Rescaling for Success: Benefits and Pitfalls of Rescaling Scores

Dr. Frank Olmos

Senior HR Analyst

Los Angeles County Office of Education

Adjunct Professor,

Division of Applied and Advanced Studies in Education

at California State University, Los Angeles

PTC-SC Conference 11.01.2018, Downey, CA

Overview

- The Typical Exam Plan
- Computing Final Scores
  - Types of Scales
    - Raw
    - Percentage
    - Rescaled (Linear Transformations in Statistics)
    - Standardized – T Scores
- Simulation: What happens to candidates' final scores when one uses scaled scores with the same, different, or similar variances?
- Findings and Recommendations

The Typical Exam Plan

PTC-SC Conference 11.01.2018, Downey, CA

Subject Matter Expert Meeting

Exam Plan

| Part Weights			100%	20%	30%	50%
Competencies	Level	Wt.	100	Written Test	Performance	Interview
Professional/Technical Expertise	3	14%	14	5	5	4
Environmental Exposure Tolerance	2	11%	11		11	
Decision Making	3	11%	11		6	5
Oral Communication	3	9%	9		4	5
Safety Focus	3	9%	9	5		4
Handling Conflict	2	8%	8		4	4
Teamwork	3	8%	8			8
Customer Focus	2	7%	7			7
Self Management	2	7%	7			7
Writing	2	7%	7	7		
Professional Impact	2	6%	6			6
Mechanical Insight	2	3%	3	3		

Exam Plan Implementation

Computing Final Scores

Key concepts and definitions

- <u>Mean (average):</u> the central tendency of scores.
- <u>Variance</u>: a measurement of score dispersion; the average of the squared differences from the Mean.
    - Standard Deviation: it is the square root of the Variance
- <u>Slope</u>: the rate of change in y as x changes

Distributions w/ Different Variances
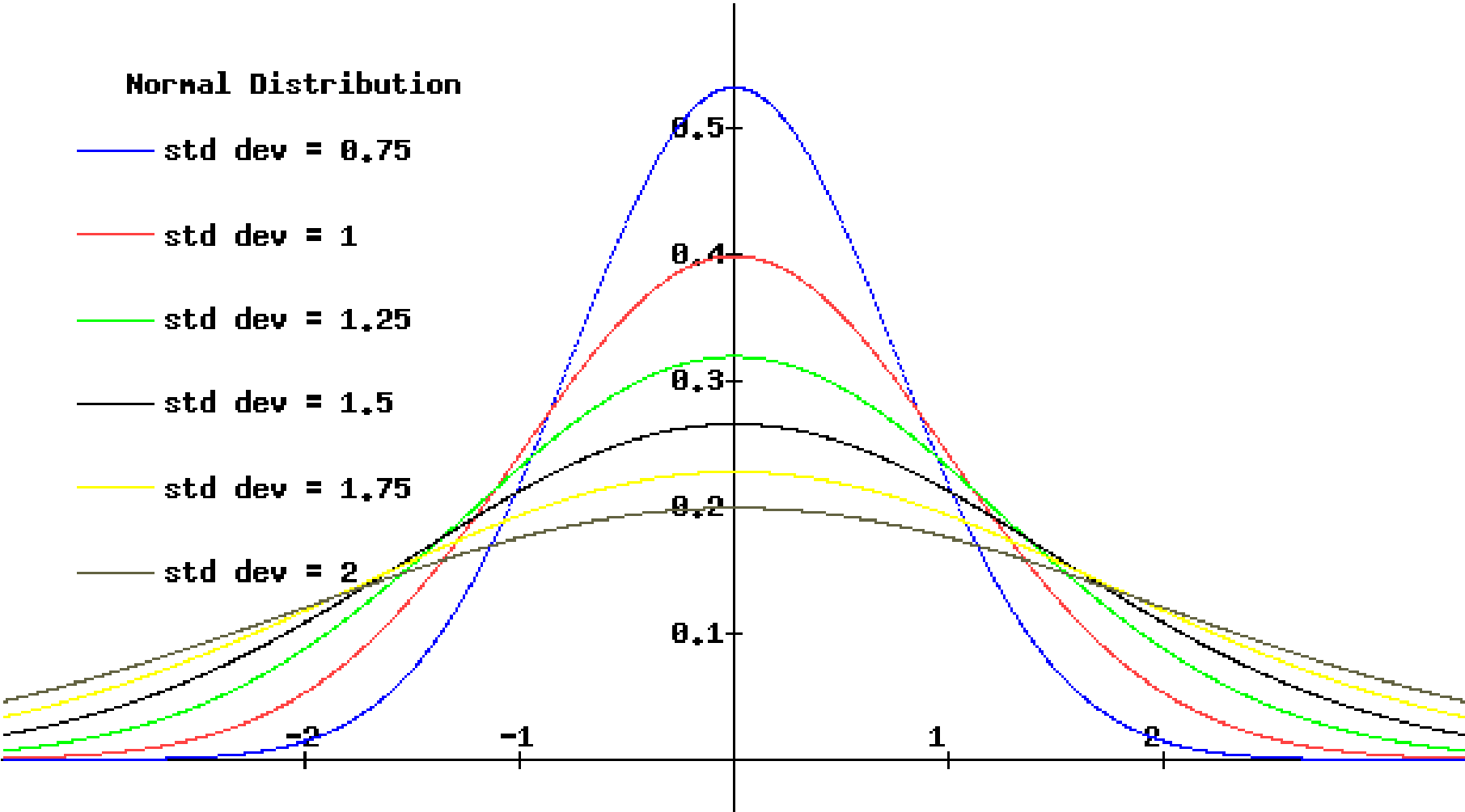
Changing Slopes

- When scores are converted, the steeper the slope, the higher the variance; the flatter the slope, the lower the variance.

$y=2x+2$

$y=1/2x +2$

$y=2/7x+2$
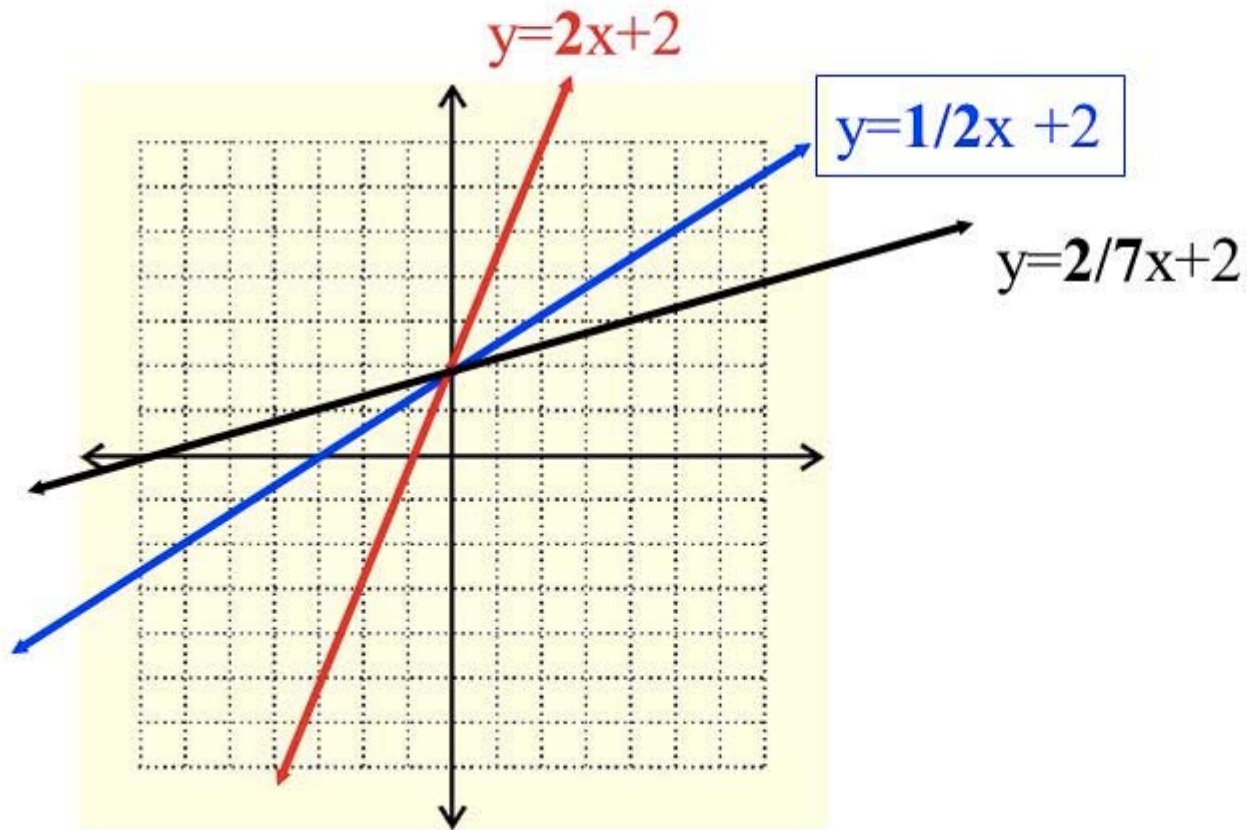
Raw Scores

Raw Scores

- Multiple Choice Exams
  - Raw points consists the number of total exam questions in the exam.
  - Candidate scores (typically) consists of the # correct answers.
- T/E, Performance, and Oral Exams
  - Non-weighted/weighted Averages
  - Wholistic Percentages
  - Rating scales
    - ex: 1=Limited, 2=Moderate, 3=Strong, 4=Compelling.

Raw Scores Limitation

- **<u>Issue #1:</u>** Scores from different metrics can not be directly added together.
  - 7000 keystrokes per hour plus 30 kilometers = ??
  - 50 WPM typing plus 35 correct on written test = ???
  - 150 points on scored essay plus 85 on interview = ???
- Raw scores must therefore be converted to another metric for proper final score summation.

Raw Score conversion Choices

- Percentage scores (a rescaled score)
- Re-scaled scores
- T-scores (standardized scores)

Percentage scores

Percentage

$$\frac{part}{whole} \times 100 = \%$$

Percentage Calculation Formula

- <u>Advantages:</u>
  - Percentage: Referenced in the Education code
  - Easy to compute and understand
  - Widely accepted
- <u>Disadvantage</u>
- The raw to % conversion will affect the range and variability of the scores, which will affect weights.

Percentage Simulations / Variances

- Random numbers were generated using different ranges (1-1000; 1-100; 1-50; 1-25; 1-4)

- These raw numbers were then converted into percentages and their corresponding variances were calculated.

- Variances ranged from 852 to 1023

Issue #2: Differences of exam part scale variances will affect the intended exam part's weight.

Re-Scaled Scores

aka: Linear Transformations

Re-Scaled Score Formula

Converts passing range of raw scores to 70 to 100.

Derived from:
$$Y = mX + b$$

Slope

Rescaled Score = Factor * (Raw - Raw Pass) + 70

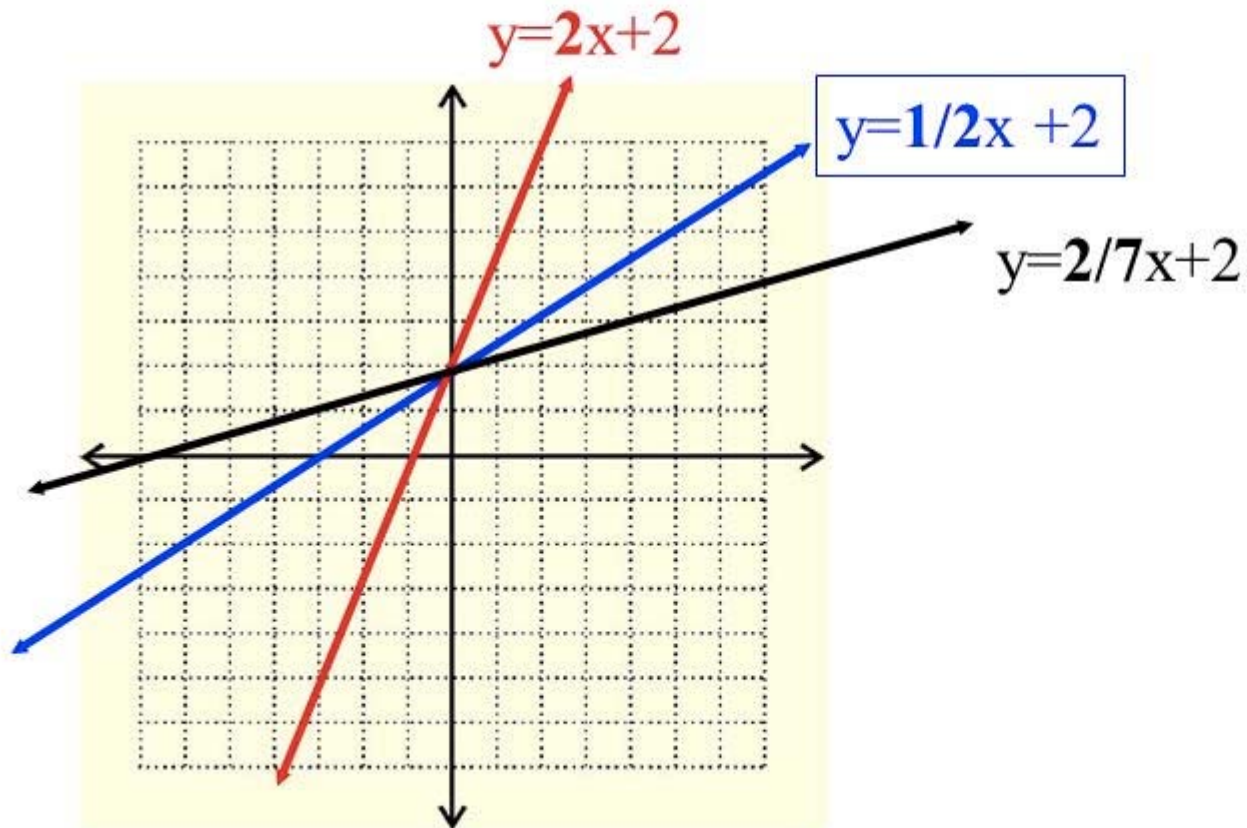**Factor =**

30 / (Raw Max – Raw Pass)

- Take note: The **Raw Pass Point will** affect the slope;
- The slope will affect the variance/dispersion of scores.
- The flatter the slope, the lower the variance

Changing Slopes

- The rescale factor, the slope, controls how quickly the line rises as we move from left to right.
- The steeper the slope, the higher the variance; the flatter the slope, the lower the variance.
- Higher Pass points = Higher Variance; Lower Pass points = Lower Variance.

$y=2x+2$

$y=1/2x +2$

$y=2/7x+2$

Re-Scaled Score Example

- Total test items = 90

- Raw pass point set at 50

- Formula: Rescaled Score = Factor * (Raw - Raw Pass) + 70

- Factor = 30/(90-50) = .75 (each raw point is worth .75)

- Candidate has raw score of 65

  65 rescales to:

  (.75*(65-50)) + 70

 (.75 * 15) + 70

  11.25 + 70 = 81.25

Re-Scaled Scores

- Advantages
  - Converts passing range of raw scores to 70 to 100
  - Necessary when mixing written test results with interviews scored on 70-100 passing range (requiring no rescaling)
  - Range of passing scores is the same across rescaled exam parts.

- Disadvantage
  - Issue #2: Variability still is not controlled – intended weights may differ from actual.

Standardized Scores: T Scores

T-Score Formula

$$T = 50 + 10 \frac{(x - m)}{s}$$
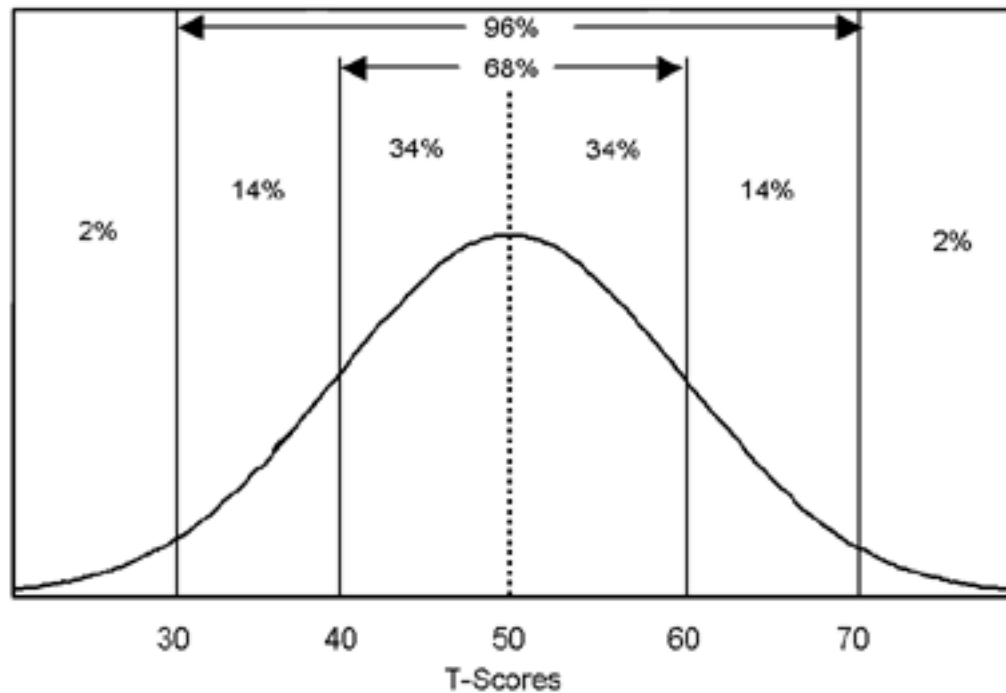
T - Score

- Scores are standardize to have a mean of 50 and a standard deviation of 10.

- To standardize:
    1. Calculate the mean and standard deviation of the score distribution
    2. Calculate the deviations of obtained raw scores from the mean and divide by the Standard Deviation
    3. Z = (Obtained - Mean) / SD
    4. Multiply the result by 10 and add 50
    5. T = 10[(Obtained - Mean)/SD] + 50   or T = 10Z + 50

T – Score Example

- Total test items = 90

- Mean of all scores = 60

- Standard Deviation = 8

- Candidate has raw score of 65

- 65 converts to a t-score of

$$10[(65 - 60)/8] + 50$$
$$10[5/8] + 50$$
$$6.25 + 50 = 56.25$$

T – Score

- Advantages
  - Most powerful and accurate way to convert scores across different units of measurement
  - **Equates the variances among methods and so preserves intended weight of the score in the final composite**

- Disadvantages
- Requires standardizing all of the exam parts (written, interview, performance, etc.)
- Interviews often have a non-linear fail point (e.g. 65) "range restriction" and can not be standardized below 70.
- Pass points will vary across parts (unlike rescaled scores)
- Highly dependent to sample size.
- Hard to explain to candidates.

Merging Lists with T – Scores

- Merging lists (scores) that are based on measurement of the same underlying competencies, but using different methods
  - Assumes candidate sampling for two lists is from the same population (talent pool) and essentially random
  - Convert both distributions to t-scores then merge and re-rank t-scores.

Re-Scale Scores Simulation

Simulation: What happens to candidates' final scores when one uses scaled scores with the same, different, or similar variances?

Exam Plan

Part Weights			100%	20%	30%	50%
Competencies	Level	Wt.	100	**Written Test**	**Performance**	**Interview**
Professional/Technical Expertise	3	14%	14	5	5	4
Environmental Exposure Tolerance	2	11%	11		11	
Decision Making	3	11%	11		6	5
Oral Communication	3	9%	9		4	5
Safety Focus	3	9%	9	5		4
Handling Conflict	2	8%	8		4	4
Teamwork	3	8%	8			8
Customer Focus	2	7%	7			7
Self Management	2	7%	7			7
Writing	2	7%	7	7		
Professional Impact	2	6%	6			6
Mechanical Insight	2	3%	3	3		

Simulation

- 3 Part Exam Plan
  - Written Exam – 20% of the total final score.
    - Raw scale: 1 to 33
  - Performance Exam - 30% of the total final score.
    - Raw scale: 1 to 114
  - Oral Exam - 50% of the total final score.
    - Raw scale: 1 to 4

- 25 Candidates applied to this process
- Each candidate received a randomized exam score for each part.

Findings and Recommendations

- Mixing final scores with <u>different variances </u>will alter the intended weighting and produce different rankings of candidates.

- If and when possible, use final scores with the same or similar **<u>variances</u>** for every weighted examination part.
  - Consider the re-scaled variance when setting passpoints.

- Ranking the Scales:
  - T-Scores (Gold - Stable Variance)
  - Percentages (Silver - Moderately Stable Variance)
  - Re-Scale (Bronze - Variable Variance)

PTC-SC Conference 11.01.2018, Downey, CA

PTC-SC Conference 11.01.2018, Downey, CA