

# การกำหนดเกณฑ์ความสามารถของผู้สอบ

## โดยอาศัยรูปแบบราสช์

สุวัฒน์ สุขมลสันต์

### ความสำคัญของปัญหา

ในการวัดและประเมินผลสัมฤทธิ์ในการเรียน (learning achievement) หรือสมรรถภาพทั่วไป (general proficiency) ของผู้เรียนนั้น ปกติแล้วผล การวัดและประเมินด้วยการทดสอบมักจะเป็นคะแนนดิบ (raw score) ซึ่งผู้วัดและประเมินผลจะใช้ในการแปล ความหมาย (interpret) ระดับความสามารถของผู้สอบ ว่ามีมากน้อยเพียงใด ปัญหาที่มักเกิดขึ้นเสมอ ๆ ในการ วัดและประเมินผลทุกครั้งก็คือคะแนนดิบเหล่านี้มีความ หมายว่าอย่างไร เช่น ผู้สอบได้คะแนน 50 จากคะแนน ดิบ 100 คะแนน หมายความว่าอย่างไร และผู้สอบได้ คะแนน 60 คะแนน มีความสามารถแตกต่างจากผู้สอบ ได้ 50 คะแนนหรือไม่ และที่สำคัญมากก็คือ ผู้ที่สอบได้ คะแนนต่าง ๆ เหล่านี้สามารถทำอะไรได้บ้าง ปัญหาพื้น ฐานเหล่านี้มักไม่มีคำตอบในเชิงวิทยาศาสตร์ แต่มัก เป็นที่เข้าใจกันทั่วไปว่า ผู้ที่ได้คะแนนดิบมากกว่ามี ความสามารถมากกว่าผู้ที่ได้คะแนนน้อยกว่า ส่วนความ สามารถของผู้ที่ได้คะแนนต่าง ๆ นั้นคืออะไร ผู้ทำการวัด และประเมินผลทั่วไปไม่สามารถตอบได้ ซึ่งเป็นประเด็น ที่เป็นจุดอ่อนมากทางการศึกษา เพราะไม่สามารถบอก ได้ว่าผู้เรียนได้เรียนรู้อะไรแล้วบ้าง และยังไม่สามารถ เรียนรู้อะไรอีก ทำให้เป้าหมายในการเรียนและการสอน ขาดเป้าหมายที่ชัดเจน อาจเกิดทั้งความซ้ำซ้อนและการ ข้ามขั้นตอนในการเรียนรู้อันทำให้ผลการเรียนรู้ไม่ประสพ

ผลดีเท่าที่ควร

ดังนั้น ประเด็นที่สำคัญมากทางการทดสอบโดย เฉพาะอย่างยิ่งสำหรับการทดสอบแบบอิงเกณฑ์ (crite- rion-referenced test) ก็คือผู้สร้างแบบทดสอบจะ กำหนดเกณฑ์ความสามารถของผู้สอบได้อย่างไร จึงจะ ได้เกณฑ์ที่มีลักษณะเป็นวิทยาศาสตร์ที่สามารถพิสูจน์ ทราบได้ ไม่ใช่เกณฑ์ที่ตั้งขึ้นเอง เช่น ผู้สอบได้คะแนน ดิบร้อยละ 80 ขึ้นไป ได้ระดับคะแนน (grade) = A หรือแม้แต่เกณฑ์ที่ตั้งขึ้นจากความเห็นของคณะกรรมการ ที่เกี่ยวข้อง แต่ขาดการพิสูจน์ก็ไม่สามารถใช้เป็นเกณฑ์ ที่ดีในการทดสอบได้

### แนวทางแก้ปัญหา

ปัจจุบันนี้ แนวคิดทางการทดสอบและประเมินผล มีความก้าวหน้ามากกว่าแต่ก่อน สาเหตุหนึ่งเกิดจาก ความก้าวหน้าทางเทคโนโลยีทางการศึกษาสมัยใหม่คือ เครื่องคอมพิวเตอร์และความก้าวหน้าทางสถิติทางการ ศึกษา ทำให้แนวความคิดทางการทดสอบและประเมิน ผลก้าวหน้ามากขึ้นจนเกิดมีทฤษฎีทางการทดสอบใหม่ขึ้น เรียกว่า ทฤษฎีการตอบสนองของข้อทดสอบ (Item Response Theory) หรือที่รู้จักกันดีทั่วไปขณะนี้ว่าทฤษฎี IRT ทฤษฎีดังกล่าวนี้ มีหลายรูปแบบ เช่น

1) รูปแบบ 1 พารามิเตอร์ (One-parameter Model) หรือที่นิยมเรียกกันทั่วไปว่า Rasch Model เพื่อ

เป็นเกียรติแก่ Georg Rasch นักคณิตศาสตร์ชาวเดนมาร์กที่คิดค้นขึ้น รูปแบบนี้ใช้ค่าความยาก (difficulty index:b) ในสูตรคำนวณหาคุณภาพของข้อทดสอบ

2) รูปแบบ 2 พารามิเตอร์ (Two-parameter Model) อาศัยค่าพารามิเตอร์ 2 ค่าในการหาคุณภาพของข้อทดสอบ คือค่าความยาก และค่าอำนาจจำแนก (discrimination index:a)

3) รูปแบบ 3 พารามิเตอร์ (Three-parameter Model) อาศัยค่าพารามิเตอร์ 3 ค่าในการหาคุณภาพของข้อทดสอบ คือค่าความยาก ค่าอำนาจจำแนก และค่าการเดาถูก (guessing index:c)

4) รูปแบบ 4 พารามิเตอร์ (Four-parameter Model) อาศัยค่าพารามิเตอร์ 4 ค่าในการหาคุณภาพของข้อทดสอบ คือค่าความยาก ค่าอำนาจจำแนก ค่าการเดาถูก และค่าความคลาดเคลื่อนต่าง ๆ (error index:e)

ในการกำหนดเกณฑ์ความสามารถของผู้สอบแบบทดสอบนั้น ผู้สร้างแบบทดสอบอาจอาศัยรูปแบบต่าง ๆ ของทฤษฎี IRT ดังกล่าวแล้วได้ ยกเว้นรูปแบบ 4 พารามิเตอร์ เพราะเป็นรูปแบบใหม่ล่าสุดที่เป็นเพียงแนวคิดและสูตรในการคำนวณหาค่าพารามิเตอร์ต่าง ๆ ยังมีปัญหาอยู่ ปกติแล้ว การคำนวณหาคุณภาพของแบบทดสอบหรือข้อทดสอบโดยอาศัยรูปแบบ 2 พารามิเตอร์ และ 3 พารามิเตอร์ เสียค่าใช้จ่ายมากเพราะว่ามักต้องอาศัยกลุ่มตัวอย่างประชากรขนาดใหญ่ (คือมากกว่า 1,000 คน) และใช้เวลาในการคำนวณหาค่าพารามิเตอร์นานกว่าจะได้ค่าพารามิเตอร์ที่คงที่ (convergent values) ดังนั้น นักทดสอบทั่วไปหรือครู-อาจารย์ทั่วไป จึงนิยมใช้รูปแบบ 1 พารามิเตอร์มากกว่าเพราะว่าสามารถใช้ได้กับกลุ่มตัวอย่างขนาดเล็กได้ดีกว่า เช่น ประมาณ 100 คนก็ใช้ได้และเสียเวลารวมทั้งค่าใช้จ่ายในการดำเนินงานเพื่อให้ได้ค่าพารามิเตอร์ที่คงที่น้อยกว่ามาก

ดังนั้นในที่นี้ผู้เขียนจึงขอแนะนำการใช้รูปแบบราสชหรือรูปแบบ 1 พารามิเตอร์ เพื่อกำหนดเกณฑ์ความสามารถของผู้สอบเพื่อใช้ในการวัดผลแบบอิงเกณฑ์ทั้งนี้

เพื่อแนะนำให้ท่านที่สนใจนำไปประยุกต์ใช้กับการวัดและประเมินผลการเรียนของท่านผู้อ่านต่อไปได้

## การตรวจสอบเอกสารที่เกี่ยวข้อง

ในทางทฤษฎีการทดสอบแล้ว การที่จะสร้างแบบทดสอบใด ๆ นั้น สิ่งที่คุณสร้างแบบทดสอบควรต้องคำนึงถึงมากที่สุดคือ ความตรงตามสภาวะสันนิษฐาน (construct validity) ซึ่งได้แก่แนวคิดเบื้องหลังในเชิงทฤษฎีว่าแบบทดสอบนั้นจะวัดอะไร มากน้อยเพียงไร และทำไมจึงต้องวัดสิ่งต่าง ๆ เหล่านั้น เช่น ในกรณีของการทดสอบความสามารถในการอ่านเข้าใจความภาษาอังกฤษ (reading comprehension) นั้น จะต้องวัดความสามารถด้านใดบ้างจึงจะทำให้ผลการสอบสะท้อนออกมาว่าผู้สอบมีความรู้ความสามารถในการอ่านเข้าใจความ ผู้สร้างแบบทดสอบอาจจะต้องการวัดความสามารถต่าง ๆ เช่น การอ้างอิงโดยตรง (direct reference) การอนุมาน (inference) การคาดคิด (supposition) และการประเมินผล (evaluation) เป็นต้น (Davies and Widdowson 1974 : 168-176) และในบางระดับผู้สร้างแบบทดสอบอาจต้องการวัดความสามารถอื่น เช่น การถอดความ (paraphrasing) การทดสอบคำศัพท์ในและนอกบริบท (vocabulary in and out of context) และการสรุปเนื้อเรื่อง (conclusion) เป็นต้น และเมื่อกำหนดขอบเขตของความตรงตามสภาวะสันนิษฐานได้แล้ว จึงกำหนดความตรงเชิงเนื้อหา (content validity) ตามประเด็น ปัญหาก็คือ เมื่อนำแบบทดสอบไปใช้แล้ว ผู้ทำการทดสอบจะทราบได้อย่างไรว่า ข้อทดสอบแต่ละข้อสามารถทำหน้าที่ในการวัดสิ่งที่มุ่งจะทำการวัดได้ และข้อที่ต้องการจะวัดสิ่งเดียวกันสามารถวัดสิ่งเดียวกันได้หรือไม่การที่จะหาคำตอบดังกล่าวนี้ได้ นักทดสอบได้อาศัยวิธีการตรวจสอบหลายอย่างดังนี้

1. วิเคราะห์องค์ประกอบ (Factor Analysis) ของผลการทดสอบเพื่อตรวจสอบว่าข้อทดสอบที่มุ่งทำการวัดความสามารถเดียวกัน สามารถจับกลุ่มรวมกันได้หรือไม่ วิธีนี้เป็นวิธีที่ง่ายในเชิงการคำนวณ เพราะใน

ปัจจุบันนี้อาจจะสามารถใช้เครื่องคอมพิวเตอร์ช่วยได้ แต่ในด้านการดำเนินการมีความจำกัด กล่าวคือ วิธีการวิเคราะห์องค์ประกอบมีหลายวิธี แต่ละวิธีมีแนวคิดเชิงทฤษฎีแตกต่างกันทำให้ผลการวิเคราะห์แตกต่างกัน นอกจากนี้ต้องอาศัยกลุ่มตัวอย่างผู้สอบจำนวนมาก คือ ประมาณ 10 เท่าของจำนวนข้อทดสอบ ซึ่งในระยะเวลาพัฒนาเครื่องมือทดสอบมักเป็นปัญหาในการดำเนินการ

2. วิธีหาความสัมพันธ์ระหว่างข้อของผลการสอบ (correlation matrix) เพื่อตรวจสอบว่าข้อทดสอบที่วัดสิ่งเดียวกันมีความสัมพันธ์กันสูงหรือไม่ วิธีนี้ง่ายในเชิงการคำนวณ แต่ไม่สามารถตอบได้ว่าข้อสอบแต่ละข้อมีความยากง่ายแตกต่างกันหรือไม่ ทำให้ไม่สามารถทราบได้ว่า ผลของคะแนนรวมแต่ละข้อมีความหมายว่าอย่างไร

3. วิธีวิเคราะห์ผลการสอบตามทฤษฎีการตอบสนองของข้อทดสอบ (Item Response Theory) ซึ่งมีหลายวิธี ดังได้กล่าวมาแล้ว

สิ่งที่วิธีการวิเคราะห์ต่าง ๆ ดังกล่าวแล้วข้างต้นนี้ สิ่งมุ่งแสวงหาก็คือลำดับที่ของความยากของข้อทดสอบ (item difficulty order) ทั้งฉบับ เมื่อคำนึงถึงค่าพารามิเตอร์ต่าง ๆ ที่เกี่ยวข้องและการรวมกลุ่มกันขององค์ความรู้ที่มุ่งทำการทดสอบ ปัจจุบันทั้ง 2 ประการนี้เป็นสิ่งสำคัญมากในการกำหนดเกณฑ์ความสามารถของผู้สอบแต่ละคน และการวิเคราะห์ข้อทดสอบด้วยรูปแบบราสซซ์จะได้คำตอบทั้งสองประการดังกล่าวนี้ด้วย

## รูปแบบราสซซ์ (Rasch Model)

### 1. ประวัติความเป็นมาโดยสังเขป

ในปี 1960 Georg Rasch นักคณิตศาสตร์ชาวเดนมาร์กได้เสนอแนวคิดเกี่ยวกับการทดสอบตามรูปแบบ 1 พารามิเตอร์ขึ้น และได้ตีพิมพ์บทความในทำนองเดียวกันอีกในปี 1961 และ 1966 ต่อมา Benjamin Wright ได้เผยแพร่แนวคิดนี้ให้เป็นที่แพร่หลายทั่วไปตั้งแต่ปี 1967 เป็นต้นมา และเรียกชื่อแนวคิดนี้ว่า Rasch Model แนวคิดที่สำคัญของ Rasch Model ก็คือแนวคิดที่เชื่อ

ว่าความเป็นปรนัยของการวัดผล (objectivity of measurement) ที่ไม่สามารถจะหาได้จากทฤษฎีการทดสอบแบบเดิม (Classical Test Theory) ทั่ว ๆ ไปก็คือ

1. ความเป็นอิสระจากกลุ่มตัวอย่าง (sample - free test calibration) กล่าวคือ การคำนวณหาค่าพารามิเตอร์ของข้อทดสอบเป็นอิสระจากกลุ่มตัวอย่างค่าต่าง ๆ ของข้อทดสอบ เช่น ค่าความยาก ( $b$ ) จะไม่แปรเปลี่ยนไปตามลักษณะกลุ่มตัวอย่าง และ

2. ความเป็นอิสระจากข้อสอบ (item - free person measurement) กล่าวคือ การคำนวณหาค่าพารามิเตอร์ของบุคคลเป็นอิสระจากข้อทดสอบ เช่น ความสามารถของบุคคล ( $\theta$ ) จะไม่แปรเปลี่ยนไปตามลักษณะของข้อทดสอบ ความสามารถของบุคคลจะคงที่ไม่ว่าจะวัดด้วยข้อทดสอบใด ข้อทดสอบหนึ่ง ๆ ที่คนเคยทำเลยแล้ว จะมีค่าคงที่ เมื่อวัดกับบุคคลเดิม ไม่ว่าข้อสอบนั้นจะไปปรากฏอยู่ที่ส่วนใดของแบบทดสอบ

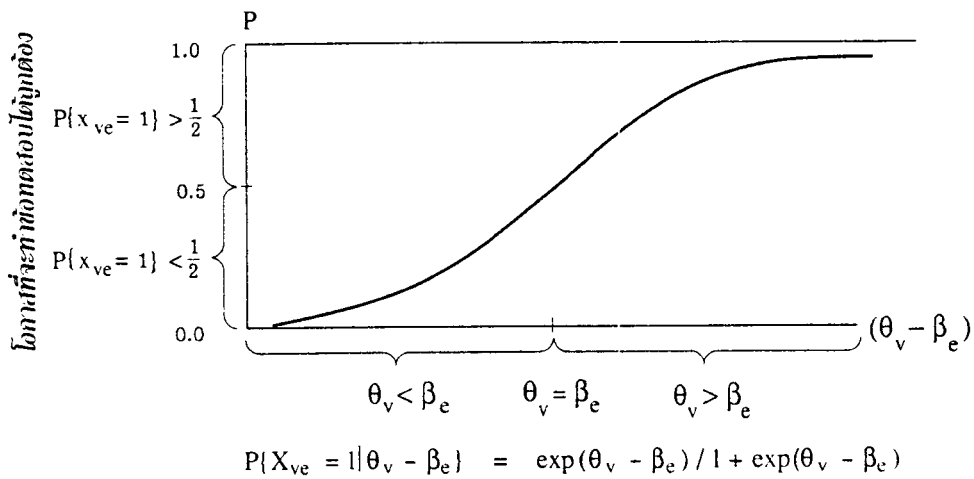
แนวความคิดดังกล่าวนี้นับว่ามีประโยชน์มากในการทดสอบและวัดผล โดยเฉพาะการทดสอบเกี่ยวกับความสามารถทางสมอง (mental measurement) ในปี 1969 Wright และ Panchapakesan ได้ทำการทดสอบแนวคิดดังกล่าวแล้ว และสนับสนุนว่าเป็นไปได้จริง ๆ และในปีต่อ ๆ มาได้มีนักทดสอบได้ทำการทดสอบแนวคิดดังกล่าวอีกหลายครั้ง และยืนยันว่า คุณลักษณะดังกล่าวนี้เป็นไปได้จริงแม้แต่กลุ่มตัวอย่างจะไม่ได้จากการสุ่มมาก็ตาม และมีความแกร่งมาก (robustness) คือสามารถทดสอบเพื่อคำนวณหาความสามารถของบุคคล ( $\theta$ ) และความยากของข้อสอบ ( $b$ ) ได้ทั้ง ๆ ที่กลุ่มตัวอย่างบุคคลและข้อทดสอบแตกต่างจากข้อตกลงเบื้องต้นของการใช้รูปแบบนี้ก็ตาม เช่น ขนาดของกลุ่มตัวอย่างมีเพียง 100 คนก็ได้ และไม่ต้องได้มาจากการสุ่มก็ได้ (Wright 1997: 219)

### 2. แนวคิดของรูปแบบราสซซ์

ตามแนวคิดของรูปแบบราสซซ์นั้น โอกาสที่คนจะทำข้อสอบได้หรือไม่ขึ้นอยู่กับระดับความสามารถของตนเอง

(ability parameter;  $\theta$ ) และระดับความยากของข้อสอบ (difficulty index :b) เช่น ถ้า  $b = .5$  และ  $\theta = .5$  โอกาสที่คนผู้นั้นจะสามารถทำข้อสอบข้อนั้นได้ถูกต้องประมาณ 50% ถ้าหากความสามารถของบุคคล ( $\theta$ ) น้อยกว่าความสามารถของข้อสอบ (ความยากหรือ

b) แล้ว โอกาสที่จะตอบข้อทดสอบนั้นได้ถูกต้องก็ย่อมจะน้อยกว่า 50% และในทำนองเดียวกัน ถ้าหากว่าค่า  $\theta$  มากกว่าค่า  $b$  แล้ว โอกาสที่คนผู้นั้นจะตอบข้อทดสอบได้ถูกต้องก็มีมากกว่า 50 % (ดูภาพประกอบที่ 1)



**ความสัมพันธ์ระหว่าง  $\theta_v$  และ  $\beta_e$  ของตัวแปรและโด่งลักษณะเฉพาะของข้อทดสอบ (item characteristic curve: ICC) ของแบบราสซ์**

ดังนั้น ตามแนวความคิดของ Rasch Model นั้น ค่าพารามิเตอร์ที่เกี่ยวข้องมีเพียง 2 ค่าเท่านั้นไม่มีอำนาจจำแนก หรือโอกาสของการเดาถูกเข้ามาเกี่ยวข้องด้วย เพราะเป็นข้อตกลงเบื้องต้นว่าข้อสอบที่นำมาใช้กับรูปแบบราสซ์นั้น จะต้องมีย่านอำนาจจำแนกเท่า ๆ กันหรือมีลักษณะที่จะทำให้เกิดการเดาได้น้อยที่สุด แต่ในทางปฏิบัติลักษณะทั้งสองนี้ ความแกร่ง (robustness) ของรูปแบบนี้สามารถจะแก้ไขได้

**3. พัฒนาของสูตรของรูปแบบราสซ์**

ดังได้กล่าวมาแล้ว แนวคิดของรูปแบบราสซ์ เกี่ยวข้องกับความสัมพันธ์ของความสามารถของบุคคล ( $\theta$ ) และความยากของข้อทดสอบ ( $b$ ) เท่านั้น ความสัมพันธ์นี้เป็นความสัมพันธ์กับเชิงคณิตศาสตร์ กล่าวคือ โอกาส

ที่บุคคล  $V$  ที่มีระดับความสามารถ  $\theta$  (หรือ  $\theta_v$ ) จะทำข้อสอบ  $e$  ที่มีระดับความยากง่าย  $b$  (หรือ  $\beta_e$ ) จะตอบข้อทดสอบได้ถูกมากน้อยเพียงใดขึ้นอยู่กับระดับความแตกต่างของ  $(\theta_v - \beta_e)$  นั่นคือ

โอกาสของความสำเร็จ (odd of success) =  $(\theta_v - \beta_e)$

แต่เนื่องจาก  $(\theta_v - \beta_e)$  นั้น มีค่าระหว่าง  $\pm\infty$  แต่ที่โอกาสของความสำเร็จมีได้ระหว่าง 1 กับ 0 เท่านั้น เพื่อให้ค่าของ  $(\theta_v - \beta_e)$  เป็นค่าที่มีหน่วยเล็กลง และคงที่เหมาะสมแก่การนำมาใช้ จึงใช้ค่า exponent ของ  $(\theta_v - \beta_e)$  แทน ซึ่งจะมีค่าอยู่ระหว่าง 0 กับ  $+\infty$  เช่น

$$\begin{aligned} \text{odd of success} &= e^{(\theta_v - \beta_e)} \\ &= \text{EXP}(\theta_v - \beta_e) \dots\dots\dots 1 \end{aligned}$$

หน่วยของการวัดเรียกว่า log-odds scale หรือ logits (ค่า  $e = 2.71828.....$ )

เพื่อให้ค่า  $\exp (\theta_v - b_e)$  มีค่าเป็นมาตราวัดอันตรภาค (interval scale) ระหว่าง 0 และ 1 สมการที่ 1 อาจเขียนได้เป็น

$$\text{odd of success} = \frac{\text{EXP} (\theta_v - b_e)}{1 + \text{EXP} (\theta_v - b_e)} \dots\dots\dots 2$$

ดังนั้น สูตรที่แสดงว่าโอกาสที่บุคคล V ที่มีความสามารถ  $\theta$  จะทำข้อสอบ e ที่มีความยาก b ได้ถูกต้อง (ได้คะแนน = 1 หรือ  $X_{v,e}$ ) คือ

$$P \{ X_{v,e} = 1 \mid \theta_v, b_e \} = \frac{\text{EXP} (\theta_v - b_e)}{1 + \text{EXP} (\theta_v - b_e)} \dots\dots\dots 3$$

นี่คือสูตรของรูปแบบราสช์ ตัวอย่างเช่น

ถ้า  $(\theta_v - b_e) = 2 \log - \text{odds}$  หน่วย โอกาสที่บุคคลจะทำข้อสอบข้อที่ e ได้ถูกต้องคือ

$$P \{ X_{v,e} = 1 \mid \theta_v, b_e \} = \frac{e^2}{1+e^2} = \frac{2.71828^2}{1+2.71828^2} = .88 \text{ หรือ } 88.1\%$$

ความสัมพันธ์ระหว่างโอกาสที่จะตอบข้อสอบได้ถูกต้องกับระดับความยากของข้อทดสอบ มีการกระจายเป็นรูปคล้ายตัวเอส (ogive distribution) ต่อมาเพื่อให้การกระจายของโค้งดังกล่าวเป็นโค้งรูปตัวเอสปกติ (normal ogive) จึงนิยมใช้ค่า 1.7 เข้ามาเกี่ยวข้องกับในการคำนวณด้วยเพื่อปรับโค้งของความสัมพันธ์นี้ ปกติค่าที่นำมาปรับนี้ทำให้ความแตกต่างของโค้งทั้ง 2 ลักษณะ แต่แตกต่างกันน้อยกว่า 1%

**4. ข้อตกลงเบื้องต้นในการใช้รูปแบบราสช์**

ข้อตกลงเบื้องต้นในการใช้รูปแบบราสช์มีดังนี้ (Hambleton and Cook, 1997: 78; Tinsley and Davis, 1975: 326)

1. ข้อทดสอบแต่ละข้อจะต้องวัดคุณลักษณะ (trait) เดียวกัน (unidimensional latent space) กล่าวคือ ข้อทดสอบจะต้องเป็นเอกพันธ์ (homogeneous items) ในการที่จะตัดสินว่าข้อทดสอบมีลักษณะดังกล่าวหรือไม่ อาจทำได้โดยการอาศัยการวิเคราะห์ตัวประกอบ (Factor Analysis) ถ้าข้อสอบเป็นวิวิธพันธ์ (heterogeneous items) จะต้องมีการวิเคราะห์ดังกล่าวเพื่อจัดกลุ่มเอกพันธ์ก่อน

2. ข้อทดสอบจะต้องมีความเป็นอิสระจากตำแหน่ง (local independence) ใน 2 ลักษณะคือ

2.1 มีความเป็นอิสระทางสถิติ (Statistically independence) กล่าวคือข้อทดสอบแต่ละข้อเป็นอิสระ ไม่เกี่ยวข้องกัน แต่ละข้อวัดความสามารถที่ไม่ซ้ำกันเลย ดังนั้น คำตอบของแต่ละข้อเป็นอิสระต่อกัน แต่รวมกันแล้วจะวัดคุณลักษณะเดียวกัน

2.2 มีความเป็นอิสระเสรี (uncorrelated independence) กล่าวคือ ข้อสอบแต่ละข้อจะปรากฏอยู่ในตำแหน่งใดของข้อสอบก็ได้ จะไม่มีผลต่อการตอบของผู้ตอบข้อทดสอบ

3. คะแนนมี 2 ลักษณะเท่านั้นคือเป็น 1 หรือ 0

4. ความเร็วในการทำข้อทดสอบไม่มีผลต่อโอกาสที่จะทำข้อสอบได้หรือไม่

5. โอกาสที่จะทำข้อทดสอบได้ถูกต้องหรือไม่ ขึ้นอยู่กับความสัมพันธ์ของความสามารถของบุคคลกับความยากของข้อทดสอบเท่านั้น

**5. ความแกร่ง (robustness) ของรูปแบบราสช์**

ในการใช้รูปแบบราสช์เพื่อการทดสอบนั้น ได้มีการทดลองใช้ในรูปแบบที่ต่าง ๆ กันเพื่อหาความแกร่งของรูปแบบนี้ ปรากฏว่ารูปแบบนี้ได้รับความนิยมนมาก เพราะมีความแกร่งมากดังนี้

1. ขนาดของกลุ่มตัวอย่างที่จะใช้กับรูปแบบราสช์ อาจจะมีขนาดเล็กประมาณ 100 คน เท่านั้นก็ได้ (Wright, 1977: 219)

2. การกระจายของความสามารถของกลุ่มตัวอย่าง หรือความยากของข้อทดสอบไม่จำเป็นต้องเป็นโค้งปกติ เพียงแต่มีแนวโน้มว่าจะเป็นโค้งปกติเท่านั้นก็ได้ และตัวอย่างไม่จำเป็นที่จะได้มาเพราะการสุ่มก็ได้ (Wright and Stone, 1979:20)

3. ข้อทดสอบไม่จำเป็นต้องมีอำนาจจำแนกเท่ากันก็ได้ และไม่จำเป็นต้องคำนึงถึงค่าการแตกก็ได้ (Wright and Panchapakesan, 1969:25)

4. ข้อทดสอบที่ให้คะแนนหลากหลายมากกว่า 0, 1 (multiple-point) ก็ใช้กับรูปแบบราสซ์ได้ (Willmott, 1980:195)

## ขั้นตอนในการกำหนดเกณฑ์ความสามารถของผู้สอบ

เพื่อให้ได้เกณฑ์ความสามารถของผู้สอบแบบทดสอบต่าง ๆ ผู้สร้างแบบทดสอบควรดำเนินการตามขั้นตอนต่าง ๆ ต่อไปนี้คือ

### 1. สร้างแบบทดสอบ

แบบทดสอบที่สร้างขึ้น ควรมีลักษณะเป็นแบบทดสอบปรนัย (objective test) ที่ข้อทดสอบแต่ละข้อมีความเป็นอิสระจากตำแหน่งของกันและกัน กล่าวคือข้อทดสอบแต่ละข้อ ต้องไม่ให้แนวคำตอบ (cue) ของข้อทดสอบข้ออื่นให้สามารถเดาคำตอบ หรือหาคำตอบได้ ข้อทดสอบดังกล่าวนี้อาจเป็นข้อทดสอบแบบเลือกตอบ (multiple - choice item) ที่มีตัวเลือกจำนวนเท่ากันตลอดส่วนข้อทดสอบแบบ cloze นั้น มักไม่มีอิสระจากตำแหน่ง จึงไม่ควรเลือกใช้ เว้นแต่ผู้สร้างข้อทดสอบจะสามารถจัดกระทำให้ข้อทดสอบมีความเป็นอิสระดังกล่าวได้

นอกจากนี้ แบบทดสอบควรมีความยาวมากเพียงพอต่อการทดสอบ เพื่อให้เกิดมีความเที่ยง (reliability) ของการทดสอบ เช่น ควรมีประมาณ 70-100 ข้อ เป็นต้น และควรคำนึงถึงเรื่องความตรง (validity) ของแบบทดสอบทั้งความตรงตามเนื้อหา (content validity)

และความตรงตามสภาวะสันนิษฐาน (construct validity) ด้วย จึงจะเป็นข้อทดสอบที่ดีได้

### 2. นำแบบทดสอบไปทดลองใช้หรือใช้จริง

ในกรณีที่ไปใช้ได้หรือในกรณีที่ผู้สร้างแบบทดสอบต้องการจะสร้างเกณฑ์วัดระดับความสามารถของผู้เรียนให้เป็นเกณฑ์มาตรฐานหรือเกณฑ์อ้างอิงได้อย่างมั่นใจ ควรนำแบบทดสอบที่สร้างขึ้นไปทดลองสอบกับกลุ่มตัวอย่างจำนวนหนึ่ง ซึ่งควรมีขนาดไม่น้อยกว่า 100 คน แต่หากได้กลุ่มตัวอย่างขนาดใหญ่กว่านี้ก็จะเป็นการดี และถ้าเป็นไปได้ กลุ่มตัวอย่างนี้ควรสุ่มมาจากประชากรเป้าหมายที่ผู้สร้างแบบทดสอบจะนำแบบทดสอบไปใช้จริงภายหลัง

แต่อย่างไรก็ตาม หากว่าไม่สามารถนำแบบทดสอบไปทดลองสอบกับกลุ่มตัวอย่างดังกล่าวแล้วได้ ก็อาจใช้สอบจริงกับกลุ่มเป้าหมายก็ได้ แม้ว่าวิธีนี้จะไม่ดีนัก แต่หากมีความจำเป็นก็สามารถกระทำได้

### 3. นำผลการสอบจากขั้นที่ 2 ไปวิเคราะห์ข้อสอบตามรูปแบบประเพณีนิยม

เพื่อปรับปรุงแบบทดสอบให้มีคุณภาพดีขึ้น ผู้สร้างแบบทดสอบควรนำผลการสอบจริงหรือผลการทดลองสอบไปวิเคราะห์ข้อสอบรายข้อ ตามรูปแบบประเพณีนิยม (Classical Model) เช่น ใช้วิธี 27% หรือ 50% โดยอาศัยโปรแกรมคอมพิวเตอร์ก็ได้ เช่น ใช้โปรแกรม CTIA/Grading ปรับปรุงคุณภาพของตัวเลือกของข้อทดสอบแต่ละข้อ และคุณภาพของแบบทดสอบอื่น ๆ ที่จำเป็น (รายละเอียดศึกษาได้จาก สุพัฒน์ สุขมลสันต์, 2538)

### 4. ศึกษาความเป็นเอกมิติของแบบทดสอบ

เพื่อให้แน่ใจว่าแบบทดสอบทั้งฉบับวัดความรู้ความสามารถของผู้สอบเพียงมิติเดียวผู้สร้างแบบทดสอบควรทำการศึกษาความเป็นเอกมิติ (unidimensionality) เช่น ทำการวิเคราะห์ผลการสอบด้วยการวิเคราะห์องค์ประกอบ

(Factor Analysis) โดยใช้โปรแกรมคอมพิวเตอร์ชื่อ SPSS/PC (Statistical Package for the Social Science for Personal Computer) หากว่าค่าเกณฑ์เจาะจง (eigenvalue) ขององค์ประกอบที่ 1 มีขนาดแตกต่างจากค่าดังกล่าวขององค์ประกอบที่ 2 มาก และค่าขององค์ประกอบที่ 2 แตกต่างจากค่าขององค์ประกอบอื่น ๆ น้อยมากก็แสดงว่าแบบทดสอบวัดความรู้ความสามารถของผู้สอบเพียงมิติเดียว หรือคำนวณหาความสัมพันธ์ระหว่างค่าสัมประสิทธิ์สหสัมพันธ์แบบพอยท์ไบซีเรียล ( $r_{pb}$ ) รายข้อที่ได้จากโปรแกรม CTIA/Grading กับค่าน้ำหนักองค์ประกอบรายข้อ (factor loading) ขององค์ประกอบที่ 1 ที่ได้จากโปรแกรม SPSS/PC หากว่าค่าสัมประสิทธิ์สหสัมพันธ์สูงกว่า 0.80 ขึ้นไป แสดงว่าแบบทดสอบวัดเพียงมิติเดียว (Warm, 1979: 101)

## 5. วิเคราะห์ผลการสอบรายข้อตามรูปแบบราชศัพท์

นำผลการทดสอบรายข้อมาวิเคราะห์ตามรูปแบบราชศัพท์ โดยใช้โปรแกรมคอมพิวเตอร์ เช่น ใช้โปรแกรม BICAL/PC (Binomial calibration for Personal Computer ซึ่งผู้เขียนได้เขียนขึ้นใช้) โปรแกรม Logist ของ ETS (Educational Testing Service) โปรแกรม Quest ของ ACER (Australian Centre for Educational Research) และโปรแกรม ASCAL ของบริษัท MicroCAT แห่งอเมริกา เป็นต้น โปรแกรมเหล่านี้จะให้ค่าสถิติที่สำคัญหลายอย่างแต่ที่สำคัญมากสำหรับการกำหนดเกณฑ์ความสามารถของผู้สอบคือ

1. ค่าความยากของข้อทดสอบรายข้อ (b) ที่เรียงจากข้อที่ง่ายที่สุดไปหาข้อที่ยากที่สุด
2. แผนภาพ (chart) แสดงการกระจายของค่าความยากของข้อทดสอบรายข้อและค่าความถี่ของการกระจายซึ่งมักเรียกกันว่าแผนที่ตัวแปร (map of variable map) และ
3. แผนภาพแสดงการตอบข้อทดสอบถูกและผิด

ของผู้ตอบแต่ละคนในการตอบข้อทดสอบแต่ละข้อ ซึ่งมักเรียกว่าแผนที่ผู้ตอบข้อสอบ (kid map)

ผลการวิเคราะห์ข้อทดสอบข้อที่ 1-2 มีประโยชน์ในการกำหนดเกณฑ์ความสามารถของผู้สอบข้อทดสอบและสารสนเทศ (information) จากข้อที่ 3 มีประโยชน์ในการตรวจสอบเกณฑ์ที่ผู้สร้างแบบทดสอบกำหนดขึ้นเมื่อมีความต้องการ

## 6. จัดกลุ่มระดับความสามารถ

นำแผนที่ตัวแปร (map of variable) มาจัดกลุ่มระดับความสามารถ โดยอาศัยความรู้เกี่ยวกับการกำหนดระดับคะแนน (grading) มาประยุกต์ใช้ (Brown, 1994) กล่าวคือ

1. ศึกษาดูว่าการกระจายของข้อทดสอบทั้งหมดมีความต่อเนื่องกันหรือไม่โดยดูที่ค่าความถี่ภายใต้ระดมภ์ (column) ชื่อ item count

2. หากที่ใดมีค่าความถี่น้อยหรือไม่มีเลยให้ศึกษาว่าข้อทดสอบที่อยู่ในกลุ่มเดียวกันทดสอบสิ่งเดียวกันหรือไม่ โดยศึกษาจากตารางกำหนดลักษณะจำเพาะของข้อทดสอบ (Table of Test Specification) ของแบบทดสอบที่สร้างขึ้น หรือศึกษาจากความมุ่งหมายของการทดสอบของข้อทดสอบแต่ละข้อในกลุ่มเดียวกันนั้น

อนึ่ง ในขั้นนี้ หากผู้สร้างแบบทดสอบสามารถวิเคราะห์ข้อทดสอบได้ว่าแต่ละข้อมุ่งทดสอบความรู้ความสามารถใดแล้วให้รหัสข้อทดสอบก่อนการนำข้อทดสอบไปวิเคราะห์ด้วยรูปแบบราชศัพท์จะดีมาก เช่น ข้อทดสอบข้อที่ 1 ทดสอบเรื่องความรู้ความสามารถในการอ้างอิงโดยตรง (direct reference) จากการอ่านเข้าใจความภาษาอังกฤษ ผู้สร้างแบบทดสอบก็อาจให้รหัสว่า DR1 เป็นต้น

3. ลองปรับปรุงการรวมกลุ่มกันของข้อทดสอบที่อยู่ในบริเวณใกล้เคียงกัน และทดสอบความรู้ความสามารถเดียวกัน หรือใกล้เคียงกัน แล้วจึงลากเส้นแบ่งกลุ่มข้อทดสอบออกเป็นกลุ่ม ๆ

4. พิจารณาตั้งชื่อกลุ่มของข้อทดสอบแต่ละกลุ่มว่า

ทดสอบความรู้ความสามารถของผู้สอบด้านใด แล้วเขียนชื่อพร้อมบรรยายสรุปสิ่งที่กลุ่มข้อทดสอบแต่ละกลุ่มมุ่งทดสอบ

ตัวอย่างเช่น แบบทดสอบวัดความสามารถในการอ่านเข้าใจความภาษาอังกฤษยาว 45 ข้อ วิเคราะห์แล้วได้แผนตัวแปรดังนี้

MAP OF VARIABLE

PERSON	RAW	MEASURE	ITEM	TYPICAL	ITEMS	(BY NAME)
STATS	COUNT	SCORE	MIDPOINT(S.E.)	COUNTS		
-5SD	1	44	4.10(.04)	1		
		11	3.50(.04)	1		
		11	3.70(.04)	1		
		43	3.50(.05)	1		
-4SD		11	3.30(.05)	1		
	2	42	3.10(.03)	1		
		11	2.90(.03)	1		
	2	41	2.70(.05)	1	1	41
		11	2.50(.05)	1		
+3SD	2	40	2.30(.05)	1		
		79	2.10(.07)	1		
	1	36	1.90(.04)	1	1	20
+2SD	6	36	1.70(.04)	1	1	45
	11	35	1.50(.03)	2	1	21 44
	13	34	1.30(.03)	1		
	14	32	1.10(.03)	1	1	19
+1SD	23	30	.90(.03)	1	1	18
	13	29	.70(.04)	5	1	3 23 25 26 42
	34	27	.50(.03)	2	1	5 27
	34	25	.30(.03)	3	1	22 40 43
MEAN	43	23	.10(.03)	3	1	31 36 39
	31	21	-.10(.03)	5	1	6 15 17 30 35
	27	19	-.30(.03)	4	1	10 11 34 38
	36	17	-.50(.04)	4	1	16 29 32 33
-1SD	12	16	-.70(.04)	6	1	7 8 13 14 24 28
	34	14	-.90(.03)	1		
	24	12	-1.10(.03)	1		
	12	11	-1.30(.03)	2	1	4 9
-2SD	3	10	-1.50(.03)	2	1	2 12
	5	8	-1.70(.04)	1	1	37
	2	7	-1.90(.04)	1		
	6	6	-2.10(.07)	1		
-3SD	5	5	-2.30(.05)	1		
		11	-2.50(.05)	1		
		4	-2.70(.05)	1	1	1
		3	-2.90(.03)	1		

ระดับวิพากษ์วิจารณ์

ระดับการสื่อสาร

ระดับการถ่ายโอน

45 ITEMS CALIBRATED ON 400 PERSONS  
400 MEASURABLE PERSONS WITH MEAN ABILITY = .04 AND STD. DEV. = .81

**การกำหนดจุดตัดของคะแนนสอบ**

ผู้สร้างแบบทดสอบอาจจัดกลุ่มข้อทดสอบได้ดังนี้

**1. ระดับที่ 1: ระดับการถ่ายโอน (Transfer Level)**

ข้อทดสอบในระดับนี้มี 16 ข้อคือ ข้อ 1, 2, 12, 37, 4, 9, 7, 8, 13, 14, 24, 28, 16, 29, 32 และ 33

ผู้สอบในระดับนี้สามารถอ่านเข้าใจข้อความที่เกี่ยวข้องกับข้อเท็จจริงที่เป็นความจริงและตัวเลข (facts and

figures) เช่น สถานที่ เวลา จำนวนบุคคล ชื่อบุคคล และข้อมูลง่าย ๆ จากเนื้อเรื่องง่าย ๆ และบทสนทนาสั้น ๆ เป็นต้น ผู้มีความรู้ความสามารถระดับนี้ได้คะแนนสอบประมาณ 1-16 คะแนน

**2. ระดับที่ 2: ระดับการสื่อสาร (Communication Level)**

ข้อทดสอบในระดับนี้มี 17 ข้อ คือ ข้อ 10, 11, 34,



38, 6, 15, 17, 30, 35, 31, 36, 39, 22, 40, 43, 5 และ 27

ผู้สอบในระดับนี้สามารถอ่านเข้าใจข้อความที่เกี่ยวข้องกับการเข้าใจความ ตีความ และการอนุมานความหมาย (inference) ซึ่งได้จากการสรุปนัยของข้อความ รวมทั้งยังเข้าใจ วลี คำศัพท์ และข้อความต่าง ๆ ที่ใช้ในเอกสารทางวิชาการได้ด้วยผู้มีความรู้ความสามารถระดับนี้มีคะแนนสอบประมาณ 17-33 คะแนน

### 3. ระดับที่ 3: ระดับวิพากษ์วิจารณ์ (Critical Level)

ข้อทดสอบในระดับนี้มี 12 ข้อ คือ 3, 23, 25, 26, 42, 18, 19, 21, 44, 45, 20 และ 41

ผู้สอบในระดับนี้สามารถอ่านเข้าใจข้อความที่เกี่ยวข้องกับการสรุปความหาใจความสำคัญ วิเคราะห์สังเคราะห์ และประเมินผลจากข้อความกึ่งวิชาการและวิชาการขนาด 200-300 คำได้ ผู้มีความสามารถระดับนี้มีคะแนนสอบประมาณ 34-45 คะแนน

ดังนั้น ผู้ที่สอบได้คะแนนระหว่าง 1-16 คะแนน ผู้สร้างแบบทดสอบก็สามารถบอกได้ด้วยความมั่นใจว่าผู้สอบนั้นมีความรู้ความสามารถในระดับที่ 1 และหากว่าได้คะแนน 17-33 คะแนน ก็บอกได้ว่าผู้สอบมีความ

รู้ระดับที่ 1 และ 2 ด้วย เป็นต้น และหากประสงค์จะกำหนดระดับคะแนน (grading) ก็สามารถกำหนดได้โดยพิจารณาว่าผู้ที่สมควรได้ระดับคะแนน A, B, C, และ D ควรมีความรู้ความสามารถในระดับใด แล้วจึงขีดเส้นแบ่งกลุ่มคะแนนสอบเพื่อใช้เป็นคะแนนจุดตัด (cutting score) เช่น หากเห็นว่า ผู้ที่สมควรได้ A ควรมีความรู้ความสามารถระดับที่ 3 (ซึ่งหมายความว่า มีความรู้ความสามารถระดับที่ 1-2 ด้วย) ก็สามารถใช่คะแนน 34 เป็นคะแนนจุดตัดได้ ซึ่งหมายความว่าผู้ที่สมควรได้ A ควรมีคะแนนตั้งแต่ 34 คะแนน เป็นต้น

ด้วยวิธีการใช้การวิเคราะห์ข้อทดสอบตามรูปแบบราชศัพท์กล่าวนี้ ผู้สร้างแบบทดสอบจะสามารถสร้างเกณฑ์ระดับความสามารถเพื่อการวัดผลอิงเกณฑ์ได้อย่างสะดวก และเป็นวิทยาศาสตร์มากยิ่งขึ้น ผู้เขียนจึงหวังว่าท่านผู้อ่านจะได้ทดลองนำแนวคิดนี้ไปใช้ในการปฏิบัติงานจริง หรือในการวิจัยเพื่อพัฒนาแบบทดสอบของท่านต่อไป ส่วนวิธีการใช้โปรแกรม BICAL/PC นั้น โปรดศึกษาจากเอกสารอ้างอิง เพราะผู้เขียนได้เขียนไว้อย่างละเอียดแล้วแต่หากยังมีปัญหาในการใช้ หรือไม่เข้าใจแนวคิดที่เกี่ยวข้อง ก็ขอได้ติดต่อผู้เขียนได้โดยตรง ผู้เขียนยินดีให้คำอธิบายท่านผู้อ่านทุกท่าน

## เอกสารอ้างอิง

- สุพัฒน์ สุขมลสันต์. การวิเคราะห์ข้อทดสอบแนวใหม่โดยใช้เครื่องคอมพิวเตอร์ กรุงเทพฯ: นำอักษรการพิมพ์, 2538
- สุพัฒน์ สุขมลสันต์ และคณะ. “ระดับความสามารถทั่วไปของนิสิต นักศึกษา และบุคคลทั่วไป” รายงานการวิจัย สถาบันภาษา จุฬาลงกรณ์มหาวิทยาลัย, 2534.
- Brown, A. and Others. “Mapping Ability and Skill Levels Using Rasch Techniques” *Occational Papers in Language Testing*, Melbourne University, 1994.
- Davies, A. and Widdowson, H. G. “Reading and Writing” in *Techniques in Applied Linguistics*, London: OUP, 1974
- Hambleton, R.K. and Cook, L.L. “Latent Trait Models and Their Use in the Analysis of Educational Test Data”, *Journal of Educational Measurement*. Vol. 14, No. 2, 1977.
- Timsley, H.A. and Davis, R. “An Investigation of the Rasch Simple Logistic Model : Sample Free Items and Test Calibration”, *Journal of Educational and Psychological Measurement*, Vol. 35, 1975.
- Warm, T.A. *A Primer of Item Response Theory*. Oklahoma : US. Coast Guard Institute, 1978.
- Willmott, A. “What does Rasch Promise?”, *Journal of NFER*, Vol. 22, No. 3, 1980.
- Wright, B.D. “Solving Measurement Problems with the Rasch Model” *Journal of Educational Measurement*, Vol. 14, No. 2 1977.
- Wright, B.D. and Panchapakesan, N. “A Procedure for Sample-free Item Analysis” *Journal of Educational and Psychological Measurement*, Vol. 29, 1969.
- Wright, B.D. and Stone, M.H., *Best Test Design*, Chicago: Mesa Press, 1979.