงและยังเป็นการแสดงให้เห็นอีกว่าข้อสอบข้อนั้นไม่ได้วัดความสามารถเดียวกันในผู้สอบแต่ละกลุ่ม วิธีการวิเคราะห์นี้จะจัดข้อสอบและกลุ่มผู้สอบลงใน Two - factor Design แล้วทำการทดสอบ Overall Interaction Effect ระหว่างข้อสอบกับกลุ่มผู้สอบ

การวิเคราะห์ความลำเอียงโดยวิธีการนี้มีข้อตกลงว่าแบบสอบฉบับนั้นมีลักษณะเป็นเอกพันธ์ นั่นคือแบบสอบฉบับนั้นต้องวัดความสามารถหรือคุณลักษณะเพียงลักษณะเดียว และยังมีข้อตกลงเบื้องต้นอีก 2 ประการคือ

1. ความแปรปรวนของข้อสอบต้องเท่ากัน
2. ความแปรปรวนร่วมของข้อสอบต้องเท่ากัน

Rudner (1977 cited in Laksana 1977) ได้ชี้ให้เห็นว่าในการศึกษาเกี่ยวกับการตอบสนองต่อข้อสอบนั้นมีแหล่งความแปรปรวนที่ควรพิจารณา ดังนี้

1. ข้อสอบ (Items) มีข้อสอบบางข้อที่ยากกว่าข้ออื่น

2. กลุ่มผู้สอบ (Groups) กลุ่มผู้สอบหนึ่งอาจจะถูกวัดคุณลักษณะที่สูงกว่าอีกกลุ่มหนึ่ง

3. ความแปรปรวนภายในกลุ่ม (Subject Within Groups) ผู้สอบในแต่ละกลุ่มมีความสามารถแตกต่างกัน

4. ผลของปฏิสัมพันธ์ระหว่างข้อสอบกับกลุ่มผู้สอบ

วิธีการหาความลำเอียงของข้อสอบโดยวิธีการวิเคราะห์ความแปรปรวนนั้นมีขั้นตอนดังนี้

1. จัดข้อสอบ (Items) และกลุ่มผู้สอบ (Subgroups) ลงในตัวแปรผลการตอบข้อสอบ และกลุ่มผู้สอบ

2. คำนวณสัดส่วนในการตอบข้อสอบแต่ละข้อถูกระหว่างกลุ่มชายและหญิง

3. ทดสอบ Overall Interaction Effect ระหว่างข้อ (Item) กับกลุ่มผู้สอบ (Subgroups) หากพบว่า Interaction Effect มีนัยสำคัญทางสถิติ (Significant) แสดงว่าแบบสอบฉบับนั้นมีความลำเอียง

4. หากการทดสอบในข้อ 3 มีนัยสำคัญทางสถิติให้ทำการทดสอบต่อเป็นรายข้อว่ามีข้อใดบ้างที่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ ข้อที่แตกต่างกันอย่างมีนัยสำคัญทางสถิติก็คือข้อที่มีความลำเอียงนั่นเอง

<u>การวิเคราะห์ด้วยโค้งลักษณะข้อสอบ</u> (ICC)

ทฤษฎีการตอบสนองต่อข้อสอบเป็นทฤษฎีทางการทดสอบที่มีข้อตกลงเบื้องต้นที่ค่อนข้างจะเข้มงวดกว่าทฤษฎีการวัดแบบคลาสสิค ทฤษฎีนี้มุ่งที่จะอธิบายหรือทำนายความสามารถของผู้สอบซึ่งเป็นคุณลักษณะที่ไม่สามารถวัดได้โดยตรง โดยอาศัยผลการสอบหรือคะแนนที่ผู้สอบแสดงออกมาให้เห็น ในการอธิบายความสัมพันธ์ระหว่างโอกาสที่จะตอบข้อสอบถูกกับระดับความสามารถนั้นจะใช้โค้งลักษณะข้อสอบ และลักษณะของโค้งนี้จะถูกกำหนดโดยค่าพารามิเตอร์ที่เกี่ยวกับความยาก ค่า

อำนาจจำแนก และค่าการเดาในกรณีที่ต้องการประมาณค่าพารามิเตอร์ 3 ตัว ส่วนในกรณีที่ต้องการประมาณค่าพารามิเตอร์เพียงตัวเดียวโค้งลักษณะข้อสอบจะถูกกำหนดโดยค่าความยาก โค้งลักษณะข้อสอบนี้จะเป็นตัวชี้ถึงโอกาสที่ผู้สอบที่มีความสามารถระดับหนึ่งจะตอบข้อสอบได้ถูก และโอกาสที่จะตอบข้อสอบถูกนี้จะเพิ่มขึ้นตามลำดับความสามารถที่เพิ่มขึ้น

วิธีการวิเคราะห์ด้วยโค้งลักษณะข้อสอบนี้จะถือว่าข้อสอบมีความลำเอียงก็ต่อเมื่อผู้สอบที่มีระดับความสามารถเท่ากันแต่อยู่ต่างกลุ่มกันมีโอกาสจะตอบถูกไม่เท่ากัน ซึ่งความลำเอียงของข้อสอบนี้อาจแสดงให้เห็นได้ด้วยโค้งลักษณะข้อสอบ กล่าวคือ ถ้าผู้สอบต่างกลุ่มกันมีโค้งลักษณะข้อสอบข้อเดียวกันแตกต่างกัน ข้อสอบข้อนั้นถือได้ว่าเป็นข้อสอบที่มีความลำเอียง ในทำนองกลับกันหากโค้งลักษณะข้อสอบของผู้สอบต่างกลุ่มกันมีลักษณะโค้งเหมือนกันก็ย่อมแสดงว่าข้อสอบข้อนั้นไม่ลำเอียง ส่วนวิธีการในการประมาณค่าโค้งลักษณะข้อสอบที่มีพารามิเตอร์ 3 ตัวนั้น จะใช้คอมพิวเตอร์โปรแกรม LOGIST และใช้โปรแกรม BICAL ในการประมาณค่าโค้งลักษณะข้อสอบที่มีพารามิเตอร์ 1 ตัว

วิธีการวิเคราะห์ความลำเอียงของข้อสอบแต่ละวิธีที่กล่าวมาต่างก็มีจุดเด่นและจุดด้อยแตกต่างกัน ซึ่งอาจสรุปจุดเด่นและจุดด้อยของแต่ละวิธีได้ดังนี้

<u>วิธีแปลงค่าความยาก</u> (Angoff cited in Berk 1982 : 106)

<u>จุดเด่น</u>	<u>จุดด้อย</u>
1. ง่ายต่อการคำนวณ	1. ค่าเดลต้าไม่คงที่
2. ค่าใช้จ่ายไม่สูง	
3. อธิบายง่าย	
4. ไม่ต้องการข้อมูลมาก	

<u>วิธีวิเคราะห์ความแปรปรวน</u> (Laksana and Coffman 1980 : 30 - 33)

จุดเด่น	จุดด้อย
1. ใช้ผลวิจัยน้อย	1. เป็นวิธีที่ผลการวิเคราะห์ยังไม่ครอบคลุม
2. เป็นวิธีที่ใช้ในทางปฏิบัติ (Practical)	2. ค่าความยาก (p) จะขึ้นอยู่กับความสามารถของผู้สอบแต่ละกลุ่ม

<u>วิธีโค้งลักษณะข้อสอบที่มีพารามิเตอร์ 1 ตัว</u> (Ironson cited in Berk 1982 : 154)

จุดเด่น	จุดด้อย
1. การประมาณค่าพารามิเตอร์และค่าความสามารถเป็นอิสระต่อกัน	1. ข้อตกลงเกี่ยวกับการเดาและอำนาจจำแนกไม่เป็นจริงในทางปฏิบัติ
2. ผลวิจัยน้อยกว่าเมื่อเทียบกับวิธี 3 พารามิเตอร์	2. ใช้ผลวิจัยขนาดใหญ่เมื่อเทียบกับวิธีการไคสแควร์ และวิธีแปลงค่าความยาก
3. สามารถหาข้อสอบที่เหมาะสม (Fit) กับวิธีการวิเคราะห์ได้	
4. มีคอมพิวเตอร์โปรแกรม BICAL	

<u>วิธีโค้งลักษณะข้อสอบที่มีพารามิเตอร์ 3 ตัว</u> (Ironson cited in Berk 1982 : 154)

จุดเด่น	จุดด้อย
1. การไม่แปรเปลี่ยนของค่าพารามิเตอร์	1. การวิเคราะห์ด้วยคอมพิวเตอร์ค่อนข้างซับซ้อน
2. ค่าพารามิเตอร์ a และ c เป็นจริงและเหมาะสมกับข้อมูลมากกว่า	2. ต้องใช้ผลวิจัยขนาดใหญ่มาก

3. มีคอมพิวเตอร์โปรแกรม LOGIST	3. ในการคำนวนโปรแกรม LOGIST นั้นค่าใช้จ่ายแพงมาก
4. มีข้อได้เปรียบในการศึกษา วิจัย ในหลาย ๆ เรื่อง	4. ความถูกต้องของประมาณค่า a และ c ยังค่อนข้างต่ำ

<u>ดัชนีความลำเอียงของข้อสอบ</u>

เนื่องจากการหาดัชนีความลำเอียงของข้อสอบจากวิธีการแปลงค่าความยากของข้อสอบ วิธีวิเคราะห์ความแปรปรวน และวิธีโค้งลักษณะข้อสอบที่มีพารามิเตอร์ 1 ตัวนั้น เป็นวิธีการที่ไม่ซับซ้อนอะไรมากนัก ทั้งนี้เพราะเป็นการนำค่าสถิติที่ใช้ทดสอบนัยสำคัญทางสถิติมาเป็นค่าดัชนีความลำเอียงของข้อสอบ ดังนั้นจึงจะไม่ขอกล่าวถึงรายละเอียด ในที่นี้จะขอกล่าวเฉพาะวิธีการหาดัชนีความลำเอียงของข้อสอบที่วิเคราะห์ด้วยวิธีโค้งลักษณะข้อสอบที่มีพารามิเตอร์ 3 ตัว ดังนี้

Rudner (1977) ได้เสนอวิธีการคำนวณหาดัชนีความลำเอียงของข้อสอบ โดยการหาความแตกต่างของพื้นที่ที่คำนวณจากโค้งลักษณะข้อสอบจากผู้สอบแต่ละกลุ่ม ดังนี้

$$\phi_i = \int_{-\infty}^{\infty} \{ | P(U_i = 1|\theta) - P'(U_i = 1|\theta) | \} \, d\theta \qquad (19)$$

ค่าที่คำนวณได้จากสมการ (19) นี้เป็นค่าสมบูรณ์ ค่านี้จึงบอกแต่เพียงว่า ข้อสอบมีดัชนีความลำเอียงเป็นเท่าใด แต่ไม่อาจบอกได้ว่าข้อสอบมีความลำเอียงต่อกลุ่มใด และหากผู้วิจัยมีความประสงค์ที่จะทราบรายละเอียดที่ลึกไปกว่านี้ เช่นข้อสอบข้อนั้นลำเอียงต่อผู้สอบกลุ่มใดในช่วงระดับความสามารถใด สมการ (19) ไม่สามารถบอกได้เช่นกัน ข้อสอบข้อเดียวกันอาจลำเอียงต่อผู้สอบกลุ่มหนึ่งในระดับความสามารถหนึ่ง แต่อาจลำเอียงต่อผู้สอบอีกกลุ่มหนึ่งในอีกระดับความสามารถหนึ่ง ดังนั้นหากผู้วิจัยต้องการที่จะทราบรายละเอียดดังกล่าว ผู้วิจัยก็จำเป็นที่จะต้องคำนวณหาพื้นที่เป็นช่วง ๆ ไปตามระดับความสามารถที่ต้องการ ส่วนการที่จะแบ่งข้อสอบ

ข้อหนึ่งออกเป็นกี่ช่วงตามระดับความสามารถนั้น ให้ดูจากโค้งลักษณะข้อสอบที่ตัดกัน เช่นถ้าโค้งลักษณะข้อสอบของผู้สอบ 2 กลุ่มตัดกัน 1 แห่ง ก็อาจคำนวณพื้นที่ 2 ช่วง หรือถ้าตัดกัน 2 แห่ง ก็อาจคำนวณพื้นที่ 3 ช่วง เป็นต้น ข้อสอบข้อหนึ่ง ๆ จึงอาจมีดัชนีความลำเอียงมากกว่าหนึ่งค่า ทั้งนี้ขึ้นอยู่กับจำนวนแห่งที่โค้งลักษณะข้อสอบตัดกัน สำหรับสมการที่ใช้คำนวณหาพื้นที่ตามช่วงระดับความสามารถหนึ่ง ๆ เป็นดังนี้

$$A_j = \int_s^m P_i(\theta)\, d\theta$$

$$= c_i \theta \Big|_s^m + \{(1-c_i)/Da_i\} \ln [1+\exp(Da_i(\theta - b_i))]\Big|_s^m \qquad (20)$$

เมื่อ A_j คือพื้นที่ของโค้งลักษณะข้อสอบของผู้สอบกลุ่ม j

s และ m คือช่วงระดับความสามารถที่ต้องการ

ฉะนั้นดัชนีความลำเอียงของข้อสอบในช่วงความสามารถ s ถึง m จึงมีค่าเท่ากับผลต่างของพื้นที่ A_j

Rudner (1977) ได้เสนอแนะวิธีการแปลความหมายของดัชนีความลำเอียงไว้ดังนี้

ดัชนีความลำเอียง	ความหมายของความลำเอียง
น้อยกว่า .40	ต่ำ
.40 - .70	ปานกลาง
มากกว่า .70	สูง

<u>รายงานการวิจัยที่เกี่ยวข้องกับการเปรียบเทียบวิธีการในการวิเคราะห์ความลำเอียง</u>

ในประเทศไทยยังไม่มีผู้ใดทำวิจัยเกี่ยวกับการเปรียบเทียบวิธีการในการวิเคราะห์ความลำเอียง ผลการวิจัยเกี่ยวกับการเปรียบเทียบวิธีการในการวิเคราะห์ความลำเอียงส่วนมากเป็นผลการวิจัยที่ทำกันในต่างประเทศ ซึ่งผู้วิจัยจะขอนำมาเสนอดังนี้

Rudner และ Convey (1978 cited in Subkoviak, et al. 1984 : 51) ได้ทำการวิเคราะห์ความลำเอียงของข้อสอบโดยใช้วิธีแปลงค่าความยากของข้อสอบ วิธีไคสแควร์ วิธีโค้งลักษณะข้อสอบ และวิธีวิเคราะห์องค์ประกอบ ข้อมูลในการวิจัยครั้งนี้เป็นข้อมูลจริงจากการใช้แบบสอบมาตรฐาน ผลการวิจัยปรากฏว่าวิธีโค้งลักษณะข้อสอบ วิธีไคสแควร์ และวิธีแปลงค่าความยากเป็นวิธีที่เหมาะสม

Ironson และ Subkoviak (1979) ได้ศึกษาถึงวิธีการในการวิเคราะห์ความลำเอียงของข้อสอบ โดยใช้วิธีแปลงค่าความยากของข้อสอบ วิธีไคสแควร์ วิธีโค้งลักษณะข้อสอบ และวิธีพอยท์ไบซีเรียล ผู้วิจัยได้นำแบบสอบมาตรฐานไปทดสอบกับชาวผิวขาวและผิวดำ แล้วนำข้อมูลที่ได้มาทำการวิเคราะห์หาสหสัมพันธ์ระหว่างวิธีการวิเคราะห์ความลำเอียงแบบต่าง ๆ ผลการวิจัยพบว่าวิธีโค้งลักษณะข้อสอบเป็นวิธีที่ดีที่สุด รองลงมาได้แก่ วิธีไคสแควร์ และวิธีแปลงค่าความยาก ตามลำดับ

Laksana (1979) ได้ประยุกต์ใช้วิธีโค้งลักษณะข้อสอบที่มี 3 พารามิเตอร์และวิธีวิเคราะห์ความแปรปรวนในการวิเคราะห์แบบสอบ Iowa Tests of Basic Skills (ITBS) ในแบบสอบคำศัพท์และแบบสอบคณิตศาสตร์ระดับ 7 เพื่อหาความลำเอียงทางเพศและผิว ผลการศึกษาพบว่าวิธีการวิเคราะห์ความแปรปรวนเป็นวิธีที่ใช้ได้ในทางปฏิบัติ แต่เป็นวิธีที่ไม่ครอบคลุม วิธีโค้งลักษณะข้อสอบเป็นวิธีที่ครอบคลุมกว่า แต่ไม่เหมาะสมในทางปฏิบัติเนื่องจากต้องใช้กลุ่มตัวอย่างขนาดใหญ่

Intasuwan (1979) ได้ใช้วิธีในการวิเคราะห์ความลำเอียงของแบบสอบ International Association for the Evaluation of Educational Achievement (IEA) วิธีการวิเคราะห์ดังกล่าว ได้แก่ วิธีโค้งลักษณะข้อสอบที่มีพารามิเตอร์ 3 ตัว วิธี Rasch และวิธีไคสแควร์ กลุ่มตัวอย่างในการวิจัยในครั้งนี้เป็นวัยรุ่นชาวอังกฤษ อเมริกัน และนิวซีแลนด์ ผลการวิจัยพบว่า สหพันธ์ระหว่างวิธีทั้ง 3 มีค่าอยู่ระหว่าง .51 - .98 โดยวิธีวิเคราะห์แบบไคสแควร์และวิธี Rasch มีค่าสหสัมพันธ์กันสูงถึง .98

Shepard และคณะ (1981 cited in Subkoviak, et al. 1984 : 50) ได้ทำการวิเคราะห์ความลำเอียงของข้อสอบโดยใช้วิธีการ 6 แบบคือ วิธีแปลงค่าความยาก วิธีไคสแควร์ 2 วิธี วิธีโค้งลักษณะข้อสอบ 2 วิธี และวิธีพอยท์ไบซีเรียล แบบสอบที่ใช้เป็นแบบสอบมาตรฐาน สำหรับกลุ่มตัวอย่างของการวิจัยครั้งนี้เป็นชาวผิวขาว ผิวดำ และ Chicano ผลการวิจัยพบว่าวิธีโค้งลักษณะข้อสอบที่มีพารามิเตอร์ 3 ตัว เป็นวิธีที่ดีที่สุด รองลงมาได้แก่ หนึ่งในสองของวิธีวิเคราะห์ด้วยไคสแควร์ อย่างไรก็ตามวิธีแปลงค่าความยากของข้อสอบเป็นวิธีที่ใช้ได้ในทางปฏิบัติ ทั้งนี้เพราะวิธีแปลงค่าความยากของข้อสอบง่ายกว่าวิธีโค้งลักษณะข้อสอบที่มีพารามิเตอร์ตัวเดียว แต่ผลการวิเคราะห์ปรากฏว่าสหสัมพันธ์ระหว่างวิธีทั้งสองค่อนข้างสูง

Ironson และ Craig (1982 cited in Subkoviak, et al, 1984 : 50) ได้ใช้วิธีวิเคราะห์ความลำเอียงของข้อสอบ 5 วิธีคือวิธีแปลงค่าความยากของข้อสอบ วิธีวิเคราะห์ด้วยไคสแควร์ 2 วิธี และวิธีโค้งลักษณะข้อสอบ 2 วิธี ในการสร้างข้อสอบนั้น ผู้วิจัยตั้งใจที่จะสร้างข้อสอบให้มีความลำเอียงต่อผู้ชาย หลังจากนั้นผู้วิจัยได้ให้ผู้สอบทั้งชายและหญิงประมาณค่า (Rate) ความลำเอียงของข้อสอบแต่ละข้อออกมาเป็น 5 ระดับ เมื่อได้ข้อมูลมาแล้วผู้วิจัยทำการวิเคราะห์ความลำเอียงของข้อสอบแต่ละข้อด้วยวิธีการ 5 วิธีดังที่ได้กล่าวมา ผลการวิจัยพบว่าการประมาณค่าความลำเอียงของข้อสอบกับการวิเคราะห์หาความลำเอียงของข้อสอบทั้ง 5 วิธี มีความสัมพันธ์กันในระดับสูง นอกจากนี้ยังพบว่าสหพันธ์ระหว่างวิธีการวิเคราะห์หาความลำเอียงของข้อสอบทั้ง 5 วิธีมีค่าสูงเช่นกัน

Subkovlak และคณะ (1984) ได้ทำการเปรียบเทียบวิธีการวิเคราะห์ความลำเอียงของข้อสอบ 3 วิธี เครื่องมือที่ใช้ในการวิจัยเป็นแบบสอบคำศัพท์แบบเลือกตอบ 4 ตัวเลือก จำนวน 50 ข้อ ในแบบสอบนี้จะประกอบไปด้วยศัพท์ภาษาอังกฤษมาตรฐาน (Standard English Vocabulary) จำนวน 40 ข้อ และเป็นศัพท์เกี่ยวกับคำสแลง (Slang) ของชาติผิวดำ 10 ข้อ ผู้วิจัยได้นำแบบสอบฉบับนี้ไปทดสอบกับนักเรียนผิวดำจำนวน 1,008 คน นักเรียนผิวขาวจำนวน 1,021 คน ผลการวิจัยปรากฏว่าวิธีโค้งลักษณะข้อสอบที่มีพารามิเตอร์ 3 ตัวเป็นวิธีที่มีประสิทธิภาพมากที่สุด รองลงมาได้แก่ วิธีไคสแควร์ ส่วนวิธีแปลงค่าความยากเป็นวิธีที่มีข้อจำกัดแต่ก็สามารถนำไปใช้ในทางปฏิบัติได้ในกรณีที่มีข้อจำกัดเกี่ยวกับคอมพิวเตอร์

ส่วนการวิเคราะห์หาความลำเอียงของข้อสอบโดยใช้ข้อมูลสมมุติ (Simulate Data) มีผู้ศึกษากันดังนี้

Rudner และคณะ (1980 b) ได้ใช้ทฤษฎีโค้งลักษณะข้อสอบในการวิเคราะห์หาความลำเอียงของข้อสอบ ในการวิจัยครั้งนี้ผู้วิจัยได้สร้างข้อมูลสมมุติ (Simulate Data) ส่วนวิธีในการวิเคราะห์นั้นมี 7 วิธีดังนี้ วิธีแปลงค่าความยากของข้อสอบ 2 วิธี วิธีไคสแควร์ 2 วิธี และวิธีโค้งลักษณะข้อสอบ 3 วิธี ผลการวิจัยปรากฏว่าวิธีโค้งลักษณะข้อสอบที่มีพารามิเตอร์ 3 ตัวเป็นวิธีการหาความลำเอียงของข้อสอบจากข้อมูลสมมุติที่ดีที่สุด รองลงมาได้แก่หนึ่งในสองของวิธีการวิเคราะห์ด้วยไคสแควร์ ส่วนวิธีแปลงค่าความยากของข้อสอบเป็นทางเลือกที่ผู้วิจัยแนะนำให้ใช้ได้ในทางปฏิบัติ

Merz และ Grossen (1979 cited in Subkovlak, et al. 1984 : 51) ใช้ทฤษฎีโค้งลักษณะข้อสอบในการหาความลำเอียงของข้อสอบโดยใช้ข้อมูลสมมุติ ส่วนวิธีการวิเคราะห์ความลำเอียงที่ใช้คือวิธีแปลงค่าความยากของข้อสอบ วิธีไคสแควร์ วิธีโค้งลักษณะข้อสอบ 2 วิธี และวิธีวิเคราะห์องค์ประกอบ ผลการวิจัยปรากฏว่า วิธีแปลงค่าความยาก

ของข้อสอบเป็นวิธีการวิเคราะห์ความลำเอียงที่ดีที่สุด วิธีโค้งลักษณะข้อสอบที่มีพารามิเตอร์ 3 ตัว และวิธีไคสแควร์ต่างก็เป็นวิธีการที่ใช้ได้

จากผลการวิจัยที่กล่าวมาปรากฏว่าข้อค้นพบที่ได้มีทั้งที่สอดคล้องและไม่สอดคล้องกัน กล่าวคือสำหรับข้อมูลสมมุตินั้น Rudner และคณะพบว่าวิธีโค้งลักษณะข้อสอบที่มีพารามิเตอร์ 3 ตัวเป็นวิธีที่ดีที่สุด ส่วน Merz และ Grossen กลับพบว่าวิธีแปลงค่าความยากของข้อสอบเป็นวิธีที่ดีที่สุด แต่สำหรับข้อมูลจริง (Real Data) นั้นปรากฏว่าข้อค้นพบที่ได้ส่วนใหญ่ค่อนข้างจะสอดคล้องกัน (Ironson and Subkoviak ; Shepard, et al, ; Subkoviak, et al.) คือต่างก็พบว่าวิธีโค้งลักษณะข้อสอบเป็นวิธีที่เหมาะสมในทางทฤษฎี ถูกต้อง และตรงที่สุดในการวิเคราะห์ความลำเอียงของข้อสอบ แต่อย่างไรก็ตาม วิธีนี้เป็นวิธีที่มีความยุ่งยากซับซ้อนทางคณิตศาสตร์ ต้องใช้กลุ่มตัวอย่างขนาดใหญ่ และข้อสอบจำนวนมาก อีกทั้งค่าใช้จ่ายในการคำนวณด้วยคอมพิวเตอร์ค่อนข้างแพง จึงเป็นสาเหตุส่วนหนึ่งที่ทำให้ผู้วิจัยสนับสนุนให้ใช้วิธีการอื่นในการวิเคราะห์ความลำเอียงของข้อสอบแม้จะไม่เหมาะสมในเชิงทฤษฎีมากนัก แต่ก็เห็นว่าเป็นวิธีการที่ใช้ได้ในทางปฏิบัติ วิธีที่ได้รับการสนับสนุนให้ใช้วิเคราะห์ความลำเอียงของข้อสอบบ่อยที่สุดคือวิธีวิเคราะห์ความแปรปรวน

จากข้อค้นพบดังกล่าวทำให้ไม่อาจสรุปความรู้ที่แน่นอนเกี่ยวกับวิธีการวิเคราะห์ความลำเอียงของข้อสอบ และเพื่อเป็นการยืนยันหรือพิสูจน์ (Identified) ข้อความรู้ดังกล่าว ผู้วิจัยจึงต้องการที่จะเปรียบเทียบผลการวิเคราะห์ความลำเอียงของข้อสอบวิธีต่าง ๆ อีกครั้งหนึ่ง