บทที่ 2
วรรณกรรมที่เกี่ยวข้อง

การวิเคราะห์แบบสอบในเมืองไทย เป็นไปในลักษณะหาค่าความยาก ค่าอำนาจจำแนก ความเที่ยง และความตรง ซึ่งใช้วิธีการวิเคราะห์แบบคลาสสิคอล ส่วนในต่างประเทศกำลังเปลี่ยนแนวการวิเคราะห์หันไปใช้ทฤษฎีการตอบข้อกระทง (Item response theary : IRT) แทน ซึ่งในเมืองไทยยังไม่มีการศึกษา IRT อย่างจริงจัง โดยเฉพาะอย่างยิ่งการประยุกต์ IRT มาใช้ในการวิเคราะห์แบบสอบ การวิจัยครั้งนี้จึงนับเป็นงานวิจัยชิ้นแรกในเมืองไทยที่นำ IRT มาประยุกต์ใช้ในการวิเคราะห์แบบสอบโดยเปรียบเทียบกับผลการวิเคราะห์แบบคลาสสิคอล

ผู้วิจัยจะเสนอรายละเอียดเกี่ยวกับทฤษฎีคลาสสิคอลกับ IRT โดยแบ่งเป็น 4 ตอน ดังนี้

ตอนที่ 1 มโนทัศน์เกี่ยวกับการวิเคราะห์แบบสอบของทฤษฎีคลาสสิคอล
ตอนที่ 2 มโนทัศน์เกี่ยวกับ IRT
ตอนที่ 3 มโนทัศน์เกี่ยวกับราสช์โมเดล และการนำราสช์โมเดลไปประยุกต์ใช้
ตอนที่ 4 งานวิจัยที่เกี่ยวกับราสช์โมเดล

ตอนที่ 1 มโนทัศน์เกี่ยวกับการวิเคราะห์แบบสอบของทฤษฎีคลาสสิคอล

ในการวัดผลนักวัดผลได้ใช้แบบสอบเป็นตัวกระตุ้นให้ผู้สอบแสดงพฤติกรรมอันเป็นความสามารถที่แท้จริงซึ่งเป็นคุณลักษณะภายในที่ไม่สามารถวัดได้โดยตรง ต้องนำเอาพฤติกรรมที่ได้จากจำนวนข้อที่ตอบถูกมาใช้หาความสามารถที่แท้จริง ดังนั้นแบบสอบที่ใช้วัดจึงจำเป็นต้องมีคุณภาพที่เชื่อถือได้ว่า คะแนนที่วัดได้นั้นเป็นคะแนนที่แท้จริง หรือมีความคลาดเคลื่อนในการวัดน้อยที่สุด และแบบสอบฉบับนั้นวัดได้ตรงตามความสามารถที่แท้จริงที่ต้องการวัด ดังนั้นจึงจำเป็นต้องมีกระบวนการในการตรวจสอบคุณภาพของแบบสอบ ซึ่งมีอยู่ 2 กระบวนการด้วยกัน คือ

1. วิเคราะห์รายข้อกระทง (Item Analysis) เป็นการตรวจสอบคุณภาพของแบบสอบเป็นรายข้อกระทง โดยพิจารณาดูคุณลักษณะที่สำคัญ 2 ประการ (Mehrens and Ebel 1969 : 375) คือ ค่าความยากของข้อกระทง (Item difficulty) และค่าอำนาจจำแนก (Discriminating power)

2. ตรวจสอบคุณภาพ โดยพิจารณาจากคุณลักษณะที่สำคัญ 2 ประการของแบบสอบ คือ ความเที่ยง (Reliability) และความตรง (validity)

ความยากของข้อกระทง หมายถึงสัดส่วนหรือร้อยละของผู้ที่เข้าสอบทั้งหมดที่ตอบข้อกระทงแต่ละข้อถูก ความยากของข้อกระทงมีค่าอยู่ระหว่าง 0 ถึง 1 ค่าความยากใกล้ 0 จะเป็นข้อกระทงที่ยาก ความยากใกล้ 1 จะเป็นข้อกระทงที่ง่าย

การหาค่าความยากของข้อกระทงเพื่อที่จะตัดข้อกระทงที่มีค่าความยากไม่เหมาะสมออก เช่น ค่าความยากที่ใกล้ 0 หรือ 1 เพราะข้อกระทงเหล่านี้จะไม่สามารถจำแนกความสามารถของผู้สอบออกจากกันได้ (Yen 1979 : 120-122)

อำนาจจำแนกของข้อกระทง หมายถึงความสามารถของข้อกระทงที่จะชี้ให้เห็นความแตกต่างระหว่างกลุ่มที่ได้คะแนนสูงกับกลุ่มที่ได้คะแนนต่ำมีค่าตั้งแต่ -1.00 ถึง 1.00 ข้อกระทงที่มีค่าอำนาจจำแนกใกล้ -1.00 หรือ 1.00 แสดงว่ามีอำนาจจำแนกสูง อำนาจจำแนกเป็นบวกหมายความว่ากลุ่มที่ได้คะแนนสูงตอบข้อกระทงนั้นได้ถูกต้องมากกว่ากลุ่มที่ได้คะแนนต่ำ อำนาจจำแนกที่มีค่าลบถือว่าจำแนกผิดทิศทาง ถ้าเป็น 0 แสดงว่าจำแนกความแตกต่างของทั้ง 2 กลุ่มไม่ได้เลย ซึ่งในทางปฏิบัติข้อกระทงที่มีค่าอำนาจจำแนกต่ำมากหรือมีค่าเป็นลบจะถูกตัดทิ้ง

ค่าอำนาจจำแนกมีความสัมพันธ์กับค่าความยาก กล่าวคือค่าความยากมีค่าเป็น .5 จะทำให้มีค่าอำนาจจำแนกสูงสุด (อนาสตาซี 2519 : 171)

ความเที่ยงของแบบสอบ (Reliability) หมายถึงความคงที่ของคะแนนที่ได้จากการสอบบุคคลกลุ่มเดียวกันสองครั้ง ด้วยแบบสอบเดิมในเวลาที่ต่างกัน หรือทดสอบคนกลุ่มเดียวกันด้วยแบบสอบต่างชุดที่เทียบเท่ากัน หรือภายใต้สภาพการสอบที่แตกต่างกัน (อนาสตาซี 2519 : 73)

<u>ทฤษฎีของความเที่ยง</u> ในการวัดทางด้านจิตภาพ ถึงแม้เครื่องมือที่ใช้จะมีความเที่ยงสูง ก็ไม่มีเครื่องมือชนิดใดที่วัดได้ถูกต้องแม่นยำโดยสมบูรณ์ จะนั้นคะแนนจากการสอบที่นักเรียนทำได้ (Observed Scores) จะประกอบด้วยส่วนที่เป็นคะแนนจริง (True Scores) และส่วนของคะแนนจากความคลาดเคลื่อน (Error Scores) ซึ่งเขียนเป็นสมการ ดังนี้ (Guilford and Fruchter 1978 : 408)

$$X_t = X_\infty + X_e$$

เมื่อ X_t = คะแนนที่ได้จากการวัด

X_∞ = คะแนนจริง

X_e = คะแนนความคลาดเคลื่อน

คะแนนจริง คือคะแนนเฉลี่ยที่ผู้สอบได้รับจากการทำแบบสอบเดิมหลาย ๆ ครั้ง โดยปราศจากอิทธิพลของการเรียนรู้จากการสอบครั้งแรก แรงจูงใจ หรือสภาพทางอารมณ์ เป็นต้น

คะแนนความคลาดเคลื่อน คือคะแนนที่เกิดจากความผิดพลาดในการวัดแบ่งเป็น 2 ประเภท คือความคลาดเคลื่อนที่เกิดจากการสุ่ม (Random Error) กับความคลาดเคลื่อนอย่างมีระบบ (Systematic Error) ซึ่งความคลาดเคลื่อนประเภทหลังนี้ไม่มีผลต่อค่าความเที่ยง แต่ความคลาดเคลื่อนที่เกิดจากการสุ่ม เช่น เกิดจากตัวข้อสอบ การดำเนินการสอบ การตรวจให้คะแนน และจากตัวผู้สอบเอง มีค่าเป็นไปได้ทั้งทางบวกและทางลบ และจะหักล้างกันหมด เพราะวัดจากทุก ๆ คนแล้วค่าเฉลี่ยของ X_e จะเท่ากับศูนย์ และคะแนนความคลาดเคลื่อนนี้ก็ไม่มีความสัมพันธ์กับคะแนนจริง หรือขนาดของคะแนนความคลาดเคลื่อนไม่เกี่ยวข้องกับขนาดของคะแนนจริงแล้ว ความแปรปรวนของคะแนนความคลาดเคลื่อน (Robert L. Ebel 1965 : 331) ก็จะเขียนเป็นสมการได้ดังนี้

$$S_t^2 = S_\infty^2 + S_e^2 \quad \ldots\ldots\ldots\ldots\ldots\ldots(1)$$

เมื่อ S_t^2 แทนความแปรปรวนของคะแนนที่วัดได้

S_∞^2 แทนความแปรปรวนของคะแนนจริง

S_e^2 แทนความแปรปรวนของคะแนนความคลาดเคลื่อน

ความเที่ยง หมายถึงอัตราส่วนระหว่างความแปรปรวนของคะแนนจริงกับความแปรปรวนของคะแนนที่วัดได้ เขียนเป็นสมการ (Guilford and Fruchter, 1978 : 410) ได้ว่า $S_t^2 = S_\infty^2 + S_e^2$

$$r_{tt} = \frac{S_\infty^2}{S_t^2} \quad \text{.....(2)}$$

ฉะนั้นจากสมการ (1) และ (2) จะได้สมการพื้นฐานในการหาสัมประสิทธิ์สหสัมพันธ์ของความเที่ยง ดังนี้

$$r_{tt} = 1 - \frac{S_e^2}{S_t^2}$$

วิธีประมาณค่าความเที่ยงจัดได้เป็น 5 วิธี (อนาสตาซี 2519 : 79)

1. ความเที่ยงแบบสอบซ้ำ (Test - Retest Reliability)
2. ความเที่ยงแบบสลับฟอร์ม (Alternate - Form Reliability)
3. ความเที่ยงชนิดแบ่งครึ่ง (Split - Half Reliability)
4. ความเที่ยงแบบคูเดอร์-ริชาร์ดสัน (Kuder - Richaradson Reliability)
5. ความเที่ยงของผู้ให้คะแนน (Scores Reliability)

<u>องค์ประกอบที่มีอิทธิพลต่อความเที่ยงของแบบสอบ</u> (Gronlund 1976 : 117) มี 4 ข้อ ดังนี้

1. ความยาวของแบบสอบ (Length of Test) แบบสอบที่มีจำนวนข้อกระทงมากจะมีค่าความเที่ยงสูงกว่าแบบสอบที่มีจำนวนข้อกระทงน้อย เนื่องจากแบบสอบที่มีจำนวนข้อกระทงมากจะประกอบด้วยข้อกระทงที่วัดพฤติกรรมที่ต้องการได้อย่างเพียงพอ คะแนนที่ได้จากอิทธิพลของการเดาลดน้อยลง

2. การกระจายของคะแนน (Spread of Scores) การกระจายของคะแนนจะขึ้นอยู่กับความสามารถของกลุ่มผู้สอบ ถ้าความสามารถของผู้สอบในกลุ่มแตกต่างกันมาก (กลุ่มวิวิธพันธ์) การกระจายของคะแนนจะมาก ทำให้ค่าความเที่ยงสูง แต่ถ้าความสามารถของผู้สอบในกลุ่มใกล้เคียงกันมาก (กลุ่มเอกพันธ์) การกระจายของคะแนนจะน้อยทำให้ความเที่ยงมีค่าต่ำ

3. ความยากของแบบสอบ (Difficulty of Test) แบบสอบที่ยากมากหรือแบบสอบที่ง่ายเกินไป จะทำให้ความแปรปรวนของคะแนนน้อย ความเที่ยงจะมีค่าต่ำ แต่ถ้าหากแบบสอบมีความยากเหมาะสมจะทำให้ความแปรปรวนของคะแนนมากขึ้น ค่าความเที่ยงจะสูง ทั้งนี้เพราะความเที่ยงของแบบสอบจะเกี่ยวข้องกับความแปรปรวนของคะแนนสอบที่ได้

4. ความเป็นปรนัย (Objectivity) แบบสอบที่มีความเป็นปรนัยสูง ได้แก่ ผู้ตรวจไม่มีอคติ ภาษาที่ใช้ถูกต้อง ชัดเจน จะทำให้ผลของการสอบตรงตามความสามารถที่ผู้สอบมีอยู่จริง การวัดนั้นก็จะมีความคงที่นั่นคือแบบสอบมีความเที่ยงสูง

<u>ความตรงของแบบสอบ</u> (Validity) หมายถึงแบบสอบนั้นสามารถวัดในสิ่งที่เราต้องการจะวัดได้ถูกต้อง ความตรงแบ่งออกเป็น 3 ชนิดด้วยกัน (แอน อนาสตาซี 2519 : 99-122) คือ

1. <u>ความตรงในเนื้อหา</u> (Content Validity) เป็นความตรงที่เกี่ยวกับการสุ่ม (Sampling) เนื้อหามาใช้ วิธีการตรวจความตรงให้เทียบกับตารางวิเคราะห์หลักสูตร (test blueprint) ดูว่าแบบสอบนั้นครอบคลุมทั้งเนื้อหาวิชา และชนิดของพฤติกรรมของวิชาที่สอบนั้นหรือไม่เพียงใด

2. <u>ความตรงตามเกณฑ์สัมพันธ์</u> (Criterion - Related Validity) คือความสัมพันธ์ระหว่างคะแนนการทดสอบกับการวัดเกณฑ์ภายนอก (external criteria) ซึ่งแบ่งตามลักษณะของเกณฑ์ที่เกี่ยวข้องได้ 2 ชนิด คือ

2.1. <u>ความตรงร่วมสมัย</u> (Concurrent Validity) เป็นการหาสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนจากการสอบกับคะแนนที่เป็นเกณฑ์ซึ่งเก็บรวบรวมได้ในเวลาเดียวกัน

2.2. <u>ความตรงตามทำนาย</u> (Predictive Validity) เป็นการหาสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนจากการสอบกับคะแนนที่เป็นเกณฑ์ซึ่งเก็บรวบรวมในภายหลัง

3. <u>ความตรงตามทฤษฎี</u> (Construct Validity) หมายถึงความสามารถของแบบสอบที่สามารถวัดคุณลักษณะประจำตามโครงสร้างทางทฤษฎี วิธีหาค่าความตรงมีหลายวิธี เช่น หาสหสัมพันธ์กับแบบสอบอื่นที่มีอยู่แล้ว ซึ่งวัดพฤติกรรมด้านเดียวกัน หรือนำไปวิเคราะห์องค์ประกอบ (Factor Analysis) เป็นต้น

ตอนที่ 2 มโนทัศน์เกี่ยวกับทฤษฎีการตอบข้อกระทง (IRT)

IRT เป็นทฤษฎีหนึ่งในทฤษฎีคุณลักษณะภายใน (Latent trait theory) บางคนจึงเรียก IRT ว่าเป็นทฤษฎีคุณลักษณะภายใน Ferguson (1942) และ Lawley (1943) เป็นผู้ริเริ่ม IRT (Warm 1979 : 19; Lord and Novick 1968 : 369) โดยมีหลักการว่า ผลการสอบของผู้สอบจากแบบสอบใด ๆ นั้นขึ้นอยู่กับความสามารถ (Latent trait or ability or skill) ของผู้สอบ ซึ่งสามารถเขียนอยู่ในรูปฟังชั่นทางคณิตศาสตร์ ซึ่งฟังชั่นเหล่านี้ประกอบด้วยค่าพารามิเตอร์ต่าง ๆ ได้แก่ ค่าความยาก อำนาจจำแนก และการเดา ซึ่งแล้วแต่ว่าฟังชั่นนั้นจะมีพารามิเตอร์กี่ตัว

ต่อมาในปี 1952 ลอร์ดได้เสนอ IRT ขึ้นใหม่ให้ชื่อว่าทฤษฎีโค้งคุณลักษณะข้อกระทง (Item Characteristic Curve Theory) โดยเสนอว่าลักษณะของโค้งผลการตอบ (Item characteristic curve) ของข้อกระทงแต่ละข้อมีลักษณะเป็น normal ogive ต่อมาจึงเรียกว่า Normal Ogive Model ซึ่งโมเดลนี้มีการคำนวณยุ่งยากมาก ดังนั้นจึงทำให้ลอร์ดขาดความสนใจใน IRT ไประยะหนึ่ง ในปี 1960 Georg Rasch ได้เสนอ IRT ในรูปพารามิเตอร์เดียวคือความยาก และในปี 1965 ลอร์ดก็หันกลับมาสนใจพัฒนา IRT ใหม่ (Warm 1979 : 19)

ปี 1968 เบิร์นบอม (Birnbaum) ได้เสนอ Logistic Model ที่ใช้ 2 พารามิเตอร์ขึ้น ซึ่งเป็นโมเดลที่ง่ายกว่าของลอร์ดจึงทำให้ IRT เป็นที่นิยมใช้กันแพร่หลาย และมีการพัฒนาขึ้นเรื่อย ๆ จนกระทั่งใช้ได้กับพารามิเตอร์ตัวเดียว 2 ตัว และ 3 ตัว (Warm 1979 : 19-21)

ข้อตกลงเบื้องต้นของ IRT (Lord and Novick 1968 : 359)

1. แบบสอบที่มีมิติเดียว (Unidimension tests) หมายความว่า ข้อกระทงแต่ละข้อในแบบสอบจะต้องวัดความสามารถหรือคุณลักษณะเดียวกัน (a single ability or latent trait) หรือมีความเป็นเอกพันธ์กัน (homogeneous)

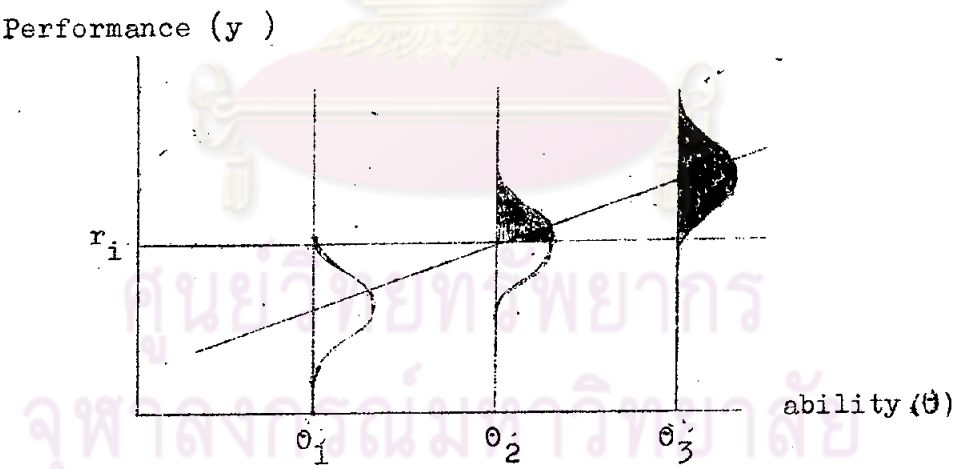
2. ข้อกระทงแต่ละข้อจะต้องเป็นอิสระจากกัน (Local independence) หมายความว่า การตอบข้อกระทงข้อหนึ่งถูกจะไม่มีผลต่อการตอบข้อกระทงข้ออื่น ๆ

3. โอกาสที่ผู้สอบจะตอบข้อกระทงถูกขึ้นอยู่กับลักษณะของโค้งผลการตอบ (Item characteristic curve : ICC) ของแต่ละโมเดลที่ใช้ ไม่ขึ้นกับการแจกแจงความสามารถของกลุ่มตัวอย่าง

โมเดลต่าง ๆ ใน IRT

โมเดลต่าง ๆ ใน IRT ได้มาจากมโนทัศน์ในเรื่องความสัมพันธ์ระหว่างความสามารถ (ability) กับการกระทำ (Performance) ถ้ามีความสามารถ (θ) สูง ก็น่าจะมี performance (y) สูงด้วย ถ้ามีเกณฑ์ (r_i) ตัวหนึ่งเป็นตัวบอกว่า y แค่ไหนจึงจะทำข้อกระทง i ถูก ดังนั้น $y \geq r_i$ แสดงว่าทำข้อกระทง i ถูก ถ้า $y < r_i$ แสดงว่าทำข้อกระทง i ผิด

รูปที่ 1 แสดงความสัมพันธ์ระหว่างความสามารถ(θ)กับ Performance (y)



ดังนั้นถ้าโอกาสที่จะตอบถูก (พ.ท.ส่วนที่แรเงา) ในระดับความสามารถต่าง ๆ มาเขียนกราฟใหม่ จะได้เป็นรูป ICC ต่าง ๆ ซึ่ง ICC เหล่านี้จะแตกต่างกันไปตามฟังชั่นทางคณิตศาสตร์ และจำนวนพารามิเตอร์ที่ใช้อธิบาย ICC ของแต่ละโมเดล ซึ่งเขียนเป็นรูปฟังชั่นทั่วไปคือ

$$P(\theta) = f(\theta)$$

$P(\theta)$ แทนความน่าจะเป็นในการตอบถูก

θ แทนความสามารถที่แท้จริง

f แทนความสัมพันธ์

ซึ่ง $f(\theta)$ จะเปลี่ยนไปตามโมเดลต่าง ๆ ดังนี้

1. <u>The Latent linear model</u> (Hambleton and Cook 1977 : 78)

$$P_g(\theta) = b_g + a_g\theta$$

$P_g(\theta)$ คือ ความน่าจะเป็นที่ของคนในระดับความสามารถ θ จะตอบข้อกระทงที่ g ถูก

a_g คือ ค่าอำนาจจำแนก

b_g คือ ค่าความยาก

2. <u>The latent distance model</u> (Torgerson 1958 : 374)

$$P_g(\theta) = \begin{cases} b_g - a_g ; & \theta \leq \theta_g \\ b_g + a_g ; & \theta > \theta_g \end{cases}$$

โมเดลนี้ใช้กับ Guttman scale และ scale อื่น ๆ

3. <u>Normal ogive model</u> ; (Lord and Novick 1968 : 366)

$$P_g(\theta) = \int_{-\infty}^{a_g(\theta - b_g)} \phi(t)\, dt \quad ; \quad g = 1, 2, 3, \ldots, n$$

$P_g(\theta)$, a_g, b_g มีความหมายเช่นเดียวกับในข้อ 1

θ คือ ระดับความสามารถของผู้สอบ

$\phi(t)$ คือ normal density function

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

4. Logistic Model จะมี ICC เหมือนกับของ Normal Ogive model เมื่อเพิ่มค่าคงที่ 1.7 เข้าไปใน model ซึ่ง Logistic Model คำนวณง่ายกว่าของ Normal Ogive มาก และยังสามารถแบ่ง model ย่อยออกเป็น 3 โมเดลดังนี้ (Hambletan and Cook 1977 : 80-82)

4.1 โมเดลที่ใช้พารามิเตอร์ 3 ตัว (Three - Parameter Logistic Model) มีสูตรการคำนวณดังนี้

$$P_g(\theta) = C_g + (1 - C_g)\frac{e^{Da_g(\theta - b_g)}}{1 + e^{Da_g(\theta - b_g)}}; \quad g = 1,2,3,\ldots n$$

C_g คือ ค่าการเดา

D คือ scaling factor มีค่า 1.7

e คือ ค่าคงที่มีค่า 2.7182818

4.2 โมเดลที่ใช้พารามิเตอร์ 2 ตัว (Two - Parameter Logistic Model) ถือว่าแบบสอบไม่มีการเดา มีสูตรการคำนวณดังนี้

$$P_g(\theta) = \frac{e^{Da_g(\theta - b_g)}}{1 + e^{Da_g(\theta - b_g)}} \quad ; \quad g = 1,2,3,\ldots,n$$

4.3 โมเดลที่ใช้พารามิเตอร์ 1 ตัว (One - Parameter Logistic Model) เป็น โมเดลที่ Geard Rasch เสนอในปี 1960 ซึ่งตรงกับ Logistic Model ของเบิร์นบอมที่คิดค่าความยากของแบบสอบเพียงอย่างเดียว จึงนิยมเรียกโมเดลนี้ว่าราสช์โมเดล (Rasch Model)

$$P_g(\theta) = \frac{e^{\theta - b_g}}{1 + e^{\theta - b_g}}$$

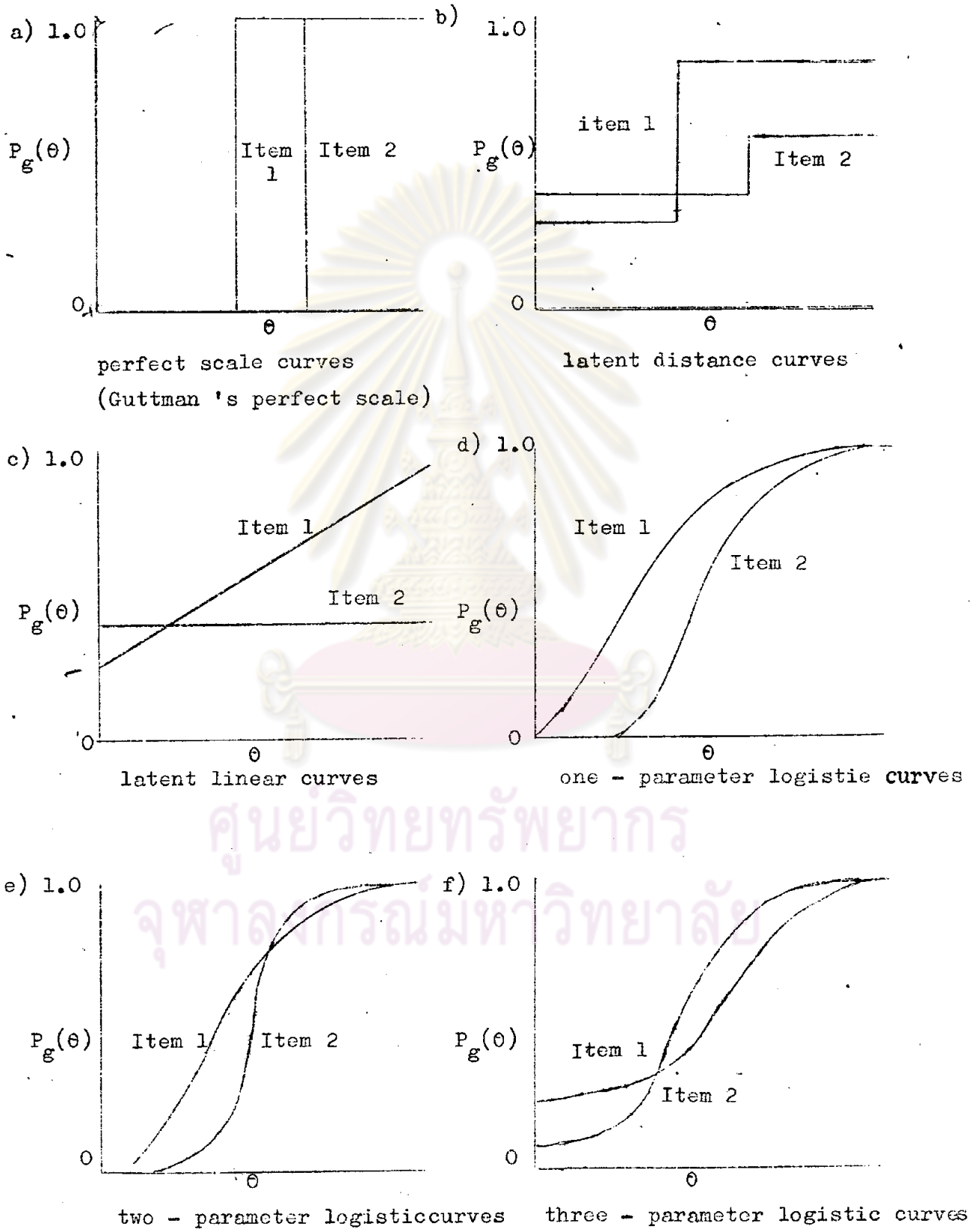
พิสัยของค่าพารามิเตอร์

θ a,b มีค่าระหว่าง $-\infty$ ถึง ∞ แต่ในทางปฏิบัติ θ มีค่าระหว่าง -3 ถึง 3 a มีค่าอยู่ระหว่าง 0.5 ถึง 2.5 b มีค่าอยู่ระหว่าง -2 ถึง 2 และ c มีค่าอยู่ระหว่าง 0 ถึง 0.3 (Ree 1979 : 372)

010595

115404705

รูปที่ 2 แสดง I.C.C ของโมเดลต่าง ๆ (Hambleton and Cook 1979 : 79)



ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

รูป 2a แสดงถึงโอกาสที่จะทำข้อกระทงแต่ละข้อได้ถูกเป็น 0 หรือ 1 เป็น Step function ใช้กับ Guttman's scale จากรูปแสดงว่าข้อกระทงที่ 2 ยากกว่าข้อ 1

รูป 2b เป็น step function เหมือน 2a แต่เป็นรูปทั่วไปที่มีโอกาสที่จะตอบถูกไม่เป็น 0 และ 1 จากรูปแสดงว่า ข้อกระทงที่ 2 ยากกว่าข้อ 1

รูป 2c เป็นรูปที่แสดงความสัมพันธ์ของโอกาสที่จะทำถูกกับระดับความสามารถเป็นเส้นตรง ซึ่งข้อกระทงที่ 1 จำแนกคนได้ดีกว่าข้อกระทงที่ 2

รูป 2d แสดงถึงโอกาสที่จะทำข้อกระทงได้ถูกขึ้นอยู่กับความยากของข้อกระทงเพียงอย่างเดียว ทุกข้อกระทงมีค่าอำนาจจำแนกเท่ากันหมดคือเป็น 1 ในรูปแสดงถึงข้อกระทงที่ 2 ยากกว่าข้อ กระทงที่ 1

รูป 2e แสดงถึงโอกาสที่จะทำข้อกระทงได้ถูกขึ้นอยู่กับค่าอำนาจจำแนก และค่าความยาก ในรูปข้อกระทงที่ 2 จำแนกคนได้ดีกว่าข้อ กระทงที่ 1

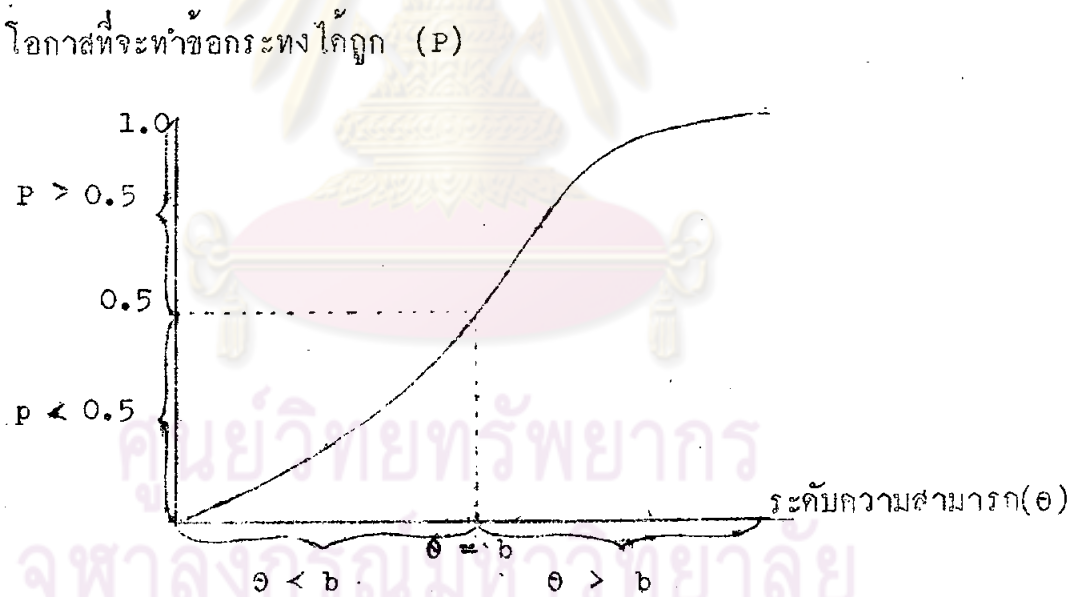
รูป 2f แสดงถึงโอกาสที่จะทำข้อกระทงได้ถูกขึ้นอยู่กับค่าอำนาจจำแนก ค่าความยาก และการเดา ในรูปข้อกระทงที่ 2 มีการเดาน้อยกว่าข้อกระทงที่ 1 และมีค่าอำนาจจำแนกสูงข้อกระทงที่ 1

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ตอนที่ 3 มโนทัศน์เกี่ยวกับราสชโมเดล และการนำราสชโมเดลไปประยุกต์

มโนทัศน์ของราสชโมเดลนั้นโอกาสที่คนจะทำข้อกระทงได้ถูกหรือไม่นั้นขึ้นอยู่กับระดับความสามารถของตนเอง (ability parameter : θ) และระดับความยากของข้อกระทง (difficulty parameter : b) เช่น ถ้า b = .5 และ θ = .5 โอกาสที่คนผู้นั้นจะสามารถทำข้อกระทงนั้นได้ถูกต้องประมาณ 50% ถ้าหากความสามารถของบุคคล (θ) น้อยกว่าความยากของข้อกระทง (b) แล้ว โอกาสที่จะทำข้อกระทงนั้นถูกย่อมจะน้อยกว่า 50% และในทำนองเดียวกันถ้าค่า θ มากกว่า b แล้ว โอกาสที่คนผู้นั้นจะทำข้อกระทงนั้นได้ถูกมีมากกว่า 50% ดูภาพประกอบในรูปที่ 3 (Wright and Stone 1979 : 14)

รูปที่ 3 ICC ของราสชโมเดลซึ่งแสดงความสัมพันธ์ระหว่าง θ และ b



หรืออีกนัยหนึ่งโอกาสที่คนจะทำข้อกระทงได้ถูกหรือไม่นั้นขึ้นอยู่กับผลต่างระหว่างความสามารถ (θ) และความยากของข้อกระทง (b)

จากสูตร $P(\theta) = \frac{e^{(\theta - b)}}{1 + e^{(\theta - b)}}$

จะเห็นได้ว่า θ > b มาก ก็ P (θ) ก็จะสูง

ข้อตกลงเบื้องต้นของราสชโมเดล

ราสชโมเดลนอกจากจะมีข้อตกลงเบื้องต้นของ IRT แล้ว ยังต้องมีข้อตกลงเบื้องต้นเพิ่มขึ้นดังนี้

1. คะแนนเป็นแบบ dichotomous : คือเป็น 1 หรือ 0

2. ข้อกระทงแต่ละข้อต้องมีค่าอำนาจจำแนกเท่ากัน เพื่อความสะดวกในการคำนวณจึงให้เป็น 1 และไม่มีการเดาเกิดขึ้น

วิธีประมาณค่าพารามิเตอร์ในราสชโมเดล มีด้วยกันหลายวิธี คือ (Wright; Afterword in Rasch 1980 : 188-192)

1. The LOG Method เป็นวิธีการประมาณค่าพารามิเตอร์ในรูปของ log ของความสำเร็จ ซึ่งเป็นวิธีที่เหมาะกับแบบสอบที่มีค่าอำนาจจำแนกเท่ากันคือ 1 และใช้จำนวนคนในการประมาณค่าพารามิเตอร์เป็นจำนวนมาก

2. The PAIR Method เป็นวิธีการประมาณค่าพารามิเตอร์ โดยใช้ข้อกระทงจับคู่เป็นคู่ ๆ เช่น 1-2, 1-3, 1-4, 2-4 .....เป็นต้น เอาจำนวนคนที่พยายามตอบ 2 ข้อ แต่ทำถูกเพียงข้อเดียวมาวิเคราะห์เพื่อหา sample free ดังนั้นจึงเหมาะที่จะใช้ประมาณค่าแบบสอบที่มีจำนวนข้อกระทงไม่มาก

3. The FCON Method เป็นวิธีประมาณค่าพารามิเตอร์ โดยใช้ข้อกระทงทุกข้อและคนทุกคนมาวิเคราะห์ เหมาะสำหรับแบบสอบที่มีข้อกระทงน้อยกว่า 30 ข้อ ถ้ามากกว่า 30 ข้อจะทำให้การประมาณค่าพารามิเตอร์มีความคลาดเคลื่อน แต่ภายหลังได้มีการปรับปรุงใหม่สามารถใช้กับแบบสอบที่มีความยาว 60-70 ข้อได้

4. The UCON Method เป็นวิธีประมาณค่าพารามิเตอร์ด้วยวิธี Unconditional Maximum Likelihood Procedure ใช้กับแบบสอบที่มีการกระจายของความสามารถของคนและความยากของแบบสอบเป็นการกระจายปกติ (normal distribution) และใช้กับแบบสอบที่มีข้อกระทงตั้งแต่ 25 ข้อขึ้นไป โดยมากจะใช้กับแบบสอบที่มี 1000 ข้อขึ้นไป ถ้าข้อกระทงน้อยใช้ FCON

5. The PROX Method เป็นวิธีประมาณค่าพารามิเตอร์ ด้วยวิธีของ Cohen's Approximation ใช้กับแบบสอบที่มีการกระจายของความสามารถของคนและความยากของแบบสอบเป็นลักษณะเบ้ หรือมีแนวโน้มเป็นปกติ โดยมาก PROX และ UCON จะให้ผลเหมือนกัน จะต่างกันที่ค่าความคลาดเคลื่อนมาตรฐานเพียงเล็กน้อย

6. The UFORM Method ใช้กับแบบสอบที่มีการกระจายความยากของข้อกระทงเป็นแบบ Uniform

การใช้เครื่องคอมพิวเตอร์ช่วยในการประมาณค่าพารามิเตอร์ในราสชโมเดล

การประมาณค่าพารามิเตอร์ในราสชโมเดลมี 2 แบบคือ ประมาณค่าด้วยมือ (Wright and Stone, 1979 : 28) กับประมาณค่าด้วยเครื่องคอมพิวเตอร์ โดยใช้โปรแกรมสำเร็จรูปที่เรียกว่า BICAL* (Wright and Mead, 1978) ซึ่งมีวิธีการประมาณค่าพารามิเตอร์อยู่ 2 วิธี คือ UCON และ PROX

ถ้าไม่ใช้ BICAL อาจจะใช้โปรแกรม LOGIST**(Wood and Lord 1976, Wingersky and Lord 1976) ประมาณค่าพารามิเตอร์แทนก็ได้ โดยกำหนดเงื่อนไขให้ค่าอำนาจจำแนกเท่ากับ 1 และค่าการเดาเป็น 0

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

---

* โปรแกรม BICAL มีอยู่ที่ อาจารย์สุพัฒน์ สุกมลสันต์ อาจารย์สถาบันภาษาจุฬาลงกรณ์ มหาวิทยาลัย

** โปรแกรม LOGIST มีอยู่ที่ อาจารย์ชูศักดิ์ ชัมภลิขิต กรมวิชาการ กระทรวงศึกษาธิการ

การนำราสชโมเดลไปใช้ (Wright; Afterword in Rasch 1980 : 194-196)

1. วิเคราะห์ข้อกระทง (Item Analysis) ซึ่งจะช่วยแก้จุดอ่อนในการวิเคราะห์แบบคลาสสิคอลได้ ในแง่ทำให้ค่าพารามิเตอร์มีค่าคงที่ไม่แปรเปลี่ยนไปตามกลุ่มตัวอย่าง

2. สร้างคลังข้อสอบ (Item Bank) เนื่องจากแบบสอบที่วิเคราะห์ด้วยราสชโมเดล จะให้ค่าพารามิเตอร์คงที่ ดังนั้นจึงสามารถสร้างแบบสอบเป็นชุด ๆ ซึ่งสามารถเลือกไปใช้สอบได้ตามต้องการ

3. สร้างแบบสอบที่ดีที่สุด (Best test design) ผลจากการวิเคราะห์ข้อกระทงด้วยราสชโมเดล สามารถนำไปใช้ในการออกแบบสอบที่มีคุณลักษณะต่าง ๆ ตามต้องการได้ เช่นให้ความยากของแบบสอบมีโค้งตามลักษณะที่ต้องการได้ ถ้าแบบสอบมีข้อกระทงที่มีความยากอยู่ระหว่าง $M \pm 2SD$. (M, SD. เป็นค่าเฉลี่ยและความเบี่ยงเบนมาตรฐานของความสามารถ (θ)) และมีความยาวเป็น $6/SE^2$ (SE คือ ความคลาดเคลื่อนมาตรฐานของการวัดตามที่ต้องการ) จะถือว่าเป็นแบบสอบที่ดี

4. วัดระดับความสามารถของแต่ละบุคคล (Self - Tailoring) โดยสุ่มข้อกระทงที่วิเคราะห์แล้ว และมีระดับความยากเรียงกันตามลำดับเพียงจำนวนเล็กน้อยมาทดสอบระดับความสามารถของแต่ละบุคคลได้

5. ตรวจสอบความเป็นอคติของข้อกระทง (Item Bias) ในการวิเคราะห์หา ICC ของแต่ละกลุ่ม ถ้า ICC ของข้อกระทงนั้น ๆ แตกต่างกันตามกลุ่มที่นำมาทดสอบ เช่น คนในเมือง คนนอกเมือง ชาย หญิง แล้ว แสดงว่าข้อกระทงนั้นมีอคติเกิดขึ้น

6. วินิจฉัยความสามารถของผู้ตอบ (Individual Diagnosis) ถ้า ICC ของข้อกระทงไม่ fit กับโมเดล แสดงว่าอาจจะมีบางสิ่งบางอย่างผิดปกติในตัวผู้สอบ

การวิเคราะห์แบบสอบด้วยราสช์โมเดล

ในการวิเคราะห์ข้อกระทงหรือประมาณค่าพารามิเตอร์นั้น ต้องพิจารณาดูว่าข้อกระทงนั้น ๆ มี ICC แตกต่างจาก ICC ของราสช์โมเดลหรือไม่ โดยการทดสอบด้วยค่า chi square หรือทดสอบ mean square ด้วย t - test (Wright and Stone 1979 : 71-77) หรืออาจจะทดสอบ mean square ด้วย อัตราส่วน F (Hashway 1978 : 21) ถ้าข้อกระทงนั้น ๆ มี ICC แตกต่างจากของราสช์โมเดลอย่างมีนัยสำคัญ แสดงว่าข้อกระทงนั้นไม่เหมาะที่จะประมาณค่าพารามิเตอร์ต่าง ๆ ด้วยราสช์โมเดล เพราะความน่าจะเป็นของการตอบข้อกระทงนั้น ๆ ถูกไม่สามารถอธิบายด้วยค่าความยากและระดับความสามารถที่แท้จริงของผู้สอบได้

สำหรับการวิเคราะห์คุณภาพของแบบสอบ หรือการพิจารณาความแน่นอนของการประมาณค่าความสามารถที่แท้จริงในราสช์โมเดลนั้น ใช้วิธีการพิจารณาจาก test information function ซึ่งหาได้จากผลรวมของ item information function (IIF) (Lord 1980 : 72)

$$I\{\theta, U_i\} = \frac{P_i'^2}{P_i Q_i} \quad \text{.....(4)}$$

$I\{\theta, U_i\}$ คือ item information function

P_i คือ ความน่าจะเป็นที่ผู้สอบที่มีความสามารถ θ จะตอบข้อกระทงที่ i ถูก

Q_i คือ $(1 - P_i)$

P_i' คือ ความชันของ ICC ที่ระดับความสามารถ θ

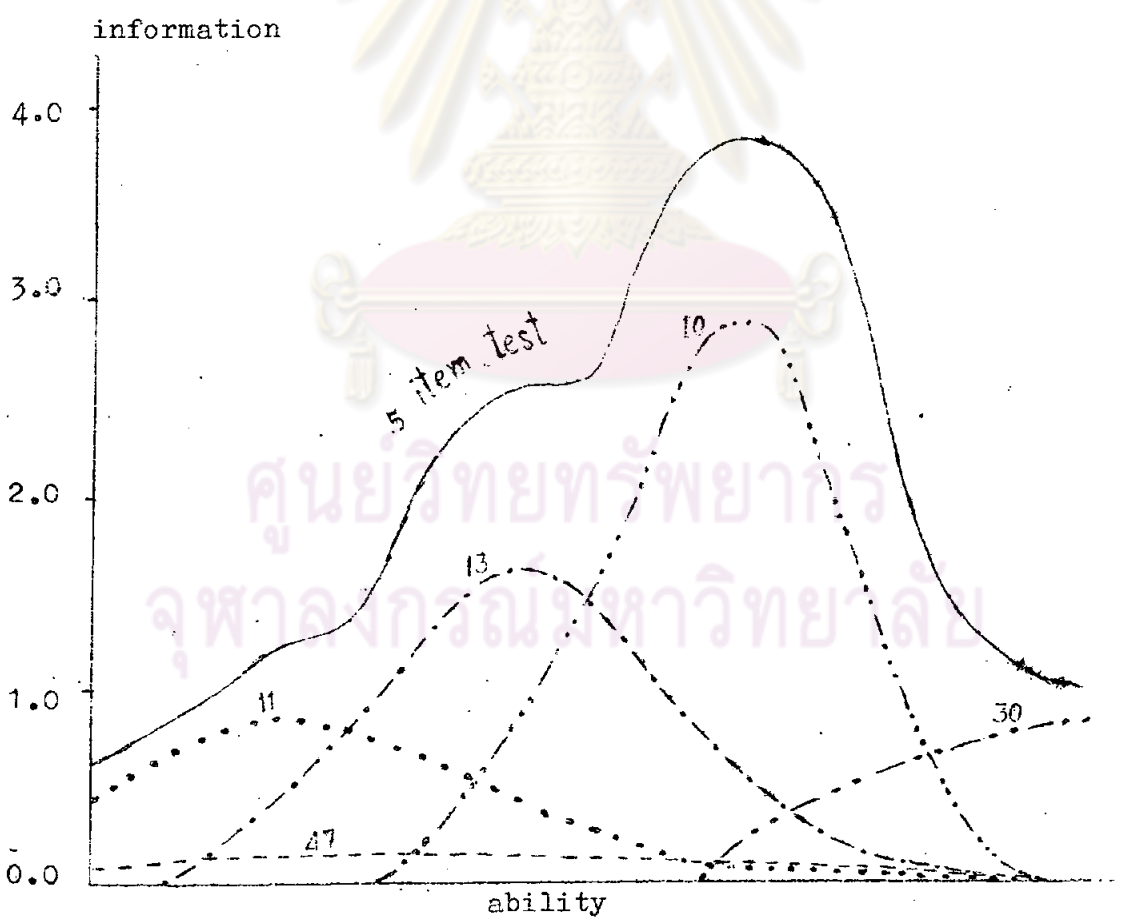
จากสมการ (4) จะเห็นได้ว่า item information function ขึ้นอยู่กับความชันของ ICC และความแปรปรวนของการตอบข้อกระทงถูกในแต่ละระดับความสามารถ θ ถ้า ICC ชันมากขึ้นในขณะที่ความแปรปรวนของการตอบข้อกระทงถูกยิ่งน้อยลง item information curve ที่ระดับความสามารถนั้น ๆ ก็จะยิ่งสูงขึ้น ซึ่งความสูงของ item information curve อยู่ที่ระดับความสามารถใด แสดงว่าสามารถจำแนกระดับความสามารถของผู้สอบได้ดี ณ ระดับความสามารถนั้น (Hambleton 1977 : 66)

$$I\{\theta\} = \sum_{i=1}^{n} \frac{P_i'^2}{P_i Q_i} \qquad \text{(Lord 1980 : 71)}$$

$I\{\theta\}$ คือ test information function

ถ้าเรามีกลุ่มของข้อกระทงที่ทราบ information curve เราสามารถสร้างแบบสอบให้มี test information curve ณ ระดับหนึ่งของความสามารถที่เราต้องการได้ เพื่อจุดประสงค์เฉพาะของการใช้แบบสอบ เช่น เพื่อคัดเลือกนักเรียนให้ได้รับทุนก็จะต้องใช้ข้อกระทงที่มีประสิทธิภาพสูงสุดที่ระดับความสามารถสูง ๆ นั้นก็คือให้มี test information สูง ณ ระดับความสามารถสูง ๆ นั่นเอง

รูปที่ 4 แสดงถึง Item information curve ของข้อกระทง 5 ข้อ และ Test information curve ของข้อกระทงทั้ง 5 ข้อนั้น (Lord 1980 : 22)



ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 1 แสดง: การเปรียบเทียบมโนทัศน์และวิธีวิเคราะห์แบบสอบของคลาสสิคอลโมเดลกับราสชโมเดล

วิเคราะห์ / โมเดลที่ใช้	ความยาก	อำนาจจำแนก	ความเที่ยง	ความตรง
คลาสสิคอล	สัดส่วนของผู้ที่เข้าสอบทั้งหมดที่ตอบข้อกระทงนั้น ๆ ถูก มีค่าระหว่าง 0 ถึง 1	ความสามารถของข้อกระทงที่จะชี้ให้เห็นความแตกต่างระหว่างกลุ่มที่ได้คะแนนสูงกับกลุ่มที่ได้คะแนนต่ำ มีวิธีประมาณค่าได้หลายวิธี เช่น พอยท์ไบซีเรียล, ไบซีเรียล เป็นต้น มีค่าตั้งแต่ -1 ถึง 1	ความคงที่ของคะแนนที่ได้จากการสอบบุคคลกลุ่มเดียวกันสองครั้งด้วยแบบสอบเดิมในเวลาที่ต่างกัน มีวิธีประมาณค่าได้หลายวิธี เช่น เพียร์สัน โพร์ดัก โมเมนต์ เป็นต้น มีค่าตั้งแต่ 0 ถึง 1	ความสามารถในการวัดในสิ่งที่เราต้องการวัดได้ถูกต้อง มีวิธีประมาณค่าหลายวิธี เช่น หาความสัมพันธ์กับเกณฑ์ เป็นต้น มีค่าตั้งแต่ 0 ถึง 1
ราสช	จุดบนแกน θ ที่สมนัยกับ $P_g(\theta)=.5$	สัดส่วนความชันของ ICC ณ จุดเปลี่ยนโค้ง มีค่าเป็น 1	ความคงที่ของการประมาณค่าความสามารถที่แท้จริง โดยพิจารณาจาก test information curve	การใช้แบบสอบให้เป็นไปตามจุดมุ่งหมายของการสอบ โดยพิจารณาจาก test information curve

ตอนที่ 4 งานวิจัยที่เกี่ยวข้อง

ไวท์ลี่ และ ดาวิส (Whitely and Dawis 1974 cited by Yen 1979 : 261) ได้นำแบบสอบอุปมาอุปไมยทางภาษา จำนวน 60 ข้อ สอบกับนักศึกษา และนักเรียนเตรียมอุดมศึกษา แล้ววิเคราะห์ด้วยราสชโมเดล ปรากฏว่ามี 30-40% ของข้อกระทงที่ไม่ fit กับโมเดล

เร็นซ์ และ ริเด็นเอาร์ (Rentz and Ridenour 1978 cited by Rentz 1979 : 3-5) ได้นำแบบสอบมาตรฐานมาทำการวิเคราะห์ด้วยราสชโมเดลใหม่มีทั้งหมด 3 ชุด ผลปรากฏว่าแบบสอบชุดแรกซึ่งเป็นชุดของแบบสอบผลสัมฤทธิ์ fit กับโมเดล 85-97% ชุดที่ 2 เป็นชุดของ Atlanta Assessment Project fit กับโมเดล 92-97% และชุดที่ 3 เป็นแบบสอบชุด Stanford Achievement Test ส่วนใหญ่ fit กับโมเดลมากกว่า 80% มีอยู่ 2 subtest ที่ fit กับโมเดลต่ำกว่า 80% คือ Primary I Vocabulary กับ Primary II Spelling

เร็นซ์ และ บาชอว์ (Rentz and Bashaw 1975 cited by Rentz 1979 : 7) ได้วิจัยพบว่าขนาดของกลุ่มตัวอย่างระหว่าง 500-1000 คน จะทำให้การประมาณค่าพารามิเตอร์มีความคงที่เพิ่มขึ้นอย่างรวดเร็ว และจะค่อย ๆ เพิ่มขึ้นเมื่อกลุ่มตัวอย่างที่ใช้ประมาณค่าพารามิเตอร์มีจำนวน 2000 และ 4000 คน

ฟอร์สเตอร์ (Forster 1976 cited by Rentz 1979 : 7) พบว่ากลุ่มตัวอย่างที่ใช้ประมาณค่าพารามิเตอร์ 200 คนก็เพียงพอที่จะให้ค่าพารามิเตอร์ที่ได้มีค่าคงที่ กลุ่มตัวอย่างที่มากกว่า 1000 คนจะไม่เพิ่มความคงที่ในการประมาณค่าพารามิเตอร์มากนัก