E-411 PRMA

LECTURE 13 - EQUATING

Christopher David Desjardins

MOTIVATION

Consider the salary of a teacher at a school now and in 1950

Is it fair to compare their salaries?

There salaries will most certainly not be the same

The krona has changed a lot, right?

How can we most fairly compare these salaries?

One possibility would be to compare the salaries against a set up comparable goods (e.g. price of milk, liter of gas, price of a stamp, etc)

HOW SHOULD WE COMPARE TEST SCORES ON DIFFERENT FORMS OF A TEST?

EQUATING IN TESTING

- We often have multiple forms of a test
  - Parallel or alternate form
- How might we compare these forms?
  - Sum up number correct
  - Calculate percent correct
  - Raw Scores
- **Problem:** These tests are composed of different questions with differences in item difficulties
- Why is this a problem?

SCALED SCORES

Tests need to be comparable across forms

We need to scale our scores to adjust for different difficulty

For each possible raw score, we will come up with a scaled score based on the difficulty of the questions

Review: What measurement framework do you think we are using?

	Scaled Scores		
Raw Score	Form A	Form B	Form C
50	130	130	130
49	130	130	128
48	129	130	126
47	127	130	124
46	126	130	122
45	124	129	120
44	121	128	118
43	119	127	115
42	118	126	114
41	117	125	113
40	116	124	110

EQUATING PROCESS

- The first form that is used to derive scale scores is the base form
- After the raw-to-scale conversion has occurred this form is on scale
- We then equate a new form to a form that is already on scale
- The form already on scale is our reference form and the form that is not yet equated is our new form
- Once our new form is on scale, we can calculate raw scores for the new test takers for the reference form and use the reference form to derive scaled scores
- **An issue** - Can derive reference scores that weren't possible because of discreteness

New Form Raw-to-Raw		**Reference Form Raw-to-Scale**	
New	Reference	Reference	Scaled
…	…	…	…
39	43.25	44	109.765
38	42.80	43	107.643
37	41.75	42	106.902
36	41	41	103.853

What should someone with a 38 on the new form get for a scaled score?

TEST TAKERS WITH A 38 ON THE NEW FORM

38's reference test score was 42.80

This is 80% of the way between 42 and 43

```
# 80% of the way between 42 and 43
(107.643 - 106.902) * .80
[1] 0.5928

# Add this to the score for 42
106.902 + 0.5928
[1] 107.4948
```

They should get a 107.4948

SCALE DECISIONS IN EQUATING

- Choosing the range of scale scores
  - Don't want scale to look like total or percent correct
- How fine should our scale be?
  - Usually, want each raw score to correspond to a unique scaled score
  - Need to be careful to not exaggerate precision
- Often truncate the scaled scores at the end
  - Allows test takers on an easier form than the reference form to get the highest possible scaled score
  - Truncate at lower end to avoid meaningless distinctions if scores are below chance alone

HOW TO CREATE THE RAW-TO-SCALE CONVERSION

Decide on the mean and standard deviation of a group of test takers

Choose two raw scores, specify their scaled scores, then linearly interpolate the other scores

GENERAL LIMITATIONS OF EQUATING

A test taker may know more answers on one form of a test

Equating is unable to adjust scores correctly for every test taker!

We strive to be approximately correct for our target population

Two groups could differ based on emphasized material (e.g. a teacher effect)

Equating results in discrete scores (well, we report them that way)

SYMMETRY OF EQUATING

A score of 20 on test form A corresponds to a score of 25 on test form B

A score of 25 on test form B corresponds to a score of 20 on test form A

This is known as symmetry

Statistical prediction isn't like this!

CARS AGAIN

```
mod1 <- lm(speed ~ dist, cars)
mod2 <- lm(dist ~ speed, cars)
predict(mod1, newdata = list(dist = 100))
       1
24.84066
predict(mod2, newdata = list(speed = 24.84066))
       1
80.10453
```

EQUATING DESIGNS

- To make scores comparable you need something similar across the forms
- This could involve …
    - Same group
        - Differences in score distributions are a function of form difficulty
    - Equivalent groups
        - Two random samples from the same population
        - Group ability, again, assumed constant and differences in score distributions are a function of form difficulty
    - Nonequivalent group
        - Two random samples from two populations
        - Common anchor items is necessary
        - Equating methods more complex

OUR FIRST DEF'N OF EQUATING

“A score on the new form and a score on the reference form are equivalent in a group of test takers if they represent the same relative position in the group.”

MEAN EQUATING

The simplest form of equating involves adjusting the scores by the difference in means between the reference and new forms

Substraction of values if the new form is easier

Addition of values if the new form is harder

EXAMPLE

Suppose the target population's mean on the reference form was 80 and their mean on the new form was 85.

1. Which form was harder?
2. What should should someone with a 90 on the new form get on the reference form if we were using mean equating?

PROBLEM WITH MEAN EQUATING (LIVINGSTON, 2014)

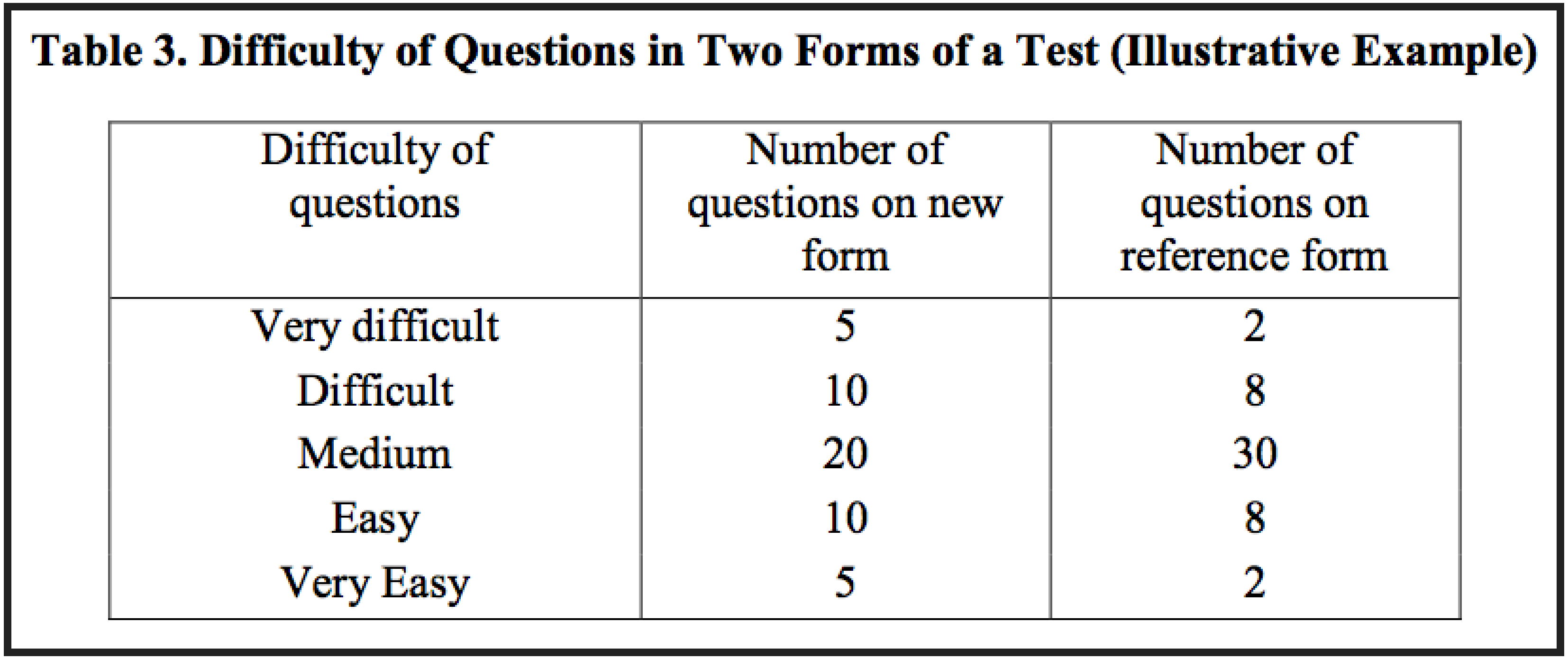


Table 3. Difficulty of Questions in Two Forms of a Test (Illustrative Example)

Difficulty of questions	Number of questions on new form	Number of questions on reference form
Very difficult	5	2
Difficult	10	8
Medium	20	30
Easy	10	8
Very Easy	5	2

LINEAR EQUATING

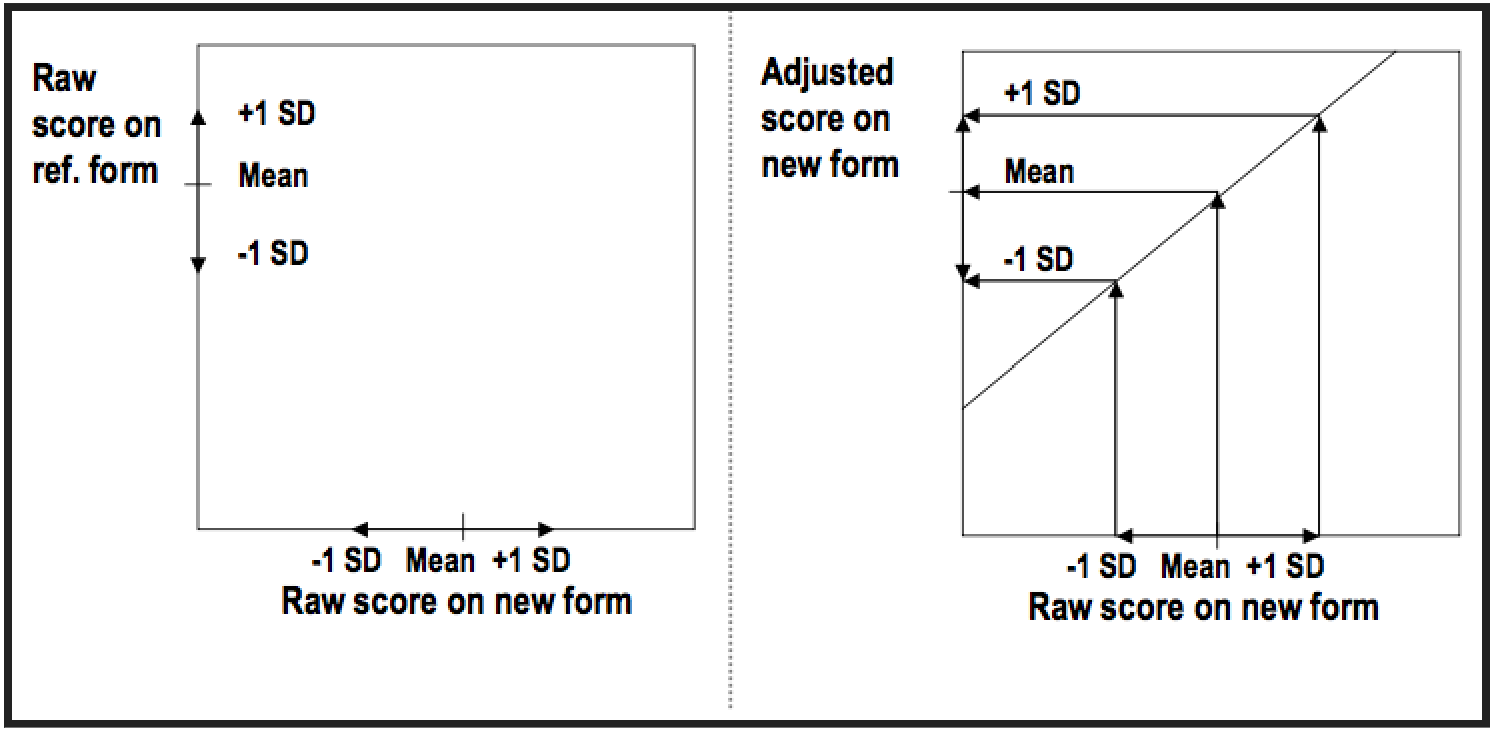
We need to adjust based on how high or low a test taker's score is from the mean

What might we consider doing?

EQUATING BETTER DEF'N

“ A score on the new form and a score on the reference form are equivalent in a group of test takers if they are the same number of standard deviations above or below the mean of the group. ”

Linear equating a harder new form (Livingston, 2014)



LINEAR EQUATING CONCEPTUALLY

- Make the adjusted new form mean equal to the reference score mean
- Same with standard deviations above and below the mean
- Do this for every possible value
- This results in a linear relationship between the new form raw and the new form adjusted scores

DOING THE MATHS!

Let *NF* stand for a score on the new form and *RF* a score on the reference form

$$\frac{RF - \bar{RF}}{sd(RF)} = \frac{NF - \bar{NF}}{sd(NF)}$$

OUR NEW ADJUSTED SCORE

$$RF = \frac{sd(RF)}{sd(NF)}NF + \bar{RF} - \frac{sd(RF)}{sd(NF)}\bar{NF} = \text{adjusted } NF$$

Note, the adjusted *NF* score is very unlikely to ever be a whole number

EXAMPLE

Form	Mean	Standard Deviation
Reference	82	15
New	79	14

If someone scored an 80 on the new form, what should there reference form score be?

$$RF = \frac{15}{14}80 + 82 - \frac{15}{14}79$$

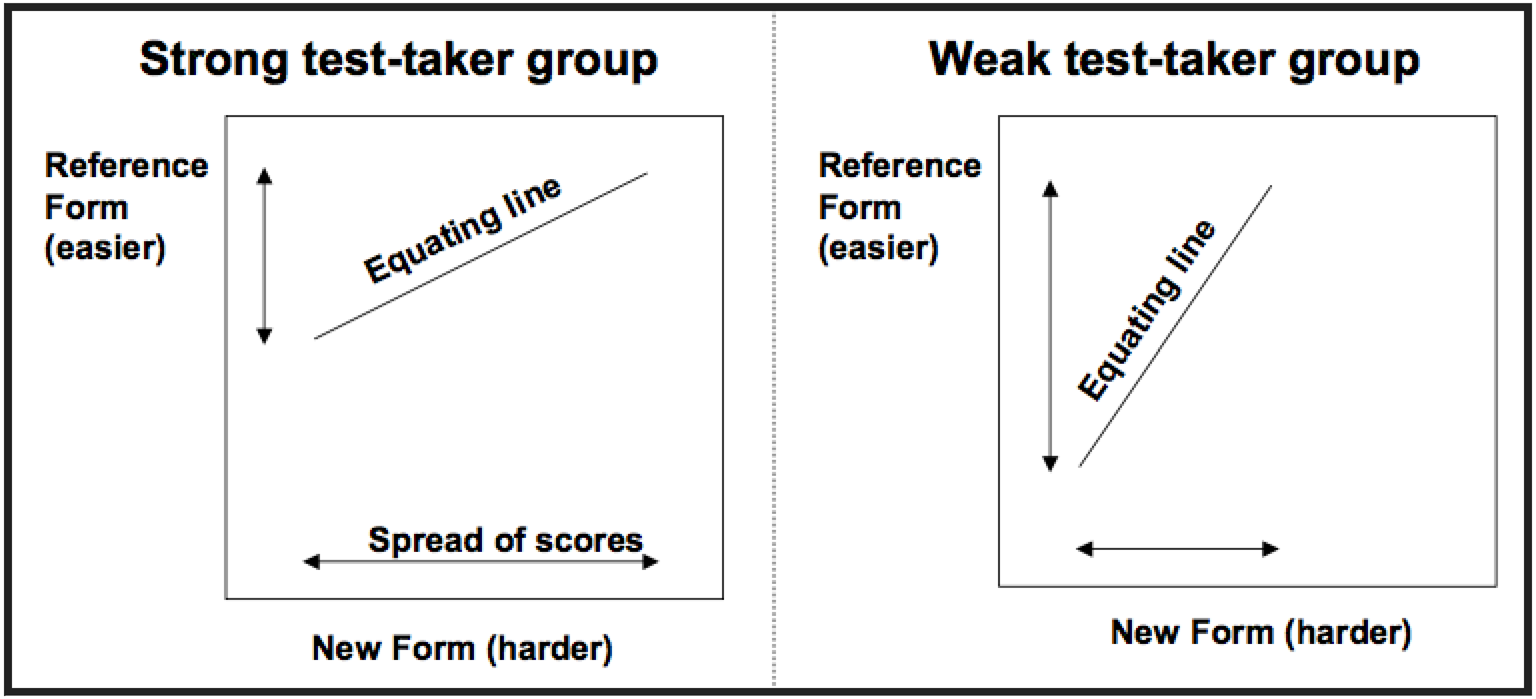
```
# Do the math in R and save it as RF
RF <- (15 / 14) * 80 + 82 - (15 / 14) * 79

# Print RF
RF
[1] 83.07143
```

Does 83.07413 seem sensible?

PROBLEMS WITH LINEAR EQUATING

A very high or very low score can equate to a score outside of the range on the reference form

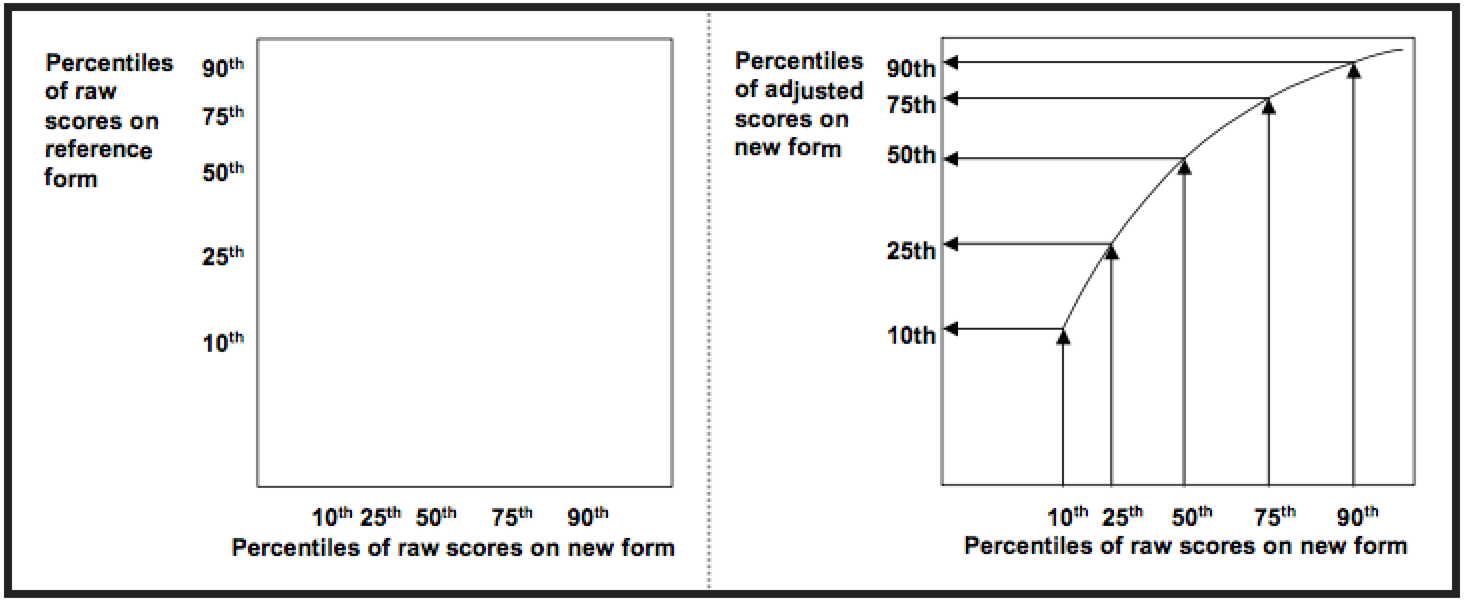
Depends heavily on the group of test takers (e.g. are they strong test takers? weak test takers?)



EQUIPERCENTILE EQUATING

"To equate scores on the new form to scores on the reference form in a group of test takers, transform each score on the new form to the score on the reference form that has the same percentile rank in that group."

EQUIPERCENTILE EQUATING WITH A HARDER NEW FORM


Percentiles of raw scores on reference form
90th
75th
50th
25th
10th
10th 25th 50th 75th 90th
Percentiles of raw scores on new form
Percentiles of adjusted scores on new form
90th
75th
50th
25th
10th
10th 25th 50th 75th 90th
Percentiles of raw scores on new form

EQUIPERCENTILE EQUATING

15th percentile of the adjusted test form corresponds (as much as possible) to 15th percentile on the reference form and so on

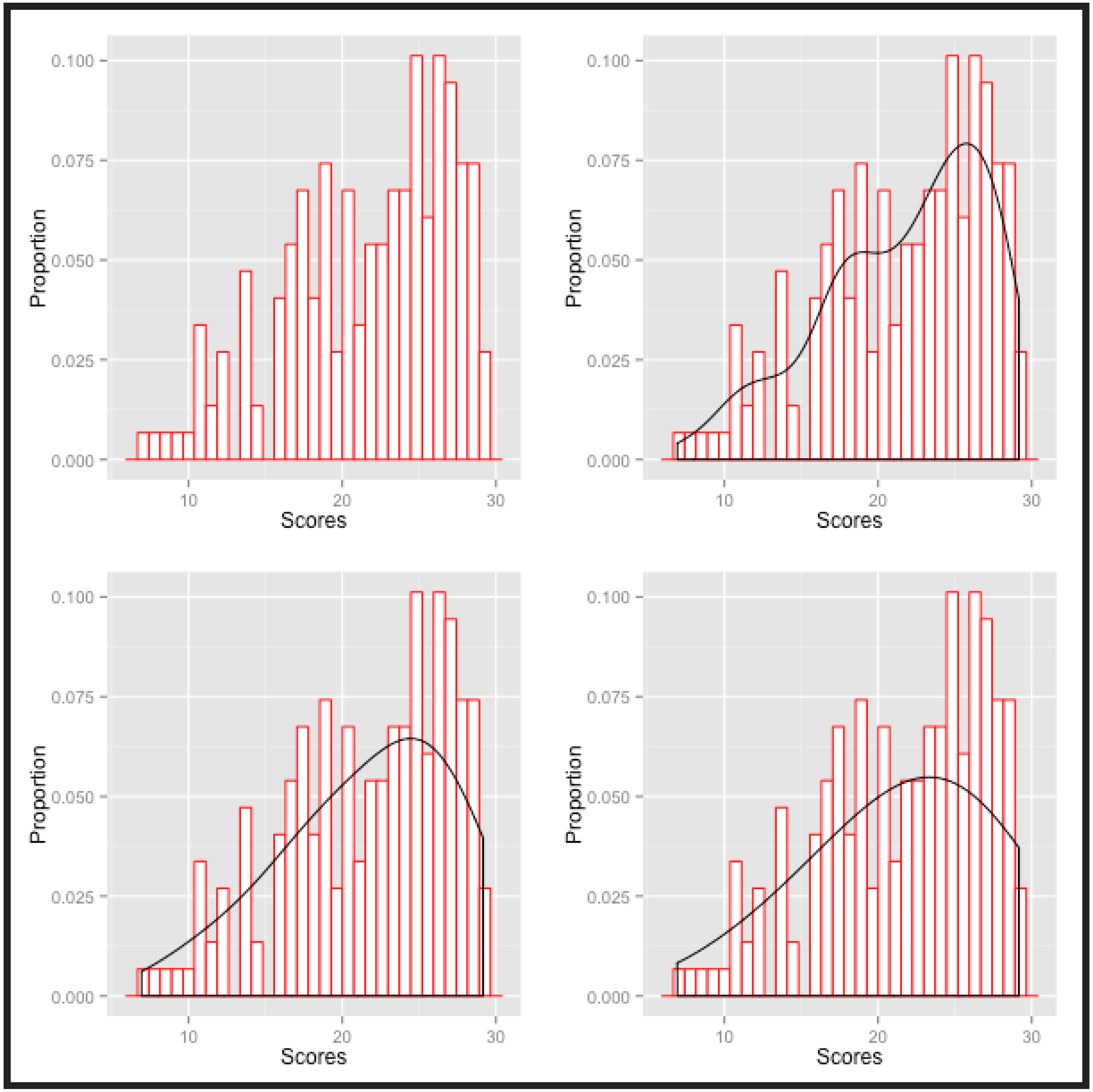
Adjusted scores will all fall within the range of possible scores on the reference form

The steepness of the slope of the curve can vary

Will result in the adjusted test scores having a similar distribution to the reference form

Will be identical to linear equating when the distribution of scores on the new form has the same shape as the distribution of the scores on the reference form

SMOOTHING



LIMITATIONS OF EQUIPERCENTILE

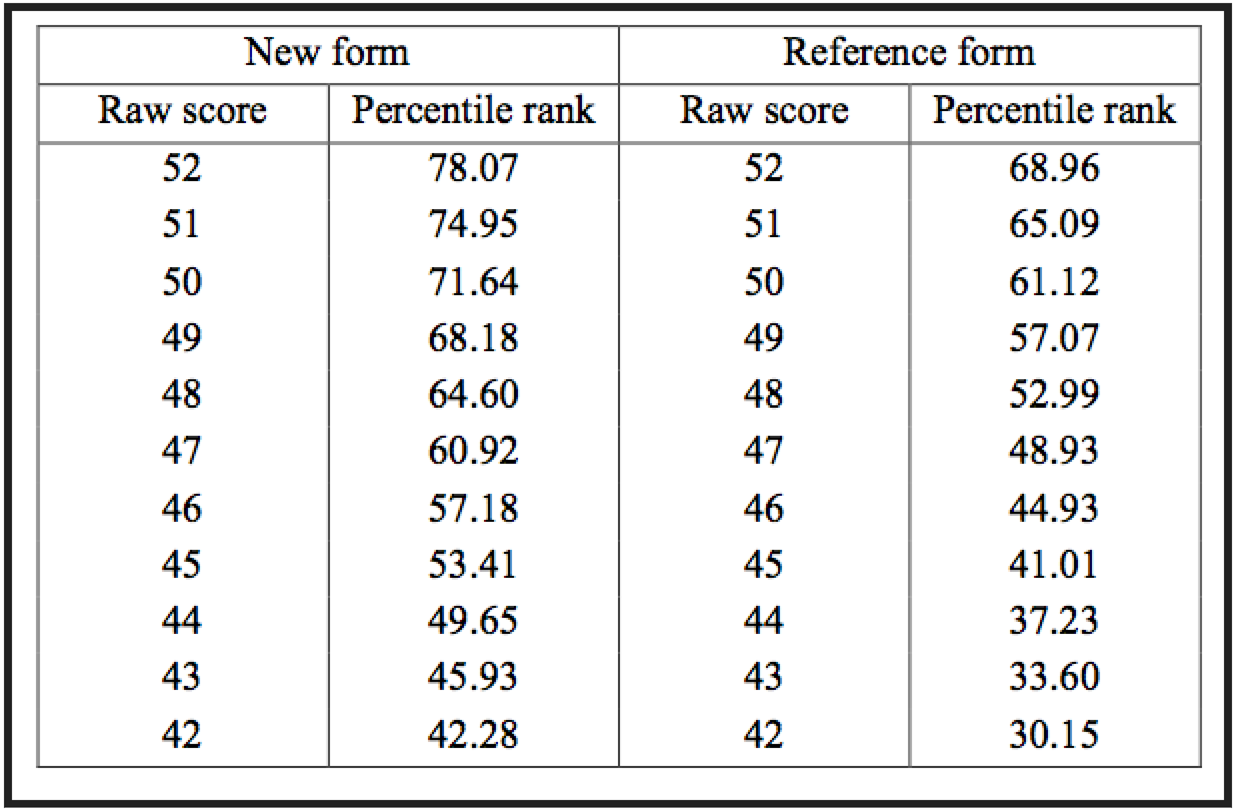
Equating relationship is bound by the highest and lowest observed score

On a difficult test, the highest possible raw score might not be observed

Future administration could result in a higher score being observed

Smoothing may help with this

AGAIN, THE DISCRETENESS PROBLEM (LIVINGSTON, 2014)



New form		Reference form	
Raw score	Percentile rank	Raw score	Percentile rank
52	78.07	52	68.96
51	74.95	51	65.09
50	71.64	50	61.12
49	68.18	49	57.07
48	64.60	48	52.99
47	60.92	47	48.93
46	57.18	46	44.93
45	53.41	45	41.01
44	49.65	44	37.23
43	45.93	43	33.60
42	42.28	42	30.15

Can use interpolation to calculate unobserved raw score

CONCLUDING REMARKS ON EQUATING

- Lots of other equating methods exist beyond these three
- Lots of other equating design exist beyond these introduced (briefly)
- Non-equivalent group designs are tricky
- The equate package in R does all of this (and more)