บทที่ 2

วรรณคดีที่เกี่ยวข้อง

ผู้วิจัยได้ศึกษาและรวบรวมวรรณคดีจากหนังสือ เอกสาร และวารสารต่าง ๆ โดยแบ่งการนำเสนอออกเป็น 3 ตอน คือ

ตอนที่ 1 การทดสอบแบบอิงเกณฑ์ แบ่งเป็น

1. แนวคิดทั่วไปของแบบสอบอิงเกณฑ์
2. ความเที่ยงของแบบสอบอิงเกณฑ์
3. การกำหนดเกณฑ์หรือคะแนนจุดตัด
4. เกณฑ์ในการประเมินวิธีกำหนดคะแนนจุดตัด

ตอนที่ 2 ทฤษฎีการตอบสนองข้อสอบ แบ่งเป็น

1. ข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบ
2. รูปแบบการตอบข้อสอบ
3. การประมาณคะแนนความสามารถ
4. ความสัมพันธ์ของคะแนนความสามารถ (θ) กับคะแนนโดเมน (π)
5. แบบสอบอิงเกณฑ์กับทฤษฎีการตอบสนองข้อสอบ
6. การใช้ทฤษฎีการตอบสนองข้อสอบ กับวิธีกำหนดคะแนนจุดตัดที่ใช้ดุลพินิจของผู้เชี่ยวชาญตัดสิน

ตอนที่ 3 งานวิจัยที่เกี่ยวข้อง

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ตอนที่ 1 การทดสอบแบบอิงเกณฑ์

แนวคิดทั่วไปของแบบสอบอิงเกณฑ์

ผู้ที่ริเริ่มแนวคิดเกี่ยวกับการทดสอบแบบอิงเกณฑ์เป็นคนแรกในปี 1963 คือเกลเซอร์ (Glaser 1963: 519-520) โดยให้ความสามารถของการเรียนรู้ในเรื่องใดเรื่องหนึ่งมีลักษณะเป็นแบบต่อเนื่องอยู่ในช่วงระหว่างไม่มีความรู้เลยกับมีความรู้อย่างสมบูรณ์ และความสามารถของผู้สอบจะตกอยู่ในช่วงของความต่อเนื่องนี้ นำระดับความสามารถของผู้สอบนี้เทียบกับเกณฑ์ที่กำหนดขึ้นเพื่อตัดสินว่าผู้สอบมีความสามารถอยู่ในระดับที่น่าพอใจหรือไม่ ในระยะเวลาใกล้เคียงกันนั้นเมเจอร์ (Mager, cited by Glass 1978 : 240) ก็เผยแพร่ความคิดเกี่ยวกับการเพิ่มเกณฑ์การปฏิบัติไว้ในการเขียนจุดประสงค์การเรียนการสอน แนวคิดของ 2 ท่านนี้เองทำให้เกิดการพัฒนาการทดสอบแบบอิงเกณฑ์ขึ้น ในปี 1969 ปอปแฮมและยูเสค (Popham & Husek 1969 : 2) ได้อธิบายการวัดแบบอิงเกณฑ์ว่า ใช้สำหรับบอกสถานะของผู้สอบโดยเทียบกับเกณฑ์ ต่อมาในปี 1971 เกลเซอร์และนิทโก (Glaser & Nitko 1971:653) ให้ความหมายของแบบสอบอิงเกณฑ์ไว้ว่า เป็นแบบสอบที่สร้างขึ้นเพื่อให้ผลการทดสอบสามารถแปลความหมายได้โดยตรง โดยเปรียบเทียบกับเกณฑ์การปฏิบัติ และ ปอปแฮม (Popham 1978:31) ได้นิยามแบบสอบอิงเกณฑ์ไว้ว่า เป็นแบบสอบที่ใช้บ่งถึงสถานภาพของผู้สอบเป็นรายบุคคล ซึ่งอ้างอิงถึงคะแนนโดเมน ภายใต้ขอบเขตของเนื้อหาที่ได้รับการกำหนดอย่างชัดแจ้ง

การทดสอบแบบอิงเกณฑ์ มักจะเน้นถึงขอบเขตของเนื้อหาที่ต้องกำหนดอย่างชัดแจ้ง การแปลความหมายของคะแนนที่ได้จากแบบสอบอิงเกณฑ์ กระทำโดยเปรียบเทียบระดับความสามารถของแต่ละบุคคลกับเกณฑ์ที่กำหนดขึ้น และมุ่งที่จะประเมินว่านักเรียนสามารถหรือไม่สามารถ รอบรู้หรือไม่รอบรู้ภายใต้ขอบเขตของเนื้อหาที่กำหนด

ลินด์วอลล์และนิตโก (Lindvall & Nitko 1975:76) ระบุถึงลักษณะของแบบสอบอิงเกณฑ์ไว้ 4 ประการ คือ

1. มีการกำหนดพฤติกรรมเฉพาะที่แสดงออกถึงระดับผลสัมฤทธิ์ได้อย่างเด่นชัดก่อนการดำเนินการสร้างแบบสอบ
2. พฤติกรรมจะถูกระบุในรูปของสถานการณ์ของการสอบที่จะทำให้พฤติกรรมทั้งหมดซึ่งต้องการศึกษาสามารถปรากฏออกมาได้
3. วางแผนการสุ่มข้อสอบ เพื่อให้ได้ข้อสอบซึ่งเป็นตัวแทนของข้อสอบที่เป็นไปได้ทั้งหมด

4. คะแนนที่ได้จากการสอบจะสามารถแปลความหมายได้อย่างเป็นปรนัย และมีความหมายในตัวของมันเอง ในการที่จะแสดงถึงความสามารถของผู้สอบแต่ละคนในพฤติกรรมที่มุ่งวัดทั้งหมด

ความเที่ยงของแบบสอบอิงเกณฑ์

แฮมเบิลตัน และคณะ (Hambleton, et al. cited by Berk 1980 : 323-349) ได้แบ่งแนวคิดเกี่ยวกับความเที่ยงของแบบสอบอิงเกณฑ์ออกเป็น 3 แนวคิด ดังนี้

1. ความเที่ยงที่แสดงถึงความเที่ยงของการตัดสินใจในการแบ่งผู้สอบออกเป็น 2 กลุ่ม คือ กลุ่มที่รอบรู้ และกลุ่มที่ไม่รอบรู้ (Reliability of mastery classification decisions) โดยการสอบซ้ำด้วยแบบสอบฉบับเดียวกัน หรือแบบสอบคู่ขนาน วิธีการหาความเที่ยงตามแนวคิดนี้มีหลายวิธี คือ

1.1 วิธีของคาร์เวอร์ (Carver Method)

1.2 วิธีของแฮมเบิลตัน และโนวิก (Hambleton and Novick)

1.3 วิธีของสวามินาธาน แฮมเบิลตัน และอัลจินา (Swaminathan, Hambleton and Algina)

1.4 วิธีของสับโคเวียก (Subkoviak)

1.5 วิธีของฮวน (Huynh)

1.6 วิธีของมาร์แชลและแฮร์เทล (Marshall and Haevtel)

วิธีการหาความเที่ยง 3 วิธีแรก ข้างต้นอาศัยการสอบ 2 ครั้ง แต่วิธีการหาความเที่ยงที่เหลืออาศัยการสอบเพียงครั้งเดียว

2. ความเที่ยงที่แสดงถึงความเที่ยงของคะแนนจากการสอบด้วยแบบสอบอิงเกณฑ์ (Reliability of Criterion-Referenced Test Scores) โดยการหาดัชนีความสอดคล้องของคะแนนของผู้สอบแต่ละคนที่เบี่ยงเบนไปจากคะแนนจุดตัดจากการสอบด้วยแบบสอบคู่ขนาน 2 ฉบับ การหาความเที่ยงตามแนวคิดนี้มีอยู่ 2 วิธี คือ

2.1 วิธีของลิฟวิงสตัน (Livingston)

2.2 วิธีของเบรนนอนและเคน (Brennan and Kane)

3. ความเที่ยงที่แสดงถึง ความเที่ยงของการประมาณค่าคะแนนมวลความรู้ หรือคะแนนจริง (Reliability of domain score estimate) โดยการหาดัชนีความสอดคล้องของคะแนนจริงของผู้สอบแต่ละคนจากการสอบด้วยแบบสอบคู่ขนาน 2 ฉบับ การประมาณค่าความเที่ยงของแบบสอบตามแนวคิดนี้ไม่ต้องใช้จุดตัดเพื่อเป็นเกณฑ์ความรอบรู้ แต่จะพิจารณาความเคลื่อนมาตรฐานของการวัด ซึ่งแบ่งได้เป็น 4 วิธี คือ

3.1 วิธีของเบอร์ก (Berk)

3.2 วิธีของลอร์ด (Lord)

3.3 วิธีของลอร์ดและโนวิค (Lord and Novick)

3.4 วิธีของเบรนนอน (Brennan)

สำหรับการวิจัยครั้งนี้ผู้วิจัยหาความเที่ยงของแบบสอบอิงเกณฑ์ซึ่งแสดงถึงความเที่ยงของการตัดสินใจในการแบ่งผู้สอบออกเป็น 2 กลุ่ม คือ กลุ่มที่รอบรู้และกลุ่มที่ไม่รอบรู้ รายละเอียดของวิธีต่าง ๆ ตามแนวคิดนี้มีดังต่อไปนี้ คือ

<u>วิธีของคาร์เวอร์</u> (Carver, กล่าวในบุญเชิด ภิญโญอนันตพงษ์ 2527 : 165) เป็นวิธีการหาความเที่ยงของการตัดสินจำแนกผู้สอบออกเป็น ผู้รอบรู้ และผู้ไม่รอบรู้จากการสอบนักเรียนกลุ่มเดียวด้วยแบบสอบคู่ขนาน 2 ฉบับ โดยหาสัดส่วนของความสอดคล้องของการตัดสินที่ตรงกันทั้ง 2 ฉบับ ได้ดังนี้

		ฉบับ ก. ไม่รอบรู้	ฉบับ ก. รอบรู้
ฉบับ ข.	รอบรู้	b	a
	ไม่รอบรู้	c	d

$$P_0 = (a + c) / (a + b + c + d)$$

เมื่อ P_0 คือ ความเที่ยงของการตัดสินความรอบรู้

a คือ จำนวนผู้สอบที่รอบรู้ทั้งฉบับ ก. และฉบับ ข.

b คือ จำนวนผู้สอบที่รอบรู้ฉบับ ข. แต่ไม่รอบรู้ฉบับ ก.

c คือ จำนวนผู้สอบที่ไม่รอบรู้ทั้งฉบับ ก. และฉบับ ข.

d คือ จำนวนผู้สอบที่รอบรู้ฉบับ ก. แต่ไม่รอบรู้ฉบับ ข.

วิธีของแฮมเบิลตันและโนวิค (Hambleton and Novick, cited by subkoviak 1976 : 266) เป็นการหาความเที่ยงของการตัดสินจำแนกผู้สอบออกเป็นผู้รอบรู้ระดับต่าง ๆ จากการสอบด้วยแบบสอบคู่ขนาน 2 ฉบับ โดยหาสัดส่วนของความสอดคล้องของการตัดสินที่ตรงกันทั้ง 2 ฉบับ ซึ่งสามารถเขียนเป็นสูตรได้ ดังนี้

$$P_o = \sum_{i=1}^{m} P_{ii}$$

เมื่อ P_o คือ ความเที่ยงของการตัดสินความรอบรู้

P_{ii} คือ สัดส่วนของผู้สอบที่ถูกตัดสินว่าเป็นผู้รอบรู้ในระดับที่ i ตรงกันจากการสอบด้วยแบบสอบคู่ขนาน 2 ฉบับ

วิธีของสวามินาทาน แฮมเบิลตัน และอัลจินา (Swaminathan Hambleton and Algina 1974 : 263-267) วิธีนี้อาศัยแนวคิดของแฮมเบิลตันและโนวิค ดังกล่าวข้างต้น มาปรับแก้โดยหักค่าความสอดคล้องของการตัดสินโดยบังเอิญออกไป ดังนี้

$$K = (P_o - P_c) / (1 - P_c)$$

เมื่อ K คือ ค่าความเที่ยวของการตัดสินซึ่งได้หักความสอดคล้องของการตัดสินโดยบังเอิญออกไปแล้ว

P_o คือ สัดส่วนของความสอดคล้องของการตัดสินความรอบรู้

P_c คือ สัดส่วนของความสอดคล้องของการตัดสินความรอบรู้ที่คาดว่าเกิดขึ้นโดยบังเอิญ

โดยที่ $P_o = \sum_{i=1}^{m} P_{ii}$

$P_c = \sum_{i=1}^{m} P_{i} \cdot P_{.j}$

วิธีของสับโคเวียก (Subkoviak 1976 : 265-276) เป็นวิธีการหาความเที่ยงของการตัดสินจำแนกผู้สอบออกเป็นผู้รอบรู้-ผู้ไม่รอบรู้ จากการสอบเพียงครั้งเดียว โดยอาศัยหลักการหาสัมประสิทธิ์ของความสอดคล้องในการตัดสินความรอบรู้ของผู้สอบแต่ละคน ซึ่งแสดงถึงความน่าจะเป็นที่ผู้สอบแต่ละคนจะถูกตัดสินให้เป็นผู้รอบรู้-ผู้ไม่รอบรู้ ได้ตรงกันจากการสอบด้วยแบบสอบคู่ขนาน 2 ฉบับ

ข้อตกลงเบื้องต้นของวิธีสับโคเวียก คือ

(1) ข้อสอบแต่ละข้อเป็นอิสระกัน

(2) ผลการสอบทั้ง 2 ครั้ง เหมือนกัน

ค่าความเที่ยงของการตัดสินตามวิธีนี้หาได้จากสูตร

$$P_c = \frac{\sum_{i=1}^{N} P_c^{(i)}}{N}$$

เมื่อ P_c คือ ค่าความเที่ยงของการตัดสิน หรือสัมประสิทธิ์ของความสอดคล้องในการตัดสินความรอบรู้ของผู้สอบ N คน

N คือ จำนวนผู้สอบ

โดยที่

$$P_c^{(i)} = [P(x_i \geqslant c)]^2 + [1 - P(x_i \geqslant c)]^2$$

เมื่อ

$$P(x_i \geqslant c) = \sum_{x_i=c}^{n} \binom{n}{x_i} P_i^{x_i} (1 - P_i)^{n-x_i}$$

P_i คือ ความน่าจะเป็นที่จะตอบข้อสอบได้ถูกของผู้สอบคนที่ i ซึ่งหาได้จาก

$$P_i = X_i/n$$

เมื่อ X_i คือ จำนวนข้อที่ตอบถูก

n คือ จำนวนข้อสอบทั้งหมดในแบบสอบ

<u>วิธีของฮวิน</u> (Huynh 1976 : 253-263) เป็นวิธีการประมาณค่าความเที่ยงของการตัดสินจากการสอบเพียงครั้งเดียว โดยพัฒนาสูตรมาจากการสอบด้วยแบบสอบคู่ขนาน 2 ฉบับ ซึ่งข้อสอบจะต้องสุ่มมาจากประชากรของข้อสอบในขอบเขตของเนื้อหาเดียวกัน แบบสอบประกอบด้วยข้อสอบจำนวน n ข้อ คะแนนสอบ X ซึ่งเป็นไปได้ตั้งแต่ 0 ถึง n ของผู้สอบคนหนึ่ง ที่มีระดับความสามารถ θ (ซึ่งแสดงถึงสัดส่วนของจำนวนข้อสอบที่สามารถตอบถูกจากประชากรข้อสอบทั้งหมดในจักรวาล) สมมุติได้ว่ามีการกระจายแบบไบโนเมียล (Binomial Density) ดังนี้

$$f(x/\theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x} \quad ; x = 0,1\ldots,n$$

โดยที่ความสามารถ θ จะมีการกระจายแบบเบต้า (Beta Density) ซึ่งมีพารา-มิเตอร์ α และ β ที่คำนวณได้จากโมเมนท์ 2 ตัวแรก ของการกระจายคะแนนสอบ ถ้า μ และ σ เป็นค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของการกระจายคะแนนสอบ และถ้าความเที่ยงแบบคูเดอร์ริชาร์ดสันสูตร 21 ซึ่งหาได้จากสูตร

$$\alpha_{21} = \frac{n}{n-1} \frac{1- \mu(n-\mu)}{n\sigma^2}$$

แล้วทำให้ $\alpha = (-1+1/\alpha_{21})\mu$

$\beta = -\alpha + n/\alpha_{21} - n$

ทำให้ได้การกระจายของคะแนนสอบ x เป็นแบบ เบต้า-ไนโนเมียล (Beta - Binomial) ดังนี้

$$f(x) = \binom{n}{x} B(\alpha+x, n+\beta-x)/B(\alpha,\beta)$$

เมื่อ B แทนฟังก์ชันของเบต้า

ถ้า x เข้าใกล้ 0 จะหา f(x) ได้จากสูตร

$$f(o) = \prod_{i=1}^{n} \frac{n+\beta-i}{n+\alpha+\beta-i}$$

และ $f(x+1) = f(x) \frac{(n-x)(\alpha+x)}{(x+1)(n+\beta-x-1)}$; $x=0,1,\ldots,n-1$

ถ้า x เข้าใกล้ n จะหา f(x) ได้จากสูตร

$$f(n) = \prod_{i=1}^{n} \frac{n+\alpha-i}{n+\alpha+\beta-i}$$

และ $f(x-1) = f(x). \frac{x(n+\beta-x)}{(n-x+1)(\alpha+x-1)}$; $x = 1,\ldots,n$

15591

ขั้นต่อมาพิจารณาถึงแบบสอบคู่ขนาน 2 ฉบับ ที่เป็นอิสระกัน ซึ่งประกอบด้วยข้อสอบ n ข้อ ที่สุ่มมาจากประชากรข้อสอบในขอบเขตของเนื้อหาที่กำหนด ทำให้ได้คะแนนสอบจากแบบสอบคู่ขนาน 2 ฉบับใด ๆ ที่เป็นอิสระกันด้วย

กำหนดให้ x และ y แทนคะแนนที่ได้จากการสอบแบบสอบฉบับ X และ Y โดยที่ x และ y เป็นอิสระกัน ทำให้ได้การกระจาย ดังนี้

$$f(x,y) = \frac{\binom{n}{x}\binom{n}{y}}{B(\alpha,\beta)} B(\alpha+x+y,\ 2n+\beta-x-y)$$

การกระจายนี้เป็นแบบสมดุลย์ (Symmetric) คือ $f(x,y) = f(y,x)$

ถ้า x และ y เข้าใกล้ 0 จะหา $f(x,y)$ ได้จากสูตร

$$f(0,0) = \prod_{i=1}^{2n} \frac{2n+\beta-i}{2n+\alpha+\beta-i} = f(0). \prod_{i=1}^{n} \frac{2n+\beta-i}{2n+\alpha+\beta-i}$$

และ
$$f(x+1,y) = f(x,y). \frac{(n-x)(\alpha+x+y)}{(x+1)(2n+\beta-x-y-1)}$$

ถ้า x และ y เข้าใกล้ n

$$f(n,n) = \prod_{i=1}^{2n} \frac{2n+\alpha-i}{2n+\alpha+\beta-i} = f(n). \prod_{i=1}^{n} \frac{2n+\alpha-i}{2n+\alpha+\beta-i}$$

และ
$$f(x-1,y) = f(x,y), \frac{x(2n+\beta-x-y)}{(n-x+1)(2+x+y-1)}$$

ในขั้นการตัดสินจำแนกผู้สอบออกเป็นผู้รอบรู้-ผู้ไม่รอบรู้ โดยเทียบคะแนน x กับคะแนนจุดตัด (c) ซึ่งเป็นไปได้ตั้งแต่ 1 ถึง n ผู้สอบที่มีคะแนน $x \geq c$ จะถูกตัดสินให้เป็นผู้รอบรู้ และ $x < c$ จะถูกตัดสินให้เป็นผู้ไม่รอบรู้ จากแบบสอบคู่ขนาน 2 ฉบับ จะจำแนกผู้สอบออกได้ดังแสดงในตารางต่อไปนี้

ฉบับ Y / ฉบับ X	ไม่รอบรู้	รอบรู้	สัดส่วนแยก
ไม่รอบรู้	P_{00}		P_0
รอบรู้		P_{11}	P_1
สัดส่วนแยก	P_0	P_1	

$$P_1 = \sum_{x=c}^{n} f(x)$$

$$P_{11} = \sum_{x,y=c}^{n} f(x,y)$$

$$P_o = \sum_{x=o}^{c-1} f(x)$$

$$P_{oo} = \sum_{x,y=o}^{c-1} f(x,y)$$

$$P = P_{oo} + P_{11}$$

เมื่อ P_1 คือ สัดส่วนของผู้สอบที่ถูกตัดสินว่ารอบรู้ จากการสอบด้วยแบบสอบฉบับเดียว

P_{11} คือ สัดส่วนของผู้สอบที่ถูกตัดสินตรงกันว่ารอบรู้ จากการสอบด้วยแบบสอบคู่ขนาน 2 ฉบับ

P_o คือ สัดส่วนของผู้สอบที่ถูกตัดสินว่าไม่รอบรู้ จากการสอบด้วยแบบสอบฉบับเดียว

P_{oo} คือ สัดส่วนของผู้สอบที่ถูกตัดสินตรงกันว่าไม่รอบรู้ จากการสอบด้วยแบบสอบคู่ขนาน 2 ฉบับ

P คือ ค่าความเที่ยงของการตัดสิน ที่รวมค่าความสอดคล้องโดยบังเอิญของการตัดสิน

การหาค่าความเที่ยงของการตัดสินจำแนกผู้สอบออกเป็น ผู้รอบรู้-ผู้ไม่รอบรู้ ด้วยดัชนีแคปปา (K) ซึ่งหักความสอดคล้องโดยบังเอิญของการตัดสินออก โดยแบ่งออกเป็น 2 กรณี ดังนี้

1. กรณีที่คะแนนจุดตัดมีค่าต่ำ หรือใกล้ 0

$$K = (P_{oo} - P_o^2)/(P_o - P_o^2)$$

2. กรณีที่คะแนนจุดตัดมีค่าสูง หรือใกล้ n

$$K = (P_{11} - P_1^2)/(P_1 - P_1^2)$$

<u>การประมาณค่า K เมื่อข้อสอบมีจำนวนข้อมาก</u>

เมื่อข้อสอบมีจำนวนข้อมาก คือมากกว่า 10 ข้อ จะทำให้การคำนวณค่า K ยุ่งยากมาก ฮวินจึงเสนอวิธีคำนวณใหม่โดยอาศัยแนวคิดของคีตส์และลอร์ด (Keats and Lord) เกี่ยวกับการแปลงคะแนนจากสูตร $x' = \sin^{-1}\sqrt{x/n}$ และมีการกระจายแบบปกติ ซึ่งมีค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐาน ดังนี้

$$\mu_{x'} = \sin^{-1}\sqrt{x/n}$$

$$\sigma_{x'} = [(\alpha_{21} + 1) / (\alpha + n)]^{\frac{1}{2}}$$

ถ้า x และ y เป็นคะแนนแปลงที่ได้จากแบบสอบคู่ขนาน 2 ฉบับ แล้วค่าสหสัมพันธ์ของคะแนนทั้งสองจะมีค่าประมาณ ดังนี้

$$\rho = \alpha_{21} [(n - 1) / (n + \alpha_{21})]^{\frac{1}{2}}$$

การประมาณค่าการแจกแจงปกติจะสามารถประมาณได้ดีเมื่อ μ/n มีค่าอยู่ระหว่าง .15 ถึง .85 และ n มีค่าน้อยที่สุดเท่ากับ 8 ข้อ

เนื่องจากฟังก์ชัน $\sin^{-1}$ จะมีเส้นกราฟสูงขึ้นแบบ monotonic ดังนั้นจึงตัดสินความรอบรู้จาก $x' \geq c'$ เมื่อ $c' = \sin^{-1}\sqrt{(c - .5)/n}$ และใช้ตาราง Univariate and Bivariate Normal Distribution ของกุปตา (Gupta 1963) ช่วย จะสามารถคำนวณค่า P_0 และ P_{00} ได้

ขั้นตอนในการคำนวณค่า K เมื่อข้อสอบมีจำนวนข้อมาก

1. คำนวณค่า $\bar{x}$ และ s^2 ของคะแนนสอบ
2. ประมาณค่า $\mu_{x'}$, $\sigma_{x'}$ และ ρ โดยแทนค่า μ ด้วย X และแทนค่า σ ด้วย s
3. คำนวณค่า $Z = (c' - \mu_{x'})/\sigma_{x'}$
4. ประมาณค่า P_0 จากค่าความน่าจะเป็นของการกระจายแบบปกติที่น้อยกว่า Z
5. ประมาณค่า P_{00} จากค่าความน่าจะเป็นของตัวแปรปกติมาตรฐาน 2 ตัว ที่มีความสัมพันธ์กันเท่ากับ ρ ซึ่งน้อยกว่า Z โดยเปิดตารางของกุปตา (Gupta)
6. คำนวณค่า P และ K จากสูตร

$$P = 1 + 2 (P_{00} - P_0)$$

$$K = (P_{00} - P_0^2) / (P_0 - P_0^2)$$

เป็งและสับโคเวียก (Peng and Subkoviak 1980 : 359-368) ได้เสนอวิธีการประมาณความเที่ยงตามวิธีของฮวินเมื่อข้อสอบมีจำนวนมาก วิธีการคล้ายกับการแปลงคะแนนแบบอาร์คไซน์ที่ฮวินเสนอ โดยให้ชื่อว่าการประมาณค่าแบบปกติอย่างง่าย (Simple Normal Approximation) มีวิธีการดังนี้คือ

1. คำนวณค่าคะแนนเฉลี่ย (μ) ส่วนเบี่ยงเบนมาตรฐาน (σ) และสัมประสิทธิ์ความเที่ยง KR_{21} (α_{21}) ของคะแนนสอบ

2. คำนวณ $Z = (c - .5 - \mu) / \sigma$

เมื่อ c คือ คะแนนจุดตัด
μ คือ ค่าเฉลี่ยของคะแนนสอบใช้ $\bar{X}$ แทน
σ คือ ส่วนเบี่ยงเบนมาตรฐานใช้ s แทน

3. ประมาณค่า P_0 จากค่าความน่าจะเป็นของการกระจายแบบปกติ ที่น้อยกว่า Z

4. ประมาณค่า P_{00} จากค่าความน่าจะเป็นของตัวแปรปกติมาตรฐาน 2 ตัว ที่มีความสัมพันธ์กันเท่ากับ α_{21} ซึ่งน้อยกว่า Z โดยการเปิดตารางของกุปตา

5. คำนวณค่า P และ K จากสูตร

$$P = 1 + 2 (P_{00} - P_0)$$

$$K = (P_{00} - P_0^2) / (P_0 - P_0^2)$$

<u>วิธีของมาร์แชลและแฮร์เทล</u> (Marshall-Haertel, cited by Subkoviak 1980 : 145-147) วิธีนี้จะมีกระบวนการคล้าย ๆ กับวิธีการของฮวิน และวิธีการของสับโคเวียก คือ ใช้ข้อมูลจากการสอบครั้งเดียว และมีข้อตกลงเบื้องต้นว่าการกระจายของคะแนนของผู้สอบคนหนึ่งจะเป็นแบบไบโนเมียล (Binomial)

หาความเที่ยงของการตัดสินจำแนกผู้สอบโดยการแบ่งครึ่งแบบสอบฉบับเดียวให้เป็น 2 ฉบับ ๆ ละ n ข้อ แล้วหาค่า P_0 ซึ่งเป็นสัดส่วนของความสอดคล้องในการจำแนกผู้สอบจากแบบสอบ 2 ฉบับ ดังมีขั้นตอนการหาดังต่อไปนี้

1. กำหนดความถี่ (N_x) ของคะแนน (x) จากการสอบด้วยแบบสอบฉบับ ก. ที่มีจำนวน n ข้อ แล้วหาความถี่ (N_w) ของคะแนน (w) จากการสอบด้วยแบบสอบที่มีความยาวเป็น 2 เท่าของฉบับ ก. จากสูตร

$$N_w = \sum_{x=0}^{n} N_x \binom{2n}{w} \left(\frac{x}{n}\right)^w \left(1-\frac{x}{n}\right)^{2n-w}$$

2. คำนวณค่าความน่าจะเป็นของความสอดคล้องกันในการตัดสินจำแนกผู้สอบจากการสอบด้วยแบบสอบที่แบ่งครึ่ง 2 ฉบับ สำหรับผู้สอบที่มีคะแนน w บนแบบสอบที่ยาวเป็น 2 เท่า คำนวณได้จากสูตร

$$\phi_w(a,b) = \sum_{j=a}^{b} \frac{\binom{w}{j}\binom{2n-w}{n-j}}{\binom{2n}{n}}$$

3. คำนวณค่าความเที่ยงของการตัดสิน (P_o) จากสูตร

$$P_o = \frac{1}{N}\left[\sum_{w=0}^{c-1} N_w + \sum_{w=c}^{2c-2} N_w \cdot \phi_w(w-(c-1),c-1) + \sum_{w=2c}^{n+c-1} N_w \cdot \phi_w(c,w-c) + \sum_{w=n+c}^{2n} N_w\right]$$

เมื่อ N คือ จำนวนผู้สอบ

n คือ จำนวนข้อสอบ

c คือ คะแนนจุดตัดของข้อสอบ n ข้อ

N_w คือ ความถี่ของคะแนน w

$\phi_w(a,b)$ คือ ความน่าจะเป็นของความสอดคล้องกันในการตัดสินจำแนกผู้สอบ

การกำหนดเกณฑ์หรือคะแนนจุดตัด

การกำหนดเกณฑ์ในยุคเริ่มต้น เกิดจากความเคลื่อนไหวของการเขียนวัตถุประสงค์การเรียนการสอนในรูปวัตถุประสงค์เชิงพฤติกรรม

เมเจอร์ (Mager, cited by Glass 1978:238) อธิบาย การกำหนดเกณฑ์ว่าเป็นส่วนหนึ่งของการเขียนวัตถุประสงค์การเรียนการสอน โดยกล่าวว่า

" ถ้าเราสามารถกำหนดผลการปฏิบัติขั้นต่ำสุดที่ยอมรับได้สำหรับวัตถุประสงค์แต่ละข้อ เราจะมีเกณฑ์การปฏิบัติที่ใช้เปรียบเทียบในการทดสอบโปรแกรมการสอนของเรา ; เราจะสามารถบอกได้ว่า โปรแกรมการสอนของเราประสบผลสำเร็จตามวัตถุประสงค์การสอนหรือไม่ สิ่งที่ต้องทำก็คือ การอธิบายว่าผลการปฏิบัติที่ยอมรับได้เป็นอย่างไรในวัตถุประสงค์การเรียนการสอนเพื่อใช้เป็นเกณฑ์ของความสำเร็จ "

และยกตัวอย่างวัตถุประสงค์เชิงพฤติกรรมที่ใช้เป็นเกณฑ์ ดังนี้

" นักเรียนต้องสามารถแก้สมการเส้นตรงได้ถูกต้องอย่างน้อย 7 ข้อ ในเวลา 30 นาที

นักเรียนต้องสามารถสะกดคำได้ถูกต้องอย่างน้อย 80 % ของคำทั้งหมดที่ใช้ในการสอบ "

ปอปแฮม (Popham 1973:3) กล่าวว่า ควรตั้งเกณฑ์การปฏิบัติโดยกำหนดเป็นระดับผลสัมฤทธิ์ขั้นต่ำสุดของนักเรียนในรูปของวัตถุประสงค์การเรียนการสอน เพื่อช่วยให้ครูสามารถวางแผนและประเมินผลการเรียนการสอน และยกตัวอย่างเกณฑ์ในรูปของวัตถุประสงค์การเรียนการสอน ดังนี้

" นักเรียนสามารถแก้โจทย์ปัญหาเกี่ยวกับเส้นรอบรูปได้ถูกต้อง 10 ข้อใน 15 ข้อ "

หอสมุดกลาง สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

นอกจากนั้นไทเลอร์ (Tyler,cited by Glass 1978:238) ก็ได้อธิบายถึงเกณฑ์การปฏิบัติที่ใช้ตัดสินความรอบรู้ในรูปของวัตถุประสงค์เชิงพฤติกรรมเช่นกัน ยกตัวอย่างเช่น " นักเรียนที่รอบรู้ในเนื้อหาใดเนื้อหาหนึ่งนั้น จะต้องสามารถตอบปัญหาที่กำหนดให้ ได้ถูกต้อง 85 % ของทั้งหมด ภายในเวลาที่กำหนด "

การกำหนดเกณฑ์การปฏิบัติในวัตถุประสงค์การเรียนการสอนดังกล่าวมาแล้ว ยังไม่เป็นค่าเชิงปริมาณที่แท้จริง ต่อมาได้มีการพัฒนาเกณฑ์ให้เป็นค่าเชิงปริมาณที่แท้จริงโดยการนิยามเกณฑ์ว่าอาจหมายถึง จำนวน หรือเปอร์เซนต์ของข้อสอบที่ตอบถูกในแบบสอบที่ใช้ทดสอบวัตถุประสงค์การเรียนการสอนหรือในขอบเขตของเนื้อหาหนึ่ง จำนวนของข้อสอบที่ตอบถูก มักหมายถึง คะแนนจุดตัดซึ่งจะแบ่งการกระจายคะแนนของผู้สอบออกเป็น 2 พวก คือ พวกที่รอบรู้และพวกที่ไม่รอบรู้คะแนนจุดตัดนี้หาได้จากการทดสอบนักเรียนด้วยข้อสอบชุดหนึ่งเท่านั้น ส่วนเปอร์เซนต์ของข้อสอบที่ตอบถูกจะเป็นเกณฑ์ที่แท้จริง ของการทดสอบวัตถุประสงค์การเรียนการสอนหรือในขอบเขตของเนื้อหาที่กำหนดโดยพิจารณาจากข้อสอบที่เป็นไปได้ทั้งหมดในวัตถุประสงค์การเรียนการสอน หรือในขอบเขตของเนื้อหาที่กำหนดนั้น (Berk 1986:138)

แต่คะแนนจุดตัดหรือเปอร์เซนต์ของข้อสอบที่ตอบถูก ก็ยังมีปัญหาเกี่ยวกับความไม่เป็นปรนัยในการกำหนดคะแนนจุดตัด จึงมีการนำเอาวิธีทางคณิตศาสตร์มาหาคะแนนจุดตัดและคะแนนของผู้สอบแต่ละคนในรูปคะแนนโดเมน หรือคะแนนจริง (π) เช่น แฮมเบิลตันและโนวิค (Hambleton & Novick 1973:159-170) กำหนดให้ π และ π_0 แทนคะแนนโดเมนและคะแนนจุดตัดตามลำดับ แล้วถ้า $\pi < \pi_0$ จะตัดสินให้เป็นผู้ไม่รอบรู้ และถ้า $\pi \geq \pi_0$ จะตัดสินให้เป็นผู้รอบรู้ นอกจากนี้ ยังมีการกำหนดคะแนนจุดตัดซึ่งพิจารณาจากความคลาดเคลื่อนในการวัด และความคลาดเคลื่อนในการจำแนกประเภทผู้สอบผิด คือ ทำให้เกิดผู้รอบรู้ไม่จริงและผู้ไม่รอบรู้ไม่จริง การคัดเลือกคะแนนจุดตัดที่เหมาะสมวิธีนี้ คือ การหาจุดตัดที่ทำให้เกิดความคลาดเคลื่อนในการจำแนกประเภทผู้สอบผิดที่ต่ำที่สุด

ในการพิจารณาเกณฑ์หรือคะแนนจุดตัดของแบบสอบอิงเกณฑ์นั้นมิลล์แมน (Millman 1973:205-215) ได้สรุปปัจจัยที่ต้องคำนึงถึงไว้ ดังนี้

1. ผลการสอบของผู้สอบคนอื่น ๆ (Performance of others) หรือผลการสอบของกลุ่มที่ผ่านการยอมรับแล้ว

2. เนื้อหาสาระของข้อสอบ (Item Content) โดยให้ผู้เชี่ยวชาญพิจารณาเนื้อหาข้อสอบแล้วกำหนดจำนวนข้อสอบที่น้อยที่สุดที่ผู้สอบต้องตอบได้ถูกต้อง หรือให้ผู้เชี่ยวชาญประมาณความน่าจะเป็นของผู้สอบที่มีความสามารถขั้นต่ำสุดที่ยอมรับได้จะตอบข้อสอบแต่ละข้อได้ถูก คะแนนจุดตัดจะเท่ากับผลบวกของความน่าจะเป็นนั้น

3. ผลทางการศึกษาที่ตามมา (Educational Consequences) คือ ถ้ากำหนดคะแนนจุดตัดต่ำเกินไปก็อาจทำให้นักเรียนไม่สามารถเรียนบทเรียนต่อไปที่ยากกว่าได้ แต่ถ้ากำหนดคะแนนจุดตัดสูงเกินไปก็จะทำให้เสียเวลาทั้งครูและนักเรียน รวมทั้งทรัพยากรต่าง ๆ

4. ผลทางด้านจิตใจและสภาพทางการเงิน (Psychological and Financial Costs) ควรคำนึงถึงผลทางด้านจิตใจที่เกิดขึ้นกับผู้สอบ เช่น แรงจูงใจ ความเบื่อหน่าย การทำลายอัตมโนทัศน์ และรวมทั้งค่าใช้จ่ายในการสอนซ่อมเสริมให้ผู้สอบที่ไม่ผ่านเกณฑ์

5. ความคลาดเคลื่อนที่เกิดจากการเดาและการสุ่มข้อสอบ (Error Due to Guessing and Item Sampling) ควรจะมีการปรับแก้การเดา และข้อสอบที่ใช้ทดสอบมีความเป็นตัวแทนของข้อสอบที่เป็นไปได้ทั้งหมดในโดเมนที่ต้องการวัด

วิธีการกำหนดเกณฑ์ที่มีผู้เสนอขึ้นใช้นั้น แกลส (Glass 1978:237-261) สรุปและรวบรวมแนวคิดของวิธีการต่าง ๆ ออกเป็น 6 แนว ดังนี้

1. การใช้ผลการสอบของคนอื่น ๆ เป็นเกณฑ์ (Performance of Others as a Criterion)

ระดับเกณฑ์สร้างขึ้นโดยอ้างอิงพารามิเตอร์ของประชากรผู้เข้าสอบ ดังนั้นเกณฑ์อาจกำหนดด้วยคะแนนมัธยฐานของคะแนนจากแบบสอบ หรือกำหนดเป็นตำแหน่งเปอร์เซนไตล์ที่เหมาะสมขึ้นมาก่อน แล้วนำไปเทียบกับคะแนนผลการสอบที่สอดคล้องกับตำแหน่งเปอร์เซนไตล์นั้น วิธีนี้นักทฤษฎีเกี่ยวกับแบบสอบอิงเกณฑ์เห็นว่าไม่เหมาะสมกับการวัดแบบอิงเกณฑ์

2. การนับถอยหลังจาก 100% (Counting Backwords from 100%)

จากการสร้างข้อสอบให้สอดคล้องกับจุดประสงค์เชิงพฤติกรรม และจุดประสงค์เชิงพฤติกรรมมีความสำคัญ ดังนั้น เกณฑ์การสอบผ่านที่พึงปรารถนา คือ 100% แต่จากเหตุผลและประสบการณ์พบว่า จะให้ได้ผลโดยสมบูรณ์นั้นย่อมเป็นไปได้ยากเนื่องจากนักเรียนอาจมีความบกพร่องในด้านต่าง ๆ เช่น ความสะเพร่าในการคิดหรือเขียนตอบ จึงยอมให้ลดเกณฑ์ลงมาจาก 100% เหลือ 95% ถึง 80% หรืออาจลดต่ำลงมากกว่านี้ การกำหนดเกณฑ์ด้วยวิธีนี้มีปัญหาว่าเป็นไปตามอำเภอใจ และไม่มีหลักการ

3. การปรับตามเกณฑ์ภายนอกอื่น ๆ (Bootstrapping on other Criterion Scores)

วิธีการนี้อาศัยเกณฑ์ภายนอกมากำหนดว่า " สำเร็จ " หรือ " รอบรู้ " เช่น ใช้เกณฑ์ว่าผู้ได้รับประกาศนียบัตรในอาชีพต่าง ๆ หรือผู้ได้รับใบขับขี่แล้ว เป็นผู้ที่มีความสามารถ ส่วนผู้ที่ไม่ได้รับประกาศนียบัตรนั้น ๆ ก็ถือว่าเป็นผู้ไม่มีความสามารถการหาคะแนนจุดตัดพิจารณาจากการกระจายของคะแนนที่ได้จากการสอบของคนเหล่านี้ แล้วกำหนดคะแนนจุดตัดที่สามารถแบ่งผู้มีความสามารถออกจากผู้ไม่มีความสามารถได้สอดคล้องกับเกณฑ์ภายนอกมากที่สุด วิธีการกำหนดจุดตัดตามแนวคิดนี้ ได้แก่ วิธีของเบอร์ก (Berk)

4. การใช้ดุลพินิจตัดสินความสามารถขั้นต่ำสุด (Judging Minimal Competence)

เป็นวิธีการที่ให้ผู้เชี่ยวชาญเนื้อหาวิชาหรือครูประจำวิชา กำหนดคะแนนจุดตัดโดยศึกษาข้อสอบแล้วระบุว่านักเรียนที่มีความสามารถขั้นต่ำสุดที่ยอมรับได้ ต้องมีคะแนนสอบผ่านเท่ากับเท่าไร การกำหนดคะแนนจุดตัดตามแนวคิดนี้ ได้แก่ วิธีของนีเดลสกี (Nedelsky) วิธีของแองกอฟ (Angoff) และวิธีของอีเบล (Ebel)

5. การใช้วิธีวิจัยเชิงปฏิบัติ (Operations Research Methods)

เป็นวิธีการหาคะแนนจุดตัดที่เหมาะสมจากผลการทดลองที่ตามมาโดยใช้คุณลักษณะที่เกิดจากการสอนอย่างใดอย่างหนึ่ง เช่น ความคงทนของการเรียนรู้ การถ่ายโยงการเรียนรู้ ความสำเร็จในการทำงาน เป็นต้น เป็นเกณฑ์ภายนอก อาศัยโค้งการแจกแจงทางคณิตศาสตร์หรือกราฟ ช่วยในการตัดสินความสัมพันธ์ของคะแนนจุดตัดจากแบบสอบกับเกณฑ์ภายนอกที่กำหนดขึ้น วิธีกำหนดจุดตัดตามแนวคิดนี้ ได้แก่ วิธีของบลอก (Block)

6. การใช้ทฤษฎีการตัดสินใจประกอบ (Decision Theoretic Procedure)

การกำหนดคะแนนจุดตัดตามแนวคิดนี้เน้นการใช้คณิตศาสตร์มาช่วยในการคำนวณและใช้ในการตัดสินใจอย่างจริงจัง ซึ่งมีอยู่หลายวิธี เช่น วิธีของเบส์ (Baysian) พิจารณาคะแนนจุดตัดโดยอาศัยฟังชันการสูญเสียที่คาดหวังของการตัดสินให้ผ่าน และไม่ผ่านคะแนนระดับต่าง ๆ ความสูญเสียที่คาดหวังของการตัดสินให้ผ่านและไม่ผ่านที่น้อยที่สุดจะเป็นตัวกำหนดคะแนนจุดตัด และวิธีของแกลส (Glass 1978:251-253) ซึ่งอาศัยวิธีการทดลองเชิงประจักษ์พิจารณาคะแนนจุดตัดโดยอาศัยฟังชันซึ่งเป็นความสัมพันธ์ของคะแนนจุดตัดกับเกณฑ์ภายนอก ได้แก่ ผู้ที่ได้รับการเรียนการสอนกับผู้ที่ไม่ได้รับการเรียนการสอน คะแนนจุดตัดที่เหมาะสมนั้นจะทำให้เกิดการจำแนกผิดน้อยที่สุด นั่นคือ นักเรียนที่ไม่ได้รับการเรียนการสอนจะถูกตัดสินให้เป็นผู้รอบรู้น้อยที่สุด และนักเรียนที่ได้รับการเรียนการสอนจะถูกตัดสินให้เป็นผู้ไม่รอบรู้มีจำนวนน้อยที่สุดด้วย

แต่แฮมเบิลตันและอิกเนอร์ (Hambleton and Eignor, cited by Hambleton 1980:103-107) รวบรวมวิธีกำหนดเกณฑ์และแบ่งออกเป็น 3 กลุ่มใหญ่ ๆ คือ

1. กลุ่มที่ใช้ดุลพินิจในการตัดสิน (Judgment method) ซึ่งแบ่งการตัดสินได้ 2 แบบ คือ

1.1 การตัดสินตามเนื้อหาข้อสอบ (Item content) ได้แก่ วิธีนีเดลสกี วิธีแองกอฟ และวิธีที่พัฒนาขึ้นจากวิธีแองกอฟ (Modified Angoff)

1.2 การตัดสินโดยการเดา (Guessing)

2. กลุ่มที่ใช้การทดลองสอบ (Empirical model) กลุ่มนี้ยังแบ่งออกเป็น 2 วิธีคือ

2.1 กำหนดเกณฑ์จากผลการสอบ (Data Criterion measure) ได้แก่ วิธีของลิฟวิงสตัน (Livingston) และวิธีของแวนเดอร์ลินเดน และเมลเลนเบอร์ก (VanderLindon & Mellenbergh)

2.2 ใช้ทฤษฎีการตัดสินใจ (Decision Theoretic) เช่น วิธีของไครวอล (Kriewal)

3. กลุ่มที่ใช้ดุลพินิจในการตัดสินและการทดลองสอบผสมผสานกัน (Combination model) ซึ่งแบ่งออกได้เป็น 3 แนวคิด ดังนี้

3.1 การตัดสินโดยอาศัยข้อมูลเชิงประจักษ์ (Judgmental empirical) แนวคิดนี้แบ่งได้ 3 วิธีการ คือ

3.1.1 วิธีการเปรียบเทียบระหว่างกลุ่มผู้สอบ (Contrasting groups) ได้แก่ วิธีของลิฟวิงสตันและชายกี (Livingston & Zieky)

3.1.2 วิธีการสอบกับกลุ่มคาบเส้น (Borderline groups) ได้แก่ วิธีของลิฟวิงสตันและชายกี

3.1.3 วิธีการใช้เกณฑ์ภายนอกประกอบ (Criterion groups) ได้แก่ วิธีของเบอร์ก (Berk)

3.2 การพิจารณาถึงผลการศึกษาที่ตามมา (Education Consequence) ได้แก่ วิธีของบลอก (Block)

3.3 ใช้วิธีการของเบส์ (Bayesian method) ได้แก่ วิธีของแฮมเบิลตันและโนวิค (Hambleton & Novick)

เบอร์ก (Berk 1986:137-172) รวบรวมวิธีการกำหนดเกณฑ์วิธีต่าง ๆ โดยแบ่งออกเป็น 2 รูปแบบ ซึ่งสอดคล้องกับเมสกอสกัส (Meskauskas 1976:133-158) ดังนี้

1. รูปแบบสถานะของตำแหน่ง (State Model) รูปแบบนี้จะตัดสินผู้สอบว่ามีความรอบรู้ / มีความสามารถ หรือยังไม่รอบรู้ / ไม่มีความสามารถเลย เกณฑ์มาตรฐานในอุดมคติ คือ 100% แต่เนื่องจากในทางปฏิบัติอาจมีความคลาดเคลื่อนเกิดขึ้นในการตัดสินว่าผู้สอบมีความรอบรู้ / มีความสามารถ หรือยังไม่รอบรู้ / ไม่มีความสามารถ จึงลดเกณฑ์มาตรฐานลงมาจาก 100% วิธีการเหล่านี้กำหนดเกณฑ์มาตรฐานได้โดย

1.1 การทดลองสอบ (Empirical-judgmental) วิธีการเหล่านี้จะหาเกณฑ์มาตรฐานจากผลการสอบของนักเรียนกลุ่มหนึ่งหรือหลาย ๆ กลุ่ม และใช้วิธีวิเคราะห์ทางสถิติประมาณความคลาดเคลื่อนของการแบ่งผู้สอบออกเป็น ผู้รอบรู้ไม่จริงและผู้ไม่รอบรู้ไม่จริงที่น้อยที่สุด เป็นวิธีการปรับเกณฑ์มาตรฐานที่เหมาะสม (Adjusting standard) วิธีเหล่านี้ ได้แก่

1.1.1 วิธีของเบอร์ก (Berk)

1.1.2 วิธีของเบอร์แกน,แคนเซลลิและลูเทน (Bergan,Cancelli, & Luiten)

1.1.3 วิธีของเอมริค (Emrick)

1.1.4 วิธีของแนป (Knapp)

1.1.5 วิธีของมาเครดีและเดทอน (Maeready & Dayton)

1.1.6 วิธีของโรดาบัส (Roudabush)

1.1.7 วิธีของวิลคอกซ์ (Wilcox)

2. รูปแบบต่อเนื่อง (Continoum Model) รูปแบบนี้เชื่อว่าความสามารถของบุคคลมีการกระจายแบบต่อเนื่อง ซึ่งสามารถกำหนด ความรอบรู้ /ความสามารถ เป็นแบบช่วงได้ โดยแบ่งวิธีการหาเกณฑ์มาตรฐานออกเป็น 3 แนวคิด คือ

2.1 ใช้ดุลพินิจของผู้เชี่ยวชาญตัดสิน (Judgmental) อาจเป็นการตัดสินของผู้เชี่ยวชาญคนเดียวหรือหลายคนก็ได้ ทำการตัดสินอย่างเป็นอิสระกัน หรือมีการประชุมกันก็ได้ โดยไม่มีข้อมูลจากการทดลองสอบนักเรียนมาก่อน ซึ่งมีอยู่ 11 วิธี คือ

2.1.1 วิธี Adjust/modified M.C. Angoff

2.1.2 วิธีแองกอฟ (Angoff)

2.1.3 วิธี Angoff-Nedelsky combination

2.1.4 วิธี Difficulty-importance estimate

2.1.5 วิธี Difficulty-relevance Ebel

2.1.6 วิธี Difficulty-taxonomy Ebel

2.1.7 วิธี Item specifications

2.1.8 วิธี Modified M-C Angoff

2.1.9 วิธีนีเดลสกี (Nedelsky)

2.1.10 วิธี Relevance-taxonomy Ebel

2.1.11 วิธีแองกอฟ 2 ตัวเลือก (Two-choice Angoff)

2.2 ใช้ดุลพินิจของผู้เชี่ยวชาญตัดสิน โดยมีข้อมูลการทดลองสอบเป็นตัวชี้นำในการตัดสิน (Judgmental-empirical) อาจทำการตัดสินอย่างเป็นอิสระกัน หรือมีการประชุมกันก็ได้ วิธีการเหล่านี้ ได้แก่

2.2.1 วิธีหาจุดสมดุลย์ระหว่างเกณฑ์สัมบูรณ์และเกณฑ์สัมพันธ์ แบบที่ 1 (Absolute-relative compromisc I)

2.2.2 วิธีหาจุดสมดุลย์ระหว่างเกณฑ์สัมบูรณ์และเกณฑ์สัมพันธ์ แบบที่ 2 (Absolute-relative compromise II)

2.2.3 วิธี Angoff-contrasting groups plus composite

2.2.4 วิธีการใช้ดุลพินิจตัดสินโดยอาศัยสารสนเทศประกอบ (Informed judgment)

2.2.5 วิธี Iterative Angoff

2.2.6 วิธี Iterative two-choice Angoff

2.2.7 วิธี Modified Angoff-empirical

2.3 ใช้การทดลองสอบ (Empirical-judgmental) ซึ่งสามารถแบ่งออกเป็น 2 แนวทาง คือ

2.3.1 ใช้ข้อมูลการทดลองสอบมาหาเกณฑ์มาตรฐานโดยการวิเคราะห์ทางสถิติ วิธีการเหล่านี้ ได้แก่

2.3.1.1 วิธีสอบกับเด็กคาบเส้น (Borderline group) เช่น วิธีของลิฟวิงสตันและซายกี (Livingston and Zieky)

2.3.1.2 วิธีการเปรียบเทียบระหว่างกลุ่มผู้สอบ (Contrasting groups) เช่น วิธีลิฟวิงสตันและซายกี (Livingston & Zieky)

2.3.1.3 วิธีใช้เกณฑ์ภายนอก (Criterion groups) เช่น วิธีของเบอร์ก (Berk)

2.3.1.4 วิธีพิจารณาผลทางการศึกษาที่ตามมา (Educational consequences) เช่น วิธีของบลอก (Block)

2.3.1.5 วิธีใช้ผลการสอบของคนอื่น ๆ เป็นเกณฑ์ (Norm-referenced criterion)

2.3.2 ใช้ข้อมูลการทดลองสอบประมาณความคลาดเคลื่อนของการแบ่งผู้สอบออกเป็นผู้รอบรู้ไม่จริง หรือผู้ไม่รอบรู้ไม่จริงที่น้อยที่สุด เป็นวิธีการปรับเกณฑ์มาตรฐานที่เหมาะสม วิธีการเหล่านี้ ได้แก่

2.3.2.1 วิธีของฮวน (Hugnh)

2.3.2.2 วิธีของไครวอล (Kriewall)

2.3.2.3 วิธีของลิฟวิงสตัน (Livingston)

2.3.2.4 วิธีของลิฟวิงสตัน 1980 (Livingston)

2.3.2.5 วิธีของโนวิคและลูวิส (Novick & Lewis)

2.3.2.6 วิธีของแวนเดอร์ลินเดนและเมลเลนเบอร์ก (Van der Linden & Mellenbergh)

2.3.2.7 วิธีของวิลคอกซ์ (Wilcox)

เมื่อพิจารณาวิธีการหาคะแนนจุดตัดวิธีต่าง ๆ ที่มีผู้รวบรวมและสรุปดังกล่าวมาแล้วข้างต้น ปรากฏว่ามีการจัดกลุ่มวิธีต่าง ๆ อย่างสอดคล้องกัน ผู้วิจัยจึงรวบรวมวิธีการหาคะแนนจุดตัดวิธีต่าง ๆ และแบ่งออกเป็น 2 รูปแบบ ตามที่เบอร์กสรุปไว้ ดังนี้ คือ

1. รูปแบบสถานะของตำแหน่ง เป็นวิธีการหาคะแนนเกณฑ์โดยการทดลองสอบ

เอมริค (Emrick 1971 : 321-326) เสนอวิธีคำนวณหาเกณฑ์ที่อยู่ในรูปร้อยละ โดยอาศัยการประมาณความผิดพลาดในการยอมรับที่ผิด (β) และความผิดพลาดในการปฏิเสธที่ผิด (α) เมื่อกำหนดให้ $\beta = 3\alpha$ และ RR = 10 จากสูตร

$$K = \frac{\log \frac{\beta}{1-2} + n^{-1} (\log RR)}{\log \frac{\alpha\beta}{(1-\alpha)(1-\beta)}}$$

เมื่อ K คือ เกณฑ์ที่อยู่ในรูปของร้อยละ

n คือ ความยาวของแบบสอบ

β คือ ค่าประมาณความน่าจะเป็นของความผิดพลาดในการยอมรับที่ผิด

α คือ ค่าประมาณความน่าจะเป็นของความผิดพลาดในการปฏิเสธที่ผิด

ค่า α และ β หาได้จากสูตร

$$\alpha + \beta = 1 - \sqrt{r_i}$$

เมื่อ $\bar{r}_i$ คือ ค่าความเที่ยงของข้อสอบ คำนวณได้โดยใช้สูตรของเสเปียร์แมน-บราวน์

RR คือ อัตราส่วนของความรุนแรงของความคลาดเคลื่อน β ต่อ α (Ratio of Regret of β to α)

2. รูปแบบต่อเนื่อง สามารถแบ่งวิธีการกำหนดเกณฑ์ออกเป็น 3 แนวคิด คือ

2.1 วิธีกำหนดจุดตัดที่อาศัยดุลพินิจของผู้เชี่ยวชาญตัดสิน ดังเช่นวิธีต่าง ๆ ต่อไปนี้

<u>วิธีของนีเดลสกี</u> (Nedelsky 1954 : 3-19) เป็นเทคนิคในการหาคะแนนผ่านต่ำสุด (minimum passing score) ดังนี้ ก่อนนำแบบสอบไปให้นักเรียนทำให้ครูผู้สอนทุกคนพิจารณาข้อสอบแต่ละข้อว่าตัวเลือกใดที่นักเรียนในระดับต่ำสุดที่จะผ่าน (พวก D student) บอกได้ว่าผิด ระดับการผ่านต่ำสุดของข้อสอบแต่ละข้อคือส่วนกลับของจำนวนตัวเลือกที่เหลืออยู่ เช่น ในข้อสอบที่มี 5 ตัวเลือก ถ้ามี 1 ตัวเลือกที่นักเรียนระดับต่ำสุดที่จะผ่านรู้ว่าผิด ส่วนกลับของจำนวนตัวเลือกที่เหลือคือ 1/4 นำค่าเศษส่วนนี้จากทุกข้อมารวมกันแล้วหาค่าเฉลี่ยของผลที่ได้จากการพิจารณาของผู้สอนแต่ละคน เป็นค่าเฉลี่ยของคะแนนเดาของนักเรียนที่อยู่ระหว่างเกรด F และ D (F - D student) แทนด้วย M_{FD} เพื่อนำไปหาค่าคะแนนต่ำสุดจากสูตร

$$\text{คะแนนผ่านต่ำสุด} = M_{FD} + k\,\sigma_{FD}$$

k เป็นค่าคงที่ ซึ่งกำหนดขึ้นจากการพิจารณาหลาย ๆ ครั้ง

σ_{FD} เป็นค่าส่วนเบี่ยงเบนมาตรฐานของค่าผลรวมเศษส่วนจากข้อสอบทุกข้อของผู้สอนแต่ละคน

ผู้สอนส่วนใหญ่ที่ใช้เทคนิควิธีนี้เห็นว่าคะแนนผ่านต่ำสุด ควรทำให้นักเรียนที่อยู่ระหว่างเกรด F และ D ส่วนใหญ่สอบตก ถ้าให้ k เป็น -1, 0, 1 และ 2 จะให้พวกที่อยู่ระหว่างเกรด F และ D สอบตก 16%, 50%, 84%, และ 98%, ตามลำดับการพิจารณาตัดสินใจเลือกค่า k ครั้งสุดท้าย จะได้หลังจากที่ผู้สอนหลาย ๆ คนเลือกคำตอบที่นักเรียนระดับต่ำสุดที่จะผ่านรู้ว่าผิดแล้ว แต่ตามการกำหนดค่ามาตรฐานสัมบูรณ์ (absolute standard) นั้น ค่า k ควรกำหนดก่อนที่จะคำนวณค่า M_{FD} และก่อนที่จะรู้คะแนนการสอบของนักเรียน

คำตอบที่นักเรียนต่ำสุดที่จะผ่านรู้ว่าผิด จะมีทั้งคำตอบที่ผิดอย่างชัดเจน และคำตอบผิดปานกลาง ซึ่งผู้ที่จะปฏิเสธได้จะต้องมีความรู้ถึงขั้นนั้นแล้ว ถ้าในแบบสอบมีคำตอบชนิดแรกมาก สามารถแก้ได้โดยการกำหนดค่า k ให้สูง แต่ถ้าชนิดหลังมากจะกำหนดค่า k ให้ต่ำในการเปลี่ยนแปลงค่า k นั้น ไม่ควรเปลี่ยนมากกว่า + 0.5 จาก ค่า k ที่กำหนดขึ้นเพื่อทดลองใช้ การตัดสินตามทฤษฎีไม่ค่อยปรากฏว่า k จะสูงถึง 2 สำหรับแบบสอบส่วนมาก ค่า k = 0 อาจจะต่ำเกินไป ข้อเสนอแนะในการพิจารณาค่า k คือควรเริ่มที่ k = 1 ก่อน และช่วงของค่า k ที่จะเปลี่ยนไปจะอยู่ระหว่าง .5 ถึง 1.5

วิธีของแองกอฟ (Angoff, cited by Glass 1978 : 248) เสนอวิธีการโดยใช้ค่าความน่าจะเป็นที่นักเรียนที่สอบผ่านในระดับต่ำสุด ควรตอบได้อย่างถูกต้องในแบบสอบแบบเลือกตอบมาเป็นคะแนนที่จุดตัดที่ยอมให้ผ่าน (passing score) วิธีการคือ ให้ผู้ตัดสินแต่ละคนเสนอค่าความน่าจะเป็นที่ผู้สอบที่มีความสามารถจะยอมรับได้ระดับต่ำสุด ควรจะตอบแต่ละข้อได้ถูกต้อง ที่จริงแล้วผู้ตัดสินจะคิดถึงจำนวนคนที่จะยอมรับได้ในระดับต่ำสุด (แทนที่จะคิดเฉพาะเพียงคนเดียว) และจะประมาณค่าสัดส่วนของผู้ที่จะยอมรับได้ในระดับต่ำสุด ซึ่งจะตอบแต่ละข้ออย่างถูกต้อง ผลรวมของความน่าจะเป็นหรือสัดส่วนเหล่านี้จะใช้เป็นคะแนนจุดตัดที่ยอมรับขั้นต่ำสุด

วิธี Modified multiple-choice Angoff (ETS, cited by Berk 1986 :48) วิธีนี้จะขอร้องให้ผู้ตัดสินเลือกเปอร์เซนต์ของผู้สอบซึ่งมีความสามารถขั้นต่ำสุดที่จะตอบข้อสอบได้ถูก โดยมีให้เลือก 7 ค่า คือ 5%, 20%, 40%, 60%. 75%. 90% และ 95% หรือ "ไม่ทราบ"

วิธี Two - choice Angoff (Nassif, cited by berk 1986 : 148) วิธีนี้จะถามผู้ตัดสินว่าผู้สอบที่มีความสามารถขั้นต่ำสุดสามารถตอบข้อสอบได้ถูกหรือไม่ คำตอบของผู้ตัดสินจะเป็น "ได้" "ไม่ได้" และ "ไม่ทราบ" เกณฑ์หาได้จากคำตอบของผู้ตัดสินที่ตอบว่า "ได้" ตรงกัน เปอร์เซนต์ของข้อสอบที่ผู้ตัดสินตอบว่า "ได้" คือ เกณฑ์ที่ต้องการ

วิธี Adjusted/modified multiple-choice Angoff (Knowledge estimation panel) (Bernknopf, Curry, and Bashaw, cited by berk 1986 : 151) วิธีนี้จะขอร้องให้ผู้ตัดสินเลือก "เปอร์เซนต์ของผู้สอบซึ่งมีความสามารถขั้นต่ำสุดที่จะตอบข้อสอบได้ถูก" สำหรับข้อสอบแต่ละข้อ จากตัวเลือก 9 ตัว แล้วปรับเปอร์เซนต์ที่ทำให้การจำแนกผู้สอบผิด คือ รอบรู้ไม่จริง หรือไม่รอบรู้ไม่จริง มีค่าต่ำสุด และปรับแก้ ความคลาดเคลื่อนที่เกิดจากการเดาเกณฑ์หาได้จากผลบวกของค่าเฉลี่ยของเปอร์เซนต์ในการตอบข้อสอบได้ถูกจากผู้ตัดสินทุกคน

วิธี Angoff-Nedelsky combination (Reid, cited by berk 1986 : 152) เกณฑ์คำนวณจากค่าเฉลี่ยของเกณฑ์ที่ได้จากวิธีแองกอฟและวิธีนีเดลสกี

วิธีของอีเบล (Ebel, cited by Glass 1978 : 247-248) อีเบลมีความเห็นว่า นิยามของคะแนนจุดตัด (passing score) ในลักษณะที่เป็นเปอร์เซนต์ของคะแนนสอบทั้งหมดอาจมีข้อบกพร่อง เนื่องจากข้อสอบอาจยากเกินไป ง่ายเกินไปหรืออำนาจจำแนกน้อยกว่าที่ผู้สร้างแบบสอบตั้งใจไว้ การสอบผ่านหรือตกของผู้เข้าสอบอาจเป็นการตัดสินโดยใช้คำถามในแบบสอบมากกว่าใช้ระดับความสามารถของเขาอีเบลจึงเสนอให้ใช้การพิจารณาคุณลักษณะของข้อสอบในแง่ของความเกี่ยวข้อง (relerance) และความยาก (difficulty) ในแง่ของความเกี่ยวข้อง แบ่งเป็น 4 ประเภท คือ จำเป็น (Essential) สำคัญ (Important) พอจะยอมรับได้ (Acceptable) และยังไม่แน่ใจ (Questionable) ส่วนค่าความยากมี 3 ระดับ คือ ง่าย ปานกลาง และยาก และเสนอเปอร์เซนต์ ที่คาดหวังซึ่งเป็นปริมาณของผู้สอบซึ่งมีความสามารถขั้นต่ำสุดที่ยอมรับได้น่าจะสอบผ่าน

ระดับ ความเกี่ยวข้อง / ความยาก	ง่าย	ปานกลาง	ยาก
จำเป็น			
สำคัญ			
ยอมรับได้			
ไม่แน่ใจ			

วิธี difficulty - taxonomy Ebel. (Skakun & Kling 1980 : 229-235) วิธีนี้อาศัยการพิจารณาข้อสอบใน 2 ประเด็น ดังนี้

1. ความยากของข้อสอบ คือ ง่าย ปานกลาง ยาก
2. ลำดับขั้นการเรียนรู้ของบลูม (Bloom Taxonomy) คือขั้นความจำ ความเข้าใจ และการนำไปใช้

ลำดับขั้นการเรียนรู้ ระดับความยาก	ง่าย	ปานกลาง	ยาก
ขั้นความจำ			
ขั้นความเข้าใจ			
ขั้นการนำไปใช้			

สามารถกำหนดความยากของข้อสอบโดยใช้ผลการทดสอบนักเรียน และให้ผู้สร้างข้อสอบเป็นผู้กำหนดลำดับขั้นการเรียนรู้ของข้อสอบ แล้วจัดแบ่งข้อสอบลงในตารางข้างต้นตามความยาก และลำดับขั้นการเรียนรู้ คะแนนเกณฑ์ หาได้โดยการตัดสินสัดส่วนของข้อสอบที่ผู้สอบซึ่งมีความสามารถขั้นต่ำสุด สามารถตอบได้ถูกต้องในแต่ละช่องของตารางข้างต้น

วิธี Relevance - taxonomy Ebel (Skakun &Kling 1980 : 229-235) วิธีนี้เหมือนกับวิธี Difficulty - Taxonomy Ebel ต่างกันที่การพิจารณาข้อสอบในด้านลำดับขั้นการเรียนรู้เปลี่ยนเป็น พิจารณาข้อสอบในด้านลักษณะข้อสอบที่ตรงกับประเด็นปัญหา จำแนกเป็น 3 ระดับ คือ 1. ระดับที่ตรงกับปัญหาและจำเป็นมาก 2. ระดับที่ตรงกับปัญหาและสำคัญ 3. ระดับที่ยอมรับว่าตรงกับปัญหา

วิธีพิจารณาลักษณะเฉพาะข้อสอบ (Item specifications) (Mills & Barr, cited by berk 1986 : 149) ให้ผู้ตัดสินอ่านลักษณะเฉพาะของข้อสอบแต่ละลักษณะ ซึ่งประกอบด้วยพฤติกรรมที่ต้องการวัด รายละเอียดของเนื้อหา และเงื่อนไขที่จะทำให้เกิดพฤติกรรมนั้น หลักในการสร้างตัวเลือก และตัวอย่างข้อสอบ

หลังจากการอ่านลักษณะเฉพาะของข้อสอบแล้ว ผู้ตัดสินจะมีมโนทัศน์เกี่ยวกับข้อสอบทั้งหมดที่เป็นไปได้ ซึ่งแทนได้ด้วยลักษณะเฉพาะของข้อสอบ ดังนั้นจึงสามารถกำหนดเกณฑ์ได้โดยการประมาณสัดส่วนของข้อสอบในจักรวาลที่ผู้สอบ ซึ่งมีความสามารถขั้นต่ำสุดจะสามารถตอบได้ถูก

วิธีการประมาณตามความสำคัญของความยาก (Difficulty - importance estimate) (Cangelosi, cited by berk 1986 : 149) ทำตามลำดับขั้นได้ดังนี้

1. ให้ผู้ตัดสินพิจารณาวัตถุประสงค์การเรียนแต่ละข้อ แล้วประมาณระดับผ่านต่ำสุดและกำหนดน้ำหนักความสำคัญของวัตถุประสงค์การเรียนแต่ละข้อ

2. หาค่าเฉลี่ยของผลคูณระหว่างระดับผ่านต่ำสุด กับน้ำหนักความสำคัญของวัตถุประสงค์การเรียนแต่ละข้อ

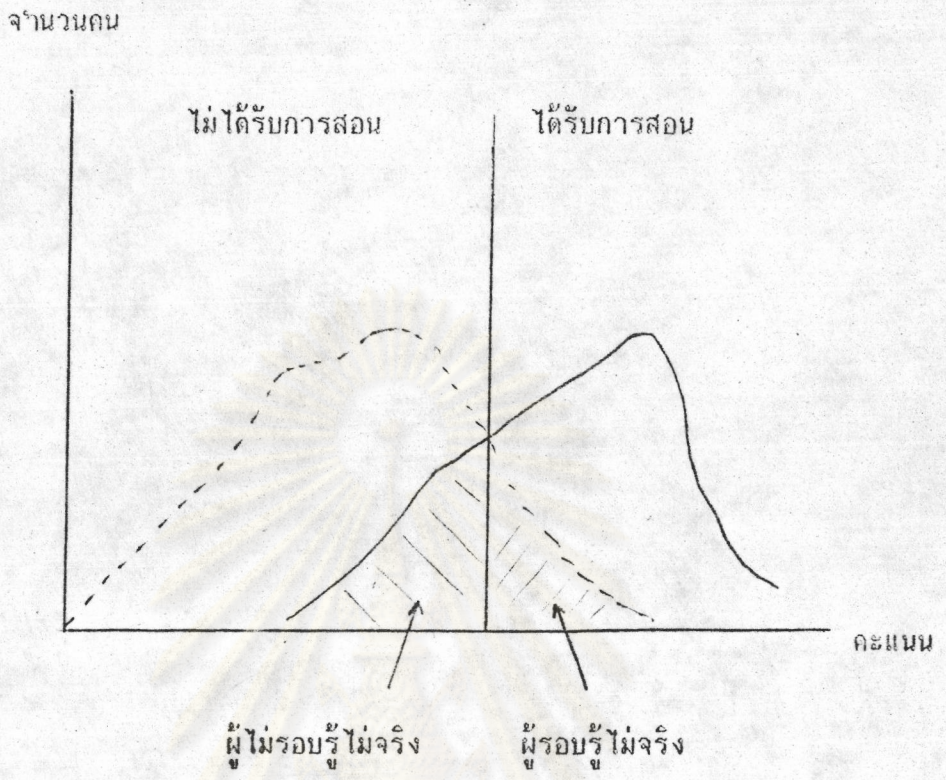
3. คำนวณคะแนนเกณฑ์โดยการหาผลคูณระหว่างค่าเฉลี่ยที่ได้จากขั้นที่ 2 กับจำนวนข้อในแบบสอบ

2.2 วิธีกำหนดจุดตัดที่อาศัยการทดลองสอบดังเช่นวิธีต่าง ๆ ต่อไปนี้

วิธีใช้กลุ่มคาบเส้น (Borderline group) (Livingston & Zieky, cited by berk 1986 : 161) วิธีการคือเลือกนักเรียนที่มีความสามารถอยู่ในระดับคาบเส้นมาทดสอบ แล้วคะแนนจุดตัดเท่ากับ ค่ามัธยฐานของคะแนนสอบของนักเรียนกลุ่มนี้

วิธีใช้การเปรียบเทียบระหว่างกลุ่มผู้สอบ (Contrasting groups) (Livingston & Zieky, cited by berk 1986 : 162) ทำการจัดประเภทผู้สอบออกเป็น "พวกรอบรู้" และ "พวกไม่รอบรู้" โดยการทดสอบ การคัดเลือกเกณฑ์จะคำนึงถึงความรู้หรือความชำนาญอย่างหนึ่งอย่างใดของผู้สอบที่นอกเหนือจากคะแนนสอบด้วย วิธีการหาคะแนนจุดตัดกำหนดจากความคลาดเคลื่อนในการจำแนกประเภทผู้สอบผิดที่มีค่าต่ำสุด

วิธีอาศัยเกณฑ์ภายนอก (Criterion groups) เบอร์ค (Berk 1978 : 4-9) ประยุกต์วิธีนี้มาใช้โดยใช้เกณฑ์ภายนอก คือ ได้รับการสอนหรือไม่ได้รับการสอนตามวัตถุประสงค์ที่ต้องการวัดมาแบ่งนักเรียนออกเป็น 2 กลุ่ม คือ กลุ่มที่ได้รับการสอนให้เป็นพวกรอบรู้ (Master : M) กลุ่มที่ไม่ได้รับการสอนให้เป็นพวกไม่รอบรู้ (Nonmaster: N) แล้วทำการสอบกับนักเรียนทั้ง 2 กลุ่ม นี้ด้วยแบบสอบชุดเดียวกัน นำการแจกแจงของคะแนนการสอบของนักเรียนทั้ง 2 กลุ่ม ไปเขียนกราฟร่วมกัน ดังนี้



จุดที่กราฟ หรือฟังก์ชันทั้งสองตัดกันจะกำหนดให้เป็นคะแนนจุดตัดนักเรียนที่ได้คะแนนเท่ากับหรือสูงกว่าคะแนนจุดตัด จัดเป็นพวกรอบรู้โดยการทำนาย (Predicted Master : PM) ส่วนนักเรียนที่ได้คะแนนต่ำกว่าคะแนนจุดตัด จัดเป็นพวกไม่รอบรู้โดยการทำนาย (Predicted Nonmaster : PN) จากการแบ่งโดยใช้เกฑฑ์ภายนอก และคะแนนจุดตัดที่กำหนดขึ้น จะได้นักเรียน 4 พวก คือ

1. พวกรอบรู้จริง (True Master : TM) คือนักเรียนที่ได้รับการสอนซึ่งได้คะแนนเท่ากับ หรือมากกว่าคะแนนจุดตัด
2. พวกรอบรู้ไม่จริง (False Master : FM) คือนักเรียนที่ไม่ได้รับการสอน ซึ่งได้คะแนนมากกว่าหรือเท่ากับคะแนนจุดตัด
3. พวกไม่รอบรู้จริง (True Nonmaster : TN) คือนักเรียนที่ไม่ได้รับการสอน ซึ่งได้คะแนนต่ำกว่าคะแนนจุดตัด
4. พวกไม่รอบรู้ไม่จริง (False Nonmaster : FN) คือนักเรียนที่ได้รับการสอน ซึ่งได้คะแนนต่ำกว่าคะแนนจุดตัด

แสดงในรูปตาราง 2 x 2 ได้ดังนี้

	เกณฑ์ภายนอก	
	ได้รับการสอน	ไม่ได้รับการสอน
รอบรู้โดยการทำนาย (คะแนนจุดตัด)	รอบรู้จริง (TM)	รอบรู้ไม่จริง (FM)
ไม่รอบรู้โดยการทำนาย	ไม่รอบรู้ไม่จริง (FN)	ไม่รอบรู้จริง (TN)

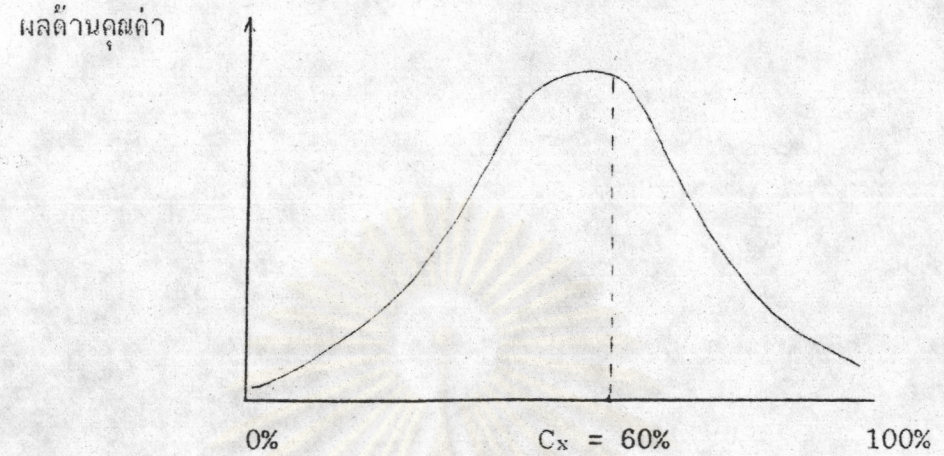
เมื่อทดลองกำหนดคะแนนจุดตัดหลาย ๆ จุด จะได้จำนวนนักเรียนแต่ละพวกแตกต่างกันตามค่าของคะแนนจุดตัด นำผลที่ได้จากการกำหนดคะแนนจุดตัดต่าง ๆ กันมาหาค่าความน่าจะเป็นของแต่ละประเภท โดยใช้สูตร ดังนี้

$$P(TM) = TM/(M + N)$$
$$P(TN) = TN/(M + N)$$
$$P(FM) = FM/(M + N)$$
$$P(FN) = FN/(M + N)$$

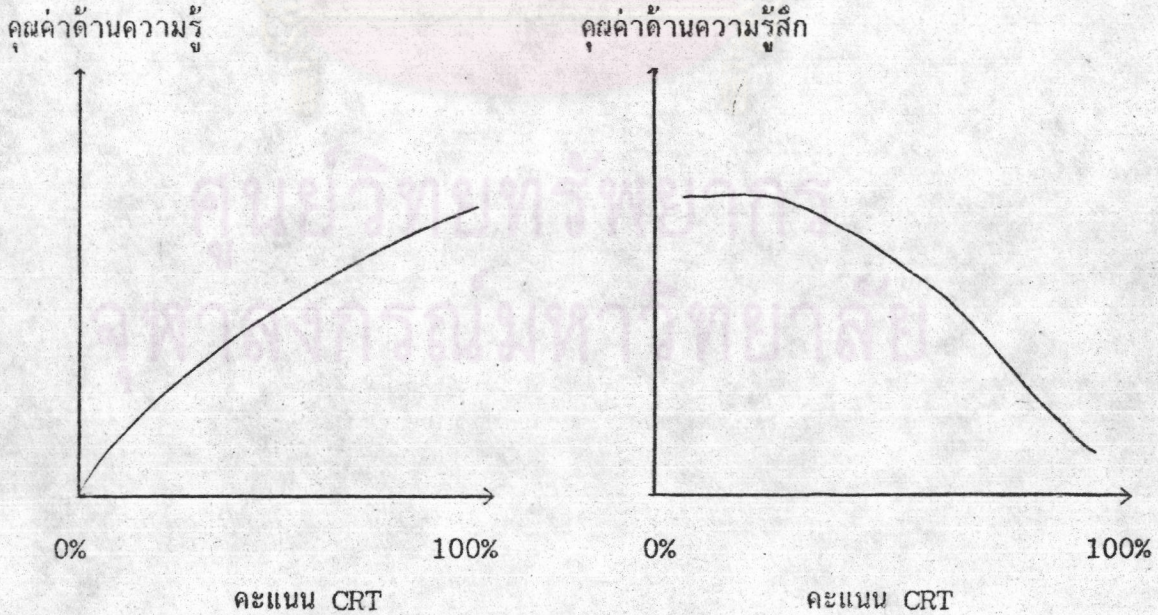
คะแนนจุดตัดที่เหมาะสม คือ จุดตัดที่ทำให้ความน่าจะเป็นในการตัดสินประเภทบุคคลได้ถูกต้อง คือ P (TM) + P (FN) มีค่าสูงสุด หรือทำให้ความน่าจะเป็นในการตัดสินประเภทบุคคลผิด คือ P (FM) + P (FN) มีค่าต่ำสุด

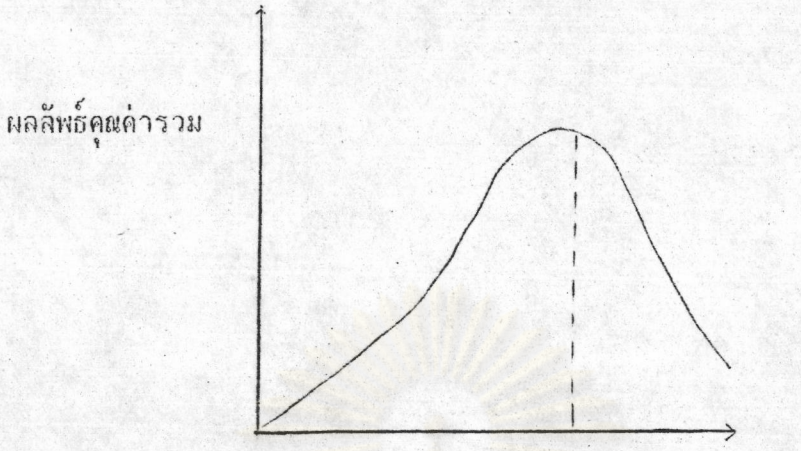
<u>วิธีพิจารณาผลทางการศึกษาที่ตามมา</u>(Educational Concequences) วิธีนี้อาศัยผลด้านคุณค่าอย่างใดอย่างหนึ่งมาช่วยในการพิจารณาเกณฑ์ โดยใช้ผลคุณค่าสูงสุดมาเป็นเครื่องชี้กำหนดคะแนนจุดตัด ซึ่งพิจารณาจากลักษณะกราฟคะแนนที่ตรงกับจุดสูงสุดของกราฟจะถือเป็นคะแนนจุดตัด บลอค (Block 1972 : 183-190) ประยุกต์วิธีการนี้มาใช้ โดยทำการสอนนักเรียนกลุ่มต่าง ๆ ที่มีความเท่าเทียมกัน จนมีผลสัมฤทธิ์จากการทำแบบสอบอิงเกณฑ์ระดับต่าง ๆกัน เช่น 10%, 15%, 20%, ..., 95%, 100% แล้ววัดผลด้านคุณค่าอย่างใดอย่างหนึ่งที่มีความสัมพันธ์กับคะแนนจากแบบสอบอิงเกณฑ์ของแต่ละกลุ่มที่มีผลสัมฤทธิ์ระดับต่าง ๆ กันนั้น นำผลที่ได้มาเขียนกราฟแสดงความสัมพันธ์เพื่อพิจารณาคะแนนจุดตัด ถ้ากราฟมีจุดโค้งกลับ (Non - monotonic) คะแนนจุดตัดจะเป็นไปได้ตั้งแต่ 0% - 100% ดังแสดงในรูปกราฟต่อไปนี้

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย



ถ้ากราฟมีลักษณะขึ้นทางเดียว (monotonic) คือไม่มีจุดโค้งกลับ คะแนนเกณฑ์จะเป็น 100% ถ้าเป็นเช่นนี้ ควรวัดผลคุณค่าด้านอื่น ๆ ที่สัมพันธ์กับคะแนนจากแบบสอบอิงเกณฑ์มาประกอบการพิจารณา เช่น วัดความสนใจ หรือทัศนคติต่อวิชานั้น ๆ เมื่อรวมผลด้านคุณค่าทั้ง 2 ด้าน เข้ากันแล้ว ใช้ระดับคะแนนที่ให้ผลลัพธ์ทางคุณค่ารวมสูงสุดเป็นคะแนนจุดตัด





คะแนน CRT

การพิจารณาคุณค่ารวม ควรมีการพิจารณากำหนดน้ำหนักของค่าการจำแนกผิดทางลบ (α) และค่าการจำแนกผิดทางบวก (β) แล้วแทนในสมการ

ผลลัพธ์คุณค่ารวม = α (คุณค่าด้านความรู้) + β (คุณค่าด้านความรู้สึก)

<u>วิธีการของแกลส</u> (Decision Theoretic Approach) (Glass 1978 : 251-253) วิธีนี้อาศัยการทดลองสอบ และเกณฑ์ภายนอกได้แก่ นักเรียนที่ไม่ได้รับการเรียนการสอน กับนักเรียนที่ได้รับการเรียนการสอน โดยมีการสมมุติคะแนนจุดตัด C_x เพื่อจำแนกผู้สอบเป็นผ่านหรือไม่ผ่าน

คะแนนจุดตัดที่เหมาะสมจะทำให้เกิดการจำแนกผิดน้อยที่สุด นั่นคือ นักเรียนที่ไม่ได้รับการเรียนการสอนจะถูกตัดสินให้ผ่านมีจำนวนน้อยที่สุด และนักเรียนที่ได้รับการเรียนการสอนจะถูกตัดสินให้ตกหรือไม่ผ่าน มีจำนวนน้อยที่สุดด้วย คะแนนจุดตัดสามารถหาได้จากฟังก์ชันความสัมพันธ์ของคะแนนจุดตัดกับเกณฑ์ภายนอกที่มีค่าต่ำสุด ดังนี้

$$f(C_x) = (P_A + P_D) / (P_B + P_C)$$

เมื่อ P_A แทนสัดส่วนจำนวนคนที่ได้รับการเรียนการสอน แต่ถูกตัดสินให้ไม่ผ่าน

P_D แทนสัดส่วนจำนวนคนที่ไม่ได้รับการเรียนการสอนแต่ถูกตัดสินให้ผ่าน

P_B แทนสัดส่วนจำนวนคนที่ไม่ได้รับการเรียนการสอนและไม่ผ่าน

P_C แทนสัดส่วนจำนวนคนที่ได้รับการเรียนการสอนและผ่าน

		เกณฑ์ภายนอก	
		ได้รับการเรียนการสอน	ไม่ได้รับการเรียนการสอน
Cx	ไม่ผ่าน	P_A	P_B
	ผ่าน	P_C	P_D

วิธีการของเบส์ (Bayesian Decision Theoretic Procedure) (Swaminathan, Hambleton & Algina 1975 : 87-98) วิธีนี้จะประมาณคะแนนโดเมนของผู้สอบจากคะแนนสอบ แล้วพิจารณาคะแนนจุดตัดโดยอาศัยฟังก์ชันการสูญเสียที่คาดหวังของการตัดสินให้ผ่านและไม่ผ่านที่ระดับคะแนนต่าง ๆ ความสูญเสียที่คาดหวังที่น้อยที่สุดจะเป็นตัวกำหนดคะแนนจุดตัด (π_0) โดยที่

1. ถ้าผู้สอบไม่รอบรู้แต่ตัดสินใจให้ผ่าน ความสูญเสียที่คาดหวัง

$$= l_{12} \text{ Prob } (\pi \leqslant \pi_0)$$

2. ถ้าผู้สอบเป็นผู้รอบรู้แต่ตัดสินใจให้ตก ความสูญเสียที่คาดหวัง

$$= l_{21} \text{ Prob } (\pi \geqslant \pi_0)$$

แต่ในสภาพความเป็นจริงไม่มีโอกาสที่จะทราบได้ว่าการตัดสินใจถูกหรือผิด เพราะเราไม่สามารถหาค่า ที่แท้จริงของผู้สอบแต่ละคน จึงถือเอาว่า

1. ให้ผู้สอบสอบผ่าน และถือว่าเป็นผู้รอบรู้ ถ้า

$$= l_{21} \text{ Prob } (\pi \leqslant \pi_0) \quad < l_{12} \text{ Prob } (\pi \geqslant \pi_0)$$

2. ให้ผู้สอบสอบตก และถือว่าเป็นผู้ไม่รอบรู้ ถ้า

$$= l_{21} \text{ Prob } (\pi > \pi_0) \quad < \quad l_{12} \text{ Prob } (\pi \leqslant \pi_0)$$

2.3 วิธีการกำหนดจุดตัด โดยอาศัยดุลพินิจของผู้เชี่ยวชาญตัดสินกับการทดลองสอบ ผสมผสานกัน ดังเช่นวิธีต่อไปนี้

วิธีหาจุดสมดุลย์ระหว่างเกณฑ์สัมบูรณ์และเกณฑ์สัมพันธ์แบบที่ 1

(Absolute - relative compromise I) (Beuk 1984 : 147-152) วิธีการนี้ผู้เชี่ยวชาญต้องเสนอ เกณฑ์ 2 ประเภท คือ

1. เกณฑ์สัมบูรณ์ (k) อยู่ในรูปของเปอร์เซนต์ของข้อสอบที่น้อยที่สุด ที่ผู้สอบควรตอบได้ถูก
2. เกณฑ์สัมพันธ์ (v) อยู่ในรูปของเปอร์เซนต์ของผู้สอบที่คาดว่าจะสอบผ่าน

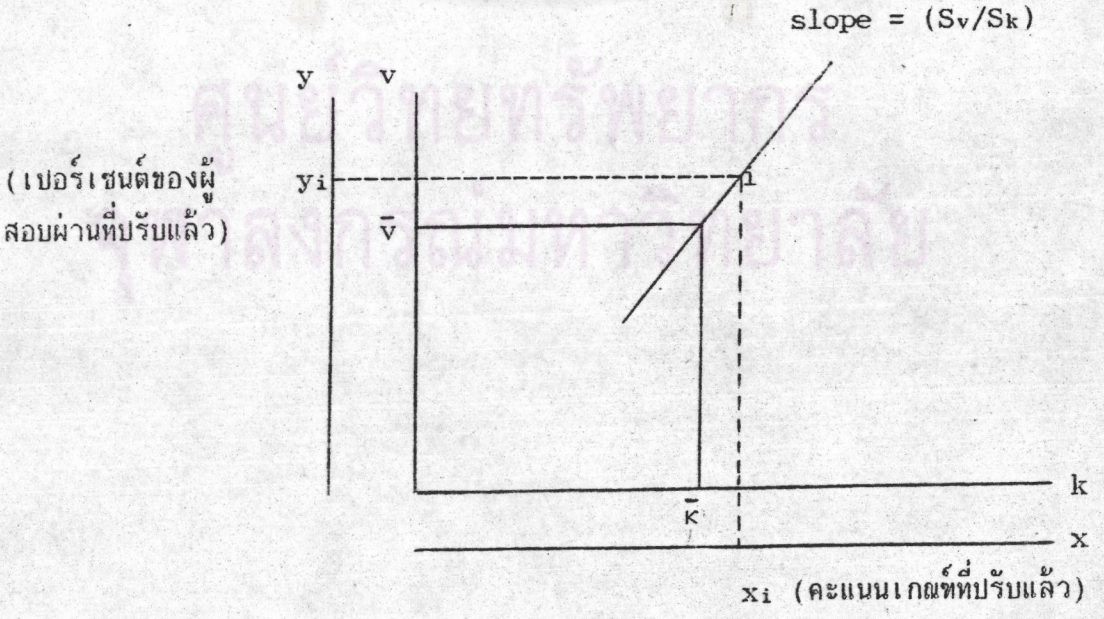
ความสัมพันธ์ระหว่าง v และ k เขียนได้ดังนี้

$$v = (Sv/Sk)(k - \bar{k}) + \bar{v}$$

การปรับเกณฑ์ขึ้นอยู่กับอัตราส่วนของค่าเบี่ยงเบนมาตรฐานของเกณฑ์สัมพันธ์ และเกณฑ์สัมบูรณ์ (Sv/Sk)

นำแบบสอบไปทดสอบกับนักเรียน เพื่อหาความสัมพันธ์ระหว่างคะแนนสอบ (x) กับเปอร์เซนต์ (y) ของผู้สอบที่จะเป็นผู้สอบได้ ถ้า x คือ คะแนนเกณฑ์ ซึ่งจะทำให้ y เป็นฟังก์ชันที่ลดลงของ x

พล็อตกราฟของความสัมพันธ์ระหว่าง v กับ k และความสัมพันธ์ระหว่าง x กับ y เพื่อหาจุดสมดุล (i) ระหว่างเกณฑ์สัมบูรณ์และเกณฑ์สัมพันธ์



วิธีหาจุดสมดุลย์ระหว่างเกณฑ์สัมบูรณ์และเกณฑ์สัมพันธ์แบบที่ 2 (Absolute - relative compromise II) (Hofstee cited by Gruijter 1985 : 263-269.) ผู้ตัดสินต้องกำหนดค่า 4 ค่า คือ

1. เปอร์เซนต์ของข้อสอบที่มากที่สุดที่ผู้สอบควรตอบได้ถูก (K_{max})
2. เปอร์เซนต์ของข้อสอบที่น้อยที่สุดที่ผู้สอบควรตอบได้ถูก (K_{min})
3. เปอร์เซนต์ของผู้สอบตกที่มากที่สุดที่ยอมรับได้ (fmax)
4. เปอร์เซนต์ของผู้สอบตกที่น้อยที่สุดที่ยอมรับได้ (fmin)

ใช้โค้งการแจกแจงสะสมของคะแนนสอบ และความสัมพันธ์ระหว่าง k และ f เพื่อกำหนดจุดสมดุลย์

คะแนนเกณฑ์ คือ จุดที่ตัดกันระหว่างค่าที่กำหนดข้างต้น กับโค้งการแจกแจงสะสมของคะแนนสอบ

วิธี Angoff - contrasting groups plus composite (Shepard, cited by Berk 1986 : 159) วิธีนี้ใช้การพิจารณา 4 อย่าง คือ

1. ใช้วิธีหาคะแนนจุดตัดโดยการตัดสินของผู้เชี่ยวชาญ เช่นวิธีของแองกอฟ
2. ใช้วิธีหาคะแนนจุดตัดโดยการทดสอบ เช่นวิธีการเปรียบเทียบระหว่างกลุ่มผู้สอบ
3. ทำการตัดสินประมาณอัตราการสอบผ่านที่ยอมรับได้
4. ปรับแก้ความไม่เที่ยงของคะแนนจุดตัด

คะแนนเกณฑ์ จะได้จากการปรับความแตกต่างที่เกิดขึ้นจากทั้ง 4 ข้อ เข้าหากัน

วิธีการใช้ดุลพินิจตัดสินโดยอาศัยสารสนเทศประกอบ (Informed judgment) (Popham & Yalow, cited by Berk 1986 : 156) วิธีนี้ผู้ตัดสินกำหนดเกณฑ์โดยอาศัยสารสนเทศ 2 อย่าง คือ

1. ผลการทดสอบของนักเรียน เช่น ค่าความยาก ค่าอำนาจจำแนกของข้อสอบและค่าเฉลี่ยของคะแนนสอบ
2. ให้ผู้ที่เกี่ยวข้อง เช่น ครู ผู้บริหารโรงเรียน ศาสตราจารย์ในวิทยาลัย และคณะกรรมการของโรงเรียน เป็นผู้เสนอเกณฑ์ของแบบสอบ

ผู้ตัดสินจะอาศัยสารสนเทศข้างต้น แล้วตัดสินเกณฑ์ของแบบสอบคะแนนเกณฑ์ของแบบสอบได้มาจาก การตัดสินที่สอดคล้องกันของกลุ่มผู้ตัดสินนี้

วิธี Iterative Angoff (Saunders & Mappus, cited by berk 1986 : 157) ในรอบแรกจะให้ผู้ตัดสินประมาณความน่าจะเป็นในการตอบข้อสอบถูก แล้วอนุญาตให้ผู้ตัดสินเปลี่ยนแปลงการตัดสินได้ 3 ครั้ง คือ

1. ภายหลังการได้ทราบผลการตัดสินของผู้ตัดสินคนอื่น ๆ
2. ภายหลังการได้ทราบการแจกแจงของคะแนนสอบของผู้สอบทั้งหมด
3. ภายหลังการได้ทราบผลการตัดสินในขั้นที่ 2 ของผู้ตัดสินทั้งหมดในรูปของสถิติภาคบรรยาย

เกณฑ์ จะถูกกำหนดโดยความสอดคล้องของการตัดสินของผู้ตัดสินทุกคน

วิธี Iterative two - choice Angoff (Gross, Impara, Frary, and Jaeger 1984 : 113-129) ขอร้องให้ผู้ตัดสินตอบคำถามว่า "ผู้สำเร็จการศึกษาชั้นเตรียมอุดมทุกคนสามารถตอบข้อสอบข้อนี้ได้ถูกต้องใช่หรือไม่" ผู้ตัดสินจะต้องตอบคำถามนี้ใน 3 สถานการณ์ คือ

1. ไม่มีสารสนเทศใด ๆ ประกอบการตัดสิน
2. มีสารสนเทศเกี่ยวกับการแจกแจงความถี่สะสมของเกณฑ์ที่ได้จากการตัดสินในสถานการณ์ที่ 1
3. มีสารสนเทศเกี่ยวกับค่าความยากของข้อสอบที่ได้จากการทดลองสอบ

เกณฑ์หาได้จากค่ามัธยฐานของเกณฑ์ของผู้ตัดสินทุกคนที่ตัดสินเกณฑ์ใน 3 สถานการณ์

วิธี Modified Angoff - empirical (Mclean & Halpin, cited by berk 1986 : 156) เป็นวิธีที่ให้ผู้ตัดสินประมาณ "ความน่าจะเป็น" สำหรับแต่ละความสามารถโดยรอบแรกจะให้ผู้ตัดสินประมาณความน่าจะเป็น และรอบที่สองจะให้ค่าเฉลี่ยของความยากของข้อสอบทั้งฉบับ เกณฑ์ของแบบสอบหาได้จากค่าเฉลี่ยของผลบวกของความน่าจะเป็นที่ได้จากการตัดสินในรอบที่สอง

เกณฑ์ในการประเมินวิธีกำหนดคะแนนจุดตัด

เบอร์ค (Berk 1986 : 140-145) ได้เสนอเกณฑ์ในการประเมินวิธีกำหนดจุดตัดที่มีมากมายหลายวิธี โดยใช้เกณฑ์ 2 อย่าง คือ

1. เกณฑ์เกี่ยวกับวิธีดำเนินการ ประกอบด้วย 6 ข้อ คือ

1.1 เป็นวิธีที่ให้จุดตัดที่เหมาะสม โดยปกติจะใช้จุดตัดในการแบ่งประเภทผู้สอบออกเป็น ผ่าน-ไม่ผ่าน มีความสามารถ-ไม่มีความสามารถ แต่บางครั้งอาจจำเป็นต้องใช้จุดตัด 2 จุด เช่นในการสอบแบบรายบุคคล (Adaptive testing) ซึ่งพิสัยระหว่างคะแนนจุดตัด 2 จุด นี้จะไม่สามารถตัดสินได้ว่า ผู้สอบ รอบรู้ หรือไม่รอบรู้

1.2 เป็นวิธีที่คำนึงถึงความสามารถของผู้สอบ และคุณภาพของข้อสอบ เช่น ค่าความยาก และค่าอำนาจจำแนกของข้อสอบ ถ้าแบบสอบประกอบด้วยข้อสอบที่มีค่าความยากเท่ากับ .3 ถึง .5 เป็นจำนวนมาก วิธีกำหนดจุดตัดที่ให้คะแนนจุดตัดมีค่าสูง จะเป็นจุดตัดที่ไม่เหมาะสมกับผู้สอบ

1.3 เป็นวิธีที่สอดคล้องกับการเรียนการสอน ถ้าแบบสอบวัดไม่ตรงกับวัตถุประสงค์การเรียนการสอนแล้ว คะแนนจุดตัดที่ได้ก็จะไม่มีความตรง หรือไม่เหมาะสม

1.4 เป็นวิธีที่ใช้สถิติซึ่งเหมาะสมและถูกต้อง เช่นในการคำนวณคะแนนจุดตัดที่ได้จากการตัดสินของผู้เชี่ยวชาญ โดยการคิดเป็นเปอร์เซนต์ การวัดแนวโน้มเข้าสู่ส่วนกลาง เป็นต้น

1.5 เป็นวิธีที่ให้คะแนนจุดตัดบนสเกลของคะแนนจริง แต่วิธีกำหนดจุดตัดส่วนใหญ่ จะหาคะแนนจุดตัดจากการแจกแจงของคะแนนสอบ และตรวจสอบความคลาดเคลื่อนของการวัด

1.6 ควรเป็นวิธีที่สามารถแสดงความตรงของการจำแนกผู้สอบออกเป็นผู้รอบรู้ และผู้ไม่รอบรู้ ด้วยคะแนนจุดตัด

2. เกณฑ์เกี่ยวกับการนำมาปฏิบัติได้จริง ประกอบด้วย 4 ข้อ คือ

2.1 เป็นวิธีที่เข้าใจง่าย ปฏิบัติได้ง่าย มีขั้นตอนการปฏิบัติที่ไม่ยุ่งยากซ้ำซ้อน ไม่เสียเวลา และเป็นระบบ

2.2 เป็นวิธีที่มีการคำนวณไม่ยุ่งยาก หรือสามารถใช้โปรแกรมสำเร็จรูปได้

2.3 เป็นวิธีที่สามารถอธิบายให้คนอื่นๆ ที่เกี่ยวข้อง เช่น นักเรียน ผู้ปกครอง ครู ผู้บริหารโรงเรียน สามารถเข้าใจได้โดยง่าย

2.4 เป็นวิธีที่น่าเชื่อถือ วิธีที่ใช้สถิติที่ยุ่งยากซับซ้อน ไม่จำเป็นต้องเป็นวิธีที่น่าเชื่อถือเสมอไป บางครั้งอาจทำให้สับสนได้มากกว่าวิธีที่ใช้การคำนวณง่าย ๆ

และเบอร์ก (Berk 1986 : 154-155) ได้ใช้เกณฑ์การประเมินวิธีกำหนดจุดตัดดังกล่าวข้างต้น 10 ข้อ ซึ่งแบ่งออกเป็นเกณฑ์การประเมินวิธีดำเนินการ 6 ข้อ และเกณฑ์การประเมินการนำมาปฏิบัติได้จริง 4 ข้อ ทำการประเมินวิธีกำหนดจุดตัดวิธีต่าง ๆ 23 วิธี ยกเว้นเฉพาะวิธีกำหนดจุดตัดที่ใช้ข้อมูลการทดลองสอบประมาณความคลาดเคลื่อน ในการจำแนกผู้สอบออกเป็นผู้รอบรู้ไม่จริง และผู้ไม่รอบรู้ไม่จริง ซึ่งเป็นวิธีที่ยุ่งยากซับซ้อนและนำมาปฏิบัติได้ยาก แล้วสรุปผลการประเมินได้ดังแสดงในตารางที่ 1

ตารางที่ 1 ผลการประเมินวิธีกำหนดจุดตัด 23 วิธี

วิธีกำหนดจุดตัด	การประเมิน										ค่าเฉลี่ยการประเมิน
	วิธีดำเนินการ						การนำมาปฏิบัติ				
	1	2	3	4	5	6	7	8	9	10	
1. วิธีที่ใช้ดุลพินิจของผู้เชี่ยวชาญตัดสิน											
1.1 วิธี Adjusted/modified M-C Angoff	++	--	-	+	+	--	++	-	++	++	+
1.2 วิธีแองกอฟ (Angoff)	++	--	-	++	++	--	+	++	++	++	+
1.3 วิธี Angoff - Nedelsky combination	++	--	-	+	+	--	--	-	-	+	-
1.4 วิธี Difficulty - importance estimate	++	--	-	++	--	--	-	+	+	+	-
1.5 วิธีอีเบล (Ebel)	++	--	-	-	++	--	-	+	-	-	-
1.6 วิธี Difficulty - taxonomy Ebel	++	+	-	-	-	--	-	+	-	-	-

ตารางที่ 1 (ต่อ)

วิธีกำหนดจุดตัด	การประเมิน										ค่าเฉลี่ยการประเมิน
	วิธีดำเนินการ					การนำมาปฏิบัติ					
	1	2	3	4	5	6	7	8	9	10	
1.7 วิธี Item specifications	++	--	-	-	--	--	+	+	+	+	-
1.8 วิธี Modified M-C Angoff	++	--	-	-	-	--	++	++	++	++	+
1.9 วิธีนีเดลสกี (Nedelsky)	++	--	-	-	++	--	--	-	-	+	-
1.10 วิธี Relevamce - taxonomy Ebel	++	--	--	-	-	--	-	+	-	-	-
1.11 วิธี Two - Choice Angoff	++	--	-	-	-	--	++	+	+	+	-+
2. วิธีที่อาศัยการทดลองสอบ											
2.1 วิธีสอบกับเด็กคาบเส้น (Borderline group)	++	++	+	+	-	--	+	+	+	-	+
2.2 วิธีการเปรียบเทียบระหว่างกลุ่มผู้สอบ (Contrasting groups)	++	++	++	+	+	++	-	+	+	-	+
2.3 วิธีใช้เกณฑ์ภายนอก (Criterion groups)	++	++	++	+	-	++	-	+	+	-	+
2.4 วิธีพิจารณาผลทางการศึกษาที่ตามมา (Educational consequences)	++	++	++	--	--	--	+	+	-	-	-+
2.5 วิธีใช้ผลการสอบของคนอื่น ๆ เป็นเกณฑ์ (Norm - referenced criterion)	++	++	-	-	--	-	+	+	-	-	-+

ตารางที่ 1 (ต่อ)

วิธีกำหนดจุดตัด	การประเมิน										ค่าเฉลี่ยการประเมิน
	วิธีดำเนินการ					การนำมาปฏิบัติ					
	1	2	3	4	5	6	7	8	9	10	
3. วิธีที่ใช้ดุลพินิจของผู้เชี่ยวชาญโดยมีข้อมูลการทดสอบ											
3.1 วิธีหาจุดสมดุลย์ระหว่างเกณฑ์สัมบูรณ์และเกณฑ์สัมพันธ์แบบที่ 1 (Absolute - relative compromise I)	++	+	+	++	+	--	-	--	-	-	-+
3.2 วิธีหาจุดสมดุลย์ระหว่างเกณฑ์สัมบูรณ์และเกณฑ์สัมพันธ์แบบที่ 2 (Absolute - relative compromise II)	++	+	+	++	+	--	--	--	-	-	-+
3.3 วิธี Angoff-contrasting groups plus composite	++	++	+	+	+	++	--	--	+	+	+
3.4 วิธีการใช้ดุลพินิจตัดสินโดยอาศัยสารสนเทศประกอบ (Informed judgment)	++	+	+	++	++	--	+	++	++	++	+/++
3.5 วิธี Iterative Angoff	++	+	+	+	++	--	--	++	+	++	+
3.6 วิธี Iterative two - choice Angoff	++	+	+	+	+	--	--	-	+	++	+
3.7 วิธี Modified Angoff - empirical	++	+	-	-	--	--	+	+	++	++	+

หมายเหตุ : -- คือ ไม่ดีอย่างยิ่ง
- คือ ไม่ดี
+ คือ ดี
++ คือ ดีอย่างยิ่ง

ผลการประเมินวิธีกำหนดจุดตัดทั้งหมด 23 วิธี ของเบอร์ก ปรากฏว่า

1. วิธีกำหนดจุดตัดที่ใช้ดุลพินิจของผู้เชี่ยวชาญตัดสินทั้ง 11 วิธี เป็นวิธีที่ไม่มีหลักการที่ชัดเจน มักเป็นไปตามอำเภอใจมากกว่า แต่ก็พบว่าวิธีของแองกอฟเป็นวิธีที่เหมาะสมที่สุด มีความสมดุลย์ทั้งวิธีการและการนำไปปฏิบัติ

2. วิธีกำหนดจุดตัดที่อาศัยการทดลองสอบ ภายใน 5 วิธีนี้ พบว่าวิธีใช้การเปรียบเทียบระหว่างกลุ่มผู้สอบ เป็นวิธีที่น่าเชื่อถือที่สุด นอกจากนี้ยังมีวิธีการที่น่าเชื่อถือที่สุดในวิธีกำหนดจุดตัดทั้งหมด 23 วิธีนี้ด้วย

3. วิธีกำหนดจุดตัดที่อาศัยการใช้ดุลพินิจของผู้เชี่ยวชาญตัดสินกับการทดลองสอบร่วมกัน ภายใน 7 วิธีนี้ พบว่าวิธีการใช้ดุลพินิจตัดสิน โดยอาศัยสารสนเทศประกอบเป็นวิธีที่น่าเชื่อถือที่สุด และยังเป็นวิธีที่น่าเชื่อถือที่สุดในวิธีกำหนดจุดตัดทั้งหมด 23 วิธีนี้ด้วย

ตอนที่ 2 ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory)

หลักการของทฤษฎีการตอบสนองข้อสอบ คือ

1 สามารถทำนายผลการสอบของนักเรียนได้โดยคุณลักษณะแฝงที่อยู่ภายในตัวบุคคล หรือที่เรียกว่าความสามารถ

2 ความสัมพันธ์ระหว่างการตอบข้อสอบกับความสามารถที่อยู่ภายในตัวผู้สอบมีอิทธิพลกับผลการตอบข้อสอบได้ถูก อธิบายได้ด้วยฟังก์ชันลักษณะข้อสอบ (Item Characteristic function) ซึ่งแสดงให้เห็นว่าผู้สอบที่มีความสามารถสูงกว่าจะมีโอกาสตอบข้อสอบได้ถูกสูงกว่าผู้ที่มีความสามารถต่ำกว่า (Hambleton 1985 : 13-14)

ข้อตกลงเบื้องต้นของทฤษฎีการตอบข้อสอบ

1. มิติของคุณลักษระแฝงที่อยู่ภายในหรือความสามารถของผู้สอบที่ใช้ในการตอบข้อสอบในแบบสอบฉบับหนึ่งต้องมีมิติเดียว คือ ข้อสอบแต่ละข้อในแบบสอบจะวัดความสามารถเดียวกัน (Unidimensional)

2. ข้อสอบแต่ละข้อเป็นอิสระกัน (Local Independence) คือ การตอบข้อสอบแต่ละข้อของผู้สอบเป็นอิสระกัน

3. โค้งลักษณะข้อสอบ (Item Characteristic Curve) เป็นฟังก์ชันทางคณิต-ศาสตร์ ที่แสดงถึงความสัมพันธ์ระหว่างโอกาสในการตอบข้อสอบข้อนั้นได้ถูกต้องกับระดับความสามารถที่วัดได้โดยชุดของข้อสอบหรือแบบสอบนั้น ถ้าแบบสอบนั้นวัดความสามารถเดียวแล้วโค้งลักษณะข้อสอบจะไม่แปรเปลี่ยนไปตามกลุ่มประชากรผู้สอบ นั่นคือโอกาสในการตอบข้อสอบได้ถูก

ของผู้สอบคนหนึ่งขึ้นอยู่กับโค้งลักษณะข้อสอบนั้น และไม่ขึ้นกับประชากรผู้สอบ โค้งลักษณะข้อสอบมีอยู่ 4 รูปแบบ ตามรูปแบบการตอบสนองข้อสอบ (Item Response Model) โดยที่จำนวนของพารามิเตอร์ที่ใช้อธิบายโค้งลักษณะข้อสอบ จะขึ้นกับรูปแบบการตอบสนองข้อสอบ (Hambleton 1985 : 15-31)

<u>รูปแบบการตอบสนองข้อสอบ</u> (Item Response Model)

รูปแบบทางคณิตศาสตร์ที่ใช้ในการวิเคราะห์ข้อสอบ อาศัยฟังก์ชันการกระจายรูปแบบโลจีสติก (logistic distribution function) ซึ่งแบ่งออกเป็น 4 รูปแบบ ตามจำนวนพารามิเตอร์ของโค้งลักษณะข้อสอบ

1. รูปแบบโลจิสติกที่มี 1 พารามิเตอร์ (One-Parameter Logistic Model) รูปแบบนี้ถือว่าข้อสอบทุกข้อมีค่าอำนาจจำแนกเท่ากัน ($a_i = 1.00$) และไม่มีค่าการเดา ($c_i = 0$) คำนึงถึงเฉพาะค่าความยากของข้อสอบ (b_i) ดังสมการ

$$P_i(\theta) = \frac{e^{D(\theta-b_i)}}{1+c^{D(\theta-b_i)}} \qquad (i = 1,2,\ldots,n)$$

เมื่อ $p_i(\theta)$ คือ ความน่าจะเป็นที่ผู้สอบซึ่งมีระดับความสามารถ θ จะตอบข้อสอบข้อ i ได้ถูก

b_i คือ ค่าความยากของข้อสอบข้อ i ที่แสดงถึงระดับความสามารถที่จุดโค้งลักษณะข้อสอบมีความชันที่สุด ในทางปฏิบัติจะมีค่าอยู่ระหว่าง -2 ถึง +2

D คือ ค่า Scaling factor = 1.7

2. รูปแบบโลจีสติกที่มี 2 พารามิเตอร์ (Two-Parameter Logistic Model) รูปแบบนี้ถือว่าข้อสอบทุกข้อไม่มีค่าการเดา คำนึงถึงค่าความยากและค่าอำนาจจำแนกของข้อสอบเท่านั้น ดังสมการ

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

เมื่อ a_i คือ ค่าอำนาจจำแนกของข้อสอบข้อ i ที่เป็นสัดส่วนโดยตรงกับความชันของโค้งลักษณะข้อสอบ ซึ่งมีค่าได้ตั้งแต่ $-\infty$ ถึง $+\infty$ แต่ในทางปฏิบัติจะใช้ค่าที่อยู่ระหว่าง 0 ถึง 2

3. รูปแบบโลจีสติกที่มี 3 พารามิเตอร์ (Three-Parameter Logistic Model) รูปแบบนี้คำนึงถึงค่าความยาก ค่าอำนาจจำแนกและค่าการเดาของข้อสอบ ดังสมการ

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

เมื่อ C_i คือ ความน่าจะเป็นที่ผู้สอบซึ่งมีความสามารถต่ำมากจะตอบข้อสอบข้อ i ได้ถูก ในทางปฏิบัติจะเลือกข้อสอบที่มีค่าการเดา $\leqslant 0.3$

4. รูปแบบโลจีสติกที่มี 4 พารามิเตอร์ (Four-Parameter Logistic Model) รูปแบบนี้คำนึงถึงความสะเพร่าในการตอบข้อสอบด้วย นอกเหนือจาก 3 พารามิเตอร์ที่กล่าวมาแล้ว ดังสมการ

$$P_i(\theta) = c_i(r_i-c_i)\frac{e^{Da_i(\theta-b_i)}}{1+c^{Da_i(\theta-b_i)}} \qquad (i=1,2,..,n)$$

เมื่อ r_i คือ ค่าความรอบคอบของข้อสอบข้อ i ซึ่งมีค่าน้อยกว่า 1

แต่รูปแบบนี้ยังไม่สามารถนำมาใช้ในทางปฏิบัติได้

<u>คุณสมบัติของคะแนนความสามารถ</u>

การประมาณคะแนนความสามารถ (θ) ด้วยค่า $\hat{\theta}$ เมื่อรู้ค่าพารามิเตอร์ของข้อสอบโดยการใช้วิธีการประมาณค่าความเป็นไปได้สูงสุด (Maximum likelihood estimation) ซึ่งค่าที่ประมาณได้จะมีคุณสมบัติ ดังนี้

1. มีความคงเส้นคงวา (Consistent) คือเมื่อขนาดของกลุ่มตัวอย่างและจำนวนข้อสอบเพิ่มขึ้น การประมาณค่าจะเข้าใกล้ค่าที่แท้จริง

2. เป็นค่าสถิติที่เพียงพอ (Sufficient) คือ มีข้อมูลทั้งหมดที่เกี่ยวกับค่าพารามิเตอร์ โดยไม่จำเป็นต้องอาศัยข้อมูลอื่น ๆ อีก

3. มีประสิทธิภาพ (Efficient) คือ การประมาณค่าด้วยวิธีนี้จะมีความแปรปรวนน้อยที่สุด

4. การกระจายของค่าสถิติเข้าใกล้การกระจายแบบปกติ (Normally distributed) โดยมีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ $[I(\theta)]^{-1}$ เมื่อ $I(\theta)$ คือ ฟังก์ชันสารสนเทศแบบสอบ (test information function) ซึ่งมีสมการ ดังนี้

$$I(\theta) = -E[\partial^2 \ln L/\partial\theta^2] = \sum_{i=1}^{n} P_i'^2/P_iQ_i$$

เมื่อ E คือ ค่าคาดหวัง

p_i' คือ อนุพันธ์ของฟังก์ชันการตอบข้อสอบ ข้อ i ณ ระดับความสามารถ

q_i คือ $1-p_i$

และ ความแปรปรวนของ $\hat{\theta}$ เขียนเป็นสมการได้ดังนี้

$$Var(\hat{\theta}/\theta) = [I(\theta)]^{-1}$$

โดยที่ $\hat{\theta}$ เป็นค่าสถิติที่เชื่อถือได้ เราจึงแทน θ ด้วย $\hat{\theta}$ ได้

เบอร์นบอม (Birnbaum 1968 : 457) ได้แสดงขอบเขตความเชื่อมั่นของการประมาณค่า θ (maximum likelihood confidence limit estimator) โดยมีช่วงความเชื่อมั่น $(1-\alpha)$ ดังนี้

$$\hat{\theta} - Z_{\frac{\alpha}{2}} [I(\hat{\theta})]^{-\frac{1}{2}} \leq \theta \leq \hat{\theta} + Z_{\frac{\alpha}{2}} [I(\hat{\theta})]^{-\frac{1}{2}}$$

ทำให้ได้ว่า $[I(\hat{\theta})]^{-\frac{1}{2}}$ คือ ความคลาดเคลื่อนมาตรฐานของการประมาณค่า θ ด้วยวิธีความเป็นไปได้สูงสุดที่เขียนเป็นสมการได้ ดังนี้ (Hambleton 1985 : 88-90)

$$SE(\theta) = \frac{1}{\sqrt{I(\theta)}}$$

ฟังก์ชันสารสนเทศแบบสอบหาได้จากผลรวมของฟังก์ชันสารสนเทศข้อสอบ $(I(\theta, u_i)$ แต่ละข้อในแบบสอบซึ่งเขียนเป็นสมการได้ดังนี้

$$I(\theta, u_i) = (P_i')^2/P_iQ_i$$

ฟังก์ชันสารสนเทศข้อสอบจะขึ้นอยู่กับความชันของฟังก์ชันการตอบข้อสอบ และความแปรปรวนที่แต่ละระดับความสามารถ (θ) ค่าสารสนเทศจะสูงขึ้นเมื่อความชันสูงขึ้น และความแปรปรวนน้อยลง และจะทำให้ความคลาดเคลื่อนมาตรฐานของการวัดมีค่าน้อยลง

<u>ความสัมพันธ์ของคะแนนความสามารถ (θ) กับคะแนนโดเมน (π)</u>

ถ้านำแบบสอบ 2 ชุด ที่วัดความสามารถเดียวกันไปดำเนินการสอบกับผู้สอบกลุ่มหนึ่ง และแบบสอบทั้ง 2 ชุด นั้นไม่ได้ขนานกันอย่างแท้จริง จะได้ผลการแจกแจงของคะแนนสอบที่ต่างกัน 2 ชุด ความแตกต่างของการแจกแจงคะแนน 2 ชุด นี้ขึ้นอยู่กับความแตกต่างของความยากของแบบสอบ 2 ฉบับ นั้นจึงไม่อาจบอกได้ว่าการแจกแจงหนึ่งดีกว่าอีกการแจกแจงหนึ่ง การแจกแจงของคะแนนสอบไม่ได้ให้ข้อมูลเกี่ยวกับการแจกแจงของคะแนนความสามารถเพราะคะแนนดิบจากแบบสอบแต่ละฉบับไม่เท่ากันและต่างกัน โดยที่เสกลความสามารถ คือ สเกลอันหนึ่ง ซึ่งผู้สอบคนหนึ่งจะมีคะแนนความสามารถเท่ากัน บนแบบสอบที่ไม่ขนานกันเหล่านั้น ดังนั้น ถึงแม้ว่าคะแนนของผู้สอบคนหนึ่งจะเปลี่ยนแปลงไปตามแบบสอบที่ไม่ขนานกัน แต่คะแนนความสามารถสำหรับผู้สอบคนหนึ่งจะเป็นอย่างเดียวกันบนแบบสอบแต่ละฉบับนั้น

มโนทัศน์ของคะแนนจริง หรือคะแนนโดเมน เป็นสิ่งสำคัญพื้นฐานในทฤษฎีคลาสสิคอล (Classical test theory) โดยถูกกำหนดให้เป็นคะแนนคาดหวังบนเซทของข้อสอบ ของผู้สอบคนหนึ่ง

ถ้าให้ x = คะแนนที่สังเกตได้
t = คะแนนจริง
แล้วให้ $t = E(x)$

ความสัมพันธ์ของคะแนนจริง หรือคะแนนโดเมน กับคะแนนความสามารถ θ หาได้ดังนี้

ถ้าให้ r = คะแนนที่สังเกตได้
n = จำนวนข้อสอบทั้งฉบับ

แล้วให้ $$r = \sum_{i=1}^{n} U_i$$

เมื่อ U_i = คำตอบของข้อสอบข้อ i ซึ่งถ้าตอบถูกจะเป็น 1 ถ้าตอบผิดจะเป็น 0

และถ้าให้ π = คะแนนโดเมน

$\hat{\pi}$ = สัดส่วนของคะแนนถูก

แล้วจะได้ $\hat{\pi} = \frac{1}{n} \sum_{i=1}^{n} U_i$

จากมโนทัศน์ของคะแนนจริงได้ว่า $E(\hat{\pi}) = \pi$

ดังนั้นจึงได้ว่า $E(\hat{\pi}) = \pi = \frac{1}{n} \sum_{i=1}^{n} E(U_i)$

และที่ระดับความสามารถ θ ทำให้

$$\hat{\pi}/\theta = \frac{1}{n} \sum_{i=1}^{n} (U_i/\theta)$$

$$E(\hat{\pi}/\theta) = \pi/\theta$$

$$= \frac{1}{n} \sum_{i=1}^{n} E(U_i/\theta)$$

ซึ่ง U_i เป็นตัวแปรเชิงสุ่มที่มีค่า 0 หรือ 1

ดังนั้นทำให้ $E(U_i/\theta) = (U_i=1)\ P_i\ (U_i=1/\theta) + (U_i=0)\ P_i\ (U_i=0/\theta)$

$= 1.P_i\ (U_i=1/\theta) + 0.P_i\ (U_i=0/\theta)$

$= P_i\ (U_i=1/\theta)$

$= P_i\ (\theta)$

ทำให้ได้ว่า $\pi/\theta = \frac{1}{n} \sum_{i=1}^{n} P_i(\theta)$

ฟังก์ชันนี้คือค่าเฉลี่ยของฟังก์ชันการตอบข้อสอบ (Item response function) ของข้อสอบ n ข้อ เรียกว่า ฟังก์ชันลักษณะแบบสอบ (Test Characteristic Function) ถ้านำไปพลอตกราฟ จะได้โค้งลักษณะแบบสอบ (Test Characteristic Curve) และความแปรปรวนสามารถแทนได้ด้วย

$$\mathrm{Var}(\hat{\pi}/\theta) = \mathrm{Var}\left\{\frac{1}{n} \sum_{i=1}^{n} (U_i/\theta)\right\}$$

จากข้อตกลงเบื้องต้นของความเป็นอิสระของข้อสอบ จะได้ว่า

$$\text{Var}\ \frac{1}{n}\sum_{i=1}^{n}(U_i/\theta) = \frac{1}{n^2}\sum_{i=1}^{n}\text{Var}\ (U_i/\theta)$$

เนื่องจาก U_i/θ เป็น binomial variable

$$\text{Var}(U_i/\theta) = P_i(\theta)\, Q_i(\theta)$$

ดังนั้น

$$\text{Var}\frac{1}{n}(U_i/\theta) = \frac{1}{n^2} P_i(\theta)\, Q_i(\theta)$$

จึงได้ว่า

$$\text{Var}(\hat{\pi}/\theta) = \frac{1}{n^2}\sum_{i=1}^{n} P_i(\theta)Q_i(\theta)$$

เมื่อแบบสอบยาวขึ้น โดยการเพิ่มจำนวนข้อสอบที่ขนานกันเข้าไปจนกระทั่ง $n \to \infty$ แล้วจะทำให้ $\hat{\pi}/\theta \longrightarrow \pi$ และ $\text{Var}(\hat{\pi}/\theta) \longrightarrow 0$ ดังนั้นจึงกล่าวได้ว่า $\hat{\pi}$ เป็นตัวประมาณค่าที่คงเส้นคงวาของ π (คะแนนโดเมน)

สรุปได้ว่า คะแนนโดเมน (π) และคะแนนความสามารถ (θ) มีความสัมพันธ์กันแบบขึ้นทางเดียว (monotonically) ความสัมพันธ์หาได้โดยฟังก์ชันลักษณะแบบสอบ คะแนนโดเมนและคะแนนความสามารถมีมโนทัศน์เหมือนกัน แต่สเกลการวัดแตกต่างกัน คือ คะแนนโดเมนจะอยู่ในช่วง (0, 1) ขณะที่คะแนนความสามารถจะอยู่ในช่วง $(-\infty, \infty)$ และที่สำคัญที่สุดคือสเกลของความสามารถเป็นอิสระจากข้อสอบ แต่สเกลของคะแนนโดเมนขึ้นกับข้อสอบที่ใช้

ดังนั้นเมื่อรู้ค่าพารามิเตอร์ (ค่าความยาก ค่าอำนาจจำแนก ค่าการเดา) ของข้อสอบทั้งหมดในโดเมน หรือของกลุ่มตัวแทนของข้อสอบชุดหนึ่งในโดเมนนั้น แล้วสามารถหาความสัมพันธ์ระหว่างคะแนนโดเมน กับคะแนนความสามารถได้ โดยที่คะแนนโดเมนจะหมายถึงคะแนนที่คาดหวังที่ได้จากข้อสอบทั้งหมดในโดเมนนั้น แต่เมื่อใช้กลุ่มตัวอย่างของข้อสอบที่มีความเป็นตัวแทนของข้อสอบทั้งหมดในโดเมนนั้น ก็จะสามารถประมาณความสัมพันธ์ระหว่าง π และ θ ได้โดย

$$\pi = \frac{1}{m}\sum_{i=1}^{m} P_i(\theta)$$

เมื่อ m = จำนวนข้อสอบในแบบสอบ

โดยทั่วไปจะกำหนดคะแนนจุดตัดบนสเกลของคะแนนโดเมน (π_o) และสามารถแปลงให้อยู่ในสเกลของคะแนนความสามารถ (θ_0) ด้วยสมการข้างต้น (Hambleton 1985 : 260)

<u>แบบสอบอิงเกณฑ์กับทฤษฎีการตอบข้อสอบ</u>

แฮมเบิลตัน (Hambleton 198. : 33-34) ได้ศึกษาเปรียบเทียบการใช้รูปแบบโลจิสติกที่มี 1, 2 และ 3 พารามิเตอร์ กับแบบสอบอิงเกณฑ์ในการประมาณคะแนนโดเมนและการตัดสินให้ผู้สอบเป็นผู้รอบรู้ หรือผู้ไม่รอบรู้ เพื่อเปรียบเทียบว่ารูปแบบใดจะให้คะแนนโดเมนที่มีความคงเส้นคงวากว่า (consistent) ทำการศึกษาโดยสร้างข้อมูล θ ต่าง ๆ ของผู้สอบด้วยเหตุการณ์จำลอง (simulate) จำนวน 2,000 คน ที่มีค่าเฉลี่ยเท่ากับ 0 และค่าเบี่ยงเบนมาตรฐานเท่ากับ 1 จากการตอบข้อสอบจำนวน 40 ข้อ แล้วแปลงระดับความสามารถ (θ) ให้เป็นคะแนนโดเมน (π) โดยอาศัยโค้งลักษณะแบบสอบ ดังสมการ

$$\pi_i = \frac{1}{n} \sum_{g=1}^{n} P_g(\theta_i) \qquad (1)$$

เมื่อ $P_g(\theta_i)$ คือ ความน่าจะเป็นที่ผู้สอบ i ที่มีระดับความสามารถ θ จะตอบข้อสอบ g ได้ถูก

n คือ จำนวนข้อสอบในแบบสอบ

π_i คือ คะแนนโดเมนของผู้สอบคนที่ i บนแบบสอบ

ทำการศึกษากับแบบสอบที่มีความยาวต่าง ๆ คือ 10, 15, 20 และ 40 ข้อ และที่คะแนนจุดตัด 3 จุด คือ O_0 = -1.50, 0 และ 1.50 ส่วนกลุ่มข้อสอบก็ได้จากการสร้างขึ้นเองโดยอาศัยคอมพิวเตอร์ ใช้จำนวนผู้สอบ 2,000 คน เพื่อหาค่าพารามิเตอร์ของข้อสอบจำนวน 40 ข้อ ข้อสอบมีค่าความยากอยู่ในช่วง (-2.00, +2.00) ค่าอำนาจจำแนกอยู่ในช่วง (.40, 2.00) และค่าการเดาอยู่ในช่วง (.15, .25)

เกณฑ์ที่ใช้เปรียบเทียบ

1. ความถูกต้องแม่นยำ (precision) ของการประมาณคะแนนโดเมนโดยอาศัยค่าเฉลี่ยของความเบี่ยงเบนสัมบูรณ์ของคะแนนโดเมนของผู้สอบ 200 คน ที่ระดับความสามารถต่าง ๆ (-2.00, -1.00, -.50, 1.00) ดังนี้

$$\sum_{j=1}^{N} \left| \pi_i - \pi_{ij} \right| / N \qquad (2)$$

เมื่อ π_i ได้จากสมการที่ 1

$$\hat{\pi}_{ij} = \frac{1}{n} \sum_{g=1}^{n} \hat{P}_g(\hat{\theta}_{ij}) \qquad (3)$$

n คือ จำนวนข้อสอบ

$\hat{\theta}_{ij}$ คือ ความสามารถของผู้สอบคนที่ i บนแบบสอบฉบับที่ j

N คือ จำนวนผู้สอบ (N = 200 คน)

จะได้ค่าเฉลี่ยของความเบี่ยงเบนสัมบูรณ์ระหว่าง คะแนนโดเมนที่เป็นคะแนนจริงของผู้สอบกับคะแนนโดเมนที่ได้จากการประมาณด้วยแต่ละรูปแบบโลจีสติกที่ระดับความสามารถต่าง ๆ

2. ความลำเอียงในการประมาณคะแนนโดเมนของแต่ละรูปแบบโลจีสติก

3. ความถูกต้องแม่นยำของการตัดสินให้ผู้สอบเป็นผู้รอบรู้หรือไม่รอบรู้ ของแต่ละรูปแบบโลจีสติก

ผลการเปรียบเทียบพบว่า

1. การเปรียบเทียบความถูกต้องแม่นยำ ของการประมาณคะแนนโดเมนที่ความยาวของแบบสอบ 4 ขนาด ด้วยรูปแบบโลจีสติก 1, 2 และ 3 พารามิเตอร์ ที่ระดับความสามารถ 5 ระดับ จาก -2.00 ถึง +1.00 ปรากฏว่ารูปแบบ 3 พารามิเตอร์ จะประมาณคะแนนโดเมนได้ดีกว่ารูปแบบ 1 หรือ 2 พารามิเตอร์

2. ความลำเอียงในการประมาณคะแนนโดเมนด้วยโลจีสติกทั้ง 3 รูปแบบ มีค่าน้อยมาก

3. ความถูกต้องแม่นยำของการตัดสินให้ผู้สอบเป็นผู้รอบรู้หรือไม่รอบรู้ พบว่ารูปแบบ 2 พารามิเตอร์ให้ความถูกต้องของการตัดสินที่ไม่แน่นอน แต่รูปแบบ 1 พารามิเตอร์ ใช้ได้ดีพอๆ กับรูปแบบ 3 พารามิเตอร์ ยกเว้นที่ระดับความสามารถต่ำ ๆ (-2.0 ถึง -.5)

<u>การใช้ทฤษฎีการตอบข้อสอบกับวิธีกำหนดคะแนนคะแนนจุดตัดที่ใช้ดุลพินิจของผู้เชี่ยวชาญตัดสิน</u>

เคน (Kane 1987 : 333-345) ได้รวบรวมวิธีการแปลงคะแนนจุดตัดที่ได้จากวิธีการกำหนดคะแนนจุดตัดโดยใช้ดุลพินิจของผู้เชี่ยวชาญตัดสิน ให้เป็นระดับความสามารถขั้นต่ำสุดที่ยอมให้ผ่าน (θ^*) โดยมีข้อตกลงว่า จะต้องทราบค่าโค้งลักษณะข้อสอบ ($P_i(\theta)$) ซึ่งเป็นหลัก

ของทฤษฎีการตอบสนองข้อสอบในการประมาณความน่าจะเป็นที่ผู้สอบซึ่งมีระดับความสามารถ θ ต่าง ๆ จะสามารถตอบข้อสอบได้ถูก ซึ่งมีอยู่ 3 วิธี คือ

<u>วิธีที่ 1</u> ประมาณค่า θ^* ของข้อสอบแต่ละข้อโดยอาศัยโค้งลักษณะข้อสอบรายข้อ แล้วจึงคำนวณหา θ^* เฉลี่ยของข้อสอบทุกข้อในแบบสอบ มี 3 ขั้นตอน ดังนี้

1. คำนวณหาค่าเฉลี่ยของความน่าจะเป็นในการตอบข้อสอบ i ได้ถูกที่ใช้ดุลพินิจของผู้เชี่ยวชาญตัดสิน (M_{iR})

2. ให้ M_{iR} เป็นค่าประมาณของความน่าจะเป็นที่ผู้สอบซึ่งมีระดับความสามารถขั้นต่ำสุดที่ยอมให้ผ่าน สามารถตอบข้อสอบได้ถูกในรูปแบบโลจีสติก ซึ่งแทนได้ด้วยโค้งลักษณะข้อสอบที่ระดับความสามารถ $\hat{\theta}^*_{iR}$ ดังสมการ

$$M_{iR} = P_i (\hat{\theta}^*_{iR})$$

3. ประมาณค่าความแปรปรวนของ $(\hat{\theta}^*_{iR})$ ในข้อ i

กำหนดให้ $$P_i (\theta_{iR}) = \frac{\Delta M_{iR}}{\Delta \theta_{iR}} \qquad (6)$$

เป็นความชันของโค้งลักษณะข้อสอบข้อ i ที่ความสามารถ (O_{iR})
แทน ด้วยความเบี่ยงเบนมาตรฐาน

$$\frac{P_i' (\theta_{iR})}{\sigma_i (M_{iR})} \quad \frac{1}{\sigma_i (\hat{\theta}^*_{iR})} \qquad (7)$$

$$\frac{[P_i' (\theta_{iR})]^2}{\sigma^2 (M_{iR})} \quad \frac{1}{\sigma^2_i (\hat{\theta}^*_{iR})} \qquad (8)$$

4. ประมาณค่า θ^* จากค่าเฉลี่ยของ $\hat{\theta}^*_{iR}$

$$\hat{\theta}^* = \frac{1}{n} \sum_i (\hat{\theta}^*_{iR}) \qquad (9)$$

ซึ่งมีความแปรปรวน $$\sigma^2(\hat{\theta}^*) = \frac{1}{n^2} \sum_i \sigma^2_i (\hat{\theta}^*_{iR}) \qquad (10)$$

โดยมีข้อตกลงเบื้องต้นว่าการตัดสินความน่าจะเป็นในการตอบถูกของข้อสอบแต่ละข้อเป็นอิสระกัน

แต่การที่ผู้ตัดสินคนหนึ่งต้องตัดสินข้อสอบทุกข้อในแบบสอบอาจทำให้ผลการตัดสินข้อสอบรายข้อไม่เป็นอิสระกัน จึงแทนสมการที่ (9) ด้วยการประมาณแบบถ่วงน้ำหนัก ดังนี้

$$\hat{\theta}^{*}_{w} = \sum_{i} w_i \hat{\theta}^{*}_{iR} \qquad (11)$$

เมื่อ $$\Sigma_i w_i = 1 \qquad (12)$$

และมีความแปรปรวนเป็น

$$\sigma^2(\theta_w) = \sum_{i} w_i 2 \sigma_i^2 (\hat{\theta}^{*}_{iR}) \qquad (13)$$

ค่าของ W_i ที่จะทำให้ความแปรปรวนในสมการที่ (13) มีค่าน้อยที่สุดและมีคุณสมบัติดังในสมการที่ (12) ทำได้โดยการใช้ Lagrange multipliers

$$w_i = \frac{1/\sigma_i^2 \ (\theta_{iR})}{{}_i(1/\sigma_i^2)\ (\theta_{iR})} \qquad (14)$$

การประมาณแบบถ่วงน้ำหนักนี้จะช่วยลดความคลาดเคลื่อนมาตรฐานของ $\hat{\theta}_i^*$ โดยข้อสอบที่ผู้ตัดสิน ตัดสินคล้าย ๆ กัน จะถูกถ่วงน้ำหนักมาก ส่วนข้อที่ผู้ตัดสิน ตัดสินต่างกันออกไปก็จะถูกถ่วงน้ำหนักน้อย

<u>วิธีที่ 2</u> เป็นวิธีที่ใช้กันมากที่สุด เป็นการประมาณ θ^* โดยอาศัยโค้งลักษณะแบบสอบ มี 3 ขั้นตอน คือ

1. คำนวณหาค่าเฉลี่ยของความน่าจะเป็นในการตอบข้อ i ถูก ที่ตัดสินโดยผู้ตัดสินทั้งหมด (M_{iR})

2. คำนวณหาคะแนนจุดตัดของแบบสอบ ซึ่งอยู่บนเสกลของคะแนนจริงจากผลรวมของ M_{iR} ในแต่ละข้อของแบบสอบ

3. ใช้โค้งลักษณะแบบสอบ (test characticsistic curve) หาระดับความสามารถ θ ซึ่งสัมพันธ์กับคะแนนจริงในขั้นที่ 2 จากสมการ

$$\sum_{i} M_{iR} = \sum_{i} P_i(\hat{\theta}^{*}) \qquad (15)$$

$$\sum_{i} [M_{iR} - P_i(\hat{\theta}^{*})] = 0 \qquad (16)$$

จากสมการที่ (5) $\sum_i [P_i(\hat{\theta}^*_{iR}) - P_i(\hat{\theta}^*)] = 0$ (17)

แต่ $\hat{\theta}^*_{iR} \rightarrow \theta^*$ ความสัมพันธ์สามารถประมาณได้ด้วย linear function (lord 1980 : 69)

$$\sum_i P'_i(\hat{\theta}^*_{iR})(\hat{\theta}^*_{iR} - \hat{\theta}^*) \simeq 0 \quad (18)$$

$$\theta^* \simeq \frac{1}{\sum_i P'_i(\hat{\theta}^*_{iR})} \sum_i P'_i(\hat{\theta}^*_{iR})\hat{\theta}^*_{iR} \quad (19)$$

จากสมการที่ (7) และ (19)

$$\theta^* \simeq \frac{1}{\sum_i \sigma_i(M_{iR})/\sigma_i(\hat{\theta}^*_{iR})} \sum_i \frac{\sigma_i(M_{iR})}{\sigma_i(\theta_{iR})} \hat{\theta}^*_{iR} \quad (20)$$

การประมาณ Θ^* จากสมการที่ 20 ยังไม่ใช่ค่าที่ดีที่สุด สามารถปรับปรุงได้โดยการถ่วงน้ำหนักด้วย $1/\sigma_i(M_{iR})$ ในสมการที่ (17) จะได้

$$\hat{\theta}^*_w \simeq \frac{1}{\sum_i P'_i(\hat{\theta}^*_{iR})/\sigma_i(M_{iR})} \sum_i \frac{P'_i(\hat{\theta}^*_{iR})}{\sigma_i(M_{iR})} \hat{\theta}^*_{iR} \quad (21)$$

จากสมการที่ (7) จะได้ว่า

$$\hat{\theta}^*_w \simeq \frac{1}{\sum_i (1/\sigma_i(\hat{\theta}^*_{iR}))} \sum_i \frac{1}{\sigma_i(\hat{\theta}^*_{iR})} \hat{\theta}^*_{iR} \quad (22)$$

<u>วิธีที่ 3</u> เป็นการประมาณ θ^* ที่เหมาะสมที่สุดระหว่างการใช้ดุลพินิจในการตัดสินความน่าจะเป็นในการตอบข้อสอบได้ถูก (M_{iR}) กับโค้งลักษณะข้อสอบ มี 3 ขั้นตอน คือ

1. คำนวณหาค่าเฉลี่ยของความน่าจะเป็นในการตอบข้อ i ถูก ที่ตัดสินโดยผู้ตัดสินทั้งหมด (M_{iR})

2. คำนวณหาความแปรปรวนของค่าเฉลี่ยความน่าจะเป็นในการตอบข้อ i ถูก ที่ตัดสินโดยผู้ตัดสินทั้งหมด $\sigma^2_i(M_{iR})$

3. คำนวณหา $\hat{\theta}^*$ ซึ่งมีผลบวกของความเบี่ยงเบนกำลังสองระหว่าง M_{iR} และ $P_i(\hat{\theta}^*)$ มีค่าน้อยที่สุด จากสมการ

(23)

แต่ $\theta_{ir} \rightarrow \theta^*$ และสมการที่ 18

$$\sum_i \frac{(M_{iR} - P_i(\hat{\theta}^*))^2}{\sigma_i^2(M_{iR})} \simeq \sum_i \frac{1}{\sigma_i^2(M_{iR})} \left[P_i'(\hat{\theta}^*_{iR})(\hat{\theta}^*_{iR} - \hat{\theta}^*)\right]^2$$

$$\simeq \sum_i \frac{P_i'^2(\hat{\theta}^*_{iR})}{\sigma_i^2(M_{iR})} (\hat{\theta}_{iR} - \hat{\theta}^*)^2 \quad (24)$$

กำหนดให้สมการที่ (24) เท่ากับ 0

$$\sum_i \frac{P_i'^2(\hat{\theta}^*_{iR})}{\sigma_i^2(M_{iR})} (\hat{\theta}^*_{iR} - \hat{\theta}^*)^2 \simeq 0 \quad (25)$$

$$\hat{\theta}^* \simeq \frac{1}{\sum_i P_i'^2(\hat{\theta}^*_{iR})/\sigma_i^2(M_{iR})} \sum_i \frac{P_i'^2(\hat{\theta}^*_{iR})}{\sigma_i^2(M_{iR})} \hat{\theta}^*_{iR} \quad (26)$$

จากสมการที่ (8)

$$\hat{\theta}^* \simeq \frac{1}{\sum_i [1/\sigma_i^2(\hat{\theta}^*_{iR})]} \sum_i \frac{1}{\sigma_i^2(\hat{\theta}^*_{iR})} \hat{\theta}^*_{iR} \quad (27)$$

สรุปวิธีการประมาณ θ^* ของทั้ง 3 วิธี จะเห็นว่าวิธีที่ 1 และ 3 จะอาศัยผลการสอบของข้อสอบแต่ละข้อประกอบกับการใช้ดุลพินิจของผู้เชี่ยวชาญ แต่สำหรับวิธีที่ 2 จะหาคะแนนจุดตัดจากดุลพินิจของผู้เชี่ยวชาญก่อน แล้วจึงอาศัยโค้งลักษณะแบบสอบ แปลงคะแนนจุดตัดในสเกลของคะแนนโดเมนให้อยู่บนสเกลของความสามารถ ซึ่งเป็นวิธีที่นิยมกัน สำหรับในการวิจัยครั้งนี้ ผู้วิจัยใช้วิธีที่ 2 เพราะต้องการคะแนนจุดตัดจากการตัดสินของผู้เชี่ยวชาญแล้วจึงแปลงให้เป็นระดับความสามารถ θ^*

ตอนที่ 3 งานวิจัยที่เกี่ยวข้อง

ปัญหาของวิธีกำหนดจุดตัดที่อาศัยการตัดสินของผู้เชี่ยวชาญสามารถแบ่งออกได้เป็น 3 ข้อ คือ

1. มโนทัศน์เกี่ยวกับความสามารถขั้นต่ำสุด
2. ความคงที่ของการตัดสิน

นักการศึกษาได้แสดงมโนทัศน์เกี่ยวกับความสามารถขั้นต่ำสุดว่าหมายถึง ระดับของความชำนาญ หรือความรู้ที่ต่ำที่สุดที่บุคคลยังสามารถทำสิ่งหนึ่งสิ่งใดได้ ซึ่งอัลเฟรด การ์วิน (Alfred garvin, cited by Glass 1978 : 251) การกำหนดเกณฑ์ของความสามารถในทางการศึกษานั้นทำได้ยาก หรืออาจทำไม่ได้เลย

ความคงที่ของการตัดสิน

ปัญหาเกี่ยวกับความคงที่ของการตัดสินมีงานวิจัยที่ศึกษาเกี่ยวกับปัญหาเรื่องนี้ คือ เมสคาคัส และเวบสเตอร์ (Meskauskas and Webster, cited by Glass 1978 : 248) ได้ทำการศึกษาโดยให้ผู้ตัดสิน 6 คน ใช้วิธีนีเดลสกีกำหนดจุดตัดในรูปของเปอร์เซนต์ของข้อสอบที่ตอบถูก ผลการศึกษาพบว่า คะแนนจุดตัดกระจายอยู่ในช่วง 36% ถึง 80% แล้วสรุปว่าการตัดสินคะแนนจุดตัดไม่มีความคงที่ นอกจากนี้ยังมีงานวิจัยที่เปรียบเทียบคะแนนจุดตัดระหว่าง วิธีนีเดลสกี วิธีอีเบล และวิธีแองกอฟ ดังต่อไปนี้

แอนดรู และเฮก (Andrew & Hecht, cited by Glass 1978 : 249) เปรียบเทียบวิธีนีเดลสกี และวิธีอีเบล โดยให้ผู้เชี่ยวชาญ 8 คน ตัดสินคะแนนจุดตัดของแบบสอบที่ขนานกัน 2 ฉบับ ฉบับแรกใช้วิธีนีเดลสกี ฉบับที่สองใช้วิธีอีเบล ผลการวิจัยพบว่าจุดตัดที่ได้จากการตัดสินด้วยวิธีอีเบลเท่ากับ 68% จะมีผู้สอบผ่านประมาณ 50% ส่วนจุดตัดที่ได้จากการตัดสินด้วยวิธีนีเดลสกีเท่ากับ 49% จะมีผู้สอบผ่านประมาณ 95% จุดตัดที่ได้จากทั้ง 2 วิธีนี้แตกต่างกันอย่างมีนัยสำคัญ ทั้งที่ใช้แนวคิดเกี่ยวกับความสามารถขั้นต่ำสุดอย่างเดียวกัน และไม่มีเหตุผลสนับสนุนอย่างเพียงพอที่จะบอกได้ว่าวิธีใดดีกว่า

สกากัน และคลิง (Kkakun & Kling 1980 : 229-235) เปรียบเทียบวิธีนีเดลสกี กับวิธีที่ประยุกต์จากวิธีอีเบล 2 วิธี ในการกำหนดคะแนนจุดตัดของแบบสอบวิชาการผ่าตัดทั่วไป ซึ่งมีข้อสอบจำนวน 194 ข้อ ผลการวิจัยพบว่าคะแนนจุดตัดที่ได้จากการตัดสินด้วยวิธีนีเดลสกี และวิธีที่ประยุกต์จากวิธีอีเบล 2 วิธี มีความแตกต่างกัน 3% และ 5% ตามลำดับ และจะมีผู้สอบตก ถ้าใช้วิธีอีเบล มากกว่าวิธีนีเดลสกี ประมาณ 23%

เบรนแนน และล็อกวูด (Brennan and Lockwood 1980 : 219-240) ได้ศึกษาเปรียบเทียบการกำหนดจุดตัดด้วยวิธีแองกอฟกับวิธีนีเดลสกี โดยใช้หฤษฎีการสรุปอ้างอิง (Generalizability Theory) การดำเนินการวิจัยใช้กลุ่มตัวอย่างเป็นผู้ตัดสิน 5 คน ตัดสินข้อสอบ 126 ข้อ ด้วยวิธีแองกอฟ และวิธีนีเดลสกี ผลการวิจัยพบว่า

1. วิธีแองกอฟ และวิธีนีเดลสกี จะให้คะแนนจุดตัดที่แตกต่างกันอย่างมีนัยสำคัญ และความเบี่ยงเบนมาตรฐานของคะแนนจุดตัดของวิธีนีเดลสกี จะเป็น 2 เท่า ของวิธีแองกอฟ

2. เมื่อทำการวิเคราะห์ความแปรปรวน (ANOVA) แบบ 3 ทาง คือ วิธีกำหนดคะแนนจุดตัด 2 วิธี X ผู้ตัดสิน 5 คน X -hvlv[126 ข้อ พบว่า ส่วนเบี่ยงเบนมาตรฐานของคะแนนจุดตัดจะใกล้เคียงกับส่วนเบี่ยงเบนมาตรฐานของคะแนนจุดตัดของการตัดสิน โดยวิธีแองกอฟ

3. การใช้ผู้ตัดสินกลุ่มเดียวกันทำการตัดสินข้อสอบทั้งวิธีแองกอฟ และวิธีนีเดลสกี พบว่า

3.1 วิธีแองกอฟ กับวิธีนีเดลสกีไม่สัมพันธ์กัน ($r = -.052$)

3.2 คะแนนจุดตัดของวิธีแองกอฟสูงกว่าของวิธีนีเดลสกี

3.3 ส่วนเบี่ยงเบนมาตรฐานของคะแนนจุดตัดของวิธีแองกอฟต่ำกว่าของวิธีนีเดลสกี

ในการกำหนดจุดตัดโดยอาศัยการตัดสินของผู้เชี่ยวชาญนี้ เชปปาร์ด (Shepard 1980) แนะนำว่าควรจะให้สารสนเทศประกอบการตัดสินของผู้เชี่ยวชาญ เพื่อลดความคลาดเคลื่อนของการตัดสิน ซึ่งมีงานวิจัยที่ศึกษาเกี่ยวกับเรื่องนี้ ดังนี้

ครอส และคณะ (Cross, et al 1984 : 113-129) ได้ศึกษาหาวิธีกำหนดคะแนนจุดตัด ที่เหมาะสมให้กับแบบสอบของ 'The National Teacher Examinations' (NTE) วิชาคณิตศาสตร์ และการศึกษาเบื้องต้นโดยให้ผู้เชี่ยวชาญ 2 กลุ่ม ๆ ละ 15 คน ตัดสินเกณฑ์มาตรฐาน กลุ่มละวิชา ในแต่ละวิชาจะใช้วิธีกำหนดคะแนนจุดตัด ทั้ง 3 วิธี คือ วิธีแองกอฟ วิธีนีเดลสกี และวิธีเจเกอร์ โดยแบ่งผู้เชี่ยวชาญออกเป็น 3 กลุ่ม ให้ตัดสินคะแนนจุดตัดกลุ่มละวิธี ซึ่งมีขั้นตอนในการดำเนินงาน แบ่งออกเป็น 3 รอบ คือ

1. ในรอบแรก ผู้ตัดสินจะได้รับข้อสอบเฉพาะข้อคี่มาพิจารณา โดยไม่มีสารสนเทศใด ๆ ประกอบ

2. ในรอบที่สอง ผู้ตัดสินจะได้รับข้อสอบเฉพาะข้อคู่มาพิจารณา แต่จะมีสารสนเทศประกอบ คือ ค่าความยากของข้อสอบแต่ละข้อ

3. ในรอบสุดท้าย ผู้ตัดสินจะได้รับการขอร้องให้พิจารณาข้ออีกครั้งหนึ่ง โดยพิจารณาเฉพาะข้อคี่ และมีสารสนเทศประกอบ คือ ผลการตัดสินเกณฑ์มาตรฐานของผู้ตัดสินคน

อื่น ๆ ในรอบแรก และการแจกแจงความถี่ของคะแนนที่ผู้สอบทำได้จากการนำแบบสอบไปใช้

ผลการศึกษาพบว่า วิธีกำหนดคะแนนจุดตัด ที่แตกต่างกัน และการให้สารสนเทศแก่ผู้ตัดสิน มีผลกระทบต่อคะแนนจุดตัด แต่เนื้อหาวิชาไม่มีผลกระทบต่อคะแนนจุดตัด สำหรับค่าความเที่ยงของการตัดสิน ผู้วิจัยคำนวณหาจากสูตรที่เสนอโดยอีเบล (Ebel formula d) และประมาณความแปรปรวนของการตัดสินจากสูตรที่เสนอโดยเบรนนัน และล็อกวู้ด (Brennan and Lockwood) จากการวิเคราะห์หาค่าความเที่ยงและความแปรปรวนของการตัดสินนี้ ทำให้สรุปได้ว่า การใช้วิธีแองกอฟโดยให้สารสนเทศประกอบแก่ผู้ตัดสินจะสามารถกำหนดคะแนนจุดตัด ที่มีความเที่ยงสูงกว่าวิธีอื่น

นอร์ซินี และคณะ (Norcini, et al. 1988 : 57-65) ได้ศึกษาถึงผลของการให้สารสนเทศเกี่ยวกับค่าความยากของข้อสอบ ต่อการกำหนดคะแนนจุดตัดด้วยวิธีแองกอฟ ดำเนินการวิจัยโดยใช้ผู้ตัดสิน 6 คน ตัดสินข้อสอบ 48 ข้อ แบ่งออกเป็น 2 รอบ ในรอบแรกผู้ตัดสินจะตัดสินคะแนนจุดตัดโดยไม่ทราบค่าความยากของข้อสอบ ในรอบที่สองผู้ตัดสินจะได้รับการขอร้องให้ตัดสินคะแนนจุดตัดซ้ำอีกครั้งหนึ่ง โดยมีสารสนเทศเกี่ยวกับค่าความยากของข้อสอบประกอบ ผลการศึกษาพบว่าการให้สารสนเทศเกี่ยวกับค่าความยากของข้อสอบไม่มีผลกระทบต่อคะแนนจุดตัด แต่ช่วยลดความคลาดเคลื่อนในการตัดสินได้

ครอส และคณะ (Cross, et al. 1985 : 137-146) ได้ทำการศึกษานำร่องหาคะแนนจุดตัดของแบบสอบอัตนัยวิชาทักษะการติดต่อสื่อสาร (the communication skills test) ของ the Nationaional Teacher Examinations (NTE) จำนวน 12 ข้อ ทำการศึกษาโดยการเปรียบเทียบคะแนนจุดตัดของผู้ตัดสิน 2 กลุ่ม กลุ่มหนึ่งจะได้รับสารสนเทศเกี่ยวกับคะแนนของข้อสอบอัตนัยทั้ง 12 ข้อ นั้น ซึ่งให้คะแนนโดย the Educational Testing Service (ETS) แล้วให้ตัดสินคะแนนที่น้อยที่สุดที่ยอมรับได้ของข้อสอบแต่ละข้อ ซึ่งเป็นไปได้ตั้งแต่ 1-6 คะแนน ส่วนกลุ่มที่ 2 จะทำการตัดสินคะแนนจุดตัดของแบบสอบฉบับนี้ 2 รอบ คือ รอบแรก ให้ตัดสินคะแนนที่น้อยที่สุดที่ยอมรับได้ของข้อสอบแต่ละข้อ โดยไม่ทราบคะแนนของข้อสอบที่ตรวจโดย ETS รอบที่สอง ดำเนินการตัดสินคะแนนจุดตัดเช่นเดียวกับกลุ่มหนึ่ง ในการศึกษานำร่องนี้ใช้ผู้ตัดสินกลุ่มละ 20 คน ผลการศึกษานำร่องพบว่า

1. คะแนนจุดตัดที่ได้จากการตัดสินของผู้ตัดสินกลุ่มที่ 2 ที่ตัดสินโดยไม่มี และโดยมีสารสนเทศเกี่ยวกับคะแนนของข้อสอบซึ่งตรวจโดย ETS นั้น เท่ากับ 2.99 และ 3.10 ตามลำดับ ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 079 เท่ากัน และเมื่อทดสอบความแตกต่าง ปรากฏว่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติที่ระดับ .05

2. คะแนนจุดตัดของผู้ตัดสินกลุ่มที่ 1 และกลุ่มที่ 2 ในรอบแรก เท่ากับ 3.30 และ 2.99 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ .66 และ .79 และเมื่อทดสอบความแตกต่าง ปรากฏว่า แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติที่ระดับ .05

ในขั้นการวิจัยได้ทำการศึกษาหาคะแนนจุดตัดของแบบสอบอัตนัยชุดใหม่ จำนวน 11 ข้อ ดำเนินการวิจัยเช่นเดียวกับการศึกษานำร่อง ทำการศึกษากับผู้ตัดสิน 2 กลุ่ม จำนวน 72 และ 49 คน ในมลรัฐเวอร์จิเนีย (Verginia) และมลรัฐเคนทักกี (Kentucky) ผู้ตัดสินแต่ละกลุ่มประกอบด้วยครูที่มาจากโรงเรียนรัฐบาล และอาจารย์มหาวิทยาลัย ผลการวิจัยพบว่า

1. คะแนนจุดตัดที่ได้จากการตัดสินของผู้ตัดสินที่ไม่มีสารสนเทศ และที่มีสารสนเทศ เท่ากับ 3.44 และ 4.0 ตามลำดับ ส่วนเบี่ยงเบนมาตรฐานเท่ากับ .80 และ .90 ตามลำดับ และเมื่อทดสอบความแตกต่าง ปรากฏว่าแตกต่างกันอย่างไม่มีนัยสำคัญ

2. คะแนนจุดตัดที่ได้จากการตัดสินของครูในโรงเรียนรัฐบาล และอาจารย์มหาวิทยาลัย แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

นอร์ซินี และคณะ (Norcini, et al. 1987:56-64) ได้ทำการศึกษาเปรียบเทียบคะแนนจุดตัดของวิธีแองกอฟใน 3 ช่วงเวลา คือก่อนการประชุม ระหว่างประชุม และหลังการประชุมผู้เชี่ยวชาญ ใช้ผู้เชี่ยวชาญซึ่งเป็นสมาชิกของ Test development Committee จำนวน 6 คน แบ่งแบบสอบออกเป็น 2 ฉบับ ที่มีความเท่าเทียมกันในเรื่องค่าความยาก ค่าอำนาจจำแนก และส่วนเบี่ยงเบนมาตรฐาน วิธีดำเนินการในระยะเวลาก่อนการประชุม 1 เดือน ให้ผู้เชี่ยวชาญประมาณความน่าจะเป็นในการตอบข้อสอบถูกของแบบสอบชุดแรก ในระหว่างการประชุมนำแบบสอบชุดแรกมาประเมินอีกครั้งหนึ่งทีละข้อ ผู้เชี่ยวชาญที่ประมาณความน่าจะเป็นแต่ละข้อมีค่าสูงสุด หรือต่ำสุด ต้องให้เหตุผลแก่ที่ประชุม หลังจากนั้นทุกคนมีสิทธิปรับความน่าจะเป็นใหม่ ทำจนครบทุกข้อ และแก้ไขจนพอใจ หลังจากนั้นผู้เชี่ยวชาญแต่ละคนจะได้ดูผลการประมาณความน่าจะเป็นซึ่งทำก่อนการประชุม แล้วหลังจากการประชุม 1 เดือน ก็จะส่งแบบสอบชุดที่สองไปให้ผู้เชี่ยวชาญชุดเดิมประมาณความน่าจะเป็นที่จะตอบข้อสอบถูก แล้วส่งกลับไปยัง The American Bord of Internal Medicine (ABIM) ทำการวิเคราะห์ความคงเส้นคงวาของการตัดสินใจของผู้เชี่ยวชาญด้วยทฤษฎีสรุปอ้างอิง (General liability Theory) ผลการวิจัยพบว่า คะแนนจุดตัดที่ได้จากก่อนการประชุม ระหว่างการประชุม และหลังการประชุมไม่แตกต่างกันอย่างมีนัยสำคัญ ทางสถิติที่ระดับ .05 แต่การประมาณความน่าจะเป็นของการตอบข้อสอบถูกของผู้เชี่ยวชาญที่ได้จากก่อนการประชุม มีความคงเส้นคงวา (consistent) น้อยกว่าระหว่างการประชุมและหลังการประชุม แสดงว่าการให้ผู้เชี่ยวชาญได้ประชุมปรึกษากัน จะทำให้การกำหนดคะแนนจุดตัดด้วยวิธีแองกอฟมีประสิทธิภาพขึ้น

แวน เดอร์ ลินเดน (Van der Linden 1982:295-308) ได้ศึกษาความไม่สอดคล้องของการตัดสินใจโดยผู้เชี่ยวชาญในการกำหนดคะแนนจุดตัดตามวิธีแองกอฟ และวิธีนีเดลสกี โดยอาศัยทฤษฎีการตอบข้อสอบ ที่แสดงความสัมพันธ์ระหว่างการตอบข้อสอบ กับความสามารถที่อยู่ภายในตัวผู้สอบ ในรูปของโค้งลักษณะข้อสอบ ($P_i(\Theta)$) ซึ่งความน่าจะเป็นในการตอบข้อสอบได้ถูกของผู้สอบที่มีระดับความสามารถ จะขึ้นอยู่กับโค้งลักษณะข้อสอบนั้น วิธีการศึกษาคือ หาระดับความสามารถ Θ_c ที่เป็นตัวกำหนดความน่าจะเป็นในการตอบข้อสอบได้ถูกของข้อสอบทุกข้อในแบบสอบ ($P_i(\Theta_c)$) และนำไปเปรียบเทียบกับความน่าจะเป็นในการตอบข้อสอบได้ถูกที่ตัดสินโดยผู้เชี่ยวชาญ สำหรับข้อสอบทุกข้อในแบบสอบ ความแตกต่างที่ได้จะแสดงถึงความไม่คงเส้นคงวาของการตัดสิน ความแตกต่างนี้อาจเนื่องมาจากว่า ผู้เชี่ยวชาญไม่ทราบคุณสมบัติของข้อสอบ และทำให้ประมาณความน่าจะเป็นในการตอบข้อสอบได้ถูกไม่ตรงกับฟังชันลักษณะข้อสอบ แสดงว่าผู้เชี่ยวชาญมีข้อมูลไม่เพียงพอในการเดา มีขั้นตอนการศึกษาดังนี้

1. หาค่าพารามิเตอร์ของข้อสอบ โดยใช้ทฤษฎีการตอบข้อสอบวิเคราะห์ข้อสอบ

2. นำข้อสอบในข้อ 1. ไปให้ผู้เชี่ยวชาญประมาณความน่าจะเป็นในการตอบข้อสอบได้ถูก แทนด้วย $P_i^{(s)}$ ตามวิธีกำหนดจุดตัดของแองกอฟ หรือ นีเดลสกี

3. คำนวณหาคะแนนจุดตัด (π_0) ของวิธีแองกอฟ หรือวิธีนีเดลสกี ซึ่งอยู่บนสเกลของคะแนนจริง

4. แปลงคะแนนจุดตัดที่อยู่บนเสกลของคะแนนจริง ให้อยู่บนเสกลของระดับความสามารถ (Θ-scale) โดยอาศัยโค้งลักษณะแบบสอบ จะได้ Θ_c

5. คำนวณหาความน่าจะเป็นในการตอบข้อสอบได้ถูกจากโค้งลักษณะข้อสอบ (P_i) ที่ระดับความสามารถ Θ_c ของข้อสอบทุกข้อ

6. คำนวณหาดัชนีความคงที่ (index of consistency : C) ของการตัดสินโดยผู้เชี่ยวชาญ ดังนี้

6.1 หาความแตกต่างระหว่าง $P_i^{(s)}$ และ P_i สำหรับข้อสอบทุกข้อ

$$e_i = P_i^{(s)} - P_i$$

6.2 หาค่าเฉลี่ยความคลาดเคลื่อนสัมบูรณ์ของการตัดสินโดยผู้เชี่ยวชาญของข้อสอบทั้งฉบับ

$$E = \left| P_i^{(s)} - P_i \right| / n$$

เมื่อ n คือ จำนวนข้อสอบ

6.3 หาดัชนีความคงที่ (C) ของการตัดสินโดยผู้เชี่ยวชาญ

$$C = \frac{M-E}{M}$$

เมื่อ $M = \sum_{i=1}^{n} e_i^{(u)}/n$

$e_i^{(u)} = \max\{p_i, 1-p_i\}$

ถ้า C เข้าใกล้ 0 แสดงว่าการตัดสินคะแนนจุดตัดโดยผู้เชี่ยวชาญไม่มีความคงที่

ผลการวิจัยพบว่า การให้ผู้เชี่ยวชาญกำหนดจุดตัดโดยวิธีของแองกอฟ จะทำให้การตัดสินมีดัชนีความคงที่ (0.77) สูงกว่าวิธีของนีเดลสกี (0.68)

อาราสมิธ (Arrasmith 1987:3405-A) ทำการศึกษาเกี่ยวกับความคลาดเคลื่อนในการกำหนดคะแนนจุดตัดของวิธีแองกอฟ และวิธีการเปรียบเทียบระหว่างกลุ่ม (Contrasting-groups) โดยอาศัยทฤษฎีสรุปอ้างอิง (generalizability theory), ทฤษฎีการตอบสนองข้อสอบ (item response theory) และ (bootstrap estimation) เพื่อเป็นแนวทางในการเลือกวิธีกำหนดจุดตัดที่เหมาะสม ผลการวิจัยพบว่า สำหรับวิธีแองกอฟ การหาค่าดัชนีความคงที่โดยทฤษฎีการตอบสนองข้อสอบ (IRT index of consistency) และการวิเคราะห์ความแตกต่างของความน่าจะเป็นในการตอบข้อสอบได้ถูกระหว่างการประมาณ โดยใช้ดุลพินิจของผู้เชี่ยวชาญ กับการประมาณ $P_i(\Theta)$ สามารถแสดงถึงความไม่สอดคล้องของการตัดสินโดยผู้เชี่ยวชาญ นอกจากนี้ทฤษฎีสรุปอ้างอิงและ bootstrap estimate ก็สามารถประมาณความคลาดเคลื่อนรวม ๆ ในการตัดสินของผู้เชี่ยวชาญได้ สำหรับวิธีการเปรียบเทียบระหว่างกลุ่มสามารถใช้วิธี bootstrap ประมาณความคลาดเคลื่อนของคะแนนจุดตัดได้

สำหรับงานวิจัยเกี่ยวกับแบบสอบอิงเกณฑ์ในประเทศไทยที่ศึกษาเกี่ยวกับวิธีการกำหนดคะแนนจุดตัดที่ใช้ดุลพินิจของผู้เชี่ยวชาญตัดสิน คือ งานวิจัยของ ประภา แก่นเพ็ม (2523) ที่ศึกษาความถูกต้องในการกำหนดคะแนนจุดตัดของวิธีกิลเบอร์ท (Guilbert) และวิธีของมหาวิทยาลัยอิลินอยส์ ซึ่งดัดแปลงมาจากวิธีของนีเดลสกี โดยการเปรียบเทียบจุดตัดที่ได้จากการตัดสินของครู กับจุดตัดที่ได้จากผลการสอบแบบสอบวิชาคณิตศาสตร์ จำนวน 40 ข้อ ของนักเรียนที่อยู่ระดับคาบเส้นจำนวน 82 คน ผลการวิจัยพบว่า จุดตัดที่ได้จากวิธีกิลเบอร์ทเท่ากับ 38% และจุดตัดที่ได้จากวิธีของมหาวิทยาลัยอิลลินอยส์ เท่ากับ 32% นำจุดตัดจากทั้งสองวิธีเปรียบเทียบกับจุดตัดที่ได้จากการทดสอบนักเรียนคาบเส้น ปรากฏว่ามีความแตกต่างกันน้อยมากไม่เกิน 10%

จากการศึกษาเอกสารและงานวิจัยทั้งหมดดังกล่าวแสดงให้เห็นว่าการกำหนดคะแนนจุดตัดยังเป็นปัญหาที่สำคัญของการทดสอบแบบอิงเกณฑ์ แต่วิธีที่นิยมใช้กันมากคือวิธีกำหนดจุดตัดโดยอาศัยการตัดสินของผู้เชี่ยวชาญ ซึ่งก็มีปัญหาเกี่ยวกับกับความคงที่ของการตัดสิน ความคลาดเคลื่อนที่เกิดขึ้นนี้อาจเนื่องมาจากผู้เชี่ยวชาญไม่มีสารสนเทศประกอบการพิจารณากำหนดคะแนนจุดตัด ดังนั้นผู้วิจัยจึงสนใจเปรียบเทียบคะแนนจุดตัด และดัชนีความคงที่ของการตัดสินคะแนนจุดตัดซึ่งตัดสินโดยผู้เชี่ยวชาญที่ไม่มีสารสนเทศและผู้เชี่ยวชาญที่มีสารสนเทศประกอบการตัดสิน

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย