

การเทียบมาตรา (Test Equating)

เรวัต อินทสระ

การทดสอบเป็นกระบวนการที่สำคัญอย่างหนึ่งในระบบการศึกษา และในการทดสอบแต่ละครั้ง เครื่องมือที่ใช้ในการทดสอบมากที่สุดคือแบบทดสอบ ซึ่งแบบทดสอบที่ดีนั้นจะต้องวัดได้ตรงกับความสามารถที่แท้จริงของแต่ละบุคคล และเกิดความยุติธรรมแก่ทุกคนด้วย จากที่มีการทดสอบบ่อยครั้ง และเป็นสิ่งจำเป็นของแต่ละบุคคล จนเป็นเหตุกระตุ้นให้ผู้สอบเกิดความกระตือรือร้น และหาหนทางเพิ่มพูนคะแนนด้วยวิธีการต่าง ๆ รวมถึงการฝึกซ้อมความชำนาญในการทำข้อสอบจากข้อสอบเก่า ๆ ที่ใช้มาแล้ว จึงทำให้ผู้ที่รับผิดชอบเกี่ยวกับการสอบจะต้องจัดทำแบบทดสอบใหม่ขึ้น และแบบทดสอบที่สร้างขึ้นใหม่นี้จะต้องสามารถวัดคุณลักษณะเดียวกันเพื่อที่จะให้เกิดความเหมาะสมกับสภาพการทดสอบในแต่ละสถานการณ์ จึงทำให้เกิดความคิดในการสร้างแบบทดสอบคู่ขนาน และในการสร้างแบบทดสอบคู่ขนานนั้นมักจะประสบปัญหาที่ไม่สามารถจะสร้างแบบทดสอบให้คู่ขนานกันได้อย่างสมบูรณ์ เพราะในแนวทางการปฏิบัติสามารถที่จะทำได้เพียงแต่ให้เกิดความใกล้เคียงด้านเนื้อหา ค่าสถิติของข้อสอบ และแบบทดสอบทั้งฉบับเท่านั้น ด้วยเหตุนี้ทำให้มีผู้คิดค้นวิธีการที่จะให้คะแนนจากแบบทดสอบที่วัดคุณลักษณะเดียวกันแต่ต่างฉบับกัน สามารถนำผลมาเปรียบเทียบกันได้อย่างมีความหมายขึ้น ทั้งนี้เพื่อประโยชน์ในการดำเนินการสอบให้เกิดความยุติธรรมและความเหมาะสมในสถานการณ์การทดสอบ วิธีการดังกล่าวคือ การเทียบมาตรา (Test Equating)

ความหมายของการเทียบมาตรา

กัลลิกเซน (Gulliksen. 1950 : 298 – 304) ได้ให้ความหมายของการเทียบมาตราว่า เป็นวิธีการทำคะแนนที่ได้จากแบบทดสอบสองชุด ในวิชาเดียวกันให้เป็นคะแนนสมมูล (Equivalent Score) ที่เปรียบเทียบกันได้โดยตรง

ฟลานาแกน (Flanagan. 1915 : 747 – 748) ได้กล่าวถึงการเทียบมาตรา ว่าเป็นวิธีการที่ทำคะแนนจากแบบทดสอบต่างชุดให้มีคุณลักษณะที่เปรียบเทียบกันได้

มาร์โก (Marco. 1981 : 105) ได้อธิบายความหมายของการเทียบมาตรา ว่าเป็นกระบวนการแปลงคะแนน ที่ได้จากแบบทดสอบฟอร์มหนึ่ง ให้มีค่าตัวเลข สมมูล (Equivalent) กับตัวเลขที่ได้จากแบบทดสอบอีกฟอร์มหนึ่ง

แองกอฟ (Angoff. 1971 : 562) ได้ให้คำจำกัดความของการเทียบมาตรา ว่าเป็นการแปลงระบบของหน่วยวัดของแบบทดสอบชุดหนึ่งไปสู่ระบบของการหน่วยวัดของแบบทดสอบอีกชุดหนึ่ง คะแนนที่ผ่านการแปลงแล้ว จะให้ความหมายของความสมมูลกันโดยตรง (directly equivalent)

ลอว์ด (Lord. 1980 : 195) ให้ความหมายของการเทียบมาตรา ว่าเป็นการแปลงคะแนนจากแบบทดสอบต่างฟอร์มให้มีความหมายให้สับเปลี่ยนกันได้ และเพื่อความเสมอภาคของบุคคลผู้รับการทดสอบ

ชูศักดิ์ ชัมภลจิต (ชูศักดิ์ ชัมภลจิต 2527 : 2) ได้สรุปความหมายของการเทียบมาตราว่า การเทียบมาตราเป็นกิจกรรมที่เกี่ยวข้องกับกิจกรรมสองประการ คือ

1. กระบวนการที่ทำให้แบบทดสอบ 2 ฉบับใด ๆ มีความทัดเทียมกันหรือเท่ากันในเชิงโครงสร้าง
2. การใช้วิธีการทางสถิติ เพื่อปรับคะแนนที่ได้จากแบบทดสอบแต่ละฉบับให้อยู่ในมาตราเดียวกัน และเทียบกันได้

จากความหมายดังกล่าว พอสรุปได้ว่า การเทียบมาตรา คือกระบวนการทางสถิติ เพื่อปรับคะแนนที่ได้จากแบบทดสอบต่างชุดที่มีโครงสร้างเดียวกันให้เปรียบเทียบกันได้

การพัฒนารูปแบบของการเทียบมาตรา

การเทียบมาตราได้มีการพัฒนาแนวความคิดมานานแล้ว จากที่ไรท์ และสโตน (Wright & Stone. 1979 : vii-x) ได้เสนอบทความเกี่ยวกับการพัฒนาการของการเทียบมาตรา สรุปได้ดังนี้

ในปี ค.ศ.1925 เทอร์สโตน (Thurstone. 1925) ได้คิดค้นวิธีการแปลงสัดส่วนจำนวนคนที่ตอบข้อสอบได้ถูก ในกลุ่มอายุ ให้เป็นหน่วยความเบี่ยงเบนปกติ (normal deviate) และใช้ค่าเหล่านี้เป็นมาตราพื้นฐาน ส่วนค่ามาตราของกลุ่มอายุอื่น ๆ จะกำหนดขึ้นจากข้อตกลงเบื้องต้น ว่าด้วยความสัมพันธ์เชิงเส้นตรง ระหว่างมาตราต่าง ๆ ของข้อสอบที่ได้จากแบบทดสอบหลาย ๆ ฟอร์ม โดยผู้ตอบมาจากหลาย ๆ กลุ่มอายุ ใช้ค่าเฉลี่ยและค่าความเบี่ยงเบนมาตรฐาน ใช้ในการแปลงคะแนนให้อยู่บนมาตราเดียวกัน (Common Scale) ปรากฏว่ามาตราของเทอร์สโตน ประสบปัญหาของความไม่แน่นอน เพราะสเกลที่ได้จะแปรเปลี่ยนไปตามการแจกแจงความสามารถของกลุ่มตัวอย่าง ต่อมาเขาได้ปรับปรุงแก้ไขวิธีการ โดยอาศัยเทคนิคเปอร์เซนต์ไคล์ และคะแนนมาตรฐาน ซึ่งในปี ค.ศ.1947 โลวิงเกอร์ (Lovinger 1947) ได้พิสูจน์ว่าวิธีของเทอร์สโตนยังไม่มีคุณภาพ โดยแสดงให้เห็นว่า การวัดที่ถูกต้องจะต้องมีเกณฑ์ที่จำเป็นคือ ความเป็นเอกพันธ์ของแบบทดสอบ และสเกลที่เพิ่มขึ้นทางเดียว และวิธีการสร้างมาตราที่ยอมรับได้ จะต้องให้ผลที่เป็นอิสระไม่ผูกพันกับหน่วยเริ่มต้นที่ใช้สร้าง หรือกลุ่มผู้รับการทดสอบกลุ่มแรก ต่อมา กุลลิคเซน (Gulliksen. 1950) ได้ย้าถึงความเข้มงวดของนิยามการเทียบมาตรา คือ เสนอวิธีการให้กลุ่มผู้สอบกลุ่มเดียว ทำแบบทดสอบสองฟอร์ม โดยให้ตรวจสอบความเป็นคู่ขนานของแบบทดสอบเสียก่อน ใช้การทดสอบด้วยสถิติของวิลส์ (Wilks) หรืออาจจะพิจารณาจากค่าสหสัมพันธ์ระหว่างแบบทดสอบกับเกณฑ์ ถ้าแบบทดสอบทั้งสองต่างมีอำนาจในการพยากรณ์แบบทดสอบอีกฉบับหนึ่ง ซึ่งก็สามารถบอกถึงความเพียงพอในการเทียบมาตรา กุลลิคสัน ได้อาศัยตามแนวความคิดของเคลเลย์ (Kelley. 1974) ในการใช้แบบทดสอบร่วม (Anchor test) เพื่อช่วยในการสร้างมาตราการเทียบสำหรับแบบทดสอบใหม่ และได้พัฒนาความคิดจากทักเกอร์ (Ledyard Tuc-

ker) ในการเทียบมาตราจากแบบทดสอบหลาย ๆ ฟอรั่ม จากผู้สอบหลาย ๆ กลุ่ม โดยใช้แบบทดสอบร่วมและสมการถดถอยเชิงเส้นตรง ในการหาค่ากะประมาณที่ได้จากคะแนนแบบทดสอบโดยตรง

ทักเกอร์ (Tucker. 1953) ได้เสนอแนวความคิด เกี่ยวกับแบบทดสอบในอุดมการณ์ของเขาว่าแบบทดสอบควรมีคุณภาพ พอที่จะกำหนดคะแนนแทนตำแหน่งบนความต่อเนื่องลักษณะใดลักษณะหนึ่ง เพื่อให้เกิดความแตกต่างที่มีลักษณะเป็นแบบฉบับนั้น แทนคะแนนสองตำแหน่ง ที่มีความสอดคล้องกับความแตกต่างของผลการสอบทุกกรณี ไม่ว่าจะเป็นคะแนนระดับใด และเขาได้เสนอให้ประเมินความเป็นเอกพันธ์ด้วย

เลวิน (Levine. 1955 citing Holland & Rubin. 1982 : 10) ได้เสนอข้อตกลงเบื้องต้นของการเทียบมาตราไว้ว่า แบบทดสอบสองฉบับใด ๆ จะเทียบคะแนนกันได้ก็ต่อเมื่อแบบทดสอบทั้งสองฉบับนั้นคู่ขนานกัน (parallel) ในด้านต่อไปนี้

1. โครงสร้าง (Structure)
2. เวลา (Timing)
3. รูปแบบ (Format)
4. ชนิดของข้อสอบ (Item type)
5. เนื้อหาวิชา (Subject matter)

แองกอฟ (Angoff. 1960 : 815 citing Wright & Stone. 1979) ได้กล่าวว่า การสร้างมาตราของแบบทดสอบ ให้มีคุณค่าต่อ ๆ ไปกับทุกกลุ่ม ผู้สอบนั้นจะต้องเป็นมาตราที่คงที่ไม่แปรผันไปกับเวลาและลักษณะของกลุ่มผู้สอบ และสามารถเก็บเป็นข้อมูลที่เหมาะสมไว้เพื่อเปรียบเทียบกันได้

ต้นศตวรรษที่ 19 ได้มีการพัฒนา ทฤษฎีคุณลักษณะแฝง (Latent trait theory) โดยมีผู้สร้างโมเดลนี้ครั้งแรก คือ เวเบอร์ และแฟกเนอร์ (Weber & Fechner) สร้างเป็นโมเดลง่าย ๆ ขึ้นอยู่กับโอกาสในการตอบข้อสอบ และต่อมา ลอเลย์ และฟินเนย์ (Lawley & Finney. 1940) ได้ใช้ นอร์มอล ออร์จัน (normal ogive) เป็นโมเดลในการตอบข้อสอบ ต่อมา ลอร์ด (Lord. 1952) เบิร์นบอม (Birnbbaum. 1968) เบเคอร์ (Baker. 1968) ราสซ (Rasch. 1968) แอนเดอร์สัน (Anderson. 1973) และ ไรท์ (Wright. 1977) ได้คิดค้นให้เกิดการนำไปใช้ได้อาศัยแนวความคิด รูปแบบความน่าจะเป็น (Probabilistic model) โดยวิเคราะห์ค่าพารามิเตอร์ของตัวข้อสอบ และประมาณค่าพารามิเตอร์ ความสามารถของคนในแต่ละระดับ และลอร์ด (Lord. 1977) ได้เสนอแนวความคิดในการเทียบมาตราโดยใช้ทฤษฎีการตอบข้อสอบ (Item Response Theory) ซึ่งเขาใช้ค่าพารามิเตอร์ของข้อสอบ และค่าความสามารถของคนตามความจริงจากแบบทดสอบ แล้วเทียบคะแนนจริงของแบบทดสอบทั้งสองฟอรั่มในระดับความสามารถที่เท่ากัน

ในระหว่างปี ค.ศ.1978—1980 สถาบันบริการทดสอบทางการศึกษาได้ตั้งทีมงานขึ้นมาชื่อว่า Program Statistics Research Project (PSRP) เพื่อทำการวิจัยและศึกษาเกี่ยวกับการเทียบมาตราในแง่

เทคนิค วิธีการที่เกี่ยวกับการใช้สถิติเพื่อการเทียบมาตรา และค้นคว้าทฤษฎีค่าความเที่ยงตรงของแบบทดสอบ (Test Validity) และ PSRT ได้จัดประชุมสัมมนาเกี่ยวกับการเทียบมาตราเมื่อเดือนเมษายน 1980 ได้จัดพิมพ์หนังสือชื่อ Test Equating โดยรวบรวมข้อเสนอ ข้อวิพากษ์วิจารณ์เกี่ยวกับแนวคิด และวิธีการเทียบมาตรา ซึ่งรวบรวมโดย ฮอลแลนด์ และรูบิน (Holland & Rubin. 1982)

การเทียบมาตราในรูปแบบต่างๆ

รูปแบบวิธีการเทียบมาตรา ในปัจจุบันได้มีผู้เสนอไว้หลายรูปแบบ พอจะสรุปได้ดังนี้

1. วิธีเทียบมาตรารูปแบบเชิงเส้นตรง

รูปแบบการเทียบมาตราเชิงเส้นตรง เป็นรูปแบบที่ได้รับความนิยมมาเป็นเวลานาน เช่น การใช้แบบทดสอบความถนัดเชิงวิชาการ (ภาวิณี ศรีสุขวัฒนานนท์ 2529 : 30 อ้างจาก Donlon and Angoff. 1971 : 32) โดยใช้สมการเส้นตรงเป็นรูปแบบอยู่ในนิยามที่ว่า คะแนนจากแบบทดสอบ 2 ชุด โดยที่แบบทดสอบทั้ง 2 ชุด ต่างเป็นแบบทดสอบที่วัดในสิ่งเดียวกัน (Parallel measures) จะพิจารณาตัดสินให้สมมูลกันได้ก็ต่อเมื่อคะแนนมาตรฐานทั้งสองชุดมีค่าเท่ากัน คือ

$$(Y - M_y) / S_y = (X - M_x) / S_x$$

เมื่อ X คือ คะแนนจากแบบทดสอบชุดที่ 1

Y คือ คะแนนจากแบบทดสอบชุดที่ 2

M_x คือ ค่าเฉลี่ยของแบบทดสอบชุดที่ 1

M_y คือ ค่าเฉลี่ยของแบบทดสอบชุดที่ 2

S_x คือ ค่าความเบี่ยงเบนของแบบทดสอบชุดที่ 1

S_y คือ ค่าความเบี่ยงเบนของแบบทดสอบชุดที่ 2

ในกรณีรูปแบบการแจกแจงของคะแนนแบบทดสอบชุดที่ 1 เหมือนกับแบบทดสอบชุดที่ 2 ใช้การเทียบรูปแบบเชิงเส้นตรงได้สะดวกง่าย และมีความเป็นปรนัย เพราะจะเป็นไปตามข้อตกลงเบื้องต้น

การเทียบมาตรารูปแบบเชิงเส้นตรงโดยใช้แบบทดสอบร่วม

ผู้คิดค้นวิธีการเทียบมาตรา รูปแบบเชิงเส้นตรงโดยใช้แบบทดสอบร่วม คือ ดร.เลดยาร์ด ทักเคอร์ (Dr. Ledyard Tucker) วิธีนี้เริ่มด้วย การมีข้อตกลงเบื้องต้นของการเทียบมาตรามีฟังก์ชันเป็นเส้นตรงดังนี้

$$e_x(Y) = M_x + \frac{\sigma_x}{\sigma_y} (Y - M_y) \quad (1)$$

เมื่อ $M_x, \sigma_x, M_y, \sigma_y$ เป็นค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐาน ของคะแนนจากแบบทดสอบชุดที่ 1 และชุดที่ 2 ตามลำดับ และในการหาค่า $M_x, M_y, \sigma_x, \sigma_y$ จะหาได้ก็ต่อเมื่อ

ได้กำหนดเงื่อนไขเป็นข้อตกลงเบื้องต้น 2 อย่าง คือ

1. ค่าเฉลี่ย และค่าความแปรปรวนอย่างมีเงื่อนไขของ x เมื่อกำหนดแบบทดสอบรวม (v) เดียวกับ y

2. เป็นข้อตกลงตามรูปแบบ (model assumption) เพราะเป็นการยอมรับ การทำค่าคาดหวังของค่าเฉลี่ยและค่าความแปรปรวน

ผลจากข้อตกลงเบื้องต้น สรุปหลักของการเทียบมาตราได้ดังนี้

ถ้าค่าคาดหวังเฉลี่ย และค่าความแปรปรวนในแบบทดสอบรวมเท่ากันแล้ว จะได้ค่า $M_x, M_y, \sigma_x, \sigma_y$ เพื่อจะนำไปแทนค่าใน (1) จะได้ฟังก์ชันของการเทียบมาตราเชิงเส้นตรงของ X ไปสู่ y (Braun and Holland. 1982 : 23-24)

ขั้นตอนของการเทียบมาตราตามรูปแบบเชิงเส้นตรง

รวบรวมข้อมูลจากกลุ่มตัวอย่าง ที่ให้กลุ่มตัวอย่างที่ 1 ทำแบบทดสอบชุดที่ 1 และกลุ่มตัวอย่างที่ 2 ทำแบบทดสอบชุดที่ 2 นำมาประกอบค่าประชากรต่าง ๆ เพื่อนำมาแทนค่าในสมการ

$$X^* = e_x(Y) = A_y + B$$

กระทำด้วย วิธีการประมาณด้วยความเป็นไปได้สูงสุด (Maximum likelihood method) (Angoff. 1984 : 105) ตามขั้นตอนดังนี้

1. หาค่าประมาณต่าง ๆ ตามสูตรดังนี้

$$\hat{M}_x = M_{x1} + b_{xv1}(\hat{M}_v - M_{v1})$$

$$\hat{\sigma}_y^2 = s_{x1}^2 + b_{xv1}^2(\hat{\sigma}_v^2 - s_{v1}^2)$$

$$\hat{M}_y = M_{y2} + b_{yv2}(\hat{M}_v - M_{v2})$$

$$\hat{\sigma}_y^2 = s_{y2}^2 + b_{yv2}^2(\hat{\sigma}_v^2 - s_{v2}^2)$$

เมื่อ $\hat{M}_x \hat{\sigma}_x^2$ คือ ค่าประมาณคะแนนเฉลี่ย และความแปรปรวนของกลุ่มตัวอย่างทั้งหมด ที่ทำแบบทดสอบชุดที่ 1

$\hat{M}_y \hat{\sigma}_y^2$ คือ ค่าประมาณคะแนนเฉลี่ย และความแปรปรวนของกลุ่มตัวอย่างทั้งหมดที่ทำแบบทดสอบชุดที่ 2

$M_{x1} s_{x1}^2$ คือ ค่าเฉลี่ย และความแปรปรวนของคะแนนชุดที่ 1 ในกลุ่มที่ 1

$M_{y2} s_{y2}^2$ คือ ค่าเฉลี่ย และความแปรปรวนของคะแนนชุดที่ 2 ในกลุ่มที่ 2

$M_{v1} s_{v1}^2$ คือ ค่าเฉลี่ยและความแปรปรวนของคะแนนแบบทดสอบรวมในกลุ่มที่ 1

$M_{v2} s_{v2}^2$ คือ ค่าเฉลี่ยและความแปรปรวนของคะแนนแบบทดสอบรวมในกลุ่มที่ 2

b_{xv1} คือ สัมประสิทธิ์ถดถอยของคะแนน จากแบบทดสอบชุดที่ 1 จากแบบทดสอบรวมในกลุ่มที่ 1

b_{yv2} คือ สัมประสิทธิ์ถดถอยของคะแนน จากแบบทดสอบชุดที่ 2 จากแบบทดสอบรวมในกลุ่มที่ 2

2. หาค่าประมาณความชัน (A) และค่าคงที่ (B) จากสูตร

$$A = \frac{b_{yv2} [s^2_{x1} + b^2_{xv1} (\hat{\sigma}^2_v - s^2_{v1})]}{b_{xv1} [s^2_{y2} + b^2_{yv2} (\hat{\sigma}^2_v - s^2_{v2})]}$$

$$B = \hat{M}_x - A \hat{M}_y$$

3. แทนค่าประมาณประชากรลงในสมการเส้นตรง

$$X^* = A_y + B$$

4. ใช้สมการสร้างตารางเทียบคะแนนจากแบบทดสอบชุดที่ 1 ไปสู่คะแนนของแบบทดสอบชุดที่ 2

2. วิธีเทียบมาตรฐานรูปแบบอิกวิเปอร์เซนไคล์

การเทียบมาตรฐานตามรูปแบบอิกวิเปอร์เซนไคล์ จะเริ่มจากการแจกแจงของคะแนนจากแบบทดสอบชุดที่ 1 และชุดที่ 2 มีลักษณะคล้ายกัน หรือจะแตกต่างกันบ้างก็เพียงเล็กน้อย การเทียบหาคะแนนสมมูลทำได้โดยใช้คะแนน ณ ตำแหน่งเปอร์เซนไคล์เดียวกันของคะแนน 2 ชุดนั้น ผลของการเทียบมาตรฐานแสดงด้วยกราฟ ขั้นตอนการแปลงคะแนนคือ เลือกกลุ่มตัวอย่างที่มีความสามารถกระจายมีทั้งเก่ง ปานกลางและอ่อน แล้วแบ่งเป็นกลุ่มย่อย โดยสุ่ม 2 กลุ่ม ให้กลุ่มหนึ่งทำแบบทดสอบชุดที่ 1 และกลุ่มสองทำแบบทดสอบชุดที่ 2 ผลการทดสอบนำมาแจกแจงเป็นคะแนนชุดที่ 1 และชุดที่ 2 คำนวณหาจุดกลางเปอร์เซนไคล์ของการแจกแจง อ่านและทำเครื่องหมายสำหรับค่าคะแนนชุดที่ 1 และชุดที่ 2 ของคะแนนการแจกแจงที่สมนัยกันบนกระดาษกราฟ ทำจุดประมาณ 30 จุด แล้วลากเส้นเชื่อมเกิดเป็นเส้นกราฟ ทำการปรับให้เรียบเส้นกราฟนี้จะใช้อ่านค่าแบบทดสอบชุดที่ 1 ที่สมนัยกับชุดที่ 2 จากนั้นทำตารางคะแนนแปลงสำเร็จ โดยปกติการเทียบมาตรฐานรูปแบบอิกวิเปอร์เซนไคล์ให้ภาพการแปลงคะแนนที่สะท้อนถึงระดับความยากง่ายของแบบทดสอบสองชุด ถ้าแบบทดสอบสองชุดมีค่าความยากใกล้เคียงกัน เส้นกราฟจะมีลักษณะใกล้เคียงเส้นตรง แต่ถ้ามีค่าความยากต่างกัน เส้นกราฟจะเป็นเส้นโค้ง คะแนนสมมูลที่เกิดขึ้น จะถูกยืดหยุ่นคะแนนดิบเพื่อให้คงรักษาไว้ให้เหมือนชุดก่อนตามต้องการ (Angoff. 1984 : 101)

การเทียบมาตรฐานตามรูปแบบอิกวิเปอร์เซนไคล์โดยใช้แบบทดสอบรวม

บรุนและฮอลแลนด์ (Braun and Holland. 1982 : 19 – 22) ได้แสดงถึงความสัมพันธ์ของแบบทดสอบรวมในการเทียบมาตรฐานตามรูปแบบอิกวิเปอร์เซนไคล์ คือ กลุ่มตัวอย่างสองกลุ่มที่สุ่มมา

อย่างอิสระจากประชากรทั้งหมด (T) โดยกลุ่มตัวอย่างที่ 1 คือ Q เขียนเป็นสมการได้

$$T = fp + (1 - f)Q$$

โดยที่ P และ Q ขึ้นอยู่กับน้ำหนักของ f และ (1-f) ตามลำดับ การคำนวณการแจกแจงของตัวอย่างรวม ทำได้โดยคำนวณการแจกแจงตามเงื่อนไขของตัวอย่างกลุ่มที่ 1 และที่ 2 ก่อนแล้วหาค่าเฉลี่ยที่ถ่วงน้ำหนักด้วย f และ 1-f ซึ่งได้จากสมการคือ

$$f = N_1 / (N_1 + N_2)$$

เมื่อ N_1 และ N_2 คือ ขนาดของกลุ่มตัวอย่างที่ 1 และที่ 2 ตามลำดับค่าของ f คือ ค่าเฉลี่ยของกลุ่มตัวอย่างที่ 1 และที่ 2

ในรูปแบบการเทียบมาตราแบบอิกวิเปอร์เซนไคล์มีข้อตกลง เพื่อหาค่าประมาณการแจกแจงของแบบทดสอบชุดที่ 1 และแบบทดสอบชุดที่ 2 ในประชากรทั้งหมด ซึ่งได้ฟังก์ชันของการเทียบมาตราเป็น

$$\hat{e}_x(Y) = \hat{F}_T^{-1}(\hat{G}_T(Y))$$

ขั้นตอนของการเทียบมาตราตามรูปแบบอิกวิเปอร์เซนไคล์

รวบรวมข้อมูลที่ได้จากการทดสอบด้วยแบบทดสอบ 2 ชุด แล้วนำมาเทียบมาตราตามรูปแบบดังต่อไปนี้

1. ทำตารางประมาณการแจกแจงความถี่ของกลุ่มตัวอย่างทั้งสองกลุ่ม
2. นำค่าความถี่ประมาณของกลุ่มตัวอย่างทั้งสองกลุ่ม หาค่าแห่งเปอร์เซนไคล์ของแต่ละชั้นคะแนน
3. เขียนกราฟแสดงความสัมพันธ์ของคะแนนชุดที่ 1 และชุดที่ 2 เพื่อแสดงคะแนนที่สมมูลโดยการลากเส้นกราฟให้ผ่านจุดต่าง ๆ ให้มากที่สุดแล้วเกล้านโค้งให้เรียบ
4. ทำตารางคะแนนสมมูลของคะแนนชุดที่ 1 และชุดที่ 2 โดยอ่านจากกราฟ

3. วิธีเทียบมาตรารูปแบบอิงทฤษฎีการตอบข้อสอบ

การเทียบมาตรารูปแบบอิงทฤษฎีการตอบข้อสอบ ลอร์ด (Lord, 1977 : 195) ได้กล่าวว่า การเทียบมาตราของแบบทดสอบ 2 ชุดใด ๆ ที่ใช้ได้กับทุกคน จะต้องเป็นไปตามความแตกต่างของบุคคลที่ระดับความสามารถ (θ) เดียวกันและได้เสนอข้อกำหนดที่สำคัญของการเทียบมาตราไว้ 3 อย่างสำหรับแบบทดสอบที่มีมิติเดียว และวัดความสามารถ (θ) เดียวกัน คือ

1) ความเสมอภาค (equity) คือถ้าพิจารณาที่ระดับความสามารถ (θ) ใด ๆ การแจกแจงความถี่อย่างมีเงื่อนไขของคะแนนแปลงด้วยเทคนิคการเทียบคะแนน ต้องเหมือนกับการแจกแจงความถี่อย่างมีเงื่อนไขของคะแนนจากแบบทดสอบที่ต้องการเทียบ

2) ความไม่แปรเปลี่ยนเมื่อเปลี่ยนกลุ่ม คือ คะแนนแปลงจะเหมือนกันโดยไม่ขึ้นกับตัวแปร

อื่น ๆ ของประชากรที่นำมาสร้างสมการเทียบคะแนน

3) ความสมมาตร (Symmetry) คือคะแนนเทียบมาตราจะต้องเหมือนกันไม่ว่าการเทียบนั้นจะเทียบจากแบบทดสอบชุดที่ 1 ไปชุดที่ 2 หรือจากแบบทดสอบชุดที่ 2 ไปชุดที่ 1

ซึ่งข้อกำหนดนี้ จะมีการเทียบคะแนนที่สอดคล้องกับข้อมูลที่เหมาะสมกับทฤษฎีการตอบข้อสอบด้วย และคะแนนจากการสอบนั้นไม่มีค่าความคลาดเคลื่อนรวมอยู่ นั่นคือมีเฉพาะคะแนนจริง

ในการคำนวณจะต้องหาค่าประมาณของประชากรในแต่ละชุดของแบบทดสอบแล้ว นำมาเทียบเป็นคะแนนที่สมมูลในค่าความสามารถ (θ) เดียวกัน และในทางปฏิบัติเราไม่สามารถหาคะแนนจริงได้ ซึ่ง ลอร์ด (Lord, 1982) ได้เสนอวิธีการเทียบโดยใช้คะแนนจากการสังเกต (Observed Score) ได้

การเทียบมาตรารูปแบบอิงทฤษฎีการตอบข้อสอบโดยใช้แบบทดสอบร่วม

การออกแบบเพื่อศึกษาการเทียบมาตราโดยใช้ทฤษฎีการตอบข้อสอบ ได้ใช้แบบทดสอบร่วมช่วยพิจารณา และปรับความแตกต่างของความสามารถ 2 กลุ่ม โดยแต่ละกลุ่มจะสอบเพียงฉบับเดียว และแบบทดสอบร่วมเท่านั้น วิธีการของการเทียบมาตรารูปแบบนี้มีการเทียบคะแนน 2 ส่วน กล่าวคือ

1) การเทียบมาตราคะแนนจริง

โดยใช้ความสัมพันธ์เชิงคณิตศาสตร์ระหว่างความสามารถและจำนวนคะแนนตามข้อที่ทำถูก ซึ่งจะได้คะแนนจริงจากแบบทดสอบ 2 ฉบับ ดังนี้

$$\xi = \xi(\theta) = \sum_{i=1}^{n_x} P_i(\theta) \quad (2)$$

$$\eta - \eta(\theta) = \sum_{j=1}^{n_y} P_j(\theta) \quad (3)$$

เมื่อ n_x , n_y คือจำนวนข้อสอบในแบบทดสอบชุด x และ y ตามลำดับ

ถ้าแทนค่าความสามารถ (θ) ลงในสมการทั้งสองก็จะได้คะแนนจริงที่สมมูลกันบนมาตราเดียวกัน และสามารถแทนค่าความสามารถ (θ) ลงในสมการที่ θ เท่ากัน จะหาคะแนนสมมูลของแบบทดสอบฉบับ x และฉบับ y ได้โดยใช้คำนวณจากค่า $P_i(\theta)$ และ $P_j(\theta)$ โดยประมาณค่า $P_i(\theta)$ และ $P_j(\theta)$ ที่ได้จากค่าพารามิเตอร์ (Lord, 1977 : 180) โดยการใช้โปรแกรมโลจิส (Logist) วิเคราะห์ข้อมูลการตอบโดยมีสมมติฐานเบื้องต้นว่า ความสามารถ (θ) ใด ๆ ของผู้สอบไม่มีความแตกต่างจากแบบทดสอบหนึ่งไปยังอีกแบบทดสอบหนึ่ง และถือว่าส่วนที่ไม่ได้สอบคือ ส่วนที่ไม่ทำไม่ถึง (Lord, 1980 : 20)

2) การเทียบมาตราคะแนนสังเกต

จากที่การกะประมาณคะแนนจริงในทางปฏิบัติมีปัญหาซึ่ง ลอร์ด (Lord, 1980 : 197) ได้กล่าว
ว่า การกะประมาณคะแนนของผู้สอบจากคำตอบ $\hat{\xi} = \sum_i \hat{p}_i(\hat{\theta})$ ที่แท้จริง คือคะแนนที่มีค่าความ
คลาดเคลื่อนรวมอยู่ด้วย ไม่มีคุณสมบัติเป็นคะแนนจริง ลอร์ดจึงเสนอการเทียบมาตราด้วยคะแนนสังเกต
เป็นวิธีการเทียบมาตรา โดยค่าประมาณที่ทำได้ในทางปฏิบัติที่สถานการณ์การสอบไม่สามารถจัดให้
ผู้สอบทำแบบทดสอบได้ทั้ง 2 ชุด โดยการกะประมาณค่าความสามารถ (θ) ของผู้สอบ ใช้การแจกแจง
ของ $\hat{\theta}$ ในกลุ่มที่กะประมาณ $r(\theta)$ คือกลุ่มค่าความยาก ค่าอำนาจจำแนก และค่าสัมประสิทธิ์การเดา
การแจกแจงของคะแนนดิบที่ได้จากสมการดังนี้

$$\phi(x, y) = \frac{1}{N} \sum_{a=1}^N \phi_x(x/\hat{\theta}_a) \phi_y(y/\hat{\theta}_a)$$

หรือโดย

$$\phi(x, y) = \int_{-\infty}^{\infty} \phi_x(x/\theta) \phi_y(y/\theta) r(\theta) d(\theta)$$

จากสมการเป็นการแจกแจงร่วมกันของตัวแปร 3 ตัว คือ θ , x และ y จากที่ θ เป็นตัวระบุ
คะแนนจริง ξ และ η การแจกแจงนี้ จึงเป็นการแจกแจงร่วมกันของตัวแปร 4 ตัว คือ ξ , η , x
และ y ซึ่งกะประมาณได้โดยการระบุนความสัมพันธ์เชิงอิควิเปอร์เซนไคล์ระหว่าง x และ y

จากทฤษฎีการเทียบคะแนนดิบของแบบทดสอบฉบับเดียว ด้วยวิธีการเทียบคะแนนจริง วิธี
การเทียบคะแนนสังเกตดังกล่าวมานั้น ซึ่งจะมีทั้งข้อดีและข้อเสีย วิธีการเทียบคะแนนจริง ξ และ η
ให้ความหมายของคะแนนสมมูลเฉพาะคะแนนที่อยู่เหนือค่าเฉลี่ยของการเดา ส่วนวิธีที่ใช้การกะประมาณ
จากคะแนนสังเกตก็มีจุดอ่อนที่เป็นเพียงการเทียบมาตราอย่างประมาณเท่านั้น ซึ่งลอร์ดได้กล่าวว่า
ทั้งสองวิธีมีความสอดคล้องกันมาก แต่การสรุปเปรียบเทียบอ้างอิงทั่วไปจะต้องทำอย่างพิถีพิถัน
(Lord, 1980 : 203) ซึ่งในปี ค.ศ. 1984 ลอร์ดและวินเกอส์กี (ภาวิณี ศรีสุขวัฒนานนท์ 2529 : 45
อ้างจาก Lord and Wingersky, 1984) ได้ศึกษาเปรียบเทียบผลการเทียบมาตราตามรูปแบบอิงทฤษฎี
การตอบข้อสอบด้วยวิธีคะแนนจริง และคะแนนสังเกต ผลปรากฏว่าทั้งสองวิธีไม่มีข้อแตกต่างกันเลย
และได้วิจารณ์ไว้ว่า วิธีคะแนนสังเกตมีความซับซ้อนในการคำนวณมากกว่าคะแนนจริง

ขั้นตอนของการเทียบมาตราตามรูปแบบอิงทฤษฎีการตอบข้อสอบ

- 1) ขั้นการประมาณคะแนนจริง โดยกะประมาณค่าพารามิเตอร์ของข้อสอบ และค่าความ
สามารถของข้อสอบรายข้อ ใช้กะประมาณค่าความเป็นไปได้สูงสุด และหาค่าประมาณคะแนนจริง
ของผู้สอบ
- 2) ขั้นการเทียบมาตรานำค่าคะแนนจริงของผู้สอบมาจับคู่ ที่ค่าความสามารถเดียวกัน และ
ปรับปรุงขยายคะแนนจริงที่ต่ำกว่าคะแนนการเดาให้มีความหมายขึ้นแล้วจึงนำคะแนนจริงของแบบทดสอบ

แต่ละชุดที่คะแนนความสามารถเดียวกันในแต่ละคู่สร้างเป็นตารางเทียบคะแนน

4. วิธีเทียบมาตรารูปแบบการใช้เทคนิคการวิเคราะห์องค์ประกอบ

วิธีการเทียบมาตรานี้ ใช้ความเป็นไปได้ของการวิเคราะห์ตัวประกอบสูงสุด (Maximum-Likelihood Factor Analysis) ในการประมาณค่าพารามิเตอร์ ในการเทียบมาตรากำหนดข้อมูลเหมาะสม (fit) กับแฟกเตอร์โมเดล (Factor Model) แล้วค่าประมาณที่เป็นไปได้สูงสุด (MLE) จะใช้ได้กับอินฟอร์เมชัน (Information) ในข้อมูลทั้งหมด และได้รับการทดสอบค่าทางสถิติกับกลุ่มตัวอย่างขนาดใหญ่ ที่มีข้อตกลงในแต่ละกลุ่มที่เหมาะสมในการทดสอบซึ่งมีวิธีการประมาณความเป็นไปได้สูงสุด (MLE) ตกกลงให้เป็นไปตามโครงสร้างดังนี้

- 1) การเทียบมาตรารูปแบบทดสอบหลาย ๆ ฉบับ
- 2) การเทียบมาตรามีพื้นฐานของส่วนประกอบของแบบทดสอบย่อย โดยใช้แบบทดสอบเปรียบเทียบทั่วไป (The Common equating test)
- 3) ผลของความถูกต้องของการประมาณค่าพารามิเตอร์ที่เปรียบเทียบได้มาจากตัวแปรอินฟอร์เมชัน (Information) ทั้งหมด
- 4) ใช้กลุ่มตัวอย่างขนาดใหญ่ จะให้ค่าพารามิเตอร์ในการเปรียบเทียบที่แน่นอน

ความสัมพันธ์ระหว่างรูปแบบองค์ประกอบกับการเทียบคะแนนของแบบทดสอบภายใต้ตัวประกอบร่วม (Common factor) เพียงอย่างเดียวก็เพียงพอที่จะอธิบายถึงความแปรปรวนของแบบทดสอบแบบทดสอบที่มีองค์ประกอบคล้ายกัน ซึ่งไม่จำเป็นจะต้องเป็นองค์ประกอบเดียว และข้อมูลจะต้องสอดคล้องกับโมเดล จึงจะใช้ได้ในการเทียบมาตรานี้ได้ (Rock, 1982 : 247)

โจเรสกอก (Joreskog, 1971 : 111) ได้สรุปไว้ว่า ถ้าคะแนนจากแบบทดสอบชนิดเดียวกัน และคะแนนจริง t_i และ t_j มีความสัมพันธ์เป็นอันเดียวกัน (Correlate Unity) ซึ่งได้จากการสุ่ม f และ t_j จะได้ความสัมพันธ์เชิงเส้นตรงดังนี้

$$t_i = v_i + \sum_i f \quad (4)$$

ถ้าแทนค่าในสมการตามทฤษฎีตั้งเดิม (Lord and Norich, 1968) จะได้ว่า

$$x_i = v_i + \sum_i f + e_i \quad (5)$$

เมื่อให้ $\sum$ เป็นความแปรปรวน และความแปรปรวนร่วมในเมตริกของ x และเป็นไดอะกอนอล (Diagonal) จะได้

$$\sum = \Lambda \Phi \Lambda' + \Psi \quad (6)$$

เมื่อ Φ คือ ความแปรปรวนของ f

ซึ่งสมการ (5) และ (6) เป็นสมการพื้นฐานของรูปแบบการวิเคราะห์ องค์ประกอบที่มีตัว

ประกอบร่วม (Common factor) ตัวหนึ่ง ถ้าแวกเตอร์ X ของคะแนนสอบมีการแจกแจงเป็นโค้งปกติ การกะประมาณความเป็นไปได้สูงสุดของพารามิเตอร์ที่ไม่ทราบค่า V, Λ, Θ และ Ψ ซึ่งเป็นค่าความคลาดเคลื่อนมาตรฐาน ก็กะประมาณได้เช่นเดียวกัน

แบบทดสอบที่มีรากฐานของการวัดในโครงสร้างที่เหมือนกันทั้งหมด โปรแกรมการวิเคราะห์ตัวประกอบคงที่ (Confirmatory factor analysis program) (Sorbom and Joreskog, 1976) จะเป็นชนิดหนึ่งของแบบทดสอบที่มีข้อตกลงเบื้องต้นที่สำคัญเกี่ยวกับรูปแบบ และการทดลองภายในกลุ่มประชากรเดียวกัน และต่างกลุ่มกัน ต่อมาได้มีการเปรียบเทียบสัดส่วนความเป็นไปได้ของการทดสอบ การกะประมาณความเป็นไปได้สูงสุด (Maximum likelihood) ของความแปรปรวนร่วมเมตริก Σ ของประชากร คือ $\Sigma = \hat{\Lambda} \hat{\Theta} \hat{\Lambda}' + \Psi$ การทดสอบร่วมกันได้เหมาะสมกับโมเดลมากที่สุด ซึ่งในการเทียบ อาจจะใช้ประชากรร่วมกัน หรือต่างกลุ่ม ที่แสดงถึงค่าความคลาดเคลื่อนมาตรฐานที่นำไปสู่ค่าพารามิเตอร์

จากการทดสอบร่วมกัน (Overall) ได้เหมาะสม (fit) กับโมเดลมากที่สุดซึ่งการเท่ากันอาจจะใช้ประกอบร่วมกัน หรือต่างกลุ่ม ที่แสดงถึงค่าความคลาดเคลื่อนมาตรฐานที่นำไปสู่ค่าพารามิเตอร์ที่ไม่ทราบค่า

การเทียบมาตรฐานรูปแบบการใช้เทคนิคการวิเคราะห์องค์ประกอบใช้แบบทดสอบร่วม

การเทียบมาตรฐานของแบบทดสอบ 2 ชุด ที่ใช้แบบทดสอบร่วมซึ่งเป็นการออกแบบนำไปสู่โมเดลของสองแฟกเตอร์ (a two-factor model)

ให้กลุ่มตัวอย่างที่ 1 ได้รับการทดสอบด้วยแบบทดสอบ X_1 และ X_2

กลุ่มตัวอย่างที่ 2 ได้รับการทดสอบด้วยแบบทดสอบ X_1 และ X'_2

ถึงแม้ว่า X_2 จะมีโครงสร้างที่คู่ขนานกับ X'_2 แต่ก็ยังแตกต่างกันในหน่วยของการวัด จุดเริ่มต้นของสเกล และความแปรปรวนคลาดเคลื่อนที่ไม่เท่ากัน ฉะนั้นความจำเป็นและเงื่อนไขสำหรับที่จะทำให้ X_2 กับ X'_2 เทียบกันจะขึ้นอยู่กับสององค์ประกอบร่วม คือรูปแบบของสององค์ประกอบเหมาะสมกับกลุ่มตัวอย่างทั้งสอง และตามข้อตกลงที่จะให้ค่าความเชื่อมั่นเท่ากันโดยทดสอบการเปลี่ยนแปลงในสมการ คือ

$$X_1 = V_i + \Lambda_i f + \Lambda_{ie} \quad (\text{Werts, 1980})$$

นั่นคือ ค่าความแปรปรวนของความคลาดเคลื่อนลดลง โดยใช้ค่าน้ำหนักองค์ประกอบ (factor loading) สำหรับความสัมพันธ์ของการวัด เมื่อค่าความเชื่อมั่นเท่ากันแล้ว จะได้ค่าคะแนนจากแบบทดสอบชุดที่ 2 (X_2 กับ X'_2) เทียบกันได้ โดยเอา V_i ซึ่งเป็นค่าคงที่ลบออกจาก X_i แต่ละตัวและหารด้วยค่าน้ำหนักของตัวประกอบ (factor loading) แล้ว ในแต่ละ X_i ก็สามารถแสดงความสมมูลของคะแนนจริงในการเทียบมาตรฐาน

ขั้นตอนวิธีเทียบมาตรฐานรูปแบบการใช้เทคนิคการวิเคราะห์องค์ประกอบ

1. รวบรวมข้อมูลที่ได้จากการสอบจากแบบทดสอบทั้งสองชุด
2. นำข้อมูลทั้งหมดไปวิเคราะห์องค์ประกอบสำคัญ เพื่อจะดูจำนวนองค์ประกอบ น้ำหนักขององค์ประกอบ ค่าความร่วมกัน
3. กะประมาณคะแนนสมมูล (f) รายข้อ และรายบุคคลจากค่าสถิติในข้อ 2 เพื่อนำมาจัดตารางสมมูลระหว่างคะแนนดิบของแบบทดสอบทั้ง 2 ฉบับ

จากรูปแบบของการเทียบมาตรฐานที่กล่าวมาทั้งหมด ได้มีผู้ศึกษาเปรียบเทียบความเหมาะสมของแต่ละรูปแบบ หลายกลุ่มด้วยกัน แต่ก็ไม่สามารถสรุปได้ว่ารูปแบบใดดีที่สุด ทั้งนี้ขึ้นอยู่กับเกณฑ์และสถานการณ์ในการสอบแต่ละครั้ง หรือความจำกัดในการวิเคราะห์ข้อมูลที่มีส่วนในการตัดสินใจที่จะเลือกใช้รูปแบบการเทียบมาตรฐาน แต่อย่างน้อยที่สุดวิธีการเทียบมาตรฐานก็สามารถนำมาเป็นวิธีการเทียบคะแนนที่ก่อให้เกิดความยุติธรรมกับผู้สอบ เพื่อให้การวัดผลประเมินผลตรงกับสภาพความเป็นจริงของผู้สอบนั่นเอง

บรรณานุกรม

- ชูชีพ พงษ์สมบุรณ์ การเปรียบเทียบการเทียบมาตรฐานระหว่างรูปแบบที่ใช้ผู้สอบร่วมกับรูปแบบที่ใช้แบบสอบร่วม
วิทยานิพนธ์ ค.ม. จุฬาลงกรณ์มหาวิทยาลัย 2528, 118 หน้า ถ่ายเอกสาร
- ชูศักดิ์ วัฒนกิจิต **Test equating procedures** เอกสารประกอบการบรรยาย ในการสัมมนาระดับชาติครั้งที่ 4 โครงการ
พัฒนาศึกษาอาเซียน 2527, 13 หน้า
- นิคม ดังคะพิภพ "Equating," ในเอกสารประกอบการเรียนของหลักสูตรดุขฎิบัณฑิต จุฬาลงกรณ์มหาวิทยาลัย
ม.ป.ป. (หน้าไม่ติดต่อกัน) จัดสำเนา
- ภาวณี ศรีสุขวัฒนานันท์ การเปรียบเทียบผลจากการใช้รูปแบบการเทียบมาตรฐานที่ต่างกันเมื่อแบบสอบร่วมมีความ
ต่างกัน วิทยานิพนธ์ ค.ด. จุฬาลงกรณ์มหาวิทยาลัย 2529, 147 หน้า ถ่ายเอกสาร
- เรวดี อินทสระระ การเปรียบเทียบความคลาดเคลื่อนของการเทียบมาตรฐานระหว่างรูปแบบ อิงทฤษฎีการตอบ
ข้อสอบกับรูปแบบการใช้เทคนิคการวิเคราะห์องค์ประกอบ ปริญญาณิพนธ์ กศ.ม. มหาวิทยาลัยศรีนครินทร
วิโรฒ ประสานมิตร, 2530
- Angoff, W.H. **Scales Norms and Equivalent Scores**. Educational Testing Service, 1984.
- _____. **Educational Measurement** 2nd ed. Washington, D.C. American Council on Education, 1971. 600 p.
- _____. "Summary and Derivation of Equating Methods Used at ETS," in **Test Equating**. p. 55-70, ed. by P.W. Holland
and D.B. Rubin. New York, Academic Press, 1982.
- Braun, Henry I. and Holland, P.W. "Observed-Score Test Equating : A Mathematical Analysis of Some ETS Equating Procedures,"
in **Test Equating**. p. 9-50, ed. by P.W. Holland and D.B. Rubin. New York, Academic Press, 1982.
- Cook, Linda L, Dunbar, S.B., and Eignor, D.R. **IRT Equating : A Flexible Alternative to Conventional Methods for Solving
Practical Testing Problems**. Los Angeles, 1981.
- Flanagan, T.C. **Educational Measurement**. Washington D.C. : American Council on Education, 1951.
- Gronlund, Norman E. **Measurement and Evaluation in Teaching**. New York, Mcmillan, 1976. 590 P.
- Gulliksen, Harold. **Theory of Mental Tests**. New York, Willey, 1950. 486 p.
- Holmes, S.E. "Unidimensionality and Vertical Equating with the Rarch Model, " **Journal of Educational Measurement**, 19 : 139-
147, Summer, 1982.
- Holland, Paul W. and Rubin Donald B. ed. **Test Equating** New York, Academic Press, 1982. 365 p.
- Jager, R.M. "Some Exploratory Indices for Selection of a test Equating Method. " **Journal of Educational Measurement** 18 : 23-
38, Spring, 1981.
- Klein, Lawrence W. and Kolen, Michael J. **Effect of Number of Common in Common-Item Equating with Nonrandom Groups**.
Chicago, 1985.
- Kolen, M.J. "Comparison of Traditional and Item Response Theory Methods for Equating Tests," **Journal of Educational Measure-
ment** 18 : 1-11, Spring, 1981.
- Kolen, M.J. and Whitney, D.R. "Comparison of Four Procedures for Equating the Tests of General Educational Development," **Journal
of Education Measurement** 19 : 279-294, Summer, 1982.
- Lord, F.M. "Practical Applications of Item Characteristic Curve Theory, " **Journal of Educational Measurement** 14 : 117-138,
Summer, 1977.
- _____. **Applications of Item Response Theory to Practical Testing Problem**. Hillsdale, N.J. : Erlbaum 1980. 247 p.
- _____. "The Standard error of Equipercntile Equating," **Journal of Educational Statistics** 7 : 165-174, 1982.
- Marco, Cary L. "Item Characteristic Curve Solutions to Three Intractable Testing Problems" **Journal of Educational Measurement**
14 : 139-160, Summer, 1977.
- _____. "Equating Tests in an Era of Test Disclosure." **New Directions for Testing and Measurement** : 105-122, September,
1981.
- Petersen, N.S., Marco, Cr.L. and Stewart, E.E. "A test of the Adequacy of Linear Score equating Methods," in **Test Equating**.
p. 71-135, ed. by P.W. Holland and D.B. Rubin. New York, Academic Press, 1982.
- Petersen, N.S., Cook, L. and Stocking, L. **IRT Versus Conventional Equating Methods : A Comparative Study of Scale Stability**.
Los Angeles, 1981.

- Rock, D.A. "Equating Using the Confirmatory Factor Analysis Model," in **Test Equating**, p. 251-254, ed. by P.W. Holland and D.B. Rubin. New York, Academic Press, 1982
- Rubin, D.B. "Discussion of observed-Score Equating : A Mathematical Analysis of Some ETS Equating Procedures," in **test Equating**, p. 51-54, ed. by P.W. Holland and D.B. Rubin. New York, Academic Press, 1982
- Slinde, J.A. and Linn, R.L. "Vertical equating test : Fact or Phantom" **Journal of Educational Measurement**, 14 : 23-35, Spring, 1978
- "The Rasch Model, Objective Measurement Equating and Robustness." **Applied Psychological Measurement** 3 : 437-452, Summer, 1979
- "An Exploration of the Adequacy of the Rasch Model for the problem of Vertical equating" **Journal of Educational Measurement** 15 : 23-35, Spring, 1978.
- Stroud, T.W. "Discussion of A test of the Adequacy of Linear Score Equating Models," in **Test Equating** p. 137-140, ed. by P.W. Holland and D.B. Rubin. New York, Academic Press, 1982.
- Thorndike, R.L. **Applied Psychometrics**, Boston : Houghton Mifflin, 1982. 390 p.
- Wright, Benjamin D. and Stone, Mark H. **Best Test Design : Rasch Measurement**, Chicago, MESA PRESS, 1979. 222 p.

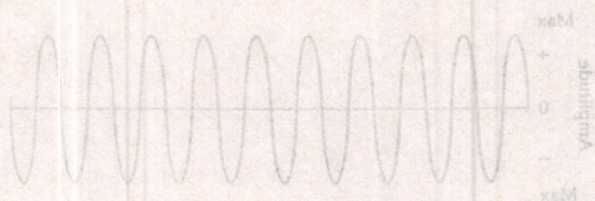


Figure 1: A periodic waveform (sine wave) plotted against time. The vertical axis is labeled 'Amplitude' and has markings for 'Max', '0', and 'Min'. The horizontal axis is labeled 'Time' and has a marking for '1 second'.

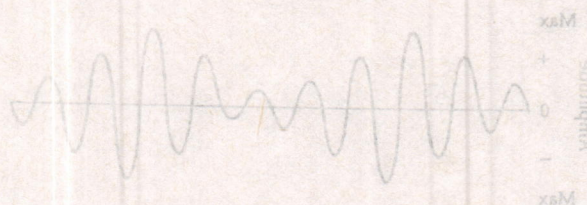


Figure 2: A periodic waveform (sine wave) plotted against time. The vertical axis is labeled 'Amplitude' and has markings for 'Max', '0', and 'Min'. The horizontal axis is labeled 'Time' and has a marking for '2 seconds'.