Evaluating Multi-item Scales

Ron D. Hays, Ph.D.
UCLA Division of General Internal Medicine/Health Services Research
drhays@ucla.edu

HS237A 11/10/09, 2-4pm, 41-235

http://twitter.com/RonDHays

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

Example Responses to 2-Item Scale

ID	Poor	Fair	Good	Very Good	Excellent
01					2
02				1	1
03		1		1	
04			1		1
05		2			

Cronbach's Alpha

01 55
02 45
03 42
04 35
05 22

Source	df	SS	MS
Respondents (BMS)	4	11.6	2.9
Items (JMS)	1	0.1	0.1
Resp. x Items (EMS)	4	4.4	1.1
Total	9	16.1	

$$\text{Alpha} = \frac{2.9 - 1.1}{2.9} = \frac{1.8}{2.9} = \boxed{0.62}$$

Computations

- Respondents SS
  - $(10^2+9^2+6^2+8^2+4^2)/2 - 37^2/10 = \underline{11.6}$
- Item SS
  - $(18^2+19^2)/5 - 37^2/10 = \underline{0.1}$
- Total SS
  - $(5^2+ 5^2+4^2+5^2+4^2+2^2+3^2+5^2+2^2+2^2)/2 - 37^2/10 = \underline{16.1}$
- Res. x Item SS= Tot. SS – (Res. SS+Item SS)

Alpha for Different Numbers of Items and Average Correlation

Number of Items (k)	Average Inter-item Correlation ($\bar{r}$)					
	.0	.2	.4	.6	.8	1.0
2	.000	.333	.572	.750	.889	1.000
4	.000	.500	.727	.857	.941	1.000
6	.000	.600	.800	.900	.960	1.000
8	.000	.666	.842	.924	. 970	1.000

$$\text{Alpha}_{st} = \frac{k * \bar{r}}{1 + (k - 1) * \bar{r}}$$

Spearman-Brown Prophecy Formula

$$\text{alpha}_y = \left(\frac{N \bullet \text{alpha}_x}{1 + (N - 1) * \text{alpha}_x} \right)$$

N = how much longer scale y is than scale x

Example Spearman-Brown Calculation

MHI-18

$$\frac{18/32\ (0.98)}{(1+(18/32\ -1)*0.98}$$

$$= 0.55125/0.57125 = 0.96$$

Number of Items and Reliability for Three Versions of the Mental Health Inventory (MHI)

Measure	Number of Items	Completion time (min.)	Reliability
MHI-32	32	5-8	.98
MHI-18	18	3-5	.96
MHI-5	5	1 or less	.90

From McHorney et al. 1992

Reliability Minimum Standards

- 0.70 or above (for group comparisons)
- 0.90 or higher (for individual assessment)
  - SEM = SD (1- reliability)$^{1/2}$

Intraclass Correlation and Reliability

Model	Reliability	Intraclass Correlation
One-way	$\frac{MS_{BMS} - MS_{WMS}}{MS_{BMS}}$	$\frac{MS_{BMS} - MS_{WMS}}{MS_{BMS} + (k-1)MS_{WMS}}$
Two-way fixed	$\frac{MS_{BMS} - MS_{EMS}}{MS_{BMS}}$	$\frac{MS_{BMS} - MS_{EMS}}{MS_{BMS} + (k-1)MS_{EMS}}$
Two-way random	$\frac{N(MS_{BMS} - MS_{EMS})}{NMS_{BMS} + MS_{JMS} + MS_{EMS}}$	$\frac{MS_{BMS} - MS_{EMS}}{MS_{BMS} + (k-1)MS_{EMS} + k(MS_{JMS} - MS_{EMS})/N}$

BMS = Between Ratee Mean Square
WMS = Within Mean Square
JMS = Item or Rater Mean Square
EMS = Ratee x Item (Rater) Mean Square

Equivalence of Survey Data

- Missing data rates were significantly higher for African Americans on all CAHPS items
- Internal consistency reliability did not differ
- Plan-level reliability estimates were significantly lower for African Americans than whites

M. Fongwa et al. (2006). Comparison of data quality for reports and ratings of ambulatory care by African American and White Medicare managed care enrollees. Journal of Aging and Health, 18, 707-721.

Missing Data Rates (%)

	Mail		Phone	
	White	AA	White	AA
Get Care Quickly	8	14	7	8
Get Needed Care	10	16	9	10
Staff	4	7	3	5
C. Service	9	19	10	10
Communication	3	6	4	5

Health-Plan Level Reliability

	White	AA
Get Care Quickly	0.93	0.90
Get Needed Care	0.94	0.91
Office Staff	0.90	0.89
Customer Service	0.88	0.83
Communication	0.90	0.86

Item-Rest Correlations

	Trait #1	Trait #2	Trait #3
Item #1	0.80*		
Item #2	0.80*		
Item #3	0.80*		
Item #4		0.80*	
Item #5		0.80*	
Item #6		0.80*	
Item #7			0.80*
Item #8			0.80*
Item #9			0.80*

*Item-scale correlation, corrected for overlap.

Item-scale correlation matrix

	Depress	Anxiety	Anger
Item #1	0.80*	0.20	0.20
Item #2	0.80*	0.20	0.20
Item #3	0.80*	0.20	0.20
Item #4	0.20	0.80*	0.20
Item #5	0.20	0.80*	0.20
Item #6	0.20	0.80*	0.20
Item #7	0.20	0.20	0.80*
Item #8	0.20	0.20	0.80*
Item #9	0.20	0.20	0.80*

*Item-scale correlation, corrected for overlap.

Item-scale correlation matrix

	Depress	Anxiety	Anger
Item #1	0.50*	0.50	0.50
Item #2	0.50*	0.50	0.50
Item #3	0.50*	0.50	0.50
Item #4	0.50	0.50*	0.50
Item #5	0.50	0.50*	0.50
Item #6	0.50	0.50*	0.50
Item #7	0.50	0.50	0.50*
Item #8	0.50	0.50	0.50*
Item #9	0.50	0.50	0.50*

***Item-scale correlation, corrected for overlap.**

Table 1.4.2 Multitrait/multi-item correlation matrix for patient satisfaction ratings

	Technical	Interpersonal	Communication	Financial
Technical				
1	0.66*	0.63@	0.67@	0.28
2	0.55*	0.54@	0.50@	0.25
3	0.48*	0.41	0.44@	0.26
4	0.59*	0.53	0.56@	0.26
5	0.55*	0.60@	0.56@	0.16
6	0.59*	0.58@	0.57@	0.23
Interpersonal				
1	0.58	0.68*	0.63@	0.24
2	0.59@	0.58*	0.61@	0.18
3	0.62@	0.65*	0.67@	0.19
4	0.53@	0.57*	0.60@	0.32
5	0.54	0.62*	0.58@	0.18
6	0.48@	0.48*	0.46@	0.24
Communication				
1	0.58@	0.59@	0.61*	0.26
2	0.47@	0.50@	0.50*	0.25
3	0.58@	0.66@	0.63*	0.23
4	0.66@	0.66@	0.67*	0.25
5	0.66@	0.71@	0.70*	0.25
Financial				
1	0.35	0.35	0.35	0.72*
2	0.17	0.14	0.15	0.65*
3	0.25	0.23	0.23	0.61*
4	0.18	0.15	0.16	0.67*
5	0.31	0.27	0.29	0.70*
6	0.24	0.23	0.22	0.73*
7	0.25	0.23	0.25	0.55*
8	0.34	0.31	0.31	0.64*
Cronbach's alpha	0.80	0.82	0.82	0.88

Note: Standard error of correlation is 0.03.

Technical = satisfaction with technical quality; Interpersonal = satisfaction with interpersonal aspects; Communication = satisfaction with communication; Financial = satisfaction with financial arrangements.

@ Correlation is within two standard errors of the correlation of the item with its hypothesized scale.

* Item-scale correlation, corrected for overlap.

Table 1.4.3 Correlations between patient satisfaction scales

	Technical	Interpersonal	Communication	Financial
Technical	1.00	0.75	0.76	0.34
Interpersonal	0.93	1.00	0.80	0.31
Communication	0.94	0.98	1.00	0.32
Financial	0.41	0.36	0.38	1.00

Note: Zero-order correlations are provided above the diagonal; correlations adjusting for unreliability of measurement are given below the diagonal.

Confirmatory Factor Analysis

- Observed covariances comapred to covariances generated by hypothesized model
- Statistical and practical tests of fit
- Factor loadings
- Correlations between factors

Fit Indices

- Normed fit index: $\dfrac{\chi^2_{null} - \chi^2_{model}}{\chi^2_{null}}$
- Non-normed fit index: $\dfrac{\frac{\chi^2_{null}}{df_{null}} - \frac{\chi^2_{model}}{df_{model}}}{\left[\frac{\chi^2_{null}}{df_{null}} - 1\right]}$
- Comparative fit index: $1 - \dfrac{\chi^2_{model} - df_{model}}{\chi^2_{null} - df_{null}}$

Table 1.4.4 Standardized parameter estimates for confirmatory factor analytic model

Factor Loadings					
	Technical	**Interpersonal**	**Communication**	**Financial**	**Unique**
Technical					
1	0.77	—	a,b	—	0.64
2	0.60	—	—	—	0.80
3	0.53	—	—	a,b	0.85
4	0.63	—	—	—	0.77
5	0.66	a,b	a	—	0.75
6	0.67	—	—	—	0.74
Interpersonal					
1	—	0.72	—	—	0.69
2	a,b	0.69	a	—	0.72
3	—	0.77	—	—	0.64
4	—	0.64	a,b	a,b	0.76
5	—	0.68	—	—	0.73
6	a	0.52	—	a,b	0.86
Communication					
1	—	—	0.69	—	0.73
2	—	—	0.54	a,b	0.84
3	—	a,b	0.73	—	0.69
4	a,b	—	0.76	—	0.66
5	—	—	0.82	—	0.58
Financial					
1	a	a,b	a	0.77	0.64
2	—	—	—	0.69	0.72
3	—	—	—	0.66	0.76
4	—	—	—	0.72	0.70
5	a	a	a,b	0.75	0.66
6	—	—	—	0.77	0.63
7	a	a	a,b	0.59	0.80
8	a,b	a	a	0.70	0.72

	Technical	**Interpersonal**	**Communication**	**Financial**
Technical	1.00			
Interpersonal	0.93	1.00		
Communication	0.94	0.97	1.00	
Financial	0.40	0.34	0.36	1.00

a – Factor loading that would lead to a significant improvement in fit according to Lagrange multiplier tests, univariate.

b – Factor loading that would lead to a significant improvement in fit according to Lagrange multiplier tests, multivariate.

Table 1.4.5 Goodness-of-fit indexes for confirmatory factor analysis models

Model	df	chi-square	delta	rho	CFI
Null	300	15038.37	—	—	—
Hypothesized four-factor model	269	1864.41	0.88	0.88	0.89
Three-factor model, fixing correlation at 1.0	270	1877.52	0.88	0.88	0.89
Modified four-factor model, adding 14 factor loadings	255	1558.31	0.90	0.90	0.91

Note: All models were statistically rejectable, $P < 0.01$.

Parameter Estimates (standardized) for Confirmatory Factor Analytic Model Estimated Separately Among English and Spanish (in parentheses) Respondents

	Physical Health	*Mental Health*	*Evaluations of Care*
Physical functioning	0.69 (0.25)		
Role–physical	0.81 (0.53)		
Pain	0.69 (0.88)	0.12 (0.02[a])	
General health perceptions	0.56 (0.39)	0.28 (0.34)	
Emotional well-being		0.86 (0.93)	
Role–emotional		0.62 (0.44)	
Energy/fatigue	0.38 (0.22)	0.56 (0.69)	
Social functioning	0.41 (0.45)	0.35 (0.36)	
Doctor			0.96 (0.97)
Overall			0.93 (0.92)
Access			0.81 (0.80)
Wait time			0.75 (0.79)
Choice			0.73 (0.78)
Plan			0.70 (0.70)

NOTE: Estimated correlations for English and Spanish respondents, respectively, were as follows: physical and mental health ($r = .45, .55$), mental health and patient evaluations of care ($r = .23, .31$), and physical health and patient evaluations of care ($r = .05, .16$[a]).

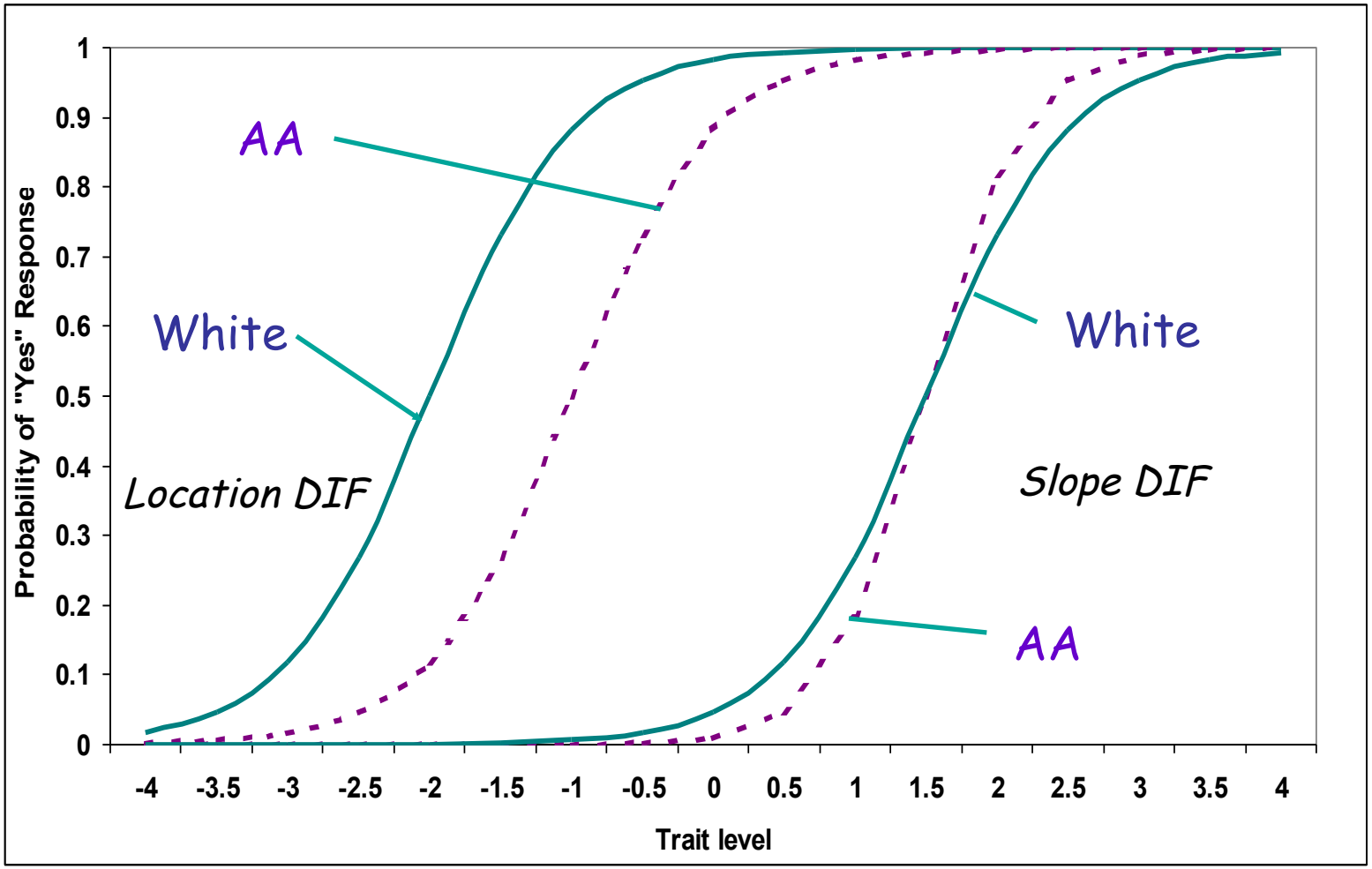
a. Not statistically significant.

Parameter Estimates (standardized) for Confirmatory Factor Analytic Model Constraining Factor Loadings and Correlations to Be Equal for English and Spanish Respondents

	Physical Health	*Mental Health*	*Evaluations of Care*
Physical functioning	0.69 (0.51)		
Role–physical	0.81 (0.78)		
Pain	0.69 (0.68)	0.12 (0.12)	
General health perceptions	0.56 (0.53)	0.28 (0.27)	
Emotional well-being		0.86 (0.80)	
Role–emotional		0.62 (0.57)	
Energy/fatigue	0.37 (0.37)	0.56 (0.56)	
Social functioning	0.41 (0.48)	0.35 (0.41)	
Doctor			0.96 (0.97)
Overall			0.93 (0.93)
Access			0.81 (0.81)
Wait time			0.75 (0.80)
Choice			0.73 (0.75)
Plan			0.70 (0.74)

NOTE: Estimated correlations for English and Spanish respondents, respectively, were as follows: physical and mental health ($r = .45$), mental health and patient evaluations of care ($r = .23$), and physical health and patient evaluations of care ($r = .06$).

Differential Item Functioning
(2-Parameter Model)



Location = uniform; Slope = non-uniform

Thank you.