

Item Response Theory Then 'til Now, and Then Some

David Thissen

L.L. Thurstone Psychometric Laboratory

The University of North Carolina at Chapel Hill

Then (1988) ...

A training session on Multilog at
Rijksuniversiteit Groningen ...



Multilog

1976 - 2010

R I P

Then ...

The Fifth Workshop on Item Response Theory

“An intellectual history of item response theory” ...

Then ...

The Fifth Workshop on Item Response Theory

“An intellectual history of item response theory” ...

A much better history has been published:

Bock, R.D. (1997). A brief history of item response theory. *Educational Measurement: Issues and Practice*, 16, 21-33.

The Fifth Workshop on Item Response Theory

“An intellectual history of item response theory” ...

Some of that material has appeared in:

Thissen, D. & Steinberg, L. (2009). Item response theory. In R. Millsap & A. Maydeu-Olivares, *The Sage handbook of quantitative methods in psychology* (Pp. 148-177). London: Sage Publications.

Thissen, D., & Orlando, M. (2001). Item response theory for items scored in two categories. In D. Thissen & H. Wainer (Eds), *Test Scoring* (Pp. 73-140). Hillsdale, NJ: Lawrence Erlbaum Associates.

Then ...

The Fifth Workshop on Item Response Theory

“Applications of loglinear item response models”

*“Multiple choice models: The distractors are also part
of the item”*

*“Trace lines for testlets: A use of multiple-categorical-
response models”*

Thissen, D. & Mooney, J.A. (1989). Loglinear item response models, with applications to data from social surveys. *Sociological Methodology* 1989, 299-330.

Thissen, D., Steinberg, L. & Fitzpatrick, A.R. (1989). Multiple choice models: The distractors are also part of the item. *Journal of Educational Measurement*, 26, 161-176.

Thissen, D., Steinberg, L. & Mooney, J.A. (1989). Trace lines for testlets: A use of multiple-categorical-response models. *Journal of Educational Measurement*, 26, 247-260.

Questions Following Reading Passages

Passage 1

Items 1-7

Passage 2

Items 8-11

Passage 3

Items 12-14

Passage 4

Items 15-22

from Thissen, Steinberg, & Mooney (1989), "Trace lines for testlets: A use of multiple-categorical-response models"

EFA loadings,
'80s Testfact,
ProMax
rotation,
---- for < 0.25

Passage 1

Items 1-7

Passage 2

Items 8-11

Passage 3

Items 12-14

Passage 4

Items 15-22

0.50	----	----	----
0.29	----	----	----
0.57	----	----	----
0.72	----	----	----
----	----	----	0.40
----	----	----	0.60
----	----	----	0.31
----	0.49	----	----
----	0.92	----	----
0.29	0.60	----	----
----	0.68	----	----
0.59	----	----	----
0.66	----	----	----
0.50	----	----	----
0.40	----	0.29	----
0.48	----	----	----
0.72	----	----	----
----	----	0.86	----
----	----	0.50	----
----	----	0.45	----
----	----	0.38	----
----	----	0.52	----

A better way to look at this came along soon after ...

Gibbons, R. D., & Hedeker, D. (1992). Full-information item bifactor analysis. *Psychometrika*, 57, 423–436.

(Bi-Factor)
 three-
 parameter
 normal
 ogive
 MIRT
 model
 slope
 parameter
 estimates
 (*almost*
 FIML)

Passage 1

Items 1-7

Passage 2

Items 8-11

Passage 3

Items 12-14

Passage 4

Items 15-22

0.52 0.34
 0.31 0.21
 0.66 0.39
 0.77 0.18
 0.39 0.46
 0.46 0.32
 0.45 0.19

0.50
 1.08
 0.78
 0.73

0.76
 0.85
 0.46

0.54
 0.71
 0.79
 0.94
 0.48
 0.54
 0.49
 0.76

0.40
 1.53
 0.58
 0.76

0.44
 0.39
 0.21

0.15
 0.05
 -0.09
 0.73
 0.34
 0.31
 0.27
 0.46

Gibbons & Hedeker
 (1992) "Full-
 information item
 bifactor analysis"

Our first attempt to estimate all of the parameters came after we had “MultiNorm”:

Edwards, M. C. (2005). *A Markov chain Monte Carlo approach to confirmatory item factor analysis.*

Unpublished doctoral dissertation, Department of Psychology, University of North Carolina at Chapel Hill.

(Bi-Factor)
 three-
 parameter
 normal
 ogive
 MIRT
 model
 slope
 parameter
 estimates
 (MCMC)

Passage 1

Items 1-7

Passage 2

Items 8-11

Passage 3

Items 12-14

Passage 4

Items 15-22

0.56	0.20
0.32	0.12
0.67	0.23
0.80	0.13
0.44	0.64
0.49	0.45
0.46	0.29

0.51	
1.02	
0.79	
0.75	

0.99	
0.87	
0.54	

0.54	
0.71	
0.77	
1.00	
0.48	
0.55	
0.50	
0.76	

0.40
1.41
0.59
0.79

0.88
0.30
0.34

0.16
0.09
0.04
0.86
0.36
0.32
0.26
0.48

Edwards (2005) “A Markov chain Monte Carlo approach to confirmatory item factor analysis”

Refined
(Bi-Factor)
three-
parameter
normal
ogive
MIRT
model
slope
parameter
estimates
(MCMC)

Passage 1

Items 1-7

Passage 2

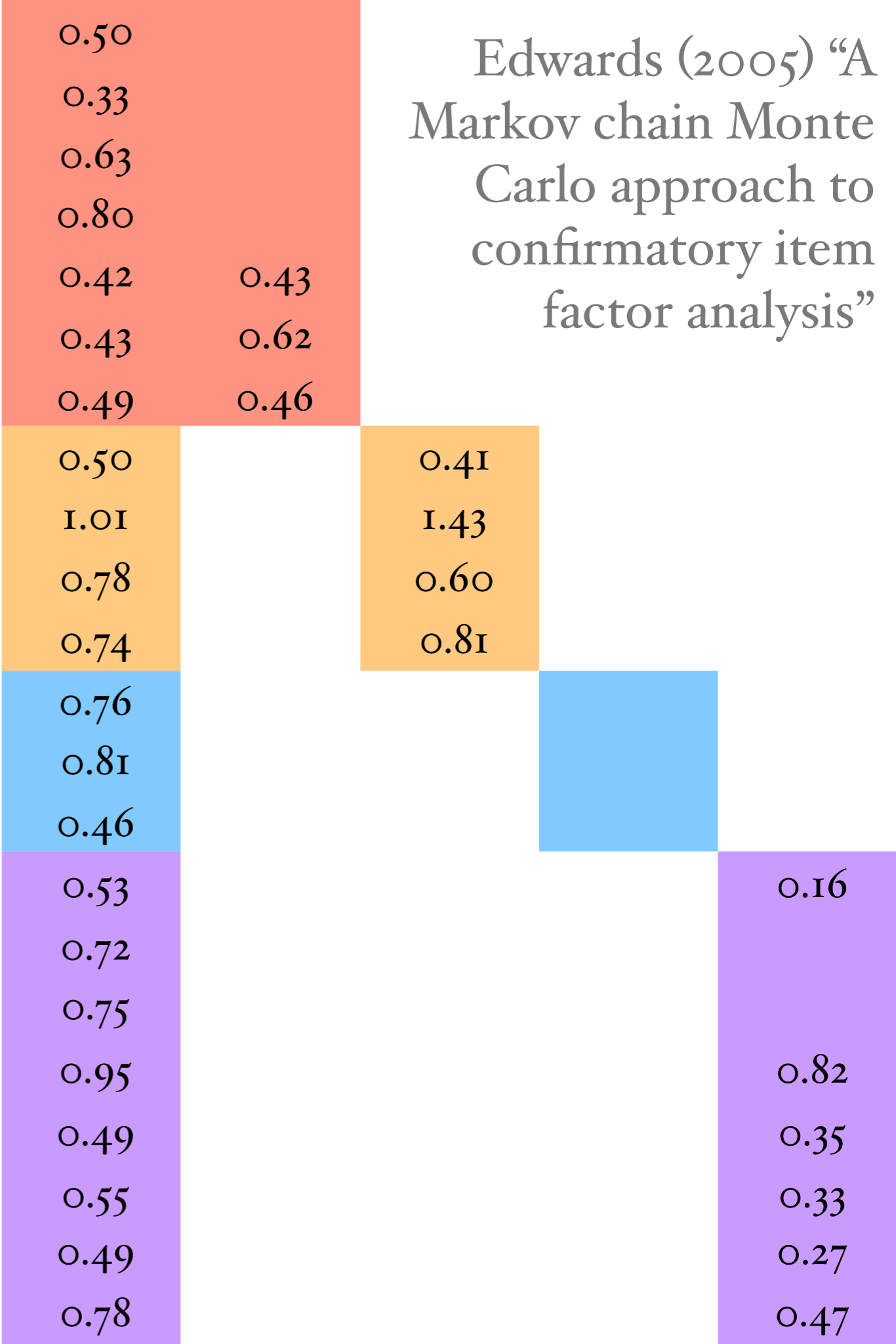
Items 8-11

Passage 3

Items 12-14

Passage 4

Items 15-22



Edwards (2005) "A Markov chain Monte Carlo approach to confirmatory item factor analysis"

Thissen, D. & Steinberg, L. (in press). Using item response theory to disentangle constructs at different levels of generality. In S. Embretson (Ed.), *Measuring psychological constructs: Advances in model-based approaches*. Washington, DC: American Psychological Association.

Subsequently Li Cai implemented in the forthcoming IRTPRO software the bifactor algorithm, along with “supplemented EM” (SEM) computation of the item parameter error covariance matrix and standard errors, from:

Cai, L. (2008). SEM of another flavour: Two new applications of the supplemented EM algorithm. *British Journal of Mathematical and Statistical Psychology*, 61, 309–329.

Meng, X., and Rubin, D. B. (1991). Using EM to obtain asymptotic variance-covariance matrices: The SEM algorithm. *Journal of the American Statistical Association*, 86, 899-909.

(Bi-Factor)
 three-
 parameter
 normal
 ogive
 MIRT
 model
 slope
 parameter
 estimates
 (FIML)

Passage 1

Items 1-7

Passage 2

Items 8-11

Passage 3

Items 12-14

Passage 4

Items 15-22

0.47 0.16
 0.31 0.11
 0.63 0.23
 0.75 0.13
 0.42 0.59
 0.51 0.45
 0.47 0.31

0.50
 1.21
 0.81
 0.76

1.64
 0.81
 0.41

0.51
 0.75
 0.78
 0.92
 0.47
 0.52
 0.50
 0.74

0.16
 0.11
 0.23
 0.13
 0.59
 0.45
 0.31

IRTPRO α
 logistic graded model
 slopes / 1.7

0.39
 1.71
 0.59
 0.76

2.17
 0.18
 0.14

0.14
 0.05
 -0.10
 0.72
 0.33
 0.29
 0.28
 0.44

Meanwhile as a closely-related alternative, Wainer, Bradlow, & Wang suggested “testlet response theory” and provided an MCMC implementation in SCORIGHT:

Bradlow, E. T., Wainer, H., & Wang, X. (1999). A Bayesian random effects model for testlets. *Psychometrika*, 64, 153-168.

...

Wainer, H., Bradlow, E. T., & Wang, X. (2007). *Testlet response theory and its applications*. New York, NY: Cambridge University Press.

Glas, Wainer, & Bradlow described MML estimation for the testlet response model, and (a bit subtly) pointed out that the testlet response model is second-order factor analysis. Put together with the relations between hierarchical (bifactor) and second order factor analysis described by Yung, McLeod, & Thissen, the forthcoming IRTPRO can do that as well.

Glas, C.A.W., Wainer, H., & Bradlow, E. T. (2000). Maximum marginal likelihood and expected a posteriori estimation in testlet-based adaptive testing. In W. J. van der Linden & C. A. W. Glas (Eds.), *Computerized adaptive testing: Theory and practice* (pp. 271-288). Boston, MA: Kluwer Academic Publishers.

Yung, Y.F., McLeod, L.D., & Thissen, D. (1999). On the relationship between the higher-order factor model and the hierarchical factor model. *Psychometrika*, 64, 113-128.

[(Constrained)
Bi-Factor]:

Passage 1

Items 1-7

0.46 0.46
0.30 0.30
0.62 0.62
0.76 0.76
0.43 0.43
0.52 0.52
0.48 0.48

IRTPRO α
logistic graded model
slopes / 1.7

$$\sigma^2(\theta_2) = 0.31 (.06)$$

Passage 2

Items 8-11

0.48
1.10
0.79
0.78

0.48
1.10
0.79
0.78

$$\sigma^2(\theta_3) = 0.90 (.14)$$

Passage 3

Items 12-14

0.82
0.88
0.39

0.82
0.88
0.39

$$\sigma^2(\theta_4) = 0.27 (.05)$$

Passage 4

Items 15-22

0.50
0.70
0.59
0.91
0.51
0.55
0.52
0.77

0.50
0.70
0.59
0.91
0.51
0.55
0.52
0.77

$$\sigma^2(\theta_5) = 0.18 (.03)$$

One can still do exploratory factor analysis (EFA) ...

Left values:
EFA loadings,
'80s Testfact,
Promax
rotation, ----
for < 0.25

Passage 1

Items 1-7

Passage 2

Items 8-11

Passage 3

Items 12-14

Passage 4

Items 15-22

Right values:
EFA loadings,
IRTPRO,
direct
Quartimin,
gray for < 0.25

0.50	0.44	----	0.01	----	-0.01	----	0.14
0.29	0.30	----	0.04	----	-0.06	----	0.14
0.57	0.50	----	0.03	----	-0.01	----	0.19
0.72	0.58	----	-0.02	----	0.03	----	0.09
----	0.20	----	0.05	----	0.04	0.40	0.44
----	0.08	----	0.16	----	0.17	0.60	0.39
----	0.10	----	0.16	----	0.16	0.31	0.29
----	0.13	0.49	0.45	----	0.00	----	0.09
----	-0.07	0.92	0.94	----	-0.01	----	0.07
0.29	0.32	0.60	0.58	----	-0.05	----	-0.12
----	0.02	0.68	0.65	----	0.16	----	-0.10
0.59	0.46	----	0.00	----	0.19	----	0.08
0.66	0.61	----	0.05	----	0.09	----	-0.14
0.50	0.36	----	-0.02	----	0.05	----	0.03
0.40	0.29	----	0.03	0.29	0.26	----	-0.13
0.48	0.40	----	0.11	----	0.18	----	-0.03
0.72	0.53	----	0.00	----	0.05	----	0.05
----	0.00	----	-0.02	0.86	0.80	----	0.01
----	0.00	----	0.10	0.50	0.45	----	-0.01
----	0.09	----	-0.01	0.45	0.43	----	0.14
----	0.18	----	-0.05	0.38	0.39	----	0.01
----	0.07	----	0.17	0.52	0.51	----	0.01

... but EFA is fragile: slow to converge in any event, and most often with unselected item tryout sets more factors are needed than can comfortably be fitted ...

When the test does not have an obvious structure (like the passage-based reading comprehension test), an alternative to EFA is to use local dependence (LD) diagnostic statistics computed after fitting a unidimensional model to start to build a multidimensional model:

Chen, W.H. & Thissen, D. (1997). Local dependence indices for item pairs using item response theory. *Journal of Educational and Behavioral Statistics*, 22, 265-289.

34 Asthma-related Quality of Life Items

Because IRTPRO was not yet ready at the time we did the data analysis for these items, we used (many repeated) model-fitting(s) with structural equations model software to (WLSMV in *Mplus*; augmented with FIML) to investigate the local dependence.

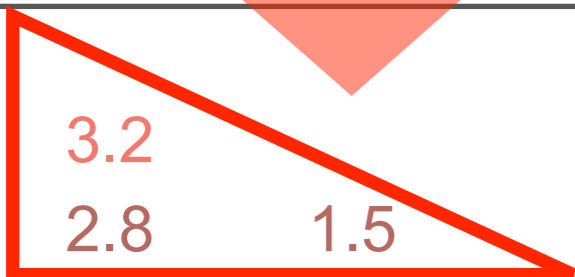
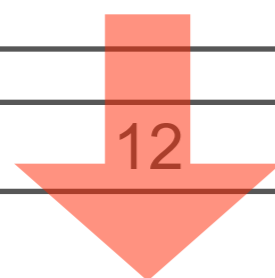
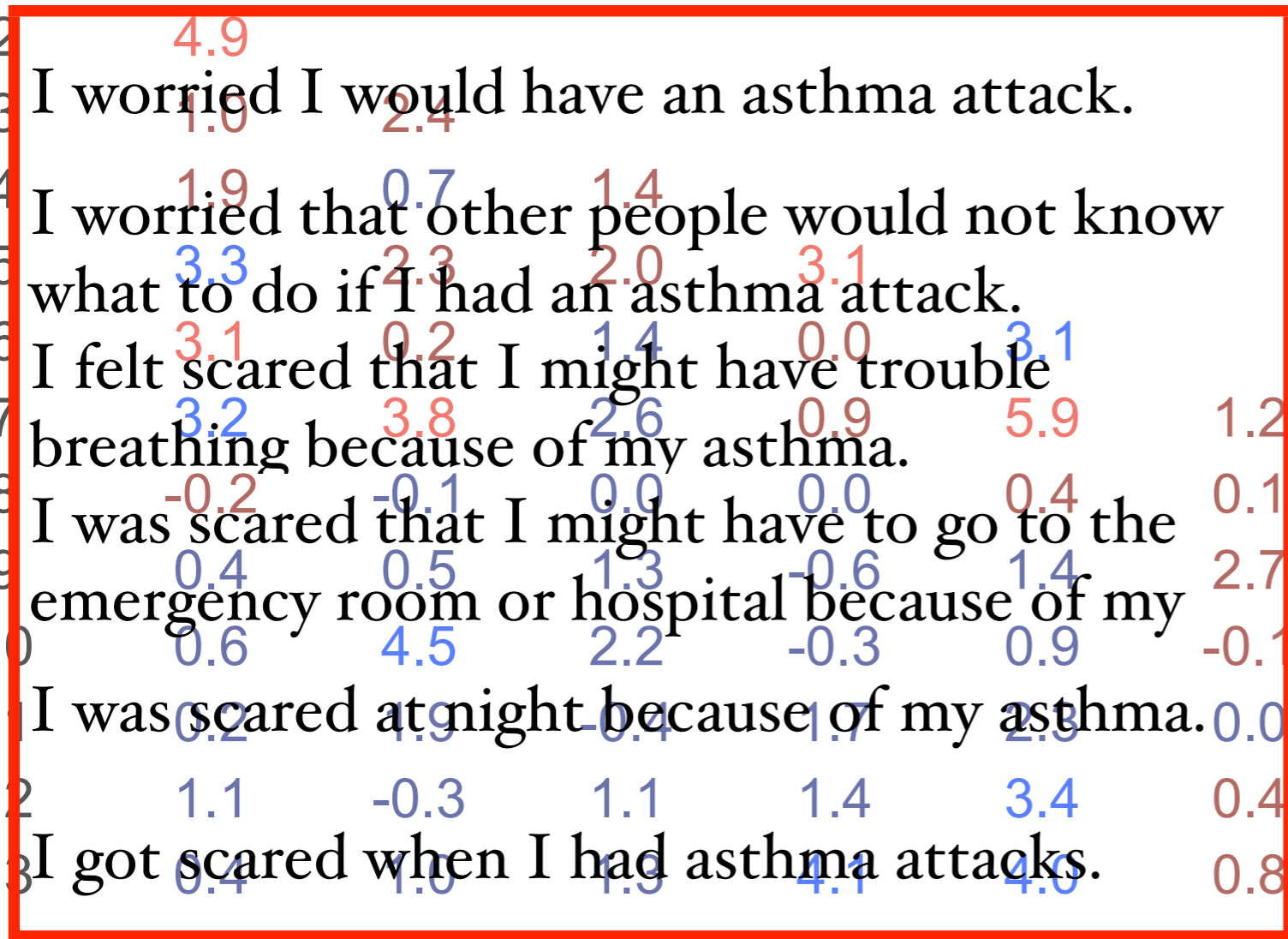
We eventually built a (nearly) bi-factor model with four factors and four additional error correlations (between pairs of items).

The LD statistics computed after fitting a unidimensional model with IRTPRO may make that process easier ...

Standardized LD χ^2 Statistics

Item	1	2	3	4	5	6	7	8	9	10
1										
2	4.9									
3	1.9	2.4								
4	1.9	0.7	1.4							
5	3.3	2.3	2.0	3.1						
6	3.1	0.2	1.4	0.0	3.1					
7	3.2	3.8	2.6	0.9	5.9	1.2				
8	-0.2	-0.1	0.0	0.0	0.4	0.1	2.6			
9	0.4	0.5	1.3	-0.6	1.4	2.7	3.2	5.6		
10	0.6	4.5	2.2	-0.3	0.9	-0.1	1.1	3.6	0.0	
11	0.2	1.9	-0.4	1.7	2.9	0.0	3.7	5.0	4.2	1.7
12	1.1	-0.3	1.1	1.4	3.4	0.4	0.1	1.8	2.1	2.0
13	0.4	1.0	1.9	4.1	4.0	0.8	2.3	5.2	2.6	2.2

Item	11	12	13	14	15	16	17	18	19	20
11										
12	3.2									
13	2.8	1.5								



Standardized LD X^2 Statistics

Item	11	12	13	14	15	16	17	18	19	20
...										
16	3.2	2.7	0.5	2.2	3.4					
17	0.0	1.4	2.6	2.7	3.1	0.1				
18	1.3	0.1	2.5	0.7	1.9	-0.2	-0.2			
19	0.9	1.3	0.5	2.2	2.3	0.2	1.0	30.1		
20	1.8	0.4	1.3	6.0	1.0	1.7	1.4	-0.2	1.5	
21	4.6	1.9	1.9	4.5	1.8	0.7	1.1	1.9	2.4	8.1
22	1.3	1.8	2.7	1.5	0.6	1.8	0.7	-0.9	2.0	3.5
23	0.6	5.3	1.8	3.8	0.4	4.9	3.0	3.1	0.5	5.9
24	0.1	1.7	12.3	1.0	1.1	0.8	3.5	0.9	2.1	6.0
25	0.6	3.4	4.5	1.4	2.6	1.5	18.3	2.1	3.4	4.7
26	2.1	1.9	3.2	4.7	3.1	0.7	-0.5	0.0	-0.3	6.8
27	-0.5	0.8	6.7	0.4	1.1	1.0	1.0	2.7	0.2	1.6
...										
Item	21	22	23	24	25	26	27	28	29	30
...										
26	2.6	0.6	0.1	2.4	2.4					
27	5.9	3.3	2.9	12.6	1.8	1.1				
...										

I got scared when I had asthma attacks.
 Asthma attacks bothered me.
 I had asthma attacks.

MML Estimation A.B.A. (after Bock-Aitkin)

for bifactor models:

Gibbons, R.D., Bock, R.D., Hedeker, D., Weiss, D.J., Segawa, E., Bhaumik, D.K., Kupfer, D.J., Frank, E., Grochocinski, V.J., & Stover, A. (2007). Full-information item bifactor analysis of graded response data. *Applied Psychological Measurement*, 31, 4-19.

MML Estimation A.B.A. (after Bock-Aitkin)

for models with moderate latent dimensionality—ADQ:

Schilling, S., & Bock, R. D. (2005). High-dimensional maximum marginal likelihood item factor analysis by adaptive quadrature. *Psychometrika*, 70, 533–555

MML Estimation A.B.A. (after Bock-Aitkin)

for models with high latent dimensionality—MHRM:

Cai, L. (in press). High-dimensional exploratory item factor analysis by a Metropolis-Hastings Robbins-Monro algorithm. *Psychometrika*.

Cai, L. (in press) Metropolis-Hastings Robbins-Monro Algorithm for confirmatory item factor analysis. *Journal of Educational and Behavioral Statistics*.

Item	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8
9	0.57	0.52						
8	0.69	0.47						
11	0.69	0.42						
12	0.75	0.36						
13	0.63	0.32	0.44					
10	0.76	0.25						
24	0.76		0.53					
27	0.71		0.40					
4	0.66			0.46				
23	0.76			0.46				
22	0.73			0.42				
16	0.76			0.32				
5	0.69			0.29				
33	0.79			0.26				
17	0.74				0.51			
25	0.78				0.47			
18	0.48					0.78		
19	0.49					0.77		
28	0.82						0.32	
20	0.86						0.29	
32	0.70							0.46
31	0.78							0.40
...	(11 more unidimensional items)							

MHRM

Factor Loadings Metric

An eight factor model
 (4 “real” factors, 4 doublets)
 for the 34 Asthma QOL items

Item	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8
9	0.56	0.52						
8	0.68	0.46						
11	0.69	0.42						
12	0.75	0.37						
13	0.62	0.33	0.45					
10	0.76	0.26						
24	0.76		0.51					
27	0.71		0.42					
4	0.66			0.47				
23	0.76			0.46				
22	0.72			0.43				
16	0.75			0.34				
5	0.69			0.30				
33	0.78			0.26				
17	0.73				0.51			
25	0.77				0.47			
18	0.54					0.72		
19	0.54					0.72		
28	0.81						0.32	
20	0.85						0.29	
32	0.70							0.46
31	0.78							0.41
...	(11 more unidimensional items)							

ADQ Factor Loadings Metric

Item	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8
9	0.01	0						
8	0.01	0.01						
11	0	0						
12	0	-0.01						
13	0.01	-0.01	-0.01					
10	0	-0.01						
24	0		0.02					
27	0		-0.02					
4	0			-0.01				
23	0			0				
22	0.01			-0.01				
16	0.01			-0.02				
5	0			-0.01				
33	0.01			0				
17	0.01				0			
25	0.01				0			
18	-0.06					0.06		
19	-0.05					0.05		
28	0.01						0	
20	0.01						0	
32	0							0
31	0							-0.01
...	(11 more unidimensional items)							

Differences
MHRM - ADQ estimates
Factor Loadings Metric

Item	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8
9	0.01	0						
8	0.01	0.01						
11	0	0						
12	0	-0.01						
13	0.01	-0.01	-0.01					
10	0	-0.01						
24	0		0.02					
27	0		-0.02					
4	0			-0.01				
23	0			0				
22	0.01			-0.01				
16	0.01			-0.02				
5	0			-0.01				
33	0.01			0				
17	0.01				0			
25	0.01				0			
18	-0.06					0.06		
19	-0.05					0.05		
28	0.01						0	
20	0.01						0	
32	0							0
31	0							-0.01
...	(11 more unidimensional items)							

Differences
MHRM - ADQ estimates
Factor Loadings Metric

Item	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8
9	0.57	0.52						
8	0.69	0.47						
11	0.69	0.42						
12	0.75	0.36						
13	0.63	0.32	0.44					
10	0.76	0.25						
24	0.76		0.53					
27	0.71		0.40					
4	0.66			0.46				
23	0.76			0.46				
22	0.73			0.42				
16	0.76			0.32				
5	0.69			0.29				
33	0.79			0.26				
17	0.74				0.51			
25	0.78				0.47			
18	0.48					0.78		
19	0.49					0.77		
28	0.82						0.32	
20	0.86						0.29	
32	0.70							0.46
31	0.78							0.40
...	(11 more unidimensional items)							

MHRM
4 minutes

Item	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8
9	0.56	0.52						
8	0.68	0.46						
11	0.69	0.42						
12	0.75	0.37						
13	0.62	0.33	0.45					
10	0.76	0.26						
24	0.76		0.51					
27	0.71		0.42					
4	0.66			0.47				
23	0.76			0.46				
22	0.72			0.43				
16	0.75			0.34				
5	0.69			0.30				
33	0.78			0.26				
17	0.73				0.51			
25	0.77				0.47			
18	0.54					0.72		
19	0.54					0.72		
28	0.81						0.32	
20	0.85						0.29	
32	0.70							0.46
31	0.78							0.41
...	(11 more unidimensional items)							

ADQ
7 hours

Item	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8
9	0.56	0.52						
8	0.68	0.46						
11	0.69	0.42						
12	0.75	0.37						
13	0.62	0.33	0.45					
10	0.76	0.26						
24	0.76		0.51					
27	0.71		0.42					
4	0.66			0.47				
23	0.76			0.46				
22	0.72			0.43				
16	0.75			0.34				
5	0.69			0.30				
33	0.78			0.26				
17	0.73				0.51			
25	0.77				0.47			
18	0.54					0.72		
19	0.54					0.72		
28	0.81						0.32	
20	0.85						0.29	
32	0.70							0.46
31	0.78							0.41
...	(11 more unidimensional items)							

Item 13, on 3 factors, is all that “breaks” the otherwise-bifactor nature of this model.

Item	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8
9	0.56	0.52						
8	0.68	0.46						
11	0.69	0.42						
12	0.75	0.37						
13	0.62	0.33	0.45					
10	0.76	0.26						
24	0.76		0.51					
27	0.71		0.42					
4	0.66			0.47				
23	0.76			0.46				
22	0.72			0.43				
16	0.75			0.34				
5	0.69			0.30				
33	0.78			0.26				
17	0.73				0.51			
25	0.77				0.47			
18	0.54					0.72		
19	0.54					0.72		
28	0.81						0.32	
20	0.85						0.29	
32	0.70							0.46
31	0.78							0.41
...	(11 more unidimensional items)							

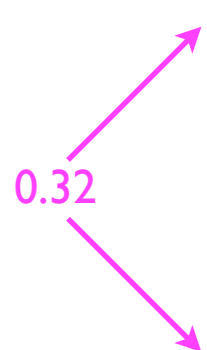
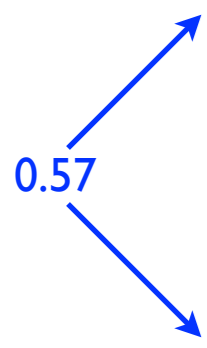
Li Cai's generalized dimension-reduction can fit this with Bock-Aitkin EM and three-dimensional quadrature.

Some Social Quality of Life Items (UNC-CH PROMIS Pediatric Project)

(Modified) “Bi-Factor” or “Hierarchical” Model

ADQ

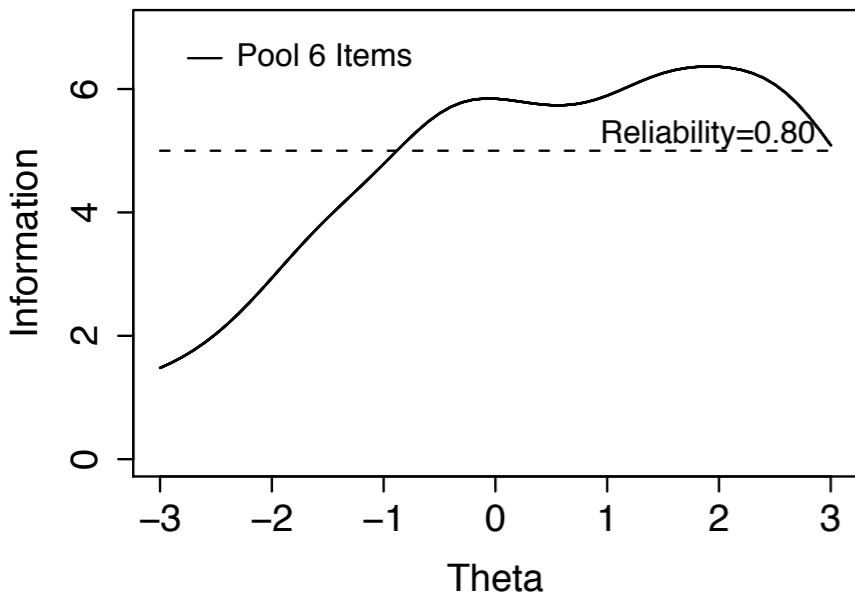
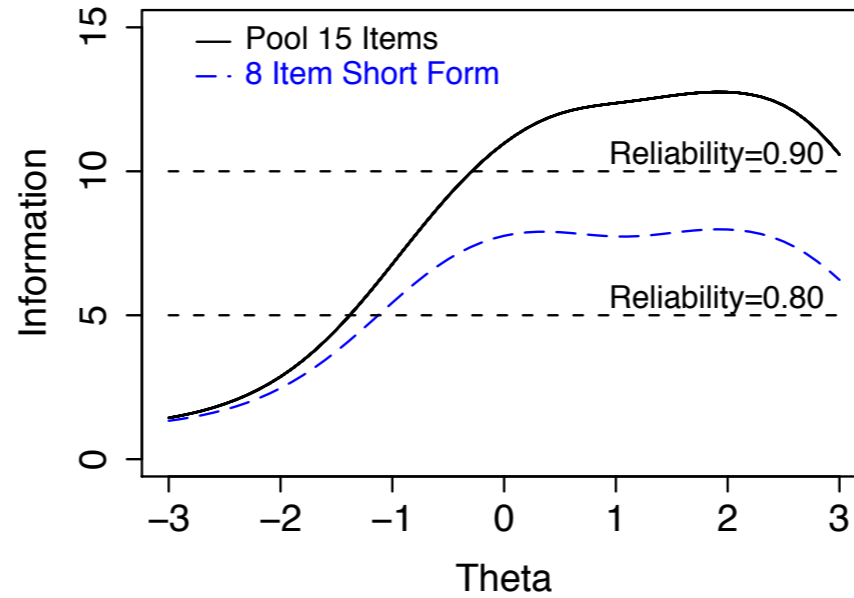
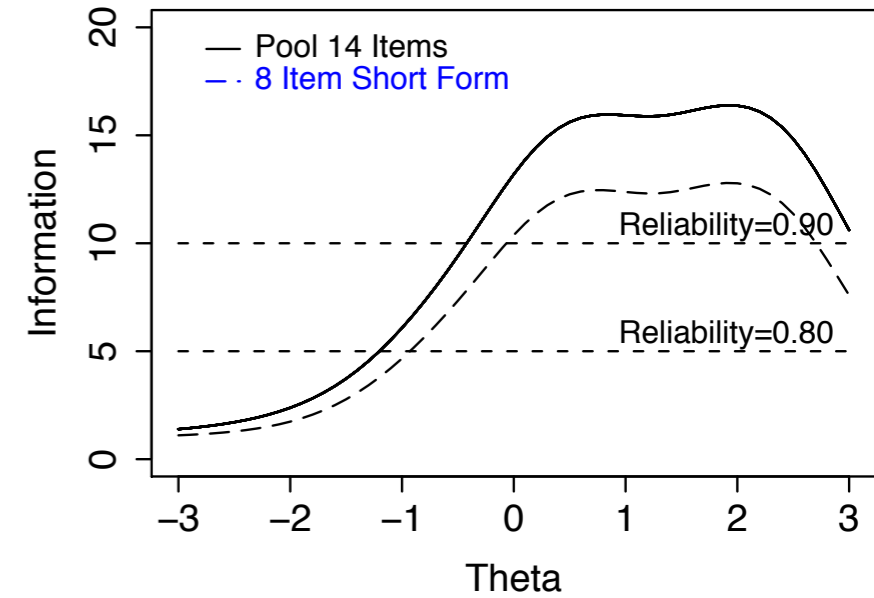
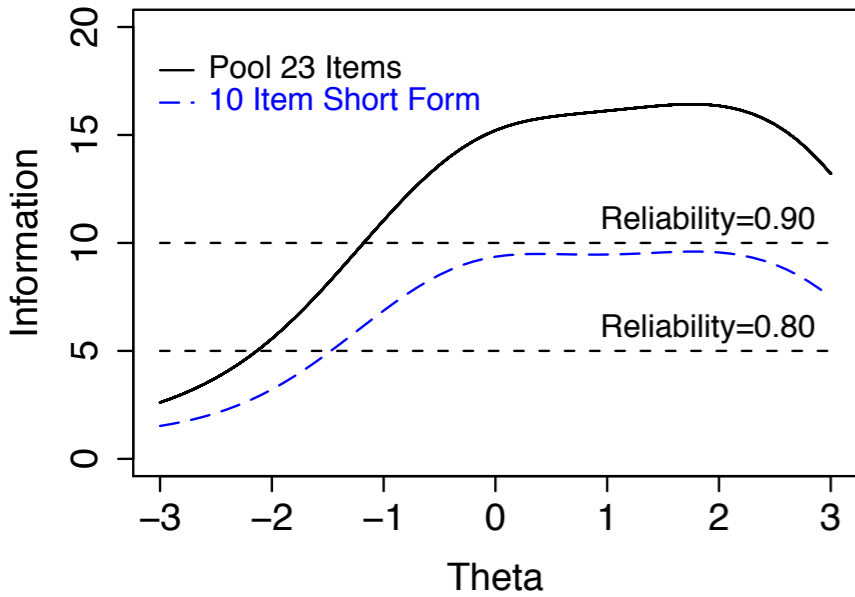
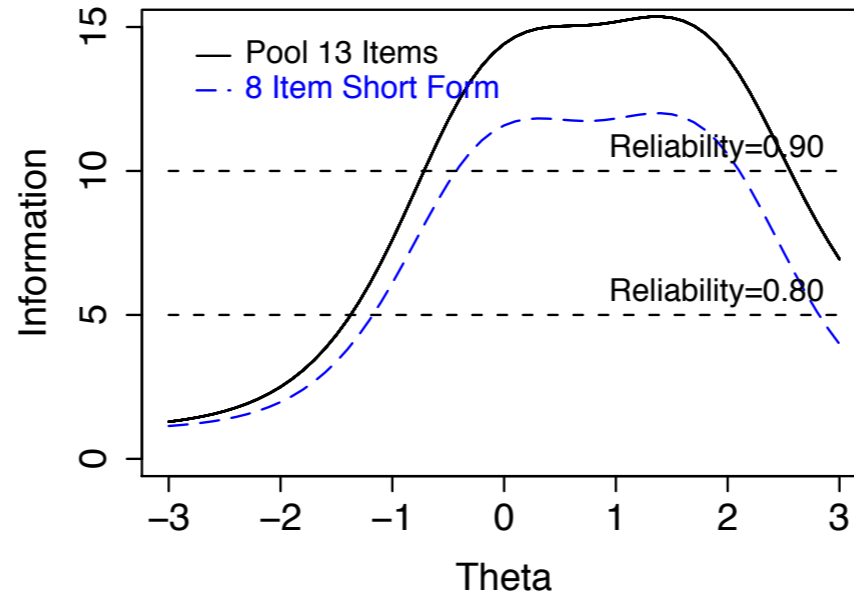
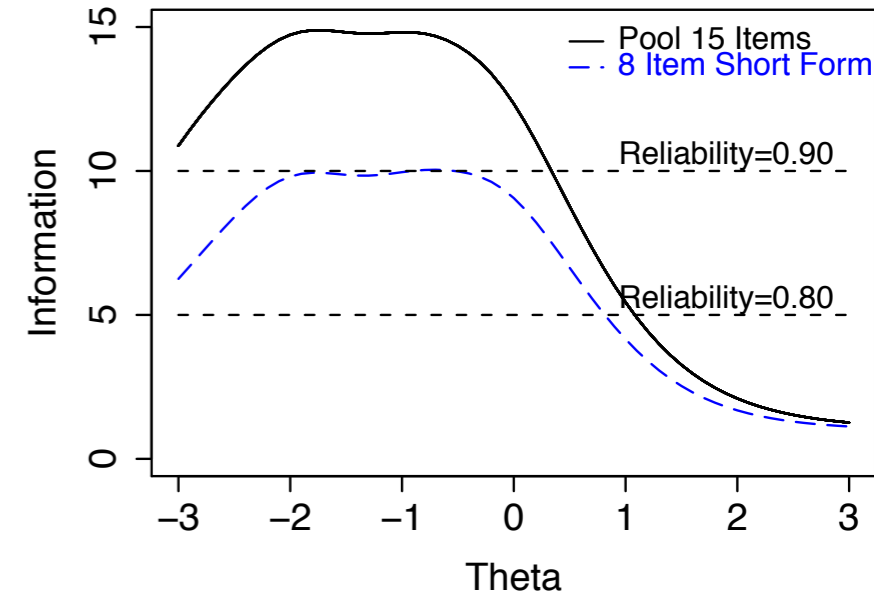
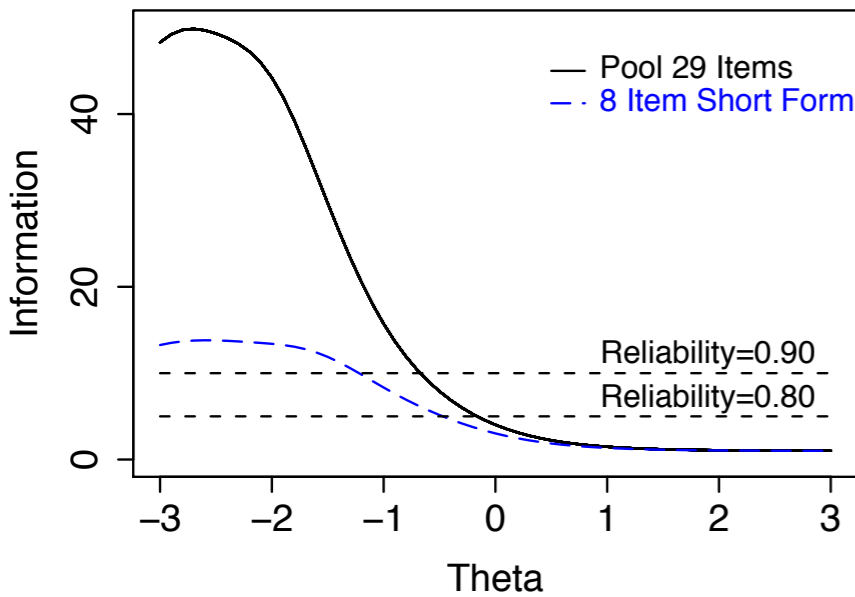
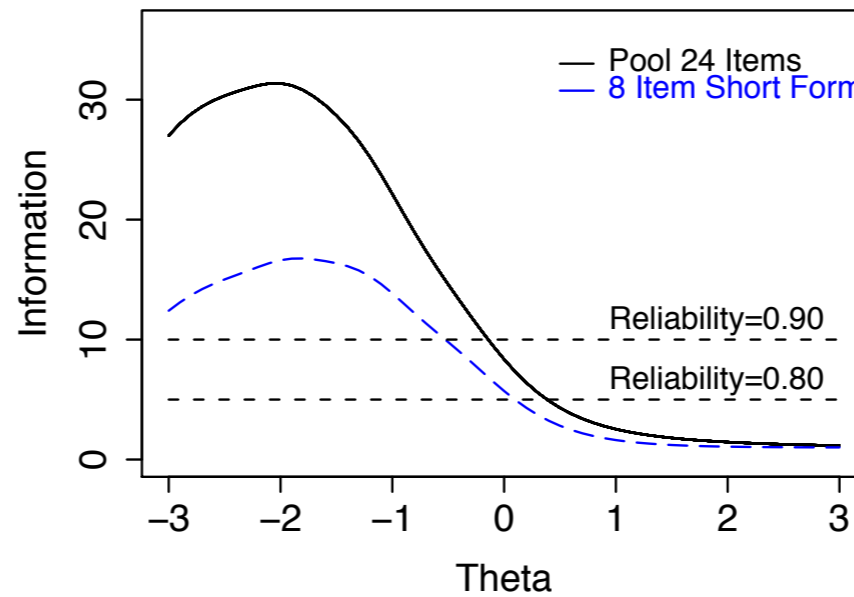
Item stem (responses “never” to “almost always”)	λ_g	λ_1	λ_2	λ_3	λ_4
I liked being around other kids my age.	0.54	0.63			
I treated other kids my age with respect.	0.43	0.46			0.41
I was able to talk about everything with my friends.	0.42	0.48		0.44	
I spent time with my friends.	0.49	0.42			
I was able to count on my friends.	0.56	0.44		0.50	
I was good at talking with other kids my age.	0.70	0.42			
I felt close to my friends.	0.55	0.44		0.54	
I had enough time to be with my friends.	0.49	0.21			
Other kids teased me.	0.48		0.70		
Other kids did not like me.	0.52		0.55		
Other kids do not want to be my friend.	0.60		0.53		
Other kids made me feel different because of how I look.	0.56		0.50		
Other kids bullied me.	0.49		0.60		
It was hard for me to make friends.	0.66		0.41		
I worried about losing a friendship.	0.39		0.33		
I felt bad about how I got along with classmates.	0.55		0.33		
I teased other kids.	0.27		0.35		0.69
I felt nervous about asking questions in class.	0.39		0.30		
I felt nervous when I was in a group of other kids my age.	0.59		0.24		
I was mean to other people.	0.28		0.28		0.75
I played alone and kept to myself.	0.52		0.21		
I got into fights (hitting, kicking, pushing) with other kids.	0.41		0.17		0.46



The UNC-CH PROMIS Pediatric Project

Beginning-to-end Summary

Domain & subdomain	Tryout Set (# of items)	Final Pool (# of items)
Emotional Distress-Anger	10	6
Emotional Distress-Anxiety	18	15
Emotional Distress-Depressive Symptoms	21	14
Fatigue-Tired	23	23
<i>Fatigue-Lack of Energy</i>	39	11
Pain	27	13
Physical Function-Lower Extremity / Mobility	32	24
Physical Function-Upper Extremity / Dexterity	38	29
Social Health / Functioning with Peers	74	15

Anger 6 Item Pool**Anxiety Pool & Short Form****Depressive Symptoms Pool & Short Form****Fatigue Pool & Short Form****Pain Pool & Short Form****Social Health Pool & Short Form****PF Upper Extremity Pool & Short Form****PF Lower Extremity Pool & Short Form**

Depressive Symptoms 8-item Short Form

I felt like I couldn't do anything right.

I felt everything in my life went wrong.

I felt unhappy.

I felt lonely.

I felt sad.

I felt alone.

I thought that my life was bad.

I could not stop feeling sad.

Response Alternatives:

Never

Almost Never

Sometimes

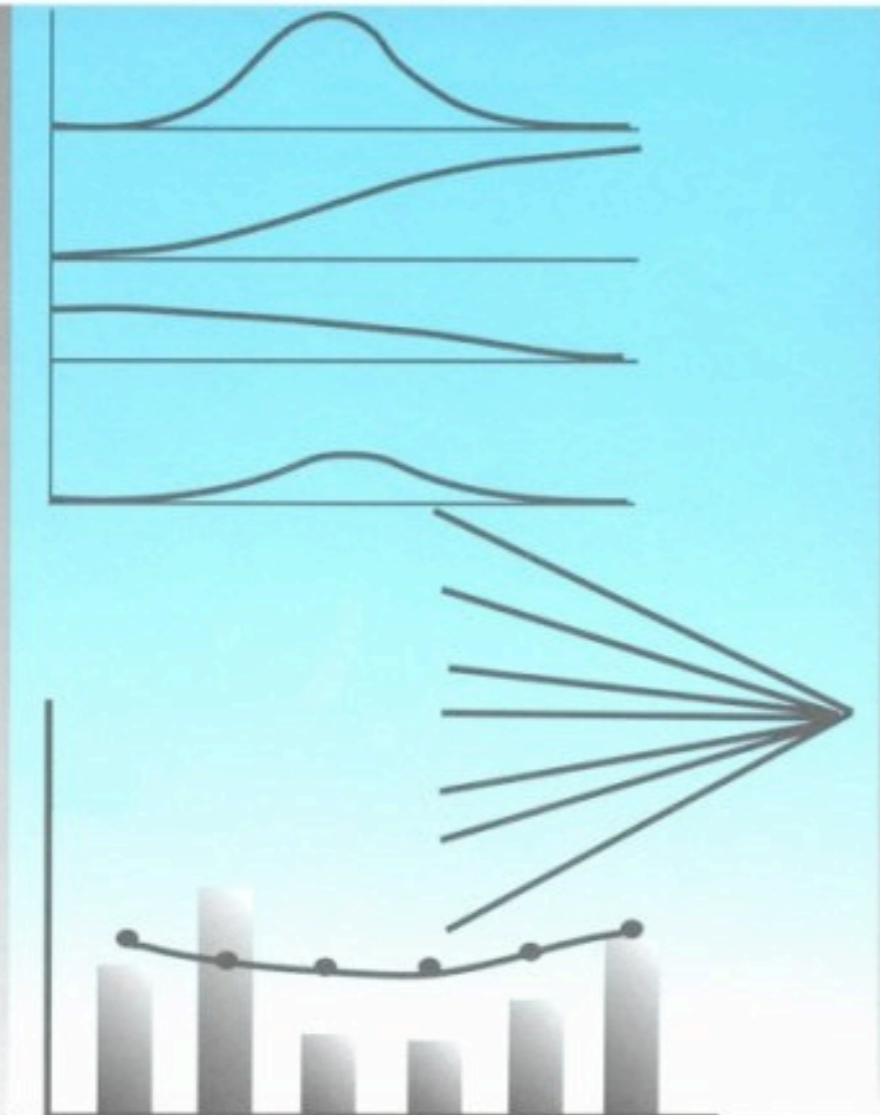
Often

Almost Always

Depressive Symptoms 8-item Short Form

Summed Score	Scaled Score (<i>T</i>)	Standard Error	Summed Score	Scaled Score (<i>T</i>)	Standard Error
0	35	6	17	63	3
1	40	5	18	64	3
2	43	4	19	65	3
3	46	4	20	66	3
4	47	4	21	67	3
5	49	3	22	68	3
6	51	3	23	69	3
7	52	3	24	70	3
8	53	3	25	71	3
9	54	3	26	72	3
10	56	3	27	73	3
11	57	3	28	75	3
12	58	3	29	76	3
13	59	3	30	78	3
14	60	3	31	79	4
15	61	3	32	82	4
16	62	3			

Test Scoring



Edited by
David Thissen
Howard Wainer

Scoring tables for summed scores are in Chapters 3 and 4.

Some elaborations for tests that mix item types are in Chapters 7 and 8.

An approach to the computation of reliable subscores is in Chapter 9 ...

... although the fact that we can now fit MIRT models may replace that soon!

One more story:
The PROMIS Pain Scale

Revicki, D.A., Chen, W.H., Harnam, N., Cook, K.F., Amtmann, D., Callahan, L.F., Jensen, M.P., Keefe, F.J. (in press). Development and psychometric analysis of the PROMIS pain behavior item bank. *Pain*.

39 items like these 8:

When I was in pain I became irritable

When I was in pain I grimaced

When I was in pain I would lie down

When I was in pain I moved extremely slowly

When I was in pain I became angry

When I was in pain I clenched my teeth

When I was in pain I tried to stay very still

When I was in pain I appeared upset or sad

...

Response Scale:

(0) had no pain

(1) never

(2) rarely

(3) sometimes

(4) often

(5) always

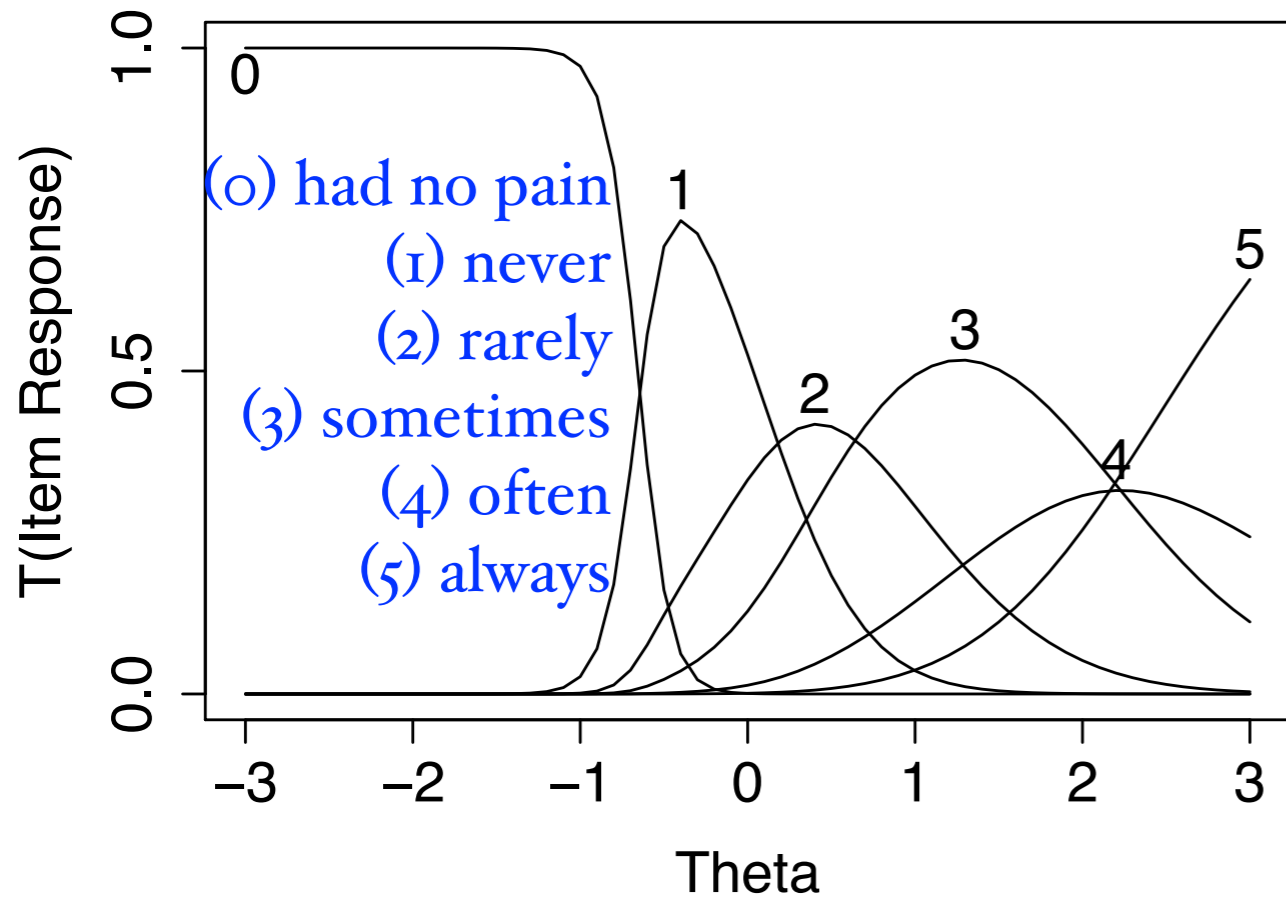
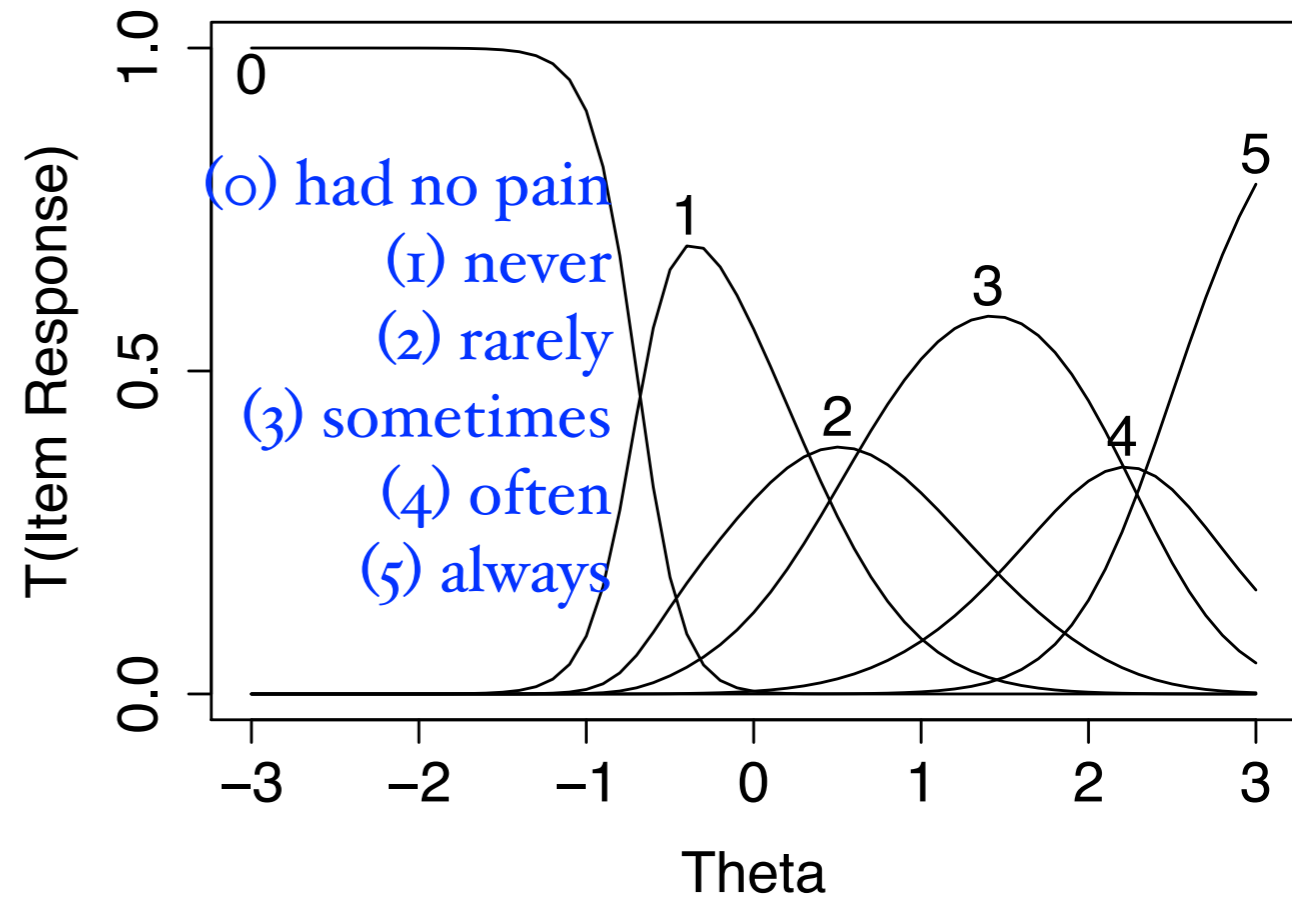
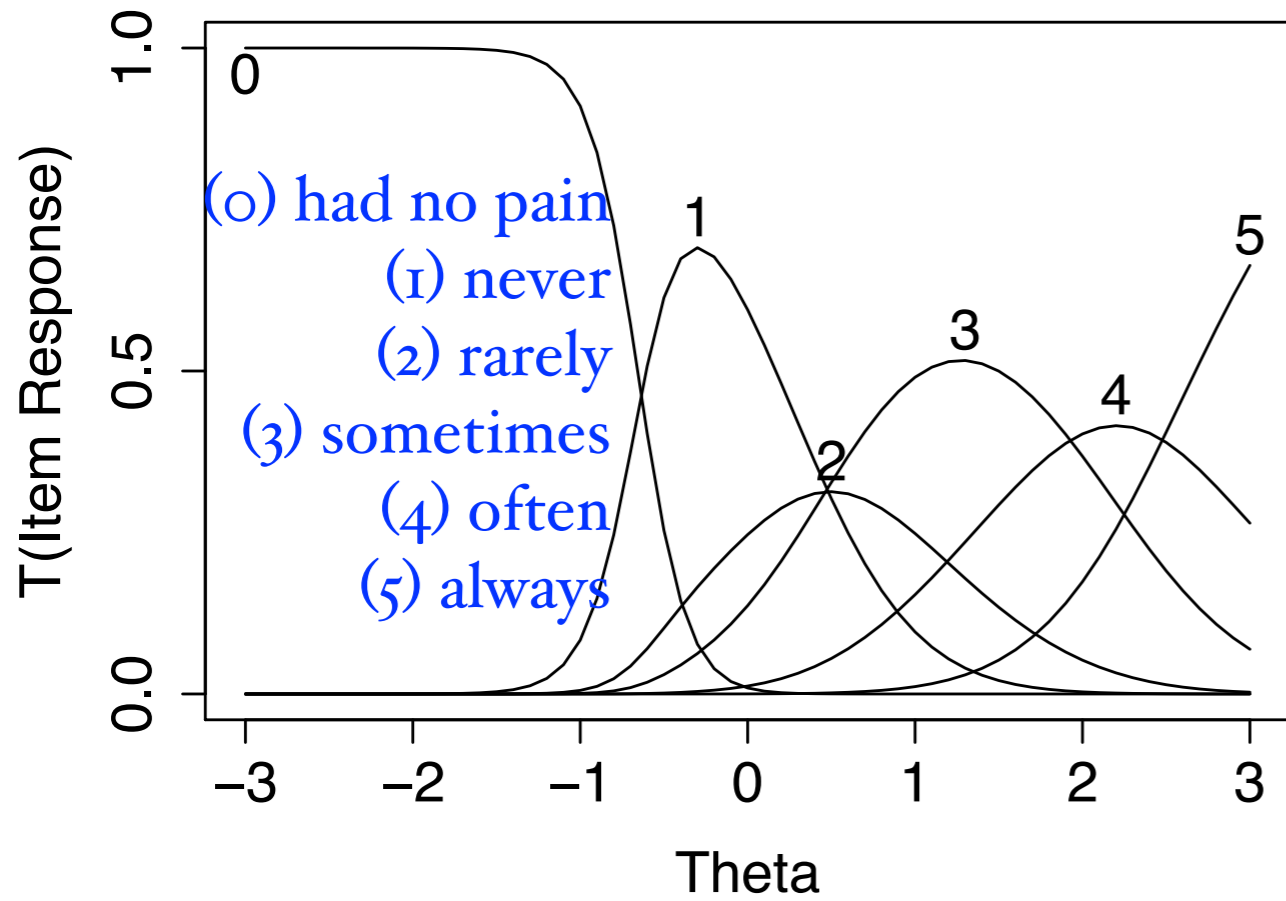
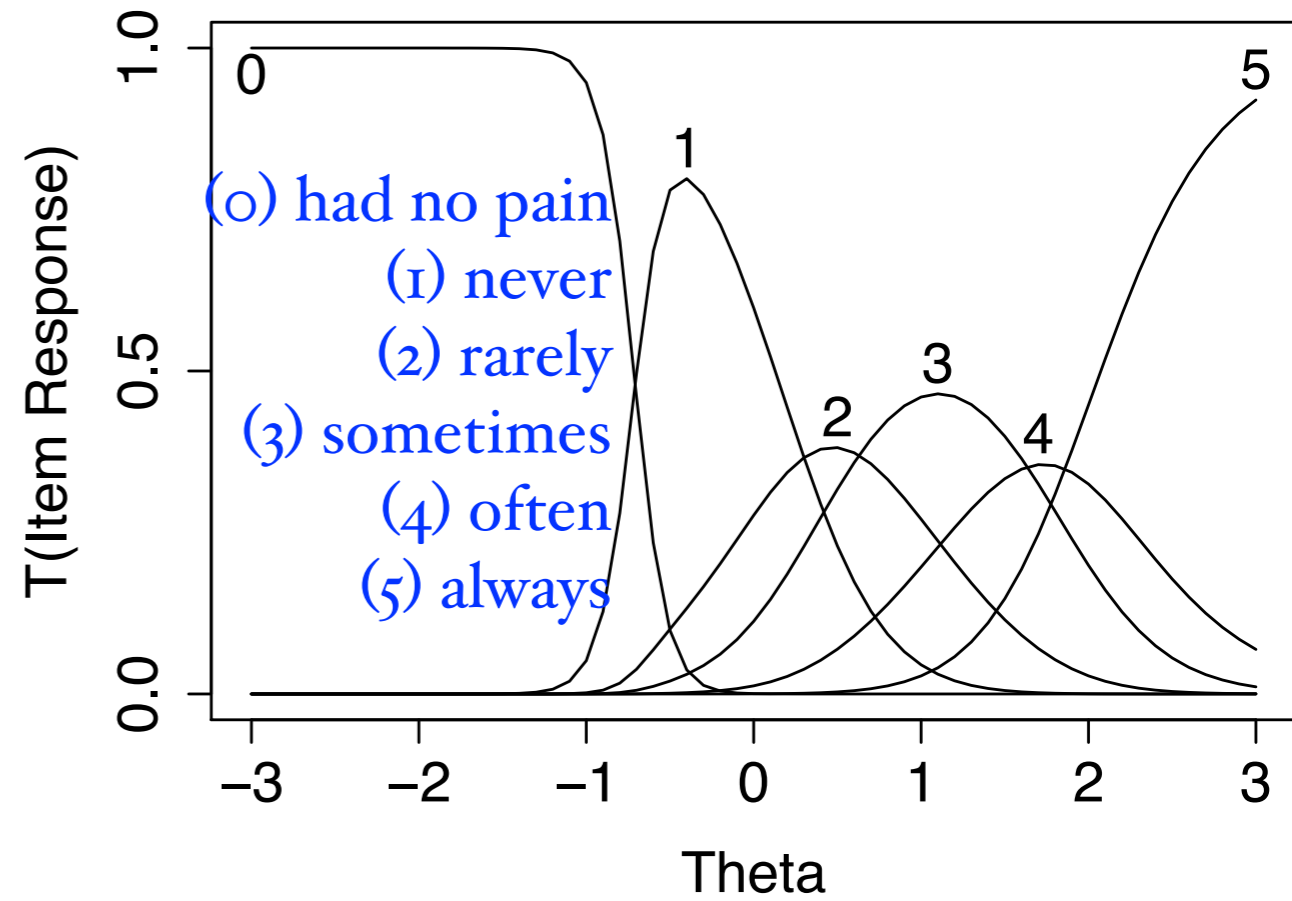
One more story:

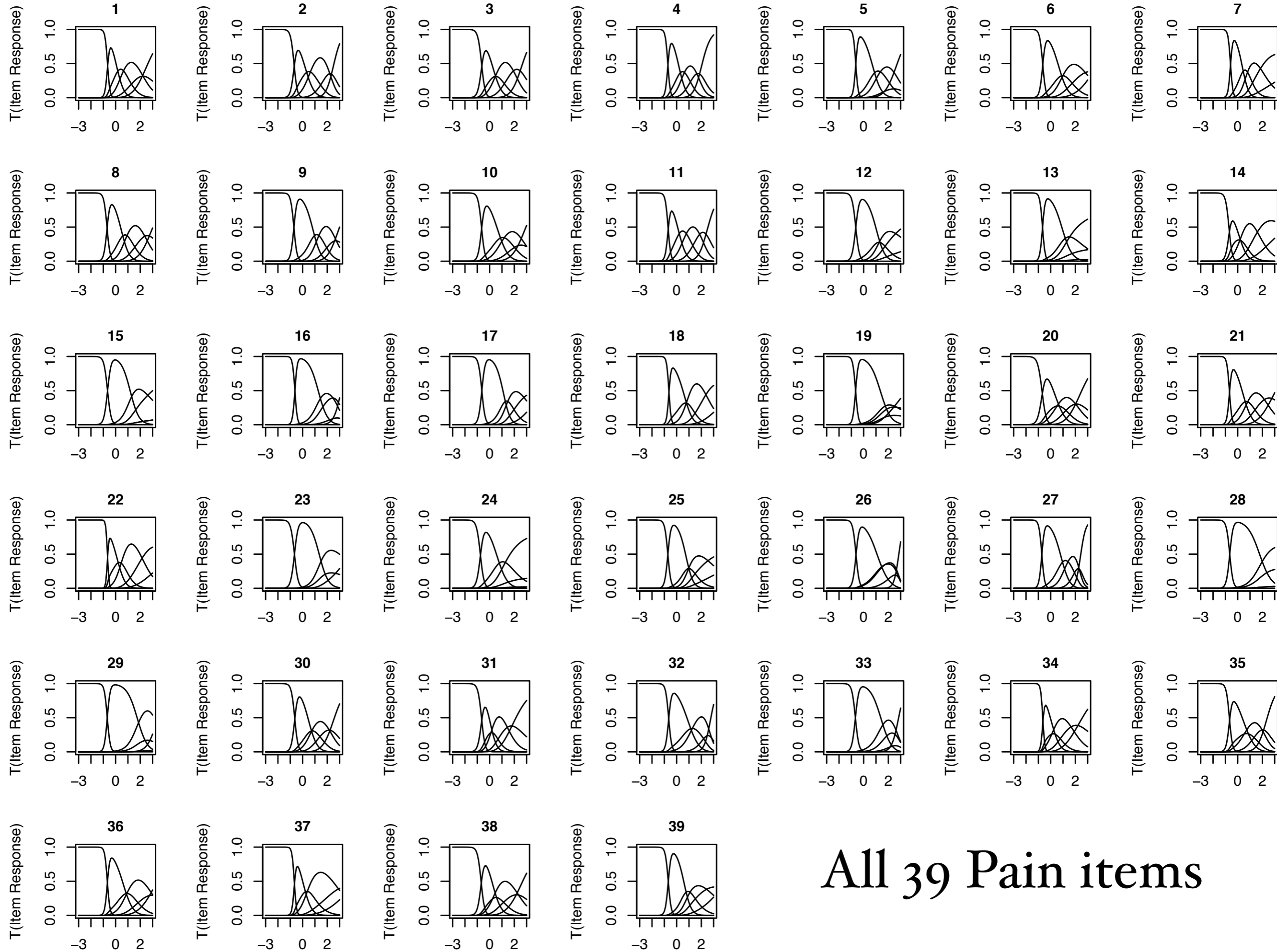
The PROMIS Pain Scale

Revicki et al. ultimately recommended the graded model as fitted with a selected sample (with some pain); but they described briefly some interesting phenomena that can be seen in an unselected sample with the nominal model ...

IRTPRO uses a revised version of the nominal model that separates the slope parameters from the scoring function values, to facilitate multidimensional analyses:

Thissen, D., Cai, L., & Bock, R.D. (in press). The nominal item response model. In M. Nering & R. Ostini (Eds.), *Handbook of polytomous item response theory models*. New York, NY: Routledge Academic.

Item 1**Item 2****Item 3****Item 4**



All 39 Pain items

Item	a1	a2	a3	a4	a5	a6	a7
31	4.00	1.52	0	0	0	0	0
20	3.00	1.40	0	0	0	0	0
4	4.16	1.11	0	0	0	0	0
19	2.40	0.67	0	0	0	0	0
14	3.31	0.81	0	0	0	0	1.17
35	3.12	0	0	0	0	0	1.17
23	3.23	0	1.18	0	0	0	0
29	2.82	0	0.84	0	0	0	0
15	2.35	0	0.75	0	0	0	0
28	1.58	0	0.36	0	0	0	0
27	3.93	0	0.80	0.89	0	0	0
24	2.34	0	0	0.89	0	0	0
6	2.84	0	0	0	1.23	0	0
36	3.33	0	0	0	1.23	0	0
34	3.53	0	0	0	0	1.06	0
38	2.95	0	0	0	0	1.06	0
22	3.71	0	0	0	0	0	0
1	3.08	0	0	0	0	0	0
7	3.07	0	0	0	0	0	0
11	2.94	0	0	0	0	0	0
2	2.83	0	0	0	0	0	0
37	2.81	0	0	0	0	0	0
...	(16 more unidimensional items)						

Logistic slope parameter estimates for a CFA model suggested by LD statistics and item content

Item	a1	a2	a3	a4	a5	a6	a7
31	4.00	1.52	0	0	0	0	0
20	3.00	1.40	0	0	0	0	0
4	4.16	1.11	0	0	0	0	0
19	2.40	0.67	0	0	0	0	0
14	3.31	0.81	0	0	0	0	1.17
35	3.12	0	0	0	0	0	1.17
23	3.23	0	1.18	0	0	0	0
29	2.82	0	0.84	0	0	0	0
15	2.35	0	0.75	0	0	0	0
28	1.58	0	0.36	0	0	0	0
27	3.93	0	0.80	0.89	0	0	0
24	2.34	0	0	0.89	0	0	0
6	2.84	0	0	0	1.23	0	0
36	3.33	0	0	0	1.23	0	0
34	3.53	0	0	0	0	1.06	0
38	2.95	0	0	0	0	1.06	0
22	3.71	0	0	0	0	0	0
1	3.08	0	0	0	0	0	0
7	3.07	0	0	0	0	0	0
11	2.94	0	0	0	0	0	0
2	2.83	0	0	0	0	0	0
37	2.81	0	0	0	0	0	0
...	(16 more unidimensional items)						

Item	a1	a2	a3	a4	a5	a6	a7
31	4.00	1.52	0	0	0	0	0
20	3.00	1.40	0	0	0	0	0
4	4.16	1.11	0	0	0	0	0
19	2.40	0.67	0	0	0	0	0
14	3.31	0.81	0	0	0	0	1.17
35	3.12	0	0	0	0	0	1.17
23	3.23	0	1.18	0	0	0	0
29	2.82	0	0.84	0	0	0	0
15	2.35	0	0.75	0	0	0	0
28	1.58	0	0.36	0	0	0	0
27	3.93	0	0.80	0.89	0	0	0
24	2.34	0	0	0.89	0	0	0
6	2.81	0	0	0	1.23	0	0
36	3.33	0	0	0	1.23	0	0
34	3.53	0	0	0	0	1.06	0
38	2.95	0	0	0	0	1.06	0
22	3.71	0	0	0	0	0	0
1	3.08	0	0	0	0	0	0
7	3.07	0	0	0	0	0	0
11	2.94	0	0	0	0	0	0
2	2.83	0	0	0	0	0	0
37	2.81	0	0	0	0	0	0
...	(16 more unidimensional items)						

IRTPRO

The IRTPRO (*Item Response Theory for Patient-Reported Outcomes*) component of the MEDPRO Project is an entirely new application for item calibration and test scoring using IRT.

Beta release of this software is anticipated soon, with full release to follow soon thereafter.



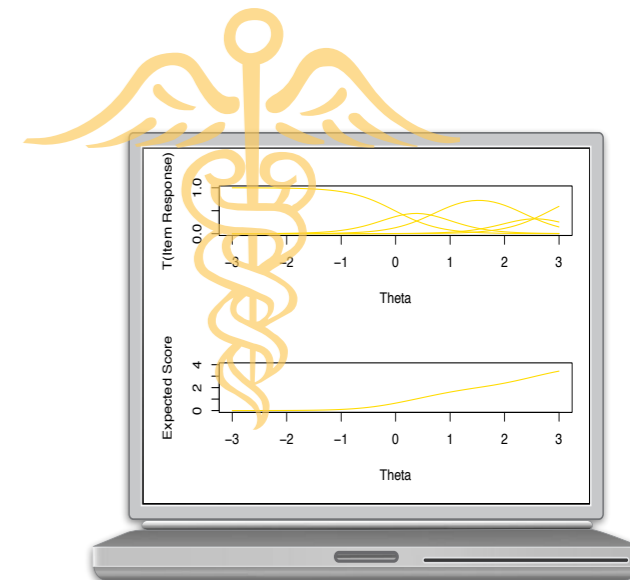
IRTPRO

Alternatively, (possibly) jokingly

IRT for PROfessionals

or

Multilog on steroids

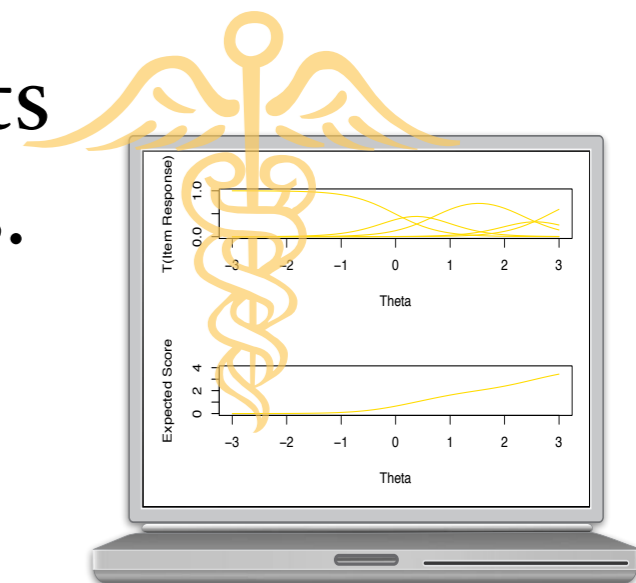


Models in IRTPRO

Unidimensional and **multidimensional** (CFA or EFA):

- Two-parameter logistic (2PL)
[equality constraints yield the one-parameter logistic (1PL)]
- Three-parameter logistic (3PL)
- Graded
- Generalized Partial Credit
[as a special case of ...]
- Nominal

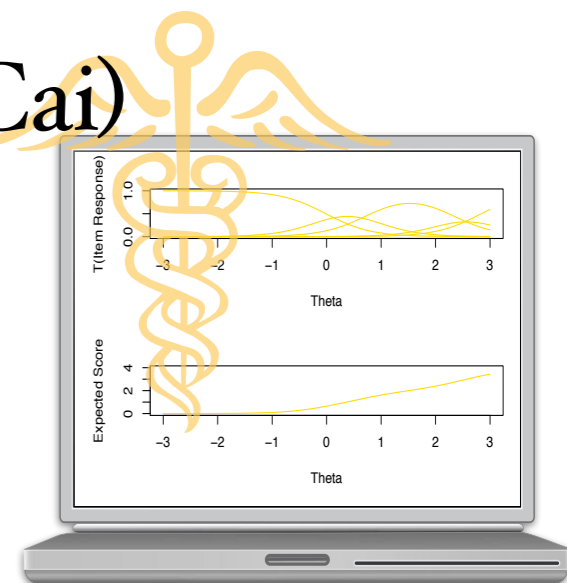
... with optional user-specified equality constraints among parameters, or fixed values for parameters.



Estimation Methods / Algorithms in IRTPRO

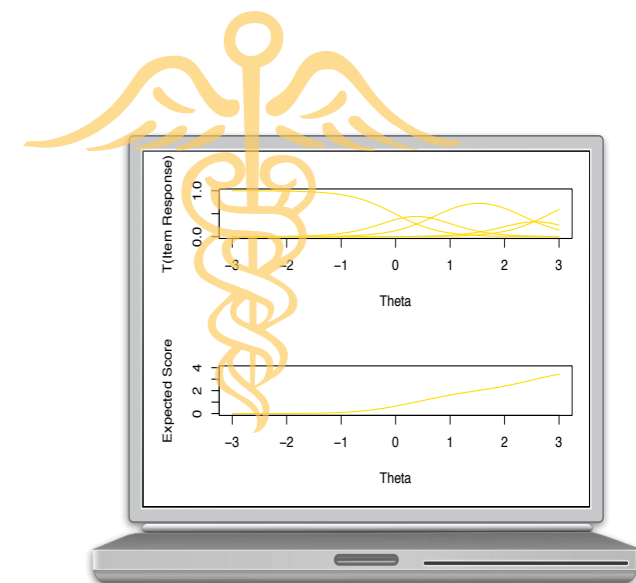
Maximum (marginal) likelihood:

- Bock-Aitkin (BAEM)
 - Bi-factor EM (Gibbons & Hedeker)
 - Now generalized dimension reduction (Cai)
- Adaptive Quadrature (ADQEM; Bock & Schilling)
- Metropolis-Hastings Robbins-Monro (MHRM; Cai)



Scoring Methods in IRTPRO

- Maximum *a posteriori* (MAP) for response patterns
- Expected *a posteriori* (EAP) for response patterns
- Expected *a posteriori* (EAP) for (weighted) summed scores



Groups in IRTPRO

Population latent variable means and variance-covariance matrices may be estimated for multiple groups.

DIF in IRTPRO

Wald tests, with accurate item parameter error variance-covariance matrices computed using the Supplemented EM (SEM) algorithm.



Fit Diagnostics in IRTPRO:

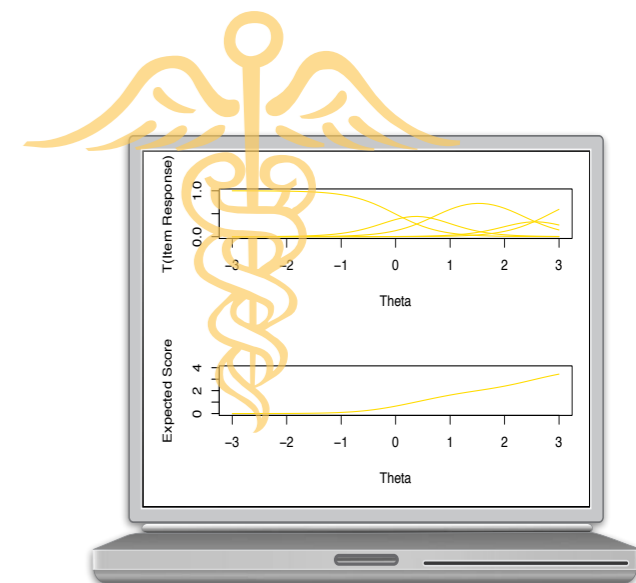
Several goodness of fit and diagnostic statistics are reported, including generalizations for polytomous responses of the Chen-Thissen local dependence (*LD*) statistic and the Orlando-Thissen $SS-X^2$ item-fit diagnostic statistic.



Fit Diagnostics in IRTPRO:

Orlando, M., & Thissen, D. (2000). Likelihood-based item fit indices for dichotomous item response theory models. *Applied Psychological Measurement*, 24, 50-64.

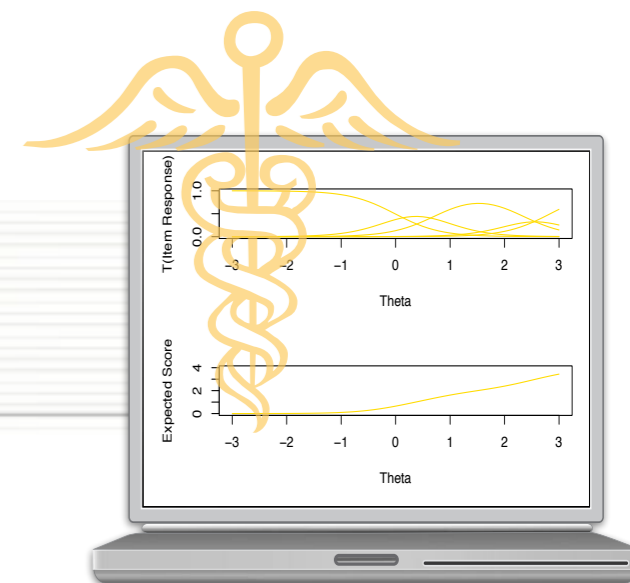
Orlando, M. & Thissen, D. (2003). Further investigation of the performance of $S-X^2$: An item fit index for use with dichotomous item response theory models. *Applied Psychological Measurement*, 27, 289-298.



IRTPRO Acknowledgments

IRTPRO (*Item Response Theory for Patient-Reported Outcomes*) is a component of the MEDPRO Project, which also includes CATPRO (software for the administration and scoring of computerized adaptive tests).

MEDPRO is being developed by Scientific Software International, Inc. (SSI), in collaboration with Assessment Systems Corporation (ASC), and has been funded by SBIR contract HHSN-2612007-00013C with the National Cancer Institute of the National Institutes of Health.



The IRTPRO Development Team

Stephen du Toit

Principal Investigator

Mathilda du Toit

Co-Principal Investigator

Li Cai

Consultant & Statistical Programmer

Shyan Lam

GUI / C++ Programmer

Johan du Toit

Output / C++ Programmer

R. Darrell Bock

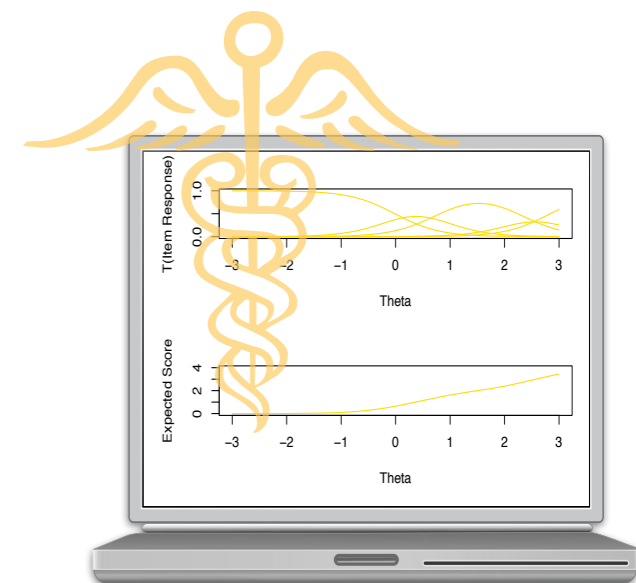
Consultant

Robert Gibbons

Consultant

David Thissen

Consultant



PROMIS at UNC

Thanks to:

Darren A. DeWalt (UNC-CH PI)

Debra E. Irwin

Kelli Scanlon

Kelli Williams

Andrea Meier

Karin Yeatts

Tasanee Walsh

James W. Varni (Texas A&M)

Esi M. DeWitt (Duke)

Jin-Shei Lai (CORE)

And at the Psychometric Lab...

Michael C. Edwards

Cheryl Hill

Michelle M. Langer

Brian D. Stucky

Taehun Lee

Jolynn Pek

Guillaume Filteau



U.S. Department of Health
and Human Services

This research was supported in part by
Grant Number 5U01AR052181.

Supported by the



**National
Institutes
of Health**



IRT has twenty or more
good years left in it ...