บทที่ 2

วรรณคดีที่เกี่ยวข้อง

ผู้วิจัยจะรายงานถึงวรรณคดีที่เกี่ยวข้องเกี่ยวกับข้อเปรียบเทียบระหว่างการประเมินแบบอิงเกณฑ์และการประเมินแบบอิงกลุ่ม โดยเฉพาะการประเมินแบบอิงเกณฑ์ซึ่งจำเป็นจะต้องใช้ข้อทดสอบแบบอิงเกณฑ์มาวัด และจะสรุปถึงวิธีการต่าง ๆ ที่จะหาจุดตัดซึ่งจะนำมาใช้เป็นเกณฑ์ในการประเมิน ในตอนสุดท้ายจะว่าด้วยความเที่ยงของแบบสอบแบบอิงเกณฑ์

สำหรับการประเมินแบบอิงเกณฑ์และการประเมินแบบอิงกลุ่ม ผลการเปรียบเทียบในแง่ต่าง ๆ มีดังนี้ คือ

1. หลักในการวัด

การวัดแบบอิงกลุ่ม (Norm-Referenced Measurement) จะขึ้นอยู่กับเกณฑ์มาตรฐานสัมพัทธ์ (Relative Standard) หรือพฤติกรรมของกลุ่มเป็นสำคัญ จึงมุ่งหาความแตกต่างระหว่างผู้เรียน เพราะเชื่อว่าความสามารถของบุคคลกลุ่มใด ๆ ในเรื่องใดเรื่องหนึ่งนั้นมีไม่เท่ากัน บางคนมีความสามารถเด่น บางคนมีความสามารถด้อย แต่บุคคลที่มีความสามารถเด่นและด้อยนี้จะมีน้อย ส่วนใหญ่จะมีความสามารถปานกลาง การกระจายของความสามารถของผู้เรียนจึงคล้ายกับโค้งรูประฆัง (Normal Curve)[1] ดังนั้นจึงยึดคนส่วนใหญ่หรือกลุ่มเป็นหลักในการเปรียบเทียบโดยพิจารณาคะแนนกับคะแนนของบุคคลอื่นภายในกลุ่ม ดังนั้นจึงเป็นแรงกระตุ้นที่จะก่อให้เกิดการเรียนรู้แบบแข่งขันระหว่างผู้เรียนด้วยกัน

---

[1]สมศักดิ์ สินธุระเวชญ์, " การประเมินผลแบบอิงกลุ่มและอิงเกณฑ์," วารสารวิจัยทางการศึกษา 8(มิถุนายน 2521) หน้า 74.

การวัดแบบอิงเกณฑ์ (Criterion-Referenced Measurement) สืบเนื่องมาจากหลักการเรียนรู้แบบ การเรียนเพื่อรอบรู้ (Mastery Learning) ของบลูมที่กล่าวว่า "การเรียนรู้ทั้งหลายควรจะเป็นการเรียนเพื่อรอบรู้ในเนื้อเรื่องต่าง ๆ ครูจึงต้องการให้นักเรียนทุกคนบรรลุสู่ความเป็นผู้รอบรู้หรือควรจะเป็น 80 - 90 % ของเนื้อหาที่เรียนรู้"[1] การประเมินผลแบบนี้จึงถือว่าเนื้อหาวิชาที่สอนจะเป็นสิ่งสำคัญและจำเป็นที่นักเรียนควรจะรู้ ดังนั้นจึงขึ้นอยู่กับเกณฑ์มาตรฐานที่ให้ความหมายไว้ชัดเจนล่วงหน้า มุ่งที่จะประเมินว่าผู้เรียนสามารถหรือไม่สามารถทำอะไรได้บ้างมากกว่าที่จะไปเปรียบเทียบกับกลุ่ม จึงผลักดันให้เกิดการเรียนรู้แบบเกื้อกูลกัน

003972

2. วัตถุประสงค์ในการประเมิน

แบบสอบอิงกลุ่ม สร้างขึ้นเพื่อใช้จัดอันดับผลสัมฤทธิ์ทางการเรียนของนักเรียนจากสูงไปต่ำ โดยมีจุดมุ่งหมายเพื่อเปรียบเทียบความแตกต่าง เพื่อตัดสินใจคัดเลือกเพื่อการจัดกลุ่มหรือการตัดเกรดเป็นต้น ดังนั้นลักษณะของแบบสอบจึงต้องมีอำนาจจำแนกให้ช่วงคะแนนกว้าง และจะต้องไม่ใช่ข้อสอบที่เด็กส่วนใหญ่ทำถูก

แต่จากการที่เกลเซอร์ (Glaser)[2] มีความเห็นว่า การวัดผลสัมฤทธิ์ควรจะเป็นเรื่องของการมีความรู้ความสามารถที่เรียงจากไม่มีเลยไปจนถึงมีอย่างสมบูรณ์ และระดับความสัมฤทธิ์ผลของผู้เรียนที่วัดได้จากพฤติกรรมที่เขาแสดงออกก็จะตกอยู่ในช่วงนี้ การจะพิจารณาว่าระดับนั้นน่าพอใจหรือไม่ก็ควรจะเทียบกับมาตรฐานของพฤติกรรมที่กำหนดไว้ ซึ่งจะเป็นพฤติกรรมขั้นต่ำหรือพฤติกรรมที่เป็นเกณฑ์นั่นเอง ลักษณะของแบบสอบก็จะไม่นำเอาระดับความยากและอำนาจจำแนกมาใช้เป็นเกณฑ์ในการเลือกข้อสอบ[3]

---

[1] เรื่องเดียวกัน, หน้า 75.

[2] Glass, "Standard and ...," p.240.

[3] สมศักดิ์ สินธุระเวชญ์, "การประเมินผลแบบอิงกลุ่มและอิงเกณฑ์," หน้า 77.

3. การแปลผลของคะแนน

การสอบแบบอิงกลุ่ม เราใช้กลุ่มเป็นหลักในการเปรียบเทียบว่าใครมีความสามารถมากน้อยปานใด ส่วนการสอบแบบอิงเกณฑ์จะเทียบกับเกณฑ์ซึ่งแปลผลออกมาในรูปที่ว่านักเรียนแต่ละคนสามารถทำอะไรได้บ้างเมื่อเทียบกับเกณฑ์ที่เราตั้งไว้

นอกจากข้อแตกต่างระหว่างการประเมินแบบอิงกลุ่มและแบบอิงเกณฑ์ดังกล่าวแล้ว ได้มีการพิจารณาถึงข้อเสียของการประเมินแบบอิงกลุ่ม[1] กล่าวคือ

ก. มาตรฐานจะเป็นไปตามยถากรรม ถ้าทุกคนในกลุ่มเรียนดีมาตรฐานจะสูง แต่ถ้าทุกคนในกลุ่มเรียนไม่ดีมาตรฐานก็จะต่ำ

ข. เราไม่สามารถเปรียบเทียบคนเก่งในกลุ่มหนึ่ง หรือรุ่นหนึ่งกับอีกกลุ่มหนึ่ง หรืออีกรุ่นหนึ่งได้

ค. ไม่สามารถบอกให้เราทราบว่านักเรียนทั้งหลุ่มมีมาตรฐานดีขึ้นหรือลดลง และยังไม่อาจบอกได้ว่า กลุ่มนี้เรียนวิชาหนึ่งดีกว่าหรือเลวกว่าวิชาอื่น ๆ

ง. คนที่ประเมินแล้วว่าเก่ง อาจเป็นเพราะเขาเก่งจริงหรืออาจเป็นเพราะคนอื่นในกลุ่มมีผลการเรียนต่ำมากก็ได้ ในทำนองตรงกันข้ามถ้าผู้สอบเป็นผู้ที่ได้มาจากการคัดเลือกแล้วเป็นอย่างมาก ถ้าประเมินผลแบบอิงกลุ่มนี้ก็จะมีคนเก่งที่ถูกประเมินว่าไม่เก่งก็ได้

แอราเซียน และมาดอส[2] วิจารณ์ถึงเรื่องการประเมินแบบอิงกลุ่มไว้ว่า

1. ความหมายของการวัดและการใช้แบบสอบให้ข้อมูลไม่เพียงพอและไม่เหมาะสมกับเนื้อหาบางประการที่ต้องการวัด

---

[1] สมโพธิ พุกกะเวส. "การให้ระดับคะแนน," หน้า 78 - 79.

[2] P.W. Airasian, and G.F. Madaus, "Criterion-Referenced Testing in the Classroom." Incrucial Issues in Testing (Mc Cutchan Publishing Corporation 1974), pp. 76 - 77.

2. การให้ระดับคะแนน ทำให้จุดมุ่งหมายของการศึกษาเปลี่ยนไปกลายเป็นการแข่งขันเพื่อให้ได้ระดับคะแนนสูง ๆ และระดับคะแนนก็ไม่ได้ให้ข้อมูลว่านักเรียนทำอะไรได้บ้าง บอกแต่เพียงว่าเขาดีหรือด้อยกว่าคนอื่น ๆ ในกลุ่มของเขาเท่านั้น

3. จากวิวัฒนาการการเคลื่อนไหวทางเทคนิคการสอนต่าง ๆ ทำให้พบว่าการวัดแบบอิงกลุ่มไม่เพียงพอที่จะใช้ประเมินพฤติกรรมเป็นรายบุคคลหรือประเมินวิธีการสอนที่ใช้ว่าดีหรือไม่เพียงใด

4. นักการศึกษาส่วนมากเชื่อว่า คนทุกคนหรือคนส่วนมากสามารถที่จะเรียนรู้หรือสัมฤทธิ์ในวิชาต่าง ๆ ได้ ถ้ามีวิธีสอนที่ดีกว่า จึงทำให้การทดสอบเปลี่ยนจากการเปรียบเทียบในกลุ่ม เป็นการตรวจสอบและให้รางวัลในรูปของการกระทำ เช่น ถ้าทุกคนทำได้ถึงเกณฑ์ ทุกคนก็อาจได้ระดับคะแนน "A" เป็นต้น

จากแนวคิดข้างต้นจะเห็นได้ว่า การวัดแบบอิงกลุ่มไม่เหมาะที่จะใช้วัดสัมฤทธิ์ผลเสมอไป เนื่องจากให้ข้อมูลเราได้ไม่ครบตามต้องการ เช่น ในกรณีที่เราต้องการจะพิจารณาว่านักเรียนได้เรียนรู้ในสิ่งที่เราสอนหรือไม่ เป็นต้น ซึ่งถ้าจะพิจารณาในแง่นี้เราก็จะต้องคำนึงถึงความแตกต่างระหว่างบุคคล (Individual differences) ในตัวผู้เรียนเป็นสำคัญ เพื่อเราจะได้นำมาใช้ค้นหายุทธวิธีที่จะพัฒนาผู้เรียนให้ถึงขีดสูงสุดนี้ ซึ่งแสดงว่ารอบรู้ในสิ่งที่เราสอนนั่นเอง

สำหรับการเรียนเพื่อรอบรู้นี้ แครอล (Carroll, 1963) เสนอโมเดลซึ่งมีสาระสำคัญว่า ถ้าเราวัดความถนัดบางวิชา เช่น คณิตศาสตร์ หรือวิทยาศาสตร์ การกระจายของคะแนนจะเป็นโค้งปกติ และถ้าเราให้การสอนที่เหมือนกันทั้งในแง่ คุณภาพ ปริมาณ และเวลา แล้ววัดผลสัมฤทธิ์ในการเรียนก็จะได้การกระจายของคะแนนสัมฤทธิ์ผลเป็นโค้งปกติเช่นกัน และค่าสหสัมพันธ์ระหว่างคะแนนความถนัดและสัมฤทธิ์ผลจะมีค่าสูง ในทางตรงข้ามถ้าคะแนนความถนัดมีการกระจายแบบปกติ แต่เราให้ชนิด คุณภาพ และเวลาในการเรียนให้เหมาะสมกับคุณลักษณะและความต้องการของนักเรียนแต่ละคนแล้ว นักเรียนส่วนใหญ่จะรอบรู้ในวิชานั้น และค่าสหสัมพันธ์ระหว่างคะแนนความถนัดและสัมฤทธิ์ผลจะเข้าใกล้ศูนย์ สรุปได้ว่าความถนัดจะเป็นแต่เพียงตัวทำนายอัตราการเรียนรู้มากกว่า

ระดับความซับซ้อนของการเรียนรู้โดยอัตราในการเรียนรู้ของแต่ละคนมีไม่เท่ากัน แต่ทุกคนก็สามารถบรรลุสู่ความเป็นผู้รอบรู้ได้ ถ้าให้เวลาตามความต้องการของแต่ละบุคคลอย่างเพียงพอ

ตัวแปรที่จะมีผลต่อการเรียนเพื่อรอบรู้อีกอย่างหนึ่ง คือ คุณภาพของการสอน แครอลคิดว่าเด็กแต่ละคนต้องการวิธีการและคุณภาพของการสอนที่จะบรรลุสู่ความเป็นผู้รอบรู้ได้ไม่เหมือนกัน ดังนั้นการประเมินคุณภาพของการสอนจึงควรประเมินในแง่ของประสิทธิภาพที่มีต่อผู้เรียนเป็นรายบุคคลมากกว่าเป็นรายกลุ่ม

ในเรื่องนี้ บลูม (Bloom, 1968) ซึ่งมีชื่อเสียงเกี่ยวกับ "การเรียนเพื่อรอบรู้" (Mastery Learning) ก็เชื่อว่า ถ้าให้เวลาที่พอเพียง และให้การช่วยเหลือที่เหมาะสมและมีประสิทธิภาพแล้ว 95 % ของนักเรียนทั้งหมด จะสามารถเรียนวิชานั้นอยู่ในระดับสูงของการรอบรู้ บลูมกล่าวว่า "การเรียนรู้ทั้งหลายควรจะเป็นการเรียนเพื่อรอบรู้ในเนื้อเรื่องต่าง ๆ ครูย่อมต้องการให้นักเรียนทุกคนบรรลุสู่ความเป็นผู้รอบรู้ซึ่งควรเป็น 80 - 90 % ของเนื้อหาที่เรียนรู้"[1]

จากแนวคิดของแครอลและบลูมนี่เอง ทำให้เกิดการประเมินผลแบบอิงเกณฑ์ (Criterion-Referenced Evaluation) ซึ่งจะใช้หลังเรียนไปแล้วหน่วยหนึ่งหรือบทหนึ่ง เพื่อปรับปรุงการเรียนการสอน หรือที่เรียกว่า การประเมินผลความก้าวหน้า (Formative Evaluation) เครื่องมือที่จะใช้ประเมินคือ "แบบสอบอิงเกณฑ์" (Criterion-Referenced Test) เมื่อทดสอบไปแล้วก็จะทำการตัดสินว่านักเรียนคนใดรอบรู้ (ผ่าน) คนใดยังไม่รอบรู้ (ไม่ผ่าน) ในเนื้อหาที่เรียนโดยเทียบกับมาตรฐานของพฤติกรรม (เกณฑ์) ที่กำหนดไว้ก่อนแล้วโดยไม่ต้องไปเทียบกับคะแนนของผู้อื่น

---

[1]สมศักดิ์ สินธุระเวชญ์, "การประเมินผลแบบอิงกลุ่มและอิงเกณฑ์" วารสารวิจัยทางการศึกษา 8(มิถุนายน 2521) หน้า 75 อ้างถึงใน Benjamin S. Bloom, Handbook on Formative and Summative Evaluation of Student Learning. New York, 1971.

เกลเซอร์ (Glaser, 1962) แสดงความเห็นไว้ว่า "ความคิดที่เป็นพื้นฐานของการวัดผลสัมฤทธิ์นั้นเขาจะถือเอาช่วงของความรู้ที่ต่อเนื่องกันจากไม่มีเลยจนถึงสามารถปฏิบัติได้โดยสมบูรณ์ทั้งหมด และระดับผลสัมฤทธิ์ของแต่ละบุคคลที่สามารถวัดได้จากพฤติกรรมที่เขาแสดงออก เช่นในการสอบจะตกอยู่ ณ จุดใดจุดหนึ่งในช่วงนี้"[1] การที่จะพิจารณาว่าระดับนั้นเป็นที่น่าพอใจหรือไม่ควรจะเทียบกับมาตรฐานของพฤติกรรมที่กำหนดไว้ การวัดผลสัมฤทธิ์จึงควรหมายถึง การวัดพฤติกรรมที่คาดหวังหรือพฤติกรรมที่เป็นเกณฑ์ (Terminal of Criterion Behavior) ซึ่งกำหนดขึ้นสำหรับการเรียนการสอนช่วงใดช่วงหนึ่ง เพื่อแสดงความงอกงามในการเรียนรู้ของผู้เรียนและพฤติกรรมนี้ควรสามารถที่จะชี้และอธิบายงานเฉพาะที่นักเรียนผู้นั้นทำได้ก่อนที่เขาจะบรรลุถึงระดับของความรู้เหล่านี้ ข้อสอบที่จะใช้วัดพฤติกรรมที่คาดหวังจากการเรียนการสอนก็คงจะได้แก่แบบสอบอิงเกณฑ์ซึ่งสร้างขึ้นเพื่อให้ได้ผลจากการวัดที่สามารถแปลความหมายได้โดยตรงในรูปของมาตรฐานของพฤติกรรม หรือมาตรฐานของการปฏิบัติที่กำหนดไว้

จากข้อความข้างต้นนี้สันนิษฐานได้ว่า

1. ช่วงของความรู้จะต่อเนื่องกันตั้งแต่ไม่มีเลยจนถึงสามารถปฏิบัติได้สมบูรณ์ทั้งหมด

2. ระดับของความสามารถนั้น จะมีนักเรียนเฉพาะบางคนเท่านั้นที่สามารถบรรลุได้

นอกจากนี้เกลเซอร์ยังได้แนะว่า "ความสามารถที่เป็นช่วงต่อเนื่องกันนั้น จะมีอยู่เพียงจุดเดียวที่ความสามารถจะเปลี่ยนเป็นความไม่สามารถ"[2] แสดงให้เห็นว่าแบบสอบอิงเกณฑ์ควรที่จะต้องกำหนดคะแนนจุดตัด (Cut-off score) ระหว่างผู้ที่มีความสามารถออกจากไม่มีความสามารถหรือการแยกเป็นได้-ตก และการรอบรู้-ไม่รอบรู้

---

[1]Glass, "Standard and...," p. 240.

[2]Ibid.

เมเจอร์ (Mager, 1962) มีความเห็นสอดคล้องกับเกลเซอร์ ดังที่เขาได้กล่าวไว้ว่า

"ในวัตถุประสงค์แต่ละข้อนั้น ถ้าเราสามารถกำหนดความสามารถขั้นต่ำสุดที่เป็นที่ยอมรับ (Minimum acceptable performance) ไว้ เราก็จะมีมาตรฐานของการปฏิบัติ (Performance standard) ที่จะใช้ตรวจสอบโปรแกรมการสอนของเราได้ว่าประสบผลสำเร็จตามจุดมุ่งหมายที่ตั้งไว้หรือไม่ ดังนั้นจากข้อความที่เราตั้งไว้จึงต้องเติมคำบรรยายเกณฑ์ของความสำเร็จไว้ด้วย"[1]

โปแฟมและยูเซค (Popham and Husek, 1969) ได้เขียนบทความเกี่ยวกับแบบสอบอิงเกณฑ์ในรูปของคะแนนมาตรฐานไว้ดังนี้ "การวัดแบบอิงเกณฑ์ เป็นสิ่งที่จะใช้ค้นหาสถานะของแต่ละบุคคล โดยเทียบกับเกณฑ์บางอย่าง คือคะแนนมาตรฐาน"[2] และในปี 1973 โปแฟมได้เขียนยืนยันความสำคัญของคะแนนมาตรฐานไว้อีกว่า "มีวิธีการเขียนวัตถุประสงค์ซึ่งจะช่วยครูในการเตรียมการสอนและการประเมินผลการสอนได้ โดยการสร้างคะแนนมาตรฐานที่ระบุระดับต่ำสุดของผลสัมฤทธิ์ของนักเรียนไว้"[3]

ไทเลอร์ (Tyler, 1973) ได้แสดงตัวอย่างมาตรฐานของการปฏิบัติเพื่อกำหนดความรอบรู้ไว้ดังนี้

"นักเรียนควรจะแสดงตัวอย่างมากเพียงพอที่จะเป็นหลักฐานที่จะเชื่อถือได้ว่าเขาจำตัวอักษรที่เป็นพยัญชนะได้ สามารถออกเสียงอักษรแต่ละตัวได้เหมาะสม และจำคำธรรมดา ๆ ได้ 100 คำ นักเรียนจะแสดงความรอบรู้ในความรู้ ความสามารถ หรือทักษะเฉพาะ เมื่อเขาปฏิบัติได้ถูกต้องในเวลา 85 % ของเวลาทั้งหมด"[4]

---

[1] Glass, "Standard and...," p. 240.

[2] Ibid., p. 241.

[3] Ibid., p. 239.

[4] Ibid.

จากแนวคิดต่าง ๆ ข้างต้น พอจะสรุปได้ว่า คะแนนจุดตัดหรือเกณฑ์ ก็คือจำนวนข้อสอบที่น้อยที่สุดที่นักเรียนจะต้องทำได้ในการที่จะได้รับการตัดสินให้เป็นผู้รอบรู้ โดยที่ระดับความรอบรู้ (Mastery level) ก็คือระดับของผลสัมฤทธิ์ของผู้เข้าสอบซึ่งตัดสินจากการเปรียบเทียบคะแนนจากแบบสอบกับคะแนนเกณฑ์ที่ตั้งไว้ และอาจมีตั้งแต่ 2 ระดับขึ้นไป ทั้งนี้ขึ้นอยู่กับวัตถุประสงค์ในการประเมินเป็นสำคัญ เช่น ถ้าต้องการจำแนกบุคคลเป็น 2 ประเภทก็จะแบ่งได้เป็นผู้รอบรู้ (Mastery) และผู้ไม่รอบรู้ (Non-Mastery) หรือถ้าต้องการจำแนกเป็น 3 ประเภท ก็จะแบ่งได้เป็น

1. ผู้ที่รอบรู้ (Mastery) คือเรียนรู้ได้ครบแล้ว ควรเรียนบทเรียนที่สูงขึ้นต่อไปได้

2. ผู้รอบรู้เป็นบางส่วน (Partial-Mastery) คือเรียนรู้ได้บางส่วน ควรให้ทบทวนเพิ่มเติม

3. ผู้ที่ไม่รอบรู้ (Non-Mastery) คือเรียนรู้ได้ต่ำกว่าเกณฑ์ ควรซ่อมเสริม[1]

ในสถานการณ์เช่นนี้ชี้ให้เห็นว่า เราจำเป็นที่จะต้องมีคะแนนจุดตัดเพื่อจำแนกประเภทของผู้เรียนได้อย่างถูกต้อง และมีความคลาดเคลื่อนน้อยที่สุด ซึ่งในต่างประเทศได้มีผู้ศึกษาวิธีหาจุดตัดที่พอดีไว้หลายท่าน และหลายวิธีด้วยกัน ดังจะได้เสนอดังต่อไปนี้

<u>วิธีกำหนดคะแนนจุดตัด</u>

กลาส (Glass)[2] ได้รวบรวมวิธีการกำหนดคะแนนจุดตัด ซึ่งเป็นวิธีการหลัก ๆ ไว้ดังนี้ คือ

---

[1]สงบ ลักษณะ, "ทฤษฎีการสร้างแบบทดสอบอิงเกณฑ์" กองวิจัยการศึกษา กรมวิชาการ กระทรวงศึกษาธิการ (ถ่ายสำเนา)

[2]Glass "Standard and...," pp. 243 - 257.

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

ระดับเกณฑ์แบบนี้สร้างขึ้นโดยอ้างอิงค่าพารามิเตอร์ของประชากรผู้เข้าสอบ ดังนั้นเกณฑ์จึงอาจกำหนดด้วยคะแนนมัธยฐาน (median) ของคะแนนจากแบบสอบ หรืออาจจะกำหนดเป็นตำแหน่งเปอร์เซนไตล์ เป็นต้น แต่วิธีนี้นักทฤษฎีแบบสอบอิงเกณฑ์หลายคนแย้งว่าไม่เหมาะสม เพราะยังคงเป็นการอิงกลุ่มอยู่นั่นเอง

2. ใช้วิธีนับถอยหลังจาก 100 % (Counting backward from 100%)

การสร้างแบบสอบ เราจะต้องสร้างให้สอดคล้องกับวัตถุประสงค์ ดังนั้นข้อความที่ทุกคนสามารถตอบคำถามในแบบสอบได้ ก็ควรเป็นไปตามวัตถุประสงค์ด้วย ดังนั้นระดับคะแนนที่เราปรารถนาก็ควรจะเป็น 100 % แต่การจะให้สมบูรณ์ถึง 100 % ย่อมเป็นไปได้ยาก เพราะย่อมมีข้อบกพร่องและปัญหาอื่น ๆ อีก เกณฑ์จึงยอมให้ลดลงมาจาก 100 % ได้บ้าง ซึ่งแล้วแต่ว่าผู้ตั้งเกณฑ์จะลดลงมากี่เปอร์เซนต์ การยินยอมเช่นนี้มีข้อเสียอยู่ตรงที่น่าจะเป็นการกำหนดกันตามอำเภอใจเสียมากกว่า

3. การเพิ่มคะแนนเกณฑ์อื่น ๆ (Bootstrapping on Other Criterion Scores)

เป็นการกำหนดโดยอาศัยเกณฑ์จากภายนอกมาระบุเพิ่มขึ้นอีกว่าสำเร็จหรือรอบรู้ เช่น อาจจะระบุว่าต้องมีใบขับขี่ หรือต้องมีใบประกอบโรคศิลป์เป็นต้น แต่การกำหนดจุดตัดวิธีนี้ อาจมีปัญหา 2 ประการ คือ

3.1 ถ้าแบบสอบที่เราต้องการกำหนดจุดตัดมีค่าสหสัมพันธ์ กับเกณฑ์ภายนอกต่ำ เราก็ไม่สามารถกำหนดจุดตัดให้สอดคล้องกับเกณฑ์ภายนอกได้อย่างสมบูรณ์ เช่น อาจจะมีผู้ที่ผ่านเกณฑ์ภายนอกบางคนได้คะแนนสูงกว่าจุดตัด แต่ก็อาจมีอีกหลายคนที่ได้คะแนนต่ำกว่าจุดตัด และเราก็ไม่สามารถจะกำหนดจุดตัดให้ผู้ที่สอบผ่านเกณฑ์ภายนอกทุกคน ให้สอบผ่านแบบสอบที่เราต้องการกำหนดจุดตัดด้วยได้

3.2 ถ้าจะกำหนดจุดตัดให้สอดคล้องกับเกณฑ์ภายนอก ก็จะมีปัญหาว่าเกณฑ์ภายนอกที่ระบุขึ้นมานั้น เลือกมาอย่างไร

เบิร์ค (Berk, 1976)[1] ได้นำวิธีการนี้มาประยุกต์ใช้โดยกำหนดเกณฑ์ภายนอกว่า ได้รับการสอน หรือไม่ได้รับการสอนตามวัตถุประสงค์ที่จะวัด แบ่งนักเรียนออกเป็น 2 กลุ่ม แล้วคำนวณหาคะแนนจุดตัดที่เหมาะสมที่สุด (Optimal Cutting Score) จากข้อมูลที่ได้จากการทำแบบสอบอิงเกณฑ์ฉบับเดียวกัน

องค์ประกอบที่เบิร์คนำมาเป็นหลักในการพิจารณากำหนดคะแนนจุดตัดคือ การจัดประเภทความน่าจะเป็นที่เกิดขึ้น และสัมประสิทธิ์ของความตรง นอกจากนี้ยังพิจารณาผลได้ผลเสียในการตัดสินจากการวิเคราะห์อรรถประโยชน์และความตรงที่เพิ่มขึ้นด้วย

สำหรับการแบ่งนักเรียนออกเป็นกลุ่มได้รับการสอน และไม่ได้รับการสอนนี้ถ้าหากไม่สามารถจะแบ่งเป็น 2 กลุ่มที่เป็นอิสระต่อกันได้ ก็ควรจะใช้กลุ่มเดียวกันแต่ใช้คะแนนก่อนและหลังการสอนจะดีกว่า จากนั้นได้ทดลองกำหนดคะแนนขึ้นมาเป็นจุดตัด โดยนักเรียนที่ได้คะแนนมากกว่าหรือเท่ากับจุดตัดจะถือเป็นพวกรอบรู้โดยการทำนาย (Predicted Master: PM) ที่ได้คะแนนต่ำกว่าจุดตัดเป็นพวกไม่รอบรู้โดยการทำนาย (Predicted Non-master: PN) เมื่อพิจารณาแบ่งประเภทนักเรียนโดยใช้เกณฑ์ภายนอกและตัวทำนายตามคะแนนที่ได้จากการทำแบบสอบแล้ว จะแบ่งนักเรียนได้เป็น 4 พวก คือ

1. พวกรอบรู้จริง (True Master : TM) คือนักเรียนที่ได้รับการสอนและได้คะแนนมากกว่าหรือเท่ากับคะแนนจุดตัด

2. พวกรอบรู้ไม่จริง (False Master : FM) คือนักเรียนที่ไม่ได้รับการสอน แต่ได้คะแนนมากกว่าหรือเท่ากับคะแนนจุดตัด จะเป็นความคลาดเคลื่อนเนื่องจากการจัดประเภทผิดแบบที่ 2 (Type II misclassification error)

---

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

3. พวกไม่รอบรู้จริง (True Non-master : TN) คือนักเรียนที่ไม่ได้รับการสอนซึ่งได้คะแนนต่ำกว่าคะแนนจุดตัด

4. พวกไม่รอบรู้ไม่จริง (False Non-master : FN) คือนักเรียนที่ได้รับการสอน แล้วได้คะแนนต่ำกว่าคะแนนจุดตัด จะเป็นความคลาดเคลื่อนเนื่องจากการจัดประเภทผิด แบบที่ 1 (Type I misclassification error)

ผลที่เกิดขึ้นนี้สามารถแสดงในรูปตาราง 2 x 2 ได้คือ

เกณฑ์ภายนอก

	ได้รับการสอน	ไม่ได้รับการสอน
รอบรู้โดยการทำนาย (PM = TM + FM)	พวกรอบรู้จริง (TM)	พวกรอบรู้ไม่จริง (FM)
ไม่รอบรู้โดยการทำนาย (PN = FN + TN)	พวกไม่รอบรู้ไม่จริง (FN)	พวกไม่รอบรู้จริง (TN)
	พวกรอบรู้ (M = TM+FN)	พวกไม่รอบรู้ (N = FM+TN)

เมื่อทดลองกำหนดคะแนนจุดตัดหลาย ๆ จุด จะได้จำนวนนักเรียนในแต่ละประเภทแตกต่างกันตามค่าของคะแนนจุดตัดแต่ละจุด นำผลที่ได้จากการกำหนดคะแนนจุดตัดต่าง ๆ กันมาหาค่าความน่าจะเป็น (Probability) ของแต่ละประเภทโดยใช้สูตรดังนี้

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

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

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

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

คะแนนจุดตัดที่ให้ค่าความน่าจะเป็นในการตัดสินอย่างถูกต้อง $[P(TM) + P(TN)]$ มีค่าสูงสุด หรือทำให้ค่าความน่าจะเป็นในการตัดสินผิด $[P(FM) + P(FN)]$ มีค่าต่ำสุด จะเป็นคะแนนจุดตัดที่เหมาะสมสำหรับแบบสอบฉบับนี้

สมหวัง พิธิยานุวัฒน์[1] ได้นำวิธีการของเบิร์ค (Berk) มาประยุกต์ใช้กับนิสิตปริญญาโทแผนกจิตวิทยา ที่เรียนวิชาสถิติประยุกต์ จำนวน 18 คน เพื่อหาจุดแบ่งที่พอดีของแบบสอบอิงเกณฑ์วิชาสถิติ เรื่อง สหสัมพันธ์ ซึ่งมีความยาว 14 ข้อ โดยให้ทดสอบก่อนและหลังการเรียน แล้วนำคะแนนที่ได้มาเขียนกราฟของการกระจายของคะแนน แล้วพิจารณากำหนดคะแนนจุดตัดจากจุดที่เส้นกราฟทั้งสองตัดกันให้เป็นจุดตัดขั้นต่ำอย่างสูง และประมวลความคิดของบลูม บลอค และเกลเซอร์ เกี่ยวกับการกำหนดระดับความรอบรู้ของผู้เรียนมากำหนดเป็นคะแนนจุดตัดขั้นสูงอย่างต่ำจากคะแนนจุดตัดที่กำหนดขึ้นทั้ง 2 นี้ ได้นำมาหาคะแนนจุดตัดที่พอดี ปรากฏผลว่าได้คะแนน 7.5 เป็นจุดตัดขั้นต่ำอย่างสูง และ 9.8 เป็นจุดตัดขั้นสูงอย่างต่ำ

4. <u>พิจารณาตัดสินจากความสามารถต่ำสุด</u> (Judging Minimal Competence)

การกำหนดจุดตัดด้วยวิธีนี้ อาศัยแนวคิดของเกลเซอร์ที่เสนอไว้ว่าความสามารถจะต่อเนื่องกันจากไม่มีเลยจนถึงปฏิบัติได้โดยสมบูรณ์ทั้งหมด และจุดตัดจะแบ่งผู้ที่มีความสามารถและไม่มีความสามารถออกจากกัน ดังนั้นหาผู้ที่มีความสามารถต่ำสุดที่จะยอมรับว่าผ่านได้ ก็จะทำให้ได้จุดตัดที่เหมาะสม สำหรับวิธีการนี้มีผู้ศึกษาและได้เสนอเทคนิคในการหาคะแนนความสามารถต่ำสุดไว้หลายคนด้วยกัน เช่น

---

[1]สมหวัง พิธิยานุวัฒน์. "จุดแบ่งที่พอดีของแบบสอบอิงเกณฑ์ทางสถิติศาสตร์, การศึกษาเบื้องต้น" ใน<u>เอกสารประกอบการประชุมชิมโปเซียม</u>, (กรุงเทพมหานคร สำนักพิมพ์จุฬาลงกรณ์มหาวิทยาลัย, 2520) หน้า 13.

อีเบล (Ebel, 1972) มีความเห็นว่าคะแนนผ่าน (passing score) ที่กำหนดเป็นเปอร์เซนต์ของคะแนนสอบทั้งหมดอาจมีข้อบกพร่องเนื่องจากข้อกระทงนั้นยาก หรือง่ายเกินไป หรืออำนาจจำแนกน้อยกว่าที่ผู้สร้างตั้งใจไว้ การที่จะสอบผ่านหรือไม่จึงเป็นการตัดสินโดยคำถามในแบบสอบมากกว่าระดับความสามารถของผู้สอบก็ได้ เขาจึงเสนอวิธีแก้ไขโดยการให้วิเคราะห์เนื้อหาของแบบสอบในแง่ของความเกี่ยวข้อง (relevance) และความยาก (Difficulty)ของข้อกระทงแต่ละข้อ แล้วกำหนดเปอร์เซนต์ความคาดหวังซึ่งเป็นปริมาณที่คาดว่าผู้ที่มีความสามารถต่ำสุดที่จะสอบผ่านควรทำได้ดังตารางต่อไปนี้

ตารางที่ 1 แสดงค่าเปอร์เซนต์ความคาดหวังในแต่ละระดับความยากของแต่ละประเภทของความเกี่ยวข้อง[1]

ประเภทของความเกี่ยวข้อง	ระดับความยาก		
	ง่าย	ปานกลาง	ยาก
จำเป็น (essential)	100 %	-	-
สำคัญ (important)	90	70 %	-
ยอมรับได้ (acceptable)	80	60	40 %
ไม่แน่ใจ (questionable)	70	50	30

อีเบลได้นำแบบสอบที่มีจำนวน 100 ข้อ ไปให้ผู้เชี่ยวชาญ 5 ท่าน พิจารณาว่าแต่ละข้อกระทงควรจะอยู่ในประเภทใด และมีความยากระดับใด แล้วนำผลรวมมาเฉลี่ย โดยเอาจำนวนข้อกระทงแต่ละประเภทคูณด้วยเปอร์เซนต์ที่คาดหวัง นำมาเฉลี่ยเพื่อคาดคะเนคะแนนผ่านที่เหมาะสม

[1]Glass, "Standard and...," p. 248.

ตารางที่ 2 แสดงจำนวนข้อกระทงในแต่ละประเภทของความเกี่ยวข้องตามระดับความยากของอีเบล[1]

ประเภทของข้อกระทง		จำนวนข้อกระทง	ความสำเร็จที่คาดหวัง	จำนวนข้อ x ความสำเร็จ
จำเป็น		94	100	9400
สำคัญ	ง่าย	106	90	9540
	ปานกลาง	153	70	10710
ยอมรับได้	ง่าย	24	80	1920
	ปานกลาง	49	60	2940
	ยาก	52	40	2080
ไม่แน่ใจ	ง่าย	4	70	280
	ปานกลาง	11	50	50
	ยาก	7	30	210
		500		37130

คะแนนผ่าน = 37130 /500 = 74.26 % ≈ 74 %

แอนกอฟ (Angoff, 1971)[2] ศึกษาวิธีหาคะแนนผ่าน (Passing score) ของแบบสอบเลือกตอบโดยใช้ค่าความน่าจะเป็นที่นักเรียนที่สอบผ่านและอยู่ในระดับต่ำสุดควรตอบได้อย่างถูกต้อง ด้วยการให้ผู้ตัดสินแต่ละคนเสนอค่าความน่าจะเป็นที่ผู้ซึ่งอยู่ระดับต่ำสุดที่

[1] Ibid., p. 248.

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

เราสามารถยอมรับได้ ควรจะตอบแต่ละข้อกระทงนั้นได้ถูกต้อง ผู้ตัดสินจะคำนึงถึงจำนวนคนที่อยู่ระดับต่ำสุดที่เราจะยอมรับได้ (แทนที่จะคิดเฉพาะเพียงคนเดียว) และประมาณค่าสัดส่วนออกมา ผลรวมของค่าความน่าจะเป็นหรือสัดส่วนเหล่านี้จะใช้เป็นคะแนนต่ำสุดที่ยอมรับได้

5. <u>การใช้ทฤษฎีการตัดสินใจ</u> (Decision Theoretic Approach)[1]

วิธีนี้จะแบ่งคนออกเป็น 2 พวก โดยอาศัยเกณฑ์ภายนอกบางอย่างที่เกี่ยวข้องกับสิ่งที่เราสนใจจะศึกษา เช่น อาจแบ่งเป็นผู้ที่จบการศึกษาจากวิทยาลัย และผู้ที่ไม่จบจากวิทยาลัย ผู้ที่ได้รับการจ้างและไม่ได้รับการจ้างเป็นต้น สัดส่วนของคน 2 กลุ่มนี้ แทนด้วย **PE** และ **1-PE** ตามลำดับ แล้วนำเอาแบบสอบอิงเกณฑ์มาทดสอบกับทั้ง 2 กลุ่ม และกำหนดคะแนนจุดตัด (**Cx**) ขึ้นมาเพื่อแบ่งคนในแต่ละกลุ่มออกเป็นผู้ที่สอบผ่าน และสอบไม่ผ่าน จะได้สัดส่วนของการตัดสินดังตารางต่อไปนี้

แบบสอบอิงเกณฑ์ \ เกณฑ์ภายนอก	ผ่าน	ไม่ผ่าน	
ไม่ผ่าน	P_A	P_B	$1-P_C$
ผ่าน	P_C	P_D	P_C
	P_E	$1-P_E$	

P_A คือ สัดส่วนของการปฏิเสธที่ผิด คือคนที่ไม่ผ่านแบบสอบอิงเกณฑ์ แต่ผ่านเกณฑ์ภายนอก

P_D คือ สัดส่วนของการยอมรับที่ผิด คือ คนที่ผ่านแบบสอบอิงเกณฑ์ แต่ไม่ผ่านเกณฑ์ภายนอก

---

[1]Ronald K. Hambleton. "Testing and Decision-Making Procedures for Selected Individualized Instructional Programs." <u>Review of Educational Research</u> 44(1974) : 371-400.

P_B คือ สัดส่วนของคนที่ไม่ผ่านทั้งแบบสอบอิงเกณฑ์และเกณฑ์ภายนอก

P_C คือ สัดส่วนของคนที่ผ่านทั้งแบบสอบอิงเกณฑ์และเกณฑ์ภายนอก

การกำหนดเกณฑ์ภายนอกจะไม่ทำให้เปลี่ยนแปลง แต่คะแนนจุดตัดในแบบสอบอิงเกณฑ์จะแปรผันไปได้หลายค่าแล้วแต่การกำหนดซึ่งจะทำให้สัดส่วนของ P_A, P_B, P_C และ P_D แปรผันตามไปด้วย จากนั้นแทนค่าในสมการ

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

ค่าของ Cx ที่ทำให้ f(Cx) มีค่าน้อยที่สุดจะถูกเลือกเป็นคะแนนจุดตัดของแบบสอบอิงเกณฑ์ แต่จะมีปัญหาอยู่ที่เราใช้ค่าของการยอมรับที่ผิด (ความคลาดเคลื่อนประเภทที่ 2) และการปฏิเสธที่ผิด (ความคลาดเคลื่อนประเภทที่ 1) โดยกำหนดให้มีความสำคัญเท่ากัน ซึ่งเป็นการกำหนดขึ้นตามความพอใจเป็นอย่างมาก (highly arbitary) ในทางตรงกันข้ามถ้าเราให้ค่าความสำคัญของ P_A เป็น α และค่าความสำคัญของ P_D เป็น β ก็จะได้ฟังชั่นที่เหมาะสมที่จะหาค่า Cx ดังนี้

$$f(Cx) = (\alpha P_A + \beta P_D) / (P_B + P_C)$$

แต่ค่าของฟังชั่นนี้มีความไวต่อค่า α และ β ซึ่งได้จากการพิจารณาของผู้ตัดสินมาก ค่านี้จึงแปรผันไปตามความเห็นของผู้ตัดสินซึ่งก็ยังเป็นการกำหนดตามความพอใจเช่นกัน ดังนั้นจึงควรพิจารณาองค์ประกอบอื่น ๆ ที่เกี่ยวข้องกับการสอบผ่านหรือไม่ผ่านด้วย ดังที่เอมริค (Emrick, 1971) ได้เสนอองค์ประกอบที่เกี่ยวข้องกับความคลาดเคลื่อนในการตัดสินไว้ 3 ประการ คือ[1]

1) ด้านค่าสถิติ เช่น ความเที่ยงของข้อกระทง ความยาวของแบบสอบ และสิ่งอื่น ๆ ที่เกี่ยวกับด้านนี้ ถ้าการสอบนี้เป็นการตัดสินได้-ตก ด้านสถิตินี้จะมีความสำคัญมาก แต่ถ้าการสอบใช้เพื่อวินิจฉัย การพิจารณาด้านสถิติก็จะลดความสำคัญลง

---

[1] John A. Emrick. "An Evaluation Model for Mastery Testing." Journal of Educational Measurement 8(1971):321-326.

2) ความสำคัญของเนื้อหา การทดสอบเพื่อการรอบรู้จะเน้นวัตถุประสงค์ที่มีความสำคัญต่อกระบวนการศึกษา ดังนั้นวัตถุประสงค์ของการเรียนบางข้อจะมีความสำคัญต่อเนื้อหามาก ถ้าวัตถุประสงค์นี้ไม่ได้วัดอย่างเพียงพอและเที่ยงตรงแล้ว ก็ย่อมจะเกิดความคลาดเคลื่อนขึ้นได้

3) คุณค่าทางด้านจิตวิทยาซึ่งเป็นผลจากความคลาดเคลื่อนในการตัดสินไม่ว่าจะเป็นกระบวนการประเมินผลใด ๆ คุณค่าทางจิตวิทยาจะต้องนำมาพิจารณาด้วย โดยเฉพาะในการประเมินผลเกี่ยวกับเด็ก คุณค่าด้านนี้จะมีความสำคัญเป็นอย่างมาก ส่วนผู้ใหญ่ความสำคัญด้านนี้จะลดลง

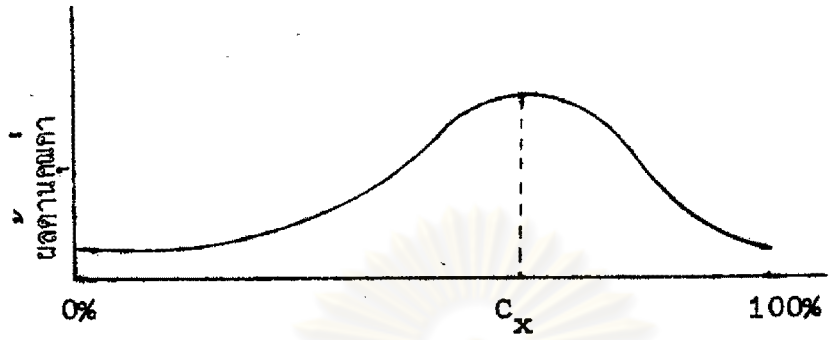
6. <u>การใช้วิธีวิจัยเชิงปฏิบัติ</u> (Operations Research Method)

วิธีนี้จะต้องอาศัยผลด้านคุณค่า (Valued Outcome) อย่างใดอย่างหนึ่งมาช่วยในการพิจารณาเกณฑ์ โดยวัดผลด้านคุณค่าของผู้ที่ได้คะแนนต่าง ๆ กันในการทดสอบด้วยแบบสอบอิงเกณฑ์ แล้วใช้คะแนนของผู้ที่มีผลด้านคุณค่านั้น สูงสุดมาเป็นคะแนนเกณฑ์ โดยพิจารณาจากลักษณะกราฟแสดงความสัมพันธ์ระหว่างผลที่ได้จากการสอบด้วยแบบสอบอิงเกณฑ์และจากการวัดผลด้านคุณค่านั้น คะแนนตรงจุดสูงสุดของกราฟจะถือเป็นคะแนนเกณฑ์ ซึ่งบลอค (Block, 1972)[1] ได้ประยุกต์วิธีนี้มาใช้ดังนี้

บลอคได้ทำการสอนกลุ่มต่าง ๆ ที่มีความเท่าเทียมกันจนมีผลสัมฤทธิ์จากการทำแบบสอบอิงเกณฑ์ระดับต่าง ๆ กัน เช่น 10%, 15%, 20%,...,95% 100% แล้ววัดผลด้านคุณค่าอย่างใดอย่างหนึ่งที่สัมพันธ์กับคะแนนจากแบบสอบอิงเกณฑ์ของแต่ละกลุ่มที่มีผลสัมฤทธิ์ระดับต่าง ๆ กันนั้น นำผลที่ได้มาเขียนกราฟแสดงความสัมพันธ์เพื่อพิจารณาค่าของคะแนนเกณฑ์ (Cx)

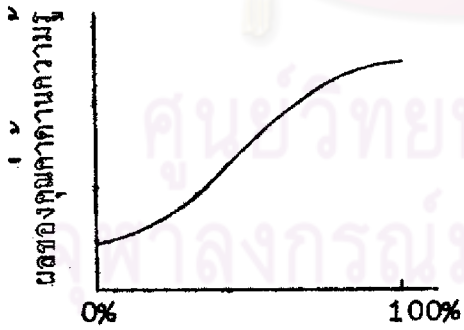
---

[1] G.V. Glass. "Standard and Criteria." <u>Journal of Educational Measurement</u>. 15(1978) : 32-36. Citing J.H. Block. "Student Learning and Setting of Mastery Performance Standard." <u>Educational Horizons</u>. 50(1972) : 183 - 190.

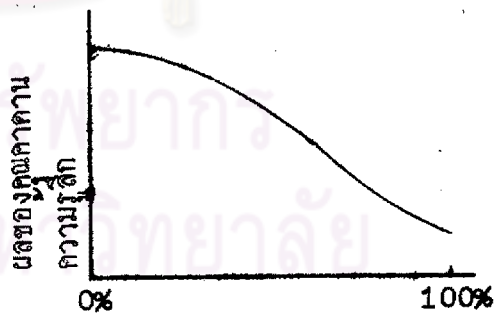


ระดับผลสัมฤทธิ์จากแบบสอบอิงเกณฑ์

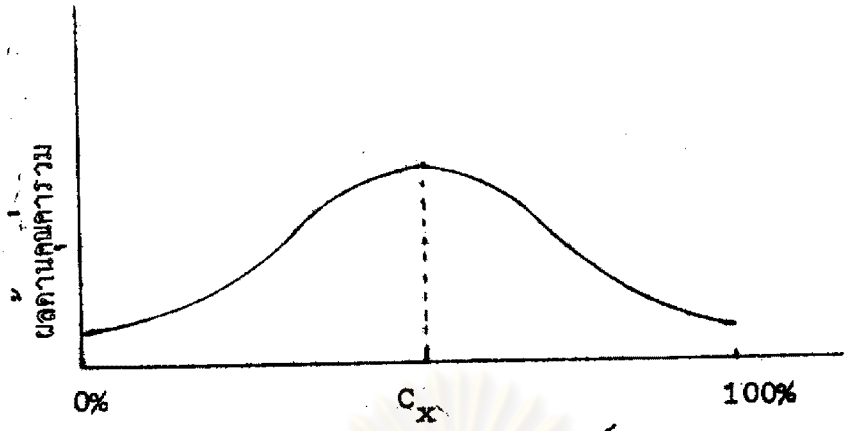
วิธีนี้ถ้ากราฟที่ได้มีลักษณะเป็นโมโนโทนิค (monotonic graph) คือ ไม่มีจุดที่เส้นกราฟจะโค้งวกกลับมาที่เส้นฐาน (base line) คะแนนเกณฑ์ก็จะเป็น 100 % ซึ่งเป็นไปไม่ได้ ถ้าทั้งแบบสอบอิงเกณฑ์และการวัดผลด้านคุณค่าเป็นการวัดด้านความรู้ (Cognitive) เส้นกราฟมักจะไม่เป็นแบบนอนโมโนโทนิค (non-monotonic) ปัญหาเช่นนี้บลอคเสนอให้ใช้ผลด้านคุณค่าด้านอื่นที่มีความสัมพันธ์กลับกันกับระดับคะแนนจากแบบสอบอิงเกณฑ์มาประกอบการพิจารณา ซึ่งมักจะเป็นการวัดด้านความรู้สึก (affective) เช่นความสนใจหรือทัศนคติต่อวิชานั้น ๆ เมื่อรวมผลด้านคุณค่าทั้งการวัดด้านความรู้และความรู้สึกเข้าด้วยกันแล้วก็จะได้ผลสัมฤทธิ์รวม คะแนนที่ให้ค่าผลด้านคุณค่ารวมสูงสุดจะถือเป็นคะแนนเกณฑ์



คะแนนจากแบบสอบอิงเกณฑ์



คะแนนจากแบบสอบอิงเกณฑ์



คะแนนจากแบบสอบอิงเกณฑ์

การพิจารณาคุณค่ารวมควรจะต้องมีการกำหนดน้ำหนักให้กับผลของคุณค่า ทั้งในด้านความรู้และความรู้สึกด้วย แล้วกำหนดค่า α และ β ในสมการ

ผลลัพธ์ด้านคุณค่ารวม = α (ผลด้านความรู้) + β (ผลด้านความรู้สึก)

นอกจากวิธีการที่กลาสได้เสนอไว้แล้ว ยังมีวิธีที่เป็นที่น่าเชื่อถืออีกวิธีหนึ่งคือ <u>การใช้ทฤษฎีการตัดสินใจของเบส์</u> (Bayesian Decision-Theoretic Procedure) ซึ่งพิจารณาจัดประเภทผู้สอบโดยใช้คะแนนโดเมน (π) ของผู้สอบมาเปรียบเทียบกับคะแนนจุดตัด (π_0) ที่กำหนดขึ้นมาใช้ โดยที่ผู้สอบที่มีคะแนน โดเมนมากกว่าหรือเท่ากับคะแนนจุดตัด จัดเป็นผู้รอบรู้ (master)และผู้สอบที่มีคะแนนโดเมนน้อยกว่าคะแนนจุดตัด จัดเป็นผู้ไม่รอบรู้ (non-master)

ในทางปฏิบัติเราไม่สามารถหาค่า π ที่แท้จริงได้ เนื่องจากไม่สามารถนำข้อกระทงในโดเมนนั้น ๆ มาใช้ได้ทั้งหมด จึงหาค่า π ได้ด้วยการประมาณเท่านั้น

แฮมเบิลตัน โนวิค และสแวมมินาธาน[1] เสนอวิธีประมาณค่าคะแนนโดเมนโดยมี

---

[1] H. Swaminathan, R.K. Hambleton, and J. Algina. "A Bayesian Decision-Theoretic Procedure for Use with Criterion-Referenced Tests." <u>Journal of Educational Measurement</u>. 12(1975) : 88-93.

ขั้นตอนดังนี้

แปลงคะแนนดิบของผู้สอบ (Xi) ให้อยู่ในรูป g_i โดยการแปลงแบบอาร์คไซน์ (arc sine) จากสูตร

$$g_i = \sin^{-1}\{(X_i + 3/8)/(n + 3/4)\}^{1/2}$$

การแจกแจงนี้มีค่าเฉลี่ยเป็น ν_i และความแปรปรวน V ซึ่ง

$$\nu_i = \sin^{-1}\sqrt{\pi_i}$$

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

การประมาณนี้จะใช้ได้ดีเมื่อ π มีค่าระหว่าง .15 ถึง .85 และจำนวนข้อสอบ (n) ไม่ต่ำกว่า 8 ข้อ แต่เนื่องจากเราไม่ทราบค่า π_i จึงไม่ทราบค่า ν_i การแจกแจงของ g_i จึงเป็นการแจกแจงที่มีเงื่อนไข คือ

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

ฉะนั้นการแจกแจงของ g_i จึงขึ้นอยู่กับ ν_i เพื่อให้ได้การแจกแจงภายหลังของ ν_i ต้องมีข้อตกลงว่า ν_i เป็นตัวอย่างที่ได้จากการสุ่มการแจกแจงอันใดอันหนึ่ง โดยสมมติว่าการแจกแจงเดิมของ ν_i เป็นปกติด้วยค่าเฉลี่ย α และความแปรปรวน ϕ ซึ่งเราไม่ทราบค่าทั้ง α และ ϕ ดังนั้นการแจกแจงเดิมของ ν_i จึงขึ้นอยู่กับค่า α และ ϕ

โนวิค (Novick,1973) แนะนำว่า เราสามารถหาค่า ϕ ได้จากการแจกแจงแบบอินเวอร์ส ไค-สแคว์ (inverse Chi-Square) ที่มีชั้นของความอิสระ ν และสเกลพารามิเตอร์ λ โดย

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

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

และยังพบว่า ค่า ν ที่เหมาะสมคือ 8 ส่วนการประมาณค่า $\bar{\phi}$ จำเป็นต้องมีข้อมูลเกี่ยวกับ π ดังนั้นจึงต้องมีการพิจารณากำหนดค่า t ซึ่งเป็นจำนวนข้อกระทงซึ่งคิดว่าจะช่วยให้ทราบข้อมูลเกี่ยวกับค่า π ได้มากเท่าที่เราต้องการ ค่าประมาณที่ได้จากการแปลงค่า π แบบอาร์คไซน์ ในการทำแบบสอบที่มี t ข้อ จะเป็นปกติด้วยความแปรปรวน $(4t+2)^{-1}$ ดังนั้นเราจึงใช้ค่า $(4t+2)^{-1}$ เป็นค่าประมาณของ ϕ และหาค่า λ ได้จากสูตรข้างต้น

เลวิสและคณะ (1973) ศึกษาเรื่องนี้พบว่า การแจกแจงภายหลังของ ν_i แต่ละค่าเป็นปกติโดยประมาณด้วยค่าเฉลี่ย μ_i และความแปรปรวน σ_i^2

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

เมื่อ

$$\mu_i = \bar{g} + \rho^* (g_i - \bar{g})$$

และ

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

โดยที่

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

เมื่อ m เป็นจำนวนผู้เข้าสอบ

n เป็นจำนวนข้อกระทงในแบบสอบ

ρ^* และ σ^{*2} ได้จากตารางแวง (Wang) เมื่อทราบค่า ν และ λ ค่าประมาณของ π หาได้จากสูตร

$$\hat{\pi}_i = (1 + 0.5/n) \sin^2 \nu_i - 0.25/n$$

เนื่องจาก π ที่นำมาใช้เป็นค่าที่ได้จากการประมาณ ดังนั้นการพิจารณาจัดประเภทผู้สอบออกเป็น 2 พวก จึงอาจมีความคลาดเคลื่อนได้ใน 2 ลักษณะ คือ

1. จัดผู้ที่ควรรอบรู้ให้เป็นผู้ไม่รอบรู้
2. จัดผู้ที่ไม่ควรรอบรู้ให้เป็นผู้รอบรู้

ถ้าให้

w_1 แทนระดับการรอบรู้ของผู้เข้าสอบที่เป็นผู้ไม่รอบรู้ คือมี $\nu < \nu_0$

w_2 แทนระดับการรอบรู้ของผู้เข้าสอบที่เป็นผู้รอบรู้ คือมี $\nu > \nu_0$

a_1 เป็นการตัดสินให้ผู้สอบเรียนซ้ำ

a_2 เป็นการตัดสินให้สอบผ่าน

ความสูญเสียเนื่องจากการจัดประเภทผิด จะมี 2 ประเภท คือ

$L(w_1, a_2) = \ell_{12}$ คือความสูญเสียเนื่องจากความคลาดเคลื่อนในการยอมรับที่ผิด (false positive)

$L(w_2, a_1) = \ell_{21}$ คือความสูญเสียเนื่องจากความคลาดเคลื่อนในการปฏิเสธที่ผิด (false negative)

ถ้าการตัดสินถูกต้อง ความสูญเสียจะไม่มีคือ

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

ในการตัดสินเราจะเลือกการตัดสินที่ให้ค่าความสูญเสียน้อยที่สุด ถ้าในการตัดสินให้สอบตก ความสูญเสียที่คาดหวัง คือ

$$\begin{aligned} Ew\ L(w, a_1) &= 0 \cdot \text{Prob}[w = w_1] + \ell_{21}\ \text{Prob}[w = w_2] \\ &= \ell_{21}\ \text{Prob}[w = w_2] \\ &= \ell_{21}\ \text{Prob}[\nu \geq \nu_0] \end{aligned}$$

ในการตัดสินให้สอบผ่าน ความสูญเสียที่คาดหวัง คือ

$$\begin{aligned} Ew\ L(w, a_2) &= \ell_{12}\ \text{Prob}[w = w_1] + 0 \cdot \text{Prob}[w = w_2] \\ &= \ell_{12}\ \text{Prob}[w = w_1] \\ &= \ell_{12}\ \text{Prob}[\nu < \nu_0] \end{aligned}$$

I16342082

ดังนั้นเราจะตัดสินให้ผู้สอบเรียนซ้ำถ้า

$$\ell_{21} \text{ Prob}[\nu \geq \nu_0] < \ell_{12} \text{ Prob}[\nu < \nu_0]$$

และจะตัดสินให้สอบผ่าน ถ้า

$$\ell_{12} \text{ Prob}[\nu < \nu_0] < \ell_{21} \text{ Prob}[\nu \geq \nu_0]$$

ถ้าการสูญเสียที่คาดหวังทั้ง 2 ประเภทมีค่าเท่ากัน การตัดสินให้ผู้สอบเรียนซ้ำหรือสอบผ่าน จะให้ผลไม่แตกต่างกัน

การหาค่า Prob $[\nu \geq \nu_0]$ และ Prob $[\nu < \nu_0]$ จะหาได้จาก

$$\text{Prob}[\nu_i \geq \nu_0 / \text{Data}] \simeq \text{Prob}[Z \geq Z_{oi} / \text{Data}]$$

$$\text{Prob}[\nu_i < \nu_0 / \text{Data}] \simeq \text{Prob}[Z < Z_{oi} / \text{Data}]$$

เนื่องจากการแจกแจงภายหลังของ ν_i เป็นปกติโดยประมาณด้วยค่าเฉลี่ย μ_i และความแปรปรวน σ_i โดยมี Zoi เป็นค่าส่วนเบี่ยงเบนมาตรฐานปกติ ที่หาได้จากสูตร

$$Z_{oi} = (\nu_0 - \mu_i)/\sigma_i$$

ดังนั้นเราจึงสามารถเปรียบเทียบในรูปของ Z ได้ คือเปรียบเทียบระหว่าง

$$\ell_{12} \text{ Prob}[Z < Z_{oi}/\text{Data}] \text{ กับ } \ell_{21} \text{ Prob}[Z \geq Z_{oi}/\text{Data}]$$

แล้วเลือกการกระทำที่เหมาะสมกับนักเรียนแต่ละคน และคะแนนจุดตัดที่เหมาะสมคือคะแนนที่ผู้สอบผ่านระดับต่ำสุดทำได้

สำหรับการพิจารณาจัดประเภทผู้สอบมากกว่า 2 ประเภท เช่น ถ้าจัดเป็น k ประเภท ระดับการรอบรู้ก็จะมี k ระดับ คือ $w_1, w_2, \ldots, w_k$ และคะแนนจุดตัดก็จะมี k -1 จุด คือ $\pi_{o1}, \pi_{o2}, \ldots, \pi_{ok-1}$ ผู้สอบจะถูกจัดให้อยู่ใน w_1 เมื่อคะแนนโดเมน (π) น้อยกว่าคะแนนจุดตัดที่ 1 (π_{o1}) และอยู่ใน w_2 เมื่อ $\pi_{o1} \leq \pi < \pi_{o2}$

หรืออาจเขียนในรูปทั่ว ๆ ไปได้ว่า ผู้สอบจะถูกจัดให้อยู่ใน w_i เมื่อ $\pi_{oi-1} \leq \pi < \pi_{oi}$ ค่าความสูญเสียในการจัดประเภทผิดคือ $L(w_i, aj) = \ell_{ij}$ และเราควรเลือกความสูญเสียที่คาดหวังสำหรับการกระทำ aj คือ

$$E_w L(w_i, aj) = \sum_{p=1}^{k} \ell_{pj} \text{ Prob}\left[\nu_{op-1} \leq \nu < \nu_{op} \ / \text{Data}\right]$$

การประทำ aj จะถูกเลือกเมื่อ

$$\sum_{p=1}^{k} \ell_{pj} \text{ Prob}\left[\nu_{op-1} \leq \nu < \nu_{op} \ / \text{Data}\right] < \sum_{m=1}^{k} \ell_{pm} \text{Prob}\left[\nu_{op-1} \leq \nu < \nu_{op} / \text{Data}\right]$$

$m = 1, 2, \ldots, k$ และ $m \neq j$

เมื่อได้การแจกแจงภายหลังของ ν ค่าความน่าจะเป็นข้างบนก็คือพื้นที่ใต้โค้งความน่าจะเป็นระหว่าง ν_{op-1} และ ν_{op} เมื่อ $p = 1, 2, \ldots, k$ ค่า $\text{Prob}[\nu_{op-1} \leq \nu < \nu_{op} / \text{Data}]$ หาได้โดยใช้วิธีการเดียวกันกับที่ใช้ในการแบ่งเป็น 2 ประเภทดังกล่าวมาแล้ว

แฮมเบิลตัน แสวมมินาธาน และอัลกินา[1] (1975) ได้ใช้วิธีของเบส์เพื่อหาเกณฑ์การจัดประเภทผู้เข้าสอบโดยพิจารณาค่าความสูญเสียน้อยที่สุด (Threshold loss) จากการใช้แบบสอบที่มีความยาว 10 ข้อ กับนักเรียน 25 คน ทดลองกำหนดคะแนนจุดตัด $\pi_0 = .8$ ความสูญเสียเนื่องจากความคลาดเคลื่อนในการยอมรับที่ผิด (ℓ_{12}) เป็น 1 หน่วย ความสูญเสียเนื่องจากความคลาดเคลื่อนในการปฏิเสธที่ผิด (ℓ_{21}) เป็น 2 หน่วย พบว่าเมื่อแบ่งผู้สอบเป็น 2 ระดับการรอบรู้ ค่าความสูญเสียจะน้อยที่สุด ถ้าตัดสินให้ผู้ทำข้อสอบได้ 9 ข้อขึ้นไป เป็นผู้รอบรู้ และผู้ที่ได้ต่ำกว่า 9 ข้อ เป็นผู้ไม่รอบรู้ และเมื่อแบ่งผู้เข้าสอบเป็น 3 ระดับการรอบรู้ โดยกำหนดคะแนนจุดตัด π_0 เป็น .6 และ

---

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

.8 พบว่าค่าความสูญเสียจะน้อยที่สุด เมื่อตัดสินให้ผู้ที่ทำข้อสอบได้ 10 ข้อเป็นผู้รอบรู้ ผู้ที่ทำได้ 7-9 ข้อ เป็นผู้รอบรู้บางส่วน ซึ่งจะต้องทบทวนให้อีกเล็กน้อย และผู้ที่ทำได้ต่ำกว่า 7 ข้อ เป็นผู้ไม่รอบรู้ ต้องเรียนใหม่

ชมพู จันทรอมรพร[1] ได้ใช้วิธีการของเบส์ เพื่อกำหนดคะแนนจุดตัดของแบบสอบอิงเกณฑ์ วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 1 เรื่องสมการด้วยเช่นกัน และหาความเที่ยงและความตรงในการตัดสินด้วยสูตรของ ซับโคเวียคและของคาร์เวอร์ตามลำดับ ปรากฏผลว่า แบบสอบฉบับที่ 1 มีคะแนนจุดตัดเป็น 6 คะแนน ค่าความเที่ยงและความตรงในการตัดสินเป็น .76 และ .84 ฉบับที่ 2 มีคะแนนจุดตัดเป็น 5 ความเที่ยงและความตรงในการตัดสินเป็น .72 และ .80 ฉบับที่ 3 คะแนนจุดตัดเป็น 5 ความเที่ยงและความตรงในการตัดสินเป็น .65 และ .79 และฉบับที่ 4 คะแนนจุดตัดเป็น 5 ความเที่ยงและความตรงในการตัดสินเป็น .67 และ .78 ตามลำดับ

สำหรับแบบสอบอิงเกณฑ์วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 1 เรื่องสมการที่ชมพูใช้เป็นเครื่องมือในการศึกษานี้ สร้างขึ้นโดย กาญจนา วัชนสุนทร[2] ซึ่งแยกเป็น 4 ฉบับย่อย ฉบับละ 10 ข้อ แล้วิเคราะห์คุณภาพของแบบสอบทั้งค่าความยาก อำนาจจำแนกและหาเกณฑ์ตัดสินที่เหมาะสมโดยการทดลองกำหนดเป็น .6, .7 และ .8 แล้วหาค่าความเที่ยงและความตรงของการตัดสินจากสูตรของลิฟวิงสตัน และคาร์เวอร์ ตามลำดับ พบว่าจุดตัดที่จะทำให้ค่าความเที่ยงและความตรงในการตัดสินสูงสุดคือ .6

---

[1]ชมพู จันทรอมรพร. "การใช้กระบวนการเชิงทฤษฎีการตัดสินใจของเบส์ ในการกำหนดคะแนนจุดตัดของแบบสอบอิงเกณฑ์วิชาคณิตศาสตร์" (วิทยานิพนธ์ปริญญามหาบัณฑิต แผนกวิจัยการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, 2522)

[2]กาญจนา วัชนสุนทร. "การสร้างแบบสอบอิงเกณฑ์วิชาคณิตศาสตร์" (วิทยานิพนธ์ปริญญามหาบัณฑิต แผนกวิจัยการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, 2521)

แฮมเบิลตันและสแวมมินาธาน[1] (1976) ได้ใช้วิธีประเมินความรอบรู้ของนักเรียนหลายวิธีเปรียบเทียบกัน แล้วพิจารณาองค์ประกอบต่าง ๆ คือความเป็นเอกพันธ์ของกลุ่ม ข้อมูลเดิมเกี่ยวกับตัวผู้สอบ คะแนนจุดตัด ขนาดของกลุ่มตัวอย่างและความยาวของแบบสอบ วิธีที่เขาใช้ได้แก่

1. วิธีหาสัดส่วนความถูกต้อง (Proportion-Correct Score Estimate)

2. วิธีคลาสสิคแบบที่ 2 (Classical Model II Estimate)

3. การประมาณค่าโดยวิธีของเบส์แบบที่ 2 (Bayesian Model II Estimate)

4. การประมาณค่ามัชฌิมของแต่ละคนโดยวิธีของเบส์ (Bayesian Marginal Mean Estimate)

5. การประมาณค่ามัชฌิมของแต่ละคนด้วยวิธีใหม่ (Modified Marginal Mean Estimate)

จากการใช้คะแนนจุดตัดเป็น .8 ขนาดของกลุ่มตัวอย่างเป็น 15, 25 และ 50 คน ความยาวแบบสอบเป็น 8, 10 และ 20 ข้อ เปรียบเทียบคะแนนความรอบรู้ที่แท้จริงกับคะแนนความรอบรู้ที่ประมาณได้จากวิธีทั้ง 5 พบว่าขนาดของกลุ่มตัวอย่างมีผลน้อยมากไม่ว่าจะใช้วิธีใด วิธีที่ให้ผลดีที่สุดคือวิธีประมาณค่ามัชฌิมของแต่ละคนโดยวิธีใหม่ ซึ่งมีข้อจำกัดอยู่ที่ว่าจะต้องมีรายละเอียดเกี่ยวกับข้อมูลเดิมของระดับความรอบรู้ และการแจกแจงของคะแนนระดับความรอบรู้ที่แท้จริงของผู้สอบอย่างแน่นอน ซึ่งเป็นไปได้ยากมาก และวิธีของเบส์เป็นวิธีที่สามารถนำไปใช้ได้ เพราะไม่จำเป็นต้องทราบคะแนนระดับความรอบรู้ที่แท้จริงของผู้สอบ

---

[1] R.K. Hambleton, L.R. Hutten and H. Swaminathan, "A Comparison of Several Methods for Assesing Student. Mastery in Objectives-Based Instructional Programs." Journal of Experimental Education. 45(1976) : 57 - 64.

จากวิธีการกำหนดคะแนนจุดตัดต่าง ๆ ดังได้กล่าวมาแล้ว จะเห็นได้ว่าแต่ละวิธีมีจุดอ่อนในตัว เช่น การใช้คะแนนของคนอื่น ๆ เป็นเกณฑ์ ยังคงเป็นการอิงกลุ่มอยู่ซึ่งไม่ได้ให้ข้อมูลเพียงพอว่านักเรียนสามารถหรือไม่สามารถปฏิบัติอะไรได้ และวิธีนับถอยหลังจาก 100 % ก็เป็นการลดระดับลงตามอำเภอใจ ไม่มีหลักเกณฑ์ที่น่าเชื่อถือพอ หรือจะปรับปรุงโดยเพิ่มคะแนนเกณฑ์อื่น ๆ เข้าไปก็ยังคงมีปัญหาอีกว่าเกณฑ์ภายนอกที่นำมาพิจารณานั้น มีความสัมพันธ์กับแบบสอบที่เราจะหาจุดตัดเพียงใด และเกณฑ์ภายนอกนั้นได้มาอย่างไร เป็นต้น ความจริงจุดบกพร่องเหล่านี้อาจแก้ไขได้โดยการใช้วิธีการกำหนดจุดตัดด้วยการใช้ทฤษฎีตัดสินใจ วิธีการวิจัยเชิงปฏิบัติ หรือวิธีการของเบส์ซึ่งใช้สถิติขั้นสูงมาช่วยในการตัดสินใจ ทำให้น่าเชื่อถือยิ่งขึ้น แต่ในแง่ของการนำไปปฏิบัติแล้ว ย่อมจะเป็นการลำบากสำหรับครูผู้สอนทั่ว ๆ ไป ซึ่งไม่ได้มีความรู้ลึกซึ้งเกี่ยวกับสถิติขั้นสูง วิธีการนี้จึงเหมาะสำหรับนักวิชาการที่จะใช้ศึกษาเท่านั้น ส่วนวิธีการที่เกี่ยวข้องกับครูผู้สอนโดยตรงเห็นจะได้แก่วิธีพิจารณาตัดสินจากความสามารถต่ำสุด นอกเหนือจากวิธีการของอีเบลและแอนกอฟ ซึ่งไม่ต้องใช้ค่าสถิติที่ยุ่งยากมาคำนวณ สะดวกในการปฏิบัติ และมีรายละเอียดดังนี้

<u>เทคนิคการหาคะแนนผ่านต่ำสุด</u> (Minimum Passing Score Techniques)

1. วิธีการของนีเดลสกี[1] (Nedelsky, 1954)

นีเดลสกีเสนอเทคนิคในการหาคะแนนผ่านต่ำสุดของแบบสอบปรนัย ที่เรียกว่า "เทคนิคการหาคะแนนการเดาระหว่างนักเรียนเกรด F และ D " (F-D Guess Score Techniques) ซึ่งจะแบ่งนักเรียนออกเป็น 2 พวก คือพวกที่อยู่เหนือจุดแบ่งเป็นพวกที่อยู่ระดับต่ำสุดที่จะยอมให้ผ่าน (D Students)และพวกที่อยู่ต่ำกว่าจุดแบ่งคือพวกที่ไม่ผ่าน (F Students) ส่วนพวกที่อยู่ตรงบริเวณเส้นแบ่ง (border-line)

---

[1] L. Nedelsky, "Absolute Grading Standards for Objective Tests." <u>Educational Psychology Measurement</u> 14(1954) : 3 - 19.

คือนักเรียนที่อยู่ระหว่างเกรด F และ D (F-D Students) เขาได้บัญญัติศัพท์ขึ้นใช้ในการคำนวณดังนี้

1.1 ตัวลวงที่นักเรียนระดับต่ำสุดของพวกที่ยอมให้สอบผ่าน (D Students) จะไม่เลือกเพราะรู้ว่าผิด และดึงดูดความสนใจของพวกที่สอบไม่ผ่าน (F Students) เรียกว่า "คำตอบที่ผิด" (F-responses)

1.2 นักเรียนที่มีความรู้เพียงพอที่จะไม่เลือกคำตอบที่ผิด แต่เลือกตัวลวงที่เหลือด้วยการสุ่ม เรียกว่า "นักเรียนที่อยู่ระหว่างเกรด F และ D" (F-D Students) ซึ่งคือผู้ที่มีความรู้อยู่ตรงเส้นแบ่งระหว่างเกรด F และ D

1.3 คะแนนเฉลี่ยของนักเรียนที่อยู่ระหว่างเกรด F และ D ทั้งหมดเรียกว่า "คะแนนการเดาระหว่างนักเรียนเกรด F และ D" (F-D Guess Score) เขียนแทนด้วย M_{FD} ซึ่งจะมีค่าเท่ากับผลรวมของส่วนกลับของจำนวนตัวลวงที่ไม่ใช่คำตอบที่ผิด

1.4 ค่า SD ที่คำนวณจาก M_{FD} เขียนแทนด้วย σ_{FD}

วิธีการของเขาคือ ก่อนนำแบบสอบไปให้นักเรียนทำ จะนำแบบสอบไปให้ครูผู้สอนวิชานั้น ๆ พิจารณาข้อกระทงแต่ละข้อว่า ตัวเลือกใดที่นักเรียนระดับต่ำสุดที่เรายอมให้ผ่านบอกได้ว่าผิด แล้วเขียนส่วนกลับของจำนวนตัวเลือกที่เหลืออยู่ไว้ทางซ้ายมือของข้อกระทง เช่น ถ้าครูเลือกตัวลวงที่นักเรียนระดับต่ำสุดที่เรายอมให้ผ่านบอกได้ว่าผิดได้ 1 ข้อ จาก 5 ข้อ ก็ให้เขียน $\frac{1}{4}$ เป็นต้น แล้วนำเศษส่วนเหล่านี้ของทุกข้อกระทงมารวมกันได้เป็น M_{FD} เพื่อนำไปคำนวณหาค่าคะแนนผ่านต่ำสุดจากสูตร

$$\text{คะแนนผ่านต่ำสุด} = M_{FD} + K\ \sigma_{FD}$$

เมื่อ k เป็นค่าคงที่ซึ่งกำหนดขึ้นจากการพิจารณาหลาย ๆ ครั้ง โดยครั้งแรกครูผู้สอนหลาย ๆ คนจะมาตกลงกันเพื่อกำหนดค่า k ขึ้นทดลองใช้ก่อน ซึ่งเขากล่าวไว้ว่าคะแนนผ่านต่ำสุดจะมีมาตรฐานพอเมื่อนักเรียนที่อยู่ระหว่างเกรด F และ D ส่วนใหญ่สอบตก และจากการศึกษาเขาพบว่า ถ้าให้ k เป็น -1, 0, 1, 2 จะทำให้พวกที่อยู่ระหว่าง

เกรด F และ D สอบตก 16 %, 50 %, 84 % และ 98 % ตามลำดับ การพิจารณาเลือกค่า k ครั้งสุดท้ายจะได้หลังจากที่ให้ผู้สอนหลาย ๆ คนเลือกคำตอบที่นักเรียนระดับต่ำสุดที่ผ่านรู้ว่าผิดแล้ว แต่การกำหนดค่ามาตรฐานสัมบูรณ์ (absolute standard) นั้นควรจะได้กำหนดค่า k ขึ้นมาก่อนจะคำนวณค่า MPD และก่อนที่จะรู้คะแนนการสอบของนักเรียน

คำตอบที่นักเรียนระดับต่ำสุดที่ผ่านรู้ว่าผิด จะมีส่วนทำให้การกำหนดค่า k เปลี่ยนแปลงไปได้ เนื่องจากในการสอบแต่ละครั้ง คำตอบที่ผิดจะมีทั้งที่ผิดอย่างชัดเจนและผิดปานกลาง ซึ่งผู้ที่จะไม่เลือกจะต้องมีความรู้ถึงขั้นนั้นแล้ว ถ้าแบบสอบมีคำตอบชนิดแรกมากก็ควรกำหนดค่า k ให้สูง ๆ แต่ถ้าคำตอบชนิดหลังมากก็ควรกำหนดค่า k ให้ต่ำ และในการพิจารณาเปลี่ยนแปลงค่า k นั้น ไม่ควรจะเปลี่ยนมากไปกว่า $\pm$ 0.5 จากค่า k ที่กำหนดขึ้นเพื่อทดลองใช้ การตัดสินตามทฤษฎีไม่คอยปรากฏว่าค่า k จะสูงถึง 2 และสำหรับแบบสอบถ้ากำหนดให้ k = 0 อาจจะต่ำเกินไป จึงมีข้อเสนอแนะว่าควรเริ่มใช้ที่ k = 1 ก่อน ดังนั้นช่วงของค่า k จะเปลี่ยนไประหว่าง 0.5 ถึง 1.5

2. วิธีการของกิลเบิร์ต[1] (Guilbert, 1976)

กิลเบิร์ตได้นำเอาวิธีการของนีเดลสกีมาดัดแปลง เพื่อคำนวณหาระดับผ่านต่ำสุด (The Minimum Pass Level:MPL) ของแบบสอบแบบเลือกตอบ โดยเขาได้ให้ความหมายของระดับผ่านต่ำสุดไว้ดังนี้

ระดับผ่านต่ำสุด (MPL) คือค่าที่ต่ำที่สุดที่จะช่วยให้เราตัดสินได้ว่าจะยอมให้นักเรียนที่มีความรู้เพียงพอ สอบผ่านหรือไม่ผ่านในวิชานั้น ๆ

การกำหนดคะแนน MPL จะขึ้นอยู่กับการตัดสินแบบสอบของครูผู้สอนเกี่ยวกับความยาก (relative difficulty) ของข้อคำถามแต่ละข้อก่อนที่จะนำแบบสอบไปใช้ ดังนั้นการใช้ MPL จะช่วยให้ทราบว่าข้อคำถามข้อใดยากหรือง่าย ซึ่งจะยังผลทำให้เราทราบว่าแบบสอบทั้งชุดมีความยากง่ายขนาดใดด้วย การคำนวณหาคะแนน MPL

---

[1] Guilbert, "Calculation of the Minimum...," pp.458-459.

ของแบบสอบจะมีความตรงก็ต่อเมื่อแบบสอบเลือกตอบมีความยาวมากกว่า 30 ข้อ

วิธีการที่จะคำนวณหาคะแนน MPL มีขั้นตอนดังนี้

2.1 กรรมการประเมินผลพิจารณาตกลงกันว่า ข้อใดเป็นตัวเลือกที่ถูกต้องของข้อกระทงแต่ละข้อ

2.2 กรรมการพิจารณาแล้วตัดสินว่า ตัวลวงข้อใดบ้างที่นักเรียนจะตัดทิ้งไม่เลือกตอบอย่างแน่นอนโดยปราศจากการเสี่ยง

2.3 กรรมการคำนวณหาค่า "ดัชนีความสามารถยอมรับ" (Acceptability Index : AI) ของแต่ละข้อกระทง

2.4 คะแนน MPL ของแบบสอบชุดนั้น คือผลบวกของดัชนีความสามารถยอมรับของข้อกระทงแต่ละข้อ

การคำนวณหาค่าดัชนีความสามารถยอมรับ ของแบบสอบเลือกตอบให้คำนวณดังนี้

พิจารณาตัวลวงแต่ละตัวอย่างระมัดระวังว่า ข้อใดที่นักเรียนที่มีความรู้เพียงพอที่จะสอบผ่านสามารถที่จะปฏิเสธคือไม่เลือกตอบเพราะรู้ว่าผิดบ้าง ตัวอย่างเช่น ในแบบสอบเลือกตอบที่มี 5 ตัวเลือกซึ่งมีข้อที่ถูกต้องเพียงข้อเดียว เมื่อพิจารณาแล้วเห็นว่ามีตัวลวงที่นักเรียนที่มีความรู้เพียงพอที่จะสอบผ่านจะตัดทิ้ง 1 ข้อ ดังนั้นนักเรียนคาบเส้นจะสามารถตอบได้ถูกโดยมีโอกาสเพียง 1 ใน 4 เท่านั้น ในกรณีนี้ค่าดัชนีความสามารถยอมรับของข้อคำถามข้อนี้จะเป็น 0.25

ดังนั้นสำหรับแบบสอบชนิด 5 ตัวเลือกจะพิจารณาได้ดังนี้

- ถ้าตัวเลือกตัดทิ้งไม่ได้เลย ค่า AI = 1/5 = 0.20
- ถ้าตัวเลือกตัดทิ้งได้ 1 ตัว ค่า AI = 1/4 = 0.25
- ถ้าตัวเลือกตัดทิ้งได้ 2 ตัว ค่า AI = 1/3 = 0.33
- ถ้าตัวเลือกตัดทิ้งได้ 3 ตัว ค่า AI = 1/2 = 0.50
- ถ้าตัวเลือกตัดทิ้งได้ 4 ตัว ค่า AI = 1/1 = 1.00

คะแนน MPL จะมีคุณค่าน้อยถ้าไม่ได้อาศัยการวิเคราะห์คำถามแต่ละข้อโดยละเอียด รวมทั้งจะต้องพิจารณาตัวเลือกที่ผิดให้มากเท่า ๆ กับตัวเลือกที่ถูกด้วย

ความตรงของการคาดคะเนระดับผ่านต่ำสุดยังขึ้นอยู่กับการตัดสินใจจากครูผู้สอนทั้งหลายซึ่งจะต้องคำนึงถึงจุดประสงค์ของการศึกษา และระดับของการสอบนั้นว่ามีจุดมุ่งหมายอย่างไร ในการกำหนดระดับผ่านต่ำสุดนี้ ถ้ายิ่งใช้ครูผู้สอนมากเท่าใดก็จะยิ่งช่วยให้เกณฑ์ระดับผ่านต่ำสุดนี้ดีขึ้น

เมื่อเกิดความแตกต่างระหว่างการพิจารณาตัดสินคะแนน MPL ขึ้น ถ้าแตกต่างกันไม่มากนักให้พิจารณาช่วงคาบเส้น (Grey zone) ดังต่อไปนี้ เช่น ครูผู้สอนคนหนึ่งได้ค่าเฉลี่ยของ MPL ของแบบสอบชุดหนึ่งเป็น 43 % ขณะเดียวกันถ้าครูผู้สอนอีก 2 คน ได้ค่าเฉลี่ยของคะแนน MPL ของแบบสอบชุดเดียวกันนั้นเป็น 45 % และ 47 % ตามลำดับ เสนอว่า คะแนนที่ได้ต่ำกว่า 43 % ถือเป็นตก คะแนนใดที่เหนือ 47 % ถือเป็นผ่าน แต่คะแนนที่อยู่ระหว่าง 43 % และ 47 % ถือว่าอยู่ในช่วงคาบเส้น (Grey zone) ซึ่งจะต้องพิจารณาต่อไปว่าจะให้ผ่านหรือตก แต่ถ้าหากมีความแตกต่างกันมากระหว่างการพิจารณาเกณฑ์ MPL ของผู้สอนหลายท่าน ควรจะต้องปรับปรุงแก้ไขจุดประสงค์การศึกษาใหม่

3. วิธีคำนวณหาระดับผ่านต่ำสุดของมหาวิทยาลัยอิลินอยส์[1]

ในการคำนวณหาระดับผ่านต่ำสุด เขาได้ให้คำนิยามไว้ดังนี้

ระดับผ่านต่ำสุด (MPL) คือ จำนวนคำถามที่นักศึกษาจะต้องตอบให้ได้เพื่อแสดงว่าสอบผ่านวิชาใดวิชาหนึ่ง

---

[1]สมโพธิ พุกกะเวส. "การให้ระดับคะแนน" คู่มืออาจารย์การวัดและประเมินผลการศึกษา. มหาวิทยาลัยมหิดล, การพิมพ์พระนคร, 2521, อ้างจาก Monograph "Setting Standards of Competence : The Minimum Pass Level" University of Illinois College of Medicine, Chicaco, January, 1964

สำหรับขั้นตอนในการกำหนดคะแนน MPL ของแบบสอบชุดใดชุดหนึ่งของเขาอาจทำได้ดังนี้ คือ

3.1 อ่านคำถามอย่างละเอียด แล้วเขียนเครื่องหมาย ✓ ลงหน้าตัวเลือกที่ถูกต้อง

3.2 กำหนดระดับความยากของตัวเลือกที่เหลือ โดยถือเอานักศึกษาระดับคาบเส้นเป็นหลักในการกำหนดความยาก ซึ่งนักศึกษาระดับคาบเส้นนี้ คือ นักศึกษาที่ถือว่าสอบผ่าน แต่อยู่ระดับต่ำสุดที่จะให้ผ่าน การกำหนดระดับความยากอาจทำได้โดยการให้ตัวเลข 0, 1 หรือ 2 กับตัวเลือกแต่ละตัว

- ตัวเลือกใดที่นักศึกษาคาบเส้นรู้ได้ทันทีว่าผิดให้เขียนเลข 0 หน้าข้อนั้น

- ตัวเลือกใดที่นักศึกษาคาบเส้นอาจจะบอกได้ว่าผิดหรือลังเลไม่แน่ใจ ให้เขียนเลข 1 หน้าข้อนั้น

- ตัวเลือกใดที่เราไม่คิดว่านักศึกษาคาบเส้นจะตอบได้ว่าเป็นคำตอบที่ผิด ให้เขียนเลข 2 หน้าข้อนั้น

ตัวเลือกใดที่ไม่แน่ใจว่าจะให้เลข 0 หรือ 1 ดี อาจจะใช้เลข 0.5 ก็ได้ และทำนองเดียวกัน ถ้าไม่แน่ใจว่าจะให้ 1 หรือ 2 ดี ก็อาจใช้เลข 1.5 ได้เช่นกัน

3.3 รวมจำนวน "✓, 0, 1 และ 2" ที่นักศึกษาเลือก

3.4 กำหนดเกณฑ์ ซึ่งอาจเลือกใช้เกณฑ์ใดเกณฑ์หนึ่งดังต่อไปนี้ เพื่อให้การตัดสิน คือ

3.4.1 กำหนดจำนวนต่ำสุดของข้อถูก (Minimal Positives) โดยถือว่านักศึกษาที่จะสอบผ่าน จะต้องเลือกข้อถูกอย่างน้อยกี่ข้อ

3.4.2 กำหนดจำนวนสูงสุดของข้อผิด (Maximal Negatives) โดยถือว่านักศึกษาที่จะสอบผ่าน ต้องเลือกข้อผิดอย่างมากไม่เกินกี่ข้อ หรืออาจระบุว่ามีข้อผิดไม่ได้เลย

การกำหนดเกณฑ์นี้มีความจำเป็นต้องอาศัยการประชุมปรึกษาหารือร่วมกันในระหว่างคณาจารย์ที่มีความเชี่ยวชาญในวิชาที่จะกำหนดเกณฑ์อย่างละเอียดถี่ถ้วนที่สุด

นอกจากนี้ยังมีวิธีคำนวณ หาระดับผ่านต่ำสุดของแบบสอบชนิดเลือกตอบที่กำหนดระดับความยากโดยให้ตัวเลข 0, 1 หรือ 2 อีกวิธีหนึ่ง โดยการหาดัชนีความสามารถยอมรับ (Acceptability Index : AI) ของข้อกระทงแต่ละข้อ จากสูตร

$$AI = \frac{\text{ตัวคูณที่}}{\text{ผลรวมของค่าตัวเลือก}}$$

ตัวอย่างเช่น ในแบบสอบเลือกตอบ 5 ตัวเลือก ผู้ตัดสินกำหนดระดับความยากของตัวเลือกข้อ ก, ข, ค, ง และ จ เป็น 0, 0, 1, 2, 2 ตามลำดับ จะคำนวณค่า AI ได้คือ

$$AI = \frac{2}{0+0+1+2+2} = \frac{2}{5} = 0.49$$

เมื่อให้ค่าตัวคูณที่เท่ากับ 2

ถ้าค่า AI สูง แสดงว่าข้อสอบข้อนั้นง่าย ถ้า AI ต่ำ แสดงว่าข้อสอบข้อนั้นยาก

จากนั้นนำเอา AI ของแต่ละข้อกระทงมารวมกัน หารด้วยจำนวนข้อกระทงทั้งหมด ก็จะได้ระดับผ่านต่ำสุดของข้อสอบชุดนั้น

$$\therefore MPL = \frac{\sum AI}{N}$$

องค์ประกอบในการพิจารณาเพื่อกำหนดคะแนนจุดตัด

มิลแมน (Milman, 1973)[1] กล่าวว่า การพิจารณาเรื่องนี้ควรพิจารณาองค์ประกอบหลาย ๆ ด้านประกอบกัน เช่น

1. คะแนนของผู้สอบคนอื่น ๆ (Performance of others)

กำหนดโดยพิจารณาจำนวนนักเรียนที่สอบผ่านไว้ก่อน แล้วเลือกจำนวนเปอร์เซนต์ของผู้ผ่านไว้ หรืออาจใช้คะแนนดิบที่ตรงกับค่าเปอร์เซนไทล์ (Percentile) ที่เราเลือกหลังจากใช้แบบสอบกับกลุ่มผู้เข้าสอบที่มีความเข้าใจอย่างแจ่มแจ้งในเนื้อหาเป็นอย่างดี แต่วิธีนี้ขัดแย้งกับปรัชญาการใช้แบบสอบอิงเกณฑ์ที่ว่า ผู้สอบควรได้รับการตัดสินว่าผ่านหรือไม่ผ่าน โดยพิจารณาจากความสามารถที่เขาทำได้ตามเกณฑ์ที่วางไว้

2. เนื้อหาของข้อกระทง (Item Content)

ข้อกระทงควรได้รับการพิจารณาตัดสินจากกลุ่มผู้เชี่ยวชาญว่า จำนวนข้อที่น้อยที่สุดที่ผู้สอบควรตอบได้ถูกต้องควรจะเป็นเท่าใด จึงจะได้รับการพิจารณาว่าเป็นผู้รอบรู้ได้ หรืออาจจะให้ผู้เชี่ยวชาญประมาณค่าความน่าจะเป็นที่ผู้เข้าสอบซึ่งมีความสามารถต่ำสุดจะตอบข้อกระทงแต่ละข้อได้ถูกต้อง แล้วนำค่าความน่าจะเป็นที่ได้จากทุกข้อกระทงรวมกัน จะได้เป็นคะแนนเกณฑ์

3. ผลทางการศึกษาที่ตามมา (Educational Consequences)

จะทำให้เห็นความสัมพันธ์ของการทำแบบสอบและการวัดโดยใช้เกณฑ์มากขึ้น เช่น การวัดวัตถุประสงค์เรื่องต้น ๆ หรือวัตถุประสงค์ที่เป็นพื้นฐานของวัตถุประสงค์ข้อต่อไป คะแนนจุดตัดควรจะสูง แต่ถ้าสิ่งที่วัดไม่มีความจำเป็นต่อการเรียนรายวิชาต่อไป คะแนนจุดตัดควรจะต่ำลงหรืออาจไม่มีเลยก็ได้ ฉะนั้นควรจะต้องพิจารณาถึงจุดมุ่งหมายในการสอนและผลที่ตามมาเป็นสำคัญด้วย เนื่องจากถ้าเราตั้งเกณฑ์ไว้ต่ำเกินไป นักเรียนที่

---

[1] J. Milman. "Passing Score and Test Lengths for Domain-Referenced Measures." Review of Educational Research. 43 (1973): 206 - 211.

ผ่านเพื่อจะไปเรียนบทเรียนต่อไป อาจเรียนมโนทัศน์และทักษะใหม่ได้อย่างขาดประสิทธิภาพ แต่ถ้าตั้งเกณฑ์ไว้สูงเกินไปก็จะทำให้เสียเวลาในการเรียนซ่อมเสริมโดยไม่จำเป็น และต้องเสียทรัพยากรไปโดยเปล่าประโยชน์ สรุปก็คือ ควรพิจารณาให้ดีถ้วนว่าคะแนนเกณฑ์ควรจะเป็นเท่าใด จึงจะทำให้เกิดประโยชน์ทางการศึกษามากที่สุด

4. คุณค่าทางจิตวิทยาและทางด้านเศรษฐกิจ (Psychological and Financial Costs)

ถ้าคุณค่าทางจิตวิทยาและทางด้านเศรษฐกิจที่เกี่ยวข้องกับโปรแกรมการสอนซ่อมเสริมสูงมาก ก็ควรจะกำหนดคะแนนจุดตัดให้ต่ำลง แต่ในสถานการณ์ที่ค่าใช้จ่ายในเรื่องนี้ต่ำ หรืออาจจะสูงแต่มีความคลาดเคลื่อนในการพิจารณาให้ผู้ที่ยังไม่ได้เรียนสอบผ่าน ซึ่งจะทำให้เกิดผลทางลบ เช่น ทำให้ประสิทธิภาพในการเรียนลดลง ก็ควรกำหนดคะแนนจุดตัดให้สูง

5. ความคลาดเคลื่อนที่เนื่องมาจากการเดาและการสุ่มข้อกระทง (Error Due to Guessing and Item Sampling)

ถ้ารูปแบบของข้อกระทงเป็นแบบชี้แนะหรือเปิดโอกาสให้นักเรียนตอบถูกได้โดยการเดา เราอาจแก้ไขได้โดยการกำหนดคะแนนจุดตัดไว้สูง ๆ ทั้งนี้เชื่อว่าถ้ากำหนดคะแนนจุดตัดไว้สูง การเดาจะไม่มีผลต่อผู้ที่ได้คะแนนใกล้เคียงกับคะแนนจุดตัด ส่วนความคลาดเคลื่อนในการสุ่มข้อกระทงเป็นความลำเอียงที่เนื่องมาจากการไม่พิจารณาชนิดของคำถามและเนื้อหาในโดเมน ทำให้การเลือกตัวอย่างไม่ครอบคลุมวัตถุประสงค์ที่ตั้งไว้ การตัดข้อกระทงที่วัดวัตถุประสงค์ข้อนั้นไป อาจเป็นเพราะโครงสร้างยาก ไม่สะดวกในการบริหารการสอบหรือข้อกระทงมีความคลุมเครือ ก็ควรจะได้มีการปรับคะแนนจุดตัดเพื่อแก้ไขจุดบกพร่องนี้

ข้อเสนอแนะเกี่ยวกับการกำหนดคะแนนจุดตัด

ในการกำหนดคะแนนจุดตัดมีสิ่งที่ควรคำนึง ดังที่แฮมเบิลตัน (Hambleton, 1978)[1]

[1] R.K. Hambleton. "Use of Cut-off Scores." Journal of Educational Measurement. 15 (1978): 288 - 289.

ได้ให้ข้อเสนอแนะที่เขาได้จากประสบการณ์ในการทำงานจากหลาย ๆ โรงเรียน ดังนี้

1. การตัดสินคะแนนจุดตัดควรทำโดยกลุ่มที่ทำงานร่วมกันหลาย ๆ กลุ่ม ได้แก่ ครู ผู้ปกครอง ผู้เชี่ยวชาญ ด้านหลักสูตร ผู้บริหารโรงเรียนและนักเรียน (ในกรณีที่เป็นนักเรียนระดับมัธยมศึกษาตอนปลาย) จำนวนคนในแต่ละกลุ่มขึ้นอยู่กับความสำคัญของแบบสอบที่จะนำมาพิจารณาและจำนวนตารางวิเคราะห์โดเมนของข้อสอบ ซึ่งน่าจะมีจำนวนคนเพียงพอที่จะแบ่งผู้พิจารณาออกเป็น 2 กลุ่ม เพื่อเปรียบเทียบความคงที่ของการตัดสินจากบุคคล 2 กลุ่มนี้

2. ควรได้นำตารางวิเคราะห์โดเมนของข้อสอบมาเสนอและให้ผู้ตัดสินอภิปรายร่วมกัน

3. วิธีการกำหนดคะแนนจุดตัดของอีเบลหรือนีเดลสกี ควรจะมีการทดลองทำกับกลุ่มตัวอย่างหลายพวก เพื่อตัดสินและแก้ปัญหาเกี่ยวกับความแตกต่างระหว่างกลุ่ม

4. การให้เวลาในการพิจารณาตารางโดเมนของข้อสอบแต่ละตารางควรจะกำหนดอย่างเหมาะสม หากตารางใดซับซ้อนหรือสำคัญมากก็ควรให้เวลามากขึ้น

5. จะต้องมั่นใจว่าผู้ตัดสินมีความรู้เกี่ยวกับการใช้แบบสอบ และรู้ว่าจะใช้กับนักเรียนกลุ่มใด

6. หากตารางวิเคราะห์โดเมนใดเป็นสิ่งที่ต้องการหรือจำเป็นกว่าตารางอื่นก็ควรกำหนดคะแนนจุดตัดให้สูงกว่า

7. พยายามใช้บุคคลตั้งแต่ 2 กลุ่มขึ้นไป เพื่อพิจารณาตัดสินคะแนนจุดตัดและศึกษาความคงที่และความแตกต่างในการกำหนดน้ำหนักของผู้ตัดสินและทำการอภิปรายกันเพื่อหาความสอดคล้องในการตัดสิน

8. ถ้าข้อมูลในการสอบครั้งก่อน ๆ มีประโยชน์ ก็สามารถนำมาใช้พิจารณาเปลี่ยนแปลงแก้ไขคะแนนจุดตัดและเวลาที่ใช้สอบได้

9. ถ้าข้อมูลที่ได้จากการสอบเหมาะที่จะนำมาศึกษาถึงเปอร์เซนต์ของผู้รอบรู้และผู้ไม่รอบรู้ในแต่ละวัตถุประสงค์ก็ควรจะทำ แต่ถ้าหากคะแนนที่ได้มีทีท่าว่าจะเป็นไปในลักษณะที่ผิดปกติ เราก็สามารถอธิบายสาเหตุได้จากการพิจารณาข้อกระทงของแบบสอบ ระดับของคะแนนจุดตัด ความแปรผันของคะแนนสอบระหว่างชั้นเรียนและเวลาที่ใช้ในการสอน

10. การเปรียบเทียบสถานภาพการรอบรู้ของผู้สอบกลุ่มที่เรียนแล้วและยังไม่ได้เรียน กลุ่มที่เรียนแล้วควรมีจำนวนผู้รอบรู้มากกว่ากลุ่มที่ยังไม่ได้เรียน และกลุ่มที่ยังไม่ได้เรียนก็ควรจะมีจำนวนผู้ไม่รอบรู้มากกว่า ถ้านักเรียนเป็นจำนวนมากได้รับการพิจารณาแบ่งกลุ่มผิด ก็ควรที่จะพิจารณาหาคะแนนจุดตัดที่มีความตรงมากกว่า

11. ถ้าหลักสูตรมีการเปลี่ยนแปลงก็จะมีผลกระทบถึงคะแนนจุดตัด ดังนั้นจึงต้องมีการตรวจสอบคะแนนจุดตัดให้มีความเหมาะสมด้วย

ความเที่ยงของแบบสอบอิงเกณฑ์

ในการประมาณค่าความเที่ยงของแบบสอบอิงเกณฑ์มีผู้ศึกษาและเสนอวิธีการไว้หลายวิธีด้วยกัน แต่พอจะแบ่งตามแนวความคิดได้เป็น 2 ฝ่าย คือ

ฝ่ายที่ 1 เห็นว่าความแปรปรวนของการวัดอิงเกณฑ์ เป็นความเบี่ยงเบนของคะแนนผู้สอบแต่ละคนออกจากคะแนนเกณฑ์ ยิ่งความเบี่ยงเบนนี้มีค่ามากเท่าใด ความถูกต้องของการตัดสินให้ผู้สอบได้หรือตกก็มีมากยิ่งขึ้นเท่านั้น ผู้ที่มีความเชื่อตามแนวความคิดนี้ ได้แก่ลิฟวิงสตัน (Livingston, 1971) เขาได้เสนอสูตรโดยมีข้อตกลงเบื้องต้นว่า เพื่อแยกคะแนนของแต่ละคนออกจากคะแนนเกณฑ์ คือ[1]

$$r_{cc} = \frac{r_{xx}s_x^2 + (\bar{X} - C)^2}{s_x^2 + (\bar{X} - C)^2}$$

เมื่อ r_{xx} คือ ความเที่ยงของแบบสอบอิงกลุ่ม

s_x^2 คือ ความแปรปรวนของคะแนน

$\bar{X}$ คือ ค่ามัชฌิมเลขคณิต

C คือ คะแนนเกณฑ์

r_{cc} คือ ความเที่ยงของแบบสอบอิงเกณฑ์

---

[1] W.A. Mehrens and J.J. Lehmann. Measurement and Evaluation in Psychology (New York: Holt, Rinehart and Winston, 1973), p. 121.

หรืออาจจะใช้เกี่ยวกับคะแนนโดเมน ดังนี้[1]

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

เมื่อ $\hat{\pi}$ คือ ค่าประมาณคะแนนโดเมน

π คือ คะแนนโดเมน

π_0 คือ คะแนนเกณฑ์

$\bar{\pi}$ คือ ค่ามัชฌิมเลขคณิตของคะแนนโดเมน

$\sigma^2_{(\pi)}$ คือ ความแปรปรวนของคะแนนโดเมนที่กระจายออกจากคะแนนจุดตัด

$\sigma^2_{(\hat{\pi})}$ คือ ความแปรปรวนของคะแนนโดเมนที่ได้จากการประมาณที่กระจายออกจากคะแนนจุดตัด

จะเห็นได้ว่า สูตรนี้ถึงแม้ความแปรปรวนของคะแนนโดเมนจะเป็นศูนย์ค่าความเที่ยงของแบบสอบก็ยังคงมากกว่าศูนย์อยู่ และค่าความเที่ยงจะเพิ่มมากขึ้นเมื่อ $(\bar{\pi} - \pi_0)^2$ เพิ่มขึ้น หรือถ้ามัชฌิมเลขคณิตของคะแนนโดเมนเบี่ยงเบนออกจากคะแนนจุดตัดมาก ความเที่ยงก็จะยิ่งสูงขึ้น สูตรนี้สามารถดัดแปลงให้อยู่ในรูป[2]

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

เมื่อ $\sigma^2_{(Tx)}$ คือ ความแปรปรวนของคะแนนจริง

$\sigma^2_{(x)}$ คือ ความแปรปรวนของคะแนนจากการสังเกต

μ_x คือ ค่ามัชฌิมเลขคณิต

C_x คือ คะแนนจุดตัด

---

[1] Hambleton, "Criterion-Referenced Testing...," p.16.

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

เพื่อให้สะดวกต่อการนำไปใช้เขาจึงแปลงสูตรนี้เป็น[1]

$$K^2(x, Tx) = \frac{\rho^2_{(x, Tx)}\sigma^2_{(x)} + (\mu_x - C_x)^2}{\sigma^2_{(x)} + (\mu_x - C_x)^2}$$

เมื่อ $\rho^2_{(x, Tx)}$ คือ ความเที่ยงของแบบสอบแบบอิงกลุ่ม

$\sigma^2_{(x)}$ คือ ความแปรปรวนของคะแนนดิบ

μ_x คือ ค่ามัชฌิมเลขคณิต

C_x คือ คะแนนจุดตัด

จากสูตรข้างต้นจะเห็นได้ว่า คะแนนเกณฑ์ยิ่งต่างจากค่ามัชฌิมเลขคณิตมากเท่าใด ค่าความเที่ยงอิงเกณฑ์ก็จะยิ่งมากขึ้นเท่านั้น หรือ ถ้าคะแนนเกณฑ์ยิ่งใกล้เคียงกับค่ามัชฌิมเลขคณิต ค่าความเที่ยงแบบอิงเกณฑ์จะยิ่งลดลง และถ้าค่าทั้งสองนี้เท่ากัน ค่าความเที่ยงบองแบบสอบแบบอิงเกณฑ์จะเท่ากับความเที่ยงแบบอิงกลุ่ม

ในกรณีที่ต้องการหาความเที่ยงของแบบสอบซึ่งข้อมูลที่ได้ไม่ใช่ข้อมูลแบบแบ่งเป็นสอง (Binary Data) อาจหาได้จากสัมประสิทธิ์แอลฟา[2]

$$\alpha = \frac{\frac{I}{I-1}\sigma^2_{(x)} - \Sigma\sigma^2_{(i)} + (\mu_x - C_x)^2}{\sigma^2_{(x)} + (\mu_x - C_x)^2}$$

เมื่อ I คือ จำนวนข้อกระทง

$\sigma^2_{(i)}$ คือ ความแปรปรวนของข้อกระทงแต่ละข้อ

ฝ่ายที่ 2 พิจารณาความเที่ยงของแบบสอบอิงเกณฑ์ในแง่ของความคงที่ในการตัดสินใจ โดยมีความเชื่อว่าแบบสอบจะมีความเที่ยงสูงถ้ามีแบบสอบคู่ขนานหลายฉบับมา

---

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

[2] Livingston. "Criterion-Referenced Application...," p.19.

ทดสอบกับคนกลุ่มเดิม แล้วจำแนกผู้สอบเป็นผู้รอบรู้หรือไม่รอบรู้ได้คงที่แน่นอนทุกครั้ง ผู้เสนอสูตรตามแนวคิดนี้ได้แก่

คาร์เวอร์ (Carver, 1970) ใช้กลุ่มตัวอย่างกลุ่มเดียวทำแบบสอบคู่ขนาน หลังจากเรียนแล้วประมาณค่าความเที่ยงจากการเปรียบเทียบสัดส่วนของความคงที่ในการตัดสินได้-ตก ดังนี้[1]

ฉบับ ก / ฉบับ ข	ตก	ได้
ได้	b	a
ตก	c	d

ความเที่ยง $= \frac{a + c}{N}$ เมื่อ $N = a + b + c + d$

จะเห็นได้ว่าสูตรของคาร์เวอร์นี้ เมื่อคะแนนเกณฑ์เปลี่ยนแปลงไป จะทำให้จำนวนนักเรียนในแต่ละช่องของตาราง 2 × 2 เปลี่ยนไปด้วย ฉะนั้นค่าความเที่ยงก็จะเปลี่ยนไปเช่นกัน

ผู้ที่เสนอการหาความเที่ยงแบบอิงเกณฑ์ตามแนวคิดนี้อีกคนหนึ่ง คือ ซับโคเวียค (Subkoviak ,1976)[2] เขาเสนอสูตรที่ใช้กับแบบสอบคู่ขนานเพื่อแบ่งระดับการรอบรู้

---

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

[2]M.J. Subkoviak. "Estimating Reliability from a Single Administration of Criterion-Referenced Test." Journal of Educational Measurement. 13(1976) : 267 - 268.

เป็น 2 ระดับ จากสูตร

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

โดยที่ $P_c^{(i)} = Prob(Xi \geq C)*Prob(Xi' \geq C)+Prob(Xi < C)*Prob(Xi' < C)$

เมื่อ Xi คือคะแนนของผู้เข้าสอบคนที่ i จากแบบสอบฉบับที่ 1

Xi' คือคะแนนของผู้เข้าสอบคนที่ i จากแบบสอบฉบับที่ 2 ซึ่งเป็นแบบสอบคู่ขนานกับฉบับที่ 1

C คือคะแนนจุดตัด

ข้อตกลงเบื้องต้นที่ซับโคเวียดเสนอสูตรนี้ กล่าวไว้ว่า

1. คะแนนจากแบบสอบฉบับที่ 1 และ ฉบับที่ 2 ของผู้สอบแต่ละคน มีการแจกแจงที่เป็นอิสระต่อกัน คือ ประสบการณ์ในการทำแบบสอบฉบับที่ 1 ไม่มีผลต่อการทำฉบับที่ 2

2. การแจกแจงของแบบสอบทั้ง 2 ฉบับของผู้เข้าสอบแต่ละคนเป็นแบบไบโนเมียล (Binomail) คือมีการให้คะแนนแบบ 0-1 และความน่าจะเป็นของคำตอบที่ถูกต้องจะคงที่ในทุกข้อกระทง นอกจากนี้ผลจากการทำข้อกระทงแต่ละข้อจะไม่มีผลต่อข้อกระทงอื่นด้วย

ดังนั้นจะเห็นได้ว่า

$$P_c^{(i)} = \left[Prob\ (Xi \geq c)\right]^2 + \left[Prob(Xi < c)\right]^2$$

$$= \left[Prob\ (Xi \geq c)\right]^2 + \left[1 - Prob\ (Xi \geq c)\right]^2$$

เมื่อ $Prob\ (Xi \geq c) = \sum_{xi=c}^{n} \binom{n}{Xi} P_i^{xi} (1 - P_1)^{n-xi}$

โดยที่ P_i คือ ความน่าจะเป็นที่แท้จริงของคำตอบในข้อกระทงที่ถูกต้องสำหรับผู้เข้าสอบคนที่ i ซึ่งสามารถประมาณได้จากคะแนนการสอบฉบับเดียว คือ

$$\hat{P_i} = X_i/n$$

วิธีการของซับโคเวียคนี้จะเห็นได้ว่า เมื่อคะแนนเกณฑ์เปลี่ยนแปลงจะทำให้ค่าความน่าจะเป็นทั้ง 4 ค่าเปลี่ยนแปลง ซึ่งจะมีผลกระทบทำให้ค่าความเที่ยงเปลี่ยนแปลงไปด้วย

สแวมมินาธาน, แฮมเบิลตัน และอัลกินา[1] เสนอวิธีหาค่าดัชนีความเที่ยงจากการหาสัมประสิทธิ์แคปป้า (Kappa-K) จากการบริหารการสอบ 2 ครั้ง แทนการใช้แบบสอบคู่ขนาน ดังมีสูตรดังนี้

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

โดยที่ $P_c = \sum_{k=1}^{m} (P_{k\cdot})(P_{\cdot k})$

เมื่อ m คือ จำนวนประเภทที่ต้องการจำแนก

$P_{k\cdot}$ แทนอัตราส่วนของผู้เข้าสอบที่จัดให้เป็นผู้รอบรู้ในการสอบครั้งที่ 1

$P_{\cdot k}$ แทนอัตราส่วนของผู้เข้าสอบที่จัดให้เป็นผู้รอบรู้ในการสอบครั้งที่ 2

ผู้ที่เสนอการคำนวณค่าความเที่ยงในรูปสัมประสิทธิ์แคปป้าอีกคนหนึ่งได้แก่ ฮวน[2] ซึ่งได้เสนอสูตรไว้ดังนี้

---

[1] R.K. Hambleton and H. Swaminathan. "Criterion-Referenced Testing and Measurement: A Review of Technical Issues and Developments." Review of Educational Research. 48(1978) : 21.

[2] H. Huynh. "On the Reliability of Decisions in Domain-Referenced Testing." Journal of Educational Measurement. 13(1976): 254 - 257.

1. เมื่อคะแนนเกณฑ์มีค่าสูงใกล้กับจำนวนข้อกระทงทั้งหมดในแบบสอบ (n) ควรใช้สูตร $K = (P_{11} - P_1^2) / (P_1 - P_1^2)$

2. เมื่อคะแนนเกณฑ์มีค่าต่ำ ห่างจากจำนวนข้อกระทงทั้งหมดของแบบสอบ (n) มาก ควรใช้สูตร

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

โดยที่

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

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

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

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

แฮมเบิลตัน และโนวิค[1] เสนอสูตรสำหรับการสอบโดยมีจุดประสงค์เพื่อจำแนกผู้สอบออกเป็น m ระดับการรอบรู้ คือ

$$P_o = \sum_{k=1}^{m} P_{kk}$$

เมื่อ P_o คือ ดัชนีความเที่ยงแบบอิงเกณฑ์

P_{kk} คือ สัดส่วนของผู้สอบที่ถูกตัดสินให้มีระดับการรอบรู้ที่ระดับ k

---

[1] R.K. Hambleton and H. Swaminathan. "Criterion-Referenced Testing...," p. 21.

นอกจากนี้ แฮริส[1] ได้เสนอวิธีหาค่าสัมประสิทธิ์ ในรูปของการวิเคราะห์ความแปรปรวน เป็นค่าสหสัมพันธ์กำลังสองระหว่างระดับการรอบรู้กับคะแนนรวม สำหรับการแบ่งระดับการรอบรู้เป็น 2 ระดับ และให้คะแนน 0 สำหรับผู้ไม่รอบรู้ ให้คะแนน 1 สำหรับผู้รอบรู้ โดยมีสูตรดังนี้

$$\mu_c^2 = \frac{SS_B}{SS_B + SS_w}$$

เมื่อ SS_B คือ ผลบวกกำลังสองระหว่างกลุ่ม

SS_w คือ ผลบวกกำลังสองภายในกลุ่ม

สำหรับการแจกแจงแบบสมมาตรค่า μ_c^2 จะมากที่สุดเมื่อ $c = \mu$

วิธีการนี้ ค่าสัมประสิทธิ์ μ_c^2 จะมีความไวต่อสัดส่วนของกลุ่มผู้รอบรู้และไม่รอบรู้มาก

ฮวน[2] (1976) ศึกษาค่าความเที่ยงของแบบสอบโดยใช้สัมประสิทธิ์แคปป้า โดยพิจารณาความแปรปรวนของคะแนนสอบ ความยาวของแบบสอบ และความสัมพันธ์ของค่าความเที่ยงและจุดตัด เขาตั้งข้อตกลงไว้ว่า การแจกแจงของคะแนนสอบเป็นแบบไบโนเมียล บริหารการสอบเพียงครั้งเดียว โดยใช้แบบสอบวิชาเลขคณิต ที่มีความยาว 5 ข้อ ทดสอบกับนักเรียน 91 คน กำหนดคะแนนจุดตัดเป็น 4 ได้ค่าสัมประสิทธิ์แคปป้า .52 และใช้แบบสอบวิชาเลขคณิตที่มีความยาว 15 ข้อ ทดสอบกับนักเรียน 96 คน ใช้คะแนนจุดตัดเป็น 12 ได้ค่าแคปป้า .57 นอกจากนี้ยังพบว่าในด้านความเป็นวิวิธพันธ์ของคะแนนสอบ ค่าแคปป้าจะมีสหสัมพันธ์ทางบวกกับความแปรปรวนของคะแนนสอบ ค่าแคปป้าจะเพิ่มขึ้นตามความยาวของแบบสอบ โดยอัตราการเพิ่มจะลดลงเมื่อความยาวของแบบสอบเพิ่มขึ้น

---

[1] Subkoviak. "Estimating Reliability from...," pp.265-266.

[2] Huynh. "On the Reliability...," pp. 253 - 264.

และค่าแคปป้านี้จะสูงขึ้นเมื่อคะแนนจุดตัดสูงขึ้นด้วย แต่ถ้าคะแนนจุดตัดสูงขึ้นมาก ๆ ค่าแคปป้ากลับลดลง

ซับโคเวียค[1] (1976) เสนอวิธีการหาค่าความเที่ยงโดยการบริหารการสอบเพียง 1 ครั้ง เช่นเดียวกันกับฮวน โดยเขาได้ทำแบบสอบที่มีความยาว 5 ข้อ ไปทดสอบกับนักเรียน 10 คน กำหนดให้คะแนนจุดตัดเป็น 4 แล้วประมาณค่าความเที่ยงจากสูตร

$$\hat{P}_i = \alpha 21/x \left(\frac{xi}{n}\right) + (1 - \alpha 21/x)\left(\frac{Mx}{n}\right)$$

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

ได้ค่า $Pc = .75$

จากการนำไปเปรียบเทียบกับค่าความเที่ยงที่ได้จากการบริหารการสอบ 2 ครั้ง ที่คำนวณจากสูตรของแฮมเบิลตัน และโนวิค (1973) เขาได้ค่า $Po = .80$

นอกจากนี้เขาได้ทดลองกับนักเรียนที่มีขนาดใหญ่ขึ้น โดยใช้แบบสอบที่มีความยาว 25 ข้อ กับนักเรียน 69 คน เปรียบเทียบค่า $\hat{P}_c$ ที่ได้จากการแจกแจงแบบซิมเปิลไบโนเมียล (Simple binomial) และการแจกแจงแบบคอมเพานด์ไบโนเมียล (Compound binomial) กับค่า $\hat{P}_o$ พบว่าค่า $\hat{P}_o$ สูงกว่า $\hat{P}_c$ จากการแจกแจงแบบซิมเปิลไบโนเมียล ด้วยค่ามัธยฐาน 6.5 % และสูงกว่า $\hat{P}_c$ จากการแจกแจงแบบคอมเพานด์ไบโนเมียล ด้วยค่ามัธยฐาน 3.5 % ปรากฏผลว่าความเที่ยงจะมีค่าสูงสุดเมื่อคะแนนจุดตัดสูงสุดและต่ำสุด และความเที่ยงจะมีค่าต่ำเมื่อจุดตัดอยู่ตรงกลาง ๆ

---

[1] Subkoviak. "Estimating Reliability from...," pp. 267 - 275.

ในปี 1978 ซับโคเวียค[1] ได้ศึกษาเปรียบเทียบวิธีการคำนวณค่าดัชนีความเที่ยง 4 วิธี ซึ่งได้แก่

1. วิธีของสแวมมินาธาน แฮมเบิลตันและอัลกินา
2. วิธีของฮวน
3. วิธีของมาร์แชลและแฮร์เทล
4. วิธีของซับโคเวียค

เขาใช้ข้อมูลที่ได้จากการให้นักเรียน 1586 คน ทำแบบสอบความถนัดทางวิชาการที่มีความยาว 10, 30 และ 50 ข้อ โดยเป็นแบบสอบคู่ขนานกัน โดยแบ่งนักเรียนออกเป็น 50 กลุ่ม กลุ่มละ 30 คน ด้วยวิธีการสุ่มแทนที่ เมื่อใช้เกณฑ์ตัดสินเป็น 50, 60, 70 และ 80 ข้อ พบว่าความคลาดเคลื่อนมาตรฐานจะลดลงเมื่อแบบสอบยาวขึ้น ค่าประมาณที่ได้จะคงที่และถูกต้องมากขึ้น เมื่อค่าความเที่ยงที่ได้จากประชากรมีค่าสูงมากหรือต่ำมาก จากการเปรียบเทียบวิธีคำนวณทั้ง 4 วิธีแล้ว พบว่าวิธีของสแวมมินาธานจะมีค่าความคลาดเคลื่อนมาตรฐานสูงกว่าวิธีอื่น ๆ แต่วิธีของฮวนจะให้ค่าประมาณที่ถูกต้องที่สุด

อัลกินาและโน[2] (1978) ได้นำสูตรของซับโคเวียคไปใช้หาค่าความเที่ยง

---

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

[2] J. Algina and M.J. Noe. "A Study of Subkoviak's Single-Administration Estimate of Coefficient of Agreement Using Two-True Score Estimates." Journal of Educational Measurement. 15(1978) : 101 - 109.

ในการบริหารการสอบเพียงครั้งเดียว โดยใช้ค่าประมาณคะแนนจริง 2 ค่าคือ P_i และ $\hat{T}_i$ ซึ่งคำนวณได้จากสูตร

$$\hat{T}_i = \hat{\beta} P_i + (1 - \hat{\beta})\hat{\mu}_P \qquad (i = 1,2,\ldots,N)$$

เมื่อ $\hat{\beta}$ คือ สัมประสิทธิ์ความเที่ยงจากสูตร KR # 20

P_i คือ คะแนนสัดส่วนความถูกต้อง ของผู้เข้าสอบคนที่ i

$\hat{\mu}_P$ คือ มัชฌิมเลขคณิตของคะแนนสัดส่วนความถูกต้อง

องค์ประกอบที่เขานำไปพิจารณาคือ คะแนนจุดตัด ความเที่ยงของแบบสอบและจำนวนผู้เข้าสอบ ได้ข้อค้นพบว่า ค่าประมาณของสัมประสิทธิ์ความเที่ยงในการตัดสินจะมีค่าสูงสุดเมื่อคะแนนจุดตัดมีค่าเข้าใกล้มัชฌิมเลขคณิตของคะแนนดิบ โดยมีค่าเป็นบวกเมื่อใช้ P_i และเป็นลบเมื่อใช้ $\hat{T}_i$ เมื่อส่วนเบี่ยงเบนระหว่างคะแนนจุดตัดกับมัชฌิมเลขคณิตของคะแนนดิบมีค่าสัมบูรณ์เพิ่มขึ้น ความคลาดเคลื่อนจะลดลงจนเครื่องหมายเปลี่ยนไปแล้วก็จะเพิ่มขึ้นอีก และในที่สุดก็จะลดลง ความเปลี่ยนแปลงนี้จะเห็นได้ชัดในแบบสอบที่มีความยาวมาก (20 ข้อ)

ในด้านความเที่ยงของแบบสอบ แบบสอบที่มีค่าความเที่ยงต่างกันจะมีผลต่อความเที่ยงในการตัดสินต่างกัน และเมื่อใช้ค่าประมาณคะแนนจริงต่างกัน ก็จะให้ผลเช่นเดียวกัน ความคลาดเคลื่อนของ P_i มีแนวโน้มจะลดลงเมื่อความเที่ยงของแบบสอบเพิ่ม ความคลาดเคลื่อนของ $\hat{T}_i$ มีแนวโน้มที่จะเพิ่มขึ้นเมื่อความเที่ยงของแบบสอบเพิ่ม

อัลกินาและคณะแนะไว้ว่า แบบสอบที่มีค่าความเที่ยงเกิน .50 ควรใช้ P_i ประมาณค่าคะแนนจริง และถ้าใช้ $\hat{T}_i$ ประมาณค่าคะแนนจริงแล้วทำให้ความเที่ยงในการตัดสินมีค่ามาก ก็อาจใช้ค่าเฉลี่ยระหว่าง P_i และ $\hat{T}_i$ เป็นค่าประมาณคะแนนจริง

เกี่ยวกับขนาดของกลุ่มตัวอย่างพบว่า ความแปรผันของค่าประมาณทั้ง 2 จะเพิ่มขึ้นเมื่อจำนวนผู้สอบลดลง แต่ไม่มากนัก และการใช้การแจกแจงแบบซิมเปิลไบโนเมียลให้ผลไม่แตกต่างจากการแจกแจงแบบคอมเพานด์ไบโนเมียล