บทที่ 9
คะแนนแปลงรูป (Derived Score)

I จุดประสงค์การเรียนรู้

หลังจากจบบทเรียนนี้แล้ว ผู้เรียนสามารถปฏิบัติดังนี้

1. บอกจุดมุ่งหมายที่ต้องแปลงคะแนนเป็นรูปแบบต่าง ๆ
2. อธิบาย คำนวณ คะแนนแปลงรูปชนิดต่าง ๆ ได้
3. เปรียบเทียบคะแนนแปลงรูปทั้ง 4 แบบได้
4. เมื่อกำหนดข้อมูลมาให้สามารถแปลงรูปคะแนนให้เหมาะสมกับข้อมูลและจุดประสงค์การใช้ได้

II รายละเอียดของเนื้อหา

Derive Score หรือคะแนนแปลงรูปนั้น เป็นรายงานทางตัวเลขของแบบทดสอบที่แสดงในลักษณะของตำแหน่งที่สัมพันธ์กันกับกลุ่มที่เกี่ยวข้อง การแปลงคะแนนดิบให้เป็นคะแนนแปลงรูปนั้น ถ้าใช้ในแบบทดสอบมาตรฐานจะเห็นได้ง่าย นั่นคือคะแนนแปลงรูปทั้งหมดจะปรากฏอยู่ในตารางเกณฑ์ปกติในคู่มือแบบทดสอบ ซึ่งเราสามารถจะเทียบคะแนนดิบให้ตรงกับคะแนนแปลงรูปนั้น

ในการใช้คะแนนแปลงรูปเพื่อทำให้คะแนนดิบมีความหมายมากขึ้นนั้น มีจุดมุ่งหมาย 2 ประการคือ

1. เพื่อนำคะแนนมาเปรียบเทียบโดยใช้ scale เดียวกัน
1. เพื่อทำให้การตีความหมายคะแนนมีความหมายมากขึ้น

คะแนนแปลงรูปนั้นสามารถจำแนกออกได้เป็น 4 รูปคือ

1. โดยการเปรียบเทียบคะแนนกับมาตรฐานหรือพิจารณาตามความยากง่ายของเนื้อหาวิชา

2. โดยการเปรียบเทียบแต่ละคนในกลุ่ม

3. โดยการเปรียบเทียบคะแนนของคน ๆ เดียวกันแต่มีคะแนน 2 ชุด

4. โดยการกล่าวถึงที่มาของพื้นฐานของคะแนนรูปแบบอื่น ๆ

การจำแนกคะแนนแปลงรูปทั้ง 4 ประเภทมีรายละเอียดดังนี้

I Comparison with "Absolute Standard"; Content Difficulty

A Percentage correct scores

B Letter Grades (sometimes)

II Interindividual Comparison

A Considering mean and standard deviation (linear standard scores)

1 Z-SCORES

2 T-scores

3 AGCT-scores

4 CEEB-scores

5 Deviation IQ's (sometimes)

(a) Wechsler IQ's

(b) Stanford-Binet IQ's

B Considering rank within groups

1 Rank

2 Percential ranks and percentile bands

3 Letter Grades (sometimes)

4 Normalized standard scores (area transformation)

(a) T-Scaled scores

(b) Stanine scores

(c) C-Scaled scores

(d) Sten-scores

(e) Deviation IQ's (sometimes)

(1) Wechsler subtests

(f) ITED-scores

5 Decile ranks

C Considering the range of scores in a group

1 Percent Placement

D Considering status of those obtaining scores

1 Age scores

(a) Mental ages

(b) Educational ages etc.

2 Grade-placement scores

(a) Full-population grade - placement

(b) Modal-age grade-placement

(c) Modal-age and modal-intelligence grade-placement

(d) Anticipated-achievement grade-placement

(e) Mental-age grade placement

III Intra-Individual Comparison

A Ratio IQ's

B Intellectual Status Index

C Educational Quotient,

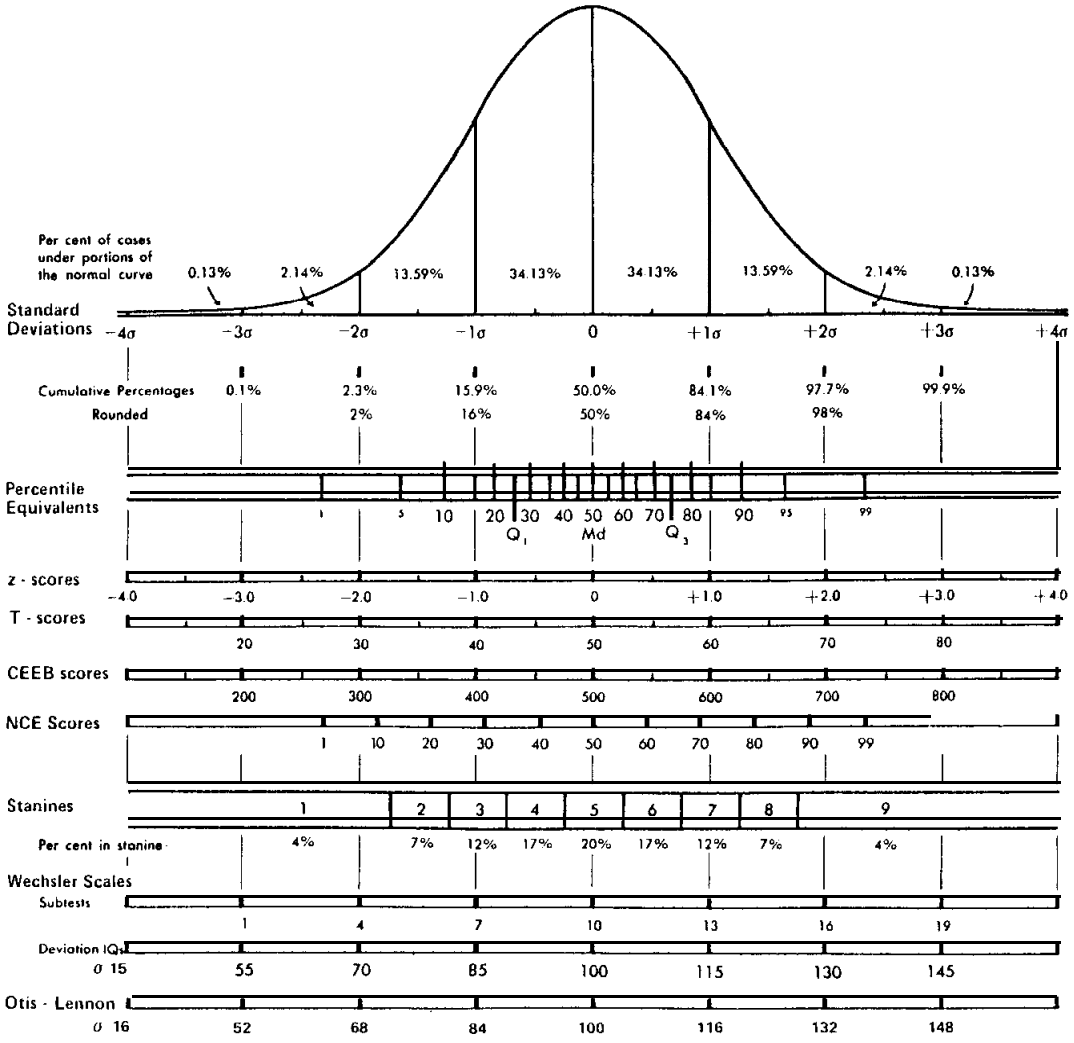
D Accomplishment Quotients

IV Assorted Arbitrary Bases

A Nonmeaningful scaled scores

B Long-range equi-unit scales

C Deviation IQ's (Otis-style)



The **Normal** Curve, Percentiles, **and Selected Standard** Scores

ภาพประกอบที่ 9.1 แสดงการเปรียบเทียบคะแนนแปลงรูปแบบต่าง ๆ

จากภาพจะเห็นว่าคะแนนใน Type II A & B มีความสัมพันธ์กัน เราสามารถจะแปลงคะแนนชนิดหนึ่งให้เป็นอีกอย่างหนึ่งได้ง่าย ๆ ถ้าการกระจายของกลุ่มตัวอย่างเป็นแบบปกติ (normal distribution) การเปรียบเทียบคะแนนเหล่านี้อยู่ภายใต้เงื่อนไข 2 ประการคือ

1. normality
2. same group

ถ้าเป็นคะแนนจากนักเรียนต่างกลุ่มกันจะเอามาเปรียบเทียบกันโดยตรงไม่ได้ และถ้าการกระจายคะแนนของกลุ่มไม่เป็นปกติ (nomal dist^{n}) ก็ไม่สามารถจะนำมาแปลงคะแนนโดยใช้ scale นี้เช่นกัน เพราะจะพบว่าความสัมพันธ์ของคะแนนนั้นจะเปลี่ยนไปเป็นอย่างอื่น

ตัวอย่าง แบบทดสอบฉบับหนึ่ง มี $\overline{X}$ = 300, s.d. = 40 ลักษณะการกระจายของคะแนนค่อนข้างจะเป็นการกระจายแบบปกติ (normal distribution) จะแปลงคะแนนเป็นอย่างไรได้บ้าง

ถ้าเด็กคนหนึ่งมีคะแนนการสอบ 300 คะแนน

จะได้ค่า Z-score	= 0
T-score	= 50
Stanine	= 5
Percentile Rank	= 50

เด็กอีกคนหนึ่งสมมุติว่ามีคะแนนการสอบจากข้อสอบฉบับนี้ 320 คะแนน

จะได้ Z-scores	= 0.5
T-score	= 55
T-score	= 6
Percentile Rank	= 69

Type I : Comparison with an absolute standard or Content Difficulty

Type I A Percentage correct

เป็นการเปรียบเทียบคะแนนของผู้เข้าสอบคนหนึ่ง ๆ กับข้อสอบ 100 ข้อ นิยมใช้ในการรายงานผลสัมฤทธิ์ทางการเรียนของนักเรียนในชั้นเรียน

สูตรที่ใช้

$$X_{\%c} = 100\ R/T$$

เมื่อ

$X_{\%c}$ = percentage correct score

R = จำนวนข้อที่ตอบถูก

T = จำนวนข้อสอบทั้งหมด

ตัวอย่าง เด็กชาย ก ทำข้อสอบถูก 44 ข้อ ในจำนวนข้อสอบทั้งหมด 50 ข้อ percentage correct score จะได้เท่ากับ

$$X_{\%c} = \frac{100 \times 44}{50}$$

$$= 88$$

percentage correct เป็นคะแนนแปลงรูปชนิดเดียว (ยกเว้น Type I Letter grade) ที่จะอนุมานว่าผู้ตอบข้อสอบมีความรู้ในเนื้อหาวิชาในแบบทดสอบเพียงใด ซึ่งจะสามารถนำมาใช้ประโยชน์ในการพิจารณาว่านักเรียนมีผลสัมฤทธิ์เพียงใด ก่อนที่จะนำคะแนนตัวนี้ไปเปรียบเทียบกับมาตรฐานที่กำหนดไว้ ทั้ง ๆ ที่ครูผู้สอนก็อาจจะตระหนักดีว่า การตัดสินเด็กล่วงหน้านั้นไม่ค่อยจะมีความเป็นปรนัยมากนัก และไม่สามารถเปลี่ยนไปตามที่ครูต้องการได้ การที่เด็กจะทำคะแนนได้มากหรือน้อยนั้นยังขึ้นอยู่กับความยากง่ายของข้อสอบในวิชานั้น ๆ ด้วย ครูบางคนอาจจะบวกคะแนนความเห็นของตนลงไปในทุก ๆ คะแนนถ้าคะแนนการสอบของนักเรียนต่ำมาก และบางครั้งก็อาจนำคะแนนจากการปฏิบัติงานอื่น ๆ มาเปรียบเทียบกับคะแนนจากการทดสอบก่อนที่จะพิจารณาตัดสินนักเรียน

Type I B Letter Grades

การให้คะแนนเด็กเป็นตัวอักษรนั้นก็เป็นวิธีการหนึ่งที่นิยมใช้กันทั่วไปทั้งระดับประถม มัธยม และอุดม พื้นฐานในการให้คะแนนเป็นตัวอักษรนั้นก็นำมาจาก percentage correct scores นั่นเอง บางครั้งอาจให้เป็นตัวอักษรแบบ Type II B 3 ก็ตาม แต่วิธีของ Type I B นั้นก็นิยมมากกว่า เช่น

90-100 จะได้เกรด A

80-89 จะได้เกรด B

ฯลฯ

การที่จะกำหนดว่าร้อยละเท่าไรจะให้เกรดอะไรนั้น ครูก็ย่อมจะปรับให้เหมาะสมกับลักษณะความยากง่ายของข้อสอบ คุณลักษณะของผู้สอบเป็นต้น การให้คะแนนแบบเป็นตัวอักษรกับแบบร้อยละนั้นก็มีข้อดีคล้าย ๆ กัน การให้คะแนนเป็นตัวอักษรจะเป็นการให้คะแนนเป็นหน่วยหยาบ ๆ ทำให้ไม่อาจมองเห็นความแตกต่างที่มีอยู่เล็ก ๆ น้อย ๆ ได้ แต่ถึงอย่างไรการให้คะแนนแบบนี้ก็ยังไม่แตกต่างจากคะแนนจริง ๆ มากนักและก็เป็นการให้คะแนนหน่วย

เดียวที่มีการแปรคะแนนออกไปอย่างกว้าง ๆ

การให้คะแนนแบบตัวอักษรนี้ ทำได้ 2 วิธีคือ

1. ให้เกรดตามคุณภาพที่ตัดสินไว้ (เช่นในแบบความเรียง)

2. แปลงจากร้อยละให้เป็นเกรด โดยทำตามข้อตกลงในการให้เกรดของโรงเรียนหรือของมหาวิทยาลัย หรือตามคำแนะนำในระเบียบการประเมินผลของกระทรวงศึกษาธิการ

เมื่อพบว่าการให้เกรดค่อนข้างจะตึงมากเกินไป ก็อาจจะนำข้อสอบมาพิจารณาความยากง่าย หรือเกณฑ์ที่กำหนด อย่างไรก็ตามการให้คะแนนเป็นเกรดนั้นก็มีจุดอ่อนตรงที่ว่าขึ้นอยู่กับความยากง่ายของข้อสอบมากกว่าจะพิจารณาจากความสามารถของกลุ่มที่ปฏิบัติได้จริง แต่ถ้าข้อสอบนั้นเป็นข้อสอบแบบอิงเกณฑ์ก็น่าจะแก้จุดอ่อนนี้ได้

Type II Interindividual Comparison

การแปลคะแนนแบบนี้มักนิยมใช้กับแบบทดสอบมาตรฐานมากกว่าข้อสอบที่ใช้ในห้องเรียน โดยทั่ว ๆ ไปแบบทดสอบมาตรฐานนิยมแปรคะแนนในรูปแบบของ Type II A, B, และ D

การแปลคะแนนแบบ Type II นี้เป็นการแปลคะแนนโดยยึดถือคะแนนของกลุ่มที่เข้ารับการทดสอบทั้งหมดเป็นเกณฑ์ การที่คะแนนจะแปรรูปไปอย่างไรจึงขึ้นอยู่กับความยากของเนื้อหาในแบบทดสอบ นักเรียนบางคนอาจทำข้อสอบฉบับหนึ่งได้คะแนนดี แต่เมื่อทำแบบทดสอบอีกฉบับหนึ่งคะแนนอาจจะน้อยกว่า ซึ่งก็ขึ้นอยู่กับความยากของเนื้อหาในข้อสอบแต่ละฉบับ แต่อย่างไรก็ตามไม่ว่าข้อสอบจะยากหรือจะง่ายก็ตาม นักเรียนทุกคนก็ทำแบบทดสอบฉบับเดียวกัน เมื่อเป็นเช่นนี้เราจึงสามารถจะนำคะแนนของเด็กแต่ละคนในกลุ่มเดียวกันมาเปรียบเทียบกันได้โดยใช้คะแนนแบบ Type II นั่นเอง คะแนนแบบ Type II แยกได้เป็น

Type II A. Considering mean and standard deviation

การพิจารณาคะแนนแบบ Type II A นี้ พิจารณาจากค่าเฉลี่ยและค่าความเบี่ยงเบนมาตรฐานเป็นหลักการพิจารณา คะแนนแบบนี้จึงมีคุณค่าทางการวิจัยมากกว่าคะแนนแปรรูปอื่น ๆ ทั้งนี้เนื่องจาก

1. มีค่าเฉลี่ยและค่าความเบี่ยงเบนมาตรฐานค่าเดียวกันทั้งกลุ่ม

2. ลักษณะการกระจายของคะแนนดิบจะคงรูปอยู่เช่นเดิมถึงแม้จำนวนจะมากขึ้นหรือน้อยลงก็ตาม

3. สามารถนำคะแนนดังกล่าวมาเปรียบเทียบกันได้

4. สามารถนำวิธีการทางคณิตศาสตร์มาจัดกระทำกับคะแนนพวกนี้ได้

คะแนนแบบ Type II A มีดังนี้คือ

1. Z-Score

เป็นคะแนนแปรรูปมาตรฐานเบื้องต้นของการแปรรูปคะแนนทุก ๆ คะแนนในรูปแบบนี้ คะแนนมาตรฐานทุกคะแนนที่เป็นการแปลงเชิงเส้น (linear transfermation) จะต้องแปลงมาจาก Z-score การแปลความหมายของคะแนนนี้ก็คือค่าที่นักเรียนแต่ละคนบ่ายเบนไปจากกลุ่ม การหาค่าคะแนน Z-score มีสูตรดังนี้

$$Z = \frac{X - \overline{X}}{S}$$

เมื่อ

Z = คะแนนมาตรฐาน Z

X = คะแนนดิบ

$\overline{X}$ = ค่าเฉลี่ยของคะแนนดิบในกลุ่ม

S = ค่าความเบี่ยงเบนมาตรฐานในกลุ่ม

ตัวอย่างเช่น ธำรงสอบได้คะแนน 49 ถ้าเปรียบเทียบคะแนนในกลุ่มซึ่งมีค่าเฉลี่ย 40 ค่าความเบี่ยงเบนมาตรฐาน 6

$$Z = \frac{49 - 40}{5}$$

$$= 1.5$$

แปลความหมายได้ว่า ธำรงมีคะแนนสูงจากค่าเฉลี่ย 1.5 ของความเบี่ยงเบนมาตรฐาน

ค่า Z-score นี้ อาจจะใช้ประโยชน์ในด้านการวิจัย แต่ในด้านการวัดผลแล้วไม่ค่อยได้นำมาใช้มากนัก แต่ก็อาจหาค่า Z-score ก่อน แล้วจึงมาหาค่า linear standard score อีกครั้งหนึ่ง

2. T-score

เป็น linear standard score ที่นิยมใช้กันมากที่สุด เหตุผลที่ใช้คะแนนชนิดนี้ก็คล้ายกับคะแนน Z-score ต่างกันแต่เพียงว่า T-score มีค่าเฉลี่ย 50 และค่าความเบี่ยงเบนมาตรฐาน = 10 มีสูตรในการคำนวณหาค่า T-score ดังนี้

$$T = 10Z + 50$$

เมื่อ $Z = \frac{X - X}{S}$

ตัวอย่างเช่น ถ้าคะแนน Z-score ของธำรง = 1.5

$$T = 10(1.5) + 50$$
$$= 65$$

ถ้าการกระจายของคะแนนกลุ่มนี้เป็นการกระจายแบบโค้งปกติแล้ว เราก็อาจสรุปได้ว่าธำรงทำคะแนนได้เท่ากับหรือดีกว่าคนในกลุ่มเดียวกัน 93% ถ้าเทียบกับภาพประกอบที่ 9.1 แล้ว คะแนน T 65 จะมีค่าเท่ากับ Z = 1.5

คะแนนแบบนี้มีข้อดีและขีดจำกัดคล้ายกับ Z-score ถึงแม้ว่าคะแนนแบบนี้ใช้ในการวิจัยแล้วจะมีประโยชน์น้อยกว่า แต่ก็ใช้ได้ดีในแง่ที่ว่าไม่มีค่าติดลบและไม่มีจุดทศนิยม

คะแนน T-score นี้มักใช้ปนกันกับคะแนนอื่น ๆ เป็นต้นว่า คะแนน T-Scaled score (คะแนนจาก Type II B score) คะแนน T ทั้งสองค่านี้จะมีค่าคล้ายกันถ้าการกระจายของคะแนนเป็นโค้งปกติ แต่ถ้าการกระจายของคะแนนเบ้ไปทางใดทางหนึ่ง ค่า T ทั้งสองค่านี้จะต่างกัน ค่า T-score นี้มักจะใช้ด้วยกันกับ Percentile rank ซึ่งอาจจะพิจารณาเปรียบเทียบดูจากภาพประกอบที่ 9.1

3. AGCT Score

คะแนนแบบนี้ย่อมาจาก Army General Classification Test มีลักษณะคล้ายกับคะแนน T-score หรือ Z-score เพียงแต่ว่า AGCT score มีค่าเฉลี่ย = 100 ค่าความเบี่ยงเบนมาตรฐาน = 20 วิธีการคำนวณคะแนนแบบนี้คือ

$$\text{AGCT score} = 20\,Z + 100$$

ตัวอย่างเช่น ถ้าธำรงได้คะแนน Z = 1.5

$$\text{AGCT} = 20\,(1.5) + 100)$$
$$= 130$$

เราอาจเทียบจากภาพประกอบ 9.1 จะเห็นว่า

Z = 1.5 = T-score = 65 และ AGCT score = 130

การใช้คะแนนแบบนี้ใช้ในการเปรียบเทียบคะแนนของแบบทดสอบที่ใช้คัดเลือกทหาร ซึ่งแบบทดสอบนี้มีค่าเฉลี่ย = 100 และค่าความเบี่ยงเบนมาตรฐาน = 20 และเราไม่นิยมใช้คะแนนแบบนี้มากนัก

4. CEEB Score

คะแนนแบบนี้ย่อมาจาก College Entrance Examination Board Test ซึ่งเป็นแบบทดสอบที่สำนักทดสอบทางการศึกษาของอเมริกา (Educational Testing Service) ใช้ในการสอบคัดเลือก

ผู้เข้าเรียนในมหาวิทยาลัย มีลักษณะเป็น linear standard score มีค่าเฉลี่ย = 500 และค่าความเบี่ยงเบนมาตรฐาน = 100

CEEB score = 100 Z + 500

ดังนั้นถ้าธำรงมีคะแนน Z-score = 1.5

CEEB score = 100 (1.5) + 500

= **650**

ปกติค่า CEEB score นี้จะเปลี่ยนไปเรื่อย ๆ ตามค่าเฉลี่ยและค่าความเบี่ยงเบนมาตรฐานของแต่ละกลุ่มที่เข้าสอบในแต่ละปี ค่าที่คำนวณได้นี้เป็นค่าที่ได้จากการทดสอบในปี ค.ศ. 1941 ดังนั้นจึงสามารถนำผลคะแนนในแต่ละปีมาเปรียบเทียบกับปีต่อ ๆ ไปได้ อย่างไรก็ตาม เนื่องจาก CEEB score ไม่ได้นำคะแนนมาจากพื้นฐานของผู้เข้าสอบที่เป็นปัจจุบัน ดังนั้น ETS จึงได้รายงานผลการสอบในรูปของ percentile rank ของผู้เข้าสอบในปัจจุบันแต่ละครั้ง

5. Deviation IQs (ในรูปแบบหนึ่ง)

คำว่า I.Q. (Intelligence Quotient) มีผู้นำมาใช้เมื่อ 70 ปีมาแล้ว โดยนักจิตวิทยาชาวเยอรมันชื่อ Stern ต่อมาในปีค.ศ. 1916 Terman ได้นำมาใช้กับแบบทดสอบของ Stanford-Binet ซึ่งต่อมาผู้สร้างแบบทดสอบอื่น ๆ ก็ใช้คำนี้ และมีคำว่า I.Q. ในแบบของคะแนน Type III ใช้อยู่ด้วย แต่ก็ไม่มากนัก I.Q. เป็นอัตราส่วนระหว่างอายุสมองและอายุปฏิทิน คะแนนแบบนี้มีประโยชน์ก็คือ ค่าความเบี่ยงเบนมาตรฐานสามารถใช้ได้กับผู้เข้าสอบทุกคนโดยไม่จำกัดอายุ

Deviation IQ ใช้เป็นค่าอธิบายคะแนน 3 ชนิด ซึ่งชนิดแรกที่จะกล่าวถึงนี้ (ชนิดที่ 2 คือ Type II B4 (e) และชนิดที่ 3 Type VI C) เป็นคะแนนแบบ linear standard score ซึ่งมีข้อดีและขีดจำกัดคล้ายกับ standard score ชนิดอื่น ๆ มีค่าเฉลี่ย = 100 ค่าความเบี่ยงเบนมาตรฐานแล้วแต่ผู้สร้างจะกำหนด ในที่นี้จะใช้ค่าเทียบกับคะแนนจาก Wechsler Intelligence Test กับ Stanford - Binet ปี 1960 ส่วนตอนแรก ๆ นั้น Stanford - Binet ใช้คะแนนแบบ Type III A

(a) Wechsler IQs มีแบบทดสอบที่นิยมมากอยู่ 3 ประเภทคือ

- Wechsler Preschool and Primary Scale of Intelligence (WPPSI)
- Wechsler Intelligence Scale for Children (WISC)
- Wechsler Adult Intelligence Scale (WAIS)

แบบทดสอบของ Wechsler ทั้ง 3 แบบนี้มีความแตกต่างกันในบางอย่าง เช่น แบบทดสอบของ WAIS มีแบบทดสอบย่อยที่วัดด้านภาษา 6 ฉบับ แบบทดสอบย่อยวัดการปฏิบัติ 5 ฉบับ ดังนั้นเวลาคิดคะแนนก็จะแยกออกมาเป็น Verbal I.Q., Performance I.Q. และ Total I.Q. กลุ่มตัวอย่างที่ใช้ทำเกณฑ์ปกติ (Norm Group) อยู่ในระหว่างอายุ 16-64 ปี แล้วนำมาแบ่งออกเป็น 7 กลุ่ม คะแนนดิบทั้ง 11 ฉบับ จะแปลงให้เป็น normal standard score (ดูจาก Type II B4 (e)) มีค่าเฉลี่ย = 10 ความเบี่ยงเบนมาตรฐาน = 3 ผลรวมของคะแนนดิบจะแปลงมาเป็น Deviation I.Q.

ในการสร้างแบบทดสอบของ WAIS นั้น จะต้องนำค่าเฉลี่ยและค่าความเบี่ยงเบนมาตรฐานของแต่ละแบบทดสอบย่อยทั้ง 11 ฉบับในแต่ละกลุ่มอายุ 7 กลุ่ม โดยผู้สร้างกำหนดค่าเฉลี่ยและค่าความเบี่ยงเบนมาตรฐานไว้ล่วงหน้าว่ามีค่าเฉลี่ย = 100 ค่าความเบี่ยงเบนมาตรฐาน = 15 ดังนั้นก็จะกำหนดสูตรว่า I.Q. = 15Z + 100 มาใช้คำนวณหาค่า I.Q. ในแต่ละกลุ่ม ซึ่งการหา I.Q. ของ WAIS จึงสามารถหาได้ทั้งจากคะแนนรวมทั้งหมด Verbal I.Q. จาก 6 แบบทดสอบย่อย และ Performance I.Q. จากแบบทดสอบย่อย 5 ฉบับ เป็นต้น

(b) 1960 Stanford - Binet I.Q.

I.Q. ชนิดนี้ได้ดัดแปลงขึ้นในปี ค.ศ. 1960 โดยกำหนดให้มีค่าความเบี่ยงเบนมาตรฐานคงที่ระหว่างอายุหนึ่งถึงอีกอายุหนึ่ง แบบทดสอบของ Stanford - Binet นี้ จะเริ่มตั้งแต่ระดับอายุ 2 ปีจนถึงเป็นผู้ใหญ่เต็มที่ ซึ่งถ้าดำเนินการสอบตามวิธีการที่กำหนดอย่างรัดกุมแล้วก็จะสามารถทราบระดับอายุสมองของผู้เข้าสอบได้ อายุสมองที่ได้เมื่อนำมาหาอัตราส่วนกับอายุจริงตามปฏิทินแล้ว ก็จะสามารถหา I.Q. ออกมาได้ โดยในแต่ละช่องของอายุจริงจะมีค่าเฉลี่ยค่าความเบี่ยงเบนมาตรฐานของอายุสมองเปรียบเทียบไว้ ค่าเฉลี่ยของอายุสมองจะมีค่าเท่ากับ

I.Q. 100 (ถ้าอายุสมองและอายุจริงเท่ากัน) ถ้าอายุสมองต่ำกว่าค่าเฉลี่ย 1 ความเบี่ยงเบนมาตรฐาน ค่า I.Q. = 84 และถ้าเหนือกว่าค่าเฉลี่ย 1 ความเบี่ยงเบนมาตรฐาน ค่า I,Q. = 116 ดังนั้น

Stanford Binet I.Q. มีสูตรหาค่า I.Q. = 16Z + 100 นั่นคือมีค่าเฉลี่ย = 100 และค่าความเบี่ยงเบนมาตรฐาน = 16

จากที่ได้กล่าวถึงคะแนนมาตรฐานเชิงเส้นตรง (Linear Standard Scores) มาตั้งแต่ต้นนั้น จะเห็นได้ว่าทุก ๆ รูปแบบของคะแนนจะทำให้เราทราบถึงตำแหน่งคะแนนดิบของผู้เข้าสอบโดยเกี่ยวข้องกับค่าเฉลี่ย และค่าความเบี่ยงเบนมาตรฐานของกลุ่มนั้น ๆ ในการกระจายของคะแนนไม่ว่าจะมีรูปร่างเป็นโค้งปกติ หรือไม่ก็ตาม ก็จะสามารถแปลงคะแนนแบบ linear standard scores ได้หมด โดยจะไม่ทำให้การกระจายของคะแนนเดิมเปลี่ยนแปลงรูปร่างไปเลย จากคุณสมบัตินี้เองเราสามารถจะดัดแปลงคะแนนได้ต่าง ๆ กันเหมือนกับที่เราสามารถจะทำได้ในคะแนนดิบ แต่ถ้าเป็นคะแนนในแบบ Type II B แล้วเราจะทำเช่นนั้นไม่ได้

Type II B Interindividual Compasion Considering Rank within Group

การแปลงคะแนนชนิดนี้นิยมใช้กันมากในการรายงานผลของแบบทดสอบมาตรฐาน วิธีการแปลงแตกต่างจาก Type II A ตรงที่ว่า คะแนนแบบนี้จะขึ้นอยู่กับจำนวนคนที่ได้คะแนนในแต่ละช่วง (ทั้งคะแนนสูงและคะแนนต่ำ) ไม่ใช่จะดูแต่ค่าของคะแนนเพียงอย่างเดียว คะแนนแบบนี้ไม่สนใจคะแนนที่ห่างจากค่าเฉลี่ยเลย แต่จะพิจารณาลักษณะการกระจายของคะแนนว่าเกือบจะเป็นการกระจายแบบโค้งปกติหรือไม่ การแปลงคะแนนแบบนี้มีดังนี้

1. Rank

การจัดเรียงอันดับ เป็นวิธีการหาความสัมพันธ์ระหว่างคะแนนที่ง่ายที่สุด คือการเรียงคนที่ได้คะแนนสูงสุดไว้เป็นคนที่ 1 แล้วเรียงลงมาตามลำดับเรื่อย ๆ ปกติมักไม่ค่อยนิยมใช้ในการรายงานผลการทดสอบทั่ว ๆ ไป

2. Percentile Rank และ Percentile Band

มีบางคนอาจเรียกว่า Centile Rank ก็ได้ คะแนนแบบนี้นิยมใช้ในการรายงานผลการทดสอบของแบบทดสอบมาตรฐาน และก็จัดได้ว่าเป็นแบบที่ดีที่สุดในการตีความหมายคะแนน อย่างไรก็ตามการใช้คะแนนแบบนี้ตีความหมายคะแนนก็มีข้อจำกัดบางอย่างดังนี้

Percentile คือตำแหน่งใด ๆ ใน 99 ตำแหน่งที่แบ่งตามการกระจายความถี่ของคะแนน ใน 100 คน Percentile Rank จึงเป็นตำแหน่งที่คน ๆ หนึ่งจะเกี่ยวพันกับกลุ่มของตน วิธีการหา Percentile Rank ของผู้เข้าสอบแต่ละคนก็โดยการดูจากคะแนนดิบของคน ๆ นั้นแล้วมาเทียบกับในตารางทีเป็น Percentile Scores

ตาราง 9.1 แสดงวิธีการหา Percentile และ Percentile Ranks

X	f	cf	$cf_{m.p.}$	$cP_{m.p.}$	PR.
225	1	50	49.5	99.0	99
224	1	49	48.5	97.0	97
223	2	48	47	94.0	94
222	4	46	44	88.0	88
221	2	42	41	82.0	82
220	5	40	37.5	75.0	75
219	6	35	32	64.0	64
218	8	29	25.0	50.0	50
217	5	21	18.5	37.0	37
216	4	16	14.0	28.0	28
215	4	12	10.0	20.0	20
214	4	8	6.0	12.0	12
213	3	4	2.5	5.0	5
212	0	1	1.0	2.0	2
211	1	1	0.5	1.0	1

วิธีทำ

1. เขียนค่าคะแนนดิบลงไป (X)
2. หาค่าความถี่ (f) ของคะแนนแต่ละคะแนน
3. **Ml** cumulative frequency (cf) ของคะแนนแต่ละตัว
4. **HI** cumulative frequency to mid point ($cf_{m.p.}$)
5. แปลงคะแนน $cf_{m.p.}$ เป็น cumulative percentage ($cP_{m.p.}$)

คือ $\frac{100}{N} \times cf_{m.p.}$

6. หา PR โดยปัดเศษทิ้ง

ตัวอย่าง จงหาค่า P_{30} ว่าคะแนนดิบที่มีคนอยู่ต่ำกว่า 30%

วิธีทำ

1. หาค่า 30% ของคน 50 คน = $30 \times \frac{50}{100} = 15$
2. นับจากข้างล่างขึ้นมา 15 คน จะเห็นว่าตรงกับ cf = 12
3. เอาจำนวนที่ต้องการจริง ๆ คือ 15 ลบด้วยค่า cf ในข้อ 2 = (15-12 = 3)
4. ที่ cf = 12 นั้น มีค่า f อยู่ 4 ค่า ให้พิจารณาค่าที่เหนือจาก cf = 12 ขึ้นไปอีก 1 ค่า จะตรงกับคะแนน 216
5. หาเศษส่วนจากข้อ 3 และข้อ 4 จะได้ $\frac{3}{4}$ หรือ = 0.75
6. ค่า raw score จะได้ 0.75 + 215.5 (lower limit) = 216.25 ดังนั้นค่า $P_{30} = 216.25$

ตัวอย่าง หาค่า P_{50} (ค่า median)

วิธีทำ

1. 50% ของกลุ่ม = $50 \times \frac{50}{100} = 25$
2. ดู $cf_{m.p.}$ ที่ 25 = 218 คะแนน

ดังนั้น P_{50} = 218 คะแนน

ตัวอย่าง หาค่า P_{80}

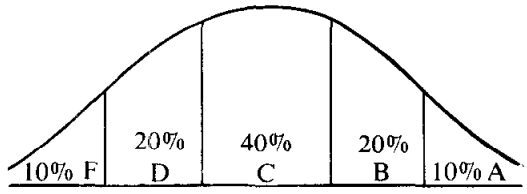
วิธีทำ

1. 80% ของกลุ่ม = $80 \times \frac{50}{100} = 40$
2. ดูค่า $cf_{m.p.} = 40 = 220$ คะแนน

ดังนั้น P_{80} = 220 หรือ 220.5 คะแนน

3. Letter Grades (บางรูปแบบ)

เป็นการให้คะแนนที่ใช้เปรียบเทียบผลการสอบในกลุ่มเด็กว่าจะมีคนได้เกรด A, B, C, D, F ตามระดับความสามารถของผู้เข้าสอบ ซึ่งมีใช้กันในแบบทดสอบมาตรฐานบางฉบับ แต่ก็ใช้กันน้อยมาก คะแนนแบบนี้ (Lyman, 1971 : 107) ได้วิจารณ์ว่าเป็นวิธีการให้คะแนนตามลักษณะโค้งปกติ ซึ่งค่อนข้างจะล้าสมัยแล้วการให้เกรดตามรูปแบบของโค้งปกติคือการกำหนดจำนวนคนที่จะได้เกรดในแต่ละเกรด เช่น 10% A, 20% B, 40% C, 20% D และ 10% F ตามรูป



ภาพประกอบที่ 9.2 แสดงการให้เกรดโดยใช้โค้งปกติ

ในปัจจุบันการให้เกรดแบบนี้ไม่นิยมใช้ยกเว้นแต่ว่ากลุ่มที่เข้าสอบนั้นไม่ได้คัดเลือกมาก่อน (unselected group) และมีจำนวนมาก Lyman ได้เสนอแนะว่าไม่ควรจะมีการกำหนดไว้ล่วงหน้าเลยว่าจะต้องมีใครได้เกรดอะไรเป็นจำนวนเท่าไรแต่ควรจะพิจารณาถึงผลงานที่ทำมากกว่า

4. Normalized Standard Scores (Area Transformation)

คะแนน normalized standard scores เป็นคะแนนแปลงรูปชนิดหนึ่งที่ถือว่าเป็นคะแนนมาตรฐาน แต่มีวิธีการคำนวณแบบ Percentile rank การใช้ linear standard score นั้น ลักษณะการกระจายของคะแนนดิบจะไม่เปลี่ยนแปลงซึ่งถ้าเราลากรูป Polygon จากฐานขึ้นไป เราจะเห็นว่าค่าต่าง ๆ ของ linear standard score จะอยู่ตามแนวเส้นตรงตามค่าของคะแนนดิบซึ่งมีรูปร่างการกระจายตามที่เป็นจริง แต่คะแนนของ normalized standard score นั้นลักษณะการกระจายของคะแนนดังกล่าวจะเป็นจริงได้ก็ต่อเมื่อการกระจายนั้นเป็นโค้งปกติเท่านั้น ซึ่งจะสังเกตได้จากภาพประกอบที่ 9.1 คะแนน normalized standard scores นี้ดูจากชื่อก็จะเห็นได้ว่าจะต้องใช้กับคะแนนที่มีลักษณะการกระจายใกล้เคียงกับการกระจายแบบปกติ ซึ่งก็เป็นลักษณะทั่ว ๆ ไปของคะแนน normalized standard scores ทุกชนิดมีวิธีการคำนวณดังนี้

1. เขียนค่าของคะแนนดิบทุกค่าลงไป

2. หาค่าความถี่ของคะแนนแต่ละคะแนน (f)

3 หาค่าความถี่สะสม (cf)

4. หาค่าความถี่สะสมของจุดกลางของคะแนนขั้นสูงกว่า 1 ขั้น (cf + ½ f)

5. แปลงค่าความถี่สะสม (cf) เป็น cumulative percentage (จากขั้นตอนที่ 1-5 จะเห็นว่าคล้ายกับวิธีการหาค่า percentile rank)

6. แทนค่า normalized standard score โดยใช้ cumulative percentage เปรียบเทียบกับค่า normal proability distribution (โดยเปิดจากในตารางท้ายเล่ม) ค่านี้จะแตกต่างกันตามชนิดของ normalized standard score แต่ละชนิด

การแปลงคะแนนในรูปของพื้นที่ (Area Transformation) แบ่งได้ดังนี้

(a) T-scaled score คะแนนแบบนี้จะคล้ายกับ T-score ถ้าการกระจายของข้อมูลเป็นแบบโค้งปกติ (มีค่าเฉลี่ย = 50 ค่าความเบี่ยงเบนมาตรฐาน = 10) ในการปฏิบัติจริงเรามักใช้ T-score มากกว่า T-scaled score ทั้งนี้เพราะการใช้ T-scaled score นั้นจะต้องกำหนดการกระจายของคะแนนเป็นแบบโค้งปกติเท่านั้นซึ่งนับเป็นขีดจำกัดของคะแนนแบบนี้

(b) Stanine score คะแนนแบบนี้ยังพบเห็นใช้กันทั่วไป เป็นรูปแบบของคะแนนที่ดัดแปลงมาจากคะแนนที่นักจิตวิทยาในสมัยสงครามโลกครั้งที่ 2 ใช้ในกองทัพอากาศสหรัฐ เหตุผลในการใช้ stanine score ก็เพื่อให้สามารถใช้กับบัตรคอมพิวเตอร์ซึ่งจะใช้ตัวเลขในบัตรเพียง 1 ตำแหน่งเท่านั้น ค่า normalized standard scores รูปแบบอื่น ๆ ซึ่งมีค่าเฉพาะลงไปจึงไม่อาจจะเขียนใส่ลงในบัตรคอมพิวเตอร์ได้สะดวก คำว่า stanine จึงมาจากคำว่า standard+ score nine unit จึงนำมาใช้เพื่อแทนค่าต่าง ๆ เหล่านั้น โดยมีค่าตั้งแต่ 1-9 ค่าของ stanine แต่ละช่องจะมีค่าความเบี่ยงเบนมาตรฐานของ Z-score = $\frac{1}{2}$ และค่าเฉลี่ยของ Z-score จะอยู่ตรง stanine ที่ 5

นักจิตวิทยาทหารอากาศสหรัฐอเมริกาบางคนใช้ stanine เป็น linear standard score มาก่อน แล้วจึงแปลงเป็น normalized standard score และใช้มาจนถึงปัจจุบัน ถ้าข้อมูลเป็นโค้งปกติแล้ว stanine จะมีค่าเฉลี่ย = 5 และความเบี่ยงเบนมาตรฐาน = 2 ดังนั้น ค่า stanine ทุกค่ายกเว้น stanine ที่ 1 และ 9 จะมีความกว้างจากกัน $\frac{1}{2}$ ของความเบี่ยงเบนมาตรฐาน การหาค่า stanine ทำดังนี้

1. จัดเรียงลำดับคะแนนการสอบ

2. กำหนดค่า stanine ตามร้อยละของคะแนนดังนี้

	ต่ำสุด	ต่อไป	ต่อไป	ต่อไป	จุดกลาง	ต่อไป	ต่อไป	ต่อไป	สูงสุด
	4%	7%	12%	17%	20%	17%	12%	7%	4%
stanine	1	2	3	4	5	6	7	8	9

ค่า stanine นี้ผู้พิมพ์แบบทดสอบบางฉบับได้สนับสนุนให้ใช้ในการตีความหมายคะแนน ในแบบทดสอบมาตรฐานจะมีการพิมพ์ตารางเกณฑ์ปกติที่มีค่าคะแนนแปลงรูป แม้ในแบบทดสอบที่ครูสร้างขึ้นใช้ในห้องเรียนก็สามารจะทำคะแนนเป็น stanine ได้เช่นเดียวกันโดยใช้วิธีการที่กล่าวไว้ข้างต้นเราก็จะสามารถบอกได้ว่าคะแนนดิบแต่ละคะแนนจะมีค่า stanine เท่าไร ในแบบทดสอบของ Flanagan Aptitude Classification Tests ได้จำแนกค่า stanine แต่ละค่าออกไปอีก 3 ค่า โดยใช้เครื่องหมาย + และ – มาช่วย เช่น stanine ที่ 1 ก็จะเป็น 1–, 1, 1+ และทุก ๆ ค่าของ stanine อื่น ๆ ก็จะแตกออกเป็น 3 ค่าเช่นเดียวกัน ดังนั้นผลการสอบจากแบบทดสอบของ Flanagan จึงมีค่าแปลงรูปเป็น 27 ค่า ซึ่งจากการดัดแปลงของ Flanagan นี้มีข้อดีกว่าค่าของ stanine เดิม แต่ก็ไม่นิยมใช้มากนัก

(c) C-scaled score Guilford ได้ดัดแปลงคะแนนแบบ stanine มาเป็น C-scale โดยขยายค่า stanine ตรงปลายสุดออกไปทั้ง 2 ข้าง C-scale จึงมี 11 ค่า โดยเริ่มตั้งแต่ 0-10 คะแนน แบบนี้ใช้ในแบบทดสอบของ Sheridan Psychological Services การคำนวณค่า C-scores คล้ายของ stanine ยกเว้นค่าบางค่าดังแสดงต่อไปนี้

ต่ำสุด	ต่อไป	ต่อไป	ต่อไป	ต่อไป	จุดกลาง	ต่อไป	ต่อไป	ต่อไป	ต่อไป	สูงสุด
1%	3%	7%	12%	17%	20%	17%	12%	7%	3%	1%
C = 0	C = 1	C = 2	C = 3	C = 4	C = 5	C = 6	C = 7	C = 8	C = 9	C = 10

(d) Sten score หลักการของ sten score ก็คล้ายกับของ stanine และ C-scaled score sten score เป็นคะแนนที่ย่อมาจากคำว่า a normalized standard score with Ten units การหา sten score ใช้วิธีการกำหนดจุดกลางแล้วกระจายออกไปทางด้านคะแนนต่ำและคะแนนสูง กล่าวคือเฉลี่ยในแต่ละช่วงคะแนนจะห่างกัน $\frac{1}{2}$ ความเบี่ยงเบนมาตรฐาน ยกเว้น sten ที่ 1 และ 10 ซึ่งเป็นปลายเปิด เนื่องจากคะแนนเหล่านี้เป็น normalized standard score ดังนั้นช่วงห่างของแต่ละคะแนนจึงเป็นการกระจายแบบปกติ คะแนนแบบนี้ใช้ในแบบทดสอบของ Cattell (Institue of Personality and Ability Testing) การคำนวณค่าของ Sten คล้ายวิธีของ Stanine ยกเว้นค่าบางค่าดังแสดงในตารางข้างล่างนี้

ต่ำสุด	ต่อไป	ต่อไป	ต่อไป	ต่ำกลาง	สูงกลาง	ต่อไป	ต่อไป	ต่อไป	สูงสุด
2%	5%	7%	15%	19%	19%	15%	9%	5%	2%
Sten	Sten	Sten	Sten	Sten	Sten	Sten	Sten	Sten	Sten
1	2	3	4	5	6	7	a	9	10

(e) Deviation IQs (อีกแบบหนึ่ง) คะแนนแบบนี้เป็นคะแนนการวัด I.Q. อีกแบบหนึ่งที่มีข้อดีและข้อจำกัดคล้ายกับ T-scaled score แต่มีค่าเฉลี่ย = 100 ค่าความเบี่ยงเบนมาตรฐานขึ้นอยู่กับผู้ออกข้อสอบจะกำหนด ซึ่งก็คล้ายกับคะแนน I.Q. แบบ II A(5) ที่ได้กล่าวมาแล้วข้างต้น ข้อสอบวัดเชาว์ปัญญาของ Wechsler ใช้คะแนนแบบ II A(5) แต่ถ้าแยกเป็น scale ออกมาในแต่ละฉบับแล้ว จะใช้ normalized standard score ซึ่งมีค่าเฉลี่ย = 10 และค่าความเบี่ยงเบนมาตรฐาน = 3 ดังนั้นการหา Deviation IQs แบบ normalized standard score จึงใช้เป็นวิธีการเบื้องต้นเพื่อหาค่า I.Q. ของคะแนนรวมซึ่งใช้วิธีของ Type II A(5) แต่ในตัวคะแนนของมันเองแล้วไม่นิยมใช้ในการแปลความหมาย

(f) ITED-score (Iowa Test of Educational Development) เป็นคะแนนที่ใช้กับแบบทดสอบของ Iowa และในปัจจุบันก็นำมาใช้กับแบบทดสอบ National Merit Scholarship Qualifying Examination รวมทั้งแบบทดสอบอื่น ๆ ด้วย คะแนนแบบนี้มีค่าเฉลี่ย = 15 ค่าความเบี่ยงเบนมาตรฐาน = 5 กลุ่มตัวอย่างนำมาจากนักเรียนทั่วสหรัฐอเมริกาในระดับ 10-11

5. Decile Rank

Cattell เป็นผู้นำคะแนนแบบนี้มาใช้กับกลุ่มเกณฑ์ปกติของเขา โดยปกติแล้ว decile ก็คือจุดใดจุดหนึ่งใน 9 จุดที่กระจายอยู่ตามลักษณะการกระจายความถี่ของ 10 กลุ่มที่มีจำนวนเท่า ๆ กัน ดังนั้น Decile แรก (D_1) จึงมีค่าเท่ากับเปอร์เซ็นต์ไทล์ ที่ 10 D_2 = เปอร์เซ็นต์ไทล์ที่ 20 ฯลฯ ต่อมา Cattell ได้เปลี่ยนแปลงแก้ไขใหม่โดยคิดจากช่วงห่าง 10% ของกลุ่ม และ 5% ที่กระจายออกไปในแต่ละข้างของจุด Decile เช่น Cattell Decile's 1. คือ ตั้งแต่ 5-15 เปอร์เซ็นต์ไทล์ (ถ้าต่ำกว่า P_5 มีค่า Decile = 0 และถ้าเกิน P_{95} มีค่า decile = 10) Cattell เชื่อว่าคะแนน Decile นี้สามารถใช้แทน percentile rank ได้ถ้าคะแนนดิบมีช่วงห่างกันมาก ๆ ดังนั้นเพื่อไม่ให้สับสนกันระหว่าง decile กับ decile score จึงเรียกคะแนนแบบนี้ว่า decile ranks

Type II C : Interindividual Comparison Considering Range

Percent Placement Score เป็นคะแนนแปลงรูปชนิดเดียวที่ขึ้นอยู่กับการพิจารณาเปรียบเทียบระหว่างผู้เข้าสอบแต่ละคนโดยใช้ช่วงห่างของคะแนนดิบ คะแนนแบบนี้ไม่ค่อยมีผู้นิยมใช้มากนัก แต่ในชั้นเรียนอาจนำมาใช้ได้ คะแนนแบบนี้เป็นการบอกให้ทราบถึงตำแหน่งของคนใน 101 จุด ซึ่งมีคะแนนสูงสุด 100 และต่ำสุด = 0

สูตร

$$X_{\%pl.} = \frac{100\ (X-L)}{(H-L)}$$

เมื่อ X = คะแนนดิบที่เด็กทำได้

H = คะแนนสูงสุดในกลุ่ม

L = คะแนนต่ำสุดในกลุ่ม

ตัวอย่าง มีแบบทดสอบฉบับหนึ่ง 300 ข้อ ผู้สอบทำคะแนนได้ตั้งแต่ 260-60 คะแนน ถ้าศรีนวลทำคะแนนได้ 60 คะแนน จะได้ Percent Placement Scores

$$X_{\%pl.} = \frac{100\ (60 - 60)}{260 - 60}$$

$$= 0$$

ถ้าศรีวรรณทำคะแนนได้ 260

$$X_{\%pl.} = \frac{100\ (260 - 60)}{260 - 60}$$

$$= 100$$

ถ้าศรีสุวรรณทำคะแนนได้ 140

$$X_{\%pl.} = \frac{100\ (140\text{-}60)}{260\text{-}60}$$

$$= 40$$

Type II D Interindividual Comparison Considering Status of those Making Same Scores

คะแนนชนิดนี้ มี 2 ชนิด คือ Age scores และ Grade Placement score ซึ่งคะแนนทั้ง 2 ชนิดนี้ใช้บรรยายคุณลักษณะของผู้เข้าสอบแต่ละคนโดยเปรียบเทียบกับค่าเฉลี่ยของกลุ่มซึ่งจะแตกต่างกันไปตามสถานภาพของกลุ่ม (เช่นอายุปฏิทินชั้นเรียน) คะแนนชนิดนี้จึงไม่ได้เป็นการเปรียบเทียบคะแนนของนักเรียนแต่ละคนกับกลุ่ม แต่เป็นคะแนนที่บอกให้รู้ว่าผู้เข้าสอบแต่ละคนมีคุณลักษณะคล้ายคลึงกับกลุ่มใด ๆ มากที่สุด คะแนนแบบนี้นิยมใช้กันมากกับแบบทดสอบมาตรฐานที่วัดผลสัมฤทธิ์และสติปัญญาของเด็กที่กำลังอยู่ในวัยเรียน แต่ถ้าเป็นผู้เข้าสอบที่นอกเหนือไปจากนี้แล้วไม่นิยมใช้ ทั้ง ๆ ที่การทำความเข้าใจกับคะแนนแบบนี้ค่อนข้างจะง่ายกว่าแบบอื่น ๆ แต่เนื่องจากมีข้อจำกัดในการใช้ดังกล่าวจึงต้องระมัดระวังการเลือกใช้ดังจะกล่าวในรายละเอียดต่อไปนี้

1. Age scores

คะแนนแบบนี้มักใช้กับแบบทดสอบทางสติปัญญา หรือแบบทดสอบวัดผลสัมฤทธิ์ของ

นักเรียนที่อยู่ในวัยเรียนหรือต่ำกว่านั้น Age scores ที่รู้จักกันมากที่สุดคือ Mental Age (MA) ซึ่ง Alfred Binet เป็นผู้นำมาใช้เมื่อ 80 ปีมาแล้ว

Age scores นี้ใช้อธิบายผลที่ได้จากการทดสอบในด้านคุณลักษณะของค่าเฉลี่ยของเด็กที่อยู่ในวัยต่าง ๆ กัน เช่น เด็กคนหนึ่งสอบด้วยแบบทดสอบวัดสติปัญญา มีอายุสมองเท่ากับ 7 ปี 6 เดือน (7.6) หมายความว่าเด็กคนนั้นมีความสามารถเท่ากับคะแนนเฉลี่ยของเด็กที่มีอายุประมาณ 7 ปี 6 เดือน การนำ Age scores มาใช้กับเด็กเล็ก ๆ จะช่วยทำให้เข้าใจง่ายขึ้น แต่ก็จะต้องระมัดระวังการตีความหมายให้อยู่ในขอบเขต เช่น เด็กคนหนึ่งอายุ 5 ปี ทดสอบด้วยแบบทดสอบได้อายุสมอง 7 ปี ผลการทดสอบนี้ไม่ได้หมายความว่าเด็กทุกคนที่มีอายุ 5 ปี จะต้องมีอายุสมอง 7 ปีทุกคนไป การสร้างแบบทดสอบเพื่อหาคะแนนชนิดนี้มีความยุ่งยากในการเลือกกลุ่มตัวอย่างที่จะนำมาทำ Age Norm ทั้งนี้เพราะเด็กบางคนอาจเรียนเก่งหรือเรียนช้ากว่าชั้นที่เรียนหรือกว่าอายุจริงของเด็ก ซึ่งการเก็บรวบรวมข้อมูลจากกลุ่มตัวอย่างเหล่านี้ทำได้ยาก เด็กบางคนอาจอยู่ต่างโรงเรียนกันเมื่อจะต้องเอามารวมกันจึงไม่ใช่เรื่องง่าย เด็กบางคนที่ฉลาดมาก ๆ อาจจะอยู่ในชั้นที่เหนือกว่าอายุจริง 2-3 ชั้น ในขณะที่เด็กอ่อนอาจต้องเรียนในห้องพิเศษของโรงเรียนอีกโรงเรียนหนึ่งทั้ง ๆ ที่อายุเท่ากัน การนำ Age scores มาใช้ก็ควรจะใช้ควบคู่กับอายุจริงหรือการวัดอื่น ๆ จึงจะทำให้การตีความหมายคะแนนสมบูรณ์ขึ้น Age scores แบ่งเป็น

(a) Mental Age เป็นการหาคะแนนตามแบบของ Binet คะแนน MA นี้ยังอาจนำมาใช้แปลความหมายอื่น ๆ ได้อีก หน่วยของ MA แต่ละหน่วยเป็นเครื่องแสดงว่าเด็กสามารถทำงานชิ้นนั้น ๆ ได้ MA จึงมีคุณสมบัติเท่ากับส่วนเฉลี่ยของอายุปฏิทินที่เด็กที่เป็นกลุ่มตัวอย่างสามารถทำได้ในการสอบวัดทางเชาว์ปัญญานั้น ค่าของ MA อาจหาได้โดยการนับจำนวนข้อที่เด็กตอบถูก แล้วนำมาหาค่าเฉลี่ยของเด็กในอายุนั้น ๆ ทำงานชิ้นนั้น ๆ ได้การหา MA ได้นั้นยังสามารถนำมาใช้ในการคำนวณหาค่า I.Q. ได้

ปัญหาของการใช้ MA ก็คือการหา MA ของผู้ใหญ่ ถึงแม้ว่าในปัจจุบันจะมีการศึกษาพบว่าเชาว์ปัญญาของคนจะพัฒนาไปเรื่อย ๆ จนถึงวัยกลางคนก็ตาม แต่มักจะปรากฏเสมอว่าความแตกต่างในการทำงานชั้นหนึ่ง ๆ ของคนที่มีอายุในวัยเด็กจะเห็นได้ชัดเจนกว่าคนที่อยู่ในวัยผู้ใหญ่ เช่น ความแตกต่างของเด็กอายุ 6-7 ปีจะมีมากกว่าเด็กอายุ 16-17 ปี หรืออายุ 26-27 ปี ดังนั้นการตีความหมาย MA ของผู้ใหญ่จึงต้องระมัดระวังให้มาก ข้อยุ่งยากอีกประการหนึ่งก็คือการตีความหมายของ MA โดยพิจารณาจากค่าความเบี่ยงเบนมาตรฐาน เรามักจะพบว่าค่า S.D. ระหว่างแบบทดสอบแต่ละฉบับ หรือระหว่างอายุหนึ่งทั้งถึงอีกอายุหนึ่งจะมีช่วงห่าง

ไม่เท่ากัน เช่น MA 13-3 กับ MA 12-3 นั้น ช่วงห่างของความเบี่ยงเบนมาตรฐานจะมีความแตกต่างจาก MA 6-3 และ 5-3

(b) Educational Age คล้ายกับ MA มาก แต่นิยมใช้ในวิชาต่าง ๆ ที่เรียนในโรงเรียน บางทีก็เรียกว่า Achievement Age, Reading Age หรือชื่อวิชาต่าง ๆ แล้วเติมด้วย Age ข้อจำกัดของ EA ก็จะคล้าย ๆ กับ MA

2. Grade Placement Scores

หลักการในการใช้คะแนนแบบนี้คล้ายกับ Age scores เพราะค่าที่ได้ก็คือค่าเฉลี่ยของคะแนนนักเรียนในแต่ละระดับชั้น มีวิธีการหาคะแนนแบบนี้ดังนี้

1. นำแบบทดสอบไปทดสอบกับนักเรียนหลาย ๆ คนในชั้นต่าง ๆ กัน โดยใช้แบบทดสอบเดียวกัน

2. หาค่าคะแนนเฉลี่ย (อาจเป็นค่าเฉลี่ยหรือค่ามัธยฐาน) ของผู้สอบในแต่ละชั้น

3. นำคะแนนจากแบบทดสอบ plot ลงบนแผ่นกระดาษกราฟ พยายามต่อจุดเหล่านั้นให้เป็นลักษณะเส้นตรง

4. ขยายเส้นตรงตามข้อ 3 ให้คลุมไปถึงคะแนนที่อยู่สูงหรือต่ำกว่าค่าเฉลี่ย

5. อ่านค่า grade equivalent ที่ใกล้ที่สุด ให้ตรงกับค่าคะแนนดิบแต่ละค่า

6. พิมพ์ค่าเหล่านั้นใส่ในตาราง

คะแนนแบบนี้มักใช้กับนักเรียนที่อยู่ในโรงเรียนมาแล้ว 10 ปี ดังนั้น คะแนน 8.2 ก็แปลว่ามีความสามารถเท่ากับอยู่ในโรงเรียนมาแล้ว 8 ปี 2 เดือน คะแนนแบบนี้ยังแบ่งออกเป็น

(a) Full-population Grade Placement Scores เป็นคะแนนที่ทำมาจากกลุ่มตัวอย่างทั้งหมดในชั้นที่เอามาทำเกณฑ์ปกติ มักจะมีค่าความเบี่ยงเบนมาตรฐานกว้างมาก เมื่อนำคะแนนดิบมาเทียบจึงทำให้มีค่า grade placement ค่อนข้างต่ำ คนไม่นิยมใช้

(b) Modal Age Grade Placement Scores เป็นวิธีการหาคะแนนโดยเอานักเรียนที่มีอายุต่ำมาก ๆ และสูงมาก ๆ ออกไป นั่นก็คือช่วยทำให้กลุ่มที่จะทดสอบมีขอบเขตจำกัดมากขึ้น ผลก็คือทำให้การเปรียบเทียบสะดวกและมีความหมายมากขึ้น

(c) Modal Age and Modal Intelligence Grade Placement Scores เป็นการหา grade placement ที่ทำเฉพาะกับนักเรียนที่มีเชาว์ปัญญาและระดับอายุใกล้เคียงกัน โดยให้เหตุผลว่านักเรียนที่มีอายุหรือเชาว์ปัญญาสูงมาก ๆ และต่ำมาก ๆ จะทำให้เกิดการสมดุลกันขึ้นเอง

(d) Anticipated Achievement Grade Placement Scores (AAGPs) คะแนนแบบนี้ทำจากแบบทดสอบแเป็นชุด ๆ และกลุ่มเกณฑ์ปกติก็อยู่ในระดับชั้นและอายุตามที่กำหนดไว้ คะแนน

แบบนี้ต่างจากคะแนนอื่น ๆ ตรงที่ว่าคะแนนที่ได้เป็นคะแนนที่คาดหวังซึ่งต้องใช้วิธีการอธิบายเป็นกรณีพิเศษเพราะไม่ได้นำมาทำนายเหมือนกับการหาค่าสหสัมพันธ์ แต่เป็นเพียงคะแนนที่คาดหวังจากอายุสมองและชั้นเรียนที่เด็กเรียนอยู่ เช่น นักเรียนคนหนึ่ง ได้คะแนน grade placement ที่ 6 เมื่อนำมาเปรียบเทียบกับคะแนน AAGPs แล้ว เขาควรจะได้คะแนนสูงกว่าหรือต่ำกว่าระดับ 6 หรือไม่

(e) Mental Age Grade Placement Scores คะแนนแบบนี้ใช้สำหรับรายงานผลการทดสอบทางเชาว์ปัญญาของนักเรียน โดยการนำอายุสมองของเด็กแต่ละคนแปลงเป็น grade placement score ซึ่งมีอายุสมองเป็นค่าเฉลี่ย ประโยชน์จากคะแนนแบบนี้คือจะต้องเขียนออกมาในรูปของ Grade Placement unit จะทำให้คนคุ้นเคยมากกว่าเขียนเป็น Mental Age

Type III Intra-Interindividual Comparison

คะแนนแปลงรูปแบบนี้เป็นการวัดผล 2 อย่างออกมาจากคน ๆ เดียวกัน โดยนำคะแนนนั้นมาทำเป็นอัตราส่วน ตามรูปแบบของ A และ B ใช้กันทั่ว ๆ ไป ส่วนแบบ C และ D นั้นไม่ค่อยมีชื่อเสียงมากนัก

Type III A Ratio IQ's

ในตอนต้นได้กล่าวถึงเรื่อง I.Q. ไว้แล้ว 2 ครั้ง แต่ในครั้งนี้จัดว่าเป็น I.Q. ต้นฉบับ ในรูปแบบของ I.Q. ที่กล่าวถึงตอนแรกนั้น Stern เป็นคนเสนอและ Terman นำมาใช้เมื่อประมาณ 60 ปีมาแล้ว ส่วน I.Q. ในรูปแบบนี้คิดมาจากสูตร

$$\text{I.Q.} = \frac{\text{MA}}{\text{CA}} \times 100$$

เมื่อ MA = อายุสมองหาได้จากการทำแบบทดสอบเชาว์ปัญญา

CA = อายุจริงในวันที่ทำการทดสอบเชาว์ปัญญา

คะแนนแบบนี้เป็นที่เข้าใจกันแพร่หลายทั่วไป แต่ก็มีข้อจำกัดบางประการที่จะทำให้คนเข้าใจคะแนนแบบนี้ไม่ค่อยถูกต้องนัก เช่น มีข้อตกลงเบื้องต้นของการใช้คะแนนแบบนี้ว่าหน่วยของ MA นั้นจะต้องมีขนาดเท่ากันซึ่งอาจจะทำได้ยาก และการใช้ I.Q. จะมีผลกับคนที่อายุระหว่าง 5-15 ปี ส่วน I.Q. ของผู้ใหญ่นั้นจะต้องขึ้นอยู่กับวิธีการวัดอายุสมองซึ่งจะต้องใช้ศิลปมากเช่นเดียวกับการปรับให้เข้ากับอายุจริงของผู้สอบ ข้อถกเถียงเกี่ยวกับอัตราส่วน I.Q. ประการหนึ่งก็คือ ค่าความเบี่ยงเบนมาตรฐาน ซึ่งจะแตกต่างไปตามระดับอายุในแต่ละช่วง ถ้าเรายอมให้ค่าความเบี่ยงเบนมาตรฐานเปลี่ยนไปได้เรื่อย ๆ ค่า I.Q. จะแสดงให้เห็นความ

เด่นและความด้อยในระดับอายุต่าง ๆ กัน ดังนั้นการใช้ diviation I.Q. ตามแบบ Type II A และ B จึงใช้ดีกว่า Ratio I.Q.

Type III B Intellectual Status Index

คะแนนชนิดนี้นำมาใช้โดย California Test Bureau เพื่อใช้กับแบบทดสอบ California Test of Mental Maturity in the Intellectual Status Index รูปของคะแนนก็คือ I.Q. ชนิดหนึ่งเพียงแต่เปลี่ยนตัวหารจากอายุจริงในปฏิทินของเด็กแต่ละคน มาเป็นอายุจริงเฉลี่ยของนักเรียนในชั้นเดียวกัน คะแนนแบบนี้จึงขึ้นอยู่กับคะแนนการทำแบบทดสอบวัดเชาว์ปัญญาในแต่ละระดับชั้นเรียนมากกว่าจะขึ้นอยู่กับอายุจริงของเด็ก

Type III C. Educational Quotients

คะแนนชนิดนี้หาได้จาก

$$EQ = \frac{EA}{CA} \times 100$$

เมื่อ EA คืออายุตามระดับการศึกษา

CA คืออายุจริง

การหา EQ นี้มีข้อดี และข้อจำกัดคล้ายกับ ratio IQs เว้นแต่ว่าในการคำนวณของ EQ นั้น จำกัด ค่า CA ตามแบบทดสอบวัดผลสัมฤทธิ์ในระดับชั้นเรียน ซึ่งของ ratio I.Q. จะคิด CA จากอายุจริง ๆ ของเด็ก นอกจากนั้นข้อจำกัดในการหา EA นั้นเราไม่อาจนำคะแนนวิชาแต่ละวิชามาเปรียบเทียบกันตรง ๆ ได้ นอกจากจะทำให้คะแนนของนักเรียนเป็นกลุ่มเดียวกันก่อน จึงไม่มีผู้นิยมใช้

Type III D Accomplishment Quotients

ไม่ค่อยมีผู้นิยมใช้คะแนนแบบนี้ซึ่งเนื่องมาจากจะต้องใช้คะแนนถึง 2 ชุดมาเปรียบเทียบกัน เมื่อมีการวัดผลมากขึ้นความผิดพลาดในการวัดก็จะมีมากขึ้น คะแนนแบบนี้เป็นการเปรียบเทียบผลสัมฤทธิ์ในการสอบกับเชาว์ปัญญา

สูตร

$$AQ = \frac{EA}{MA} \times 100$$

EA คือ อายุทางการศึกษาพิจารณาจากผลการสอบแบบทดสอบผลสัมฤทธิ์

MA คือ อายุสมองพิจารณาจากแบบทดสอบวัดเชาว์ปัญญา

Type IV Assorted Arbitrary Bases

คะแนนแบบนี้มี 3 รูปแบบที่ใช้อธิบายคะแนนการสอบที่ยังใช้กันอยู่ดังนี้

Type IV A : Nonmeaningful Scaled Scores

มีผู้พิมพ์แบบทดสอบหลาย ๆ คนที่ยังนิยมใช้คะแนนที่ไม่มีความหมายในตัวของมันเอง แต่สามารถนำคะแนนเหล่านี้มาใช้เป็นพื้นฐานในการแบ่งแบบทดสอบออกเป็นหลาย ๆ ฉบับ เพื่อให้แบบทดสอบย่อย ๆ เหล่านั้นมีระดับเท่า ๆ กัน จำนวนเท่า ๆ กัน ตัวอย่างเช่นคะแนนจาก CEEB ก็เป็นคะแนนที่มีรากฐานมาจาก linear standard score ซึ่งมีค่าเฉลี่ย 500 ค่าความเบี่ยงเบนมาตรฐาน = 100 ซึ่งต่อมาอีก 30 ปีหลังจากนั้นก็ไม่มีผู้ใช้คะแนนแบบนี้แต่มาใช้ percentile rank แทน

แบบทดสอบ SCAT (School and College Abilities Tests) ซึ่ง E.T.S. เป็นผู้ปรับปรุงก็เป็นแบบทดสอบอีกชนิดหนึ่งที่มีการคิดคะแนนแบบนี้ โดยกำหนดให้คะแนน 300 มีค่าเท่ากับ percentage correct score = 60 และให้ scaled score 260 มีค่าเท่ากับ percentage correct score = 20 ดังนั้น scaled score จึงมีประโยชน์ในทางสถิติแก่ผู้สร้าง ส่วน percentile จึงมีประโยชน์ในการตีความหมาย

Type IV B : Longe-Range Equi-Unit Scores

คะแนนต่าง ๆ ที่กล่าวมาข้างต้นไม่มี scale ของหน่วยเท่ากันทุกหน่วย นอกจากพิสัยคะแนนจะแคบหรือมีข้อตกลงเบื้องต้นที่แน่นอนแล้ว แต่บางครั้งเราก็ต้องการให้ scale 1 หน่วยสามารถครอบคลุมอายุออกไปกว้าง ๆ ด้วย คะแนนแบบนี้จึงมักใช้ประโยชน์ในการทำวิจัย เพราะคะแนนเหล่านี้ไม่สามารถจะนำมาตีความหมายโดยตัวของมันเองได้ดีพอ แต่จะต้องใช้หลักของเหตุผลมาใช้ตีความหมายที่ยุ่งยากซับซ้อนให้

Type IV C : Deviation IQ's (Otis-style)

มีสูตรคำนวณว่า

$$IQ = 100 + (X - \bar{X}_{age\ norm})$$

เมื่อ X = คะแนนที่สอบจากแบบทดสอบวัดเชาว์ปัญญาของ Otis

$\bar{X}_{age\ nom}$ = ส่วนเฉลี่ยของคะแนนดิบในกลุ่มทดสอบที่มีอายุปฏิทินเท่ากับผู้ทดสอบนั้น

Deviation IQ แบบนี้มีค่าเฉลี่ย = 100 ส่วนค่าความเบี่ยงเบนมาตรฐานไม่อาจทำให้เหมือนกับ Type II ได้ เพราะความเบี่ยงเบนมาตรฐานของ Otis อาจแปรไปตามอายุ ซึ่งทำให้การตีความหมายทำได้ลำบาก จึงไม่มีผู้นิยมใช้

สรุปท้ายบท

1. การใช้คะแนนแปลงรูปก็เพื่อนำคะแนนมาเปรียบเทียบกันจะทำให้การตีความหมายของคะแนนมีความหมายมากขึ้น

2. คะแนนแปลงรูปมี 4 รูปแบบ โดยแบ่งตามที่มาของคะแนนและการนำผลไปใช้

3. รูปแบบที่ 1 เป็นการนำคะแนนไปเปรียบเทียบกับมาตรฐานหรือความยากของเนื้อหาวิชาซึ่งอาจทำในรูปของร้อยละ และให้เป็นเกรด

4. รูปแบบที่ 2 เป็นการนำคะแนนไปเปรียบเทียบกันในกลุ่ม ส่วนใหญ่จะใช้กับแบบทดสอบมาตรฐาน ซึ่งสามารถจะพิจารณาโดยใช้ค่าเฉลี่ยและค่าความเบี่ยงเบนมาตรฐานเป็นเกณฑ์ เช่น Z-score, T-score, AGCT score, CEEB score, Deviation IQ ส่วนการพิจารณาโดยนำคะแนนแต่ละคนมาเปรียบเทียบกันโดยใช้ตำแหน่งในกลุ่มก็ได้แก่ การจัดลำดับตำแหน่งเปอร์เซ็นต์ไทล์ การให้เกรดโดยกำหนดจากกลุ่ม Normalized Standard Score แบบที่พิจารณาในด้านของ area transformation ที่นิยมใช้ได้แก่ Stanine, C-Scaled score, และ Decile rank อีกรูปหนึ่งเป็นการพิจารณาเปรียบเทียบภายในกลุ่มโดยดูจากพิสัยของคะแนนเป็นหลักได้แก่ Grade Placement Score นอกจากนั้นยังมีคะแนนที่เปรียบเทียบในกลุ่มโดยดูจากคุณลักษณะของกลุ่ม เช่น Age scores, Grade placement score

3. รูปแบบที่ 3 เป็นการนำคะแนนที่วัดจากคน ๆ เดียวกัน แต่วัด 2 อย่าง มาเปรียบเทียบกันในรูปของอัตราส่วน เช่น Ratio IQ's Intellectual Status Index, Educational Quotient และ Accomplishment Quotients แต่ใน 2 อย่างหลังนี้ไม่ค่อยมีผู้ใช้

4. รูปแบบที่ 4 เป็นคะแนนที่คะแนนอื่น ๆ นำมาใช้เป็นพื้นฐานในการตีความหมายอื่น ๆ ซึ่งในตัวของมันเองแล้วจะไม่มีความหมายอะไร

5. ข้อพึงระมัดระวังในการตีความหมายคะแนนแปลงรูปจะต้องพิจารณาคู่กันไปกับคุณลักษณะของแบบทดสอบด้วย ถ้าแบบทดสอบไม่ดีจะทำให้การตีความหมายคะแนนผิดพลาดด้วย

แบบฝึกหัดท้ายบท

1. จงอธิบายเหตุผลที่ต้องมีการแปลงคะแนน
2. แบบทดสอบฉบับหนึ่งมี 150 ข้อ สอบกับนักเรียน 50 คน ได้คะแนนเฉลี่ย 100 คะแนน และค่าความเบี่ยงเบนมาตรฐาน 15 คะแนน ถ้านักเรียนคนหนึ่งสอบได้ 90 คะแนน จงหาคะแนนแปลงรูปที่ท่านคิดว่าจะต้องหา
3. คะแนนแปลงรูปแบบที่ 1 จะสามารถนำมาใช้ในกรณีใดได้บ้าง จงอธิบายพร้อมทั้งยกตัวอย่างประกอบ
4. คะแนนแปลงรูปแบบที่ 3 มีข้อดีและข้อจำกัดในการใช้อย่างไรบ้าง
5. คะแนนแปลงรูปแบบ linear standard score กับ normalized standard score มีข้อแตกต่างกันอย่างไรบ้าง จงเปรียบเทียบ
6. จงอธิบายความแตกต่างของ Deviation IQ ในแต่ละรูปมาให้เข้าใจ