

# บทที่ 3

## การกำหนดมาตรฐานคะแนนโดยพิจารณา จากตัวคำถาม

### I จุดประสงค์การเรียนรู้

หลังจากจบบทเรียนนี้แล้ว นักศึกษาสามารถปฏิบัติได้ดังนี้

1. อธิบายความหมาย การตัดสินใจ มาตรฐาน และการพิจารณา
2. เปรียบเทียบความแตกต่างของมาตรฐานสัมบูรณ์และมาตรฐานสัมพัทธ์
3. อธิบายผลของความผิดพลาดจากการตัดสินใจ 2 รูปแบบ
4. อธิบายขั้นตอนของการกำหนดมาตรฐานคะแนนจากตัวคำถามตามวิธีของ

- Nedelsky
- Angoff
- Ebel

### II รายละเอียดของเนื้อหา

#### 1. การตัดสินใจ (Decisions) มาตรฐาน (Standards) และการพิจารณา (Judgement)

การตัดสินใจ (decision) หมายถึงการนำข้อมูลชุดหนึ่งของผู้เข้าสอบมาพิจารณา ร่วมกับความคิดเห็นของครู แล้วจึงตัดสินใจผลการสอบนั้น การใช้วิธีพิจารณาผู้สอบที่ละคนนับว่าเป็นวิธีที่วิธีหนึ่งที่ครูยังไม่จำเป็นจะต้องตัดสินใจโดยกำหนดเกณฑ์ลงไปทันที แต่จะสามารถนำข้อมูลอื่น ๆ ที่เกี่ยวข้องมาร่วมพิจารณาเพื่อตัดสินใจด้วย อย่างไรก็ตามการพิจารณาผู้สอบจากผู้พิจารณา 2 คนอาจจะแตกต่างกันทั้ง ๆ ที่ข้อมูลที่ใช้พิจารณานั้นเป็นข้อมูลชุดเดียวกัน ทั้งนี้ อาจเนื่องมาจากปัญหาในการอธิบายเกณฑ์อาจจะเป็นภาษาที่เข้าใจยาก และอาจมีปัญหามาจากความรู้สึกส่วนตัวของผู้พิจารณาเข้ามาปะปนอยู่ด้วย เช่น ความชอบพอ ความรู้สึกรังเกียจ

ฯลฯ ด้วยเหตุผลนี้เองจึงจำเป็นจะต้องมีการกำหนดเกณฑ์ในการตัดสินใจที่จะสามารถทำให้เกิดความยุติธรรมและประสิทธิภาพของผลจากการตัดสินใจให้มากที่สุด กฎเกณฑ์ที่จะนำมาใช้นั้นอาจจะต้องระบุชัดเจนตามลักษณะข้อมูลที่นำมาใช้ตลอดถึงวิธีการที่จะตัดสินผู้เข้าสอบแต่ละคนด้วย

รูปแบบหนึ่งที่ย่อยและใช้กันทั่ว ๆ ไป คือ การแบ่งผู้สอบออกเป็น 2 กลุ่ม คือกลุ่มที่มีคะแนนสอบสูงและกลุ่มที่มีคะแนนสอบต่ำ รูปแบบนี้มักจะนำมาใช้ตัดสินใจในหลาย ๆ สถานการณ์ เช่น

1. กลุ่มที่ได้คะแนนสูงสามารถขึ้นไปเรียนหน่วยการเรียนรู้ต่อไปได้ ส่วนกลุ่มที่ได้คะแนนต่ำจะต้องเรียนซ้ำชั้นหรือเรียนซ้ำในหน่วยการเรียนรู้เดิม
2. กลุ่มที่ได้คะแนนสูงจะได้รับวุฒิปัตร์ ประกาศนียบัตร์ ปริญญาบัตร ในขณะที่กลุ่มต่ำจะไม่ได้รับ
3. กลุ่มที่ได้คะแนนสูงจะได้รับใบประกอบอาชีพ ส่วนกลุ่มต่ำจะไม่ได้รับ
4. กลุ่มที่ได้คะแนนต่ำจะได้รับการสอนซ่อมเสริม ส่วนกลุ่มสูงจะไม่ได้รับ
5. กลุ่มที่ได้คะแนนสูงจะได้รับคัดเลือกเข้าเรียน ส่วนกลุ่มต่ำจะไม่ได้รับ
6. กลุ่มที่ได้คะแนนสูงจะได้รับหน่วยกิตการเรียนโดยไม่ต้องเรียน ส่วนกลุ่มต่ำจะไม่ได้รับ

ฯลฯ

นอกจากนั้นอาจจะมีการกำหนดกฎเกณฑ์ต่าง ๆ ให้ละเอียดมากขึ้น เช่น คณะกรรมการอาชีพหนึ่ง ๆ อาจกำหนดเกณฑ์ว่านอกจากจะได้คะแนนสูงแล้วยังต้องมีประสบการณ์อย่างน้อย 2 ปี เป็นต้น การนำคะแนนการสอบมาใช้เป็นแนวทางการตัดสินใจนี้ เบื้องต้นจะต้องแยกกลุ่มผู้สอบออกมาเป็นกลุ่มผู้ได้คะแนนสูง และกลุ่มผู้ได้คะแนนต่ำ หรือกลุ่มที่ผ่านและกลุ่มที่ไม่ผ่าน คะแนนตรงจุดนี้เอง เราเรียกว่า จุดผ่านของคะแนน (passing score)

### มาตรฐาน (Standard)

คำว่ามาตรฐาน เป็นคำตอบคำถามที่ว่า “เท่าไรจึงจะพอ” คำ ๆ นั้นนอกจากจะใช้กับคะแนนแล้ว ยังนำมาใช้กับของหลาย ๆ อย่าง เช่น มาตรฐานสินค้า อาหาร การประกันอัคคีภัย เป็นต้น การที่เรากำหนดจุดผ่านของคะแนนก็คือการกำหนดมาตรฐานผลจากการทำแบบทดสอบ

ในการกำหนดจุดผ่านของคะแนนนั้นมักไม่มีปัญหาสำหรับผู้ที่อยู่ในกลุ่มสูงหรือกลุ่มต่ำอย่างแท้จริงถ้าคนในกลุ่มนั้น ๆ สอบได้คะแนนตรงนั้นจริง ๆ แต่จะมีคนอีกกลุ่มหนึ่งมีคะแนน

จริง ๆ อยู่ใกล้ ๆ จุดผ่านนั้น ซึ่งก็ย่อมจะมีทั้งคนโชคร้ายและโชคดี ที่เราไม่สามารถจะหาจุดผ่านของคะแนนได้ชัดเจนนั่นเอง

มาตรฐานนั้นอาจกำหนดให้เด็ดขาดลงไป หรืออาจจะต้องเอาไปสัมพันธ์กับสิ่งอื่น ๆ ด้วย มาตรฐานจึงแบ่งเป็น 2 ประเภทคือ

1. **มาตรฐานสัมบูรณ์ (absolute standard)** เป็นมาตรฐานที่สามารถกำหนดได้เลยโดยไม่ต้องเอาคะแนนไปเปรียบเทียบกับคนอื่น ๆ

2. **มาตรฐานสัมพัทธ์ (relative standard)** เป็นมาตรฐานที่จะต้องเปรียบเทียบคะแนนระหว่างผู้เข้าสอบด้วยกันจึงจะตอบคำถามได้ว่า “ดีพอแล้ว” เช่น

1. ถ้าคะแนนอยู่เหนือคะแนนคนอื่น 5% ถือว่าดีพอแล้ว
2. ถ้าคะแนนอยู่เหนือคะแนนเฉลี่ยของกลุ่มถือว่าดีพอแล้ว
3. ถ้าคะแนนต่ำกว่าคะแนนเฉลี่ยไม่เกิน 25 แต้ม ถือว่าดีพอแล้ว
4. ถ้าคะแนนสูงกว่าคนที่อยู่ล่างสุด 2% ถือว่าดีพอแล้ว

จากตัวอย่างข้างต้นทั้ง 4 ข้อนี้จะเห็นได้ว่าควรที่จะกำหนดว่าคุณภาพของผู้เข้าสอบแต่ละคนที่จะถือว่าดีพอแล้วนั้นจะต้องเอาคะแนนไปเปรียบเทียบกับคะแนนของผู้เข้าสอบคนอื่น ๆ ด้วย ถ้าในกลุ่มเข้าสอบทำคะแนนได้สูง มาตรฐานก็ย่อมจะสูงตามไปด้วย ส่วนมาตรฐานสัมบูรณ์นั้น เป็นมาตรฐานที่ไม่ต้องขึ้นกับคะแนนของผู้เข้าสอบคนอื่น ๆ แต่จะอยู่ที่คะแนนของผู้เข้าสอบแต่ละคนเปรียบเทียบกับมาตรฐานสัมบูรณ์

การที่จะรู้ว่าจุดผ่านของคะแนนตรงไหนเป็นมาตรฐานสัมบูรณ์หรือเป็นมาตรฐานสัมพัทธ์นั้น ก็จำเป็นจะต้องรู้ว่าคะแนนการสอบอยู่ในรูปของ absolute หรือ relative เช่น

กำหนดว่า จุดผ่านของคะแนน = 60 จากคะแนนเต็ม 100 ก็จะต้องรู้ว่า 60 คะแนนหมายความว่าอย่างไร ถ้าหมายความว่าตอบข้อสอบถูก 60% จุดผ่านของคะแนนก็เป็นมาตรฐานสัมบูรณ์ แต่ถ้าหมายความว่า ดีกว่าคนที่ทำข้อสอบเดียวกัน 60% หรืออยู่ต่ำกว่าค่าเฉลี่ย 2 standard deviation จุดผ่านของคะแนนก็เป็นมาตรฐานสัมพัทธ์

การเลือกจุดผ่านของคะแนนที่เป็นมาตรฐานสัมพัทธ์นั้นทำได้ง่าย คือเลือกคะแนนที่ผ่านตามจำนวนที่ต้องการหรือตามร้อยละของผู้เข้าสอบ เช่น จะสอบคัดเลือกคนเพียง 30 คน จุดผ่านของคะแนนก็คือคะแนนที่ คนที่ 30 สอบได้ ส่วนการเลือกจุดผ่านของคะแนนที่เป็นมาตรฐานสัมบูรณ์เป็นเรื่องที่ละเอียดและมีความยุ่งยากในการพิจารณา

### การพิจารณา (Judgement)

มาตรฐานใด ๆ ก็ตามไม่ว่าจะเป็นมาตรฐานสัมบูรณ์หรือมาตรฐานสัมพัทธ์ต่างก็ขึ้น

อยู่กับการพิจารณา (Judgement) ทั้งนี้เพราะการที่จะใช้มาตรฐานตอบคำถามว่า “ดีพอแล้วหรือยัง” นั้น จะต้องผ่านการพิจารณาก่อน การเลือกจุดผ่านของคะแนนจึงรวมเอาการพิจารณาไว้ด้วย ปกติการพิจารณาที่ดีควรมีลักษณะดังนี้

1. ผลการพิจารณาจากผู้พิจารณาที่มีความเหมาะสม
2. มีความสำคัญสำหรับผู้พิจารณา
3. มีวิธีการดำเนินการที่จะตอบสนองเป้าหมายของแบบทดสอบ

ลักษณะสำคัญทั้ง 3 ประการนี้จะต้องเกี่ยวพันกัน การใช้วิธีการเลือกจุดผ่านของคะแนนที่ต่างกันย่อมจะต้องใช้วิธีพิจารณาที่ต่างกัน รวมทั้งจะต้องมาจากผู้พิจารณาที่มีคุณลักษณะต่างกันด้วย ในการอธิบายวิธีพิจารณาแต่ละวิธีนั้นจะต้องอธิบายถึงความจำเป็นที่จะต้องกำหนดคุณสมบัติของผู้พิจารณา ทั้งนี้เพื่อจะได้กระตุ้นให้ผู้พิจารณาเหล่านั้นรู้จุดมุ่งหมายของการทดสอบเพื่อประกอบการพิจารณาด้วย

## 2. ความผิดพลาดที่เกิดจากการตัดสินใจ (Two Types of Wrong Decisions)

ความผิดพลาดในการจำแนกผู้สอบออกเป็น 2 กลุ่ม จะมีความผิดพลาดเกิดขึ้นได้ 2 แบบคือ

1. ผู้เข้าสอบที่ความจริงแล้วอยู่ในกลุ่มต่ำแต่ผิดใจได้คะแนนอยู่เหนือจุดผ่านของคะแนน
2. ผู้เข้าสอบที่ความจริงแล้วอยู่ในกลุ่มสูง แต่ผิดใจได้คะแนนอยู่ต่ำกว่าจุดผ่านของ

คะแนน

ความผิดพลาดเช่นนี้เกิดขึ้นเพราะแบบทดสอบไม่ได้วัดตรงจุดที่ต้องการวัด ทั้งนี้เนื่องจาก

1. เกิดความเปลี่ยนแปลงภายในตัวผู้สอบตลอดเวลา
2. แบบทดสอบไม่ครอบคลุมเนื้อหาและทักษะที่ต้องการวัด
3. ผู้สอบเดาข้อสอบและไม่สามารถตรวจสอบได้ว่าเด็กเดาข้อสอบไปแล้วก็คะแนน
4. แบบทดสอบเป็นเพียงตัวอย่างของความรู้และทักษะทั้งหมดที่จะวัด ดังนั้นอาจเกิด

ความเข้าใจผิดของสภาพที่แท้จริงในด้านทักษะของผู้เข้าสอบบางคนได้

จากที่กล่าวมานี้จะเห็นได้ว่าแบบทดสอบส่วนใหญ่ไม่สามารถหลีกเลี่ยงปัญหาความผิดพลาดจากการเลือกจุดผ่านของคะแนนซึ่งนำไปสู่การตัดสินใจผิดพลาดได้ บางครั้งอาจจะแก้ปัญหานี้โดยการยกระดับจุดผ่านให้สูงขึ้น หรือลดจุดผ่านลงมา การปรับปรุงแบบทดสอบก็เป็นวิธีการหนึ่งที่จะช่วยลดปัญหาความผิดพลาดในเรื่องนี้ลงได้บ้าง แต่ถึงอย่างไรก็ตามก็ยังไม่มีทางที่จะกำจัดปัญหานี้ให้หมดไปได้ ในกรณีที่ถ้าการตัดสินใจพลาดนั้นไม่มีผลอะไรมาก

นัก ก็อาจไม่จำเป็นต้องกลับไปพิจารณาแบบทดสอบ แต่อาจจะใช้วิธีง่าย ๆ คือ ให้ทุกคนสอบได้ หรือให้ทุกคนสอบตก เช่น ถ้าให้ผู้ที่ได้คะแนนต่ำ ๆ สอบได้จะไม่เกิดอันตรายหรือข้อเสียหายใด ๆ แล้วการตัดสินใจให้ทุกคนสอบผ่านหมดก็จะเป็นวิธีที่ดีที่สุด แต่ถ้าเราจะเลือกจุดผ่านของคะแนนแล้วความผิดพลาดทั้ง 2 ประการที่กล่าวมาแล้วก็จะเกิดขึ้นอย่างแน่นอน

### 3. วิธีการกำหนดมาตรฐานโดยพิจารณาจากตัวคำถาม

(Methods Based on Judgements About Test Questions)

วิธีการกำหนดมาตรฐาน 3 วิธีที่จะกล่าวต่อไปนี้จะมาจากแนวคิดในเรื่องที่มีผู้เข้าสอบบางคนที่อยู่ติดกับจุดผ่านของคะแนน ซึ่งบุคคลเหล่านี้จะมีทั้งที่อยู่เหนือและใต้จุดผ่าน วิธีการนี้มีความคิดเบื้องต้นว่าผู้ที่อยู่ในกลุ่มสูงน่าจะมีคะแนนสูงกว่าผู้ที่อยู่ในกลุ่มต่ำ จุดผ่านของคะแนนจึงควรเป็นจุดที่จะสามารถตอบคำถามได้ใกล้เคียง การพิจารณาวิธีนี้จึงต้องย้อนกลับไปดูที่คำถามในแบบทดสอบซึ่งอาจจะทำได้ทั้งก่อนสอบและหลังสอบ การพิจารณาก็จะต้องตรวจสอบเนื้อหาในแบบทดสอบ และการตอบคำถามของผู้สอบที่อยู่ใกล้ ๆ จุดผ่านของคะแนน วิธีการนี้มีขั้นตอนดังนี้ (Livingston & Zicky, 1982:16)

1. เลือกผู้ที่จะต้องเป็นผู้พิจารณา
2. ให้นิยามผู้ที่อยู่ติดกับจุดผ่านของคะแนน (boderline)
3. ทำความเข้าใจให้ตรงกันในการเลือกวิธีพิจารณากับผู้พิจารณา
4. รวบรวมผลการพิจารณา
5. เอาข้อพิจารณาเหล่านั้นมากำหนดจุดตัด

วิธีการทั้ง 3 วิธีที่จะกล่าวต่อไปนั้นจะมีขั้นตอนที่ 1 และ 2 เหมือนกัน ส่วนขั้นตอนอื่น ๆ ก็แตกต่างกันไปตามแต่ละวิธีการขั้นตอนทั้ง 5 ขั้นตอน มีรายละเอียดดังนี้

#### ขั้นตอนที่ 1 การเลือกผู้พิจารณา

ผู้ที่จะพิจารณาแบบทดสอบควรจะเป็นผู้ทรงคุณวุฒิที่สามารถตัดสินใจได้ว่าความรู้หรือทักษะที่จำเป็นควรมีในระดับใด ตัวอย่างเช่น แบบทดสอบวัดความรู้เกี่ยวกับอาชีพเพื่อที่จะออกไปอนุญาตในการปลูกพืช โดยใช้พลังงานนิวเคลียร์ ผู้พิจารณาก็ควรจะทรงคุณวุฒิในด้านความรู้ที่จำเป็นในการป้องกันสาธารณชนให้พ้นภัยของนิวเคลียร์ แต่ถ้าเป็นแบบทดสอบเกี่ยวกับการอ่านในระดับมัธยมศึกษา ผู้พิจารณาก็ควรจะเป็นผู้ที่มีความรู้เรื่องความสามารถในการอ่านของนักเรียนระดับมัธยม ดังนั้นการคัดเลือกหาผู้พิจารณาอาจจะมีปัญหามีน้อยเกินไปหรือบางสาขาอาจมีมากเกินไป ในกรณีที่ผู้พิจารณามีน้อยก็อาจให้มาร่วมกันพิจารณาพร้อมกันทีเดียว

อย่างไรก็ตามควรจะให้แน่ใจว่าในบรรดาผู้พิจารณาเหล่านั้นเป็นผู้ที่มีความสามารถอย่างแท้จริงและสามารถทำงานร่วมกับคนอื่น ๆ ได้ด้วย

การกำหนดจำนวนผู้พิจารณาให้เหมาะสมก็เป็นข้อที่ควรพิจารณาอีกข้อหนึ่ง ถ้ามีน้อยเกินไปในที่ประชุมก็อาจจะถูกผู้พิจารณา 1-2 คน มีอิทธิพลในการกำหนดมาตรฐานได้ ซึ่งก็อาจจะสูงหรือต่ำกว่าที่ควรจะเป็น ในกรณีเรื่องจำนวนผู้พิจารณานี้อาจยึดหลักว่า ถ้ามีมากเท่าไรก็ยิ่งเป็นผลดี โดยปกติควรจะใช้ประมาณ 5 คน อย่างไรก็ตามผลของการพิจารณาของผู้พิจารณานั้นก็ยังไม่ใช่ว่าผลสรุปสุดท้ายในการกำหนดจุดผ่านของคะแนน ดังนั้นถ้าจะสามารถหาผู้พิจารณาได้มากคนก็น่าจะมีผลดีมากกว่า

ในการพิจารณาของผู้พิจารณาแต่ละคนซึ่งต่างก็มีอิสระที่จะพิจารณาเป็นเอกเทศไม่เกี่ยวข้องกัน แต่ก็ควรจะได้จัดให้ผู้พิจารณาได้มีโอกาสร่วมประชุมกันสักครั้งหนึ่ง (ถ้ามีมากกว่า 20 คน อาจจัดประชุมเป็นกลุ่มเล็ก ๆ แยกกันพิจารณา) ในการประชุมก็ควรให้ผู้พิจารณากำหนดนิยามของกลุ่มที่อยู่ติดกับจุดผ่านของคะแนน (borderline group) แล้วให้ผู้พิจารณาได้ทำความเข้าใจเกี่ยวกับวิธีการที่เลือกใช้ด้วย

## ขั้นตอนที่ 2 การให้นิยาม borderline group

ขั้นตอนนี้จะต้องให้ผู้พิจารณาเข้าใจก่อนว่า แบบทดสอบวัดอะไรคะแนนการสอบจะเอามาใช้อะไรได้บ้าง แล้วลองให้ผู้พิจารณาลองอธิบายด้วยภาษาของตัวเองว่า ใครบ้างที่ควรจะอยู่ตรงจุดผ่านของคะแนนทั้งกลุ่มคนที่อยู่สูงหรือต่ำกว่าจุดผ่าน ซึ่งจะช่วยให้ผู้พิจารณาเข้าใจกลุ่ม borderline ดีขึ้น ตัวอย่างเช่น แบบทดสอบวัดการอ่านในโรงเรียนมัธยม ซึ่งมีจุดมุ่งหมายจะให้ผู้ที่สอบผ่านเรียนต่อในด้านการอ่าน แบบทดสอบนี้อาจจะตั้งคำถามว่า ผู้สอบควรจะอ่านหัวข้อเรื่องใดได้บ้างจากหนังสือพิมพ์ข้อความใดเป็นข่าว ข้อความใดเป็นความคิดเห็น คนที่อยู่ตรง borderline จะต้องจับใจความสำคัญของเรื่องได้หรือไม่ จะให้เขียนเป็นภาษาของตนเองหรือไม่ ฯลฯ สิ่งเหล่านี้จะต้องให้ผู้พิจารณาตกลงกันก่อนว่ากลุ่ม borderline ควรจะทำอะไรได้บ้าง ถ้าความเห็นของผู้พิจารณาเหล่านั้นขัดแย้งกันมาก ก็อาจดำเนินการไปก่อนโดยไม่จำเป็นจะต้องนิยามคำนั้นออกมาเพียงข้อเดียว แต่ถ้าจะสามารถตกลงกันได้ก็ควรจะดำเนินการก่อน หลังจากนั้นก็ให้ผู้พิจารณานับที่นิยามนั้นไว้

จากขั้นตอนนี้ วิธีการแต่ละวิธีก็จะแตกต่างกันออกไป ชื่อของวิธีการต่อไปนี้เป็นชื่อของผู้เสนอวิธีการ 3 วิธี ดังนี้คือ

#### 4. Nedelsky's Method

วิธีการนี้เสนอโดย Leo Nedelsky ในปี ค.ศ. 1954 วิธีการนี้ใช้ได้เฉพาะกับแบบทดสอบแบบเลือกตอบ การพิจารณาจะดูเฉพาะตัวลวงของข้อสอบ งานที่ผู้พิจารณาจะต้องทำคือพิจารณาตัวคำถาม แล้วกำหนดคำตอบที่ผู้เข้าสอบในกลุ่ม borderline จะต้องทำผิดว่าเป็นข้อใดบ้างตัวอย่างเช่น ในข้อสอบทักษะภาษาอังกฤษที่ให้ผู้สอบเลือกคำหรือวลีที่จะทำให้ประโยคนั้นสมบูรณ์

My music teacher thinks that Marian Anderson sings \_\_\_\_\_ any other contralto he has ever heard.

- (a) more well than
- (b) better than
- (c) the best of
- (d) more better over

ผู้พิจารณาจะต้องพิจารณาว่าข้อสอบข้อนี้พวก borderline อาจไม่เลือกตัวเลือก (a) และ (d) ส่วนตัวเลือก (c) ซึ่งเป็นตัวลวง และตัวถูก (b) อาจจะยากเกินไปสำหรับเขา หรือผู้พิจารณาอาจจะเห็นว่าตัวลวง (a) และ (d) ค่อนข้างจะผิดเด่นชัดจนพวก borderline อาจไม่เลือกก็ได้

ขั้นตอนที่ 1 และที่ 2 ที่ได้กล่าวไปแล้วเป็นวิธีการที่วิธีการนี้ดำเนินการตาม ส่วนขั้นตอนที่ 3 ที่แตกต่างจากวิธีการของคนอื่น ๆ มีขั้นตอนการดำเนินการดังนี้

##### การรวบรวมผลการพิจารณา

วิธีของ Nedelsky นั้น อาจให้ผู้พิจารณาแยกไปพิจารณาเป็นเอกเทศ หรืออาจพิจารณาร่วมกันก็ได้ถ้ามีผู้พิจารณาไม่มากนัก การจะนัดหมายผู้พิจารณามาประชุมร่วมกันอาจจะมีปัญหา และจะต้องใช้เวลามาก อาจตกลงกันไม่ได้ในบางข้อ แต่การให้ผู้พิจารณาร่วมกันก็จะช่วยให้ผลการพิจารณามีความเที่ยงตรงมากขึ้น การให้ผู้พิจารณาประชุมร่วมกันควรดำเนินการดังนี้

1. ผู้พิจารณาแต่ละคนตรวจสอบล่วงหน้ามาก่อนว่าข้อสอบข้อใดบ้างที่กลุ่ม borderline จะไม่เลือกตัวลวงตัวใดบ้าง

2. ในการจัดอภิปรายในกลุ่มผู้พิจารณาควรดำเนินการดังนี้

2.1 จับประเด็นสำคัญที่ผู้พิจารณาสนใจตัวลวงตัวแรกแล้วจึงถามว่า มีผู้พิจารณา

ผู้ใดคิดว่ากลุ่ม borderline จะไม่เลือก และมีใครบ้างที่คัดค้านคือเห็นว่าพวก borderline อาจจะเลือก

2.2 ถ้าผู้พิจารณามีความเห็นไม่ตรงกัน ก็ให้ถามคนที่พิจารณาล่วงหน้ามาก่อนว่ามีเหตุผลอะไร แล้วจึงถามผู้พิจารณาคนอื่น ๆ ว่าทำไมจึงเห็นไม่ตรงกัน เราไม่จำเป็นจะให้ผู้พิจารณาทุกคนตกลงกันได้ เพียงแต่ให้แต่ละคนได้แสดงเหตุผลซึ่งกันและกัน และเมื่อแสดงความคิดเห็นกันแล้วผู้พิจารณาบางคนก็อาจจะเปลี่ยนใจได้ แต่ควรจะให้แน่ใจว่าผู้พิจารณาแต่ละคนเข้าใจประเด็นที่จะพิจารณาความสามารถของกลุ่ม borderline

2.3 หลังจากนั้นก็เริ่มพิจารณาข้ออื่น ๆ ต่อไป

3. หลังจากอภิปรายข้อสอบทุกข้อจบแล้ว ก็อาจจะถามผู้พิจารณาใหม่ว่าจะมีใครเปลี่ยนใจแก้ไขข้อคิดเห็นของตนเองบ้าง เพื่อเป็นการทบทวนความมั่นใจอีกครั้งหนึ่ง

4. รวบรวมความคิดเห็น ในขั้นตอนนี้อาจใช้เทคนิคต่อไปนี้ดำเนินการเพื่อเป็นการประหยัดเวลา

4.1 ถามผู้พิจารณาว่ามีใครบ้างที่ได้กาเครื่องหมายคำตอบที่กลุ่ม borderline คิดว่าเป็นตัวลวง ครบหมดทุกข้อแล้ว

4.2 ถามผู้พิจารณาว่าใครกาเครื่องหมายข้อ 1 เสร็จแล้วและถามไปเรื่อย ๆ จนครบ

4.3 ถามผู้พิจารณาที่กาเครื่องหมายให้อธิบายเหตุผลในการกาเครื่องหมายนั้น

4.4 ถามผู้พิจารณาที่มีความเห็นต่างออกไป

4.5 ให้มีการอภิปรายในประเด็นที่คิดว่ามีประโยชน์แล้วอนุญาตให้ผู้พิจารณาเปลี่ยนแปลงผลการพิจารณาได้

4.6 เริ่มคำถามใหม่ต่อไป

วิธีการนี้มีข้อจำกัดอยู่ที่ผู้พิจารณาจะต้องร่วมกันพิจารณาในเวลาเดียวกันและสถานที่เดียวกัน นอกจากนั้นข้อจำกัดอีกประการหนึ่งคือเรื่องของเวลา ซึ่งบางครั้งอาจต้องใช้มากในการอภิปรายกัน ดังนั้นวิธีแก้ก็คืออาจจะให้ผู้พิจารณาตรวจสอบข้อคำถามมาล่วงหน้าก่อน ซึ่งถ้าจะทำอย่างนี้ก็ต้องให้ผู้พิจารณาแต่ละคนเข้าใจวิธีการพิจารณาด้วย เช่นอาจให้ลองทำเป็นตัวอย่างสัก 2-3 ข้อ โดยการอภิปรายกับคนอื่น ๆ หลังจากนั้นส่วนที่เหลือก็ให้เอากลับไปทำตามลำพัง รูปแบบของ Nedelsky อาจจะมีปัญหาถ้าข้อสอบแบบเลือกตอบใช้คำถามเป็นเชิงนิเสธ เช่น



อาหารต่อไปนี้ข้อใด **ไม่มี** วิตามินซี?

ก. นม

ค. กะหล่ำปลี

ข. น้ำส้ม

ง. มันฝรั่งเผา

การที่จะให้หาว่าข้อใดผิด ผู้พิจารณาจะต้องจำว่าอาหารข้อใดให้วิตามินซีดีกว่า ข้อใดเป็นคำตอบที่ไม่ถูก ซึ่งส่วนใหญ่แล้วพวก borderline มักจะจำเพียงว่าข้อใดผิด

แบบทดสอบอีกแบบหนึ่งที่มีมักจะเกิดปัญหาตามวิธีการของ Nedelsky ก็คือ แบบเลือกตอบให้เลือกถูกผิด เช่น

ประเทศใดหรือคู่ใดเป็นฝ่ายตรงข้ามกับอเมริกาในสงครามโลกครั้งที่ 2

I เยอรมัน

II รัสเซีย

III อิตาลี

IV ญี่ปุ่น

ก. I

ข. II

ค. I และ IV

ง. I, III และ IV

จ. I, II, III และ IV

ข้อสอบประเภทนี้ที่จริงก็คือแบบถูกผิด 4 ข้อ ผู้พิจารณาจะต้องดำเนินการดังนี้

พิจารณาว่ากลุ่ม borderline จะเลือกข้อใดเป็นตัวถูกและข้อใดเป็นตัวลวง และตัวเลือกใดที่ผู้ตอบไม่แน่ใจ ซึ่งจากการพิจารณาผู้พิจารณาก็จะพบว่าผู้ตอบตอบข้อใดบ้างที่เป็นข้อผิด สมมติว่าผู้พิจารณาเห็นว่า พวก borderline รู้ว่าข้อ I และ IV ถูก ข้อ II ผิด พวก borderline ก็จะกาผิดที่ตัวเลือก ก. เพราะไม่มีญี่ปุ่นอยู่ด้วย และกาผิดที่ตัวเลือก ข เพราะไม่มีเยอรมันหรือญี่ปุ่นและรวมรัสเซียไว้ด้วย และจะกาตัวเลือก จ. เพราะมีรัสเซียอยู่ด้วย

ถ้าจะใช้วิธีการของ Nedelsky กับคำถามแบบนิเสธ หรือแบบถูกผิดตามตัวอย่างข้างบนนี้ ผู้พิจารณาจะต้องแน่ใจว่าพวก borderline จะคิดอย่างไรและควรจะฝึกวิธีการพิจารณา ร่วมกับผู้อื่นก่อนจะแยกไปพิจารณาตามลำพัง เมื่อพิจารณาเสร็จสิ้นแล้วก็ต้องถามเหตุผลของการพิจารณาคำถามบางข้อเพื่อให้แน่ใจว่าผู้พิจารณาเข้าใจวิธีการจริง ๆ

ข้อสอบอีกประเภทหนึ่งที่อาจมีปัญหในการพิจารณาตามวิธีการของ Nedelsky เช่น ข้อสอบที่มีการคำนวณทางคณิตศาสตร์ ทั้งนี้เพราะคำตอบผิดในข้อสอบประเภทนี้มักจะเป็น

คำผิดทั่ว ๆ ไป การพิจารณาจึงต้องเสียเวลามาก ดังนั้นจึงอาจจะต้องลอกแบบของความผิดต่าง ๆ ของคำตอบผิดแต่ละข้อให้ผู้พิจารณาเป็นแนวทางด้วยเช่น

คนงานคนหนึ่งทำคอนตกรจากหลังคาบ้านสูง 36 ฟุต คอนจะตกถึงพื้นดินกินเวลานานเท่าไร ( $g = 32$ )

ก. 1.06 วินาที

ข. 1.125 วินาที

ค. 1.5 วินาที

ง. 2.25 วินาที

ผู้พิจารณาจะต้องรู้ว่าสูตรการคำนวณโจทย์ข้อนี้คือ  $S = 1/2gt^2$  ซึ่งก็จะได้คำตอบข้อ ค. เป็นคำตอบที่ถูก ปัญหาจึงเกิดว่าการคิดตัวลง ก, ข และ ง จะคิดอย่างไร วิธีการที่จะทำให้ง่ายและเร็วขึ้นคือจดที่มาของคำตอบแต่ละคำตอบลงไปดังนี้

ก. 1.06 วินาที ( $S = gt^2$ )

ข. 1.125 วินาที ( $S = gt$ )

ค. 1.5 วินาที ( $S = 1/2 gt^2$ )

ง. 2.25 วินาที ( $S = 1/2 gt$ )

ประเด็นที่ควรพิจารณาวิธีการของ Nedelsky (รวมทั้งของ Angoff และ Ebel) ก็คือ ควรจะบอกคำตอบที่ถูกของข้อสอบแต่ละข้อให้กับผู้พิจารณาหรือไม่ ซึ่งก็จะมีทั้งข้อดีและข้อเสียคือ

**ข้อดี** ถ้าให้คำตอบไปเลยจะทำให้การพิจารณาง่ายขึ้น

**ข้อเสีย** เป็นการชักจูงให้ผู้พิจารณาโน้มเอียงไปตามคำตอบที่ถูกที่ให้มา และมักจะพบว่าจุดผ่านของคะแนนจะสูงกว่าปกติ

ถ้าไม่บอกคำตอบไปก็อาจจะมีผลเสียตรงที่ว่าผู้พิจารณาอาจจะคิดว่าคำตอบที่เป็นตัวลงนั้นเป็นคำตอบถูก แต่ก็อาจจะมีผลดี ถ้าผู้พิจารณาหลาย ๆ คนเห็นว่าคำตอบที่ถูกนั้นไม่ใช่คำตอบถูก แสดงให้เห็นว่าข้อสอบขาดประสิทธิภาพ แต่ถ้าผู้พิจารณาภาตัวลงหลาย ๆ ตัวก็แสดงว่าผู้พิจารณานั้นขาดประสิทธิภาพเช่นกัน นอกจากนั้นการไม่บอกคำตอบถูกไปผู้พิจารณาก็อาจจะคิดว่ากำลังถูกลองภูมิก็ได้ นอกจากนั้นยังทำให้เสียเวลาเพราะจะต้องมาพิจารณาข้อที่ถูกก่อน วิธีที่จะแก้ไขได้ก็คืออาจให้ผู้พิจารณาตรวจสอบแบบทดสอบก่อนจะถึงเวลาพิจารณาร่วมกัน และเมื่อถึงเวลาพิจารณาจึงเฉลยคำตอบนั้น

## การเลือกจุดตัด

วิธีของ Nedelsky เกิดจากแนวคิดที่ว่ากลุ่ม borderline จะเลือกการตอบคำถามโดยค่อย ๆ ขจัดคำตอบผิดออกไป หลังจากนั้นจึงใช้วิธีเดาสุ่มจากตัวเลือกที่เหลือ ถ้าเป็นแบบทดสอบที่ให้คะแนนโดยไม่หักคะแนนการเดา ก็จะสามารถดำเนินการตรวจให้คะแนนง่ายขึ้น ตามวิธีการดังนี้

1. คะแนนที่ได้ในแต่ละข้อ คือ 1 หากด้วยจำนวนข้อที่เดา
2. คะแนนรวมทั้งฉบับคือผลรวมของคะแนนที่คาดหวังของข้อคำถามทุกข้อ

ตัวอย่างเช่น ถ้ากลุ่ม borderline ตัดตัวเลือกข้ออื่น ๆ ออกเว้นไว้ 3 ตัวเลือก โอกาสที่จะตอบถูกในข้อนั้นก็จะมีอยู่ใน 3 ตัวเลือกนั้น คะแนนคาดหวังก็จะเท่ากับ  $1/3$  หรือ .33 แต่ถ้ามีการหักคะแนนการเดาก็จะต้องใช้วิธีการอื่นเพิ่มเติมอีก วิธีการคำนวณดังแสดงในตาราง 3.1

ตาราง 3.1 ตัวอย่างการคำนวณคะแนนโดยวิธีของ Nedelsky ที่ไม่หักคะแนนการเดา

| ข้อสอบข้อที่ | คำตอบ                                                                                                                                                                          | จำนวนข้อที่ไม่ได้ตัดออก | คะแนนคาดหวัง |
|--------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------|--------------|
| 1            | ก <input checked="" type="radio"/> ข <input checked="" type="radio"/> ค <input checked="" type="radio"/> ง <input checked="" type="radio"/> จ                                  | 2                       | $1/2 = .50$  |
| 2            | <input checked="" type="radio"/> ก <input checked="" type="radio"/> ข <input checked="" type="radio"/> ค <input checked="" type="radio"/> ง <input checked="" type="radio"/> จ | 1                       | $1/1 = 1.00$ |
| 3            | <input checked="" type="radio"/> ก <input checked="" type="radio"/> ข ค <input checked="" type="radio"/> ง <input checked="" type="radio"/> จ                                  | 2                       | $1/2 = .50$  |
| 4            | ก <input checked="" type="radio"/> ข ค <input checked="" type="radio"/> ง <input checked="" type="radio"/> จ                                                                   | 3                       | $1/3 = .33$  |
| 5            | <input checked="" type="radio"/> ก <input checked="" type="radio"/> ข <input checked="" type="radio"/> ค <input checked="" type="radio"/> ง <input checked="" type="radio"/> จ | 1                       | $1/1 = 1.00$ |
| 6            | ก ข <input checked="" type="radio"/> ค ง จ                                                                                                                                     | 5                       | $1/5 = .20$  |
| 7            | ก ข ค <input checked="" type="radio"/> ง <input checked="" type="radio"/> จ                                                                                                    | 4                       | $1/4 = .25$  |
| 8            | <input checked="" type="radio"/> ก ข <input checked="" type="radio"/> ค ง จ                                                                                                    | 4                       | $1/4 = .25$  |
| 9            | ก <input checked="" type="radio"/> ข ค ง จ                                                                                                                                     | 5                       | $1/5 = .20$  |
| 10           | ก <input checked="" type="radio"/> ข ค ง จ                                                                                                                                     | 5                       | $1/5 = .20$  |

ผลรวม = 4.43

ผลรวมของคะแนนคาดหวัง = 4.43

0 คือข้อถูก

x คือข้อที่คาดว่าผู้เข้าสอบจะตัดออก

จากตาราง 3.1 เป็นตารางของผู้พิจารณา 1 คน แต่ละคนก็จะมีตารางเช่นนี้คนละ 1 แผ่น การพิจารณาก็จะแตกต่างกันไป หลังจากได้แผ่นพิจารณาเช่นนี้แล้วก็อาจนำมาหาค่าเฉลี่ย (mean)

ซึ่งก็มักจะมีข้อเสียว่า ถ้าผู้พิจารณาคนใดคนหนึ่งให้คะแนนสูงหรือต่ำกว่าของผู้พิจารณาคนอื่น ๆ มาก ๆ แล้วจะมีผลต่อการหาค่าเฉลี่ยมาก ถ้าเป็นกรณีนี้เราอาจใช้วิธีหาค่ามัธยฐาน (median) โดยเรียงคะแนนจากมากที่สุดมาหาคะแนนน้อยที่สุด ถ้าจำนวนผู้พิจารณาเป็นเลขคี่ ค่ามัธยฐานคือค่าของคะแนนตรงจุดกึ่งกลาง แต่ถ้าจำนวนผู้พิจารณาเป็นเลขคู่ ค่ามัธยฐานคือค่าตรงกลางของคะแนน 2 ตัว การหาค่ามัธยฐานมักจะมีข้อเสียตรงที่ว่าถ้ามีผู้พิจารณา มาก ๆ คน การจะหาคะแนนจุดกลางก็จะลำบาก ดังนั้นเพื่อแก้ไขปัญหานี้เราอาจจะใช้ trimmed mean วิธีการคำนวณคือเอาค่าที่สูงสุดหรือต่ำสุดมาก ๆ ออกก่อน แล้วจึงหาค่าคะแนนเฉลี่ยตามวิธีการปกติ การจะตัดคะแนนที่มีค่าต่างจากกลุ่มมาก ๆ นั้น อาจจะใช้ 1, 2 หรือ 3 ของคะแนนตัวบนสุดและล่างสุดก็ได้ ถ้าจะตัดสินใจใช้ trimmed mean ก็ควรจะบอกผู้พิจารณาให้ทราบล่วงหน้าเพื่อที่ว่าผู้พิจารณาจะได้ไม่เกิดความรู้สึกว่าคะแนนที่เขาพิจารณามานั้นถูกตัดออกไป ดังนั้นจากตาราง 3.1 จึงแสดงให้เห็นค่าต่าง ๆ ดังแสดงในตาราง 3.2

ตาราง 3.2 ตัวอย่างการคำนวณคะแนนจากผู้พิจารณา 3 แบบ

|                  |        |         |        |
|------------------|--------|---------|--------|
| คนที่ 1 (สูงสุด) | 92.50  | คนที่ 2 | 77.25  |
| คนที่ 2          | 77.25  | คนที่ 3 | 67.00  |
| คนที่ 3          | 67.00  | คนที่ 4 | 66.67  |
| คนที่ 4          | 66.67  |         |        |
| คนที่ 5 (ต่ำสุด) | 65.33  |         |        |
| รวม              | 368.75 | รวม     | 210.92 |

$$\text{ค่าเฉลี่ย} = 368.75 \div 5 = 73.75$$

$$\text{ค่ามัธยฐาน} = 67.00$$

$$\text{Trimmed mean} = 210.92 \div 3 = 70.31$$

จากตาราง 3.2 นี้แสดงว่าถ้าการพิจารณาของผู้พิจารณาทุกคนถูกต้อง ค่าคาดหวังที่กลุ่ม borderline ควรจะได้ก็คือค่าเฉลี่ย, ค่ามัธยฐาน หรือค่า trimmed mean ถ้าเราเลือกตรงจุดนี้เป็นจุดผ่านของคะแนนกลุ่ม borderline ก็จะมีโอกาสสอบผ่าน 50% (ถ้าวิธีของ Nedelsky สามารถอธิบายแนวคิดของผู้สอบได้จริง) ดังนั้นในกลุ่ม borderline ก็จะมี 50% สอบผ่าน และ 50% สอบตก

## 5. Angoff's Method

William H. Angoff ได้เสนอวิธีการนี้เมื่อปีค.ศ. 1971 ซึ่งก็คล้าย ๆ กับวิธีของ Nedelsky

แตกต่างกันแต่เพียงว่าวิธีของ Angoff ไม่ได้ใช้กับแบบทดสอบแบบเลือกตอบ การคำนวณจุดผ่านของคะแนนก็คล้าย ๆ กับของ Nedelsky ผิดกันแต่ว่าวิธีของ Angoff นี้ผู้พิจารณาจะต้องตรวจสอบข้อคำถามทั้งหมดแล้วพิจารณาว่าความน่าจะเป็นของกลุ่ม borderline ที่จะตอบข้อสอบถูกต้องควรจะเป็นเท่าไร วิธีนี้ถ้าผู้พิจารณาไม่เข้าใจเรื่องของความน่าจะเป็นก็อาจจะให้ผู้พิจารณาคิดเทียบเอาจาก 100 ส่วนว่าพวก borderline จะตอบถูกสักเท่าไร นั่นก็คือถ้าคำถามง่าย โอกาสจะตอบถูกก็จะมาก ถ้าคำถามยากโอกาสจะตอบถูกก็น้อย ความน่าจะเป็นของการตอบถูกจึงมีค่าอยู่ระหว่าง 0.00 ถึง 1.00 ถ้าคำถามนั้นเป็นแบบเลือกตอบ ความน่าจะเป็นของการตอบคำถามถูกก็จะเท่ากับโอกาสในการเดาข้อสอบถูก (คือเท่ากับ 1 หารด้วยจำนวนตัวเลือก)

### การรวบรวมผลการพิจารณา

ในการพิจารณาตัวคำถามผู้พิจารณาอาจทำตามลำพัง หรืออาจจะนำมารวมอภิปรายก็ได้ อย่างไรก็ตามก็ควรดำเนินการดังนี้

1. ผู้พิจารณาควรตรวจสอบคำถามแรก ๆ 2-3 ข้อมาก่อน
2. ให้มีการอภิปรายคำถามแต่ละข้อควรดำเนินการดังนี้
  - 2.1 ให้ผู้พิจารณาแต่ละคนบอกความน่าจะเป็นของการตอบคำถามถูกแต่ละข้อ แล้วเขียนตัวเลขลงบนกระดานดำหรือกระดานแผ่นใหญ่ ๆ ให้ผู้ร่วมอภิปรายทุกคนได้เห็น ถ้าตัวเลขของผู้พิจารณาแต่ละคนใกล้เคียงกันก็ทำข้ออื่น ๆ ต่อไป
  - 2.2 ถ้าตัวเลขของความน่าจะเป็นในการตอบคำถามถูกของผู้พิจารณาแต่ละคนแตกต่างกันมาก ก็ให้ถามเหตุผลของผู้ให้คะแนนสูงสุดหรือต่ำสุดว่ามีเหตุผลอะไร
  - 2.3 หลังจากได้มีการอภิปรายกันแล้ว ถ้าผู้พิจารณาบางคนต้องการเปลี่ยนคะแนนก็ให้เปลี่ยนได้ โดยต้องแน่ใจว่าผู้พิจารณาคนนั้นมีความเข้าใจประเด็นที่จะเปลี่ยนใจดีพอ
3. หลังจากอภิปรายคำถามข้อแรก ๆ แล้ว ก็ให้พิจารณาข้อที่เหลือต่อไป
4. การอภิปรายข้อที่เหลือก็ดำเนินการเช่นเดียวกับ ข้อ 2 แล้วให้โอกาสผู้พิจารณาเปลี่ยนคะแนนได้หลังจากอภิปราย
5. รวบรวมผลการพิจารณา

มีบางคนเสนอแนะว่าวิธีของ Angoff นี้อาจจะกำหนดความน่าจะเป็นมาให้หลาย ๆ ตัวเลือกให้ผู้พิจารณาตัดสินใจ แต่ก็อาจมีข้อเสียที่จะทำให้เกิดความลำเอียงได้ถ้าคำถามบางข้อเป็นคำถามแบบปลายเปิด เช่น กำหนดความน่าจะเป็นของคำถามข้อหนึ่งออกมาเป็นดังนี้

.10 .20 .30 .40 .50 .75

ผู้พิจารณาที่เห็นว่ากลุ่ม borderline ทั้งหมดจะสามารถตอบคำถามได้ถูกหมดก็จะไม่ยอมเลือกตามตัวเลือกที่กำหนดมาให้ นอกจากนั้นการกำหนดความน่าจะเป็นมาให้ผู้พิจารณาเลือกก็ขัดแย้งกับแนวคิดพื้นฐานของ Angoff's Method ทั้งนี้เพราะวิธีการนี้เชื่อว่าผู้พิจารณาแต่ละคนจะสามารถพิจารณาได้อย่างเที่ยงธรรม การกำหนดความน่าจะเป็นมาให้เป็นตัวเลือกก็เท่ากับไม่ไว้วางใจว่าผู้พิจารณาจะให้ความเที่ยงธรรมนั่นเอง ซึ่งถ้าเป็นความเชื่อเช่นนั้นก็ไม่ควรจะใช้วิธีของ Angoff การกำหนดตัวเลือกอาจทำได้ถ้าเราเชื่อว่าผู้พิจารณาจะสามารถตัดสินด้วยความเที่ยงธรรมจากตัวเลือกที่กำหนดให้ แต่ไม่ใช่เพราะไม่มีตัวเลือกเหล่านั้นแล้วจะทำให้ผู้พิจารณาตัดสินพลาดไป

### การเลือกจุดตัด

วิธีการหาคะแนนความคาดหวังของกลุ่ม borderline ก็ใช้วิธีการเดียวกับของ Nedelsky ถ้าการให้คะแนนข้อสอบไม่มีการหักคะแนนการเดา ความน่าจะเป็นของคำตอบที่ถูกต้องก็คือความคาดหวังคะแนนจากคำถามนั้น วิธีการง่าย ๆ ก็คือรวมความน่าจะเป็นของข้อสอบแต่ละข้อเป็นความคาดหวังของคะแนนทั้งฉบับ แล้วคำนวณหาค่าเฉลี่ย ค่ามัธยฐาน หรือค่า Trimmed mean เช่นเดียวกับวิธีของ Nedelsky ตามตารางที่ 3.3

ตาราง 3.3 ตัวอย่างการคำนวณโดยวิธีการของ Angoff เมื่อการให้คะแนนไม่มีการหักคะแนนการเดา

| คำถามข้อที่            | ความน่าจะเป็นของคำตอบถูก |
|------------------------|--------------------------|
| 1                      | .95                      |
| 2                      | .80                      |
| 3                      | .90                      |
| 4                      | .60                      |
| 5                      | .75                      |
| 6                      | .40                      |
| 7                      | .50                      |
| a                      | .25                      |
| 9                      | .25                      |
| 10                     | .40                      |
| คะแนนที่คาดหวัง = 5.80 | รวม 5.80                 |

## 6. Ebel's Method

วิธีการของ Ebel นี้ไม่เหมือนกับ 2 วิธีข้างต้น Ebel มีขั้นตอนอยู่ 2 ขั้นตอนคือ

1. ผู้พิจารณาจำแนกข้อคำถามออกมาเป็นกลุ่ม ๆ แล้วให้คะแนนเป็นตัวเลขในแต่ละกลุ่มคำถาม การจำแนกข้อคำถามนั้นจำแนกข้อคำถามแต่ละข้อเป็น 2 ประเภทคือ

- จำแนกตามความยากง่าย แบ่งเป็น ง่าย ปานกลาง ยาก
- จำแนกตามความสำคัญของคำถาม แบ่งเป็น 4 กลุ่มคือ สำคัญมาก สำคัญบ้าง พอยอมรับ และ ยังสงสัย

2. พิจารณาข้อคำถามลงในตาราง

หลังจากนั้นจึงสรุปผลการพิจารณาลงในตาราง ตามตัวอย่างตาราง 3.4

ตาราง 3.4 ตัวอย่างการจำแนกคำถามและการพิจารณาตามวิธีของ Ebel

| ความสำคัญ | ความยากง่าย                           |                                  |                              |
|-----------|---------------------------------------|----------------------------------|------------------------------|
|           | ง่าย                                  | ปานกลาง                          | ยาก                          |
| สำคัญมาก  | ข้อ 1, 4, 8, 18, 13<br>ถูกประมาณ 95 % | ข้อ 11, 15, 20<br>ถูกประมาณ 85 % | ข้อ 25<br>ถูกประมาณ 80 %     |
| สำคัญ     | ข้อ 2, 6, 9<br>ถูกประมาณ 90 %         | ข้อ 10, 14, 20<br>ถูกประมาณ 75 % | ข้อ 16, 25<br>ถูกประมาณ 60 % |
| พอยอมรับ  | ข้อ 5<br>ถูกประมาณ 80 %               | ข้อ 12, 18<br>ถูกประมาณ 55 %     | ข้อ 19, 23<br>ถูกประมาณ 35 % |
| ยังสงสัย  | ข้อ 3<br>ถูกประมาณ 50 %               | ไม่มี<br>ไม่มี                   | ข้อ 17, 24<br>ถูกประมาณ 20 % |

### การรวบรวมผลการพิจารณา

เนื่องจากวิธีของ Ebel ค่อนข้างละเอียดมากกว่าของ Nedelsky และของ Angoff ทั้งนี้เพราะจะต้องพิจารณาตัดสินใจคำถามข้อหนึ่งเป็น 2 ประเด็น คือความยากง่าย และความสำคัญ และจะต้องเลือกพิจารณาใส่ลงใน 12 กลุ่ม การรวบรวมผลการพิจารณาคือดำเนินการตามขั้นตอนดังนี้

1. การให้ผู้พิจารณาตรวจสอบข้อคำถามใส่ลงใน 12 ช่องนั้นให้ทำตามลำพังก่อน
2. หลังจากนั้นจัดให้มีการอภิปรายคำถามแต่ละข้อ โดยดำเนินการอภิปรายตามขั้นตอนดังนี้

2.1 ให้ผู้พิจารณาบอกจำนวนข้อของคำถามที่จัดอยู่ในกลุ่มง่าย ปานกลาง และยาก ถ้าจำนวนข้อในแต่ละช่องของผู้พิจารณาคนใดแตกต่างจากผู้พิจารณาคนอื่นมากก็ให้อธิบายเหตุผล

2.2 ให้ผู้พิจารณาบอกจำนวนข้อของคำถามที่จัดอยู่ในกลุ่มสำคัญมาก สำคัญ พอ ยอมรับ และยังสงสัย ว่ามีช่องละกี่ข้อ ถ้าความเห็นของใครแตกต่างออกไปมากก็ให้อธิบายเหตุผล

- 2.3 หลังจากอภิปรายแล้ว ให้ผู้พิจารณามีโอกาสเปลี่ยนใจใหม่ได้
3. ให้ผู้พิจารณาตรวจสอบว่าในแต่ละช่องพวก borderline จะมีโอกาสตอบถูกกี่เปอร์เซ็นต์
4. อภิปรายคำถามในช่องต่าง ๆ รวม 12 ช่อง โดยดำเนินการดังนี้
- 4.1 ให้ผู้พิจารณาบอกจำนวนเปอร์เซ็นต์ในแต่ละช่อง
- 4.2 ให้ผู้พิจารณาที่ให้คะแนนสูงสุดหรือต่ำสุดอธิบายเหตุผล
- 4.3 ให้ผู้พิจารณาแต่ละคนมีโอกาสเปลี่ยนใจได้หลังจากอภิปรายแล้ว
5. รวบรวมความคิดเห็นในการพิจารณา

### การเลือกจุดตัด

ดำเนินการดังนี้

1. เอาจำนวนเปอร์เซ็นต์ที่คาดว่าจะตอบถูกในช่องแรกคูณกับจำนวนข้อในช่องนั้น
2. ทำต่อไปจนครบ 12 ช่อง
3. รวมคะแนนที่คาดหวังทั้ง 12 ช่อง ดังแสดงในตาราง 3.5

ตาราง 3.5 ตัวอย่างการคำนวณตามวิธีการของ Ebel เมื่อคะแนนการสอบไม่มีการตัดคะแนนการเดา

| กลุ่ม    | เปอร์เซ็นต์ที่จะทำถูก | จำนวนข้อ | คะแนนที่คาดหวังในแต่ละช่อง |
|----------|-----------------------|----------|----------------------------|
| สำคัญมาก |                       |          |                            |
| ง่าย     | 95                    | 5        | $.95 \times 5 = 4.75$      |
| ปานกลาง  | 85                    | 3        | $.85 \times 3 = 2.55$      |
| ยาก      | 80                    | 1        | $.80 \times 1 = 0.80$      |



| กลุ่ม           | เปอร์เซ็นต์ที่<br>จะทำถูก | จำนวนข้อ | คะแนนที่คาดหวังใน<br>แต่ละข้อ |
|-----------------|---------------------------|----------|-------------------------------|
| <b>สำคัญ</b>    |                           |          |                               |
| ง่าย            | 90                        | 3        | $.90 \times 3 = 2.70$         |
| ปานกลาง         | 75                        | 3        | $.75 \times 3 = 2.25$         |
| ยาก             | 60                        | 2        | $.60 \times 2 = 1.20$         |
| <b>พอยอมรับ</b> |                           |          |                               |
| ง่าย            | 80                        | 1        | $.80 \times 1 = 0.80$         |
| ปานกลาง         | 55                        | 2        | $.55 \times 2 = 1.10$         |
| ยาก             | 35                        | 2        | $.35 \times 2 = 0.70$         |
| <b>ยังสงสัย</b> |                           |          |                               |
| ง่าย            | 50                        | 1        | $.50 \times 1 = 0.50$         |
| ปานกลาง         | —                         | 0        | $0 = -$                       |
| ยาก             | 20                        | 2        | $.20 \times 2 = 0.40$         |

รวม 17.75

คะแนนที่คาดหวัง = 17.75

หลังจากได้คะแนนตามตัวอย่าง 3.5 ของแต่ละคนแล้ว ก็นำมาหาค่าเฉลี่ย ค่ามัธยฐาน หรือค่า trimmed mean

# บทสรุป

1. การตัดสินใจ (Decision) เป็นการนำข้อมูลของผู้เข้าสอบ มาพิจารณาแล้วตัดสินใจผ่านหรือตกตามมาตรฐานของคะแนนที่กำหนด
2. มาตรฐาน (Standard) เป็นการกำหนดจุดผ่านของคะแนนว่าผู้สอบที่สามารถทำคะแนนได้ถึงจุดผ่านจึงจะยอมรับว่ามีความสามารถเพียงพอตามที่ต้องการ
3. มาตรฐานมี 2 แบบคือ มาตรฐานสัมบูรณ์ เป็นมาตรฐานที่กำหนดขึ้นเองโดยไม่ต้องเทียบกับคนอื่น ๆ มาตรฐานสัมพัทธ์เป็นมาตรฐานที่จะต้องเอาคะแนนไปสัมพันธ์กับคะแนนคนอื่น ๆ
4. การพิจารณา (Judgement) เป็นวิธีการหนึ่งในการนำข้อมูลมาตรวจสอบก่อนที่จะกำหนดจุดผ่านของคะแนน ซึ่งจะนำไปสู่การตัดสินใจในขั้นสุดท้าย
5. ความผิดพลาดในการตัดสินใจเกิดขึ้น 2 รูปแบบ คือการให้ผู้สอบสอบผ่านทั้ง ๆ ที่ความจริงแล้วไม่ผ่าน และการที่ให้ผู้สอบสอบไม่ผ่าน ทั้ง ๆ ที่ความจริงแล้วควรจะผ่าน ซึ่งความผิดพลาด 2 ประการนี้จะต้องเกิดขึ้นทันทีที่มีการกำหนดจุดผ่านของคะแนน
6. ขั้นตอนการพิจารณาจุดผ่านของคะแนนตามวิธีของ Nedelsky ข้อสอบจะต้องเป็นแบบเลือกตอบมีการกำหนดเลือกผู้ทรงคุณวุฒิที่เกี่ยวข้องกับข้อสอบเป็นผู้พิจารณาจุดตัด ควรจะเลือกหลาย ๆ คนให้ผู้พิจารณาเป็นผู้ให้นิยามกลุ่ม borderline อาจจะทำตามลำพังหรือพิจารณาร่วมกัน หลังจากนั้นจึงให้ผู้พิจารณาตรวจสอบดูว่า ข้อสอบแต่ละข้อกลุ่ม borderline จะไม่เลือกตัวเลือกใดบ้าง นำคะแนนความคาดหวังใส่ในตารางของแต่ละคน หลังจากนั้นจึงหาค่าเฉลี่ยหรือค่ามัธยฐาน หรือค่า trimmed mean เพื่อกำหนดจุดผ่านของคะแนน
7. วิธีการพิจารณาจุดผ่านของ Angoff ต่างจากของ Nedelsky ตรงที่ข้อคำถามไม่ต้องเป็นข้อสอบแบบเลือกตอบ มีการคำนวณหาค่าคาดหวังของคะแนนโดยใช้หลักของความน่าจะเป็นที่ผู้สอบจะตอบข้อสอบข้อนั้นถูก แล้วเอามารวมกันเป็นตารางความคาดหวังของผู้พิจารณาแต่ละคน แล้วนำมาหาค่าเฉลี่ย หรือค่ามัธยฐานหรือค่า trimmed mean เป็นจุดผ่านของคะแนน
8. วิธีการพิจารณาจุดผ่านของ Ebel ต่างจากของ Nedelsky และ Angoff ตรงที่ผู้พิจารณาจะต้องจำแนกคำถามตามความยากง่ายและตามความสำคัญ แล้วกำหนดจำนวนข้อและโอกาสที่พวก borderline จะตอบถูกลงเป็นเปอร์เซ็นต์ลงในตาราง

หลังจากนั้นก็คำนวณความคาดหวังของแต่ละช่องรวม 12 ช่อง เป็นตารางของผู้พิจารณาแต่ละคน นำคะแนนมาหาค่าเฉลี่ย หรือค่ามัธยฐาน หรือค่า trimmed mean เป็นจุดผ่านของคะแนน

## คำถามท้ายบท

1. จงอธิบายความสัมพันธ์กันของการพิจารณา มาตรฐาน และการตัดสินใจ
2. จงเปรียบเทียบความแตกต่างกันและความคล้ายคลึงกันของมาตรฐานสัมพัทธ์ และมาตรฐานสัมบูรณ์
3. เหตุใดถ้ากำหนดจุดผ่านของคะแนนแล้วจะต้องเกิดความผิดพลาดขึ้น 2 รูปแบบเสมอ
4. จงอธิบายขั้นตอนของการกำหนดจุดผ่านของคะแนนของ Nedelsky's Method
5. ปัญหาที่มักเกิดกับการหาจุดผ่านของคะแนนตามวิธีของ Angoff คืออะไรบ้าง จะแก้ไขได้อย่างไร
6. วิธีการของ Ebel มีขั้นตอนที่แปลกออกไปอย่างไรจงอธิบาย