

“ผมสอบวิชาเลขได้ 80 คะแนนครับ”

ถามว่าเด็กคนนั้นเก่งหรืออ่อน ไม่มีใครตอบ คำถามนี้ได้เลยถ้าไม่มีสิ่งเปรียบเทียบ ในปัจจุบันนี้ เกณฑ์เปรียบเทียบที่ใช้กันมากได้แก่

1. คะแนนเต็ม
2. 50% ของคะแนนเต็ม
3. กลุ่มนักเรียนในชั้นนั้น ๆ

ดังนั้น เพื่อที่จะตอบคำถามข้างบนนี้ เราจะ ต้องถามว่า คะแนนเต็มเท่าไร (ตามเกณฑ์ข้อ 1) ถ้าเขาตอบว่าคะแนนเต็ม 100 เราก็พอจะประเมิน ได้ว่าเขา “เก่ง” เพราะได้เกือบเต็ม และแปลว่า “สอบได้” เพราะได้เกิน 50% ของคะแนนเต็ม (ตามข้อ 2) แต่ถ้าถามต่อไปว่าคนสูงสุดและต่ำ สุดได้เท่าไรก็ได้คำตอบว่า 90 กับ 70 ตามลำดับ (เปรียบเทียบกับนักเรียนในชั้นนั้นตามข้อ 3) ก็ แปลว่าเด็กคนนั้นเป็นพวกปานกลาง ไม่ได้เก่ง ะไร แต่ถ้าสูงสุดได้ 95 ต่ำสุดได้ 80 ก็แปลว่า เขาเป็นพวกอ่อน และยิ่งกว่านั้น ยังชี้ให้เห็นว่า ข้อสอบนั้นไม่ดี เพราะเด็กทั้งชั้นได้คะแนนใกล้เคียงกัน มี Range เพียง 15 คะแนน (95-80= 15) เท่านั้น ข้อสอบจำแนกเด็กเก่งเด็กอ่อนให้

เด่นชัดไม่ได้ เด็กจึงได้คะแนนเกาะกลุ่มกัน หรือ อาจจะแปลว่าเด็กกลุ่มนั้นเก่งพอ ๆ กัน จึงทำ ให้ได้คะแนนเกาะกลุ่มกัน ถ้าจะแปลตามกรณีหลัง นี้ต้องพิจารณาดูข้อสอบว่า “ยากหรือง่าย” ถ้า ข้อสอบยากและเด็กยังได้คะแนนเกาะกลุ่ม กันเช่น นี้ก็แปลว่าเด็กเก่งพอ ๆ กันได้ แต่ถ้าข้อสอบ ง่ายก็ไม่เห็นว่าจะเก่งได้อย่างไร เพียงแต่แปลว่า เด็กมีความสามารถพอๆ กัน จะเก่งหรืออ่อนยัง ตอบไม่ได้ การแปลคะแนนที่จะให้ความหมาย จริง ๆ โดยไม่คำนึงถึงว่าข้อสอบจะยากหรือง่าย ควรเปรียบเทียบกับ “ตำแหน่งที่เด็กอยู่ในชั้น นั้น ๆ” จึงตารางสมมติต่อไปนี้

คะแนนดิบ (x)	ความถี่ (f ₁)	ความถี่ (f ₂)
55	1	1
50	2	1
45	4	2
40	7	2
35	10	3
30	20	6
25	10	12
20	8	30
15	5	10
10	2	2
5	1	1
	70	70

ขอให้พิจารณาคณะที่ได้คะแนนดิบ 25 เขา
งกว่าเพื่อน ๆ เพียง 16 คน (ความถี่ 1) แต่
ดูในคอลัมน์ความถี่ (2) จะเห็นว่าเขาเก่งกว่า
เพื่อนในชั้นถึง 43 คน ซึ่งจะเห็นว่า ถ้าดูเพียง
คะแนนดิบ 25 เราจะได้ความหมายของคะแนน
มีชัดเจนเท่ากับเมื่อดูความถี่ควบคู่กันไปด้วย

ตามที่กล่าวมาข้างต้นชี้ให้เห็นว่า คะแนนดิบ
ของความถี่ เป็นเครื่องช่วยในการพิจารณาประเมิน
ว่าใครเก่งหรือไม่เพียงไร โดยเฉพาะความถี่
ของแต่ละคะแนนนั้นมีความ ปรหลาดแผ่งอยู่ใน
วันนั้นเอง ถ้าผู้อ่านพิจารณาให้ละเอียดก็จะพบว่า
คนที่ได้คะแนนมาก (คนเก่ง) จะมีน้อยคน คน
อื่น ๆ ที่ได้คะแนนน้อย ๆ ก็มักจะมีน้อยคน
วนคนที่ได้คะแนนกลาง ๆ ซึ่งเป็นพวกปาน-
กลางมีมาก ถ้าเรานำความถี่เหล่านี้มาจากราฟจะ
เห็นว่า จะมีลักษณะตรงกลางโป่งออกคล้าย ๆ
ระฆังคว่ำ หรือระฆังคว่ำ แต่ถ้าข้อสอบยากไป
หรือกลุ่มนักเรียนมีคนที่เก่ง คนอ่อน ปะปนอยู่ ไม่ได้
มีส่วนที่จะเป็นตัวแทนของกลุ่มประชากร (popu-
lation) ได้จริง ๆ โคง หรือระฆังคว่ำนั้นก็จะมี
มีรูปร่างที่สวยงามแต่จะบิดเบี้ยว หรือทาง
ภาษาสถิติเรียกว่า “เบ้” ตั้งในคอลัมน์ 1 2
ลักษณะของการแจกแจงที่ หัวท้าย แผลบ ตรง กลาง
โป่งออกเหมือนเจดีย์ทรายนี้ จะพบจากคะแนนที่
ได้จาก การวัดความสามารถทางค่าน สมอง ของคน
งเช่นคะแนนที่ได้จากการสอบเป็นประจำ (ยก-

เว้นกลุ่มผู้สอบน้อย ๆ หรือข้อสอบไม่ดี) ปรากฏ
การณ์เช่นนี้ สอดคล้องกับการแจกแจงทางทฤษฎี
ทางคณิตศาสตร์ที่เรียกว่า “โค้งปรกติ” (Normal
Curve) ดังนั้น นักสถิติจึงนำการแจกแจงของ
คะแนนไป “ประกบ” เข้ากับโค้งปรกติทาง
คณิตศาสตร์ แล้วมีกฎเกณฑ์ของโค้งปรกติมาใช้
ในการคำนวณ และอธิบายปรากฏการณ์ธรรมชาติ
ซึ่งถ้าเงื่อนไขสอดคล้องกันก็จะใช้กันได้ดี และ
สามารถพยากรณ์ปรากฏการณ์ธรรมชาตินั้น ๆ ได้
อย่างแม่นยำ จนสามารถทำให้มนุษย์สามารถควบคุม
ธรรมชาติได้ดียิ่งขึ้นทุกขณะ อนึ่ง ผู้เขียนได้
กล่าวมาแล้วว่าการ แจกแจง ของคะแนน ที่วัดจาก
ความสามารถด้านสมองมักจะแจกแจง แบบหัวท้าย
แผลบตรงกลางโป่งเสมอ จนทำให้เราลงความเห็น
ได้ว่าถ้าเราสอบกันคนเป็นหมื่นเป็นแสนแล้ว การ
แจกแจงของ คะแนน จะใกล้เคียง แจกแจงแบบโค้ง
ปรกติเข้าทุกที (ข้อสอบต้องดีด้วย) นั่นคือ
Population ของความสามารถนั้น ๆ ของคน จะ
แจกแจงเป็นแบบ Normal Curve นั่นก็คือ
นักเรียนกลุ่ม 30-40 คน ในชั้นเราก็อาจถือได้ว่า
เป็นส่วนหนึ่งของ Population แม้ว่าการแจกแจง
ของคะแนนดังกล่าวจะเบ้หรือบิดเบี้ยว ไม่เป็นแบบ
Normal Curve อันเนื่องมาจากสาเหตุหลักการ
Random ก็ตามเราอาจจะปรับการแจกแจงนั้นให้
เป็นแบบคล้าย Normal Curve ได้ (วิธีการ
ดังกล่าวผู้อ่านจะหา อ่านได้จาก หนังสือการ วัดผล

และการประเมินผลของโสม เลี่ยมแก้ว หรือหนังสือสถิติทางการศึกษาและจิตวิทยา ได้ทั่วไป) หลังจากการปรับโค้งบุคคลเบียวให้เป็นแบบ Normal Curve ได้แล้วก็แปลงเปอร์เซ็นต์เหล่านั้น (Percentile) ให้เป็นคะแนนมาตรฐานต่าง ๆ เช่น Z-score, T-score หรือ Stanine Score เป็นต้น จากนั้น จึงใช้คะแนน มาตรฐาน เหล่านี้ ตัดเกรดเกรดที่ได้จะมีหลัก มีเกณฑ์ ที่เหมาะสมกับ โลกเรากำลังก้าวหน้าไปสู่ยุคอวกาศ หรือยุควิทยาศาสตร์ขึ้นบ้าง ถึงตารางต่อไป

คะแนนดิบ ความถี่ แปลงเป็นคะแนนมาตรฐาน T-ปรกติ

x	f1	f2	T1	T2
55	1	1	74	74
50	2	1	69	71
45	4	2	65	68
40	7	2	60	63
35	10	3	56	62
30	20	8	50	60
25	10	12	45	55
15	5	10	36	38
10	2	2	32	32
5	1	1	27	27

(วิธีแปลงคะแนนดิบให้เป็นคะแนน T-ปรกติ อาจหาอ่านได้จากหนังสือการวัดผลและการประเมินผลของ โสม เลี่ยมแก้ว, ชวาล แพร์ตกุล หรือสถิติทางการศึกษาทั่วไป)

วิธีการตัดเกรด

วิธีการตัดเกรดมีวิธีแปลก ๆ ออกไปหลายวิธี แต่ละวิธีก็อาศัยหลักวิชาต่างกันบ้าง คล้าย

กันบ้าง หรือบางรายตัดเกรดโดยไม่อาศัยหลักวิชาใดเลยก็มี ซึ่งก็ช่วยเหลืออะไรไม่ได้ถือว่าเป็นเคราะห์กรรมของผู้ถูกตัดเกรดก็แล้วกัน ในที่นี้ผู้เขียนจะเสนอวิธีง่าย ๆ หนึ่งวิธี วิธีการทำดังนี้

1. คะแนน T สูงสุดลบด้วย T ต่ำสุด ตามตารางข้างต้นก็จะได้ $74-27=47$ คะแนน 47 นี้เราถือเป็น Range หรือพิสัย ของคะแนน T ชุดนี้ (บางรายใช้ $(74-27)+1$ ซึ่งก็ได้ผลพอ ๆ กัน)

2. ทกลงใจว่าจะตัดกี่เกรด เกณฑ์ที่จะช่วยในการตกลงใจว่าจะตัดกี่เกรดนี้ อาจพิจารณาได้จาก

2.1 การสอบควรวินิจฉัยสอบยาก หรือง่ายถ้าข้อสอบยากแต่ยังได้คะแนนสูงๆ และ Range ต่างกันมาก เช่นคะแนนดิบสูงสุดได้ 55 ต่ำสุดได้ 5 ดังนั้นควรตัดเกรด 5 เกรด คือ A,B,C,D,E ถ้าสูงสุดได้ 55 ต่ำสุดได้ 30 อาจจะตัด A,B,C, D,C,D หรือ C,D,E ก็ได้ถ้าคะแนนเต็ม 100 และข้อสอบไม่ยากไม่ง่าย (จะรู้ได้อย่างไร) แล้วคนได้สูงสุดยังได้ 55 ต่ำสุด 5 อาจจะตัดเพียง B,C,D,E ก็ได้ นั่นคือเราแปลเอาว่าเด็ก กลุ่มนี้ถึงกลุ่มอ่อน แต่ถ้าข้อสอบยากก็ควรให้ A,B,C,D,E ตั้งนี้เป็นต้น ฉะนั้น ในตอนนั้นคะแนนดิบมีบทบาทช่วยเราในการลง A,B,C,D,E ได้อย่างสำคัญ

2.2 สังเกตจากการเรียนในชั้นเรียนชั้น ชักตามครู อาจารย์ สนใจทำการบ้านสม่ำเสมออะไรก็มักจะตอบได้ฉลาด ฉลาด เป็นเช่นนี้

กลุ่ม หรือว่าแตกต่างกันมาก คนยอศก็ยอศเอาจริง ๆ คนแยก็แยเอาจริง ๆ ชาติเรียนประจำ การงานและทะเล คั้งทำให้เราประเมินค่าได้ว่า กลุ่มนี้ควรมี A และ E แต่ทงนครู อาจารย์ต้องเข้าสอนทุกชั่วโมง ไม่ใช่ว่าครูไม่ค่อยสอนแล้วไปโยนความผิดให้ลูกศิษย์ หนาวลูกศิษย์ ไร่ ๓๓ ๕ ครั้งช่น อย่างนี้ไมควร

ในทั้นสมมติว่าเราตกลงใจตัด 5 เกรด คือ A,B,C,D,E

3. ให้เอาจำนวนเกรดไปหาร Range ของคะแนน T ปรกติในทั้นจะได้ $47/5=9.4$ ทีเสียว่า 9 ก็แล้วกัน

4. หน้า 9 ไปบวกกับคะแนน T ค่าสุด ทอยยช่นไปเรื่อย ๆ จะได้คั้งนี้

ช่วงคะแนน T	เกรด	f1	f2
64-74	A	7	4
55-63	B	17	23
46-54	C	20	30
37-45	D	18	10
27-36	E	8	3

* ฉลาดเคลื่อนบ้างเล็กน้อยเพราะปัดเศษทง

การประเมินค่า

ในระยที่เรว "กำหนดคะแนน" ให้แต่ละข้อเรตกลงกันเรียกว่า "การวัด" (Measurement) กระบวนการนี้จะติดต่อกันไปจนถึง ตัดคะแนน T ออกเป็นช่วง ๆ ส่วนการใส่ A,B,C,D,E ลงไปให้

กับช่วงคะแนน T แต่ละช่วงนั้น เราเรียกว่า "ประเมินค่า" (Evaluation) เพราะในช่วงนี้เราต้องคิดแล้วก็คิดว่าควรจะให้เกรดอะไรสูงสุด เกรดอะไรต่ำสุด จะตัดก็เกรด คั้งกล่าวแล้วในข้อ 2.1 คั้งนี้ "การวัดผล" กับ "การประเมินค่า" จึงมักคั้งกันเสมอ จะวัดโดยไม่ประเมินค่าก็ไม่รู้จะวัดไปทำไม ประเมินค่าโดยไม่วัดก็ไม่ม้หลักเกณฑ์ที่จะให้การประเมินค่าเป็นไปโดยเที่ยงตรงได้ แต่การกำหนด A,B,C,D,E ลงไปในช่วงคะแนน T นั้น ยังไม่ถือว่าเป็นการประเมินค่าที่สมบูรณ์ (ในบางกรณี) ปัญหาที่เกิดขึ้นตามมาในขณะนั้นก็คือ "พวกคาบเส้น" ที่อยู่ระหว่างเกรดต่าง ๆ เราจะทำอย่างไร ในกรณีเช่นนี้ผู้เขียนใคร่ขอเสนอแนะว่า ให้อยู่ในดุลยพินิจของอาจารย์ผู้สอนว่าเด็กคนนั้นเป็นเช่นไร ชาติเรียนบ่อยไหม การแสดงความกิดเห็นในชั้น การส่งงานย่อย ๆ ที่ไม่คิดเกรดการตอบคำถามของอาจารย์ ฯลฯ เป็นเช่นไร เมื่อพิจารณาอย่างรอบคอบแล้วก็อาจจะเลื่อนลงได้ว่าควรจะอยู่ในเกรดบนหรือเกรดล่าง การทำเช่นนี้ก็เพราะเรามีความคิดว่า ข้อสอบของไมรู้ว่าค่า p,r (ค่าความยาก, ค่าอำนาจจำแนก ตามลำดับ) เป็นเช่นไรค่าความเชื่อมั่น (r_{tt}) เป็นเช่นไร เรา จะหลงเชื่อเฉพาะข้อสอบที่เราไม่มั่นใจ แน่นอนนัก เช่นนั้นก็อาจเสียความเที่ยงตรงได้ ในทางกลับกัน ถ้าข้อสอบของเราดีแล้วการทำเช่นนั้น ก็นับว่าไม่ถูก ทางที่ดีครูจะต้องรู้จักเด็กของตนพอเพียง

ด้วยเหตุที่ตนเอง "Measurement" จึงเป็นพวก
"Science" แต่ "Evaluation" เป็นพวก "Art"
เมื่อเป็นเช่นนี้ "อาจารย์หรือครูผู้ใดสอนชั้นใดก็
ควรจะคัดเกรดเช่นนั้น" จะให้คนอื่นคัดเกรดให้
เป็นการไม่เหมาะสมอย่างยิ่ง เพราะเขาจะ Evaluation
เด็กที่เขาไม่เคยสอนเลยให้ถูกต้องได้อย่างไร

ข้อเสนอแนะในการคัดเกรด

1. การคัดเกรดก็คล้าย ๆ กับผู้พิพากษา
ตัดสินความเขาจะต้องให้ "พยานหลายปาก" มา
เล่าให้ฟัง หลังจากได้รับฟังจากพยานปากต่าง ๆ
แล้วจึงวินิจฉัยลงไปว่า "ติดคุก" หรือ "ปล่อย"
ครูก็ทำนองเดียวกันจะให้ A,B หรือ E กับเด็ก
คนใดโดยดูแต่ "คะแนนจากการสอบ" เพียงอย่าง
เดียวไม่ดู "พยาน" จากท่านอื่น ๆ บ้างด้วยแล้ว
ก็จะทำให้การตัดสินใจใส่เกรดควรวานนั้นผิดพลาดได้
ง่ายครูจึงควรใช้อย่างอื่นมาพิจารณาด้วย เช่น งาน
การอภิปรายในชั้น ความสนใจ การขาดเรียน

2. ควรคัดเกรดหลายเกรด แต่ข้อนี้อาจจะ
ใช้พิสัย (Range) ของคะแนนดิบช่วยพิจารณาได้
ดังกล่าวมาครั้งหนึ่งแล้วในข้อ 2.1 (โปรดดูข้อ
2.1)

3. ในขณะที่คัดเกรด ผู้ที่คัดเกรดควร

คำนึงถึง

3.1 ความเที่ยงตรง

3.2 ความยุติธรรมและคุณธรรมอันสูง
ยิ่ง

4. เกรดเป็นดัชนีซึ่งบ่งถึง "คุณภาพ" ของ
สถาบันนั้น ๆ คุณภาพของเด็ก คุณภาพของครู
และการบรรลุเป้าหมายของวิชานั้น ๆ ถ้าครูสอน
ไม่เต็มที่เด็กมีความรู้ไม่พอหลักสูตร แต่ได้เกรด
A เป็นแฉวง A นั้นมีค่าต่ำ อาจจะมีเกรด C
จากตอนที่เขาคำนึงถึงคุณภาพไม่ได้ด้วยซ้ำไป

5. การสอนคนเรามุ่งสอนให้มีความรู้ และ
ให้คิดเป็น แต่ถ้าข้อสอบของเราวัดแต่ความจำใน
Method, Principles, แต่อย่างเดียวไม่ได้วัดความ
คิดบ้าง A นั้น ๆ ก็เป็น A จำไม่ใช่ A ปัญหาเรา
จะให้ A จำหรือ A คิดก็ขึ้นอยู่กับ "ข้อสอบ"
ดังนั้น ครูจึงควรฝึกเขียนข้อสอบให้เป็นเข้าใจ
เป็นดี การที่เห็นใคร ๆ เขาเขียนข้อสอบแบบเลือก
ตอบ เราเห็นว่าตรงง่าย เขียนบ้าง แต่เขียนไม่
ถูกหลักวิชานั้น จะวัดได้แต่ความจำ และเป็น
อันตรายอย่างยิ่งสำหรับค่าของ "เกรด" □