

การเทียบคะแนนแบบทดสอบ

ความหมาย

การเทียบคะแนนแบบทดสอบ (Test Equating) คือ การสร้างฟังก์ชันเทียบคะแนนที่เหมาะสมเพื่อใช้สำหรับเทียบคะแนนแบบทดสอบที่วัดคุณลักษณะอย่างเดียวกันตั้งแต่สองฉบับขึ้นไปให้อยู่บนระบบหน่วยวัดเดียวกัน

ที่มาของการเทียบคะแนน

เนื่องจากแบบทดสอบแต่ละฉบับที่สร้างขึ้นเพื่อวัดคุณลักษณะอย่างเดียวกันมีระดับความยากไม่เท่ากัน เป็นผลให้ระบบหน่วยวัดของแบบทดสอบแต่ละฉบับเป็นอิสระจากกัน คะแนนที่ได้จากหน่วยวัดแต่ละฉบับจึงไม่สามารถเปรียบเทียบกันได้โดยตรง แต่ถ้าจำเป็นต้องนำคะแนนจากการวัดมาเทียบกันต้องทำคะแนนให้อยู่ในระบบหน่วยวัดเดียวกัน

รูปแบบและสาระสำคัญ

รูปแบบการเทียบคะแนนแบ่งออกเป็น 4 รูปแบบใหญ่ คือ

1. การเทียบคะแนนเชิงเส้นตรง (Linear Equating)
2. การเทียบคะแนนแบบไม่เป็นเส้นตรง (Non-Linear Equating)
3. การเทียบคะแนนตามวิธีการของเทอร์สโตน (Thurstone Equating)
4. การเทียบคะแนนตามทฤษฎีตอบสนองข้อสอบ (Item Response Theory Equating)

การเทียบคะแนนแต่ละรูปแบบมีสาระสำคัญดังนี้

1. การเทียบคะแนนเชิงเส้นตรง คือ การเทียบคะแนนซึ่งถือข้อตกลงที่สำคัญดังนี้

- (1) แบบทดสอบแต่ละฉบับที่นำมาเทียบคะแนนต้องวัดคุณลักษณะเดียวกัน
- (2) แบบทดสอบแต่ละฉบับที่นำมาเทียบคะแนนต้องมีความเชื่อมั่นเท่ากัน
- (3) แบบทดสอบแต่ละฉบับที่นำมาเทียบคะแนนต้องมีการแจกแจงเหมือนกัน

วิธีการเทียบคะแนนเชิงเส้นตรงกระทำได้โดยการใช้ค่าเฉลี่ย (\bar{X}) และความเบี่ยงเบนมาตรฐาน (S.D) ซึ่งได้จากกลุ่มตัวอย่างสร้างฟังก์ชันเทียบคะแนนให้ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของแบบทดสอบทั้งสองฉบับเท่ากัน และใช้ฟังก์ชันดังกล่าวปรับคะแนนดิบทุกๆ จุดภายใต้การแจกแจงของคะแนนของแบบทดสอบทั้งสองฉบับให้สามารถเปรียบเทียบกันได้แบบคะแนนต่อคะแนน รูปแบบของการเทียบคะแนนเชิงเส้นตรง มีดังนี้

(1) กรณีกลุ่มตัวอย่างกลุ่มเดียว คือ การเทียบคะแนนโดยนำแบบทดสอบทั้ง 2 ฉบับ ไปทดสอบกลุ่มตัวอย่างเพียงกลุ่มเดียว แล้วนำค่า \bar{X} และ S.D. มาสร้างฟังก์ชันเทียบคะแนน

(2) กรณีกลุ่มตัวอย่างชนิดสุ่ม การเทียบคะแนนโดยนำแบบทดสอบ 2 ฉบับไปทดสอบคือใช้กลุ่มตัวอย่าง 2 กลุ่ม ซึ่งสุ่มมาจากประชากรเดียวกัน หรือใช้แบบทดสอบฉบับที่ 1 กับกลุ่มหนึ่งและฉบับที่ 2 กับกลุ่มสอง แล้วแต่กรณี แล้วนำค่า \bar{X} และ S.D. มาสร้างฟังก์ชันเทียบคะแนนทั้งหมดที่ได้จากกลุ่มตัวอย่าง

(3) กรณีใช้แบบทดสอบร่วม คือ การเทียบคะแนนโดยการนำแบบทดสอบ 2 ฉบับ ไปทดสอบกลุ่มตัวอย่าง 2 กลุ่ม ซึ่งสุ่มมาจากประชากรที่มีความสามารถแตกต่างกันแล้วนำ \bar{X} , S.D. และค่าเชื่อมั่นทั้งหมดที่ได้จากกลุ่มตัวอย่างมาสร้างฟังก์ชันเทียบคะแนน

(4) กรณีใช้บุคคลร่วมกัน คือ การเทียบคะแนนโดยนำแบบทดสอบ 2 ฉบับไปทดสอบกลุ่มตัวอย่าง 3 กลุ่ม ซึ่งสุ่มมาจากประชากรเดียวกัน ฉบับที่หนึ่งทดสอบกลุ่ม 1 และฉบับที่สองทดสอบกลุ่ม 2 ส่วนกลุ่ม 3 ให้เป็นกลุ่มร่วมที่ทดสอบทั้งฉบับที่ 1 และฉบับที่ 2 แล้วนำ \bar{X} และ S.D. ของแบบทดสอบแต่ละฉบับมาสร้างฟังก์ชันเทียบคะแนน

2. การเทียบคะแนนแบบไม่เป็นเส้นตรง หรือเรียกอีกอย่างหนึ่งว่าการเทียบคะแนนระบบเปอร์เซ็นต์ไทล์เท่ากัน คือการเทียบคะแนนกรณีที่ใช้แบบทดสอบทั้ง 2 ฉบับ วัดคุณลักษณะเดียวกัน มีค่าความเชื่อมั่นเท่ากัน แต่การแจกแจงของคะแนนดิบแตกต่างกันมาก การเทียบคะแนนของทั้ง 2 ฉบับ สามารถเทียบกันได้คะแนนต่อคะแนนก็ต่อเมื่อตำแหน่งเปอร์เซ็นต์ไทล์ของคะแนนแบบทดสอบทั้ง 2 ฉบับมีค่าเท่ากัน

3. การเทียบคะแนนตามวิธีการของ เซอร์สโตน คือการเทียบคะแนนความสามารถทางสติปัญญาแนวตั้ง ตามระดับชั้นหรือระดับอายุของกลุ่มตัวอย่าง การสร้างฟังก์ชันการเทียบคะแนนของเซอร์สโตนใช้ค่าระดับความยากของข้อสอบร่วมระหว่างแบบทดสอบที่ใช้สอบกลุ่มตัวอย่างแต่ละระดับโดยการแปลงค่าระดับความยากของข้อคำถามร่วมให้เป็นคะแนนมาตรฐาน (z - score) และสร้างฟังก์ชันการเทียบคะแนนโดยถือข้อตกลงที่สำคัญ 3 ข้อดังนี้

(1) คะแนนจากแบบทดสอบมีการแจกแจงปกติ

(2) แบบทดสอบแต่ละฉบับสามารถวัดได้ถึงระดับอันตรภาค

(3) ค่าระดับความยากอยู่ในรูปคะแนนมาตรฐานของข้อสอบร่วมของแบบทดสอบแต่ละคู่มิมีความสัมพันธ์กันเชิงเส้นตรง และสหสัมพันธ์เท่ากับ 1.00

4. การเทียบคะแนนตามทฤษฎีตอบสนองข้อสอบ (Item Response Theory Equating) คือการเทียบคะแนนที่ถือข้อกำหนดที่สำคัญว่า การเทียบคะแนนของแบบทดสอบ X และ Y จะสัมฤทธิ์ผลสำหรับผู้สอบแต่ละคนก็ต่อเมื่อผลการเทียบคะแนนจะต้องไม่แตกต่างกันในทุกๆระดับความสามารถของผู้สอบ ไม่ว่าผู้สอบจะสอบแบบทดสอบ X หรือ Y ก็ตาม นอกจากนั้นยังมีเงื่อนไขเพิ่มเติมที่สำคัญดังนี้

(1) แบบทดสอบที่นำมาเทียบคะแนนแต่ละฉบับ ต้องวัดเพียงคุณลักษณะเดียว (unidimensionality)

(2) แบบทดสอบที่นำมาเทียบคะแนนแต่ละฉบับต้องวัดคุณลักษณะอย่างเดียวกัน

(3) แบบทดสอบที่นำมาเทียบคะแนนแต่ละฉบับต้องมีความเชื่อมั่นเท่ากัน

(4) ข้อคำถามของแบบทดสอบแต่ละฉบับต้องมีคุณลักษณะตรงตามตัวแบบของทฤษฎีตอบสนองข้อสอบ (model fit) ตัวแบบใดตัวแบบหนึ่ง

วิธีเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบมีสองลักษณะดังนี้

(1) การเทียบค่าพารามิเตอร์ ความสามารถของบุคคล ซึ่งประมาณค่าจากคะแนนดิบของแบบทดสอบ 2 ฉบับซึ่งวัดคุณลักษณะเดียวกันให้อยู่บนระบบวัดเดียวกัน

(2) การเทียบค่าพารามิเตอร์ ข้อคำถามของแบบทดสอบแต่ละฉบับให้อยู่บนระบบหน่วยวัดเดียวกัน และใช้วิธีคำนวณคะแนนจริงของบุคคลซึ่งสอบแบบทดสอบแต่ละฉบับ ณ ระดับความสามารถเดียวกัน คะแนนจริงจากแบบทดสอบทั้งสอง คือ คะแนนซึ่งสามารถเทียบกันได้

ตามปกติแล้วการเทียบคะแนนตามทฤษฎีตอบสนองข้อสอบนั้น ต้องใช้โปรแกรมคอมพิวเตอร์ BILOG หรือ LOGIST

รูปแบบการเทียบคะแนนตามทฤษฎีตอบสนองข้อสอบมี 3 รูปแบบคือ

- (1) กรณีกลุ่มตัวอย่างกลุ่มเดียว
- (2) กรณีใช้แบบทดสอบร่วม
- (3) กรณีใช้บุคคลร่วม

การนำวิธีการเทียบคะแนนไปใช้ในการวัดผลการศึกษา

ในวงการวัดผลการศึกษา มีการเทียบคะแนน 2 ลักษณะคือ การเทียบคะแนนแนวนอน และการเทียบคะแนนแนวตั้ง

1) การเทียบคะแนนแนวนอน คือ การเทียบคะแนนแบบทดสอบหลายๆ ฉบับซึ่งวัดความสามารถระดับเดียวกัน เช่น การเทียบคะแนนแบบทดสอบ 2 ฉบับที่วัดความสามารถในการใช้ภาษาไทยระดับชั้นประถมศึกษาปีที่ 6 การเทียบคะแนนกรณีนี้สามารถใช้รูปแบบการเทียบคะแนนต่างๆ ได้ดังนี้

(1) การเทียบคะแนนเชิงเส้นตรง

(2) การเทียบคะแนนแบบไม่เป็นเส้นตรง

(3) การเทียบคะแนนตามทฤษฎีตอบ

สนองข้อสอบ

2) การเทียบคะแนนแนวตั้ง คือ การเทียบคะแนนแบบทดสอบหลายๆ ฉบับซึ่งวัดความสามารถต่างระดับกัน เช่น การเทียบคะแนนแบบทดสอบวัดความสามารถในการใช้ภาษาไทยระดับป. 4- ป. 6 เป็นต้น การเทียบคะแนนกรณีนี้สามารถใช้รูปแบบการเทียบคะแนนต่างๆ ได้ดังนี้

(1) การเทียบคะแนนตามวิธีการของเธอร์สตัน

(2) การเทียบคะแนนตามทฤษฎีตอบสนองข้อสอบ

การเลือกรูปแบบการเทียบคะแนนสำหรับการเทียบคะแนนแนวนอนและการเทียบคะแนนแนวตั้ง ให้พิจารณาความสอดคล้องระหว่างข้อมูลจากกลุ่มตัวอย่างและข้อตกลงหรือเงื่อนไขที่สำคัญของการเทียบคะแนนแต่ละรูปแบบ

ในประเทศไทยได้มีการทำการเทียบคะแนนของแบบทดสอบ เช่น สุรรัตน์ ขวัญบุญจันทร์ ได้สร้างแบบทดสอบและตารางเทียบคะแนนในแนวนอนวิชาวิทยาศาสตร์ สุจินดา ผ่องอักษร ได้เทียบคะแนนเพื่อศึกษาความก้าวหน้าของผลสัมฤทธิ์ทางการเรียนชั้นประถมศึกษาปีที่ 6 ที่เรียนจบตามหลักสูตรประถมศึกษา พุทธศักราช 2521 ในช่วงเวลาระหว่างปีการศึกษา 2529 ถึง 2531 เป็นต้น

ระพีพันธ์ โพธิ์ศรี

บรรณานุกรม

- สุจินดา ผ่องอักษร. การศึกษาความก้าวหน้าของผลสัมฤทธิ์ทางการเรียนในกลุ่มทักษะของนักเรียนชั้นประถมปีที่ 6 ที่เรียนจบตามหลักสูตรประถมศึกษา พุทธศักราช 2521 ในช่วงระยะเวลา 3 ปีการศึกษา (ปีการศึกษา 2529 ถึง 2531) โดยใช้การเทียบคะแนนรูปแบบราชส์. ปรินญาณินพนธ์ กศ.ค. กรุงเทพฯ : มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร, 2533.
- สุรัตน์ ขวัญบุญจันทร์. การสร้างแบบทดสอบและตารางเทียบคะแนนในแนวนอนวิชาวิทยาศาสตร์ที่วิเคราะห์ด้วยราชส์โมเดล. ปรินญาณินพนธ์ กศ.ม. กรุงเทพฯ : มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร, 2531.
- Allen, M.J. & Wendy M.Y. **Introduction to Measurement Theory.** Monterey : Brooks/Cole., 1979.
- Angoff, W.H. "Scales, norms and equivalent score." in R.L. Thordike (Ed.), **Educational measurement.** 2 nd ed. Washington DC : American Council on Education, 1971.
- Braun, H.I. & Holland, P.W. "Observed - Score Test Equation : A Mathematical Analysis of Some ETS Equating Procedures." in P.W. Holland & D.W. Rubin (Ed.) **Test Equating.** New York : Academic Press. 1982.
- Flanagan, J.C. "Units, Scores and Norms." in E.F. Lindquist (Ed.). **Educational measurement.** Washington DC : American Council On Education, 1961.
- Gulliksem, H. **Theory of Mental tests.** New York : Wiley, 1950.
- Hambleton, R.K. & Swaminaman, H. **Item Respones Theory.** Boston : Kluwer - Nijoff. 1985.
- Lord. F.M. "Note on comparable scales for test score." **Research Bulletin.** No. 48. 1995.
- _____. "Education test score : a maximum likellhood solution." **Psychometrika.** 20, 193-200.; 1955.
- _____. **Application of Item Response Theory to Practical Testing Problems.** New Jersey : Lawrence Erlbaum. 1980.
- _____. & Novick. M.R. **Statistical Theories of Mental Tests Scores.** Massachusetts : Addison - Wesley. 1968.

Magnusson, D. **Test Theory**. Stockholm : Stockhome University Press, 1966.

Nunnally, J.C. **Psychometric Theory**. New York : McGraw - Hill, 1976.

Popham, W.J. **Modern Educational Measurement**. New Jersey : Prentice - Hall, 1981.

Posrie, R. **The Differences among reliability estimates of randomly Parallel tests and their effects on Tucker nonrandom group linear equating**. Unpublished doctoral dissertation. The Florida State University, Florida, 1991.

Thurstone, L.L. "A method of scaling psychological and educational tests," **The Journal of Educational Psychology**. 16 : 433 - 451; 1925.