

วรรณคดีที่เกี่ยวกับของ

ในการวัดผลสัมฤทธิ์ เกลเซอร์ (Glaser, 1963) กล่าวว่า
มโนทัศน์ที่เน้นในการวัดผลสัมฤทธิ์ คือ ความคิดเกี่ยวกับความต่อเนื่องของความรู้
ที่ได้เรียนมา ซึ่งจะเรียงลำดับจากการไม่มีเลยไปจนถึงการมีอย่างสมบูรณ์ ระดับผลสัมฤทธิ์
ของบุคคลจะตกอยู่ที่จุดใดจุดหนึ่ง ในช่วงของความต่อเนื่องนี้ซึ่งจะชี้ได้โดยพฤติกรรมที่แสดงออก
ในระหว่างการสอบ ระดับผลสัมฤทธิ์ที่จะแสดงว่าเป็นการปฏิบัติที่น่าพอใจหรือไม่จะประเมินได้
ด้วยการวัดแบบอิงเกณฑ์... เกณฑ์ที่ใช้ในลักษณะนี้ไม่จำเป็นต้องอ้างอิงถึงพฤติกรรมสุดท้ายหลังจาก
จากเรียนจบแล้ว ระดับของเกณฑ์สามารถกำหนดที่จุดใด ๆ ของการสอบก็ได้ถ้าจำเป็น เพื่อให้
ให้ได้ข้อมูลเกี่ยวกับการปฏิบัติของแต่ละคนอย่างพอเพียง... คะแนนของนักเรียนที่ได้จากวัดแบบ
อิงเกณฑ์จะให้ข้อมูลที่แสดงอย่างชัดเจนว่า แต่ละคนสามารถหรือไม่สามารถทำอะไรได้บ้าง¹

จากแนวคิดเกี่ยวกับการวัดแบบอิงเกณฑ์ของเกลเซอร์ แสดงว่าจะต้องมีจุดตัด
ระหว่างความสามารถและความไม่สามารถของบุคคล คือมีจุดหนึ่งซึ่งความสามารถจะเปลี่ยน
เป็นไม่สามารถ คะแนนที่จุดนี้จะใช้ในการตัดสินให้ผู้สอบเป็นนุรอนรุ (masters)

¹ R. Glaser, "Instructional Technology and the Measurement of Learning Outcomes, American Psychologist 18 (1963) : 519 - 520.

และไมรอบรู (mcn maters) หรือเป็นผู้สอบผ่านและผู้ลอบคต ในปี 1963 เกลเซอร์ ได้เขียนข้อความที่มีความหมายถึงคะแนนจุดตัดว่า "เราต้องการกำหนดระดับต่ำสุดของการปฏิบัติ (minimum level of performance) ที่อธิบายความสามารถปริมาณน้อยที่สุดที่คาดว่านักเรียนจะบรรลุหลังจากเรียนจบแล้ว หรือที่เขาต้องการเพื่อที่จะเรียนวิชาต่อไป" ¹ เมเจอร์ (Mager ,1962) ได้ใช้คำว่า "มาตรฐานการปฏิบัติ" ในการเขียนเกี่ยวกับภาคเตรียมวัตถุประสงค์เชิงพฤติกรรมว่า "ถ้าเราสามารถกำหนดขั้นต่ำสุดของการปฏิบัติที่ยอมรับ (minimum acceptable performance) สำหรับแต่ละวัตถุประสงค์ได้ เราจะมีมาตรฐานการปฏิบัติ (performance standard) ที่จะใช้ตรวจสอบว่าโปรแกรมการสอนประสบผลสำเร็จตามความมุ่งหมายหรือไม่" ² ในปี 1969 โฟแพมและฮูเชก (Popham and Husek) ได้เขียนเกี่ยวกับการวัดแบบอิงเกณฑ์ และใช้คำว่า มาตรฐานการปฏิบัติเช่นเดียวกับเมเจอร์ ในการอธิบายความหมายของการวัดแบบอิงเกณฑ์ว่า "การวัดอิงเกณฑ์เป็นสิ่งที่ใช้ค้นหาสถานะของแต่ละบุคคล โดยอาศัยเกณฑ์บางอย่างคือ มาตรฐานการปฏิบัติ"³ ดังนั้น มาตรฐานการปฏิบัติตามความหมายของโฟแพม และฮูเชก ก็คือคะแนนจุดตัดนั่นเอง คำว่ามาตรฐานการปฏิบัตินี้ สมศักดิ์ สินธุระเวชญ์ ให้ความหมายว่า

¹ Ibid. p 520.

² Robert F.Mager, Preparing Instructional Objectives.(Palo Alto, Ca.Feardon Publishers,1962), p.44 Cited by Gene V.Glass, "Standard and Criteria" Journal of Educational Measusement 15 (1978):240.

³ W.J..Popham and T.R. Husek, " Implication of Criterion - Referenced Measurement, "Journal of Educational Measurement 6 (1969):2.

"หมายถึงภาระกิจทั้งหมดที่ผู้เรียนควรปฏิบัติได้เช่น ในการสอนให้นักเรียนบวกเลข 2 จำนวน ซึ่งมีผลลัพธ์ไม่เกิน 100 ภาระกิจทั้งหมดคือเลขทุกคู่ที่บวกกันแล้วผลลัพธ์ไม่เกิน 100"¹ ความหมายนี้ต่างจากความหมายของ โฟแฟม และ ฮูเซ็ค และพิจารณาแล้วน่าจะเป็นความหมายของคำว่า โคเมนมากกว่า มาร์ตูซา (Martuza) ใช้คำว่ามาตรฐานการปฏิบัติ ในความหมายที่เป็นคะแนนจุดตัดอย่างชัดเจนโดยให้ความหมายว่า "มาตรฐานการปฏิบัติ หรือคะแนนจุดตัด (cut score) คือคะแนนโคเมน ซึ่งผู้สอบจะต้องทำได้ ในการเป็นผู้บรรลุวัตถุประสงค์หรือเป็นผู้รอบรู้ในเนื้อหาของโคเมนที่เฉพาะ"² เกลเซอร์และนิตโก (Glaser and Nitko, 1971) ได้กล่าวถึงมาตรฐานการปฏิบัติในการวัดแบบอิงเกณฑ์ว่า " มาตรฐานการปฏิบัติสามารถบ่งชี้ได้โดยกำหนดโคเมนของงานบางอย่างที่นักเรียนควรทำได้ ตัวอย่างที่เป็นตัวแทนของงานจากโคเมนนี้ จะถูกจัดเข้าในแบบสอบ นอกจากการวัดที่กระทำจะแสดงถึงการปฏิบัติของแต่ละคนโดยเทียบกับโคเมนที่กำหนดไว้ "³ กรอนลันด์ (Gronlund) ใช้คำว่า มาตรฐาน (standard) ในความหมายว่า "เป็นคะแนนที่คาดหวังว่านักเรียนจะทำได้ถึง และนักเรียนที่ทำคะแนนได้ไม่ถึง จะได้รับการฝึกฝนจน

¹ สมศักดิ์ สินธุระเวชัญ , "การประเมินผลแบบอิงกลุ่มและอิงเกณฑ์" วารสาร การวิจัยทางการศึกษา ๕ (มิถุนายน 2521) หน้า 77 .

² Victor R.Martuza, Applying Norm - Referenced and criterion - Referenced Measurement in Education (Boston: Allyn and Bacon, Inc, 1977) pp. 269 - 270.

³ Ronald K.Hambleton, " Testing and Decision. . . , " p.373. Citing R.Glaser and A.J.Nitko, "Measurement in Learning and Instruction," In:R.L. Thorndike (Ed) Educational Measurement (Washington, D.C.:American Council on Education, 1971), p.653.

สามารถหาคะแนนใดอย่างสมบูรณ์หรือเกือบสมบูรณ์" 1

นอกจากคำว่าเกณฑ์ มาตรฐานการปฏิบัติ มาตรฐาน และคะแนนจุดตัดแล้ว ยังมีผู้ใจค้ำอื่น ๆ อีก เช่น แอมเบิลตันและโนวิก (1973) ใช้คำว่าคะแนนจุดตัดหรือคะแนนระดับความรอบรู้ (mastery level score) ดึงข้อความที่กล่าวมาแล้วในบทที่ 1 แม้ว่าค่าที่ใช้จะแตกต่างกันแต่ส่วนใหญ่แล้วความหมายที่มุ่งใช้เหมือนกันคือ เป็นคะแนนที่ใช้ในการเปรียบเทียบกับคะแนนที่ผู้สอบทำได้เพื่อจัดประเภทผู้สอบออกเป็นผู้รอบรู้และผู้ไม่รอบรู้ สำหรับกรณีที่มีระดับการรอบรู้มากกว่า 2 ระดับ จะมีคะแนนจุดตัดมากกว่า 1 ค่า

วิธีการกำหนดคะแนนจุดตัด

จากการที่นักการศึกษาหันมาสนใจการวัดแบบอิงเกณฑ์ ทำให้มีผู้ทำการศึกษาค้นคว้าและหาวิธีการที่จะกำหนดคะแนนจุดตัดของแบบสอบอิงเกณฑ์กันอย่างแพร่หลายในการกำหนดคะแนนจุดตัดนี้ ครอนบาช (Cronbach) เสนอแนะว่า "ไม่ควรกำหนดคะแนนจุดตัดจากคะแนนดิบโดยตรง...การกำหนดคะแนนจุดตัดต้องการการพิจารณาคุณค่า" 2 วิธีการกำหนดคะแนนจุดตัดมีหลายวิธี ~~แต่ยังมีผู้หาว่าวิธีใดเป็นวิธีที่ดีและเหมาะสมที่สุด~~ ~~ยังไม่มีใครสามารถแก้มันจนได้~~ เดี๋ยวจะมีนักการศึกษาบางคนเสนอให้เลิกใช้คะแนนจุดตัดเนื่องจากมีปัญหาในการกำหนดค่าของคะแนนนี้ และเสนอให้ใช้การเปลี่ยนแปลงของการ

1

Norman E.Gronlund; Measurement and Evaluation in Teaching.

2 d.ed (New York : Macmillan Company, 1971), p.17.

2

Lee J.Cronbach, Essentials of Psychological Testing, 3 d.ed.

(New York : Harper Row, 1960) p.422.

ปฏิบัติของผู้เข้าสอบแทน แต่ส่วนใหญ่ยังคงเห็นว่าคะแนนจุดตัดเป็นสิ่งจำเป็นสำหรับการวัดแบบอิงเกณฑ์ วิธีการกำหนดคะแนนจุดตัดที่มีผู้เสนอขึ้นใช้นั้น กลาส (Glass, 1978) ได้สรุปและรวบรวมได้เป็น 6 วิธี ดังนี้¹

1. ใช้ผลที่ได้จากการปฏิบัติของคนอื่น ๆ เป็นเกณฑ์ (Performance of Others as a Criterion) ระเบียบเกณฑ์สร้างขึ้นโดยอ้างอิงพารามิเตอร์ของประชากรผู้เข้าสอบ ดังนั้น เกณฑ์หรือระดับความรู้ในแบบสอบ จึงอาจกำหนดขึ้นโดยใช้ค่ามัธยฐาน (median) ของคะแนนจากแบบสอบ เช่น การสอบเพื่อวัดความสามารถของนักศึกษาชั้นปีที่ 4 ที่ อริโซนา ระดับของเกณฑ์ตั้งขึ้นโดยใช้เกรดเทียบเท่าเกรด 9 ซึ่งได้จากการใช้แบบสอบมาตรฐานวัดผลสัมฤทธิ์ ในการอ่านเกรดเทียบเท่าเกรด 9 นี้ ได้มาจากค่ามัธยฐานของคะแนนของเด็กเกรด 9 การกำหนดเกณฑ์ด้วยวิธีนี้เป็นการอ้างอิงถึงนอร์ม (Norm) อย่างแท้จริง นักทฤษฎีเกี่ยวกับแบบสอบอิงเกณฑ์จำนวนมากจึงเห็นว่าไม่เหมาะสมกับการวัดแบบอิงเกณฑ์

2. นับถอยหลังจาก 100% (Counting Backwards from 100 %) ในการสร้างข้อสอบจะสร้างให้สอดคล้องกับวัตถุประสงค์ เนื่องจากวัตถุประสงค์มีความสำคัญผู้เขียนข้อสอบจึงเห็นพ้องกันว่า ทุกคนควรสามารถตอบปัญหาในแบบสอบที่วัดวัตถุประสงค์นั้น ๆ ได้ นั่นคือระดับการปฏิบัติที่พึงปรารถนาจะเป็น 100 % แต่จากเหตุผลและประสบการณ์พบว่าจะให้ใครโดยสมบูรณ์นั้นยอมเป็นไปไม่ได้ จึงยอมให้ลดระดับลงมาสำหรับความไม่สมบูรณ์ของสมองและความคลาดเคลื่อนต่าง ๆ บางแห่งอาจลงมา 5 % บางแห่งลดลง 20 % หรือมากกว่า เกณฑ์ที่ตั้งขึ้นในลักษณะนี้จะเกิดปัญหาในการกำหนดกันตามอำเภอใจ เช่น ถ้าผู้เชี่ยวชาญ A กำหนดคะแนนเกณฑ์เป็น 95 % ผู้เชี่ยวชาญ B กำหนดเป็น 90 % เปอร์เซนต์ของผู้เข้าสอบที่จะบรรลุระดับเกณฑ์ 2 ระดับนี้จะแตกต่างกันมาก

1

Gene V. Glass, "Standard and Criteria," Journal of Educational Measurement 15 (1978) : 243 - 257.

3. การเพิ่มคะแนนเกณฑ์อื่น ๆ (Bootstrapping on other Criterion Scores) วิธีใส่คะแนนเกณฑ์ในแบบสอบพิจารณาโดยอาศัยการระบุจากภายนอก กำหนดว่า "สำเร็จ" หรือ "รอบรู้" อย่างชัดเจน เช่น อาจใช้เกณฑ์ความได้เปรียบภาคนี้บัตร ในอาชีพต่าง ๆ เช่น คัดคน นายหนา นักสถิติ เป็นต้น ถือเป็นผู้มีความสามารถ ส่วนผู้ที่ไม่ได้รับภาคนี้บัตรถือเป็นผู้ไม่มีความสามารถ พิจารณาการกระจายของคะแนนที่ได้จากการทำแบบสอบของคนเหล่านี้ แล้วกำหนดคะแนนเกณฑ์เพื่อแยกผู้มีความสามารถออกจากผู้ไม่มีความสามารถ

เบอร์ค¹ประยุกต์วิธีนี้มาใช้โดยใช้เกณฑ์ภายนอกคือ ได้รับการสอนหรือไม่ได้รับการสอนตามวัตถุประสงค์ที่จะวัดมาแบ่งนักเรียนออกเป็น 2 กลุ่มคือ กลุ่มที่ได้รับการสอน ให้เป็นพวกรอบรู้ (Master :M) กลุ่มที่ไม่ได้รับการสอน ให้เป็นพวกไม่รอบรู้ (Nonmaster : N) หรืออาจใช้นักเรียนกลุ่มเดียว เป็นกลุ่มก่อนเรียนและกลุ่มหลังเรียนแทนก็ได้ หลังจากให้นักเรียนทั้ง 2 กลุ่มทำแบบสอบแล้ว จะได้คะแนนซึ่งมีพิสัย ตั้งแต่ 0 - i (เมื่อ i เป็นจำนวนข้อกระทง) แบ่งนักเรียนที่ได้คะแนนต่าง ๆ เหล่านี้ออกเป็น 2 กลุ่มโดยใช้คะแนนที่ตกลงกำหนดขึ้นมาเป็นคะแนนจุดตัด นักเรียนที่ได้คะแนนเท่ากับหรือสูงกว่าคะแนนจุดตัดจัดเป็นพวกรอบรู้ โดยการทำนาย (Predicted Master : PM) ส่วนนักเรียนที่ได้คะแนนต่ำกว่าคะแนนจุดตัดเป็นพวกไม่รอบรู้ โดยการทำนาย (Predicted Nonmaster : PN) จากการแบ่งโดยใช้เกณฑ์ภายนอกและคะแนนจุดตัดที่กำหนดขึ้น จะได้นักเรียน 4 พวกคือ

1. พวกรอบรู้จริง (True Master:TM) คือนักเรียนที่ได้รับการสอนซึ่งได้คะแนนเท่ากับหรือมากกว่าคะแนนจุดตัด

1

Ronald A. Berk, " Determination of Optimal Cutting' Score in Criterion - Referenced Measurement, " Journal of Experimental Education 45 (1976) : 4 - 9.

2. พวกครอบรู้ไม่จริง (False Master :FM) คือนักเรียนที่ไม่ได้รับการสอน ซึ่งได้คะแนนมากกว่าหรือเท่ากับคะแนนจุดตัด

3. พวกไม่ครอบรู้จริง (True Nonmaster:TN)คือนักเรียนที่ไม่ได้รับการสอนซึ่งได้คะแนนต่ำกว่าคะแนนจุดตัด

4. พวกไม่ครอบรู้ไม่จริง (False Nonmaster:FN) คือนักเรียนที่ได้รับการสอน ซึ่งได้คะแนนต่ำกว่าคะแนนจุดตัด

ผลที่เกิดขึ้นสามารถแสดงในรูปตาราง 2×2 ได้ดังนี้

เกณฑ์ภายนอก

	ได้รับการสอน	ไม่ได้รับการสอน
ครอบรู้โดยการทำนาย	พวกครอบรู้จริง (TM)	พวกครอบรู้ไม่จริง (FM)
ไม่ครอบรู้โดยการทำนาย	ไม่ครอบรู้ไม่จริง (FN)	ไม่ครอบรู้จริง (TN)

เมื่อทดลองกำหนดคะแนนจุดตัดหลาย ๆ จุด จะได้จำนวนนักเรียนแต่ละพวก แยกต่างกันตามค่าของคะแนนจุดตัด นำผลที่ได้จากการกำหนดคะแนนจุดตัดต่าง ๆ กัน มาหาค่าความน่าจะเป็น (Probability) ของแต่ละประเภทโดยใช้สูตรดังนี้

$$P (TM) = TM / (I + N)$$

$$P (FM) = FM / (I + N)$$

$$P (TN) = TN / (M + N)$$

$$P (FN) = FN / (M + N)$$

คะแนนจุดตัดที่ เหนือค่าความน่าจะเป็น ในการตัดสินอย่างถูกต้อง คือให้ค่า $P(TH) + P(FH)$ สูงสุดหรือให้ค่าความน่าจะเป็น ในการตัดสินผิด คือให้ค่า $P(FT) + P(FN)$ ต่ำสุดจะเป็นคะแนนจุดตัดที่เหมาะสมสำหรับแบบสอบฉบับนี้

วิธีนี้ กลาส (Glass) กล่าวว่าปัญหาที่อาจเกิดขึ้นคือ ถ้าแบบสอบที่ต้องการ กำหนดคะแนนเกณฑ์ค่าสหสัมพันธ์กับ เกณฑ์ที่ใช้ ในการ ระบุจากภายนอกน้อยกว่าค่าสหสัมพันธ์ อย่างสมบูรณ์ เราจะไม่สามารถกำหนดคะแนน เกณฑ์ใช้ สอดคล้องกับการ ระบุจากภายนอก อย่างสมบูรณ์ได้ คืออาจมีผู้สอบผ่าน เกณฑ์ภายนอกบางคนได้คะแนนน้อยกว่าคะแนน เกณฑ์ที่กำหนดขึ้น

4. การพิจารณาตัดสินจรรยาบรรณความสามารถต่ำสุด (Judging Minimal Competence) วิธีนี้จะใหญ่ เชี่ยวชาญศึกษาแบบสอบหรือข้อกระทงแล้ว เสนอค่าความสามารถ ขั้นต่ำสุดที่ผู้สอบควรทำได้ วิธีนี้มีผู้ประยุกต์มาใช้หลายคน คือ

นีเดิลสกี (Nedelsky) เสนอเทคนิคในการหาคะแนนผ่านต่ำสุด (minimum passing score) ดังนี้ ก่อนนำแบบสอบไปให้นักเรียนทำให้ครูผู้สอนทุกคนพิจารณาแต่ละข้อกระทงว่าคัดเลือกใดที่นักเรียนในระดับต่ำสุดที่จะผ่าน (พวก D student) บอกไว้ว่านิค ระดับการผ่านต่ำสุดของแต่ละข้อกระทง คือส่วนกลับของจำนวนตัว เลือกที่เหลืออยู่ เช่น ในข้อกระทงที่มี 5 ตัวเลือก ถ้ามี 1 ข้อที่นักเรียนระดับต่ำสุดที่จะผ่านรู้ว่านิค ส่วนกลับของจำนวนตัว เลือกที่เหลือคือ $1/4$ นำค่าเศษส่วนนี้จากทุกข้อมารวมกันแล้วหาค่าเฉลี่ยของผลที่ได้จากการพิจารณาของผู้สอนแต่ละคนเป็นค่าเฉลี่ยของคะแนนเดาของนักเรียนที่อยู่ระหว่างเกรด F และ D (F - D student) แทนด้วย M_{FD} เพื่อนำไปหาค่าคะแนนต่ำสุดจากสูตร

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

k เป็นค่าคงที่ ซึ่งกำหนดขึ้นจากการพิจารณาหลาย ๆ ครั้ง
 σ_{FD} เป็นค่าส่วนเบี่ยงเบนมาตรฐานของค่ารวมเศษส่วนจากทุกข้อกระทงของผู้สอนแต่ละคน

ผู้สอนส่วนใหญ่ที่ใช้เทคนิควิธีนี้เห็นว่าคะแนนผ่านค่าสุทธควรทำให้นักเรียนที่อยู่ระหว่างเกรด F และ D ส่วนใหญ่ตอบตก ถ้าให้ k เป็น $-1, 0, 1$ และ 2 จะให้พวกที่อยู่ระหว่างเกรด F และ D สอดตก $16\%, 50\%, 84\%$ และ 98% ตามลำดับ การพิจารณาตัดสินใจเลือกค่า k ครั้งสุดท้าย จะได้อีกหลังจากที่ผู้สอนหลาย ๆ คนเลือกคำตอบที่นักเรียนระดับค่าสุทธจะผ่านรูวาคแล้ว แต่ตามการกำหนดคามาตรฐานสัมบูรณ์ (absolute standard) นั้น ค่า k ควรกำหนดก่อนที่จะคำนวณค่า M_{FD} และก่อนที่จะระบุคะแนนการสอบของนักเรียน

คำตอบที่นักเรียนค่าสุทธจะผ่านรูวาค จะมีทั้งคำตอบที่ผิดอย่างชัดเจนและคำตอบผิดปานกลางซึ่งผู้ที่จะปฏิเสธได้ จะต้องมีความรู้ถึงขั้นนั้นแล้ว ถ้าในแบบสอบมีคำตอบชนิดแรกมาก สามารถแก้ได้โดยการกำหนดค่า k ให้สูง แต่ถาชนิดหลังมากจะกำหนดค่า k ให้ต่ำ ในการเปลี่ยนแปลงค่า k นั้น ไม่ควรเปลี่ยนมากกว่า ± 0.5 จากค่า k ที่กำหนดขึ้นเพื่อทดลองใช้ การตัดสินตามทฤษฎีไม่ค่อยปรากฏว่า k จะสูงถึง 2 สำหรับแบบสอบส่วนมาก ค่า $k=0$ อาจจะทำเกินไป ขอเสนอแนะในการพิจารณาค่า k คือควรเริ่มที่ $k = 1$ ก่อน และช่วงของค่า k ที่จะเปลี่ยนไปจะอยู่ระหว่าง .5 ถึง 1.5

อีเบล (Ebel) มีความเห็นวานิยามของคะแนนที่ตองผ่าน (passing score) ในลักษณะที่เป็นเปอร์เซ็นต์ของคะแนนสอบทั้งหมดอาจมีข้อบกพร่อง เนื่องจากข้อกระทงอาจยากเกินไป ง่ายเกินไป หรืออำนาจจำแนกน้อยกว่าที่ผู้สร้างแบบสอบตั้งใจไว้ การสอบผ่านหรือตกของผู้เข้าสอบอาจเป็นการตัดสินโดยใช้คำถามในแบบสอบมากกว่าใช้ระดับความสามารถของเขา อีเบลจึงเสนอให้ใช้การพิจารณาคูณลักษณะของข้อกระทงในแง่ของความเกี่ยวข้อง (relevance) และความยาก (difficulty) ในแง่ของความเกี่ยวข้อง แบ่งเป็น 4 ประเภทคือ จำเป็น (Essential) สำคัญ (Important) พอจะยอมรับได้ (Acceptable) และยังไม่แน่ใจ (Questionable) ส่วนค่าความยากมี 3 ระดับคือ ง่าย ปานกลาง และยาก และเสนอเปอร์เซ็นต์ ที่คาดหวังซึ่งเป็นปริมาณที่ต่ำกว่าผู้สอบที่มีความสามารถค่าสุทธ ในการที่จะสอบผ่านควรทำได้ โดยเขียนในรูปตาราง

3x4 ดังนี้

ตารางที่ 1 แสดงค่า เบอรเซนคท์คาคหวังในแกละระคัมความบากของแถและประเมท
ความเกี้ยวของของอึเบล

ประเมทของความเกี้ยวของ	ระคัมความบาก		
	งาย	ปานกลาง	บาก
จำเป็น (Essential)	100 %	—	—
สำคัญ (Important)	90 %	70 %	—
ยอมรับได้ (Acceptable)	30 %	60 %	40 %
ไม่แน่ใจ (Questionable)	70 %	50 %	30 %

วิธีนคองนำแบบสอบไปใหญ่เชียวชาญพิจารณาจคัประเมทขอกระทงทุกขอ แลวรวมจำนวน
ขอกระทงในแถละประเมท คณควยไปอรเซนคท์คาคหวัง นำมาหาคาเฉลี่ยเป็นคะแนนคองผ่านใน
รูปรอยละ เช่น นำแบบสอบทมิจำนวนขอกระทง 100 ขอ ไปใหญ่เชียวชาญ 5 ทานพิจารณา จำนวน
ขอกระทงในแถละประเมทแสดงในคอลัมน์ที่ 2 ของตาราง คะแนนคองผ่านคไจจากการหารผลรวม
ของผลคูณในคอลัมน์ที่ 4 ควย 500

ศูนย์วิจัยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 2 แสดงจำนวนข้อกระทงในแต่ละระดับความยากของประเภทของความถี่ของของอ็เบล

ประเภทของข้อกระทง	จำนวนข้อกระทง	ความสำเร็จที่คาดหวัง	จำนวนข้อ	ความสำเร็จ
จำเป็น	94	100	9400	
สำคัญ	ง่าย	90	9540	
	ปานกลาง	70	10710	
ยอมรับได้	ง่าย	80	1920	
	ปานกลาง	60	2940	
	ยาก	40	2080	
ไม่แน่ใจ	ง่าย	70	280	
	ปานกลาง	50	50	
	ยาก	30	210	
	500		37130	

คะแนนที่คงผ่าน = $37130/500 = 74.26\% \approx 74\%$

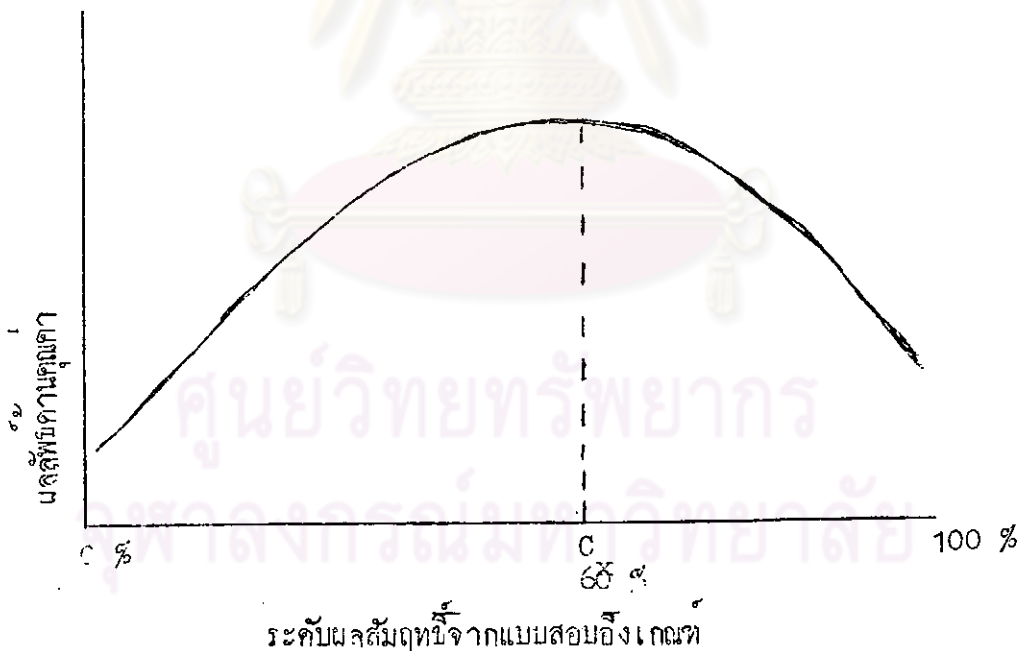
แองกอฟ (Angoff)¹ เสนอวิธีการ โดยใช้ค่าความน่าจะเป็นที่นักเรียนที่สอบผ่านในระดับต่ำสุดควรตอบได้อย่างถูกต้อง ในแบบสอบแบบเลือกตอบมาเป็นคะแนนที่คงผ่าน (passing score) วิธีการคือ ให้ผู้ตัดสินแต่ละคนเสนอค่าความน่าจะเป็นที่ผู้ที่สามารถจะยอมรับได้ระดับต่ำสุดควรจะตอบแต่ละข้อใดถูกต้อง ที่จริงแล้วผู้ตัดสินจะคิดถึงจำนวนคนที่ จะยอมรับได้ในระดับต่ำสุด

¹W. H. Angoff Scales, "Norms and Equivalent Scores," in Educational Measurement, ed. R. L. Thorndike (Washington : American Council Education, Council on Education, 1971), p. 656.

(แทนที่จะคิดเฉพาะเพียงคนเดียว) และจะประมาณค่าสัดส่วนของผู้ที่จะยอมรับได้ในระดับค่าสูงสุดซึ่งจะตอบแต่ละข้ออย่างถูกต้อง ผลรวมของความน่าจะเป็นหรือสัดส่วนเหล่านี้จะใช้เป็นคะแนนที่ยอมรับขั้นค่าสุด

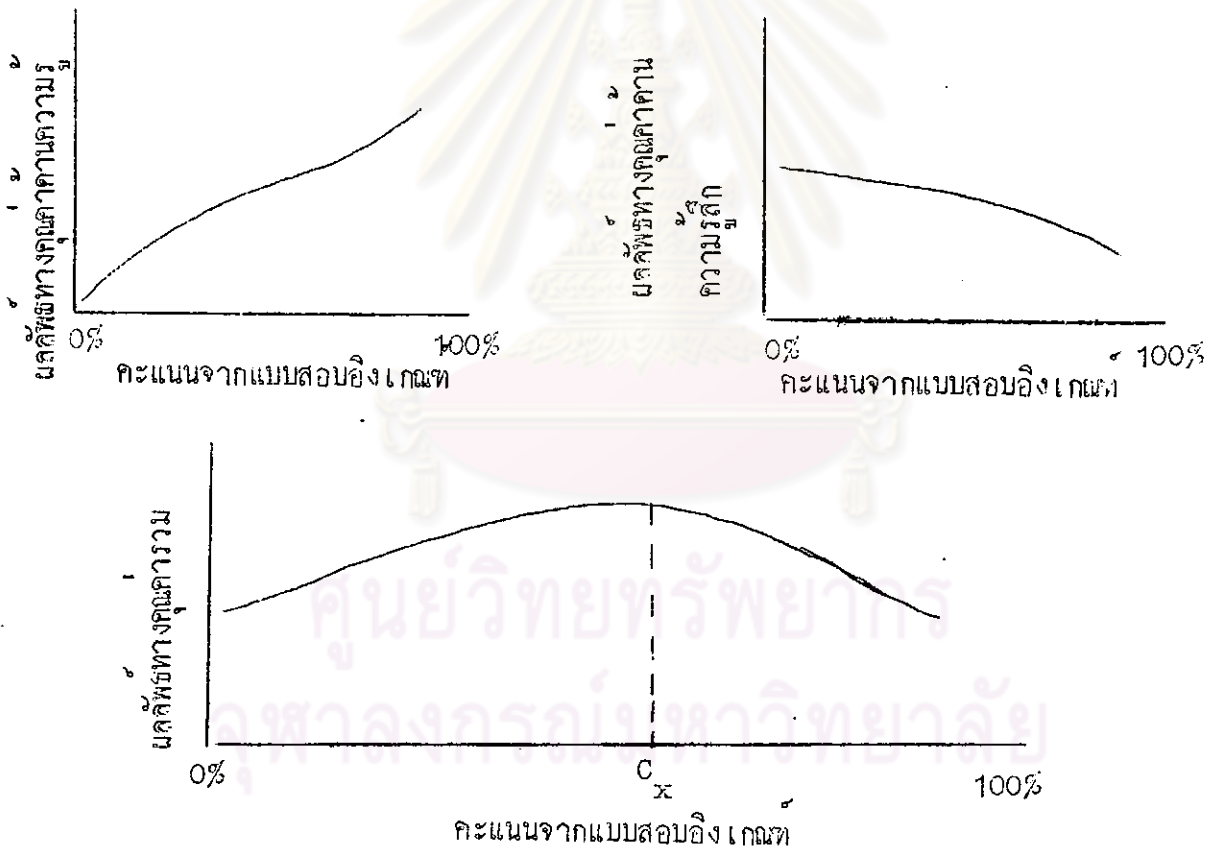
5. วิธีวิจัยเชิงปฏิบัติ (Operations Research Methods) วิธีนี้อาศัยผลลัพธ์ทางคุณค่าอย่างใดอย่างหนึ่งที่เกิดขึ้นมาช่วยในการพิจารณาคะแนนเกณฑ์ โดยวัดผลลัพธ์ทางคุณค่าของผู้ที่ไต่คะแนนต่าง ๆ กัน ในการทดสอบควบบแบบสอบอิงเกณฑ์ แล้วใช้คะแนนของผู้ที่มีผลลัพธ์ทางคุณค่าสูงสุดเป็นคะแนนเกณฑ์ วิธีการนี้บล็อก¹ (Block) ประยุกต์มาใช้ดังนี้

ทำการสอนกลุ่มต่างๆ ที่มีความเท่าเทียมกันจนมีผลสัมฤทธิ์จากการทำแบบสอบอิงเกณฑ์ระดับต่าง ๆ กัน เช่น 10 % , 15 % , 20 % , 95 % , 100 % , แล้ววัดผลลัพธ์ทางคุณค่าที่มีความสัมพันธ์กับระดับคะแนนจากแบบสอบอิงเกณฑ์ ของแต่ละกลุ่มที่มีผลสัมฤทธิ์ระดับต่าง ๆ กันนั้น นำผลที่ได้มาเขียนแกนแสดงความสัมพันธ์ เพื่อพิจารณาคะแนนเกณฑ์ (C_x)



¹Gene V Glass , "Standard and Criteria ," pp.32 - 36. Citing J.H.Block, "Student Learning and the Setting of Mastery Performance Standard," Educational Horizons,(1972):50, 183 - 190.

สำหรับวิธีการนี้ ถ้าไม่มีจุดที่เส้นกราฟจะโค้งวกกลับมาที่เส้นฐาน (base line) คือ มีลักษณะเป็นแบบโมโนโทน (monotonic graph) คะแนนเกณฑ์จะเป็น 100% ซึ่งเป็นเกณฑ์ที่ เป็นไปไม่ได้ และกราฟมักจะไม่เป็นแบบ นอนโมโนโทน (non - monotonic graph) ถ้าทั้ง แบบสอบอิงเกณฑ์และการวัดผลลัพธ์ทางคุณค่าเป็นการวัดด้านความรู้ (cognitive) ในการแก้ปัญหา นี้ รอคอยให้ใช้ผลลัพธ์ทางคุณค่าอีกด้านหนึ่ง ซึ่งมีความสัมพันธ์ในทางลบกับระดับคะแนนจาก แบบสอบอิงเกณฑ์มาประกอบการพิจารณา ซึ่งมักจะเป็นผลลัพธ์ทางคุณค่าด้านความรู้สึก (affective) เช่น ความสนใจหรือทัศนคติต่อวิชานั้น ๆ เมื่อรวมผลลัพธ์ทางคุณค่าทั้ง 2 ด้าน เข้ากันแล้วใช้ระดับคะแนนที่ให้ผลลัพธ์ทางคุณค่ารวมสูงสุดเป็นคะแนนเกณฑ์



การพิจารณาทางคุณค่ารวมควรมีการพิจารณากำหนดน้ำหนักให้กับผลลัพธ์ทางคุณค่าทั้ง 2 ด้าน ตามความสำคัญ แล้วกำหนดค่า α และ β เพื่อหาผลลัพธ์ทางคุณค่ารวมในสมการ

$$\text{ผลลัพธ์ทางคุณค่ารวม} = \alpha (\text{ผลลัพธ์ด้านความรู้}) + \beta (\text{ผลลัพธ์ด้านความรู้สึก})$$

นำคะแนนที่ได้จากการทำแบบสอบของผู้นิเทศ (xi) มาแปลงเป็น gi โดยการแปลงแบบอาร์คไซน์ซึ่งโนริคและคณะ (1973) และจะช่วยให้การคำนวณสะดวกขึ้นเพราะคะแนนที่ได้จากการแปลง gi มีการแจกแจงแบบปกติ โดย

$$g_i = \sin^{-1} \sqrt{(x_i + 3/8) / (n + 3/4)}$$

การแจกแจงนี้มีค่ามัธยฐานเลขคณิตเป็น $\sqrt{\pi}$ และความแปรปรวน V โดยที่

$$x_i / r_i = \sin^{-1} \sqrt{\pi}$$

$$V = (4n + 2)^{-1}$$

การประมาณนี้จะใช้ได้ก็เมื่อ π มีค่าอยู่ระหว่าง .15 ถึง .85 และจำนวนข้อกระทง (n) ไม่ต่ำกว่า 8 ข้อ แต่เนื่องจากไม่ทราบค่า π จึงไม่ทราบค่า V การแจกแจงของ g_i จึงเป็นการแจกแจงที่มีเงื่อนไข คือ

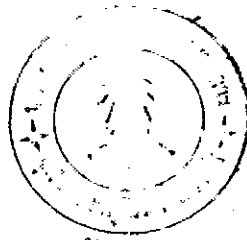
$$g_i / r_i \sim N(r_i, V)$$

การแจกแจงของ g_i จึงขึ้นกับ V ในการหาการแจกแจงของ V มีข้อตกลงว่า Y_i ของแต่ละคนเป็นตัวอย่างที่ได้มาโดยการสุ่มจากการแจกแจงอันใดอันหนึ่ง โดยเฉพาะในที่นี้กำหนดว่าการแจกแจงเดิมของ Y_i เป็นปกติด้วยมัธยฐานเลขคณิต ∞ และความแปรปรวน σ^2 ซึ่งไม่ทราบค่าทั้ง ∞ และ σ^2 ดังนั้นการแจกแจงเดิมของ Y_i จึงขึ้นอยู่กับค่าของ ∞ และ σ^2 ความรู้เกี่ยวกับค่า ∞ ไม่สำคัญเท่ากับค่า σ^2 โนริคและคณะ (1973) และเราสามารถหาค่า σ^2 ได้จากการแจกแจงแบบอินเวอร์สไคสแควร์ (inverse chi - square) ที่มี ν ของความอิสระ ν และสเกลพารามิเตอร์ λ โดยมีมัธยฐานเลขคณิตเป็น $\bar{\sigma}^2$ โดยที่

$$\bar{\sigma}^2 = \lambda / (\nu - 2)$$

จากการศึกษาของ โนริคและคณะพบว่าค่า ν ที่เหมาะสมคือ 3 ส่วนการประมาณค่า $\bar{\sigma}^2$ นั้นต้องมีข้อมูลเกี่ยวกับค่า π ซึ่งต้องอาศัยความรู้เกี่ยวกับตัวนักเรียบเนมาพิจารณาจาก ข้อกระทงที่คิดว่า จะช่วยให้ทราบข้อมูลเกี่ยวกับค่า π โดยวัดเวลาที่เรากลองการแปลงค่า π แบบอาร์คไซน์ในการทำแบบสอบที่มีจำนวนข้อกระทง t จะเป็นปกติด้วยความแปรปรวน $(4t + 2)^{-1}$ ดังนั้นจึงใช้ $(4t + 2)^{-1}$ เป็นค่าประมาณของ σ^2 และสามารถหาค่า λ ได้โดย

$$\lambda = (\nu - 2) \bar{\sigma}^2$$



ในการสอบแบบอิงเกณฑ์เราต้องการตัดสินผู้สอบแยกเป็นรายบุคคล จึงต้องการแจกแจงของ Y_i แต่ละค่า (marginal distribution of Y_i) ซึ่งต้องอาศัยค่า μ และ σ^2

จากการศึกษาของเจวิสและคณะ (1973) พบว่าการแจกแจงภายหลังของ Y_i แต่ละค่าเป็นปกติโดยประมาณควมมีขั้วมีเลขคณิต μ_i และความแปรปรวน σ_i^2

$$Y_i / \text{Data} \sim N(\mu_i, \sigma_i^2)$$

เมื่อ

$$\mu_i = \bar{g} + P^*(g_{i-1} - \bar{g})$$

$$\sigma_i^2 = \frac{1 + (m-1)P^{*2} + (g_i - \bar{g})^2}{(4m+2)m} \sigma^{*2}$$

เมื่อ

$$\bar{g} = m^{-1} \sum_{i=1}^m g_i$$

m

เป็นจำนวนผู้เข้าสอบ

n

เป็นจำนวนชอกระหว่งในแบบสอบ

P*

และ σ^{*2} ได้จากตารางของแวง (Wang) เมื่อทราบค่า μ , λ และ $S_{g_i}^2$ $S_{g_i}^2$

$$= m^{-1} \sum (g_i - \bar{g})^2$$

ค่า μ_i และ σ_i^2 ที่ได้จะนำไปใช้ในการจัดประเภทผู้สอบต่อไป

การพิจารณาจัดประเภทผู้สอบ

เนื่องจากวิธีการนี้ใช้การแปลงค่าแบบอาร์คไซน์โดยแปลง π_i เป็น Y_i ดังนั้น π_0 จึงต้องแปลงเป็น Y_0 ด้วยเพื่อให้สอดคล้องกันโดย $Y_0 = \sin^{-1} \sqrt{\pi_0}$ แล้วจัดประเภทโดยเปรียบเทียบ Y กับ Y_0 แทน π กับ π_0 ในการจัดประเภทผู้สอบใช้สัญลักษณ์ต่าง ๆ ดังนี้

ให้ w แทนระดับการยอมรับของผู้เข้าสอบ ซึ่งมีได้ 2 ค่าคือ w_1 และ w_2

w_1 แทนระดับการยอมรับของผู้สอบที่เป็นผู้ไม่ยอมรับ คือมี $Y < Y_0$

w_2 แทนระดับการยอมรับของผู้สอบที่เป็นผู้ยอมรับ คือมี $Y \geq Y_0$

ให้ a เป็นประเภทของการตัดสินใจซึ่งมีค่าได้ 2 ค่าคือ a_1 และ a_2

a_1 เป็นการตัดสินใจให้สอบไม่ผ่าน

a_2 เป็นการตัดสินใจให้สอบผ่าน

ในการตัดสินใจอาจมีความคลาดเคลื่อนเกิดขึ้น 2 ประเภท คือ ความคลาดเคลื่อนประเภทที่ 1 ซึ่งเกิดจากการปฏิเสธผิด (false negative) และความคลาดเคลื่อนประเภทที่ 2 ซึ่งเกิดจากการยอมรับที่ผิด (false positive) ผู้ตัดสินใจควรพิจารณาความสูญเสียที่เกิดจากความคลาดเคลื่อนทั้ง 2 ประเภทนี้ด้วย

ให้ $L(w_1, a_2) = l_{12}$ = ความสูญเสียที่เกิดจากความคลาดเคลื่อนในการยอมรับที่ผิด

$L(w_2, a_1) = l_{21}$ = ความสูญเสียที่เกิดจากความคลาดเคลื่อนในการปฏิเสธที่ผิด

ถ้าการตัดสินใจต้องความสูญเสียจะไม่มีคือ

$$L(w_1, a_1) = L(w_2, a_2) = 0$$

ในการตัดสินใจจะเลือกการตัดสินใจที่ให้ค่าความสูญเสียที่คาดหวังน้อยที่สุด

ถ้าให้ $E_w L(w, a)$ เป็นความสูญเสียที่คาดหวัง

ในการตัดสินใจให้สอบตก ความสูญเสียที่คาดหวังคือ

$$\begin{aligned} E_w L(w, a_1) &= 0 \cdot \text{Prob}[w = w_1] + l_{21} \text{Prob}[w = w_2] \\ &= l_{21} \text{Prob}[Y \geq Y_0] \end{aligned}$$

และในการตัดสินใจให้สอบผ่านความสูญเสียที่คาดหวังคือ

$$\begin{aligned} E_w L(w, a_2) &= l_{12} \text{Prob}[w = w_1] + 0 \cdot \text{Prob}[w = w_2] \\ &= l_{12} \text{Prob}[Y < Y_0] \end{aligned}$$

ดังนั้นเราจะตัดสินใจให้สอบตกถ้า

$$l_{21} \text{Prob}[Y \geq Y_0] < l_{12} \text{Prob}[Y < Y_0]$$

และจะตัดสินใจให้สอบผ่านถ้า

$$l_{12} \text{Prob}[Y < Y_0] < l_{21} \text{Prob}[Y \geq Y_0]$$

แต่ถ้าความสูญเสียที่ค่าครึ่งทั้ง 2 ค่าเท่ากัน การตัดสินใจหรือผ่านจะไหลลื่นเท่ากัน

ในการหาค่า $\text{Prob} [Y < Y_0]$ และ $\text{Prob} [Y \geq Y_0]$ เนื่องจากการตัดสินใจจะกระทำหลังจากได้ข้อมูลการสอบแล้ว และการแจกแจงภายหลังของ Y_i เป็นปกติโดยประมาณ ด้วยมัธยฐานเลขคณิต μ_i และความแปรปรวน σ_i^2 เราสามารถหาค่าส่วนเบี่ยงเบนมาตรฐานปกติของ Y_i แต่ละค่าจากสูตร

$$z_{01} = (Y_0 - \mu_i) / \sigma_i$$

ดังนั้นเราจึงสามารถเปรียบเทียบในรูปของคะแนน z ได้โดย

$$\text{Prob} [Y_i \geq Y_0 | \text{Data}] \approx \text{Prob} [z \geq z_{01} | \text{Data}]$$

$$\text{และ} \quad \text{Prob} [Y_i < Y_0 | \text{Data}] \approx \text{Prob} [z < z_{01} | \text{Data}]$$

หลังจากทำการตัดสินใจประเภทแล้ว จะได้คะแนนจุดตัดซึ่งเป็นคะแนนที่ผู้สอบผ่านในระดับต่ำสุดทำได้

สวามินาธานและคณะ (1975) ได้สรุปลำดับขั้นในการประมาณดังนี้

1. แปลงคะแนน x_i ของผู้สอบแต่ละคนเป็น z_i โดยใช้สูตร

$$z_i = \sin^{-1} \sqrt{(x_i + 3/8)(n+3/4)}$$

2. กำหนดคะแนนจุดตัด (π_0) และเปลี่ยนเป็น Y_0 โดย

$$Y_0 = \sin^{-1} \sqrt{\pi_0}$$

3. หากการแจกแจงภายหลังของ ϕ โดยระบุค่าพารามิเตอร์ λ และ θ

4. หาค่า p^* และ σ^{*2} จากตารางของวง แล้วหาค่ามัธยฐานเลขคณิต μ_i และความแปรปรวน σ_i^2 ของการแจกแจงภายหลังของ Y_i

5. หาค่าส่วนเบี่ยงเบนมาตรฐานปกติจาก

$$z_{0i} = (Y_0 - \mu_i) / \sigma_i$$

6. หาค่า $\text{Prob} [z < z_{0i} | \text{Data}]$ และ $\text{Prob} [z \geq z_{0i} | \text{Data}]$

7. จัดประเภทผู้สอบโดยพิจารณาจากความสูญเสียดังกล่าวแล้ว

การพิจารณาจัดประเภทผู้สอบมากกว่า 2 ประเภท

สวามินาธานและคณะ (1975) ได้ประยุกต์การใช้ทฤษฎีการตัดสินใจของเบส
มาใช้ในการจัดผู้สอบเป็นหลายพวก เช่น ถ้าจัดเป็น 3 พวก ก็อาจจัดเป็นพวกรอบรู้
(master) พวกรอบรู้เป็นบางส่วน (partial master) และพวกไม่รอบรู้ (non-
master)

ถ้าให้ k เป็นจำนวนประเภทที่จะจัดระดับการรอบรู้ จะมี k ระดับคือ w_1, w_2, \dots, w_k และจะต้องมีคะแนนจุดตัด $k-1$ ค่าคือ $\pi_{0i-1}, \dots, \pi_{0i-k}$ ผู้สอบจะถูกจัดให้อยู่
ในระดับ w_i เมื่อคะแนนโคเมน π น้อยกว่าคะแนนจุดตัดที่ $i-1$ (π_{0i-1}) และอยู่ในระดับ w_i เมื่อ
คะแนนโคเมนมากกว่าหรือเท่ากับ π_{0i-1} แต่ต่ำกว่า π_{0i} หรือเขียนในรูปทั่วไปได้ว่า ผู้สอบ
จะอยู่ในระดับ w_i เมื่อ $\pi_{0i-1} \leq \pi < \pi_{0i}$

ค่าความสูญเสียในการจัดประเภทผิดคือ $L(w_i, a_j) = \pi_{ij}$

ค่าความสูญเสียที่คาดหวังสำหรับการกระทำ a_j คือ

$$E_w L(w, a_j) = \sum_{i=1}^k \pi_{ij} \text{Prob} [Y_{0i-1} < Y_0 < Y_{0i} \mid \text{Data}]$$

การกระทำ a_j จะถูกเลือกเมื่อ

$$\sum_{j=1}^k \pi_{ij} \text{Prob} [Y_{0i-1} \leq Y < Y_{0i} \mid \text{Data}] < \sum_{j=1}^k \pi_{ij} \text{Prob} [Y_{0i-1} \leq Y < Y_{0i} \mid \text{Data}], P=1, \dots, k$$

ค่าของ $\text{Prob} [Y_{0i-1} \leq Y < Y_{0i} \mid \text{Data}]$ หาได้ในทำนองเดียวกันกับการจัดเป็น 2 ประเภท $P \neq j$)

ดังกล่าวแล้ว

องค์ประกอบในการพิจารณาคะแนนจุดตัด

ในการกำหนดจุดตัดนั้น มิลแมน (Millman, 1973) กล่าวว่ายังไม่มีวิธีการ
ใดที่ไม่ต้องใช้การพิจารณาตัดสิน¹ มิลแมนไม่ได้แนะนำให้ใช้วิธีใดวิธีหนึ่งในการกำหนดคะแนน
จุดตัด แต่ให้พิจารณาวิธีการและข้อมูลต่าง ๆ อย่างระมัดระวังในการกำหนดจุดตัด องค์
ประกอบที่เสนอให้พิจารณามีดังนี้²

¹J. Millman, "Passing Score and Test Lengths for Domain-Referenced Measures" Review of Educational Research 43 (1973) :206.

²Ibid., 206 - 311.

1. คะแนนของผู้สอบคนอื่น ๆ (Performance of others) การกำหนดคะแนนจุดตัดโดยพิจารณาองค์ประกอบค่านนี้ อาจทำได้โดยกำหนดคะแนนจุดตัดโดยพิจารณาจำนวนนักเรียนที่จะสอบผ่านไว้ก่อน แล้วเลือกจากผู้มีความสามารถสูงสุดลงมา นอกจากนี้อาจใช้ระดับคะแนนของผู้ที่ได้รับการรับรองแล้วว่ามีความรู้ในเนื้อหา นั้นมากำหนดเป็นคะแนนจุดตัด
2. เนื้อหาของข้อกระทง (Item Content) ข้อกระทงในแบบสอบจะได้รับการพิจารณาโดยผู้เชี่ยวชาญเพื่อตัดสินว่ามีความสำคัญที่จะต้องให้สอบตอบโต้บ้างถูกต้องหรือไม่ เช่น ในการกำหนดมาตรฐานการปฏิบัติที่น่าพอใจสำหรับเด็กเกรด 12 ชั้นแรกต้องศึกษานิวทอนและข้อควบความระมัดระวัง แล้วตัดสินว่ามีปัญหจำนวนกี่ข้อที่ควรจะทำได้ คือกำหนดในรูปคะแนนดิบที่ผู้สอบควรทำได้ องค์ประกอบค่านนี้เหมือนกับวิธีพิจารณาที่กำหนดคะแนน เกณฑ์จากความสามารถต่ำสุดที่กลาสเสนอไว้นั้นเอง
3. ผลทางการศึกษาที่ตามมา (Educational Consequences) ในค่านนี้ถ้าตั้งระดับเกณฑ์ค่าเกินไปนักเรียนที่ผ่านไปเรียนบทเรียนต่อไป อาจเรียนไม่ทันและทักษะใหม่ ๆ ใดอย่างไม่มีประสิทธิภาพ ถ้าตั้งเกณฑ์สูงเกินไปจะทำให้นักเรียนเสียเวลาในการเรียนซ่อมเสริมโดยไม่จำเป็น สิ่งที่ต้องพิจารณาคือ คะแนนจุดตัดเท่าใดจึงจะทำให้เกิดประโยชน์ทางการศึกษามากที่สุด ถ้าความรู้และทักษะที่วัดเป็นพื้นฐานของสิ่งที่จะเรียนต่อไป ควรตั้งเกณฑ์ให้สูง แต่ถ้าเนื้อหา นั้นไม่ใช่สิ่งที่จะต้องใช้ในการเรียนต่อไป คะแนนจุดตัดควรต่ำลง
4. คุณค่าทางจิตวิทยาและทางเศรษฐกิจ (Psychological and Financial costs) ในการพิจารณากำหนดคะแนนจุดตัด ถ้าคุณค่าทางจิตวิทยาและทางเศรษฐกิจที่เกี่ยวข้องกับโปรแกรมการสอนซ่อมเสริมสูง ควรกำหนดคะแนนจุดตัดให้ต่ำ คุณค่าเหล่านี้อาจรวมทั้งแรงจูงใจและควมเบื่อ การทำลายความรู้สึกส่วนตัวของผู้เรียน เงินและเวลาที่ต้องใช้ในการสอนซ่อมเสริม ถ้าคุณค่าเหล่านี้ไม่มากนักจุดตัดควรสูงขึ้น หรือถ้าการให้เด็กสอบผ่านทำให้เกิดผลในทางลบ เช่น เกิดความสับสน หรือประสิทธิภาพในการเรียนลดลง ก็ควรกำหนดคะแนนจุดตัดให้สูงเช่นกัน
5. ความคลาดเคลื่อนเนื่องจากการเดาและการสุ่มข้อกระทง (Error: Due to Guessing and Item Sampling) การประมาณความสามารถของนักเรียนจะมีความคลาดเคลื่อนที่เป็นระบบ (systematic error) เมื่อรูปแบบของข้อกระทงในแบบสอบเปิดโอกาส

ให้นักเรียนตอบถูกได้โดยการเดา การกำหนดคะแนนจุดตัดที่สูงจะช่วยลดล้างการเดาที่ค่าความถี่อยู่ในการทำแบบข้อสอบ อีกวิธีหนึ่งคือคะแนนของนักเรียนควรมีการปรับตามสูตรการแก้การเดามาตรฐานแล้ว จึงนำมาเปรียบเทียบกับมาตรฐานที่กำหนด การปรับคะแนนของผู้สอบหรือการปรับมาตรฐานนี้จะช่วยควบคุมการเดาได้เพียงบางส่วนเท่านั้น

ความคลาดเคลื่อนที่เชื่อว่าอาจเกิดขึ้นได้อีกนั้นมาจากความยุ่งยากในการสร้างแบบสอบ ความไม่สะดวกในการบริหารการสอบหรือการละเลยบางสิ่งบางอย่าง คำถามและเนื้อหาต่าง ๆ ที่เป็นตัวแทนใน โดเมนไม่ได้นำมาใช้ในแบบสอบ เมื่อเกิดความสงสัยว่าเนื้อหาในแบบสอบจะไม่เป็นตัวแทนควรมีการปรับคะแนนจุดตัดที่สูงขึ้นหรือต่ำลง เพื่อป้องกันความคลาดเคลื่อนที่เกิดจากการจัดประเภทผิดทั้งประเภท 1 และประเภท 2

ข้อเสนอแนะเกี่ยวกับคะแนนจุดตัด

แฮมเบิลตัน (Hambleton, 1978)¹ ได้ให้ข้อเสนอแนะเกี่ยวกับการกำหนดคะแนนจุดตัดซึ่งได้มาจากประสบการณ์ในการทำงานในหลาย ๆ โรงเรียนดังนี้

1. การพิจารณาคะแนนจุดตัดควรทำโดยกลุ่มที่ทำงานร่วมกันหลาย ๆ กลุ่ม กลุ่มเหล่านี้ได้แก่ ครู ผู้ปกครอง ผู้เชี่ยวชาญด้านหลักสูตร ผู้บริหารโรงเรียนและนักเรียน (ในกรณีที่นักเรียนในระดับมัธยมศึกษาตอนปลาย) จำนวนคนในแต่ละกลุ่มขึ้นกับความสำคัญของแบบสอบที่นำมาพิจารณา และจำนวนของตารางวิเคราะห์ โดเมนของข้อสอบ โดยอย่างน้อยที่สุดควรมีจำนวนคนเพียงพอที่จะแบ่งผู้พิจารณาเป็น 2 กลุ่มเพื่อเปรียบเทียบความคงที่ของการตัดสินจากบุคคล 2 กลุ่มนี้
2. ผู้ทำการตัดสินควรนำตารางวิเคราะห์ โดเมนของข้อสอบมาเสนอและอภิปรายร่วมกัน
3. การใช้วิธีของอีเบลหรือเอ็มเคลสกี ควรทำการทดลองกับกลุ่มตัวอย่างหลายพวกเพื่อตัดสินและแก้ปัญหาเกี่ยวกับความแตกต่างระหว่างกลุ่ม

¹Ronald K. Hambleton, "Use of Cut-off Scores," Journal of Educational Measurement 15(1978): 285-289.

4. พยายามจัดเวลาให้เหมาะสมกับการวางวิเคราะห์โคเมนข้อสอบแต่ละตาราง หากตารางวิเคราะห์โคเมนใดมีความซับซ้อนหรือมีความสำคัญมาก ควรให้เวลามากขึ้น

5. พยายามทำให้เกิดความมั่นใจว่าผู้ตัดสินมีความรู้เกี่ยวกับการใช้แบบสอบและรู้ว่าจะใช้กับนักเรียนกลุ่มใด

6. หากมีความสัมพันธ์ระหว่างตารางวิเคราะห์โคเมนควรนำมาพิจารณาด้วย เช่น ถ้าตารางวิเคราะห์โคเมนใดเป็นสิ่งที่ต้องการหรือจำเป็นมากกว่าควรกำหนดคะแนนจุดตัดให้สูงกว่า

7. เมื่อใดก็ตามที่เป็นไปก็พยายามใช้บุคคล 2 กลุ่มหรือมากกว่าในการตัดสินคะแนนจุดตัด ควรพิจารณาความคงที่ในการกำหนดค่าน้ำหนักและความแตกต่างในการกำหนดค่าน้ำหนักของผู้ตัดสิน และทำการอภิปรายเพื่อหาการตัดสินที่สอดคล้องกัน

8. ถ้าข้อมูลเกี่ยวกับการสอบครั้งก่อน ๆ มีประโยชน์ก็สามารถนำมาใช้เพื่อพิจารณาแก้ไขเกี่ยวกับการกำหนดคะแนนจุดตัดและ เวลาที่ใช้สอนได้

9. หากข้อมูลที่ได้จากการสอบเหมาะที่จะนำมาใช้พิจารณาเปอร์เซ็นต์ของผู้รอบรู้และผู้ไม่รอบรู้ในแต่ละวัตถุประสงค์ ก็ควรนำมาศึกษา หากคะแนนที่ได้จากการทำแบบสอบมีที่ทาวจะเป็นไปในลักษณะผิดปกติ (Out of line) เราก็สามารถอธิบายสาเหตุได้โดยการพิจารณาข้อกระทงของแบบสอบ (ซึ่งอาจไม่มีความตรงก็ได้) รัศมีของคะแนนจุดตัด ความแปรผันของคะแนนสอบระหว่างชั้นเรียนและปริมาณเวลาที่ใช้สอน

10. เปรียบเทียบสภาพการรอบรู้ของผู้สอบที่ยังไม่ได้เรียน และที่เรียนแล้ว โดยกลุ่มที่เรียนแล้วควรมีจำนวนผู้รอบรู้มาก กลุ่มที่ยังไม่ได้เรียนควรมีจำนวนผู้ไม่รอบรู้มาก หากมีนักเรียนจำนวนมากได้รับการจัดประเภทผิดต้องพิจารณาคะแนนจุดตัดที่มีความตรงมากกว่าคะแนนจุดตัดที่ใช้

11. หากหลักสูตรมีการเปลี่ยนแปลงซึ่งจะมีผลต่อคะแนนจุดตัด จำเป็นต้องมีการตรวจสอบคะแนนจุดตัดใหม่ให้เหมาะสม



ความตรงของแบบสอบอิงเกณฑ์

ความตรงตามเนื้อหา (Content Validity) เป็นสิ่งจำเป็นอย่างยิ่งสำหรับแบบสอบอิงเกณฑ์ นอกจากนี้จะต้องมีความตรงของการตัดสิน คือแบ่งผู้สอบออกเป็นผู้รอบรู้และผู้ไม่รอบรู้ อย่างถูกต้องตามเหตุการณ์ที่มุ่งวัด ผู้เสนอวิธีการในการพิจารณาความตรง คือ

คาร์เวอร์ (Carver)¹ เสนอวิธีการโดยให้กลุ่มที่เรียนแล้วและกลุ่มที่ยังไม่ได้เรียนทำแบบสอบ แล้วคำนวณค่าความตรงจากผลที่ได้จากการตัดสินผู้เข้าสอบทั้ง 2 กลุ่มเป็นผู้รอบรู้และไม่รอบรู้ ดังนี้

	กลุ่มที่ยังไม่ได้เรียน	กลุ่มที่เรียนแล้ว
ผ่าน	b	a
ไม่ผ่าน	c d	d e

สัมประสิทธิ์ความตรง = $a + c/N$; $N = a + b + c + d$

ดัชนีความตรงอีกตัวเ้ามาจากการหาค่าสัมประสิทธิ์สหสัมพันธ์แบบ พี มีสูตรว่า²

$$r = \frac{ad - bc}{(a+b)(c+d)(a+c)(b+d)}$$

เมื่อ

- a คือจำนวนผู้สอบผ่านหลังจากเรียน
- b คือจำนวนผู้สอบผ่านก่อนเรียน
- c คือจำนวนผู้สอบไม่ผ่านหลังจากเรียน
- d คือจำนวนผู้สอบไม่ผ่านก่อนเรียน

¹K.D.Crohan, "Item Analysis for...", p.256. Citing R.P.Carver, "Special Problem in Measuring Change with Psychometric Devices," in Evaluation Research: Strategies and Method, (Pittsburgh, American Institutes for Research, 1970)

²Jason Millman, "Criterion-Referenced Measurement, in Evaluation Education, ed. W.J. Popham (Berkeley: Mecutchan Publishing Corporation, 1974)

เบอร์ค (1976) ได้หาความตรงของการตัดสินใจเพื่อเลือกคะแนนจุดตัดของแบบ
สอบโดยอาศัยผลการตัดสินใจในการทำแบบสอบของกลุ่มที่เรียนแล้วและยังไม่ได้เรียนดังนี้

	กลุ่มที่เรียนแล้ว	กลุ่มที่ยังไม่ได้เรียน
รอบรู้โดยการทำนาย	พวกรอบรู้จริง (TM)	พวกรอบรู้ไม่จริง (FM)
ไม่รอบรู้โดยการทำนาย	พวกไม่รอบรู้ไม่จริง (FN)	พวกไม่รอบรอบรู้จริง (TN)

ค่าสัมประสิทธิ์ความตรงได้จากค่าสหสัมพันธ์แบบ ϕ

$$\phi \text{ v c} = \frac{P(TM) - BR(SR)}{\sqrt{BR(1-BR) SR(1-SR)}}$$

เมื่อ $P(TM)$ คือความน่าจะเป็นของพวกรอบรู้จริง = $\frac{TM}{(M + N)}$, ($M + N$
คือจำนวนคนทั้งหมด)

BR คือ ค่าความน่าจะเป็นของผู้รอบรู้ในประชากรมีค่า = $P(FN) + P(TM)$

SR คือ ค่าความน่าจะเป็นของผู้รอบรู้โดยการทำนายในประชากร
มีค่า = $P(TM) + P(FM)$

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

¹Ronald A. Berk, "Determination of Optional...", pp. 5 - 7.

ความเที่ยงของแบบสอบอิงเกณฑ์

ในการหาความเที่ยงของแบบสอบอิงเกณฑ์ พอบแฮมและฮูเชค (1969) กล่าวว่า แม้วาความสอดคล้องภายใน (internal consistency) และความคงที่ทางเวลา (temporal stability) จะเป็นลักษณะสำคัญของคะแนนที่ได้จากการทำแบบสอบอิงเกณฑ์ การใช้ทฤษฎีคลาสสิกมาหาค่าความเที่ยงของแบบสอบอิงเกณฑ์อาจไม่เหมาะสม ทั้งนี้เพราะคะแนนที่ได้จากการใช้แบบสอบอิงเกณฑ์อาจมีการกระจายน้อยทำให้ได้ค่าความเที่ยงต่ำ¹

นักการศึกษาได้คิดวิธีการในการหาค่าความเที่ยงของแบบสอบอิงเกณฑ์ซึ่งแยกออกเป็น 2 ฝ่ายคือ ฝ่ายที่ 1 พิจารณาในรูปความเที่ยงของคะแนนที่ได้จากการทำแบบสอบอิงเกณฑ์ อีกฝ่ายหนึ่งพิจารณาจากความคงที่ในการตัดสินระหว่างการสอบ 2 ครั้ง

การหาค่าความเที่ยงของคะแนนที่ได้จากการทำแบบสอบอิงเกณฑ์ ผู้เสนอสูตรซันไซคือ ลิวิงสตัน (Livingston, 1972) โดยกำหนดว่าจุดประสงค์ของแบบสอบอิงเกณฑ์คือแยกคะแนนโดเมนที่ได้จากการประมาณของผู้สอบแต่ละคนออกจากคะแนนจุดตัด จึงควรใช้การหาค่าความเที่ยงแบบของคะแนนโดเมนที่ได้จากการประมาณและคะแนนโดเมนของผู้สอบออกจากคะแนนจุดตัดมากกว่าคะแนนหนึ่งเป็นค่ามัธยฐาน เลขคณิตของคะแนนโดเมนซึ่ง เป็นวิธีที่ทฤษฎีคลาสสิกใช้ สูตรในการหาค่าความเที่ยงจึงเป็น 2

$$(\hat{\pi}, \pi) = \frac{\sigma_{(\pi)}^2 + (\bar{\pi} - \pi_0)^2}{\sigma_{(\hat{\pi})}^2 + (\bar{\pi} - \pi_0)^2}$$

¹W. J. Popham and T. R. Husck, "Implications of Criterion - Referenced Measurement," p. 4

²R. K. Hambleton, "Criterion - Referenced Testing...", p. 16

เมื่อ	$\hat{\pi}$	คือคะแนนโคเมนที่ได้จากการประมาณ
	π	คือคะแนนโคเมน
	$\bar{\pi}$	คือมัธยิมเลขคณิตของคะแนนโคเมน
	$\hat{\sigma}(\hat{\pi})$	คือความแปรปรวนของคะแนนโคเมนที่ได้จากการประมาณออก จากคะแนนจุดตัด
	$\hat{\sigma}(\pi)$	คือความแปรปรวนของคะแนนโคเมนออกจากคะแนนจุดตัด
	π_0	คือคะแนนจุดตัด

จากสูตรนี้จะเห็นว่า แม้ว่าความแปรปรวนของคะแนนโคเมนจะเป็นศูนย์ ค่าความเที่ยงก็ยังคงมากกว่า 0 (ซึ่งตามทฤษฎีคลาสสิกค่าความเที่ยงจะเป็นศูนย์) และค่าความเที่ยงจะเพิ่มขึ้นเมื่อ $(\hat{\pi} - \pi_0)^2$ เพิ่มขึ้น หรือ ถ้ามัธยิมเลขคณิตของคะแนนโคเมนของกลุ่มยิ่งเบี่ยงเบนออกจากคะแนนจุดตัดมาก ค่าความเที่ยงจะยิ่งสูงขึ้น สูตรการหาค่าความเที่ยงของลิวิงสตันสามารถจัดแปลงให้อยู่ในรูป¹

$$K^2(x, T_x) = \frac{\hat{\sigma}^2(T_x) + (\mu_x - C_x)^2}{\hat{\sigma}^2(x) + (\mu_x - C_x)^2}$$

เมื่อ	$\hat{\sigma}^2(T_x)$	คือความแปรปรวนของคะแนนจริง
	$\hat{\sigma}^2(x)$	คือความแปรปรวนของคะแนนจากการสังเกต
	μ_x	คือค่ามัธยิมเลขคณิต
	C_x	คือคะแนนจุดตัด

¹Chester W. Harries, "An Interpretation of Livingston's Reliability Coefficient for Criterion - Referenced Tests," Journal of Educational Measurement 9(1972):27.

หรือ¹

$$K^2(x, T_x) = \frac{\rho^2(x, T_x) \sigma^2(x) + (\mu_x - c_x)^2}{\sigma^2(x) + (\mu_x - c_x)^2}$$

เมื่อ $\rho^2(x, T_x)$

คือความเที่ยงแบบอิงกลุ่ม

 $\sigma^2(x)$

คือความแปรปรวนของคะแนนดิบ

 μ_x

คือมัชฌิม เลขคณิต

 c_x

คือคะแนนจุดตัด

แฮร์ริส (Harris)² เห็นว่าสูตรที่ลิวิงสตันใช้นั้นค่าความเที่ยงเหมือนกับได้มาจากประชากร 2 กลุ่มที่มีมัชฌิมเลขคณิตสูงกว่าคะแนนจุดตัดและต่ำกว่าคะแนนจุดตัดเท่ากัน ซึ่งความแปรปรวนของคะแนนจริงและความแปรปรวนของความคลาดเคลื่อนจะต้องเหมือนกัน ในประชากรทั้ง 2 กลุ่ม นั่นคือการใช้สูตรของลิวิงสตันไม่มีการพิจารณาว่าคะแนนจริงสูงกว่าหรือต่ำกว่าคะแนนจุดตัด อีกประการหนึ่งความแปรปรวนของความคลาดเคลื่อนในการวัดอาจใช้เป็นดัชนีในการแสดงความมั่นใจว่าแต่ละคนมีคะแนนจริงสูงกว่าหรือต่ำกว่าคะแนนจุดตัดได้ก็ว่าขนาดของสัมประสิทธิ์ความเที่ยง

เชฟเวลสัน, บล็อกและเรวิทซ์ (Shavelson, Block and Ravitch)³

เห็นว่าสูตรที่ลิวิงสตันใช้นั้นค่าความแตกต่างระหว่างมัชฌิมเลขคณิตและคะแนนจุดตัดมากจะมีอิทธิพลต่อค่าสัมประสิทธิ์ความเที่ยงสูงกว่าความแตกต่างภายในตัวผู้สอบแต่ละคน ในการทำแบบสอบบุขนานเนื่องจากความคลาดเคลื่อนของการวัด เชฟเวลสันและคณะให้ข้อเสนอแนะว่าควรแบ่งแบบสอบเป็นฉบับย่อย ที่วัดแต่ละวัตถุประสงค์ในแบบสอบแล้วความเที่ยงของแต่ละฉบับซึ่งค่าความเที่ยงของแบบ

¹Samuel A. Livingston, "Criterion-Referenced Application of Classical Test Theory," Journal of Educational Measurement 9(1972):17-18.

²C. W. Harris, "An Interpretation of Livingston's...", pp.27-29.

³R. A. Shavelson, James H. Block and M. M. Ravitch, "Criterion Referenced Testing; Comment on Reliability," Journal of Educational Measurement 9(1972) : 133-139.

สอบฉบับย่อยจะต่ำกว่าค่าความเที่ยงของคะแนนรวมทั้งหมด แต่จะให้ข้อมูลที่ต้องการในการวินิจฉัยแต่ละคนในแบบสอบฉบับย่อย

ในการหาความเที่ยงของแบบสอบที่มีความยาวเพิ่มขึ้น ลิวิงสตันเสนอให้ใช้สูตรของ สเพียร์แมน - บราวน์ คือ¹

$$\text{ความเที่ยง} = \frac{nK^2(x, Tx)}{1 + (n-1)K^2(x, Tx)}$$

เมื่อ n เป็นจำนวนเหาของข้อสอบที่เพิ่มขึ้นจากเดิม

ในกรณีที่ข้อมูลที่ได ไม่ใช่อข้อมูลแบบแบ่งเป็นสอง (Binary Data) ลิวิงสตันเสนอให้ใช้สัมประสิทธิ์แอลฟา²

$$\alpha = \frac{\frac{I}{I-1} \sum \sigma_i^2(x) - \sum \sigma_i^2(i) + (\mu_x - Cx)^2}{\sum \sigma_i^2(x) + (\mu_x - Cx)^2}$$

เมื่อ I คือจำนวนข้อกระทง

σ_i^2 คือความแปรปรวนของข้อกระทงแต่ละข้อ

การหาความเที่ยงของการตัดสินใจ มีผู้เสนอวิธีการไว้หลายวิธีด้วยกัน ซึ่งมีทั้งวิธีการที่ต้องใช้การบริหารการสอบ 2 ครั้ง และการบริหารการสอบเพียงครั้งเดียว ดังนี้

ก. การหาความเที่ยงของการตัดสินใจโดยการบริหารการสอบ 2 ครั้ง มีผู้เสนอสูตรขึ้นไว้คือ

¹Samuel A Livingston, "Criterion-Referenced Application...",

²Ibid.

คาร์เวอร์(1970) เสนอการประมาณค่าความเที่ยง โดยใช้กลุ่มตัวอย่างกลุ่มเดียวทำแบบสอบคุณาน แล้วเปรียบเทียบสัดส่วนของความคงที่ในการตัดสินดังนี้

ฉบับ ก ฉบับ ข	ตก	ได้
ได้	b	a
ตก	c	d

ค่าความเที่ยง $= (c+c)/N$, เมื่อ $N = a+b+c+d$

แอมเบิลตันและโนวิก (1973) เสนอสูตรที่ใช้ในการหาค่าความเที่ยงของการตัดสินแบ่งผู้เรียนตามระดับการรอบรู้เป็น 3 ระดับ จากการบริหารการสอบ 2 ครั้ง โดยใช้แบบสอบฉบับเดียวกันหรือแบบสอบคุณานดังนี้²

$$P_0 = \sum_{i=1}^m P_{ii}$$

เมื่อ P_0 คือดัชนีความเที่ยง
 P_{ii} คือสัดส่วนของผู้สอบที่ถูกตัดสินให้อยู่ในระดับการรอบรู้ดีในการบริหารการสอบทั้ง 2 ครั้ง

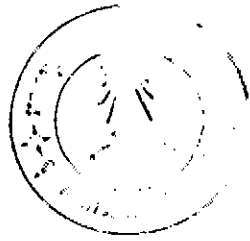
สวามินาธานแอมเบิลตันและอัลกิงา (1974) มีความเห็นว่าค่า P_0 ของแอมเบิลตันและโนวิกนั้นอาจรวมการตัดสินตรงกันโดยบังเอิญไว้ด้วยเขาเสนอให้ใช้สัมประสิทธิ์ K เป็นดัชนีความเที่ยง

$$K = (P_0 - P_c)/(1 - P_c)$$

¹Kolvin D. Crehan, "Item Analysis for Teacher Made Mastery Tests," "Journal of Educational Measurement 11(1974):256.

²Michael J. Sbkowiak, "Estimating Reliability from Single Administration of a Criterion-Referenced Tests," "Journal of Educational Measurement 13, 1976 : 266

³Ronald K. Hambleton, "Criterion-Referenced Testing...", p. 21.



$$\text{เมื่อ } P_c = \sum_{k=1}^m P_k \cdot P_{.k}$$

$P_{.k}$ และ P_k คือสัดส่วนของผู้เข้าสอบที่ถูกคัดสินให้อยู่ในระดับการรอบรู้ที่ k ในการบริหารการสอบครั้งที่ 1 และ 2 ตามลำดับ

ข. การหาความเที่ยงของการคัดสินโดยการบริหารการสอบ 1 ครั้ง ผู้คิดหาวิธีการในการหาค่าความเที่ยงของแบบสอบอิงเกณฑ์โดยวิธีนี้ มีความเห็นว่าในทางปฏิบัติการสอบซ้ำหรือการใช้แบบสอบคู่ขนานมีความยุ่งยาก จึงพยายามหาวิธีที่ใช้การบริหารการสอบเพียงครั้งเดียว ผู้เสนอสูตรขึ้นใช้ คือ

ซิมโคเวียค¹(1976) เสนอวิธีการที่ประยุกต์มาจากการบริหารการสอบ 2 ครั้ง โดยแบ่งระดับการรอบรู้เป็น 2 ระดับ โดยใช้สูตร

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

$$\text{โดยที่ } P_c^{(i)} = P(x_i \geq c) \cdot P(x'_i \geq c) + P(x_i < c) \cdot P(x'_i < c)$$

เมื่อ x_i คือคะแนนที่ได้จากแบบสอบฉบับที่ 1
 x'_i คือคะแนนที่ได้จากแบบสอบฉบับที่ 2 ซึ่งเป็นแบบสอบคู่ขนานกับฉบับที่ 1
 c คือคะแนนจุดตัด

¹Michael J. Subkowiak, "Estimating Reliability from...", pp.267- 268.

วิธีการของซิมโคเวียคมีข้อตกลง 2 ข้อคือ

1. คะแนนจากแบบสอบฉบับที่ 1 (x_i) และจากแบบสอบฉบับที่ 2 (x'_i) ของผู้สอบแต่ละคนมีการแจกแจงที่เป็นอิสระต่อกัน หมายความว่าประสิทธิภาพในการทำแบบสอบฉบับที่ 1 ไม่มีผลต่อผลที่ได้จากการทำแบบสอบฉบับที่ 2

2. ถ้าไร้แจกแจงของคะแนนจากแบบสอบฉบับที่ 1 และฉบับที่ 2 ของผู้เข้าสอบแต่ละคนเป็นแบบโบโนเมียล ซึ่งมีลักษณะดังนี้

ก. ข้อกระทงแต่ละข้อในแบบสอบมีการให้คะแนนแบบ 0 หรือ 1

ข. ผลจากการทำข้อกระทงหนึ่งไม่มีผลต่อข้อกระทงอื่น

ค. ความน่าจะเป็นของคำตอบที่ถูกต้องจะคงที่ในทุกข้อกระทง

จากข้อตกลงนี้จะเห็นได้ว่า

$$\begin{aligned} P_c(i) &= [P(x_i \geq c)]^2 + [P(x_i < c)]^2 \\ &= [P(x_i \geq c)]^2 + [1 - P(x_i \geq c)]^2 \end{aligned}$$

$$\text{เมื่อ } P(x_i \geq c) = \sum_{x_i=c}^n \binom{n}{x_i} P_i^{x_i} (1-P_1)^{n-x_i}$$

P_i คือ ความน่าจะเป็นที่แท้จริง ของคำตอบในข้อกระทงที่ถูกต้องสำหรับผู้เข้าสอบคนที่ i ซึ่งสามารถประมาณได้จากคะแนนที่ได้ในการทำแบบสอบ

$$\text{ฉบับเดียวคือ } \hat{P}_i = x_i/n$$

ฮวน¹ (1976) คำนวณค่าความเที่ยงในรูปสัมประสิทธิ์สหสัมพันธ์ โดยแยกการพิจารณาเป็น

2 กรณีคือ

¹Huynh Huynh, "On the Reliability of Decisions in Domam-Referenced Testing," Journal of Educational Measurement 13(1976): 254 6 257.

1. ถ้าคะแนนจุดตัดมีค่าสูงใกล้เคียงค่า μ (จำนวนข้อกระทง) ไชสุตร

$$\begin{aligned}
 K &= (P_{11} - P_1^2) / (P_1 - P_1^2) \\
 \text{เมื่อ } P_{11} &= \sum_{x,y=c}^n f(x,y) \\
 &= f(n,n) + f(n,n-1) + f(n-1,n) + \dots + f(c,c) \\
 P_1 &= \sum_{x=c}^n f(x) \\
 &= f(n) + \dots + f(c)
 \end{aligned}$$

ข. ถ้าคะแนนจุดตัด (c) มีค่าต่ำ ไชสุตร

$$\begin{aligned}
 K &= (P_{00} - P^2_0) / (P_0 - P^2_0) \\
 P_{00} &= \sum_{x,y=0}^{c-1} f(x,y) \\
 &= f(0,0) + f(0,1) + \dots + f(c-1,c-1) \\
 P_0 &= \sum_{x=0}^{c-1} f(x) \\
 &= f(0) + \dots + f(c-1)
 \end{aligned}$$

แบริส¹(1972) เสนอวิธีการในกรณีที่แบ่งระดับการรอบรู้เป็น 2 ระดับ ($k=2$) โดยไชคาสสัมพันธ์กำลังสองระหว่างระดับการรอบรู้กับคะแนนรวม โดยให้คะแนน 0 สำหรับผู้ไม่รอบรู้ และคะแนน 1 สำหรับผู้รอบรู้ ในรูปของการวิเคราะห์ความแปรปรวนดัชนีนี้แสดงความสัมพันธ์หาได้จาก

$$\mu_c^2 = \frac{SSB}{SSB + SSW}$$

เมื่อ SSB คือผลบวกกำลังสองระหว่างกลุ่ม

SSW คือผลบวกกำลังสองภายในกลุ่ม

แบริสกล่าวว่า สำหรับการแจกแจงแบบสมมาตรค่า μ_c^2 จะมากที่สุดเมื่อ $c = \mu$

สำหรับวิธีการนี้ ชัยโตเว็คกล่าวว่า สมประสิทธิ์ของแบริสมีความไวต่อสัดส่วนของกลุ่มที่เป็นผู้รอบรู้และผู้ไม่รอบรู้

¹Subkwick, "Estimating Reliability from...", pp 205-266. Citing C.W.Harris, "An Index of Efficiency for Fixed-Length Master Tests," Paper Present at the Annual Meeting of the American Educational Research Association (Chicago: 1972, (a).

มาแชล และ เฮอร์ทซ์¹(1975) เสนอวิธีที่ใช้การบริหารการสอบ 1 ครั้ง แจกแบ่งแบบ
สอบเป็นแบบสอบ 2 ฉบับที่ชูนานกัน หากค่าเฉลี่ย แล้วใช้สูตรสเปียร์แมน - บราวน์ หา
ความเที่ยงของแบบสอบทั้งฉบับ มารแชลกล่าวว่า สัมประสิทธิ์นี้จะเพิ่มขึ้นหรือลดลงอย่างเหมาะสม
เมื่อคะแนนจุดตัด ถูกกำหนดที่จุดต่าง ๆ ของคะแนนที่ได้

การวิเคราะห์รายข้อ

การวิเคราะห์รายข้อของแบบสอบอิงเกณฑ์แตกต่างจากแบบสอบอิงกลุ่ม โดยที่แบบสอบ
อิงกลุ่มใช้วิธีแบ่งผู้สอบเป็นกลุ่มสูง กลุ่มต่ำ แต่แบบสอบอิงเกณฑ์ใช้หลักการก่อนเรียนกับหลังเรียน
ผู้เสนอแนวคิดนี้เกี่ยวกับการวิเคราะห์รายข้อของแบบสอบอิงเกณฑ์ คือ

ชู² (1971) เสนอ D % ซึ่งเป็นความแตกต่างระหว่างระดับความยากของข้อสอบก่อน
เรียนและหลังเรียนที่ได้มาจากการวิเคราะห์แบบอิงกลุ่ม

คริสตินและเพลยูสัน³ (1974) เสนอดัชนีเอส (Sensitivity Index) โดยมีสูตรว่า

¹Ibid.,p.266. Citing J.L. Marshall and E.H.Haertal,"A Single -
Administration Reliability Index for Criterion-Referenced Tests:The Mean
Split-Half Coefficient of Agreement,"Paper Presented at the Annual Meeting
of the American Educational Research Association(Washington D.C.:1975).

²Thomas M.Haladyna,"Effects of Different Samples on Item and Test
Charac teristics of Criterion-Referenced Test,"Journal of Educational Mea-
surement 11(1974):94. Citing Hsu,Empirical Data on Criterion-Referenced Tests
Paper Presented at the Annual Meeting of the American Educational Research
Association , New York : 1971.

³สมศักดิ์ สีนุระเวชญ์, " การประเมินผลแบบอิงกลุ่มและอิงเกณฑ์,W.J.Krypin, and
Feldhusen,Dereloping Classroom Test(Minneapolis,Minnesota :Burgess Publishing
Company,1974).

$$S = \frac{R_{\text{pos}} - R_{\text{pre}}}{T}$$

เมื่อ R_{pos} คือจำนวนผู้ตอบถูกหลังจากเรียน
 R_{pre} คือจำนวนผู้ตอบถูกก่อนเรียน
 T คือจำนวนผู้เข้าสอบทั้งหมด

จากสูตรนี้จะเห็นว่า ถ้านักเรียนทุกคนทำข้อสอบผิดก่อนสอนและทำถูกหลังจากการสอน คำนวณค่า S จะเป็น 1 ข้อที่มีค่านี้อาจแสดงถึงความไวในการวัดผลการสอน งานวิจัยที่เกี่ยวข้อง

เอมริค¹ (1971) ทำการศึกษาโดยใช้ข้อมูลจากแบบสอบ IPI Math Placement "Numerical D" ซึ่งมีความยาว 16 ข้อ คำนวณหาเกณฑ์ที่เหมาะสมโดยใช้สูตร

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

เมื่อ K คือคะแนนจุดตัดแสดงในรูปร้อยละ
 n คือความยาวของแบบสอบ
 α คือค่าประมาณความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1
 β คือค่าประมาณความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 2
 RR คืออัตราส่วนของความรุนแรงของความคลาดเคลื่อนประเภทที่ 2 ต่อความคลาดเคลื่อนประเภทที่ 1

ค่า α และ β หาได้จาก

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

เมื่อ r_i คือค่าความเที่ยงของข้อระหวังกำนวนจากสูตรของสเปียร์แมน - บราวน์ เมื่อ กำหนดให้ $\beta = 3\alpha$ และ $RR = 10$ คำนวณค่าต่าง ๆ แทนในสูตรจะได้

53.5 % ค่าของคะแนนที่ใกล้เคียงที่สุด คือ 9 คะแนน หรือคิดเป็น 56.25 %

¹John A. Emrick, "An Evaluation Model...", pp. 321 - 326.

สวามินาธานแฮมเมิลตันและอัลกินา¹ (1975) ได้แสดงการหาเกณฑ์ในการจัดประเภท ผู้สอบโดยวิธีพิจารณาความสูญเสียที่น้อยที่สุด (Threshold loss) ในการตัดสินด้วยวิธีของเบส์ ใช้ข้อมูลจากแบบสอบที่มีความยาว 10 ข้อ ทดสอบกับนักเรียน 25 คน กำหนดคะแนนจุดตัด $T_0 = .8$ ความสูญเสียเนื่องจากความคลาดเคลื่อนในการยอมรับที่ผิด (1_{12}) เป็น 1 หน่วย และความสูญเสียเนื่องจากความคลาดเคลื่อนในการปฏิเสธที่ผิด (1_{21}) เป็น 2 หน่วย

เมื่อแบ่งผู้สอบตามระดับการรอบรู้เป็น 2 ระดับ ค่าความสูญเสียจะน้อยที่สุดเมื่อตัดสินให้ ผู้ที่ทำข้อสอบได้ 9 ข้อขึ้นไปเป็นผู้รอบรู้และผู้ที่ได้ต่ำกว่า 9 ข้อเป็นผู้ไม่รอบรู้ และเมื่อแบ่งผู้สอบเป็น 3 ระดับ กำหนดจุดตัด เป็น .6 และ .8 พบว่าค่าความสูญเสียจะน้อยที่สุดเมื่อตัดสินให้ผู้ที่ทำได้ 10 ข้อเป็นผู้รอบรู้ ทำได้ 7 - 9 ข้อ เป็นผู้รอบรู้เป็นบางส่วนคือต้องสอนทบทวนอีกเล็กน้อย ส่วน ผู้ที่ได้ต่ำกว่า 7 ข้อ เป็นผู้ที่ไม่รอบรู้ต้องเรียนใหม่

แฮมเบิลตัน และ สวามินาธาน² (1976) ได้เปรียบเทียบวิธีต่าง ๆ ในการประเมิน ความรอบรู้ของนักเรียนคือ

1. วิธีหาสัดส่วนความถูกต้อง (Proportion-Correct Score Estimate)
2. วิธีคลาสสิกแบบที่ 2 (Classical Model II Estimate)
3. การประมาณค่าโดยวิธีของเบส์แบบที่ 2 (Bayesian Model II Estimate)
4. การประมาณค่ามัธยฐานของแต่ละคนโดยวิธีของเบส์ (Bayesian Marginal Mean Estimate)
5. การประมาณค่ามัธยฐานของแต่ละคนโดยวิธีใหม่ (Modified Marginal Mean Estimate)

องค์ประกอบที่นำมาพิจารณาคือความเป็นเอกพันธ์ของกลุ่ม ข้อมูลเดิมเกี่ยวกับตัวผู้สอบ คะแนนจุดตัด ขนาดกลุ่มตัวอย่าง และความยาวของแบบสอบ จากการใช้คะแนนจุดตัดเป็น .8 ขนาดกลุ่มตัวอย่างเป็น 15 25 และ 50 คน ความยาวของแบบสอบเป็น 8 10 และ 20 ข้อ

¹H.Swaminathan, R.K Hambleton and J-Algina, "A Bayesian Decision-Theoretic...", pp. 87 - 97.

²R.K Hambleton, Leah R.Hutten and H.Swaminathan, "A Comparison of Several Method for Assessing Student Mastery in Objectives-Based Instructional Programs" Journal of Exerimental Educational 45(1976): 57 - 64.

เปรียบเทียบ คะแนนความรอบรู้ที่แท้จริงกับคะแนนความรอบรู้ที่ได้จากการประมาณด้วยวิธีต่าง ๆ พบว่าขนาดกลุ่มตัวอย่างมีผลน้อยมากในทุก ๆ วิธี วิธีที่ใหม่ลึที่สุดคือการประมาณตามมิติของแต่ละคนโดยวิธีใหม่ แต่วิธีนี้ต้องมีรายละเอียดที่แน่นอนเกี่ยวกับข้อมูลเดิมของระดับความรอบรู้และการแจจแจงของคะแนนระดับความรอบรู้ที่แท้จริงของผู้สอบซึ่งในทางปฏิบัติเป็นไปได้ยาก วิธีที่สามารถนำไปใช้ได้คือวิธีการของเบส ซึ่ง เป็นวิธีที่ไม่จำเป็นต้องทราบคะแนนระดับความรอบรู้ที่แท้จริงของผู้สอบ

เครเฮน¹ (1974) ศึกษาการใช้เทคนิคการเลือกข้อกระทงต่าง ๆ เพื่อแบ่งเป็นแบบสอบ
คุณาน คือ

1. วิธีของคอกซ์ - แวกาส (Cox - Vagas) เรียงลำดับข้อกระทงตามความแตกต่างของความยากระหว่างการสอบก่อนเรียนและหลังเรียนของผู้สอบกลุ่มเดียว
2. วิธีของเบรนนัน (Brennan) เรียงลำดับข้อกระทงตามความแตกต่างของความยากระหว่างการสอบก่อนเรียนของกลุ่มที่ทำกาสอบทั้งก่อนเรียนและหลังเรียนและคำตอบของกลุ่มที่สอบเฉพาะหลังเรียน
3. เรียงตามสัดส่วนในการตอบของผู้สอบหลังจากเรียน
4. เรียงตามการเลือกของผู้สร้างแบบสอบ
5. เรียงตามค่าสัมประสิทธิ์สหสัมพันธ์แบบพอยท์ไบซี เรียบของการสอบหลังเรียน
6. เรียงลำดับตามการสุ่ม

ผลจากการเปรียบเทียบในด้านความเที่ยงและความตรงพบว่าแต่ละวิธีให้ผลต่อค่าความเที่ยงไม่แตกต่างกันมากนัก ในด้านความตรงพบว่าวิธีของคอกซ์ - แวกาสและวิธีของเบรนนันให้ค่าความตรงสูงสุด

¹ Kelvin D. Crehan, "Item Analysis for...", pp. 255 - 262.

ฮวน¹ (1976) ศึกษาค่าความเที่ยงของแบบสอบในรูปของสัมประสิทธิ์แคปลา โดยมิชอบ คกกลงว่าการแจกแจงของคะแนนที่ได้จากการสอบเป็นแบบไบนอมิเยล ทำการทดสอบครั้งเดียวโดยใช้แบบสอบเลขคณิตที่มีความยาว 5 ข้อ ทดสอบกับนักเรียน 91 คน กำหนดคะแนนจุดตัด = 4 ไคค่าแคปลา = .52 ในกรณีที่ข้อกระทงมีจำนวนมาก ฮวนใช้แบบสอบเลขคณิตที่มีความยาว 15 ข้อทดสอบกับนักเรียน 96 คน ใช้คะแนนจุดตัด = 12 ไคค่าแคปลา = .57 นอกจากนี้ได้พิจารณาองค์ประกอบที่มีผลต่อค่าแคปลาพบว่าในด้านความเป็นวิวิธพันธ์ของคะแนนสอบค่าแคปลา มีสหสัมพันธ์ทางบวกกับความแปรปรวนของคะแนนสอบ ในด้านความยาวของแบบสอบค่าแคปลาจะเพิ่มตามความยาวของแบบสอบโดยอัตราการเพิ่มจะลดลงเมื่อความยาวของแบบสอบเพิ่มขึ้น นอกจากนี้ค่าแคปลาจะเพิ่มขึ้นเมื่อคะแนนจุดตัดสูงขึ้น แต่เมื่อคะแนนจุดตัดสูงมาก ๆ ค่าแคปลาจะลดลง ในที่เดียวกันซิปโคเวียค² ได้เสนอวิธีหาค่าความเที่ยงโดยใช้การบริหารการสอบเพียง 1 ครั้งเช่นกัน ใช้แบบสอบที่มีความยาว 5 ข้อทดสอบกับนักเรียน 10 คนกำหนดคะแนนจุดตัด = 4 ประมาณค่าสัดส่วนของคะแนนจริงจาก

$$p_i = \alpha_{21/x} \left(\frac{x_i}{n} \right) + (1 - \alpha_{21/x}) \left(\frac{Mx}{n} \right)$$

$$\alpha_{21/x} = \frac{n}{n-1} \left[1 - \frac{Mx(n-Mx)}{nS_x^2} \right]$$

ไคค่า $P_c = .75$ เปรียบเทียบกับค่าความเที่ยงที่ได้จากการบริหารการสอบ 2 ครั้งคำนวณโดยใช้สูตรของแฮมเมลตันและโนวิก (1973) ไคค่า $P_o = .80$

ในกรณีที่แบบสอบมีความยาวและกลุ่มตัวอย่างมีขนาดใหญ่ได้ใช้แบบสอบที่มีความยาว 25 ข้อทดสอบกับนักเรียน 69 คนเปรียบเทียบค่า \hat{p}_c ที่ได้จากการแจกแจงแบบซิมเปิลไบนอมิเยล (Simple binomial) และที่ได้จากการแจกแจงแบบคอมพอนด์ไบนอมิเยล (Compound binomial) กับค่า \hat{P}_o พบว่าค่า P_o สูงกว่า \hat{P}_c จากการแจกแจงแบบซิมเปิลไบนอมิเยลด้วยค่ามีเกิน 6.5 % และสูงกว่า \hat{P}_c จากการแจกแจงแบบคอมพอนด์ไบนอมิเยลด้วยค่ามีเกิน 3.5 %

¹Huynh. Huynh, "On the Reliability of...", pp. 252 - 263.

²Subkoviak, "Estimating Reliability from...", pp 267-275

ผลของคะแนนจุดตัดที่มีต่อค่าความเที่ยงทั้ง 3 มีลักษณะเดียวกันคือ ความเที่ยงมีค่าสูงเมื่อคะแนนจุดตัดที่สูงที่สุดและต่ำสุด และมีค่าต่ำเมื่อคะแนนจุดตัดอยู่ตรงกลาง ๆ

อัลกินาและ โน¹ (1978) ได้ศึกษาความเชื่อถือได้ของการประมาณค่าความเที่ยงในการบริหารการสอบครั้งเดียวจากสูตรของซัทโคเว็ค โดยใช้เวลาประมาณคะแนนจริง 2 ค่าคือ p_i และ \hat{T}_i โดยที่

$$\hat{T}_i = \hat{\beta} p_i + (1 - \hat{\beta}) \hat{\mu}_p$$

เมื่อ $\hat{\beta}$ คือสัมประสิทธิ์จากสูตรคูเคอร์ - ริชาร์ดสันสูตรที่ 20
 p_i คือคะแนนสัดส่วนความถูกต้องของผู้เข้าสอบที่
 $\hat{\mu}_p$ คือมัธยัม เลขคณิตของคะแนนสัดส่วนความถูกต้อง

นอกจากนี้ยังพิจารณาองค์ประกอบในด้านคะแนนจุดตัด ความเที่ยงของแบบสอบและจำนวนผู้เข้าสอบพบว่าค่าประมาณของสัมประสิทธิ์ความเที่ยงในการตัดสิน (p_e) จะคลาดเคลื่อนมากขึ้นเมื่อคะแนนจุดตัดมีค่าใกล้ $n/2$ โดยมีค่าเป็นบวกเมื่อใช้ p_i และเป็นลบเมื่อใช้ \hat{T}_i เมื่อส่วนเบี่ยงเบนระหว่างคะแนนจุดตัดกับ $n/2$ เพิ่มค่าสัมบูรณ์ของความคลาดเคลื่อนจะลดลงจนเครื่องหมายเปลี่ยนไป แล้วเพิ่มอีกจนในที่สุดจะลดลงการเปลี่ยนแปลงนี้เห็นได้ชัดในแบบสอบที่ยาวที่สุด (คือ 20 ข้อ) ในด้านความเที่ยงของแบบสอบ (p^{2xt}) ความคลาดเคลื่อนของ p_i มีแนวโน้มที่จะลดลงเมื่อ p^{2xt} เพิ่มส่วนของ \hat{T}_i มีแนวโน้มเพิ่มขึ้นเมื่อ p^{2xt} เพิ่ม อัลกินาและคณะแนะนำสำหรับแบบสอบที่มีค่า p^{2xt} เกิน .50 ควรใช้ \hat{T}_i ประมาณค่าคะแนนจริง และถ้า \hat{T}_i มีค่ามากอาจใช้ค่าเฉลี่ยระหว่าง p_i และ \hat{T}_i เป็นค่าประมาณของคะแนนจริงสำหรับขนาดของกลุ่มตัวอย่างพบว่าความผันแปรของค่าประมาณทั้ง 2 ตัวจะเพิ่มขึ้นเมื่อจำนวนผู้สอบลดลงแต่ไม่มากนัก นอกจากนี้ยังพบว่าการใช้การแจกแจงแบบซิมเพิลไบโนเมียลให้ผลไม่แตกต่างจากการใช้การแจกแจงแบบคอมพิวเตอร์ไบโนเมียล ในปีเดียวกันนี้ ซัทโคเว็ค² ได้เปรียบเทียบค่าความเที่ยงที่ได้จากการคำนวณด้วยวิธีต่าง ๆ 4 วิธีคือ

¹James Algina and Michael J. Noe, "A Study of Subkoviak's Single Administration Estimate of the Coefficient of Agreement Using Two-true Score Estimates," Journal of Educational Measurement 15(1978):101-109.

²Michael J. Subkoviak, "Empirical Investigation of Procedures for Estimating Reliability for Mastery Tests," Journal of Educational Measurement 15(1978): 111-115.

1. วิธีของสแวมมินาธาน แยมเบิลตัน และอัลกินา
2. วิธีของฮวน
3. วิธีของมาร์แชล และแอสเทล
4. วิธีของซิมโคเวียต

วิธีของสแวมมินาธานและคณะ เป็นวิธีเดียวที่ได้จากการบริหารการสอบ 2 ครั้ง นอกนั้นใช้การบริหารการสอบเพียงครั้งเดียว ในการเปรียบเทียบไข้อมูลจากการให้นักเรียน 1586 คน ทำแบบสอบความถนัดทางวิชาการซึ่งเป็นแบบสอบคุณชนาน มีความยาว 10 , 30 และ 50 ข้อ แบ่งนักเรียนออกเป็น 50 กลุ่ม กลุ่มละ 30 คน โดยวิธีสุ่มแบบแทนที่ เมื่อใช้เกณฑ์ตัดสินเป็นร้อยละ 50, 60 , 70 และ 80 พบว่าค่าความคลาดเคลื่อนมาตรฐานจะลดลงเมื่อแบบสอบยาวขึ้น เมื่อเปรียบเทียบทั้ง 4 วิธีแล้ว วิธีของสแวมมินาธานและคณะมีค่าความคลาดเคลื่อนมาตรฐานสูงกว่าวิธีอื่น เมื่อพิจารณาทุกด้านแล้ว วิธีของฮวนให้ค่าประมาณที่ถูกต้องที่สุด

สมหวัง พิธิยานุวัฒน์¹ ได้ประยุกต์วิธีของเบอร์คมาหาคะแนนจุดตัดของแบบสอบวิชาสถิติเรื่อง: สหสัมพันธ์ จำนวน 14 ข้อ ไข้อมูลจากกลุ่มตัวอย่างซึ่งเป็นนิสิตปริญญาโท แผนกวิชาจิตวิทยา จำนวน 18 คน ทำแบบสอบก่อนเรียนและหลังเรียน นำคะแนนที่ได้จากการสอบมาเขียนกราฟเพื่อดูการกระจายของคะแนนที่ได้จากการสอบก่อนเรียนและหลังเรียน การกำหนดคะแนนจุดตัดพิจารณาจากจุดที่กราฟแสดงการแจกแจงของคะแนนในการสอบทั้ง 2 ครั้งตัดกัน เป็นจุดตัดขั้นต่ำ และประมวลความคิดของบรูม บล๊อค และเกลเซอร์ เกี่ยวกับการกำหนดเกณฑ์ความรอบรู้ของผู้เรียนมาหาคะแนนจุดตัดขั้นสูง จากจุดตัดขั้นต่ำและจุดตัดขั้นสูงนำมาหาคะแนนจุดตัดที่เหมาะสมของแบบสอบวิชาสถิติศาสตร์ ผลจากการศึกษาได้ 70 % เป็นจุดตัดขั้นสูงอย่างต่ำและ 58.6 % เป็นจุดตัดขั้นต่ำอย่างสูง

¹สมหวัง พิธิยานุวัฒน์, "จุดแบ่งที่พอดีของแบบสอบอิงเกณฑ์ทางสถิติศาสตร์ การศึกษาเบื้องต้น," ใน เอกสารประกอบการประชุมสัมมนาไปเชียงใหม่, (กรุงเทพมหานคร สำนักพิมพ์จุฬาลงกรณ์มหาวิทยาลัย, 2520), หน้า 13.

กาญจนา วัฒนสุนทร¹ ได้สร้างแบบสอบอิงเกณฑ์วิชาคณิตศาสตร์ชั้นมัธยมศึกษาปีที่ 1 เรื่องสมการ จำนวน 40 ข้อ แยกเป็น 4 ฉบับ ฉบับละ 10 ข้อ ทำการวิเคราะห์คุณภาพของแบบสอบโดยใช้ดัชนีเอสควานวดค่าอำนาจจำแนกส่วนค่าความยากค่าความยากจากสัดส่วนของผู้ตอบข้อกระทงนั้น ๆ ถูก ในการหาเกณฑ์การตัดสินที่เหมาะสมได้ตกลงกำหนดเกณฑ์การตัดสินเป็น .6, .7, และ .8 แล้วหาค่าความเที่ยงและความตรงของการตัดสินโดยใช้สูตรของจิตวิงส์ตัน¹ และสูตรของคาร์เวอรัลตามลำดับ พบว่าเมื่อใช้ .6 เป็นคะแนนเกณฑ์จะได้ค่าความเที่ยงและความตรงในการตัดสินสูงสุด

จากทฤษฎี แนวคิดและงานวิจัยเหล่านี้ ผู้วิจัยได้นำมาประกอบกันเป็นแนวทางในการศึกษาครั้งนี้ กล่าวคือใช้วิธีการของ เบสส์ โดยอาศัยวิธีการของ เบอรัลมากำหนดคะแนนจุดตัด และอาศัยแนวคิดและข้อเสนอแนะของนักการศึกษาหลาย ๆ ท่าน เช่น เอมริค มิกแมน แอม-เบิร์ตสัน และเบอรัล เป็นต้น มาใช้ในการพิจารณาค่าความสูญเสียประเภทที่ 1 และประเภทที่ 2 ทั้งนี้ผู้วิจัยเชื่อว่าจะช่วยให้ผลที่ได้มีความน่าเชื่อถือและสอดคล้องกับสภาพความเป็นจริงมากยิ่งขึ้น

ศูนย์วิจัยทรัพยากร จุฬาลงกรณ์มหาวิทยาลัย

¹กาญจนา วัฒนสุนทร, "การสร้างแบบสอบอิงเกณฑ์วิชาคณิตศาสตร์" (วิทยานิพนธ์ปริญญาโทบริหารศึกษาด้านจิต แผนกวิชา วิชาการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, 2521.