Example 5: Binary 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 635 older adults (age 80–100 years) self-reporting on seven binary 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!*

1. Housework (cleaning and laundry): 1=64%
2. Bedmaking: 1=84%
3. Cooking: 1=77%
4. Everyday shopping: 1=66%
5. Getting to places outside of walking distance: 1=65%
6. Handling banking and other business: 1=73%
7. Using the telephone 1=94% → **Instability!**

Two versions of a response format were available for these data:

Binary → 0 = needs help, 1 = does not need help
Categorical → 0 = can't do it, 1 = big problems, 2 = some problems, 3 = no problems

Higher scores indicate greater function. We will examine each response format in turn.

Comparing Tetrachoric vs. Pearson Correlation Matrices for 7 Binary Item Responses

(see online files for code and output of saturated model that generated these correlations)

Tetrachoric Correlation Estimates

	DIA1	DIA2	DIA3	DIA4	DIA5	DIA6
DIA1						
DIA2	.916					
DIA3	.921	.921				
DIA4	.882	.873	.863			
DIA5	.837	.845	.771	.924		
DIA6	.786	.790	.795	.888	.861	
DIA7	.679	.796	.805	.844	.755	.818

Pearson Correlation Estimates

	DIA1	DIA2	DIA3	DIA4	DIA5	DIA6
DIA1						
DIA2	.562					
DIA3	.661	.692				
DIA4	.680	.555	.614			
DIA5	.627	.533	.525	.747		
DIA6	.551	.529	.563	.666	.631	
DIA7	.279	.438	.398	.341	.309	.379

Tetrachoric Correlation Standard Errors

	DIA1	DIA2	DIA3	DIA4	DIA5	DIA6
DIA1						
DIA2	.026					
DIA3	.020	.021				
DIA4	.023	.030	.028			
DIA5	.028	.033	.037	.017		
DIA6	.035	.039	.035	.023	.026	
DIA7	.078	.055	.056	.174	.066	.054

Pearson Correlation Standard Errors

	DIA1	DIA2	DIA3	DIA4	DIA5	DIA6
DIA1						
DIA2	.028					
DIA3	.023	.021				
DIA4	.022	.028	.025			
DIA5	.024	.029	.029	.018		
DIA6	.028	.029	.028	.022	.024	
DIA7	.038	.033	.034	.036	.036	.034

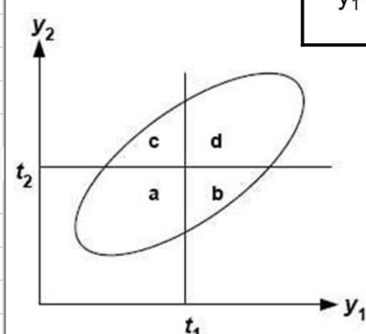
	DIA4	
DIA 7	0	1
0	.057	**.000** Empty cell: Everyone who could shop could also answer the phone
1	.290	.653

Remember that the maximum possible Pearson correlation will be smaller than ±1 for any two variables with means for the probability of y=1 ≠ .50:

$$r_{x,y} = \sqrt{\frac{p_x(1-p_y)}{p_y(1-p_x)}}$$

Below is a review of where tetrachoric correlations come from (bivariate normal): what correlation would have created the proportion of responses in each quadrant?

Data	$y_2 = 0$	$y_2 = 1$
$y_1 = 0$	a	c
$y_1 = 1$	b	d



Mplus Syntax for Binary 2-PL Model Syntax (left) and 1-PL Model (right) using Full-Information ML and a Logit Link:

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

DATA:   FILE = Example5.csv; ! Don't need path if data in same folder
        FORMAT = free;       ! Default
        TYPE = INDIVIDUAL;   ! Default

VARIABLE:   NAMES = case dia1-dia7;     ! All vars in data
            USEVARIABLES = dia1-dia7;   ! All vars in model
            CATEGORICAL = dia1-dia7;    ! 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 = Thetas2P.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:
! Factor loadings all estimated in 2PL
  IADL BY dia1-dia7*;
! Item thresholds all estimated
  [dia1$1-dia7$1*];
! Factor variance=1 and mean=0 for identification
  IADL@1; [IADL@0];
```

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

DATA:   FILE = Example5.csv; ! Don't need path if data in same folder
        FORMAT = free;       ! Default
        TYPE = INDIVIDUAL;   ! Default

VARIABLE:   NAMES = case dia1-dia7;     ! All vars in data
            USEVARIABLES = dia1-dia7;   ! All vars in model
            CATEGORICAL = dia1-dia7;    ! 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 = Thetas1P.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:
! Factor loadings all held EQUAL in 1PL
  IADL BY dia1-dia7* (loading);
! Item thresholds all estimated
  [dia1$1-dia7$1*];
! Factor variance=1 and mean=0 for identification
  IADL@1; [IADL@0];
```

Binary 2-Pameter Model Fit (left) and 1-Parameter Model Fit (right) using Full-Information ML and a Logit Link:

```
MODEL FIT INFORMATION – 2PL

Number of Free Parameters                        14

Loglikelihood

          H0 Value                        -1454.634

Information Criteria

          Akaike (AIC)                     2937.268
          Bayesian (BIC)                   2999.619
          Sample-Size Adjusted BIC         2955.170
            (n* = (n + 2) / 24)

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

          Pearson Chi-Square

          Value                             340.829
          Degrees of Freedom                    113
          P-Value                            0.0000

          Likelihood Ratio Chi-Square

          Value                             120.273
          Degrees of Freedom                    113
          P-Value                            0.3023
```

Linda Muthén (and others) have suggested that if these two χ^2 values don't match, they should not be used to assess model fit.

Further, the possible total DF for the χ^2 is calculated based on # possible response patterns. Here, for 7 binary items:

2PL model: 2^7 = 128 possible – 7 loadings – 7 thresholds – 1 = 113
1PL model: 2^7 = 128 possible – 1 loading – 7 thresholds – 1 = 119

However, the 1PL only has df=118 because of the deleted cell.

```
MODEL FIT INFORMATION – 1 PL

Number of Free Parameters                         8

Loglikelihood

          H0 Value                        -1464.457

Information Criteria

          Akaike (AIC)                     2944.915
          Bayesian (BIC)                   2980.544
          Sample-Size Adjusted BIC         2955.144
            (n* = (n + 2) / 24)

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

          Pearson Chi-Square

          Value                             296.199
          Degrees of Freedom                    118
          P-Value                            0.0000

          Likelihood Ratio Chi-Square

          Value                             126.354
          Degrees of Freedom                    118
          P-Value                            0.2828

** Of the 630 cells in the latent class indicator table, 1
   were deleted in the calculation of chi-square due to extreme values.
```

This error message indicates that these 2 sets of chi-squares for the 2-PL and 1-PL are not on the same scale because they are not based on the same data. So we can't compare the chi-squares to test the difference in model fit, but we can still compare LL values.

Does the 2-PL fit better than the 1-PL?

−1454.634*-2 = 2909.258 −2LL difference = 19.946, df = 6, *p* = .0032
−1464.457*-2 = 2928.914 AIC (but not BIC) is smaller for 2PL, too

3 differently scaled 2-Parameter solutions from ML logit provided by Mplus—all provide the exact same model predictions!

UNSTANDARDIZED MODEL RESULTS (IFA MODEL SOLUTION)

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

IADL BY	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
DIA1	4.328	0.560	7.725	0.000
DIA2	4.978	0.808	6.159	0.000
DIA3	4.323	0.570	7.579	0.000
DIA4	7.511	1.696	4.429	0.000
DIA5	4.248	0.527	8.062	0.000
DIA6	3.451	0.401	8.600	0.000
DIA7	3.283	0.601	5.467	0.000

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

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
DIA1$1	-1.629	0.295	-5.516	0.000
DIA2$1	-5.202	0.770	-6.754	0.000
DIA3$1	-3.462	0.441	-7.842	0.000
DIA4$1	-3.120	0.744	-4.193	0.000
DIA5$1	-1.833	0.298	-6.158	0.000
DIA6$1	-2.442	0.292	-8.368	0.000
DIA7$1	-5.962	0.858	-6.951	0.000

STDYX MODEL RESULTS (STANDARDIZED IFA MODEL SOLUTION)

FACTOR LOADINGS IN STANDARDIZED METRIC = loading*SD(Theta)/SD(Y)

IADL BY	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
DIA1	0.922	0.018	51.712	0.000
DIA2	0.940	0.018	52.557	0.000
DIA3	0.922	0.018	50.622	0.000
DIA4	0.972	0.012	80.380	0.000
DIA5	0.920	0.018	52.291	0.000
DIA6	0.885	0.022	39.729	0.000
DIA7	0.875	0.037	23.380	0.000

THRESHOLDS IN STANDARDIZED METRIC = threshold/SD(Y)

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
DIA1$1	-0.347	0.048	-7.303	0.000
DIA2$1	-0.982	0.056	-17.409	0.000
DIA3$1	-0.739	0.051	-14.373	0.000
DIA4$1	-0.404	0.045	-8.928	0.000
DIA5$1	-0.397	0.048	-8.348	0.000
DIA6$1	-0.626	0.050	-12.558	0.000
DIA7$1	-1.590	0.080	-19.949	0.000

R-SQUARE = standardized loading^2

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
DIA1	0.851	0.033	25.856	0.000
DIA2	0.883	0.034	26.278	0.000
DIA3	0.850	0.034	25.311	0.000
DIA4	0.945	0.024	40.190	0.000
DIA5	0.846	0.032	26.145	0.000
DIA6	0.784	0.039	19.865	0.000
DIA7	0.766	0.066	11.690	0.000

(output from same 2PL model continued)

IRT PARAMETERIZATION IN TWO-PARAMETER LOGISTIC METRIC WHERE THE LOGIT = DISCRIMINATION*(THETA - DIFFICULTY)

Item Discriminations = SLOPE OF ICC AT P=.50 (difficulty location)

IADL BY	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
DIA1	4.328	0.560	7.725	0.000
DIA2	4.978	0.808	6.159	0.000
DIA3	4.323	0.570	7.579	0.000
DIA4	7.511	1.696	4.429	0.000
DIA5	4.248	0.527	8.062	0.000
DIA6	3.451	0.401	8.600	0.000
DIA7	3.283	0.601	5.467	0.000

Item Difficulties = LOCATION OF ITEM ON LATENT TRAIT at P=.50, LOGIT=0

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
DIA1$1	-0.376	0.052	-7.298	0.000
DIA2$1	-1.045	0.065	-15.978	0.000
DIA3$1	-0.801	0.059	-13.562	0.000
DIA4$1	-0.415	0.047	-8.849	0.000
DIA5$1	-0.432	0.052	-8.296	0.000
DIA6$1	-0.708	0.060	-11.889	0.000
DIA7$1	-1.816	0.126	-14.454	0.000

USING RESULTS FROM IFA MODEL (LEFT PANEL):

IFA model: Logit(y) = -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

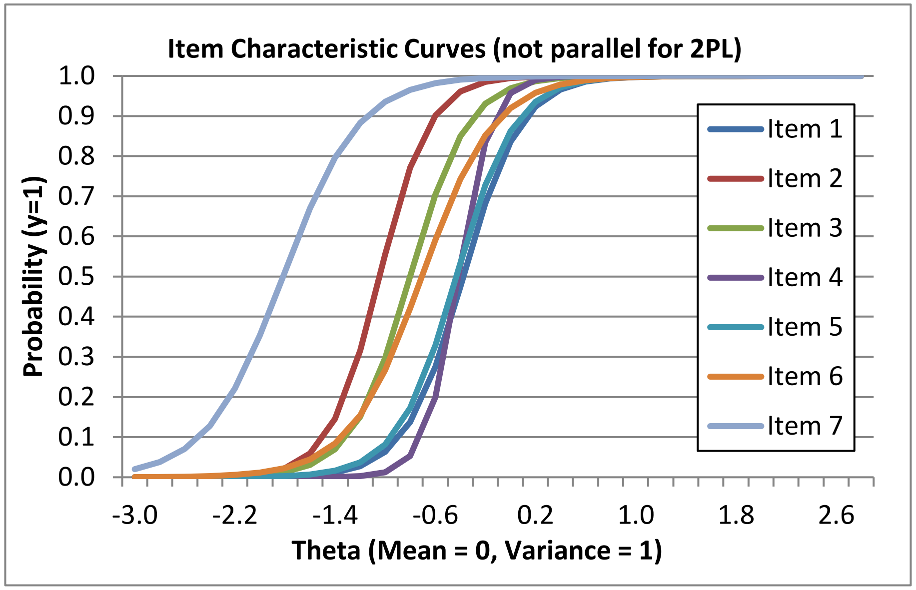
Loading = regression of item logit on Theta
= change in logit(y) for a one-unit change in Theta

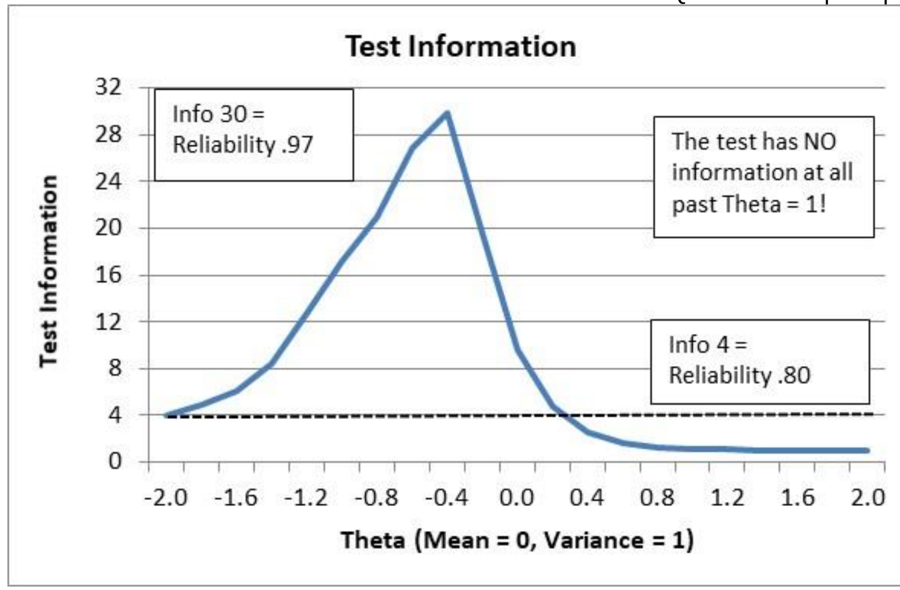
IFA Models:
Logit (DIA1=1) = 1.629 + 4.328(Theta) → if Theta=0, prob(y=1)= .836
Logit (DIA7=1) = 5.962 + 3.283(Theta) → if Theta=0, prob(y=1)= .997

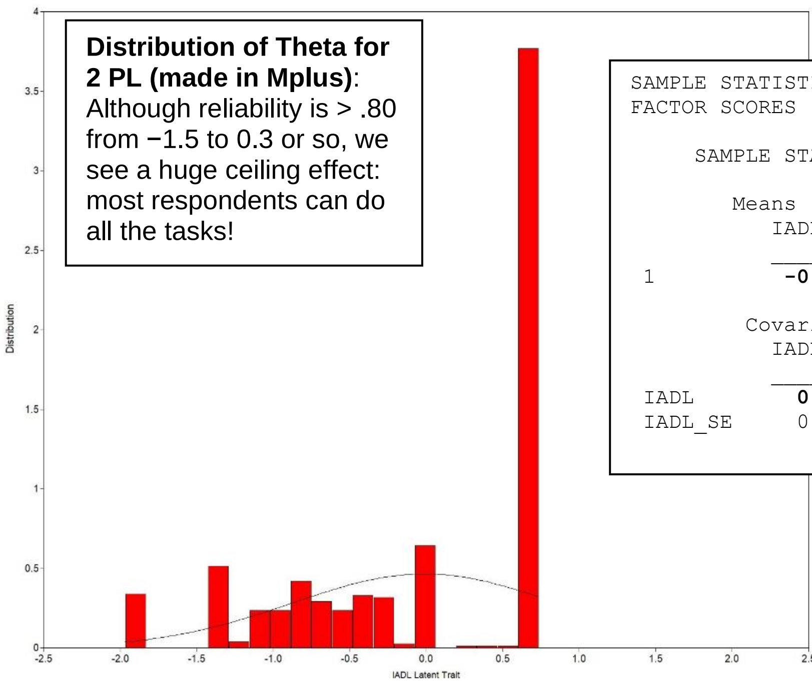
USING RESULTS FROM IRT MODEL (RIGHT PANEL):

IRT model: Logit(y=1) = a(theta - difficulty)
a = discrimination (rescaled slope) = loading/1.7
b = difficulty (location on latent metric) = threshold/loading

IRT Models:
Logit (DIA1=1) = 4.328*(Theta - -0.376)→ if Theta=0, prob(y=1)= .836
Logit (DIA7=1) = 3.283*(Theta - -1.816)→ if Theta=0, prob(y=1)= .997

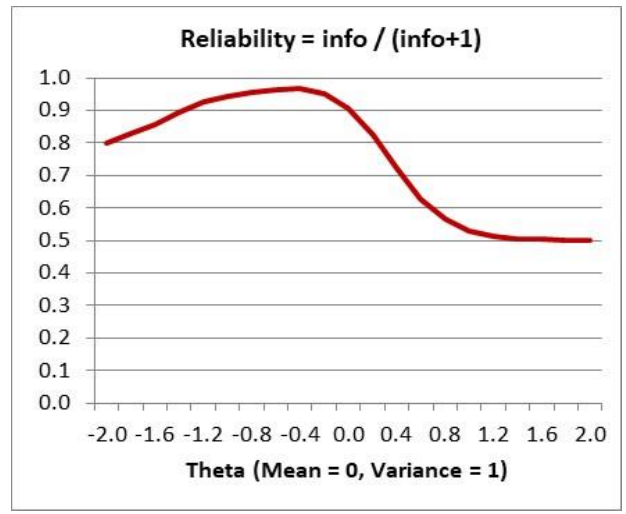




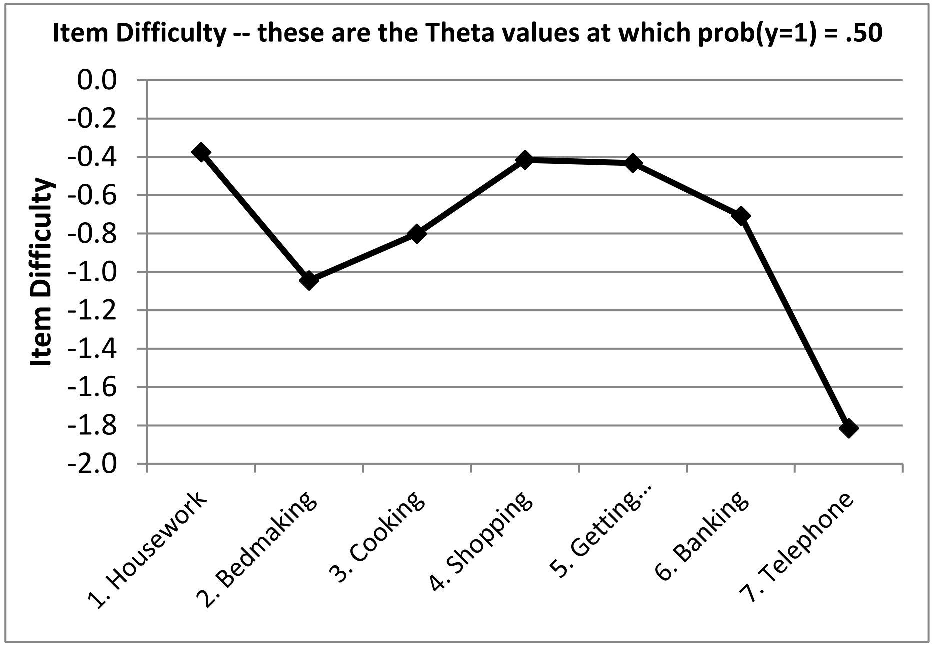


Distribution of Theta for 2 PL (made in Mplus): Although reliability is > .80 from −1.5 to 0.3 or so, we see a huge ceiling effect: most respondents can do all the tasks!

```
SAMPLE STATISTICS FOR ESTIMATED
FACTOR SCORES

     SAMPLE STATISTICS

        Means
           IADL          IADL_SE
        ________      ________
 1        -0.009         0.471

        Covariances
           IADL          IADL_SE
        ________      ________
 IADL        0.741
 IADL_SE     0.124         0.038
```



The predicted theta values are supposed to have a mean of 0 and a variance of 1, but this table shows that they have a variance of only .741 instead. Such shrinkage is why it can be problematic to use these estimated theta scores as observed variables in other analyses.



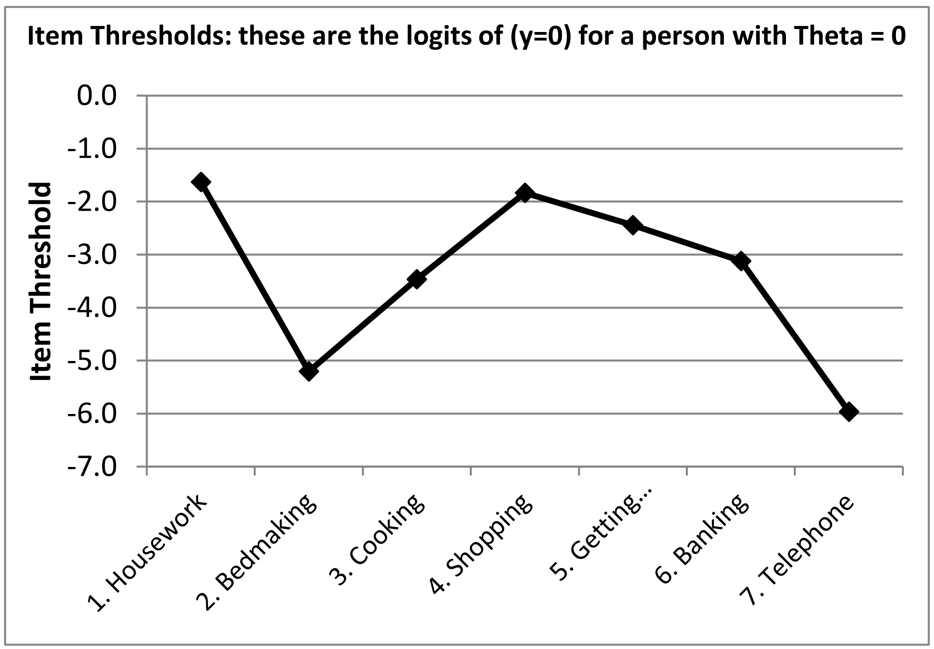
Plots of item parameters and predicted probabilities of item responses (made in excel):

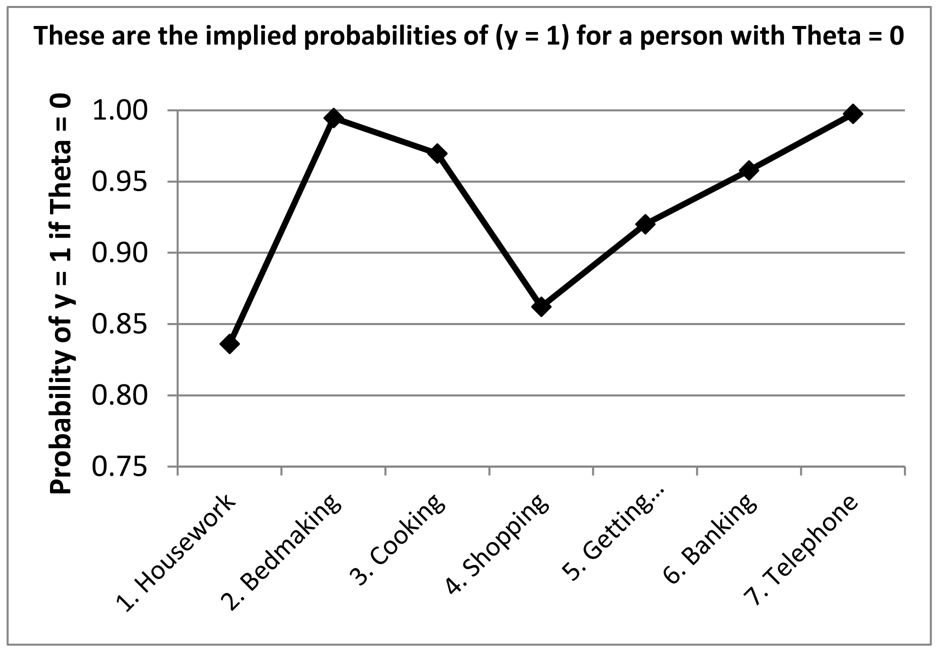
Top Left: Note that no items are available to measure above-average abilities well! The item difficulty for most items covers values of Theta between −1.0 to −0.5.

Bottom Left: These are the thresholds for each item, or the logit of (y=0) if Theta=0. These are hard to interpret as is….

Bottom Right: These are the probability of y=1 if Theta=0, as given by 1 – [exp(threshold) / (1+(exp(threshold))]

See excel workbook for calculations and plots





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:  Binary Models using Limited-Info WLSMV
DATA:   FILE = Example5.csv;  ! Don't need path if data in same folder
VARIABLE:   NAMES = case dia1-dia7;      ! All vars in data
            USEVARIABLES = dia1-dia7;    ! All vars in model
            CATEGORICAL = dia1-dia7;     ! 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:
! Factor loadings all estimated in 2PL
  IADL BY dia1-dia7*;
! Item thresholds all estimated
  [dia1$1-dia7$1*];
! Item error variances fixed at 1 for identification
  dia1-dia7@1;
! Factor variance=1 and mean=0 for identification
  IADL@1; [IADL@0];

MODEL FIT INFORMATION
Number of Free Parameters                        14
Chi-Square Test of Model Fit
          Value                              54.820*
          Degrees of Freedom                     14
          P-Value                            0.0000

*   The chi-square value for MLM, MLMV, MLR, ULSMV, WLSM and WLSMV
cannot be used for chi-square difference testing in the regular way.
MLM, MLR and WLSM chi-square difference testing is described on the
Mplus website.  MLMV, WLSMV, and ULSMV difference testing is done
using the DIFFTEST option.

RMSEA (Root Mean Square Error Of Approximation)
          Estimate                            0.068
          90 Percent C.I.                     0.049  0.087
          Probability RMSEA <= .05            0.055
CFI/TLI
          CFI                                 0.997
          TLI                                 0.995
Chi-Square Test of Model Fit for the Baseline Model
          Value                            12351.798
          Degrees of Freedom                     21
          P-Value                            0.0000
SRMR (Standardized Root Mean Square Residual)
          Value                               0.037
```

```
TITLE:  Binary Models using Limited-Info WLSMV
DATA:   FILE = Example5.csv;  ! Don't need path if data in same folder
VARIABLE:   NAMES = case dia1-dia7;      ! All vars in data
            USEVARIABLES = dia1-dia7;    ! All vars in model
            CATEGORICAL = dia1-dia7;     ! 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:
! Factor loadings all EQUAL in 1PL
  IADL BY dia1-dia7* (loading);
! Item thresholds all estimated
  [dia1$1-dia7$1*];
! Item error variances fixed at 1 for identification
  dia1-dia7@1;
! Factor variance=1 and mean=0 for identification
  IADL@1; [IADL@0];

MODEL FIT INFORMATION
Number of Free Parameters                         8
Chi-Square Test of Model Fit
          Value                              64.889*
          Degrees of Freedom                     20
          P-Value                            0.0000


Chi-Square Test for Difference Testing
          Value                              17.874
          Degrees of Freedom                      6
          P-Value                            0.0066

RMSEA (Root Mean Square Error Of Approximation)
          Estimate                            0.059
          90 Percent C.I.                     0.044  0.076
          Probability RMSEA <= .05            0.154
CFI/TLI
          CFI                                 0.996
          TLI                                 0.996
SRMR (Standardized Root Mean Square Residual)
          Value                               0.056
```

The Chi-Square for Difference Testing tells us directly that the 2P version of the binary model fits significantly better than 1P (now using WLSMV, but same results as when using ML).

Here are the parameter estimates under WLSMV Theta Parameterization (Probit) for the 2P model for binary items

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
DIA1	2.686	0.317	8.461	0.000
DIA2	2.937	0.491	5.979	0.000
DIA3	2.806	0.386	7.274	0.000
DIA4	3.659	0.577	6.338	0.000
DIA5	2.485	0.294	8.457	0.000
DIA6	1.990	0.223	8.943	0.000
DIA7	1.570	0.299	5.250	0.000

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

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
DIA1$1	-1.004	0.179	-5.607	0.000
DIA2$1	-3.093	0.479	-6.458	0.000
DIA3$1	-2.224	0.308	-7.227	0.000
DIA4$1	-1.584	0.299	-5.303	0.000
DIA5$1	-1.057	0.174	-6.073	0.000
DIA6$1	-1.390	0.166	-8.360	0.000
DIA7$1	-2.944	0.397	-7.409	0.000

STDYX MODEL RESULTS (STANDARDIZED IFA MODEL SOLUTION)

FACTOR LOADINGS IN STANDARDIZED METRIC = loading*SD(Theta)/SD(Y)

IADL BY	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value	Scale Factors
DIA1	0.937	0.013	69.490	0.000	
DIA2	0.947	0.016	57.546	0.000	
DIA3	0.942	0.015	64.560	0.000	
DIA4	0.965	0.011	91.204	0.000	
DIA5	0.928	0.015	60.668	0.000	
DIA6	0.894	0.020	44.369	0.000	
DIA7	0.843	0.046	18.191	0.000	

Thresholds IN STANDARDIZED METRIC = threshold/SD(Y)

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value	Scale Factors
DIA1$1	-0.350	0.052	-6.790	0.000	
DIA2$1	-0.997	0.061	-16.472	0.000	
DIA3$1	-0.746	0.056	-13.331	0.000	
DIA4$1	-0.417	0.052	-8.041	0.000	
DIA5$1	-0.395	0.051	-7.674	0.000	
DIA6$1	-0.624	0.054	-11.647	0.000	
DIA7$1	-1.582	0.081	-19.624	0.000	

R-SQUARE = standardized loading^2

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value	Scale Factors
DIA1	0.878	0.025	34.745	0.000	0.349
DIA2	0.896	0.031	28.773	0.000	0.322
DIA3	0.887	0.027	32.280	0.000	0.336
DIA4	0.931	0.020	45.602	0.000	0.264
DIA5	0.861	0.028	30.334	0.000	0.373
DIA6	0.798	0.036	22.185	0.000	0.449
DIA7	0.711	0.078	9.095	0.000	0.537

(output from same 2P model continued)

IRT PARAMETERIZATION IN TWO-PARAMETER PROBIT METRIC WHERE THE PROBIT IS DISCRIMINATION*(THETA - DIFFICULTY)

Item Discriminations

IADL BY				
DIA1	2.686	0.317	8.461	0.000
DIA2	2.937	0.491	5.979	0.000
DIA3	2.806	0.386	7.274	0.000
DIA4	3.659	0.577	6.338	0.000
DIA5	2.485	0.294	8.457	0.000
DIA6	1.990	0.223	8.943	0.000
DIA7	1.570	0.299	5.250	0.000

Item Difficulties

DIA1$1	-0.374	0.055	-6.743	0.000
DIA2$1	-1.053	0.069	-15.358	0.000
DIA3$1	-0.793	0.062	-12.867	0.000
DIA4$1	-0.433	0.054	-7.982	0.000
DIA5$1	-0.425	0.056	-7.606	0.000
DIA6$1	-0.699	0.063	-11.084	0.000
DIA7$1	-1.875	0.154	-12.188	0.000

Logit = 1.7*probit, or Probit = Logit/1.7

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

IRT model: Probit(y=1) = a(theta - difficulty)
a = discrimination (rescaled slope) = loading/1
b = difficulty (location on latent metric) = threshold/loading

LOCAL FIT VIA RESIDUALS FOR CORRELATION LEFTOVER TETRACHORIC CORRELATION (HOW FAR OFF MODEL PREDICTIONS ARE FROM ESTIMATED DATA CORRELATIONS)

Residuals for Covariances/Correlations/Residual Correlations

	DIA1	DIA2	DIA3	DIA4	DIA5	DIA6
DIA1						
DIA2	0.029					
DIA3	0.038	0.029				
DIA4	-0.022	-0.040	-0.046			
DIA5	-0.032	-0.033	**-0.103**	0.029		
DIA6	-0.052	-0.056	-0.046	0.026	0.032	
DIA7	**-0.111**	-0.002	0.010	0.031	-0.027	0.064

Extensive Results Section (in which model fit via WLSMV is reported first, followed by full-information MML as "better" version of the 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 associations among seven binary items measuring physical capacity was conducted using Item Factor Analysis (IFA) in M*plus* v 8.8 (Muthén and Muthén, 1998–2017). These models use a link function (i.e., logit or probit) and a conditional Bernoulli response distribution to predict the conditional probability of a response = 1 (instead of 0) from a linear model as $Link(y_{is} = 1) = -\tau_i + \lambda_i F_s$. In this item model, $-\tau_i$ is the item-specific threshold, which when multiplied by −1 becomes an intercept that gives the link-transformed probability of response $y_{is} = 1$ (for item i and subject s) at a latent trait score F for subject s of 0, and λ_i is an item-specific factor loading 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' binary response formats.

The current gold standard of estimation for IFA models is marginal maximum likelihood (MML), in which the term *marginal* refers to the full-information process of marginalizing over all possible trait values for each person in the analysis using adaptive Gaussian quadrature with 15 quadrature points per latent trait. 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 2^7 = 128 possible cells. Consequently, no measures of absolute fit would be valid for the current sample of 635 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 $\chi 2$ (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 tetrachoric correlation matrix using the cross-tabulation of responses for each possible pair of items. The IFA models are then fitted to the estimated tetrachoric correlation matrix, such that traditional measures of global and local absolute fit (i.e., traditional in confirmatory factor analyses of continuous responses) can be computed by comparing the model-predicted and data-estimated tetrachoric correlation matrices. In addition to $\chi 2$ tests of absolute fit, WLSMV also provides 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 tetrachoric 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 tetrachoric correlations), in which SRMR and RMSEA values ≤ .06 traditionally indicate excellent 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 tetrachoric 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 estimated for the 7 binary items using WLSMV, in which the latent trait mean and variance were fixed for identification to 0 and 1, respectively, and separate thresholds and factor loadings were estimated for each item. This model exhibited acceptable fit by every measure except the $\chi 2$ test of absolute fit, $\chi 2$ (14) = 54.820, p < .001, CFI = .997, SRMR = .037, RMSEA = .068 [CI = .049–.087, p = .055]. Examination of local misfit revealed all discrepancies between the model-predicted and data-estimated tetrachoric correlations were less than .113 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) = 17.874, p = .007, 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 (given the presence of missing item-level responses).

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 $b_i = \tau_i / \lambda_i$; 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 ~1.8 SD below the mean to 0.20 SD above the mean, after which point reliability drops off precipitously due to a lack of items with difficulty levels above 0.

(See Example 5 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.