บทที่ 2

วรรณคดีที่เกี่ยวข้อง

การศึกษาเป็นองค์ประกอบที่สำคัญในการพัฒนาประเทศและการศึกษาจะบรรลุเป้าหมายหรือไม่ ยังขึ้นอยู่กับองค์ประกอบหลายอย่าง เช่น หลักสูตร การเรียนการสอน การวัดผลการศึกษาฯลฯ (ชวาล แพรัตกุล, 2516 : 2-4) ในการวัดผลการศึกษานั้นจะถูกต้องเที่ยงตรงมากน้อยเพียงใดขึ้นอยู่กับคุณภาพของข้อสอบเป็นสำคัญ ฉะนั้นในปัจจุบันจึงได้มีการคิดค้นและนำเอาทฤษฎีใหม่ที่เรียกว่า ทฤษฎีการตอบข้อสอบ (Item Response Theory : IRT) มาใช้ในการตรวจสอบคุณภาพของข้อสอบ โดยมีความเชื่อว่าทฤษฎีนี้จะสามารถแก้ไขปัญหาที่เกิดจากทฤษฎีเดิม (The Classical Test Theory) ได้

ทฤษฎีเดิม (The Classical Test Theory)นั้น มีการวิเคราะห์หาคุณภาพของข้อสอบอยู่ 2 ขั้นตอน คือ

1. การวิเคราะห์หาคุณภาพของข้อสอบเป็นรายข้อ (Item Analysis) โดยพิจารณาคุณลักษณะที่สำคัญ 2 ประการของข้อสอบ (Mehrens and Ebel, 1969 : 325) คือ

1.1 ค่าความยากของข้อสอบ (Item difficulty)

1.2 ค่าอำนาจจำแนกของข้อสอบ (Discriminating power)

ค่าความยากของข้อสอบ (Item difficulty) หมายถึงสัดส่วนของผู้ที่เข้าสอบทั้งหมดที่ตอบข้อสอบถูก สัญลักษณ์ที่ใช้แทนค่าความยากคือ p ซึ่งระดับค่าความยากของข้อสอบจะมีค่าอยู่ระหว่าง 0 ถึง 1 โดยที่

ข้อสอบที่มีค่า p มาก หมายถึง ข้อสอบข้อนั้นมีสัดส่วนผู้ตอบถูกมาก แสดงว่าเป็นข้อสอบที่ง่าย

ข้อสอบที่มีค่า P น้อย หมายถึง ข้อสอบข้อนั้นมีสัดส่วนผู้ตอบถูกน้อย แสดงว่าเป็นข้อสอบที่ยาก

การพิจารณาค่าความยากของข้อสอบในการคัดเลือกข้อสอบที่ไม่เหมาะสมออก (Yen, 1979 : 120-122) ได้แก่ข้อสอบที่มีค่าความยากต่ำกว่า 0.20 หรือสูงกว่า 0.80 เนื่องจากข้อสอบเหล่านี้ไม่สามารถจำแนกความสามารถของผู้สอบออกจากกันได้

ค่าอำนาจจำแนกของข้อสอบ (Discriminating Power) หมายถึง คุณลักษณะของข้อสอบที่สามารถจำแนกผู้ที่ได้คะแนนสูงและผู้ที่ได้คะแนนต่ำในแต่ละข้อ สัญลักษณ์ที่ใช้แทนค่าอำนาจจำแนกคือ r ซึ่งระดับค่าอำนาจจำแนกจะมีค่าระหว่าง -1.00 ถึง +1.00 โดยที่

ข้อสอบที่มีค่า r เป็นบวก หมายถึง ผู้ที่ได้คะแนนรวมสูงจะมีแนวโน้มที่จะตอบข้อสอบข้อนั้นได้ถูก ส่วนผู้ที่ได้คะแนนรวมต่ำจะมีแนวโน้มที่จะตอบข้อสอบข้อนั้นผิด และถ้าค่า r เป็นบวกเข้าใกล้ 1 มากเท่าใด ก็แสดงว่าผู้ที่ได้คะแนนรวมสูงจะมีแนวโน้มที่จะตอบข้อนั้นถูกมากขึ้นเท่านั้น ส่วนผู้ที่ได้คะแนนรวมต่ำจะมีแนวโน้มที่จะตอบข้อสอบข้อนั้นผิดมากขึ้นเช่นกัน

ข้อสอบที่มีค่า r เป็นลบ หมายถึง ผู้ที่ได้คะแนนรวมสูงจะมีแนวโน้มที่จะตอบข้อสอบข้อนั้นผิด ส่วนผู้ที่ได้คะแนนรวมต่ำจะมีแนวโน้มที่จะตอบข้อสอบข้อนั้นได้ถูก และถ้าค่า r เป็นลบเข้าใกล้ 1 มากเท่าใด ก็แสดงว่าผู้ที่ได้คะแนนรวมสูงจะมีแนวโน้มที่จะตอบผิดมากขึ้นเท่านั้น ส่วนผู้ที่ได้คะแนนรวมต่ำจะมีแนวโน้มที่จะตอบข้อสอบข้อนั้นถูกมากขึ้นเช่นกัน

ข้อสอบที่มีค่า r เป็น 0 หรือใกล้ 0 หมายถึง ผู้ที่ได้คะแนนรวมสูง หรือ ต่ำอาจจะตอบข้อสอบข้อนั้นถูกหรือผิดก็ได้ไม่แน่นอน

ค่าอำนาจจำแนกจะมีความสัมพันธ์กับค่าความยาก (อนาสตาซี, 2519 : 171) กล่าวคือ ค่าอำนาจจำแนกจะมีค่าสูงสุด เมื่อค่าความยากมีค่าเท่ากับ 0.50

2. การวิเคราะห์หาคุณภาพของข้อสอบทั้งฉบับ โดยพิจารณาคุณลักษณะที่สำคัญ 2 ประการ ของข้อสอบคือ

2.1 ค่าความเที่ยงตรง (Validity)

2.2 ค่าความเชื่อมั่น (Reliability)

ความเที่ยงตรงของข้อสอบ (Validity) หมายถึงข้อสอบนั้นสามารถวัดได้ในสิ่งที่ต้องการวัดได้ถูกต้อง หรือความสามารถในการให้ความหมายในสิ่งที่วัดได้อย่างถูกต้อง (Lindquist, 1942 : 213)

ความเที่ยงตรงแบ่งออกเป็น 3 ชนิด คือ (Anasstasi, 1968 : 134)

1. ความเที่ยงตรงเชิงเนื้อหา (Content Validity) เป็นความเที่ยงตรงที่เกี่ยวกับเนื้อหาที่ใช้ในการสร้างข้อสอบ การตรวจสอบความเที่ยงตรงโดยการเทียบกับตารางวิเคราะห์หลักสูตรว่าข้อสอบฉบับนั้นครอบคลุมเนื้อหาของวิชานั้นหรือไม่เพียงใด

2. ความเที่ยงตรงเชิงเกณฑ์สัมพันธ์ (Criterion-Related Validity) เป็นความเที่ยงตรงที่มีความสัมพันธ์ระหว่างคะแนนการสอบกับการวัดเกณฑ์ภายนอก (External Criteria) ซึ่งเกณฑ์นี้แบ่งตามลักษณะที่เกี่ยวข้องได้ 2 ชนิดคือ

2.1 ความเที่ยงตรงเชิงสภาพ (Concurrent Validity) เป็นความสัมพันธ์ระหว่างคะแนนจากการสอบกับคะแนนที่เป็นเกณฑ์ซึ่งได้ในเวลาเดียวกัน

2.2 ความเที่ยงตรงเชิงพยากรณ์ (Predictive Validity) เป็นความสัมพันธ์ระหว่างคะแนนจากการสอบกับคะแนนที่เป็นเกณฑ์ซึ่งได้ภายหลัง

3. ความเที่ยงตรงเชิงโครงสร้าง (Construct Validity) เป็นความสามารถของข้อสอบที่วัดขอบเขตหรือคุณลักษณะประจำตามโครงสร้างทางทฤษฎี การประมาณค่าความเที่ยงตรงเชิงโครงสร้าง (Construct Validity) นี้ทำได้หลาย ๆ วิธี โดยอาศัยเทคนิคเฉพาะต่าง ๆ เช่น หาสหสัมพันธ์กับข้อสอบฉบับอื่นที่มีอยู่แล้ว ซึ่งวัดพฤติกรรมด้านเดียวกัน หรือใช้วิธีวิเคราะห์องค์ประกอบ (Factor Analysis) เป็นต้น

ความเชื่อมั่น (Reliability) หมายถึง ความคงที่ของคะแนนที่ได้จากการวัด โดยใช้ข้อสอบฉบับเดียวไปสอบกับผู้สอบคนเดียวกันหลาย ๆ ครั้ง หรือด้วยข้อสอบสองฉบับที่คู่ขนานกัน (Equivalent item) หรือภายใต้เงื่อนไขของตัวแปรอื่นในการวัดนั้น (Anastasi, 1968 : 71)

การประมาณค่าความเชื่อมั่นของแบบทดสอบ โดยอาศัยคะแนนที่ได้จากการวัดนำมาวิเคราะห์ ซึ่งแยกเป็น 2 วิธี (บุญเชิด ภิญโญอนันตพงษ์, มปป. : 238) คือ

1. การประมาณค่าความเชื่อมั่นโดยอาศัยความแปรปรวนของคะแนนที่วัดได้จากกลุ่มผู้สอบกลุ่มเดียวกันโดยการวัดหลาย ๆ ครั้ง จะถือว่าการวัดที่เชื่อมั่นได้ต้องให้ผลคงที่หรือมีความแม่นยำ แต่ในการวัดใด ๆ ย่อมมีความคลาดเคลื่อนเสมอ โดยเฉพาะอย่างยิ่งการวัดทางการศึกษา ซึ่งถือว่าคะแนนที่วัดได้ (Observed Score) จะประกอบด้วย คะแนนจริง (True Score) และคะแนนความคลาดเคลื่อน (Error Score) ซึ่งเขียนเป็นสมการได้ดังนี้ (Frederic M.Lord, 1980 : 4)

$$X \quad = \quad T + E$$

เมื่อ X คือ คะแนนที่วัดได้

T คือ คะแนนจริง

E คือ คะแนนความคลาดเคลื่อน

คะแนนจริง คือ คะแนนที่แท้จริงของสิ่งที่วัดโดยไม่ได้รับอิทธิพลจากสิ่งอื่น

คะแนนความคลาดเคลื่อน คือ คะแนนที่เกิดจากความผิดพลาดในการวัด ซึ่งมีอยู่ 2 ชนิด คือ

(1) ความคลาดเคลื่อนอย่างมีระบบ (Systematic Error) เป็นความคลาดเคลื่อนที่เกิดขึ้นกับผู้สอบทุกคนเหมือนกันหมด เช่น เกิดจากการพิมพ์ข้อสอบผิด ซึ่งความคลาดเคลื่อนชนิดนี้ จะไม่มีผลต่อความเชื่อมั่น

(2) ความคลาดเคลื่อนอย่างสุ่ม (Random Error) เป็นความคลาดเคลื่อนที่เกิดขึ้นโดยบังเอิญ เช่น เกิดจากการดำเนินการสอบ การตรวจให้คะแนนหรือจากตัวผู้สอบ ทำให้ค่าที่วัดได้มีค่าเพิ่มขึ้นหรือลดลงได้ ซึ่งกล่าวได้ว่า ความคลาดเคลื่อนนี้มีค่าเป็นได้ทั้งบวกและลบ ขึ้นอยู่กับสภาพการณ์และโอกาสของการวัดแต่ละคนหลาย ๆ ครั้ง เนื่องจากคะแนนความคลาดเคลื่อนเป็นได้ทั้งบวกและลบ ดังนั้นผลรวมของคะแนนความคลาดเคลื่อนจึงมีค่าเป็น 0 และในการวัดกับผู้สอบจำนวนมาก ๆ จะได้ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนจริงกับคะแนนความคลาดเคลื่อนมีค่าเป็น 0 ($r_{TE} = 0$) หรือกล่าวได้ว่า ผู้ที่สอบได้คะแนนสูง ไม่จำเป็นต้องมีคะแนนความคลาดเคลื่อนสูงตามไปด้วย (Guilford, 1954 : 349-350)

จากสมการ $X = T + E$ สามารถเขียนเป็นสมการความแปรปรวนของคะแนนได้ ดังนี้

$$\sigma_X^2 = \sigma_T^2 + \sigma_E^2$$

เมื่อ σ_X^2 คือ ความแปรปรวนของคะแนนที่วัดได้

σ_T^2 คือ ความแปรปรวนของคะแนนจริง

σ_E^2 คือ ความแปรปรวนของคะแนนความคลาดเคลื่อน

การประมาณค่าความเชื่อมั่นโดยยึดความคงที่ หรือความแน่นยำของการวัดเป็นหลักนี้ จะถือว่าถ้าความแปรปรวนของคะแนนที่วัดได้มีค่าใกล้เคียงกับความแปรปรวนของคะแนนจริงมากเพียงใด ผลที่ได้จากการวัดก็จะมีความเชื่อมั่นมากขึ้นเท่านั้น ซึ่งเขียนเป็นสมการได้ดังนี้ (Guildford, 1954 : 350)

$$r_{tt} = \sigma_T^2 / \sigma_X^2$$

หรือ $$r_{tt} = 1 - \sigma_E^2/\sigma_X^2$$

เมื่อ r_{tt} คือ ความเชื่อมั่นของแบบทดสอบ

2. การประมาณค่าความเชื่อมั่นโดยอาศัยสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนของผู้สอบกลุ่มเดียวกัน โดยการวัดหลาย ๆ ครั้ง แนวคิดนี้สืบเนื่องมาจากการไม่สามารถหาความเชื่อมั่นที่แท้จริงจาก $r_{tt} = \sigma_T^2/\sigma_X^2$ ได้ (เพราะไม่มีทางทราบ σ_T^2) ดังนั้นจึงอาศัยแนวคิดที่ว่าการวัดที่เชื่อมั่นได้นั้น ลำดับที่ของผู้สอบในกลุ่มเดียวกันที่ได้จากการวัดหลาย ๆ ครั้งต้องสอดคล้องกัน นั่นคือ ค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนผู้สอบกลุ่มเดียวกันที่ได้จากการวัดหลาย ๆ ครั้ง

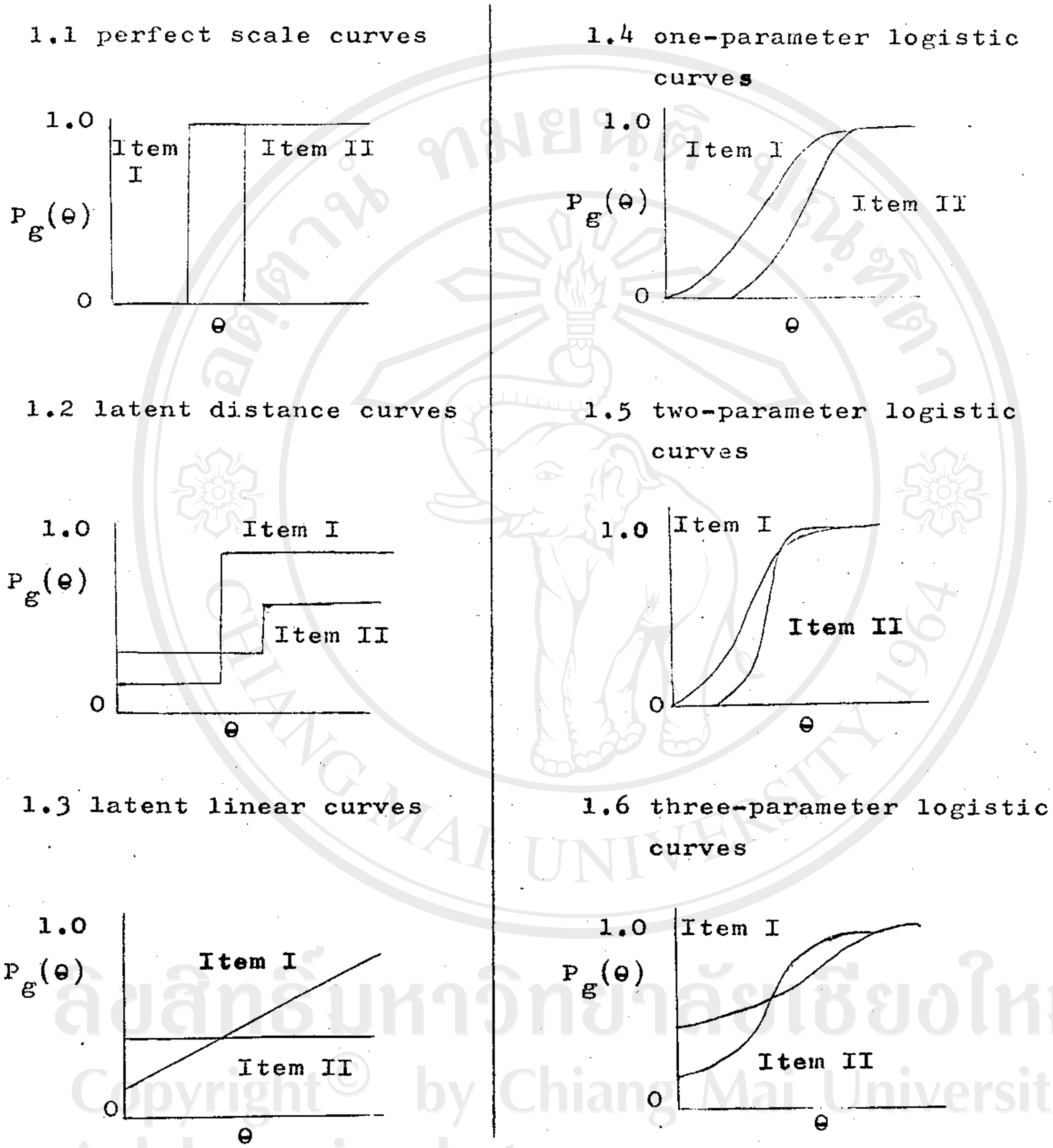
ทฤษฎีการตอบข้อสอบ (Item Response Theory : IRT) หรือที่มีชื่อเรียกอย่างอื่นคือทฤษฎีคุณลักษณะแฝง (Latent Trait Theory) หรือทฤษฎีรูปโค้งลักษณะของข้อสอบ (Item Characteristic Curve Theory) หรือทฤษฎีใหม่ (Modern Test Theory) เป็นทฤษฎีที่เฟอร์กูสัน (Ferguson) และลอเลย์ (Lawley) เป็นผู้ริเริ่มในปี 1942 และ 1943 (Warm, 1979 : 19, Lord and Novick, 1968 : 369) ต่อมาลอร์ด (Frederic M. Lord) ได้ศึกษาทฤษฎีนี้อีกครั้งหนึ่งในปี 1952 โดยให้ชื่อว่า ทฤษฎีรูปโค้งลักษณะของข้อสอบ (Item Characteristic Curve Theory) ใจความสำคัญของทฤษฎีนี้กล่าวว่า โอกาสในการตอบข้อสอบถูกหรือไม่นั้นขึ้นอยู่กับความยากของข้อสอบข้อนั้นกับระดับความสามารถของผู้สอบ (Trait or Ability) ความสัมพันธ์นี้กำหนดได้ด้วยฟังก์ชั่นทางคณิตศาสตร์ที่เรียกว่าเส้นโค้งลักษณะรายข้อ (Item Characteristic Function หรือ Item Characteristic Curve : ICC)

ลอร์ด (Frederic M. Lord) ได้เสนอลักษณะโค้งการตอบข้อสอบ (Item Characteristic Curve : ICC) ของแต่ละข้อว่ามีลักษณะเป็นแบบโอไจฟ์ปกติ (Normal Ogive หรือที่เรียกว่า Normal Ogive Model) ซึ่งโมเดลนี้จะกล่าวถึงพารามิเตอร์ของข้อสอบ 2 ตัว คือ ค่าความยาก (b) และค่าอำนาจจำแนก (a) แต่ความก้าวหน้าของเรื่องนี้ในระยะแรกนั้นเป็นไปอย่างเชื่องช้ามาก เนื่องจากความสลับซับซ้อนทางคณิตศาสตร์ที่เป็นพื้นฐานของทฤษฎี และการขาดแคลนเครื่องคอมพิวเตอร์ที่จำเป็นในการวิเคราะห์ข้อสอบ ความยุ่งยากดังกล่าวทำให้ลอร์ด (Frederic M. Lord) พัฒนาทฤษฎีนี้ได้ช้าอยู่ระยะหนึ่ง

ในปี 1960, 1966 ราชค์ (Georg Rasch) ได้ศึกษาและเสนอแนวความคิดของทฤษฎีนี้ในรูปแบบของการใช้พารามิเตอร์ของข้อสอบเพียงตัวเดียวคือ ค่าความยาก (b) หรือที่เรียกกันว่าราชค์โมเดล (Rasch Model) ซึ่งเป็นรูปแบบอย่างหนึ่งของโลจิสติคโมเดล (Logistic Models) ที่เรียกกันว่าแบบพารามิเตอร์ตัวเดียว (One - parameter Logistic Model) และในปี 1965 ลอร์ด (Fredetic M. Lord) ได้กลับมาสนใจศึกษาทฤษฎีนี้ใหม่ (Warm, 1979 : 19)

ในปี 1968 เบิร์นบอม (Birmbaum) ได้เสนอแนวความคิดของทฤษฎีนี้ในรูปของโลจิสติคโมเดล ที่ใช้พารามิเตอร์ของข้อสอบ 2 ตัว คือ ค่าความยาก (b) และค่าอำนาจจำแนก (a) ซึ่งเป็นโมเดลที่ง่ายกว่าที่ลอร์ด (Frederic M. Lord)ได้เสนอไว้ จึงทำให้เป็นที่นิยมใช้กันอย่างกว้างขวางและมีการพัฒนาขึ้นเรื่อย ๆ จนสามารถดัดแปลงให้ใช้ได้กับพารามิเตอร์ตัวเดียวคือ ค่าความยาก (b) และพารามิเตอร์สามตัวคือ ค่าความยาก (b) ค่าอำนาจจำแนก (a) และค่าการเดา (c) ได้ (Warm, 1979 : 19-21)

รูปที่ 1 แสดงโค้งลักษณะของข้อสอบ (ICC) ของโมเดลต่าง ๆ ในทฤษฎีการตอบข้อสอบ (IRT)(Hambleton and Cook, 1979 : 79)



หมายเหตุ θ คือ ระดับความสามารถของผู้สอบ

$P_g(\theta)$ คือ โอกาสในการตอบข้อสอบข้อที่ g ได้ถูกต้อง

รูปที่ 1.1 แสดงถึงโอกาสที่จะตอบข้อสอบแต่ละข้อได้ถูกเป็น 0 หรือ 1 มีลักษณะเป็นฟังชั่นแบบสเตป (Step function) ใช้กับมาตราของกัตต์แมน (Guttman's scale) จากรูปแสดงว่าข้อสอบข้อที่ 2 ยากกว่าข้อที่ 1

รูปที่ 1.2 แสดงถึงโอกาสที่จะตอบข้อสอบแต่ละข้อได้ถูกไม่เป็น 0 หรือ 1 มีลักษณะเป็นฟังชั่นแบบสเตป (Step function) เช่นเดียวกับรูปที่ 1.1 จากรูปแสดงว่าข้อสอบข้อที่ 2 ยากกว่าข้อที่ 1

รูปที่ 1.3 แสดงถึงความสัมพันธ์ของโอกาสที่จะทำข้อสอบได้ถูกกับระดับความสามารถมีลักษณะเป็นเส้นตรง จากรูปแสดงว่า ข้อสอบข้อที่ 1 จำแนกคนได้ดีกว่าข้อสอบข้อที่ 2

รูปที่ 1.4 แสดงถึงโอกาสที่จะทำข้อสอบได้ถูกขึ้นอยู่กับความยากของข้อสอบ (b) เพียงอย่างเดียว และทุกข้อมีค่าอำนาจจำแนก (a) เท่ากับ 1 จากรูปแสดงว่า ข้อสอบข้อที่ 2 ยากกว่าข้อที่ 1

รูปที่ 1.5 แสดงถึงโอกาสที่จะทำข้อสอบได้ถูกขึ้นอยู่กับค่าความยาก (b) และค่าอำนาจจำแนก (a) ของข้อสอบ จากรูปแสดงว่า ข้อสอบข้อที่ 2 จำแนกคนได้ดีกว่าข้อที่ 1

รูปที่ 1.6 แสดงถึงโอกาสที่จะทำข้อสอบได้ถูกขึ้นอยู่กับค่าความยาก (b) ค่าอำนาจจำแนก (a) และค่าการเดาของข้อสอบ (c) จากรูปแสดงว่า ข้อสอบข้อที่ 2 มีค่าการเดาน้อยกว่าข้อที่ 1 และมีค่าอำนาจจำแนกสูงกว่าข้อที่ 1

<u>ข้อตกลงเบื้องต้นของทฤษฎีการตอบข้อสอบ</u> (IRT) (Lord and Novick, 1968:359)

1. ข้อสอบแต่ละข้อในแบบทดสอบ จะต้องวัดความสามารถหรือคุณลักษณะเดียวกัน (Unidimension)

2. การตอบข้อสอบข้อใดข้อหนึ่งถูกนั้น ต้องเป็นอิสระจากการตอบข้ออื่น ๆ

3. โอกาสที่ผู้สอบจะตอบข้อสอบได้ถูกหรือไม่ ไม่ขึ้นอยู่กับการแจกแจงความสามารถของกลุ่มตัวอย่าง แต่จะขึ้นอยู่กับโค้งลักษณะของข้อสอบ (Item Characteristic Curve : ICC) ของแต่ละโมเดลที่ใช้

ทฤษฎีการตอบข้อสอบ (Item Response Theory : IRT) ได้มีการพัฒนาไปในรูปแบบต่าง ๆ ทำให้เกิดเป็นโมเดลเฉพาะขึ้น โดยที่แต่ละโมเดลก็จะแตกต่างกันด้วยรูปแบบของฟังชั่นทางคณิตศาสตร์ และจำนวนของพารามิเตอร์ที่ใช้ในการอธิบายฟังชั่นโค้งลักษณะของข้อสอบ (Item Characteristic Function) ในที่นี้จะกล่าวถึงโลจิสติคโมเดลที่กำลังได้รับการพัฒนาไปมาก และมีการนำไปใช้กันมากที่สุดในหลาย ๆ สถานการณ์ของการทดสอบ (ผจงจิต อินทสุวรรณ, 2525 : 59) โดยสามารถเลือกใช้กับพารามิเตอร์ได้ตามความเหมาะสม ซึ่งโลจิสติคโมเดลนี้มีโมเดลย่อยอยู่ 3 โมเดล คือ

1. แบบพารามิเตอร์ตัวเดียว (One-parameter Logistic Model)
2. แบบพารามิเตอร์สองตัว (Two-parameter Logistic Model)
3. แบบพารามิเตอร์สามตัว (Three-parameter Logistic Model)

<u>แบบพารามิเตอร์ตัวเดียว</u> (One-parameter Logistic Model)

เบิร์นบอม (Birnbuam) ได้พัฒนาทฤษฎีการตอบข้อสอบ (Item Response Theory : IRT) โดยที่ฟังชั่นของโมเดลนั้นสามารถอธิบายได้ด้วยพารามิเตอร์ของข้อสอบเพียงตัวเดียวคือ ค่าความยากของข้อสอบ (b) (One-parameter Logistic Model) หรือ ที่นิยมเรียกกันว่าราชต์โมเดล (Rasch Model) ซึ่งตรงกับที่เบิร์นบอม (Birnbaum) ได้พัฒนาขึ้น โอกาสหรือความน่าจะเป็นที่ผู้สอบจะตอบข้อสอบได้ถูกหรือไม่นั้นขึ้นอยู่กับผลต่างระหว่างความสามารถของผู้สอบ (θ) และค่าความยากของข้อสอบข้อนั้น (b) ส่วนค่าอำนาจจำแนก (a) จะมีค่าคงที่ทั้งฉบับและไม่มีการเดา (c) สามารถเขียนเป็นฟังชั่นแบบพารามิเตอร์ตัวเดียว (One-parameter Logistic Function) ได้ดังที่เบิร์นบอมให้สูตรไว้ดังนี้ (Birnbaum, 1968 : 402)

$$P_i(\theta) = DL_i(\theta) \; ; \; i = 1, 2, 3, \ldots\ldots, n$$

เมื่อ (X) เป็น logistic cummulation distribution function

และ $L_i(\theta) = \bar{a}(\theta - b_i)$ (Birnbaum, 1968 : 402)

หรือเขียนได้ดังนี้

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

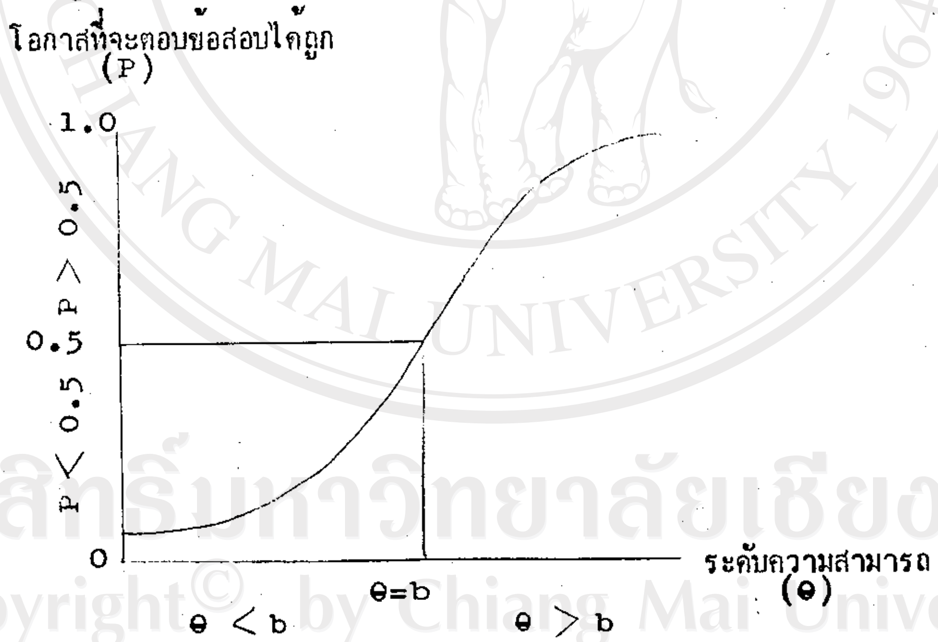
สำหรับโมเดลนี้ถือว่า ค่าอำนาจจำแนกของข้อสอบทุกข้อมีค่าเท่ากัน และถ้ากำหนดให้

$\bar{a}_i = 1$ และ $D = 1$ แล้ว

ดังนั้น จะได้รูปสมการที่ง่ายขึ้นดังนี้ คือ

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

รูปที่ 2 แสดงรูปโค้งลักษณะของข้อสอบ (ICC) ของแบบพารามิเตอร์ตัวเดียว (One-parameter Logistic Model)



รูปที่ 2 แสดงความสัมพันธ์ระหว่างความยากของข้อสอบและระดับความสามารถของผู้สอบ ซึ่งโอกาสที่จะตอบข้อสอบได้ถูกหรือไม่นั้น ขึ้นอยู่กับผลต่างระหว่างความยากของข้อสอบ (b) และระดับความสามารถของผู้สอบ (θ)

แบบพารามิเตอร์สองตัว (Two-parameter Logistic Model)

ในปี 1968 เบิร์นบอม (Birnbaum) เป็นผู้เสนอโลจิสติคโมเดลที่ใช้กับพารามิเตอร์ 2 ตัว คือค่าความยาก (b) และค่าอำนาจจำแนก (a) โดยที่โมเดลนี้สามารถดัดแปลงให้ใช้กับพารามิเตอร์ตัวเดียวคือ ค่าความยาก (b) และพารามิเตอร์สามตัวคือ ค่าความยาก (b) ค่าอำนาจจำแนก (a) และค่าการเดา (c) ซึ่งโมเดลที่ใช้กับพารามิเตอร์สองตัวนี้จะมีรูปแบบ (Form) เดียวกับแบบโอไจฟ์ปกติ (Normal Ogive Model) โดยที่สามารถอธิบายได้ด้วยพารามิเตอร์สองตัวเช่นเดียวกัน และจะมีโค้งลักษณะของข้อสอบ (Item Characteristic Curve : ICC) คล้ายคลึงกัน ซึ่งโค้งลักษณะของข้อสอบ (ICC) ของโลจิสติคโมเดลที่ใช้พารามิเตอร์สองตัวนี้เป็นฟังชั่นความถี่สะสมของโลจิสติค (Logistic cummulative distribution function) ซึ่งเขียนได้ดังนี้ (Birnbaum, 1968 : 399)

$$P_i(\theta) = \psi\left[DL_i(\theta)\right] \quad ; \quad i = 1,2,3, \ldots, n$$

เมื่อ $\psi(X)$ เป็น Logistic cummulative distribution function

$$\psi(X) = \left[1 + e^{-x}\right]^{-1} \quad ; \; L_i(\theta) = a_i(\theta - b_i)$$

และ D = 1.7 เป็น Scaling factor (Birnbaum, 1968 : 400) หรือเขียนเป็นรูปเต็มของฟังชั่นแบบพารามิเตอร์สองตัว (Two-parameter Logistic Function) ได้ดังนี้

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

เมื่อโค้งลักษณะของข้อสอบ (ICC) ของโอไจฟ์ปกติ (Normal Ogive Model) ได้มาจากการแจกแจงความถี่ของโค้งปกติ (Normal frequency distribution) ซึ่งมีค่าเฉลี่ยเป็น 0 และส่วนเบี่ยงเบนมาตรฐานเป็น 1.7 นั้น โค้งลักษณะของข้อสอบ (ICC) ในแบบพารามิเตอร์สองตัว (Two-parameter

Logistic Model) จะใกล้เคียงกับโค้งลักษณะของข้อสอบ (ICC) ในแบบโอไจฟ์ปกติ (Normal Ogive Model) มากที่สุด (Birnbaum, 1968 : 400) ดังนั้นจึงกล่าวได้ว่าโค้งลักษณะของข้อสอบ (ICC) ในแบบพารามิเตอร์สองตัว (Two-parameter Logistic Model) มีรูปแบบ (Form) เดียวกับโค้งลักษณะของข้อสอบ (ICC) ในแบบโอไจฟ์ปกติ (Normal Ogive Model) ซึ่งต่างอธิบายได้ด้วยพารามิเตอร์สองตัว โดยที่แต่ละโมเดลสามารถดัดแปลงใช้ได้กับพารามิเตอร์ตัวเดียวหรือสามตัว แต่ต่างกันที่โลจิสติคโมเดลคำนวณได้ง่ายกว่า และสามารถเลือกใช้ตามความเหมาะสมได้ดีกว่า

<u>แบบพารามิเตอร์สามตัว</u> (Three-parameter Logistic Model)

สำหรับกรณีที่เป็นข้อสอบเลือกตอบ (Multiple-choices) โดยที่แต่ละข้อคำถามจะมีตัวเลือกตอบ A ตัวเลือก นั่นคือโอกาสหรือความน่าจะเป็นที่ผู้ตอบอาจเดาคำตอบที่ถูกหรือเลือกตอบอย่างสุ่มจะมีค่าเท่ากับ 1/A แสดงว่าผู้ที่มีระดับความสามารถต่ำมาก ๆ ก็ไม่ได้หมายความว่าผู้นั้นจะมีความน่าจะเป็นในการตอบถูกเท่ากับ 0 หรือกล่าวได้ว่า เมื่อมีการเดาเกิดขึ้นปลายโค้งทางต่ำ (Lower asymtote) ของโอกาสในการตอบข้อสอบข้อที่ i ได้ถูก ณ ความสามารถ θ $[P_i(\theta)]$ ไม่เป็น 0 ดังนั้นพารามิเตอร์ของการเดา (Guessing parameter : c) ย่อมมีความสำคัญในการนำมาพิจารณา กรณีเช่นนี้ควรใช้โมเดลที่มีค่าการเดาด้วยโดยการดัดแปลงจากฟังชั่นแบบพารามิเตอร์สองตัว (Two-parameter Logistic Function) เพียงเล็กน้อยก็จะได้เป็นฟังชั่นแบบพารามิเตอร์สามตัว (Three-parameter Logistic Function) ซึ่งความน่าจะเป็นในการตอบข้อที่ i ถูก ณ ความสามารถ θ เขียนได้ดังนี้ (Birnbaum, 1968 : 405)

$$P_i(\theta) = c_i + (1 - c_i)\ \psi\left[DL_i(\theta)\right]$$

เมื่อ $\psi(X) = \left[1 + e^{-x}\right]^{-1}$; $L_i(\theta) = a_i(\theta - b_i)$

หรือเขียนเป็นรูปเต็มของฟังชั่นแบบพารามิเตอร์สามตัว (Three-parameter Logistic Function) ได้ดังนี้

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

ซึ่งพอจะสรุปได้ว่า ในการคำนวณค่าพารามิเตอร์นี้มี 2 ขั้นตอนใหญ่ ๆ คือ (นิสิตดุษฎีบัณฑิต จุฬาลงกรณ์มหาวิทยาลัย, 2525 : 3)

ขั้นที่ 1 ประมาณค่าความสามารถของผู้สอบโดยใช้จำนวนข้อที่ถูก แล้วกำหนดค่าความสามารถของผู้สอบ (θ) ให้มีค่าคงที่และนำไปคำนวณหาค่าพารามิเตอร์ของข้อสอบโดยให้ค่าอำนาจจำแนก (a) มีค่าเท่ากับ 1 และค่าความยาก (b) มีค่าเท่ากับสัดส่วนของจำนวนคนที่ตอบข้อสอบข้อนั้นถูก

ขั้นที่ 2 กำหนดค่าพารามิเตอร์ของข้อสอบให้มีค่าคงที่ แล้วจะประมาณค่าความสามารถ (θ) ของผู้สอบ ซึ่งจะประมาณค่ากลับไปกลับมาเช่นนี้ จนกระทั่งได้โค้งลักษณะของข้อสอบ (Item Characteristic Curve : ICC) ที่เหมาะ (Fit) กับโมเดล

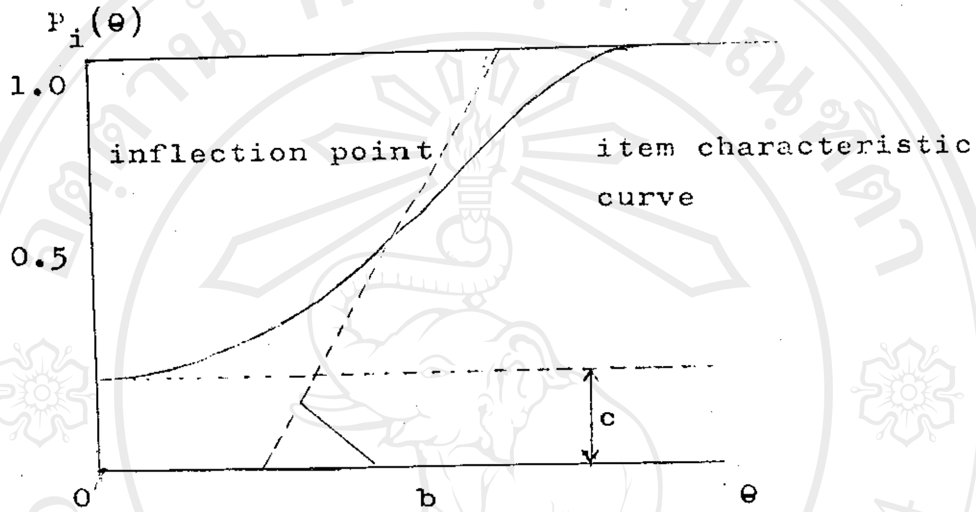
ในการประมาณค่าพารามิเตอร์โดยการวิเคราะห์แบบพารามิเตอร์ตัวเดียว นอกจากจะประมาณค่าด้วยเครื่องคอมพิวเตอร์โดยใช้โปรแกรมโลจิส (LOGIST) ซึ่งกำหนดเงื่อนไขให้ค่าอำนาจจำแนก (a) เท่ากับ 1 และค่าการเดา (c) เท่ากับ 0 แล้ว ยังสามารถประมาณค่าด้วยเครื่องคอมพิวเตอร์โดยใช้โปรแกรมไบคาล (BICAL) (Wright and Stone, 1979 : 46) หรืออาจจะประมาณค่าพารามิเตอร์แบบง่าย ๆ ด้วยมือก็ได้ (Wright and Stone, 1979 : 28)

2. การวิเคราะห์หาคุณภาพของข้อสอบหรือการพิจารณาความแน่นอนในการประมาณค่าความสามารถที่แท้จริงตามทฤษฎีการตอบข้อสอบ (Item Response Theory : IRT) นั้นใช้วิธีการพิจารณาจากอินฟอเมชั่นฟังชั่น (Test information function) แต่ในทฤษฎีเดิม (The Classical Test Theory) ใช้วิธีการพิจารณาจากค่าความเชื่อมั่น (Reliability) และค่าความคลาดเคลื่อนมาตรฐานในการวัด (Standard Error of Measurement) ซึ่งค่าที่ได้จะเปลี่ยนแปลงไปตามกลุ่มผู้สอบ ดังนั้นในทฤษฎีการตอบข้อสอบ (Item Response Theory : IRT) จึงได้ศึกษาในเรื่องอินฟอเมชั่นฟังชั่น (Test information function) แทนความเชื่อมั่น (Reliability) (Hambleton, 1977 : 64)

เมื่อ b_i เป็นระดับความสามารถของความน่าจะเป็นที่ตอบถูกเป็น $(1 + c_i)/2$

และ c_i เป็น lower asymtote ของ ICC

รูปที่ 3 แสดงโค้งลักษณะของข้อสอบ (ICC) ของแบบพารามิเตอร์สามตัว (Three-parameter Logistic Model)



รูปที่ 3 แสดงความสัมพันธ์ของโอกาสที่จะตอบข้อสอบได้ถูก $[P_i(\theta)]$ กับระดับความสามารถของผู้สอบ (θ) โดยผู้ที่มีระดับความสามารถสูง จะมีโอกาสที่จะตอบข้อสอบได้ถูกมากกว่าผู้ที่มีระดับความสามารถต่ำ ซึ่งค่าความสามารถและค่าความยากจะถูกประมาณค่าอยู่บนสเกลเดียวกัน

<u>การวิเคราะห์หาคุณภาพของข้อสอบ</u>

ในทฤษฎีนี้มีการวิเคราะห์หาคุณภาพของข้อสอบอยู่ 2 ขั้นตอน คือ

1. การวิเคราะห์หาคุณภาพของข้อสอบเป็นรายข้อหรือการประมาณค่าพารามิเตอร์ของข้อสอบ

ในการวิเคราะห์ข้อสอบเป็นรายข้อ หรือการประมาณค่าพารามิเตอร์ในโลจิสติคโมเดลนี้ ปกติคำนวณโดยใช้เครื่องคอมพิวเตอร์ ซึ่งมีโปรแกรมสำเร็จรูปคือโปรแกรมโลจิส(LOGIST) ซึ่งวูด วินเกอร์สกี และ ลอร์ด (Wood, Wingersky and Lord, 1968 : 9)ได้พัฒนาโปรแกรมนี้ขึ้นมาใช้ในการประมาณค่าความสามารถของผู้สอบ และค่าพารามิเตอร์ของข้อสอบ โดยที่โปรแกรมนี้จะประมาณค่าความสามารถของผู้สอบแต่ละคน และค่าพารามิเตอร์ของข้อสอบแต่ละข้อ

ในเวลาเดียวกันโดยจะประมาณค่ากลับไปกลับมา (Iterate) จนกว่าจะเหมาะ (Fit) กับโมเดล รวมค่าพารามิเตอร์ที่ต้องประมาณค่าทั้งหมดมีจำนวน N+3n ตัว และจำนวนค่าพารามิเตอร์ทั้งหมดนี้จะมีค่าสูงสุดเท่ากับ L^*

$$L^* = \sum_{a=1}^{N} \sum_{i=1}^{n_a} P_{ia}^{V_{ia}} Q_{ia}^{(1-V_{ia})}$$

เมื่อ P_{ia} คือ $P_g(\theta)$ คือ ความน่าจะเป็นของผู้สอบคนที่ตอบข้อสอบข้อที่ i หรือ g ถูก

Q_{ia} คือ $1 - P_{ia}$

N คือ จำนวนผู้สอบ

n_a คือ จำนวนข้อสอบที่ผู้สอบคนที่ a ทำถูก

V_{ia} คือ คะแนนของผู้สอบคนที่ a ตอบข้อสอบข้อที่ i โดยตอบถูกมีค่าเป็น 1 ตอบผิดมีค่าเป็น 0

ค่าพารามิเตอร์ทั้งหมดใน L^* จะถูกปรับหรือประมาณค่าใหม่ให้มีค่าต่ำสุดเท่ากับ F (Wood, Wingersky and Lord, 1976 : 10)

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

โดยกำหนดให้ $\underset{\sim}{Z}$ เป็นอนุพันธ์ที่ 1 ของ F มีค่าเป็น 0 และเป็นเวคเตอร์ของพารามิเตอร์ที่ประกอบด้วยเวคเตอร์ของความสามารถและเวคเตอร์ของพารามิเตอร์ของข้อสอบทุกข้อ

เนื่องจากสมการ F ไม่เป็นสมการเส้นตรง ดังนั้นการคำนวณค่าพารามิเตอร์จึงต้องใช้วิธีประมาณค่าของนิวตัน (Iteration Procedure or Newton's Method)โดยประมาณค่าซ้ำหลาย ๆ ครั้ง (Wood, Wingersky and Lord, 1976 : 10)จากสมการ

$$\underset{\sim}{Z}^{r+1} = \underset{\sim}{Z}^{r} - \left\{ H^{-1} \frac{\partial F}{\partial Z} \right\} \Bigg| \, Z^{r}$$

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

H คือ แมทริกซ์ของอนุพันธ์ที่ 2 ของ F

การพิจารณาความแน่นอนในการประมาณค่าความสามารถที่แท้จริงจากอินฟอเมชั่นฟังชั่น (Test information function) นั้นหาได้จากผลรวมของอินฟอเมชั่นฟังชั่นรายข้อ (Item information function) (Lord, 1980 : 72) ซึ่งกำหนดไว้ดังนี้

$$I\left\{\theta, U_i\right\} = P_i'^2/P_iQ_i \quad \ldots\ldots\ldots\ldots\ldots\ldots(A)$$

เมื่อ $I\left\{Q, U_i\right\}$ คือ Item information function

P_i' คือความชันของ Item Characteristic Curve ที่ระดับความสามารถ θ

P_i คือความน่าจะเป็นของผู้ตอบที่มีความสามารถ θ จะตอบข้อสอบข้อที่ i ได้ถูก

Q_i คือ $1 - P_i$

จากสมการ (A) แสดงว่าอินฟอเมชั่นฟังชั่นแต่ละข้อ (Item information function) ขึ้นอยู่กับความชันของโค้งลักษณะของข้อสอบ (ICC) และความแปรปรวนของการตอบข้อสอบข้อนั้นถูกที่แต่ละระดับของความสามารถ θ นั่นคือ ถ้าโค้งลักษณะของข้อสอบ (ICC) มีความชันมากขึ้นในขณะที่ความแปรปรวนของการตอบข้อสอบข้อนั้นถูกมีค่าลดลงจะทำให้โค้งของอินฟอเมชั่นฟังชั่น (Item information curve) ที่ระดับความสามารถนั้นมีค่าสูงขึ้น ซึ่งความสูงของโค้งอินฟอเมชั่นฟังชั่นที่ระดับของความสามารถในการตอบข้อสอบใด ๆ จะสามารถจำแนกระดับความสามารถของผู้ตอบได้ในการวัดความสามารถระดับนั้น(Hambleton, 1977 : 66)

ดังนั้นจากสมการ (A) สามารถสรุปได้ว่า (Lord, 1980 : 71)

$$I\left\{\theta\right\} = \sum_{i=1}^{n} P_i'^2/P_iQ_i \;;\quad i = 1,2,3 \ldots\ldots, n$$

เมื่อ $I\left\{\theta\right\}$ คือ Test information function

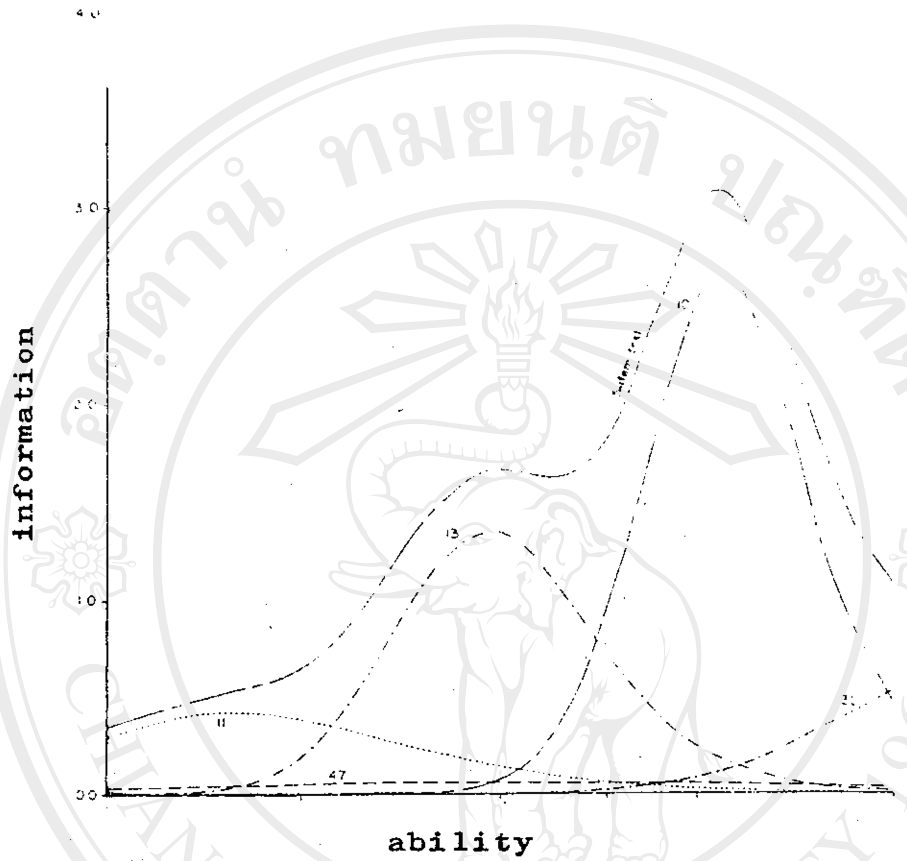
ค่าอินฟอเมชั่นฟังชั่น (Test information function) นี้ จะเป็นดัชนีชี้ให้เห็นถึงคุณภาพในการวัดของข้อสอบในกรณีที่มีกลุ่มของข้อสอบที่ทราบ (Information curve) จะสามารถสร้างข้อสอบชุดนั้นให้มีโค้งของอินฟอเมชั่น (Test information curve) ที่ระดับหนึ่งของความสามารถที่ต้องการได้ เพื่อประโยชน์ในการใช้แบบทดสอบได้ตรงตามจุดประสงค์ เช่น เพื่อคัดเลือกบุคคลให้ได้รับทุน ก็ใช้ข้อสอบที่มีประสิทธิภาพสูงสุดที่ระดับความสามารถสูง ๆ หรือเลือกข้อสอบที่ให้อินฟอเมชั่น (Test information) สูง ที่ระดับความสามารถสูง ๆ ดังตัวอย่างในตารางที่ 1 (สำเริง บุญเรืองรัตน์, 2525 : 11)

ตารางที่ 1 แสดงค่าพารามิเตอร์ของข้อสอบ 5 ข้อ

ข้อที่	ค่าความยาก (b)	ค่าอำนาจจำแนก (a)	ค่าการเดา (c)
10	1.1	2.0	.05
11	-1.5	0.9	.20
13	0.1	1.6	.16
30	2.4	1.1	.09
47	-0.4	0.4	.20

จากตารางที่ 1 เป็นค่าพารามิเตอร์ของข้อสอบวัดความถนัดทางการเรียนด้านภาษา 5 ข้อ เมื่อคำนวณค่าความโค้งของอินฟอเมชั่น (Item information curve) ของแต่ละข้อ และคำนวณค่าความโค้งของอินฟอเมชั่น (Item information curve) รวมทั้ง 5 ข้อ แล้วสามารถสร้างกราฟได้ดังรูปที่ 4 (อ้างถึงใน Lord, 1980 : 22)

รูปที่ 4 แสดงโค้งอินฟอเมชั่น (Item information curve) ของข้อสอบ 5 ข้อ และโค้งอินฟอเมชั่น (Test information curve) ของข้อสอบทั้ง 5 ข้อ (Lord, 1980 : 22)



จากรูปที่ 4 แสดงว่าข้อสอบทั้ง 5 ข้อ เหมาะกับการทดสอบกับผู้สอบที่มีความสามารถสูง ๆ

รายงานวิจัยภายในประเทศ

ปัจจุบันนักการศึกษาและนักวัดผลบางคนได้ให้ความสนใจ และนำเอาทฤษฎีการตอบข้อสอบ (Item Response Theory : IRT) มาใช้ในทางการศึกษา เช่น จากการศึกษาของ ชัยชัย เผ่าพงศ์ (ชัยชัย เผ่าพงศ์, 2527 : 73-74) ได้นำเอาวิธีการวิเคราะห์ข้อสอบแบบพารามิเตอร์สามตัว (Three-parameter Logistic Model) มาใช้ในการศึกษาเรื่อง การวิเคราะห์ความลำเอียงของข้อสอบจากแบบทดสอบความถนัดทางการเรียนด้านคณิตศาสตร์และภาษาในระดับชั้นมัธยมศึกษาตอนต้น ซึ่งได้วิเคราะห์โดยการเปรียบเทียบโค้งลักษณะของข้อสอบ

(Item Characteristic Curve : ICC) ผลปรากฏว่า ข้อสอบส่วนใหญ่มีความลำเอียงทางการทดสอบ และมีขีดจำกัดเรื่องการเปรียบเทียบผลการสอบระหว่างนักเรียนกลุ่มชายและหญิง เพราะผลการสอบที่วัดได้ส่วนใหญ่มาจากความสามารถต่างชนิดกัน ซึ่งไม่สามารถชี้ให้เห็นถึงความแตกต่างเกี่ยวกับระดับความสามารถด้านเดียวกันได้อย่างชัดเจน

ชูศักดิ์ ขัมภลิขิต (ชูศักดิ์ ขัมภลิขิต, 2525 : 33-43) ได้นำเอาวิธีการวิเคราะห์ข้อสอบแบบพารามิเตอร์สามตัว (Three-parameter Logistic Model)มาใช้ในการศึกษาเรื่อง IRT กับผลการสอบคัดเลือก โดยศึกษากับบุคคลที่สอบเข้าศึกษาต่อในระดับอุดมศึกษา ผลปรากฏว่า ในสถานการณ์การสอบคัดเลือกโดยทั่ว ๆ ไปที่ข้อสอบส่วนใหญ่เป็นข้อสอบที่เขียนขึ้นมาใหม่ และยังไม่มีการทดลองใช้มาก่อนนั้น สามารถที่จะนำข้อมูลจากการทดสอบไปใช้กับลาเท้น-เทรทโมเดล (Latent Trait Model) ได้

อวยพร วิบูลย์กาญจน์ (อวยพร วิบูลย์กาญจน์, 2526 : 47-48) ได้นำเอาวิธีการวิเคราะห์ข้อสอบแบบราชค์โมเดล มาใช้ในการศึกษาเรื่อง การเปรียบเทียบผลการวิเคราะห์แบบสอบอุปมาอุปไมยด้วยคลาสสิคอลโมเดลกับราชค์โมเดล ผลปรากฏว่า

1. ข้อสอบที่คัดเลือกไว้ด้วยคลาสสิคอลโมเดลมีจำนวนข้อสอบมากกว่าข้อสอบที่คัดเลือกไว้ด้วยราสชโมเดลอย่างมีนัยสำคัญที่ระดับ .05

2. ค่าความเชื่อมั่นของแบบทดสอบที่ประกอบด้วยข้อสอบที่คัดเลือกไว้ด้วยคลาสสิคอล-โมเดล มีค่าสูงกว่าของราชค์โมเดล อย่างมีนัยสำคัญที่ระดับ .01

รายงานวิจัยต่างประเทศ

ฮัทเทน (Hutten, 1982 : 4799A) ได้ทำการศึกษาถึงความสัมพันธ์ระหว่างราชค์โมเดล (Rasch Model) กับการวิเคราะห์แบบพารามิเตอร์สามตัว (Three-parameter Logistic Model)ตามข้อตกลงเบื้องต้นของแต่ละโมเดล ผลปรากฏว่า

1. ข้อสอบที่ทำการวิเคราะห์โดยวิธีราชค์โมเดล (Rasch Model)และการวิเคราะห์แบบพารามิเตอร์สามตัว (Three-parameter Logistic Model) จะมีข้อ

สอบที่เหมาะสม (Fit) กับโมเดลทั้งสองไม่แตกต่างกัน กล่าวคือ จะมีจำนวนข้อที่เหมาะ (Fit) กับโมเดลประมาณ 80 %

2. ข้อสอบที่ไม่เหมาะ (Fit) กับโมเดลเป็นเพราะข้อสอบนั้นไม่ได้วัดความสามารถด้านเดียว (Unidimension)

3. เมื่อค่าอำนาจจำแนกไม่เท่ากัน จำนวนข้อที่เหมาะ (Fit) กับโมเดลทั้งสองจะเปลี่ยนไป

4. การวิเคราะห์ข้อสอบโดยวิธีราชค์โมเดล (Rasch Model) นั้นใช้กลุ่มตัวอย่างประมาณ 250 คน ก็สามารถใช้การประมาณค่าความยากของข้อสอบได้ แต่ในการประมาณค่าการเดาของการวิเคราะห์แบบพารามิเตอร์สามตัว (Three-parameter Logistic Model)ควรใช้กลุ่มตัวอย่างตั้งแต่ 1000 คนขึ้นไป

เมเชสชา (Meshesha, 1984 : 2121A) ได้ทำการศึกษาโดยนำข้อสอบวัดผลสัมฤทธิ์มาให้ผู้เชี่ยวชาญ (Expert) เป็นผู้คัดเลือกข้อสอบที่มีประสิทธิภาพและเปรียบเทียบกับข้อสอบที่จัดไว้ว่าเหมาะ (Fit)กับราชค์โมเดล (Rasch Model) ผลปรากฏว่า

1. ข้อสอบที่คัดเลือกโดยผู้เชี่ยวชาญ (Expert)กับข้อสอบที่เหมาะ (Fit) กับราชค์โมเดล (Rasch Model) มีความสอดคล้องกันประมาณ 65-75 %

2. ข้อสอบที่คัดเลือกโดยผู้เชี่ยวชาญ (Expert)จะมีค่าความเชื่อมั่นสูงกว่าข้อสอบที่เหมาะ (Fit) กับราชค์โมเดล (Rasch Model)

จากผลงานการวิจัยที่มีผู้นำเอาทฤษฎีการตอบข้อสอบ (Item Response Theory IRT) มาทำการศึกษาดังกล่าว จะเห็นได้ว่าในปัจจุบันทฤษฎีนี้ ได้เริ่มเข้ามามีบทบาทต่อวงการวัดผลการศึกษาในสถานการณ์ต่าง ๆ และในการศึกษาครั้งนี้จะทำการวิเคราะห์ข้อสอบวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 1 ซึ่งผู้วิจัยสร้างขึ้น โดยให้ผู้เชี่ยวชาญ (Expert) เป็นผู้คัดเลือกข้อสอบและยังไม่มีการทดลองใช้มาก่อน มาทำการศึกษากับกลุ่มตัวอย่างจำนวน 1,687 คน โดยนำเอาวิธีการวิเคราะห์ข้อสอบแบบโลจิสติคโมเดลมาใช้ในการศึกษา และเปรียบเทียบผลที่ได้กับวิธีการวิเคราะห์ข้อสอบแบบเดิม