

Evaluating the Significance of Individual Change

May 14, 2018 (3:00 - 4:00 PDT)

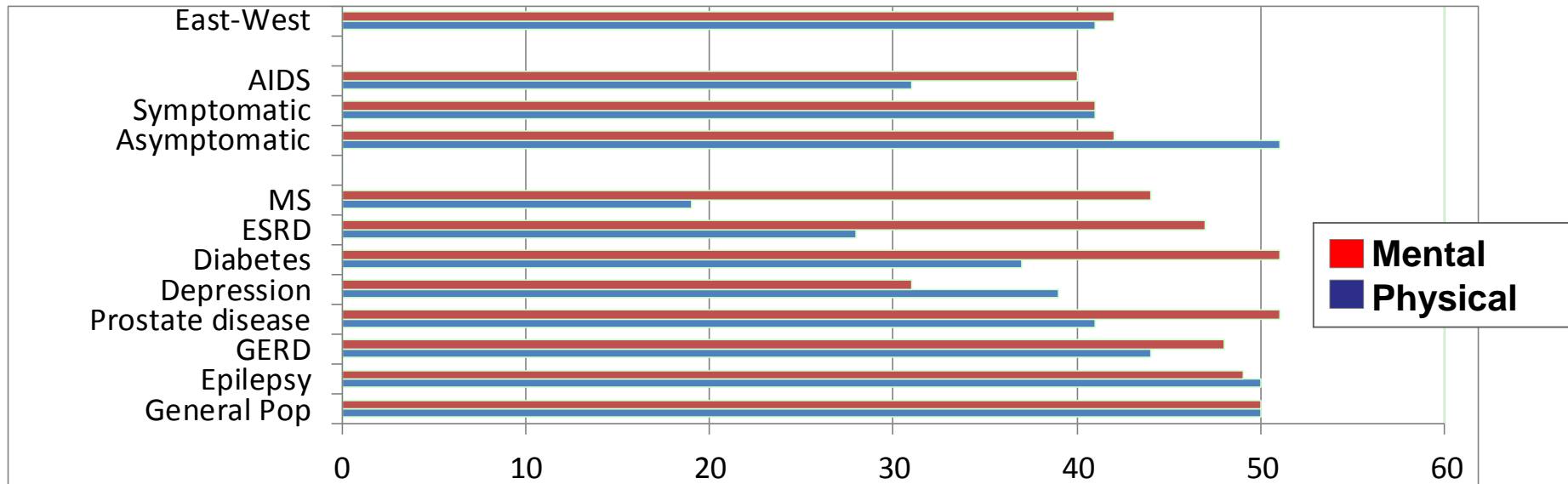
RCMAR Methods Seminar Series

Hays, R. D., Brodsky, M., Johnston, M. F., Spritzer, K. L., & Hui, K. (2005). Evaluating the statistical significance of health-related quality of life change in individual patients. Evaluation and the Health Professions, 28, 160-171.



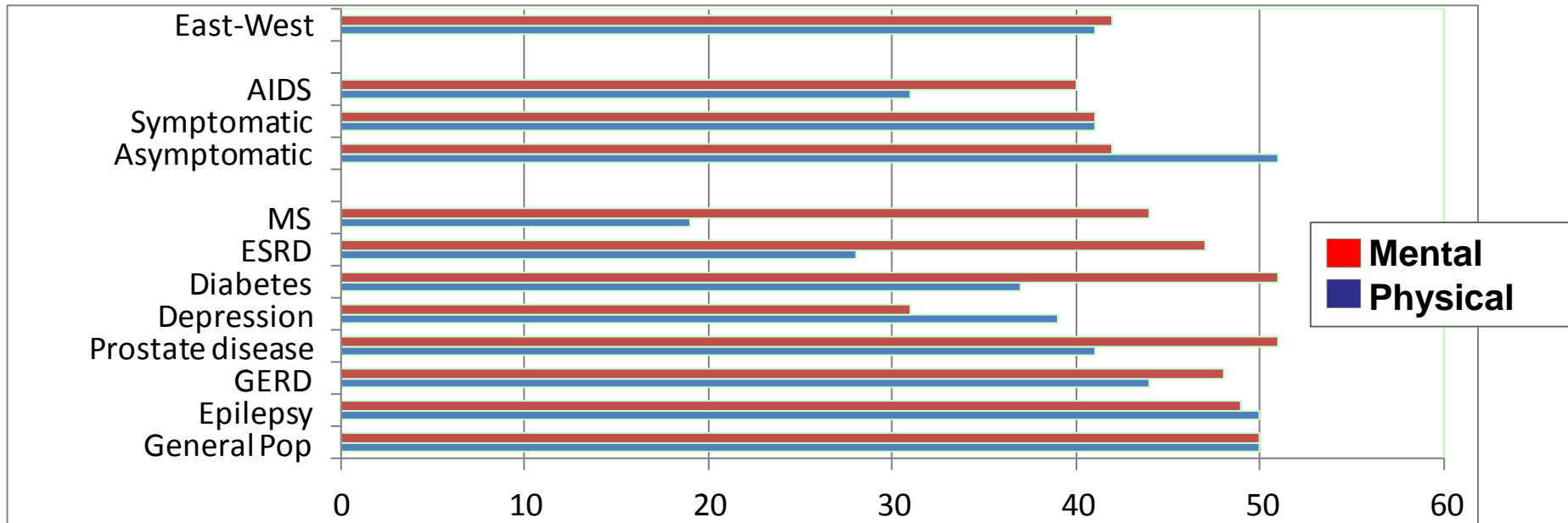
Ron D. Hays, Ph.D.
drhays@ucla.edu

Physical Functioning and Emotional Well-Being at Baseline for 54 Patients at UCLA-Center for East West Medicine



MS = multiple sclerosis; ESRD = end-stage renal disease; GERD = gastroesophageal reflux disease.

Physical Functioning and Emotional Well-Being at Baseline for 54 Patients at UCLA-Center for East West Medicine



MS = multiple sclerosis; ESRD = end-stage renal disease; GERD = gastroesophageal reflux disease.

Significant Improvement in all but 1 of SF-36 Scales (Change is in T-score metric)

	Change	t-test	prob.
PF-10	1.7	2.38	.0208
RP-4	4.1	3.81	.0004
BP-2	3.6	2.59	.0125
GH-5	2.4	2.86	.0061
EN-4	5.1	4.33	.0001
SF-2	4.7	3.51	.0009
RE-3	1.5	0.96	.3400 ←
EWB-5	4.3	3.20	.0023
PCS	2.8	3.23	.0021
MCS	3.9	2.82	.0067

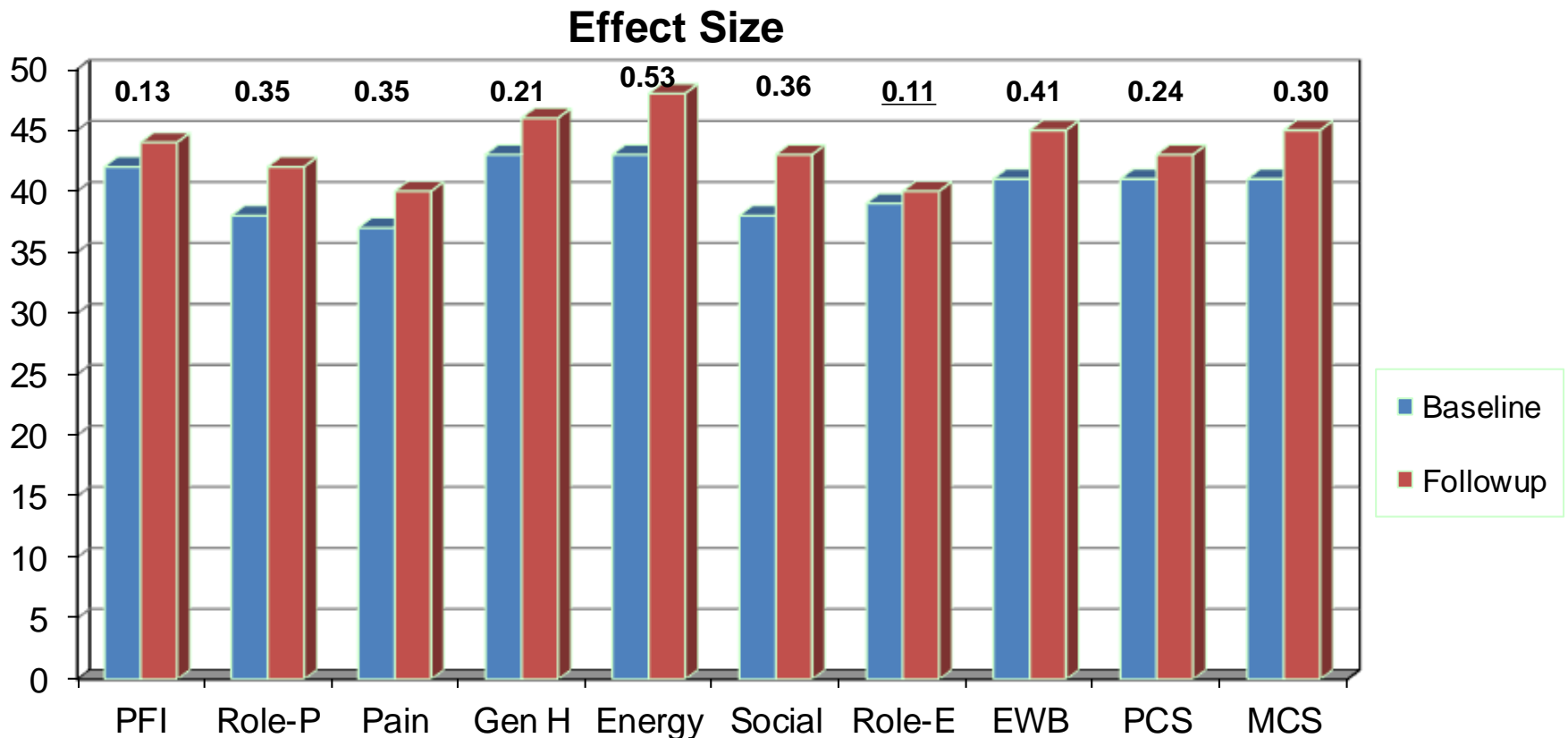
Effect Size

$$(\text{Follow-up} - \text{Baseline}) / \text{SD}_{\text{baseline}}$$

Cohen's Rule of Thumb:

- ✓ ES = 0.20 Small
- ✓ ES = 0.50 Medium
- ✓ ES = 0.80 Large

Effect Sizes for Changes in SF-36 Scores



PFI = Physical Functioning; Role-P = Role-Physical; Pain = Bodily Pain; Gen H=General Health; Energy = Energy/Fatigue; Social = Social Functioning; Role-E = Role-Emotional; EWB = Emotional Well-being; PCS = Physical Component Summary; MCS =Mental Component Summary.

0.11 0.13 0.21 0.24 0.30 0.35 0.35 0.36 0.41 0.53

Defining a Responder: Reliable Change Index (RCI)

$$\frac{X_2 - X_1}{(\sqrt{2}) (SEM)}$$

$$SEM = SD_{bl} \times \sqrt{1 - r_{xx}}$$

Note: SD_{bl} = standard deviation at baseline
 r_{xx} = reliability

Significant Change

$$\frac{X_2 - X_1}{(\sqrt{2}) (SD) \sqrt{(1 - r_{xx})}} \geq 1.96$$

Amount of Change in Observed Score Needed To be Statistically Significant

$$(\sqrt{2}) (SD) \sqrt{(1 - r_{xx})} (1.96)$$

$$= 2.77 * SD_{bl} * SQR (1 - r_{xx})$$

“Coefficient of reproducibility”

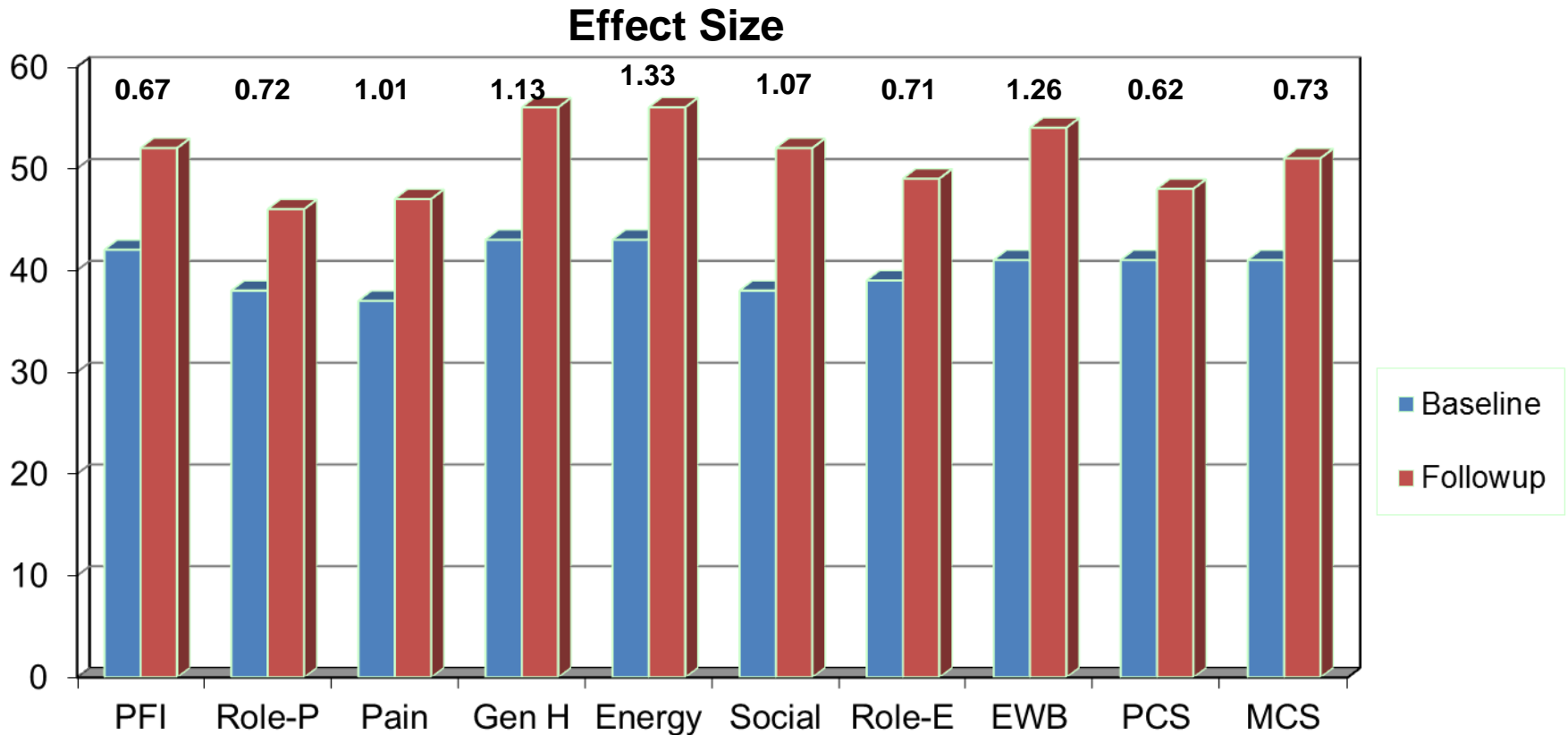
Note: SD_{bl} = standard deviation at baseline and r_{xx} = reliability

How Reliability Relates to Amount of Individual Change Needed

Reliability	SQR (1- r _{xx})	Change Needed
0.70	0.55	1.5 SD
0.80	0.44	1.2 SD
0.90	0.32	0.9 SD
0.95	0.22	0.6 SD
0.97	0.17	0.5 SD

$$2.77 * \text{SQR} (1- r_{xx}) * \text{SD}_{bl}$$

Amount of Change Needed for Significant Individual Change



PFI = Physical Functioning; Role-P = Role-Physical; Pain = Bodily Pain; Gen H=General Health; Energy = Energy/Fatigue; Social = Social Functioning; Role-E = Role-Emotional; EWB = Emotional Well-being; PCS = Physical Component Summary; MCS =Mental Component Summary.

7-31% Improve Significantly

	% Improving	% Declining	Difference
PF-10	13%	2%	+ 11%
RP-4	31%	2%	+ 29%
BP-2	22%	7%	+ 15%
GH-5	7%	0%	+ 7%
EN-4	9%	2%	+ 7%
SF-2	17%	4%	+ 13%
RE-3	15%	15%	0%
EWB-5	19%	4%	+ 15%
PCS	24%	7%	+ 17%
MCS	22%	11%	+ 11%

Observational Study of ~1800 Chiropractic Patients in Treatment for Chronic Low Back or Neck Pain

Reliability and Baseline Means for PROMIS-29 v2.0 Scales

Scale	Reliability	Baseline Mean
Physical functioning	0.86	46
Pain	0.88	44
Fatigue	0.93	47
Sleep Disturbance	0.85	48
Social Health	0.93	52
Emotional distress	0.93	50
PROMIS-29 Physical	0.90	46
PROMIS-29 Mental	0.97	48

Baseline and Endpoint (3 months later) Means and

Within Group t-tests

Scale	Baseline Mean	Endpoint Mean	t-test of change	Effect Size
Physical functioning	46	47	4.15	.07
Pain	44	46	9.48	.19
Fatigue	47	48	7.11	.15
Sleep Disturbance	48	50	8.47	.18
Social Health	52	53	7.61	.15
Emotional distress	50	50	0.04	.01
PROMIS-29 Physical	46	47	5.80	.10
PROMIS-29 Mental	48	50	9.06	.16

Note: Net benefit from manipulation at 3 months in UK BEAM (back pain exercise and manipulation) randomized trial was about 3 points on the SF-36 PCS and MCS (BMJ, 2004).

Significance of Individual Change in PROMIS-29 v2.0 Scales from Baseline to Endpoint (3 months later)

Scale	Effect Size	Got worse	Same	Got better
Physical functioning	.07	9%	78%	13%
Pain	.19	9%	74%	17%
Fatigue	.15	13%	64%	23%
Sleep Disturbance	.18	6%	80%	14%
Social Health	.15	12%	67%	21%
Emotional distress	.01	16%	68%	16%
PROMIS-29 Physical	.10	9%	76%	14%
PROMIS-29 Mental	.16	18%	52%	30%

Note: Change (Got worse or Got better) was determined by coefficient of reproducibility = $2.77 * \text{standard error of measurement}$. “Responders” were those in the “Got Better” subgroup.

Thank you

- Finkleman, M. D. et al. (2010). Item selection and hypothesis testing for the adaptive measurement of change. *Applied Psychological Measurement*, 34, 238-254.
- Jabrayilov, R. et al. (2016). Comparison of classical test theory and item response theory in individual change assessment. *Applied Psychological Measurement*, 40, 559-572.
- Wang, C., & Weiss, D. J. (2018). Multivariate hypothesis testing methods for evaluating significant individual change. *Applied Psychological Measurement*, 42, 221-239.

drhays@ucla.edu (310-794-2294)

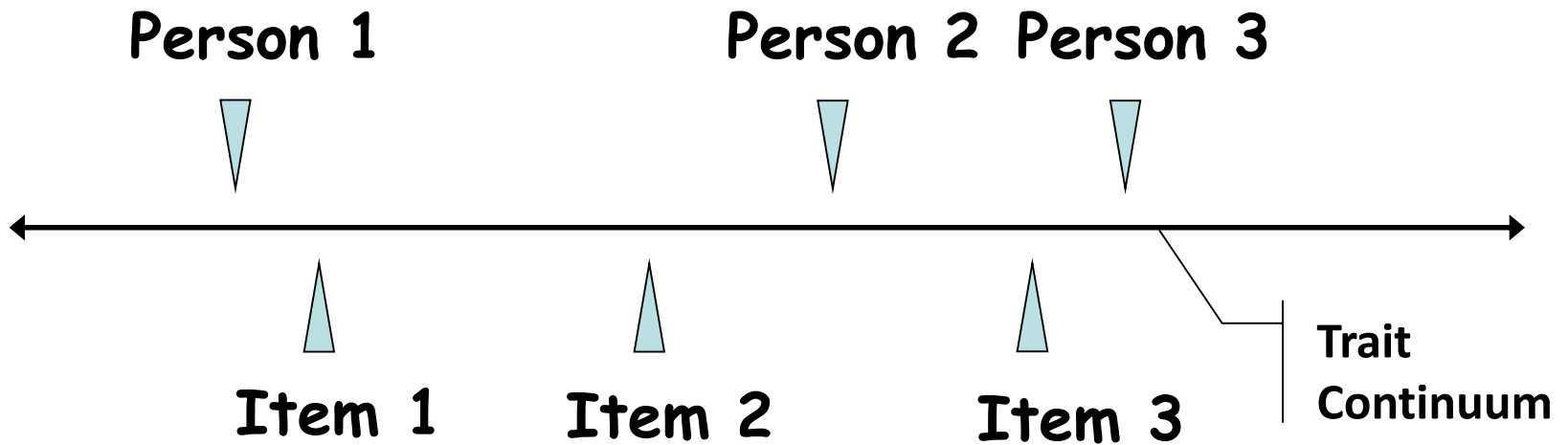
Powerpoint file available for downloading at:

<http://gim.med.ucla.edu/FacultyPages/Hays/>

Supplemental Material



Item Responses and Trait Levels



Computer Adaptive Testing (CAT)



Graduate Record Examinations®



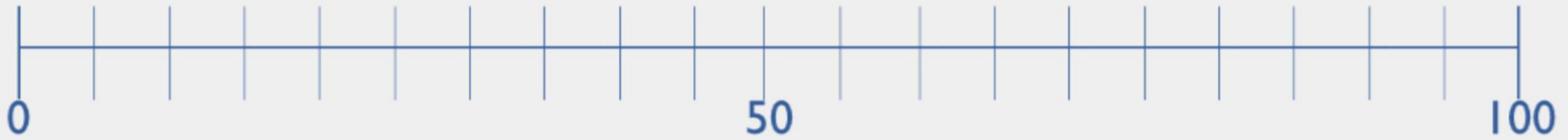
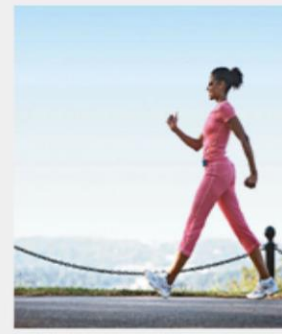
National Council
of State Boards of Nursing, Inc.



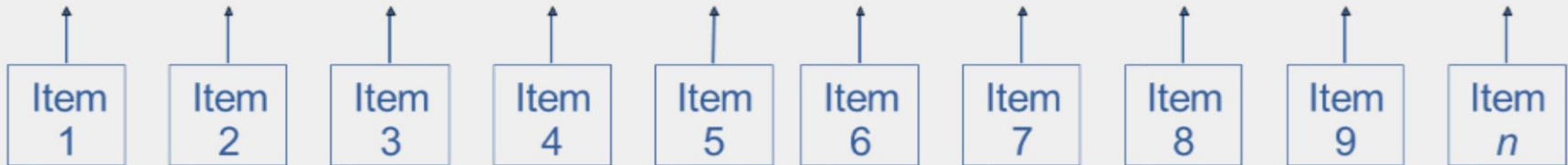
The PROMIS Metric

- T Score
 - Mean = 50
 - SD = 10
 - Referenced to US General Pop.
 - $T = 50 + (z * 10)$

www.healthmeasures.net



Physical Functioning Item Bank



Are you able to get in and out of bed?

Are you able to stand without losing your balance for 1 minute?

Are you able to walk from one room to another?

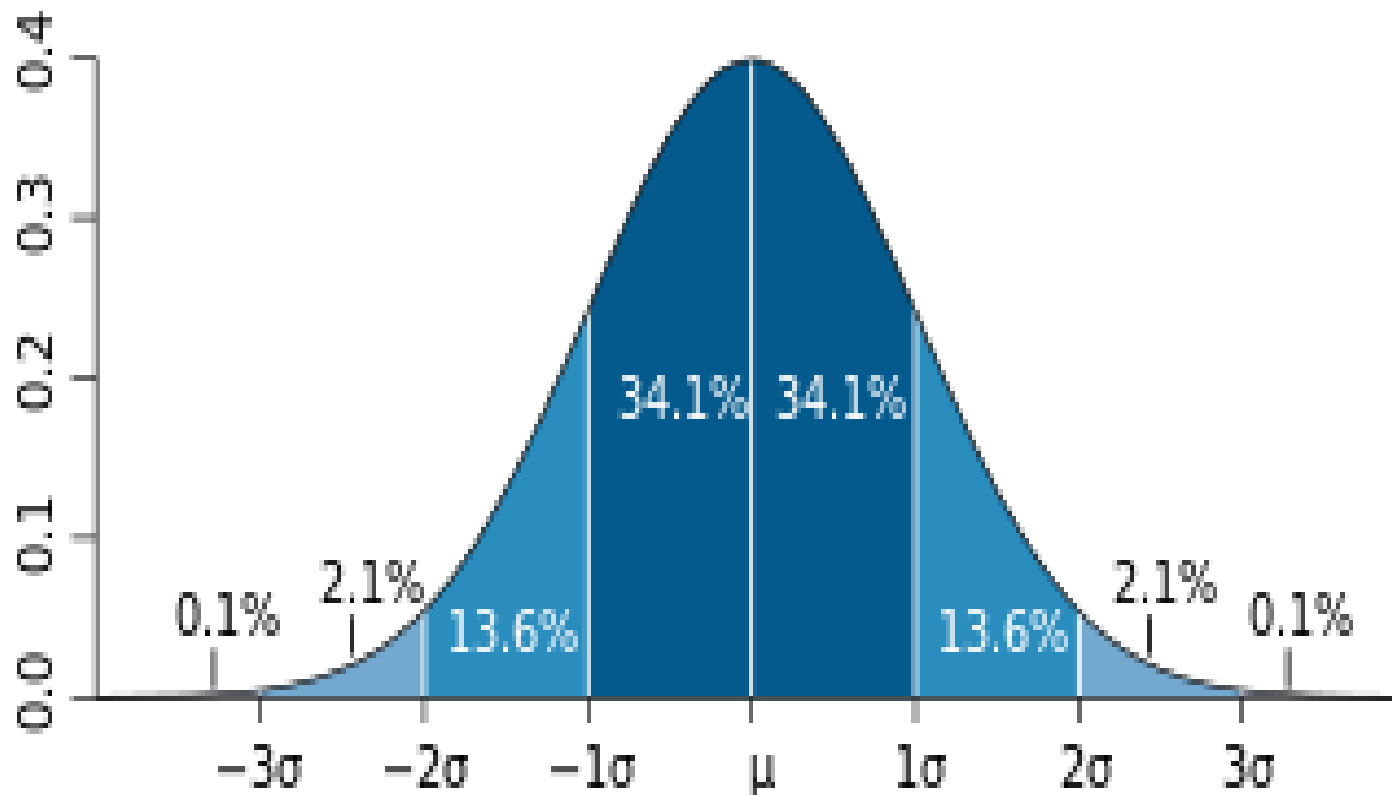
Are you able to walk a block on flat ground?

Are you able to run or jog for two miles?

Are you able to run five miles?

Normal Distribution

Within 1 SD = 68.2%, 2 SDs = 95.4%; 3 SDs = 99.6%



Reliability Target for Use of Measures with Individuals

- Reliability ranges from 0-1
 - 0.90 or above is goal
- $SE (SEM) = SD (1 - \text{reliability})^{1/2}$
- Reliability (T-score) = $1 - (SE/10)^2$
 - Reliability = 0.90 when SE = 3.2
 - 95% CI = true score +/- 1.96 x SE

In the past 7 days ...

I was grouchy [1st question]

- Never [39]
- Rarely [48]
- Sometimes [56]
- Often [64]
- Always [72]

Estimated Anger = 56.1

SE = 5.7 (rel. = 0.68)

In the past 7 days ...

I felt like I was ready to explode

[2nd question]

- Never
- Rarely
- Sometimes
- Often
- Always

Estimated Anger = 51.9

SE = 4.8 (rel. = 0.77)

In the past 7 days ...

I felt angry [3rd question]

- Never
- Rarely
- Sometimes
- Often
- Always

Estimated Anger = 50.5

SE = 3.9 (rel. = 0.85)

In the past 7 days ...

I felt angrier than I thought I should

[4th question]

- Never
- Rarely
- Sometimes
- Often
- Always

Estimated Anger = 48.8

SE = 3.6 (rel. = 0.87)

In the past 7 days ...

I felt annoyed [5th question]

- Never
- Rarely
- Sometimes
- Often
- Always

Estimated Anger = 50.1

SE = 3.2 (rel. = 0.90)

In the past 7 days ...

I made myself angry about something just by thinking about it. [6th question]

- Never
- Rarely
- Sometimes
- Often
- Always

Estimated Anger = 50.2

SE = 2.8 (rel = 0.92)