

Psychometric Evaluation of Multi-Item Scales



Questionnaire Design and Testing Workshop

February 6, 2013, 1-2:30pm

10940 Wilshire Blvd. Suite 710

Los Angeles, CA

Reliability and Validity



Reliable
Not Valid



Low Validity
Low Reliability



Not Reliable
Not Valid



Both Reliable
and Valid

Evaluating Validity

Scale	Age	Obesity	ESRD	Nursing Home Resident
Physical Functioning	Medium (+).	Small (-)	Large (-)	Large (-)
Depressive Symptoms	?	Small (+)	?	Small (+)

Cohen effect size rules of thumb ($d = 0.2, 0.5$, and 0.8):

Small correlation = 0.100

Medium correlation = 0.243

Large correlation = 0.371

$$\underline{r} = \underline{d} / [(\underline{d}^2 + 4)^{\cdot 5}] = \underline{0.8} / [(0.8^2 + 4)^{\cdot 5}] = 0.8 / [(0.64 + 4)^{\cdot 5}] = 0.8 / [(4.64)^{\cdot 5}]$$

$$= 0.8 / 2.154 = \underline{0.371}.$$

Note: Often r' s of $0.10, 0.30$ and 0.50 are cited as small, medium, and large.

Reliability Minimum Standards

- 0.70 or above (for group comparisons)
- 0.90 or higher (for individual assessment)
 - $SEM = SD (1 - \text{reliability})^{1/2}$
 - $95\% \text{ CI} = \text{true score} \pm 1.96 \times SEM$
 - if $z\text{-score} = 0$, then CI: $-.62$ to $+.62$ when reliability = 0.90
 - Width of CI is 1.24 z-score units

Two Raters' Ratings of GOP Debate Performance on *Excellent to Poor* Scale

[1 = Poor; 2 = Fair; 3 = Good; 4 = Very good; 5 = Excellent]

1=Bachman Turner Overdrive (Good, Very Good)

2=Ging Rich (Very Good, Excellent)

3=Rue Paul (Good, Good)

4=Gaylord Perry (Fair, Poor)

5=Romulus Aurelius (Excellent, Very Good)

6=Sanatorium (Fair, Fair)

(*Target = 6 candidates; assessed by 2 raters*)

Cross-Tab of Ratings

		Rater 1					Total
		P	F	G	VG	E	
Rater 2	P	0	1				1
	F		1				1
	G			1			1
	VG			1	0	1	2
	E				1	0	1
Total		0	2	2	1	1	6

Calculating KAPPA

$$P_c = \frac{(0 \times 1) + (2 \times 1) + (2 \times 1) + (1 \times 2) + (1 \times 1)}{(6 \times 6)} = 0.19$$

$$P_{\text{obs.}} = \frac{2}{6} = 0.33$$

$$\text{Kappa} = \frac{0.33 - 0.19}{1 - 0.19} = \mathbf{0.17} \quad (0.52, 0.77)$$

Weighted Kappa (Linear and Quadratic)

	P	F	G	VG	E
P	1	.75 (.937)	.50 (.750)	.25 (.437)	0
F	.75 (.937)	1	.75 (.937)	.50 (.750)	.25 (.437)
G	.50 (.750)	.75 (.937)	1	.75 (.937)	.50 (.750)
VG	.25 (.437)	.50 (.750)	.75 (.937)	1	.75 (.937)
E	0	.25 (.437)	.5 (.750)	.75 (.937)	1

$$W_l = 1 - (i / (k - 1))$$

$$W_q = 1 - (i^2 / (k - 1)^2)$$

i = number of categories ratings differ by

k = n of categories

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

N = n of ratees

WMS = Within Mean Square

k = n of items or raters

JMS = Item or Rater Mean Square

EMS = Ratee x Item (Rater) Mean Square

01 13
01 24
02 14
02 25
03 13
03 23
04 12
04 21
05 15
05 24
06 12
06 22

Reliability of Performance Ratings

Source	df	SS	MS
Candidates (BMS)	5	15.67	3.13
Raters (JMS)	1	0.00	0.00
Cand. x Raters (EMS)	5	2.00	0.40
Total	11	17.67	

$$\text{2-way } R = \frac{6 (3.13 - 0.40)}{6 (3.13) + 0.00 - 0.40} = \boxed{0.89}$$

ICC = 0.80

GOP Presidential Candidates Responses to Two Questions about Their Health

1. Bachman Turner Overdrive (Good, Very Good)
 2. Ging Rich (Very Good, Excellent)
 3. Rue Paul (Good, Good)
 4. Gaylord Perry (Fair, Poor)
 5. Romulus Aurelius (Excellent, Very Good)
 6. Sanatorium (Fair, Fair)
- (Target = 6 candidates; assessed by 2 items)*

01 34
02 45
03 33
04 21
05 54
06 22

Two-Way Fixed Effects (Cronbach's Alpha)

Source	df	SS	MS
Respondents (BMS)	5	15.67	3.13
Items (JMS)	1	0.00	0.00
Resp. x Items (EMS)	5	2.00	0.40
Total	11	17.67	

$$\text{Alpha} = \frac{3.13 - 0.40}{3.13} = \frac{2.93}{3.13} = \boxed{0.87} \quad \text{ICC} = 0.77$$

Overall Satisfaction of 12 Patients with 6 Doctors (2 patients per doctor)

1. Dr. Overdrive (p1: Good, p2: Very Good)
 2. Dr. Rich (p3: Very Good, p4: Excellent)
 3. Dr. Paul (p5: Good, p6: Good)
 4. Dr. Perry (p7: Fair, p8: Poor)
 5. Dr. Aurelius (p9: Excellent, p10: Very Good)
 6. Dr. Sanatorium (p11: Fair, p12: Fair)
- (Target = 6 doctors; assessed by 2 patients each)*

01 13
01 24
02 14
02 25
03 13
03 23
04 12
04 21
05 15
05 24
06 12
06 22

Reliability of Ratings of Doctor

Source	df	SS	MS
Respondents (BMS)	5	15.67	3.13
Within (WMS)	6	2.00	0.33

Total	11	17.67
-------	----	-------

$$1\text{-way} = \frac{3.13 - 0.33}{3.13} = \frac{2.80}{3.13} = \boxed{0.89}$$

Candidates Perceptions of the U.S. Economy in November & December, 2011

1. Bachman Turner Overdrive (Good, Very Good)
 2. Ging Rich (Very Good, Excellent)
 3. Rue Paul (Good, Good)
 4. Gaylord Perry (Fair, Poor)
 5. Romulus Aurelius (Excellent, Very Good)
 6. Sanatorium (Fair, Fair)
- (target = ?; assessed by ?)

How would you estimate reliability?

Two Possibilities

- 1) If target is *GOP* Presidential candidates, then estimate is same as earlier example about responses to two questions about their health
 - $\text{Alpha} = 0.87$
- 2) If target is month (November vs. December), then data structure for 2 observations is:
Nov 3 4 3 2 5 2
Dec 4 5 3 1 4 2

8 Hypothesized GI Distress Scales (101 items)

1. Gastroesophageal reflux (29 items)
2. Disrupted swallowing (7 items)
3. Diarrhea (7 items)
4. Bowel incontinence/soilage (4 items)
5. Nausea and vomiting (13 items)
6. Constipation (15 items)
7. Belly pain (8 items)
8. Gas/bloat/flatulence (18 items)

Gastroesophageal Reflux (29 items)

In the past 7 days ...

- How often did you have regurgitation—that is, food or liquid coming back up into your throat or mouth without vomiting?

- Never
- One day
- 2-6 days
- Once a day
- More than once a day

Disrupted swallowing (7 items)

In the past 7 days ...

- How often did food get stuck in your throat when you were eating?
 - Never
 - Rarely
 - Sometimes
 - Often
 - Always

Diarrhea (7 items)

In the past 7 days ...

- How many days did you have loose or watery stools?

- No days
- 1 day
- 2 days
- 3-5 days
- 6-7 days

Bowel incontinence (4 items)

In the past 7 days ...

- How often did you have bowel incontinence—that is, have an accident because you could not make it to the bathroom in time?

- No days
- 1 day
- 2-3 days
- 4-5 days
- 6-7 days

Nausea and vomiting (13 items)

In the past 7 days ...

- How often did you have nausea—that is, a feeling like you could vomit?
 - Never
 - Rarely
 - Sometimes
 - Often
 - Always

Constipation (15 items)

In the past 7 days ...

- How often did you strain while trying to have bowel movements?

- Never
- Rarely
- Sometimes
- Often
- Always

Belly pain (8 items)

In the past 7 days ...

- How often did you have belly pain?
 - Never
 - Rarely
 - Sometimes
 - Often
 - Always

Gas/bloat/flatulence (18 items)

In the past 7 days ...

- How often did you feel bloated?
 - Never
 - Rarely
 - Sometimes
 - Often
 - Always

Recoding of Item Based on Mean Score for other 28 Items

Gastroesophageal Reflux (Gisx22)

In the last 7 days ...

How often did you burp?

- Never {1}
- One day {2}
- 2-6 days {3}
- Once a day {3}
- More than once a day {4}

Recoding of Item Based on Mean Score for other 14 Items

Constipation (Gisx62)

In the last 7 days ...

**How often did you have a bowel
movement?**

- | | |
|------------------------|-----|
| – Never | {3} |
| – One day | {2} |
| – 2-6 days | {1} |
| – Once a day | {1} |
| – More than once a day | {1} |

Recoding of Item Based on Mean Score for other 17 Items

Gas/Bloat/Flatulence (Gisx105)

In the last 7 days ...

How often did you pass gas?

- Never {1}
- Rarely (only once or twice a day) {2}
- About every 3-4 hours {3}
- About every 2 hours {4}
- About every hour {4}

Internal Consistency Reliability for 8 Scales (100 items)

Scale	Number of items	Alpha
Gastroesophageal Reflux	29	0.94
Disrupted swallowing	7	0.90
Diarrhea	7	0.89
Bowel incontinence	4	0.87
Nausea/vomiting	13	0.93
Constipation	14+	0.93
Belly pain	8	0.93
Gas/bloat	18	0.96

Reliability Standards:

0.70 or above for group comparisons; 0.90 for individual assessment

+ *How often did you have a bowel movement? (Gisx62)* did not go with *Constipation* scale.

Item-scale correlation matrix

	<u>Depress</u>	<u>Anxiety</u>	<u>Anger</u>
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

	<u>Depress</u>	<u>Anxiety</u>	<u>Anger</u>
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.

MULTI -- MULTITRAIT SCALING PROGRAM: gisx105rec item-scale correlations based on separate MULTI.
 SAMPSIZE = 573 SEFCORR = 0.04

item	Gastro	Disrup	Diarr	Bowel	Nause	Consti	Belly	Gas
gisx2	0.70*	0.50	0.18	0.15	0.52	0.23	0.30	0.29
gisx3	0.59*	0.40	0.18	0.17	0.55	0.19	0.31	0.29
gisx4	0.63*	0.49	0.14	0.15	0.50	0.17	0.22	0.22
gisx5	0.65*	0.52	0.11	0.18	0.39	0.20	0.22	0.22
gisx6	0.71*	0.52	0.20	0.17	0.55	0.19	0.27	0.31
gisx7	0.66*	0.48	0.12	0.19	0.42	0.17	0.18	0.20
gisx8	0.74*	0.54	0.19	0.18	0.54	0.24	0.28	0.31
gisx9	0.75*	0.59	0.21	0.17	0.54	0.30	0.31	0.30
gisx10	0.51*	0.42	0.16	0.13	0.33	0.19	0.20	0.14
gisx11	0.68*	0.51	0.11	0.11	0.41	0.22	0.23	0.26
gisx12	0.42*	0.31	0.09	0.11	0.35	0.25	0.21	0.18
gisx13	0.71*	0.55	0.15	0.12	0.40	0.27	0.28	0.31
gisx14	0.65*	0.48	0.13	0.12	0.38	0.26	0.30	0.31
gisx15	0.72*	0.52	0.15	0.15	0.45	0.30	0.34	0.39
gisx16	0.74*	0.51	0.14	0.15	0.44	0.28	0.35	0.36
gisx17	0.56*	0.43	0.07	0.05	0.35	0.26	0.21	0.31
gisx18	0.70*	0.50	0.11	0.11	0.40	0.28	0.25	0.35
gisx19	0.71*	0.50	0.14	0.16	0.40	0.28	0.31	0.37
gisx20	0.44*	0.25	0.12	0.07	0.25	0.11	0.21	0.24
gisx21	0.74*	0.60	0.15	0.16	0.40	0.21	0.30	0.32
gi22rec	0.30*	0.22	0.09	0.05	0.19	0.11	0.12	0.19
gisx23	0.57*	0.39	0.14	0.13	0.39	0.23	0.20	0.31
gisx24	0.59*	0.43	0.14	0.07	0.44	0.31	0.23	0.34
gisx25	0.41*	0.29	0.12	0.19	0.26	0.19	0.15	0.19
gisx26	0.42*	0.29	0.17	0.15	0.36	0.20	0.15	0.19
gisx27	0.46*	0.28	0.11	0.11	0.31	0.19	0.16	0.20
gisx28	0.63*	0.64	0.11	0.13	0.36	0.28	0.25	0.28
gisx29	0.62*	0.62	0.18	0.20	0.42	0.27	0.23	0.27
gisx30	0.63*	0.63	0.14	0.13	0.39	0.27	0.24	0.27
gisx31	0.59	0.71*	0.13	0.18	0.37	0.26	0.18	0.19
gisx32	0.60	0.80*	0.14	0.24	0.38	0.27	0.18	0.25
gisx33	0.68	0.68*	0.14	0.18	0.41	0.32	0.29	0.29
gisx34	0.60	0.81*	0.15	0.26	0.37	0.25	0.18	0.23
gisx35	0.56	0.74*	0.13	0.26	0.35	0.22	0.17	0.25
gisx36	0.56	0.70*	0.09	0.19	0.31	0.24	0.18	0.21
gisx37	0.48	0.62*	0.09	0.19	0.32	0.24	0.16	0.23
gisx38	0.06	0.02	0.69*	0.34	0.20	0.04	0.26	0.18
gisx39	0.06	0.05	0.63*	0.39	0.20	0.11	0.24	0.17
gisx40	0.23	0.14	0.84*	0.46	0.37	0.22	0.40	0.33
gisx41	0.20	0.13	0.83*	0.40	0.34	0.21	0.37	0.31
gisx42	0.20	0.17	0.71*	0.51	0.26	0.20	0.32	0.27
gisx43	0.26	0.20	0.82*	0.47	0.33	0.28	0.38	0.36
gisx44	0.23	0.19	0.80*	0.44	0.32	0.29	0.37	0.34
gisx45	0.13	0.19	0.42	0.76*	0.22	0.10	0.08	0.12
gisx46	0.15	0.23	0.46	0.84*	0.20	0.10	0.04	0.11
gisx47	0.13	0.21	0.41	0.80*	0.17	0.07	0.03	0.09
gisx48	0.28	0.28	0.49	0.60*	0.28	0.19	0.19	0.27

item	Gastro	Disrup	Diarr	Bowel	Nause	Consti	Belly	Gas
gisx49	0.52	0.39	0.30	0.19	0.79*	0.31	0.46	0.37
gisx50	0.51	0.35	0.30	0.16	0.84*	0.32	0.48	0.38
gisx51	0.53	0.35	0.31	0.16	0.82*	0.32	0.49	0.36
gisx52	0.45	0.37	0.23	0.17	0.73*	0.27	0.33	0.31
gisx53	0.52	0.36	0.35	0.16	0.84*	0.35	0.46	0.41
gisx54	0.52	0.35	0.32	0.15	0.84*	0.34	0.49	0.40
gisx55	0.39	0.34	0.31	0.25	0.52*	0.28	0.35	0.30
gisx56	0.40	0.32	0.19	0.21	0.59*	0.21	0.25	0.18
gisx57	0.43	0.34	0.21	0.22	0.63*	0.26	0.26	0.23
gisx58	0.46	0.37	0.20	0.21	0.63*	0.26	0.29	0.25
gisx59	0.47	0.29	0.16	0.18	0.62*	0.15	0.26	0.21
gisx60	0.49	0.31	0.22	0.21	0.67*	0.18	0.28	0.25
gisx61	0.41	0.26	0.21	0.21	0.61*	0.11	0.24	0.23
gisx63	0.22	0.19	-.04	-.00	0.13	0.44*	0.17	0.21
gisx64	0.31	0.29	0.02	0.01	0.22	0.66*	0.32	0.38
gisx65	0.24	0.24	0.09	0.05	0.25	0.74*	0.32	0.35
gisx66	0.25	0.25	0.08	0.04	0.25	0.78*	0.33	0.35
gisx67	0.23	0.22	0.14	0.05	0.25	0.82*	0.37	0.40
gisx68	0.28	0.26	0.21	0.11	0.26	0.79*	0.34	0.37
gisx69	0.29	0.27	0.23	0.09	0.27	0.78*	0.38	0.37
gisx70	0.31	0.25	0.24	0.12	0.30	0.79*	0.39	0.39
gisx71	0.28	0.26	0.24	0.08	0.27	0.82*	0.37	0.37
gisx72	0.24	0.22	0.30	0.16	0.27	0.63*	0.42	0.44
gisx73	0.24	0.17	0.36	0.15	0.31	0.68*	0.45	0.50
gisx74	0.27	0.30	0.17	0.25	0.25	0.43*	0.18	0.23
gisx75	0.24	0.17	0.20	0.21	0.29	0.48*	0.25	0.24
gisx76	0.25	0.18	0.21	0.16	0.30	0.53*	0.26	0.31
gisx77	0.34	0.23	0.35	0.09	0.42	0.41	0.89*	0.55
gisx78	0.25	0.15	0.31	0.09	0.39	0.33	0.81*	0.49
gisx79	0.36	0.21	0.37	0.08	0.45	0.39	0.87*	0.53
gisx80	0.37	0.23	0.35	0.06	0.46	0.43	0.89*	0.52
gi8l89ct	0.31	0.21	0.29	0.04	0.33	0.31	0.72*	0.49
gisx90	0.35	0.21	0.44	0.14	0.48	0.43	0.84*	0.57
gisx91	0.35	0.22	0.39	0.10	0.46	0.47	0.89*	0.58
gisx92	0.37	0.24	0.36	0.11	0.46	0.41	0.83*	0.61
gisx93	0.39	0.30	0.24	0.17	0.32	0.39	0.53	0.82*
gisx94	0.38	0.28	0.24	0.17	0.34	0.38	0.52	0.84*
gisx95	0.37	0.26	0.24	0.15	0.34	0.36	0.51	0.84*
gisx96	0.41	0.28	0.30	0.17	0.41	0.41	0.57	0.81*
gisx97	0.38	0.28	0.24	0.11	0.37	0.42	0.54	0.85*
gisx98	0.31	0.23	0.27	0.14	0.31	0.41	0.53	0.82*
gisx99	0.37	0.26	0.26	0.15	0.35	0.41	0.52	0.88*
gisx100	0.34	0.23	0.23	0.10	0.32	0.41	0.53	0.86*
gisx101	0.38	0.27	0.24	0.10	0.36	0.43	0.55	0.87*
gisx102	0.25	0.23	0.14	0.05	0.26	0.34	0.39	0.65*
gisx103	0.38	0.26	0.26	0.14	0.35	0.43	0.53	0.85*
gisx104	0.38	0.26	0.22	0.08	0.33	0.43	0.51	0.87*
gisx106	0.30	0.18	0.34	0.22	0.25	0.36	0.32	0.56*
gisx107	0.30	0.17	0.35	0.21	0.23	0.40	0.34	0.55*
gisx108	0.21	0.09	0.35	0.11	0.27	0.24	0.42	0.48*
gisx109	0.21	0.11	0.34	0.09	0.28	0.29	0.43	0.49*
gisx110	0.29	0.17	0.33	0.14	0.31	0.37	0.44	0.61*
gi105rec	0.18	0.07	0.19	0.05	0.10	0.25	0.24	0.39*

Five Item-Scale Correlations Issues

1. Gisx22 (burp) correlates only 0.30 with *Gastroesophageal Reflux* scale
- 2, 3, 4. Three items correlate as highly with *Disrupted Swallowing* as with *Gastroesophageal Reflux* scale:
 - Gisx28 (lump in throat frequency)
 - Gisx29 (lump in throat interfere)
 - Gisx30 (lump in throat bother)
5. Gisx33 (pain in chest when swallowing food) correlates as highly with *Gastroesophageal Reflux* as with *Disrupted swallowing* scale

Confirmatory Factor Analysis

	<u>Depress</u>	<u>Anxiety</u>	<u>Anger</u>
Item #1	0.80*	0.00	0.00
Item #2	0.80*	0.00	0.00
Item #3	0.80*	0.00	0.00
Item #4	0.00	0.80*	0.00
Item #5	0.00	0.80*	0.00
Item #6	0.00	0.80*	0.00
Item #7	0.00	0.00	0.80*
Item #8	0.00	0.00	0.80*
Item #9	0.00	0.00	0.80*

*Factor loading.

Confirmatory Factor Analysis Fit Indices

- Normed fit index:
$$\frac{\chi_{null}^2 - \chi_{model}^2}{\chi_{null}^2} = \frac{\frac{\chi_{null}^2}{df_{null}} - \frac{\chi_{model}^2}{df_{model}}}{\left[\frac{\chi_{null}^2}{df_{null}} - 1 \right]}$$
 - Non-normed fit index:
$$\left[\frac{\chi_{null}^2}{df_{null}} - 1 \right]$$
 - Comparative fit index:
$$1 - \left[\frac{\chi_{model}^2 - df_{model}}{\chi_{null}^2 - df_{null}} \right]$$
- RMSEA = SQRT ($\lambda^2 - df$)/SQRT ($df (N - 1)$)

Confirmatory Factor Analysis

- Eight-factor categorical model fit the data well (CFI=0.95 and RMSEA=0.049).
- Lower loadings on *Gastroesophageal Reflux* for:
 - *Gisx22* (burp frequency): 0.38
 - *Gisx20* (pain behind breastbone): 0.39
- Secondary loadings on *Gas* for:
 - *GIsx48* (thought gas but liquid; BI): 0.22
 - *GIsx72* (unfinished frequency; Con.): 0.22
 - *GIsx73* (feeling unfinished bother; Con.): 0.26

Correlated Residuals

- **$r = 0.58$** (Gas/bloat/flatulence)
 - GI108 (freq. gurgling in belly) with GI109 (freq. gurgling when not hungry)
- **$r = 0.53$** (Constipation)
 - GI75 (freq. pain in anus) with GI76 (bother pain in anus)
- **$r = 0.50$** (Gas/bloat/flatulence)
 - GI106 (gas interference) with GI107 (gas bother)

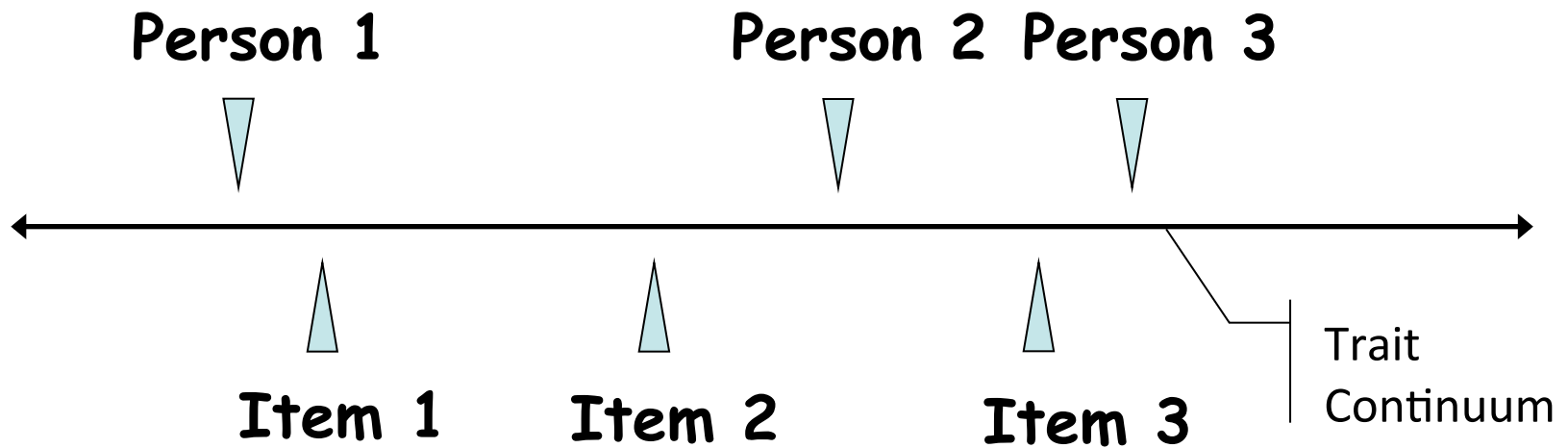
Item Response Theory (IRT)

IRT models the relationship between a person's response Y_i to the question (i) and his or her level of the latent construct θ being measured by positing

$$\Pr(Y_i \geq k) = \frac{1}{1 + \exp(-a_i \theta + b_{ik})}$$

b_{ik} estimates how difficult it is for the item (i) to have a score of k or more and the discrimination parameter a_i estimates the discriminatory power of the item.

Item Responses and Trait Levels



Bowel Incontinence Items

- How often did you soil or dirty your underwear before getting to a bathroom?
- How often did you have bowel incontinence—that is, have an accident because you could not make it to the bathroom in time?
- How often did you leak stool or soil your underwear?
- How often did you think you were going to pass gas, but stool or liquid came out instead?

Graded Response Model Parameters (Incontinence)

Item	Slope	Threshold 1	Threshold 2	Threshold 3	Threshold 4
Gisx46	11.4	0.7	1.3	1.7	2.1
Gisx45	5.0	0.9	1.5	1.9	2.4
Gisx47	4.8	0.7	1.2	1.7	2.1
Gisx48	2.0	0.1	1.0	2.0	3.1

1. How often did you soil or dirty your underwear before getting to a bathroom?
2. How often did you have bowel incontinence—that is, have an accident because you could not make it to the bathroom in time?
3. How often did you leak stool or soil your underwear?

[No days; 1 day; 2-3 days; 4-5 days; 6-7 days]

4. How often did you think you were going to pass gas, but stool or liquid came out instead?

[Never; Rarely; Sometimes; Often; Always]



www.nihpromis.org

- Patient Reported Outcomes Measurement Information System (PROMIS®)
- Funded by the National Institutes of Health
- One domain captured is “anger”
 - Mood (irritability, frustration)
 - Negative social cognitions (interpersonal sensitivity, envy, disagreeableness)
 - Needing to control anger

Reliability (0-1)

- 0.70 or above for group comparisons
- 0.90 or above for individual assessment

z-scores (mean = 0 and SD = 1):

- Reliability = $1 - SE^2$
- So reliability = **0.90** when $SE = 0.32$

T-scores (mean = 50 and SD = 10): $T = 50 + (z * 10)$

- Reliability = $1 - (SE/10)^2$
- So reliability = **0.90** when $SE = 3.2$

In the past 7 days ...

I was grouchy [1st question]

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

Theta = 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

Theta = 51.9 SE = 4.8 (rel. = 0.77)

In the past 7 days ...

I felt angry [3rd question]

- Never
- Rarely
- Sometimes
- Often
- Always

Theta = 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

Theta = 48.8 SE = 3.6 (rel. = 0.87)

In the past 7 days ...

I felt annoyed [5th question]

- Never
- Rarely
- Sometimes
- Often
- Always

Theta = 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

Theta = 50.2 SE = 2.8 (rel = 0.92)

Theta, SEM, and 95% CI

- 56 and 6 (reliability = .68) $W = 22$
- 52 and 5 (reliability = .77) $W = 19$
- 50 and 4 (reliability = .85) $W = 15$
- 49 and 4 (reliability = .87) $W = 14$
- 50 and 3 (reliability = .90) $W = 12$
- 50 and <3 (reliability = .92) $W = 11$

Thank you.

Powerpoint file posted at URL below (freely available for you to use, copy or burn):

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

Contact information:

drhays@ucla.edu 310-794-2294

For a good time call 867-5309 or go to:

<http://twitter.com/RonDHays>

Appendices

ANOVA Computations

- Candidate/Respondents SS
$$(7^2+9^2+6^2+3^2+9^2+4^2)/2 - 38^2/12 = \underline{15.67}$$
- Rater/Item SS
$$(19^2+19^2)/6 - 38^2/12 = \underline{0.00}$$
- Total SS
$$(3^2+ 4^2+4^2+5^2+3^2+3^2+2^2+1^2+5^2+4^2+2^2+2^2) - 38^2/10 = \underline{17.67}$$
- Res. x Item SS= Tot. SS – (Res. SS+Item SS)

```
options ls=130 ps=52 nocenter;  
options nofmterr;
```

```
data one;  
  input id 1-2 rater 4 rating 5;  
CARDS;  
01 13  
01 24  
02 14  
02 25  
03 13  
03 23  
04 12  
04 21  
05 15  
05 24  
06 12  
06 22  
;  
run;  
*****.  
;
```

```

proc freq;
tables rater rating;
run;
*****.
,

proc means;
var rater rating;
run;
*****.
,

proc anova;
class id rater;
model rating=id rater id*rater;
run;
*****.
,

```

```

data one;
input id 1-2 rater 4 rating 5;
CARDS;
01 13
01 24
02 14
02 25
03 13
03 23
04 12
04 21
05 15
05 24
06 12
06 22
;
run;
*****.
%GRIP(indata=one,targetv=id,repeatv=rater,dv=rating,
      type=1,t1=test of GRIP macro,t2=);

```

GRIP macro is available at: <http://gim.med.ucla.edu/FacultyPages/Hays/util.htm>

```
data one;  
  input id 1-2 rater1 4 rater2 5;  
  control=1;  
CARDS;  
01 34  
02 45  
03 33  
04 21  
05 54  
06 22  
;  
run;  
*****.  
DATA DUMMY;  
INPUT id 1-2 rater1 4 rater2 5;  
CARDS;  
01 11  
02 22  
03 33  
04 44  
05 55  
RUN;
```

```

DATA NEW;
SET ONE DUMMY;
PROC FREQ;
TABLES CONTROL*RATER1*RATER2
/NOCOL NOROW NOPERCENT AGREE;
*****.
,

data one;
set one;
*****.
,

proc means;
var rater1 rater2;
run;
*****.
,

proc corr alpha;
var rater1 rater2;
run;

```

Guidelines for Interpreting Kappa

Conclusion	Kappa	Conclusion	Kappa
Poor	< .40	Poor	< 0.0
Fair	.40 - .59	Slight	.00 - .20
Good	.60 - .74	Fair	.21 - .40
Excellent	> .74	Moderate	.41 - .60
		Substantial	.61 - .80
		Almost perfect	.81 - 1.00
Fleiss (1981)		Landis and Koch (1977)	