Example 6a: Graded Response Ordinal IFA-IRT Models in M*plus* v. 8.8 (complete syntax and output available electronically)

This example comes from the Octogenarian Twin Study of Aging in Sweden. The current analysis includes 634 older adults (age 80–100 years) self-reporting on seven four-category items assessing the Instrumental Activities of Daily Living (IADL). *Note: I have also included R syntax in the online files, but the lavaan default of listwise deletion must be switched to pairwise deletion for the WLSMV results to match those of Mplus!*

Proportion of responses per category:

1. Housework (cleaning and laundry)
2. Bedmaking
3. Cooking
4. Everyday shopping
5. Getting to places outside of walking distance
6. Handling banking and other business
7. Using the telephone

Item	0=Can't Do It	1=Big Problems	2=Some Problems	3=Can Do It
1	0.09	0.08	0.26	0.58
2	0.07	0.04	0.12	0.77
3	0.09	0.05	0.15	0.72
4	0.10	0.09	0.19	0.62
5	0.06	0.16	0.21	0.57
6	0.06	0.08	0.12	0.74
7	0.01	0.03	0.08	0.88

Comparing Polychoric vs. Pearson Correlation Matrices for 7 Ordinal Item Responses
(see online files for code and output of saturated model that generated these correlations)

Polychoric Correlation Estimates

	CIA1	CIA2	CIA3	CIA4	CIA5	CIA6
CIA1						
CIA2	.937					
CIA3	.925	.924				
CIA4	.913	.891	.870			
CIA5	.849	.829	.796	.904		
CIA6	.814	.794	.813	.873	.842	
CIA7	.680	.694	.708	.723	.637	.673

Pearson Correlation Estimates

	CIA1	CIA2	CIA3	CIA4	CIA5	CIA6
CIA1						
CIA2	.820					
CIA3	.835	.835				
CIA4	.840	.768	.766			
CIA5	.753	.679	.672	.822		
CIA6	.686	.660	.669	.749	.708	
CIA7	.464	.487	.496	.472	.402	.469

Polychoric Correlation Standard Errors

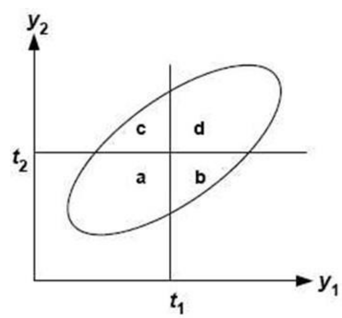
	CIA1	CIA2	CIA3	CIA4	CIA5	CIA6
CIA1						
CIA2	.010					
CIA3	.010	.012				
CIA4	.012	.016	.018			
CIA5	.018	.024	.026	.012		
CIA6	.024	.029	.027	.018	.021	
CIA7	.045	.048	.046	.046	.052	.050

Pearson Correlation Standard Errors

	CIA1	CIA2	CIA3	CIA4	CIA5	CIA6
CIA1						
CIA2	.013					
CIA3	.012	.012				
CIA4	.012	.017	.017			
CIA5	.018	.022	.022	.013		
CIA6	.022	.023	.024	.019	.020	
CIA7	.032	.031	.031	.032	.034	.032

Polychoric correlation is analogous to tetrachoric correlation: They are both based on a bivariate normal distribution, and they both try to represent the correlation that would have created the proportion of responses in each section.

They differ in the number of cells of each pairwise contingency table (and the corresponding degree of division of the bivariate normal distribution at the thresholds).



I found this website that provides a more thorough description with some helpful examples.

Graded Response Model Syntax for 2PL-ish model (left) and 1PL-ish model (right) using ML and a logit link function:

```
TITLE:  Ordinal Models using Full-Info ML

DATA:   FILE = Example6a.csv;  ! Don't need path if in same directory
        FORMAT = free;         ! Default
        TYPE = INDIVIDUAL;     ! Default

VARIABLE:   NAMES = case cia1-cia7;      ! All vars in data
            USEVARIABLES = cia1-cia7;    ! All vars in model
            CATEGORICAL = cia1-cia7;     ! All ordinal outcomes
            MISSING = ALL (99999);       ! Missing value code
            IDVARIABLE = case;           ! Person ID variable

ANALYSIS:   TYPE = GENERAL;                  ! Default
            ESTIMATOR = ML; LINK = LOGIT;    ! Full-info ML in logits
            CONVERGENCE = 0.0000001;         ! For OS comparability

OUTPUT:     STDYX;       ! Standardized solution
            TECH10;      ! Local misfit for full-info ML

SAVEDATA:   SAVE = FSCORES;             ! Save factor scores (thetas)
            FILE = Thetas2Pish.dat;     ! File factor scores saved to
            MISSFLAG = 99999;           ! Missing data value in file

PLOT:   TYPE = PLOT1;     ! PLOT1 gets you sample descriptives
        TYPE = PLOT2;     ! PLOT2 gets you the IRT-relevant curves
        TYPE = PLOT3;     ! PLOT3 gets you descriptives for theta

MODEL: ! Original Graded Response Model (separate loadings per item)

! Original GRM: Factor loadings all estimated and labeled
  IADL BY cia1-cia7* (L_I1-L_I7);
! Item thresholds all estimated and labeled
! You don't have to type these unless you want to make difficulties
! If any listed are not observed, Mplus will throw an error
  [cia1$1-cia7$1*] (T1_I1-T1_I7);
  [cia1$2-cia7$2*] (T2_I1-T2_I7);
  [cia1$3-cia7$3*] (T3_I1-T3_I7);
! Will become Factor mean=0 and variance=1 below for identification
  [IADL*] (FactMean); IADL*  (FactVar);

MODEL CONSTRAINT: ! Factor identification here so can use below
FactMean=0; FactVar=1;


! Creating new IRT parameters
! A = discrimination, B1=y>0, B2=y>1, B3=y>2
  NEW(A_I1-A_I7 B1_I1-B1_I7 B2_I1-B2_I7 B3_I1-B3_I7);
! DO (begin, end), replace # with index
! Discriminations
  DO (1,7) A_I# = L_I# * SQRT(FactVar);
! Difficulties
  DO (1,7) B1_I# = (T1_I#-(L_I#*FactMean)) / (L_I#*SQRT(FactVar));
  DO (1,7) B2_I# = (T2_I#-(L_I#*FactMean)) / (L_I#*SQRT(FactVar));
  DO (1,7) B3_I# = (T3_I#-(L_I#*FactMean)) / (L_I#*SQRT(FactVar));
```

```
TITLE:  Ordinal Models using Full-Info ML

DATA:   FILE = Example6a.csv;  ! Don't need path if in same directory
        FORMAT = free;         ! Default
        TYPE = INDIVIDUAL;     ! Default

VARIABLE:   NAMES = case cia1-cia7;      ! All vars in data
            USEVARIABLES = cia1-cia7;    ! All vars in model
            CATEGORICAL = cia1-cia7;     ! All ordinal outcomes
            MISSING = ALL (99999);       ! Missing value code
            IDVARIABLE = case;           ! Person ID variable

ANALYSIS:   TYPE = GENERAL;                  ! Default
            ESTIMATOR = ML; LINK = LOGIT;    ! Full-info ML in logits
            CONVERGENCE = 0.0000001;         ! For OS comparability

OUTPUT:     STDYX;       ! Standardized solution
            TECH10;      ! Local misfit for full-info ML

SAVEDATA:   SAVE = FSCORES;             ! Save factor scores (thetas)
            FILE = Thetas1Pish.dat;     ! File factor scores saved to
            MISSFLAG = 99999;           ! Missing data value in file

PLOT:   TYPE = PLOT1;     ! PLOT1 gets you sample descriptives
        TYPE = PLOT2;     ! PLOT2 gets you the IRT-relevant curves
        TYPE = PLOT3;     ! PLOT3 gets you descriptives for theta

MODEL: ! Constrained Graded Response Model (same loading for all items)

! Factor loadings constrained equal to single label
  IADL BY cia1-cia7* (L);
! Item thresholds all estimated and labeled
! You don't have to type these unless you want to make difficulties
! If any listed are not observed, Mplus will throw an error
  [cia1$1-cia7$1*] (T1_I1-T1_I7);
  [cia1$2-cia7$2*] (T2_I1-T2_I7);
  [cia1$3-cia7$3*] (T3_I1-T3_I7);
! Will become Factor mean=0 and variance=1 below for identification
  [IADL*] (FactMean); IADL* (FactVar);

MODEL CONSTRAINT: ! Factor identification here so can use below
FactMean=0; FactVar=1;
  NEW(L_I1-L_I7); DO (1,7) L_I# = L; ! For 1PL model

! Creating new IRT parameters
! A = discrimination, B1=y>0, B2=y>1, B3=y>2
  NEW(A_I1-A_I7 B1_I1-B1_I7 B2_I1-B2_I7 B3_I1-B3_I7);
! DO (begin, end), replace # with index
! Discriminations
  DO (1,7) A_I# = L_I# * SQRT(FactVar);
! Difficulties
  DO (1,7) B1_I# = (T1_I#-(L_I#*FactMean)) / (L_I#*SQRT(FactVar));
  DO (1,7) B2_I# = (T2_I#-(L_I#*FactMean)) / (L_I#*SQRT(FactVar));
  DO (1,7) B3_I# = (T3_I#-(L_I#*FactMean)) / (L_I#*SQRT(FactVar));
```

Graded Response Model 2PL-ish Model Fit (left) and 1PLish Model Fit (right) using ML logit:

```
MODEL FIT INFORMATION

Number of Free Parameters                       28

Loglikelihood
          H0 Value                       -2523.585

Information Criteria
          Akaike (AIC)                    5103.171
          Bayesian (BIC)                  5227.828
          Sample-Size Adjusted BIC        5138.931
            (n* = (n + 2) / 24)

Chi-Square Test of Model Fit for the Binary and Ordered Categorical
(Ordinal) Outcomes**

          Pearson Chi-Square
          Value                           1876.488
          Degrees of Freedom                 16317
          P-Value                           1.0000

          Likelihood Ratio Chi-Square
          Value                            676.937
          Degrees of Freedom                 16317
          P-Value                           1.0000

** Of the 48600 cells in the latent class indicator table, 38
  were deleted in the calculation of chi-square due to extreme values.
```

```
MODEL FIT INFORMATION

Number of Free Parameters                       22

Loglikelihood
          H0 Value                       -2591.310

Information Criteria
          Akaike (AIC)                    5226.620
          Bayesian (BIC)                  5324.565
          Sample-Size Adjusted BIC        5254.717
            (n* = (n + 2) / 24)

Chi-Square Test of Model Fit for the Binary and Ordered Categorical
(Ordinal) Outcomes**

          Pearson Chi-Square
          Value                           2650.119
          Degrees of Freedom                 16321
          P-Value                           1.0000

          Likelihood Ratio Chi-Square
          Value                            803.028
          Degrees of Freedom                 16321
          P-Value                           1.0000

** Of the 48600 cells in the latent class indicator table, 40
   were deleted in the calculation of chi-square due to extreme values.
```

This error message indicates that these 2 sets of chi-squares are not on the same scale. We need to test the −2LL difference instead.

Does the 2PL-ish version of the GRM (original with separate loadings) fit better than the 1PL-ish version (constrained same loading)?

−2523.585*-2 = 5047.170 −2ΔLL = 135.45, df = 6, *p* < .0001
−2591.310*-2 = 5182.620 AIC and BIC are smaller for original GRM with separate loadings, too

3 differently scaled solutions from ML logit—all provide the exact same predictions!

UNSTANDARDIZED MODEL RESULTS (IFA MODEL SOLUTION)

FACTOR LOADINGS = CHANGE IN LOGIT(Y) PER UNIT CHANGE IN THETA

IADL BY

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
CIA1	**6.846**	**0.841**	**8.140**	**0.000**
CIA2	5.200	0.555	9.363	0.000
CIA3	4.613	0.456	10.119	0.000
CIA4	5.701	0.612	9.312	0.000
CIA5	3.556	0.298	11.950	0.000
CIA6	2.897	0.261	11.094	0.000
CIA7	1.778	0.209	8.512	0.000

THRESHOLDS = EXPECTED LOGIT(Y=0) WHEN THETA IS 0 (MEAN OF SAMPLE)

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
CIA1$1	**-9.808**	**1.138**	**-8.620**	**0.000**
CIA1$2	**-6.460**	**0.799**	**-8.088**	**0.000**
CIA1$3	**-1.238**	**0.384**	**-3.226**	**0.001**
CIA2$1	-8.145	0.794	-10.257	0.000
CIA2$2	-6.313	0.618	-10.219	0.000
CIA2$3	-3.737	0.441	-8.480	0.000
CIA3$1	-6.841	0.613	-11.162	0.000
CIA3$2	-5.194	0.480	-10.810	0.000
CIA3$3	-2.572	0.330	-7.792	0.000
CIA4$1	-7.454	0.747	-9.975	0.000
CIA4$2	-4.635	0.514	-9.026	0.000
CIA4$3	-1.426	0.327	-4.366	0.000
CIA5$1	-6.578	0.494	-13.314	0.000
CIA5$2	-3.041	0.273	-11.155	0.000
CIA5$3	-0.681	0.203	-3.354	0.001
CIA6$1	-5.538	0.411	-13.486	0.000
CIA6$2	-3.583	0.285	-12.554	0.000
CIA6$3	-2.044	0.219	-9.344	0.000
CIA7$1	-5.810	0.472	-12.315	0.000
CIA7$2	-4.398	0.322	-13.673	0.000
CIA7$3	-2.951	0.237	-12.457	0.000

USING RESULTS FROM IFA MODEL:

IFA model: Logit(y=1) = −threshold + loading(Theta)
Threshold = expected logit of (y=0) for someone with Theta=0
When *-1, threshold becomes intercept: expected logit for (y=1) instead
Loading = regression of item logit on Theta

For 4-category responses, the submodels look like this:
Logit(y= 0 vs 123) = -threshold$1 + loading(Theta)
Logit(y= 01 vs 23) = -threshold$2 + loading(Theta)
Logit(y= 012 vs 3) = -threshold$3 + loading(Theta)

EXAMPLE IFA Model FOR CIA1:
$1 Logit(CIA1=0 vs 123)= 9.808 + 6.846(Theta) → if Theta=0, prob=.99994
$2 Logit(CIA1=01 vs 23)= 6.460 + 6.846(Theta) → if Theta=0, prob=.99844
$3 Logit(CIA1=012 vs 3)= 1.238 + 6.846(Theta) → if Theta=0, prob=.77522

(output from same model continued)

RESULTS FROM IRT MODEL GIVEN BY NEW PARAMETERS:

New/Additional Parameters

DISCRIMINATIONS = SLOPE AT EACH DIFFICULTY VALUE (= LOADING HERE)

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
A_I1	**6.846**	**0.841**	**8.140**	**0.000**
A_I2	5.200	0.555	9.363	0.000
A_I3	4.613	0.456	10.119	0.000
A_I4	5.701	0.612	9.312	0.000
A_I5	3.556	0.298	11.950	0.000
A_I6	2.897	0.261	11.094	0.000
A_I7	1.778	0.209	8.512	0.000

DIFFICULTIES = THETA AT WHICH PROB OF NEXT OPTION = .50

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
B1_I1	**-1.433**	**0.079**	**-18.127**	**0.000**
B1_I2	-1.566	0.088	-17.807	0.000
B1_I3	-1.483	0.086	-17.205	0.000
B1_I4	-1.308	0.076	-17.175	0.000
B1_I5	-1.850	0.104	-17.748	0.000
B1_I6	-1.911	0.120	-15.976	0.000
B1_I7	-3.268	0.320	-10.223	0.000
B2_I1	**-0.944**	**0.059**	**-16.004**	**0.000**
B2_I2	-1.214	0.072	-16.870	0.000
B2_I3	-1.126	0.070	-16.068	0.000
B2_I4	-0.813	0.058	-14.128	0.000
B2_I5	-0.855	0.063	-13.574	0.000
B2_I6	-1.237	0.083	-14.933	0.000
B2_I7	-2.474	0.215	-11.507	0.000
B3_I1	**-0.181**	**0.049**	**-3.714**	**0.000**
B3_I2	-0.719	0.055	-13.083	0.000
B3_I3	-0.558	0.054	-10.386	0.000
B3_I4	-0.250	0.050	-5.029	0.000
B3_I5	-0.192	0.054	-3.548	0.000
B3_I6	-0.705	0.063	-11.169	0.000
B3_I7	-1.660	0.136	-12.244	0.000

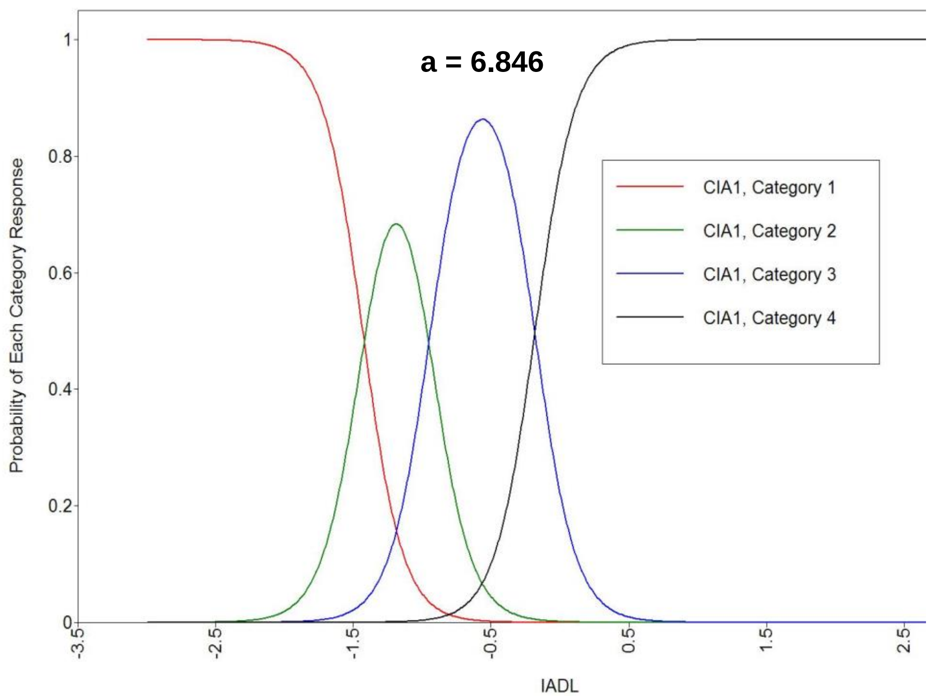
USING RESULTS FROM IRT MODEL WHEN THETA~N(0,1):

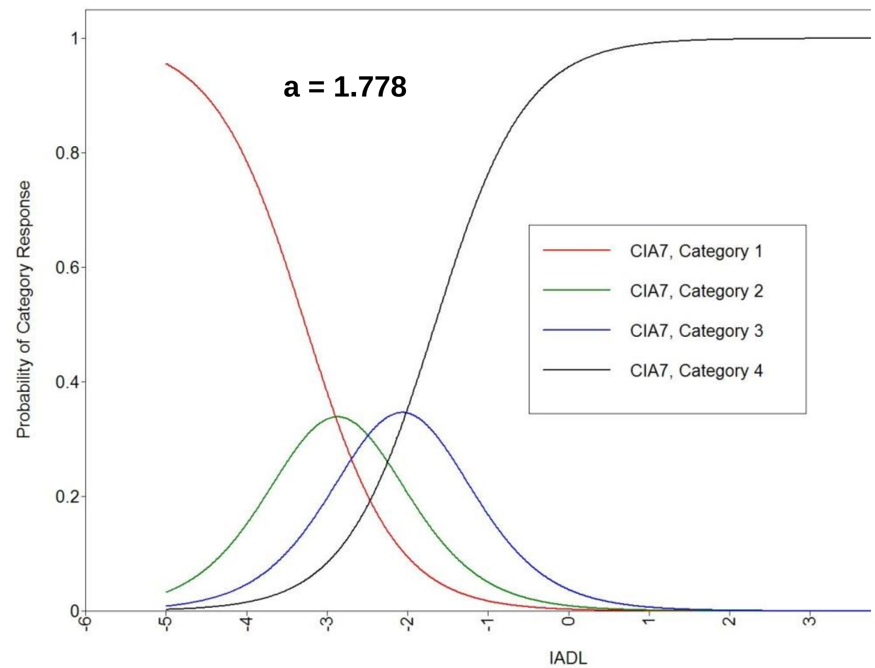
IRT model: Logit(y) = a(theta − difficulty)
a = discrimination (rescaled slope) = loading
b = difficulty (location on latent metric) = threshold/loading

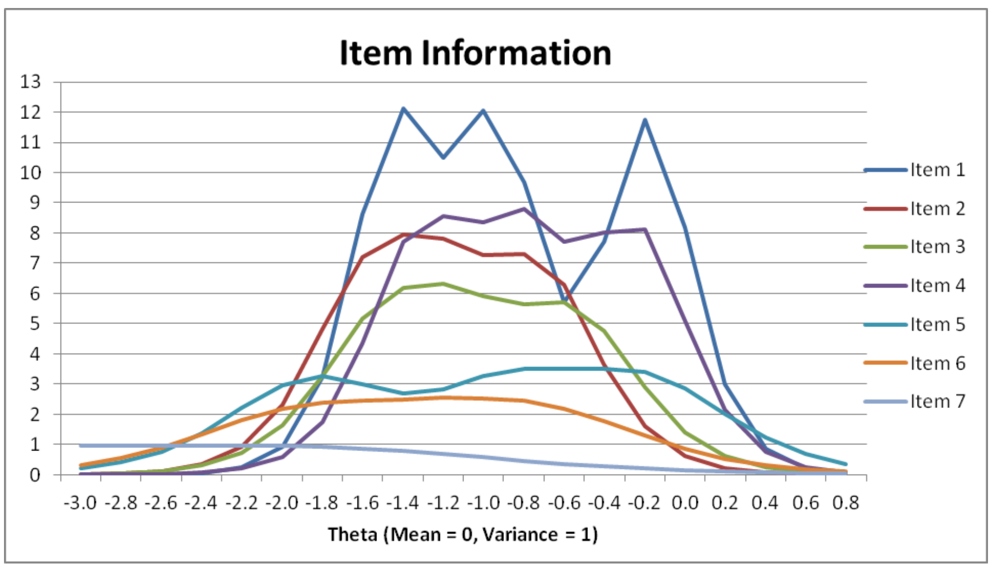
For 4-category responses, the submodels look like this:
$1 Logit(y= 0 vs 123) = a(Theta − difficulty$1)
$2 Logit(y= 01 vs 23) = a(Theta − difficulty$2)
$3 Logit(y= 012 vs 3) = a(Theta − difficulty$3)

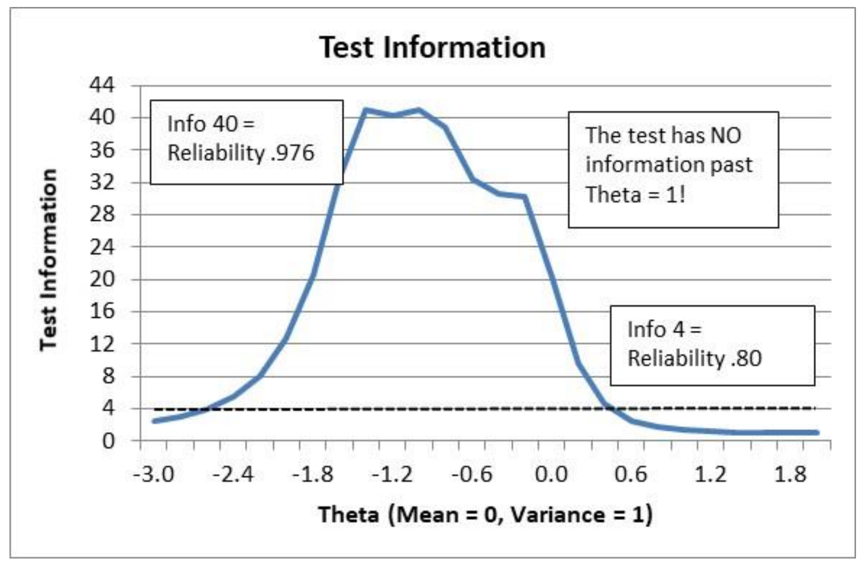
EXAMPLE IRT Model FOR CIA1:
$1 Logit(CIA1=0 vs 123)= 6.846(Theta - -1.433)
$2 Logit(CIA1=01 vs 23)= 6.846(Theta - -0.944)
$3 Logit(CIA1=012 vs 3)= 6.846(Theta - -0.181)

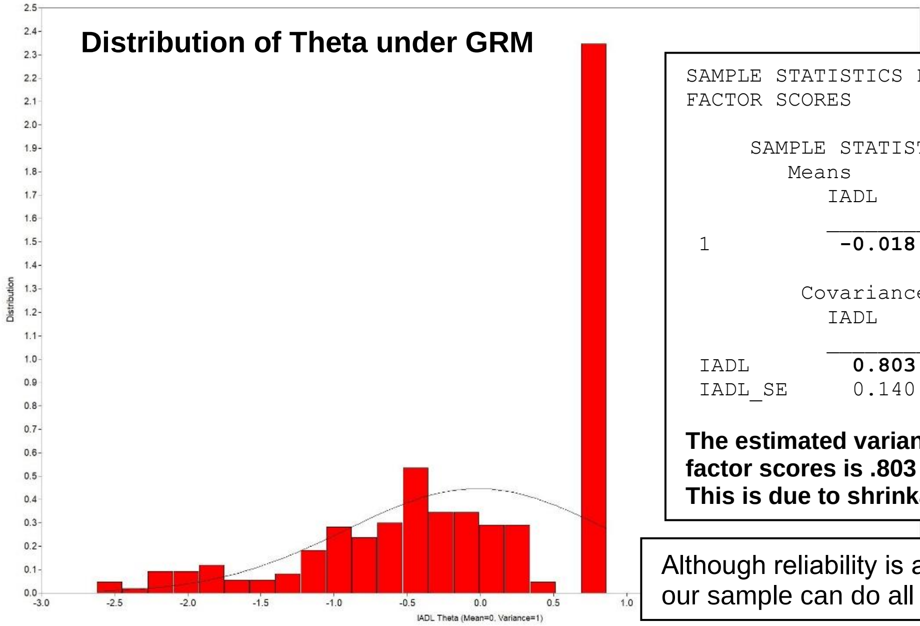
Mplus Category Response Curves – Item 1 (is good with steep discrimination) and Item 7 (is less good because is less steep)





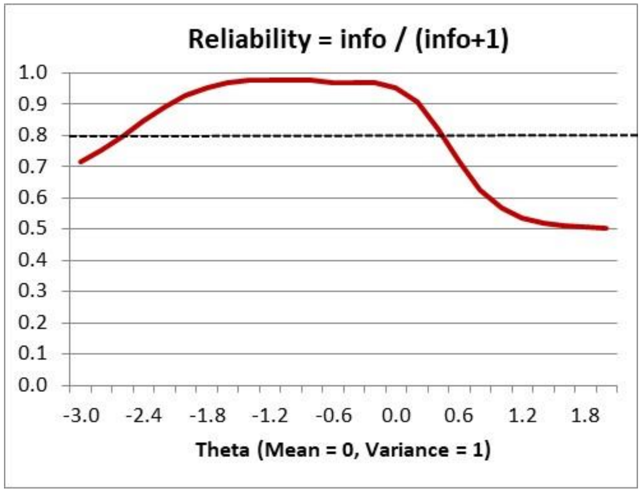




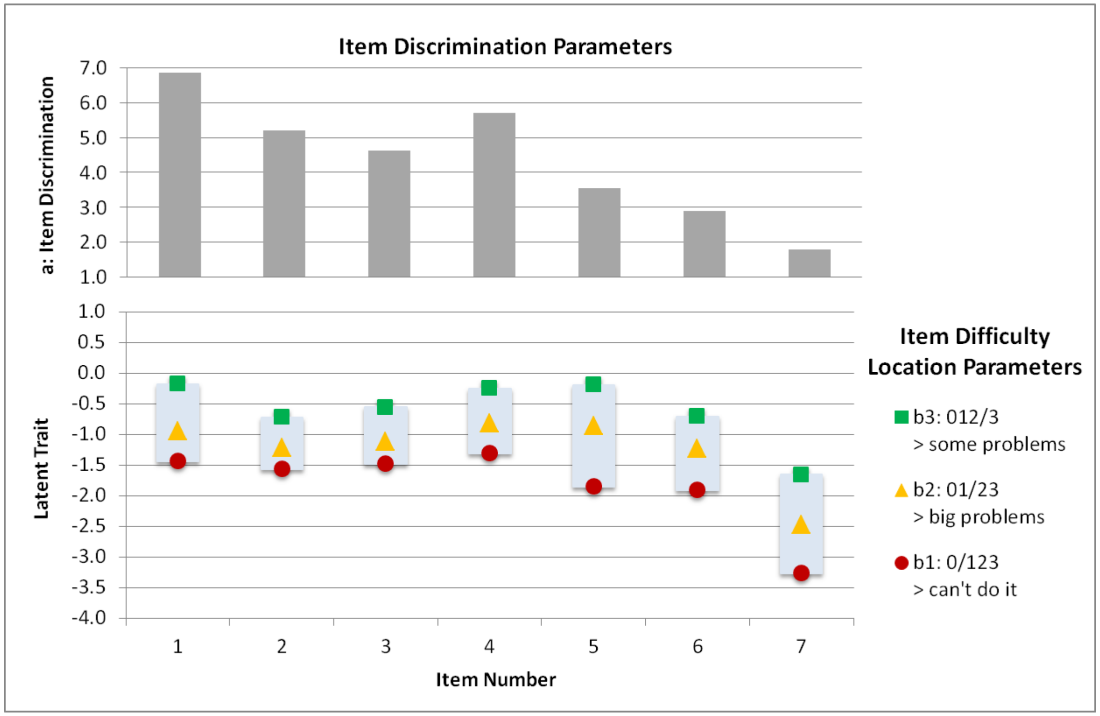



```
SAMPLE STATISTICS FOR ESTIMATED
FACTOR SCORES

     SAMPLE STATISTICS
        Means
           IADL          IADL_SE
        ________       ________
 1        -0.018          0.394

         Covariances
           IADL          IADL_SE
        ________       ________
 IADL      0.803
 IADL_SE   0.140          0.042
```

The estimated variance of the factor scores is .803 instead of 1. This is due to shrinkage.



Although reliability is above .80 from about −2.6 to 0.4 or so, we see a huge ceiling effect: most of our sample can do all the tasks. To measure higher thetas better, we need more difficult items!



Spread of Item Difficulty (made in excel):
Some items (5, 6, and 7) have a wider spread of their b1 and b2 category thresholds, whereas they are closer together for the others. This suggests that those options are less differentiable for item 1–4. Besides adding more difficult items, another way to improve measurement of high thetas would be to expand the higher response options (e.g., from "can do it" to "can do it sometimes" or "can do it always").

What do consider when making a short form:
If we wanted to improve our test by adding more difficult items but keep it the same length, then we'd need to remove some of the current items. These plots show why one must consider the combination of discrimination and difficulty in selecting which items could be removed. For instance, item 7 has the lowest discrimination (slope), but it covers a range of low theta that none of the other items do, so we should keep it for that reason. Instead, items 2 and 3 might be good candidates for removal, as they have lower discriminations than other items in their theta range.

Here is another estimation approach: a 2P vs. a 1P for Binary Responses using WLSMV and a Probit Link
(see the online syntax and output files for the corresponding lavaan version using pairwise deletion as in Mplus WLSMV)

```
TITLE:  Ordinal items using limited-info WLSMV

DATA:   FILE = Example6a.csv; ! Don't need path if data in same folder
VARIABLE:    NAMES = case cia1-cia7;       ! All vars in data
             USEVARIABLES = cia1-cia7;     ! All vars in model
             CATEGORICAL = cia1-cia7;      ! All ordinal outcomes
             MISSING = ALL (99999);        ! Missing value code
             IDVARIABLE = case;            ! Person ID variable

ANALYSIS:    ESTIMATOR = WLSMV;            ! Limited-info in probits
             PARAMETERIZATION = THETA;     ! Error vars=1 scaling
             CONVERGENCE = 0.0000001;      ! For OS comparability

OUTPUT:      STDYX RESIDUAL;      ! Standardized solution, local misfit
             MODINDICES (6.635); ! Cheat codes for p<.01 for df=1

PLOT:        TYPE = PLOT1 PLOT2 PLOT3;     ! Get all IRT plots

SAVEDATA:    DIFFTEST=2P.dat;  ! Save info from bigger model

MODEL: ! Original Graded Response Model (separate loadings per item)

! Factor loadings all estimated and labeled
  IADL BY cia1-cia7* (L_I1-L_I7);
! Item thresholds all estimated and labeled
! If any listed are not observed, Mplus will throw an error
  [cia1$1-cia7$1*] (T1_I1-T1_I7);
  [cia1$2-cia7$2*] (T2_I1-T2_I7);
  [cia1$3-cia7$3*] (T3_I1-T3_I7);
! Item error variances fixed at 1 for identification
  cia1-cia7@1;
! Direct Factor mean=0 and variance=1 for identification (because we
! are using DIFFTEST, which does not allow MODEL CONSTRAINTS)
    [IADL@0]; IADL@1;
_____________________________________________________________________
! If not using DIFFTEST, then can get IRT parameters as before
! Will become Factor mean=0 and variance=1 for identification
    [IADL*] (FactMean);
     IADL*  (FactVar);
MODEL CONSTRAINT: ! Identification here so can use below
FactMean=0; FactVar=1;


! Creating new IRT parameters
! A = discrimination, B1=y>0, B2=y>1, B3=y>2
  NEW(A_I1-A_I7 B1_I1-B1_I7 B2_I1-B2_I7 B3_I1-B3_I7);
! DO (begin, end), replace # with index
! Discriminations
  DO (1,7) A_I# = L_I# * SQRT(FactVar);
! Difficulties
  DO (1,7) B1_I# = (T1_I#-(L_I#*FactMean)) / (L_I#*SQRT(FactVar));
  DO (1,7) B2_I# = (T2_I#-(L_I#*FactMean)) / (L_I#*SQRT(FactVar));
  DO (1,7) B3_I# = (T3_I#-(L_I#*FactMean)) / (L_I#*SQRT(FactVar));
```

```
TITLE:  Ordinal items using limited-info WLSMV

DATA:   FILE = Example6a.csv; ! Don't need path if data in same folder
VARIABLE:    NAMES = case cia1-cia7;       ! All vars in data
             USEVARIABLES = cia1-cia7;     ! All vars in model
             CATEGORICAL = cia1-cia7;      ! All ordinal outcomes
             MISSING = ALL (99999);        ! Missing value code
             IDVARIABLE = case;            ! Person ID variable

ANALYSIS:    ESTIMATOR = WLSMV;            ! Limited-info in probits
             PARAMETERIZATION = THETA;     ! Error vars=1 scaling
             CONVERGENCE = 0.0000001;      ! For OS comparability
             DIFFTEST=2P.dat;  ! Use saved info from bigger model

OUTPUT:      STDYX RESIDUAL;      ! Standardized solution, local misfit
             MODINDICES (6.635); ! Cheat codes for p<.01 for df=1

PLOT:        TYPE = PLOT1 PLOT2 PLOT3;     ! Get all IRT plots

MODEL: ! Constrained Graded Response Model (same loading for all items)

! Factor loadings constrained equal to single label
  IADL BY cia1-cia7* (L);
! Item thresholds all estimated and labeled
! If any listed are not observed, Mplus will throw an error
  [cia1$1-cia7$1*] (T1_I1-T1_I7);
  [cia1$2-cia7$2*] (T2_I1-T2_I7);
  [cia1$3-cia7$3*] (T3_I1-T3_I7);
! Item error variances fixed at 1 for identification
  cia1-cia7@1;
! Direct Factor mean=0 and variance=1 for identification (because we
! are using DIFFTEST, which does not allow MODEL CONSTRAINTS)
    [IADL@0]; IADL@1;
_____________________________________________________________________
! If not using DIFFTEST, then can get IRT parameters as before
! Will become Factor mean=0 and variance=1 for identification
    [IADL*] (FactMean);
     IADL*  (FactVar);
MODEL CONSTRAINT: ! Identification here so can use below
FactMean=0; FactVar=1;
  NEW(L_I1-L_I7); DO (1,7) L_I# = L; ! For 1PL model

! Creating new IRT parameters
! A = discrimination, B1=y>0, B2=y>1, B3=y>2
  NEW(A_I1-A_I7 B1_I1-B1_I7 B2_I1-B2_I7 B3_I1-B3_I7);
! DO (begin, end), replace # with index
! Discriminations
  DO (1,7) A_I# = L_I# * SQRT(FactVar);
! Difficulties
  DO (1,7) B1_I# = (T1_I#-(L_I#*FactMean)) / (L_I#*SQRT(FactVar));
  DO (1,7) B2_I# = (T2_I#-(L_I#*FactMean)) / (L_I#*SQRT(FactVar));
  DO (1,7) B3_I# = (T3_I#-(L_I#*FactMean)) / (L_I#*SQRT(FactVar));
```

Graded Response Model 2PL-ish Model Fit (left) and 1PLish Model Fit (right) using WLSMV and probit link:

```
MODEL FIT INFORMATION

Number of Free Parameters                       28

Chi-Square Test of Model Fit
          Value                             96.262*
          Degrees of Freedom                    14
          P-Value                           0.0000

RMSEA (Root Mean Square Error Of Approximation)
          Estimate                           0.096
          90 Percent C.I.                    0.079  0.115
          Probability RMSEA <= .05           0.000

CFI/TLI
          CFI                                0.997
          TLI                                0.995

SRMR (Standardized Root Mean Square Residual)
          Value                              0.021
```

```
MODEL FIT INFORMATION

Number of Free Parameters                       22

Chi-Square Test of Model Fit
          Value                            202.568*
          Degrees of Freedom                    20
          P-Value                           0.0000

Chi-Square Test for Difference Testing (analog to LRT in ML)
          Value                             93.825
          Degrees of Freedom                     6
          P-Value                           0.0000

RMSEA (Root Mean Square Error Of Approximation)
          Estimate                           0.120
          90 Percent C.I.                    0.105  0.135
          Probability RMSEA <= .05           0.000

CFI/TLI
          CFI                                0.993
          TLI                                0.993

SRMR (Standardized Root Mean Square Residual)
          Value                              0.077
```

Right: the Chi-Square for Difference Testing tells us directly that the 2P version of the polytomous model fits significantly better (now using WLSMV, but same conclusion as when using ML).

Here are the parameter estimates under WLSMV Theta Parameterization (Probit) for the 2P version of ordinal responses

UNSTANDARDIZED MODEL RESULTS (IFA MODEL SOLUTION)

FACTOR LOADINGS = CHANGE IN PROBIT(Y=1) PER UNIT CHANGE IN THETA

IADL BY	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
CIA1	3.655	0.330	11.086	0.000
CIA2	3.347	0.388	8.630	0.000
CIA3	2.923	0.269	10.881	0.000
CIA4	3.286	0.299	11.008	0.000
CIA5	2.222	0.159	13.963	0.000
CIA6	1.907	0.169	11.305	0.000
CIA7	1.075	0.130	8.280	0.000

THRESHOLDS = EXPECTED PROBIT(Y=0) WHEN THETA IS 0

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
CIA1$1	-5.150	0.424	-12.140	0.000
CIA1$2	-3.657	0.347	-10.536	0.000
CIA1$3	-0.734	0.217	-3.383	0.001
CIA2$1	-5.097	0.497	-10.252	0.000
CIA2$2	-4.254	0.445	-9.550	0.000
CIA2$3	-2.620	0.353	-7.424	0.000
CIA3$1	-4.193	0.327	-12.825	0.000
CIA3$2	-3.403	0.296	-11.486	0.000
CIA3$3	-1.762	0.232	-7.592	0.000
CIA4$1	-4.379	0.342	-12.794	0.000
CIA4$2	-2.987	0.269	-11.106	0.000
CIA4$3	-1.024	0.211	-4.863	0.000
CIA5$1	-3.866	0.233	-16.615	0.000
CIA5$2	-1.892	0.160	-11.857	0.000
CIA5$3	-0.424	0.130	-3.275	0.001
CIA6$1	-3.451	0.235	-14.697	0.000
CIA6$2	-2.354	0.184	-12.804	0.000
CIA6$3	-1.400	0.154	-9.071	0.000
CIA7$1	-3.282	0.249	-13.171	0.000
CIA7$2	-2.577	0.181	-14.232	0.000
CIA7$3	-1.757	0.137	-12.841	0.000

For 4-category responses, the sub-models look like this:
Probit(y= 0 vs 123) = -threshold$1 + loading(Theta)
Probit(y= 01 vs 23) = -threshold$2 + loading(Theta)
Probit y= 012 vs 3) = -threshold$3 + loading(Theta)

For 4-category responses, the sub-models look like this:
$1 Probit(y= 0 vs 123) = a(theta – difficulty$1)
$2 Probit(y= 01 vs 23) = a(theta – difficulty$2)
$3 Probit(y= 012 vs 3) = a(theta – difficulty$3)

In requesting predicted factor scores using WLSMV, their sample mean was -0.199 (not 0) and the sample variance was 0.538 (not 1). Whereas ML provided EAP (expected a posteriori = mean) estimates, WLSMV provides MAP (maximum a posteriori = mode) estimates, which are less stable with fewer items. Use the ML versions instead!

RESULTS FROM IRT MODEL GIVEN BY NEW PARAMETERS:

New/Additional Parameters

DISCRIMINATIONS = SLOPE AT EACH DIFFICULTY VALUE

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
A_I1	3.655	0.330	11.086	0.000
A_I2	3.347	0.388	8.630	0.000
A_I3	2.923	0.269	10.881	0.000
A_I4	3.286	0.299	11.008	0.000
A_I5	2.222	0.159	13.963	0.000
A_I6	1.907	0.169	11.305	0.000
A_I7	1.075	0.130	8.280	0.000

DIFFICULTIES = THETA AT WHICH PROB OF NEXT OPTION = .50)

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
B1_I1	-1.409	0.080	-17.669	0.000
B1_I2	-1.523	0.087	-17.606	0.000
B1_I3	-1.435	0.084	-17.013	0.000
B1_I4	-1.333	0.078	-17.089	0.000
B1_I5	-1.740	0.100	-17.385	0.000
B1_I6	-1.809	0.113	-16.054	0.000
B1_I7	-3.054	0.284	-10.735	0.000
B2_I1	-1.001	0.065	-15.310	0.000
B2_I2	-1.271	0.074	-17.066	0.000
B2_I3	-1.165	0.073	-16.020	0.000
B2_I4	-0.909	0.064	-14.124	0.000
B2_I5	-0.852	0.064	-13.232	0.000
B2_I6	-1.234	0.081	-15.174	0.000
B2_I7	-2.397	0.207	-11.555	0.000
B3_I1	-0.201	0.054	-3.730	0.000
B3_I2	-0.783	0.059	-13.333	0.000
B3_I3	-0.603	0.058	-10.391	0.000
B3_I4	-0.312	0.054	-5.734	0.000
B3_I5	-0.191	0.055	-3.467	0.001
B3_I6	-0.734	0.064	-11.550	0.000
B3_I7	-1.635	0.138	-11.888	0.000

LOCAL FIT VIA STANDARDIZED RESIDUAL CORRELATIONS
LEFTOVER POLYCHORIC CORRELATION (HOW FAR OFF FROM DATA)

Residuals for Covariances/Correlations/Residual Correlations

	CIA1	CIA2	CIA3	CIA4	CIA5	CIA6
CIA1						
CIA2	0.013					
CIA3	0.012	0.017				
CIA4	-0.010	-0.025	-0.036			
CIA5	-0.030	-0.045	**-0.067**	0.032		
CIA6	-0.040	**-0.055**	-0.025	0.026	0.035	
CIA7	-0.026	-0.007	0.016	0.022	-0.031	0.025

The largest correlation discrepancy is < .07 in absolute value, which is pretty good. Rather than follow the cheat codes, I am calling it done.

Extensive Results Section (in which model fit via WLSMV is reported first, followed by full-information MML as "better" version of model parameters). Note this is *way* more text than one would typically write, but I provide it here for completeness:

Psychometric assessment for the extent to which a single latent trait could predict that pattern of association among seven items was conducted using Item Factor Analysis (IFA) in M*plus* v 8.8 (Muthén and Muthén, 1998–2017). These models use a cumulative link function (i.e., logit or probit) and a multinomial conditional response distribution, such that the four-category response outcomes (i.e., response y for item i and subject s) are predicting using three binary submodels: $Link[p(y_{is} > 0)] = -\tau_{i1} + \lambda_i F_s$, $Link[p(y_{is} > 1)] = -\tau_{i2} + \lambda_i F_s$, and $Link[p(y_{is} > 2)] = -\tau_{i2} + \lambda_i F_s$. In each model, τ_i is an item-specific and category-specific threshold. When multiplied by −1, it becomes an intercept that gives the link-transformed probability of the submodel's item response (for item i and subject s) at a latent trait score F for subject s of 0, and λ is a factor loading for item i for the expected change in the link-transformed response for a one-unit change in F_s. No separate item-specific residual variances can be estimated given these items' multinomial response options.

The current gold standard of estimation for such IFA models is marginal maximum likelihood (MML), in which the term *marginal* refers to the full-information process of marginalizing over the possible trait values for each person in the analysis using adaptive Gaussian quadrature (here, with 15 points per factor). Accordingly, measures of model fit when using MML involve the contingency table of all possible responses to all items. In our 7 items, the full contingency table generates up to 4^7 = 16,384 possible cells. Consequently, no measures of absolute fit would be valid for the current sample of 634 respondents (which would need a minimum expected count of 5 respondents within each possible cell). Instead, we conducted assessment of model fit via a limited-information diagonally weighted least squares estimator using a mean- and variance-corrected $\chi2$ (i.e., WLSMV in M*plus* with the THETA parameterization and a probit link function). In the WLSMV estimator, the item responses are first summarized into an estimated polychoric correlation matrix using the cross-tabulation of responses for each possible pair of items. The IFA models are then fitted to the estimated polychoric correlation matrix, such that measures of global and local absolute fit (i.e., as traditional in confirmatory factor analyses of continuous responses) can be computed from the discrepancy of the model-predicted and data-estimated polychoric correlation matrices. In addition to $\chi2$ tests of absolute fit, results also include the Comparative Fit Index (CFI), the Standardized Root Mean Square Residual (SRMR), and the Root Mean Square Error of Approximation (RMSEA). The CFI indexes the fit of the specified model relative to a null model (of no polychoric correlations across items), in which CFI values ≥ .95 traditionally indicate excellent fit. Conversely, the SRMR and RMSEA index the fit of the specified model relative to a saturated model (i.e., the data-estimated polychoric correlations), in which SRMR and RMSEA values ≤ .06 traditionally indicate good fit. RMSEA also offers a 90% confidence interval and a significance test of "close fit" with a null hypothesis of .05. Local misfit can be diagnosed by examining the specific sources of discrepancy between the model-predicted and data-estimated polychoric correlations (i.e., as available using the RESIDUAL option in M*plus*). Finally, the fit of nested models can be compared using the DIFFTEST procedure in M*plus*.

A single-trait model was first fit for the seven ordinal items using WLSMV, in which the latent trait mean and variance were fixed for identification to 0 and 1, respectively, a separate factor loading was estimated for each item, and separate thresholds were estimated for each binary submodel per item. This model exhibited acceptable fit by CFI = .997 and SRMR = .021, but unacceptable fit by the $\chi2$ test of absolute fit, $\chi2$ (14) = 96.262, p < .001, and RMSEA = .096 [CI = .079–.115, p < .001]. However, examination of local misfit revealed all discrepancies between the model-predicted and data-estimated polychoric correlations were less than .07 in absolute value, indicating no practically significant bivariate item misfit. A reduced model in which all loadings were constrained equal across items fit significantly worse, DIFFTEST(6) = 93.825, p < .001, indicating differences in item discrimination (i.e., the extent to which each item was related to the latent trait). Thus, the original model was retained for further examination using full-information marginal maximum likelihood (MML) estimation instead.

Model parameters obtained using MML and a logit link are shown in Table 1, which includes the IFA item parameters (thresholds and loadings), as well as their Item Response Theory (IRT) analogous parameter of item difficulty, computed as $b_{ic} = \tau_{ic}/\lambda_i$; IRT discrimination a_i is the same as the loading λ_i in this case. The net result of these item parameters can be described more succinctly by examining the overall reliability with which the latent trait has been measured. In IFA or IRT models—as in any kind of psychometric model with a nonlinear relationship between the item response and the latent trait—reliability is trait-specific, most often characterized by a quantity known as *test information*. For ease of interpretation, the test information function created by the items was converted to a traditional measure of reliability that ranges from 0 to 1 as reliability = information / (information +1). Figure 1 shows that test reliability is ≥.80 only from ~2.6 SD below the mean to 0.40 SD above the mean, after which point reliability drops off precipitously due to a lack of items with difficulty levels above 0.

(See Example 6a spreadsheet for Table 1 and Figure 1)

Reference: Muthén, L. K., & Muthén, B.O. (1998–2017). *Mplus user's guide* (8th ed.). Los Angeles, CA: Muthén & Muthén.