This is a supplementary resource to Measurement Theory and Applications for the Social Sciences, by Deborah L. Bandalos.
Copyright © 2018 by The Guilford Press.

Estimating IRT parameters Using Flexmirt

In this document I explain how to use the Flexmirt program to estimate IRT models. A free 2-week trial of Flexmirt is available at: https://vpgcentral.com/software/irt-software/purchase/. I have included a comprehensive set of analyses in this document, and most of these go beyond what we were able to present in one book chapter. Before running IRT software, I recommend that you gain a greater understanding of through study of resources such as Bock, R. D., & Gibbons, R. D. (2021). *Item response theory*. Hoboken, NJ: Wiley or de Ayala, R. J. (2009). *The Theory and Practice of Item Response Theory*. New York: Guilford Press.

The data for these examples are in the simulated dataset “IRTexample.dat”. There were 5000 examinees. The format is space-delimited, with ID followed by 20 four-option multiple-choice items (scored 0 = wrong, 1 = correct) and 5 three-point constructed response options (scored as wrong, partly correct, correct). This test is very short, so the ability estimates will not be very reliable, but a short test makes cleaner, shorter output.

First I will show the 3PL, 2PL, and 1PL models for the 20 dichotomously-scored items. I will add in the graded response and partial credit models in the last sections.

Following the conventions in the other online documents, Flexmirt commands are bolded; entries shown in unbolded text should reflect the specifications appropriate to the user’s data and model.

In Flexmirt comments are preceded by //, so that convention is followed here.

As in the SAS and M*plus* programs, commands in Flexmirt must end with a semi-colon.

Example 1: 3PL Model

Flexmirt syntax (ex3PL.flexmirt):

<Project>

The line above (**<Project>**) is one of four section headers that must be included in a Flexmirt syntax file. The **<Project>** section header contains the **Title** and **Description** commands. The other sections headers (**<Options>, <Groups>**, and **<Constraints>)** and their commands are shown subsequently.

Title = "Simulated Data";
Description = "3PL";

// The Title and Description can have any words within the quotation marks

<Options>
Mode = Calibration;

//The command **Mode=Calibration** is the specification used for estimating item parameters. **Model = Scoring** is used to score tests.

This is a supplementary resource to Measurement Theory and Applications for the Social Sciences, by Deborah L. Bandalos.
Copyright © 2018 by The Guilford Press.

Quadrature = 17, 4.0;

//The command above indicates that 17 quadrature points ranging from -4 to 4, at equal intervals should be used in estimation.

Score=EAP**;**

// This command indicates that Expected A-Posteriori (EAP) estimates should be used (alternatives are Maximum Likelihood (ML) or Modal A-Posteriori (MAP)).

SaveSco=Yes**;**

//This command indicates that scores should be saved to a file.

Processors=4;

//This command indicates that 4 processors should be used for computing. The default is to use only use 1 processor.

NormalMetric3PL = Yes;

//This commands puts the logistic parameters into the normal metric, with D = 1.7 (see Equation 14.4 in the text).

GOF=Complete;

//The Goodness of Fit (GOF) command controls which model fit indices will be printed. **GOF=Complete** will print several fit indices.

<Groups>

// The commands under **<Groups>** are needed for multiple group modeling options, such as DIF analyses. These commands allow the researcher to define their groups and group names. Here I only have one group. For a multiple-group model, data for each group must be in a separate file.

%Group1%

// Any group name can be inside the % %;

File = "IRTexample.dat";
Varnames = ID, I1-I20;

/The two command above provide the file name and variable names. Variables in the data file must be separated by spaces.

This is a supplementary resource to Measurement Theory and Applications for the Social Sciences, by Deborah L. Bandalos.
Copyright © 2018 by The Guilford Press.

Select = I1-I20;

//The **Select** command is used to select variables for analyses. Here, only the 20 items are selected as the ID number should not be analyzed as an item!

N = 5000;
NCATS(I1-I20)=2;

//The two commands above indicate the sample size and the number of categories for each item. The data are already scored as either 0 or 1—if the original multiple-choice categories (ABCD) were maintained, there would be four valid categories.

Model(I1-I20)=ThreePL;

//The **Model** command is used to choose the desired model. Here I request a 3PL model.

EmpHist=Yes;

//This command indicates that the shape of the theta distribution should be estimated, instead of assuming it is normal.

<Constraints>

Prior (I1-I20), **Slope:** Normal(1.5,0.5);

//The **Prior** command is used to specify the prior distributions for the item parameters. These specifications were beyond the scope of the text, but analysts may put prior distributions on the item parameters (similar to using priors on the ability distribution when estimating examinees' abilities). Priors generally make little difference in the 1PL and 2PL models, but some 3PL estimates will generally be unreasonable without priors.

Here, the prior distributions for the slope parameters of all 20 items (I1 – I20) are specified to be normal with a mean of 1.5 and a standard deviation of 0.5 (Normal(1.5,0.5)).

The prior is expressed for a model with D = 1. I specified D = 1.7 (with the **NormalMetric3PL = Yes** command), so the mean is really .88 (1.5/1.7), and not the values of 1.5 that is specified. The researcher can specify a different prior for each item; I used the same prior for all items for simplicity.

Prior (I1-I20), **Guessing:** Beta(21,81);

//In the commands above, the prior distribution for the guessing parameter is specified as a beta distribution. The parameters (21,81) will yield a mean of .20. Typically, researchers choose a value somewhat lower than chance guessing, which for these 4-option multiple choice items would be .25.

This is a supplementary resource to Measurement Theory and Applications for the Social Sciences, by Deborah L. Bandalos.
Copyright © 2018 by The Guilford Press.

Running these commands will create an output file with the extension "-irt.txt". Because my syntax file was named "ex3PL.flexmirt," the output file is named "ex3PL-irt.txt". The file begins by confirming information in the syntax file.

The table below shows the iteration history. The 1st column is the iteration number, the 2nd column is the maximum change between iterations, multiplied by 10,000, the 3rd column indicates the number of the parameter that changed the most, and the 4th column is the marginal log-likelihood of the data given the current item parameter estimates.

1 :	-27058.8213	(29) :	128785.4774
2 :	-7814.6852	(29) :	102495.0297
3 :	-3925.5447	(29) :	102119.2479
	Iterations 4 – 110 are omitted to save space		
111 :	-1.0172	(11) :	101904.3618
112 :	-1.0096	(11) :	101904.3646
113 :	-0.9923	(11) :	101904.3676

This output indicates that the algorithm took 113 iterations to reach the default stopping criterion (maximum parameter change < 0.0001). In the last iteration parameter #11 changed by −0.00009923 (the number displayed in the window is multiplied by 10,000). From the output in the next table, we can see that parameter #11 is the difficulty estimate for item 4. This is a very difficult (*b*=1.83) and not very discriminating (*a*=.77) item, so it may have been a bit harder to estimate than other items.

In the table below, the 1st column shows the item number and the 2nd column shows the item label. Following these, the information for each item parameter is displayed as follows: the parameter number assigned by Flexmirt, the item parameter estimate, and its standard error.

For example, the *a* parameter for item 4 is numbered as P#12 and its estimate is 0.77 with a standard error of .10. The *b* parameter for this item is numbered P#11 and is estimated as 1.83 with a standard error of .09. Finally, the *c* parameter (labeled “*g*” by Flexmirt) is numbered as P#10 with an estimate of .20 and standard error of .02.

This is a supplementary resource to Measurement Theory and Applications for the Social Sciences, by Deborah L. Bandalos.
Copyright © 2018 by The Guilford Press.

Parameter #, corresponding to parameters in the iteration history above

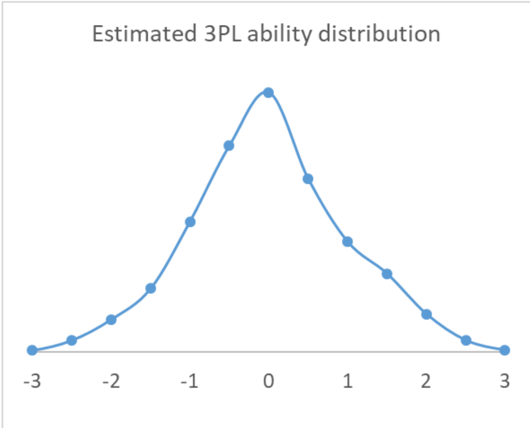
			estimate of *a*			estimate of *b*			estimate of *c*	
Item	Label	P#	a	s.e.	P#	b	s.e.	P#	g	s.e.
1	I1	3	0.54	0.03	2	-1.88	0.14	1	0.20	0.04
2	I2	6	1.11	0.11	5	1.40	0.05	4	0.19	0.01
3	I3	9	0.65	0.06	8	0.72	0.09	7	0.21	0.03
4	I4	12	0.77	0.10	11	1.83	0.09	10	0.20	0.02
5	I5	15	0.60	0.04	14	-1.46	0.12	13	0.21	0.04
6	I6	18	0.81	0.05	17	0.00	0.07	16	0.18	0.03
7	I7	21	1.34	0.09	20	0.31	0.04	19	0.18	0.02
8	I8	24	1.22	0.08	23	-1.34	0.07	22	0.20	0.04
9	I9	27	1.13	0.08	26	0.19	0.05	25	0.18	0.02
10	I10	30	2.10	0.19	29	1.32	0.03	28	0.09	0.01
11	I11	33	1.23	0.09	32	0.85	0.04	31	0.19	0.01
12	I12	36	1.39	0.09	35	0.10	0.04	34	0.18	0.02
13	I13	39	1.49	0.10	38	-1.20	0.06	37	0.19	0.03
14	I14	42	1.23	0.11	41	1.23	0.05	40	0.19	0.01
15	I15	45	1.17	0.07	44	-1.24	0.07	43	0.21	0.04
16	I16	48	1.22	0.09	47	-0.71	0.07	46	0.24	0.03
17	I17	51	1.08	0.07	50	-0.66	0.06	49	0.18	0.03
18	I18	54	0.68	0.04	53	-1.00	0.10	52	0.20	0.04
19	I19	57	0.63	0.04	56	-1.75	0.13	55	0.22	0.04
20	I20	60	0.50	0.05	59	0.46	0.14	58	0.23	0.04

In the process of estimating the item parameters, Flexmirt also estimates the shape of the ability distribution (although obtaining point estimates for individual examinees is a separate, optional step). I graphed the distribution in Excel from the data below that is provided by Flexmirt; Flexmirt does not provide graphs.

This is a supplementary resource to Measurement Theory and Applications for the Social Sciences, by Deborah L. Bandalos.
Copyright © 2018 by The Guilford Press.

```
Quadrature
    Point             Ordinate

    -4.000        9.8325857e-036
    -3.500        8.8384760e-005
    -3.000        1.6469556e-003
    -2.500        1.0513528e-002
    -2.000        2.9106122e-002
    -1.500        5.6591390e-002
    -1.000        1.1646411e-001
    -0.500        1.8398207e-001
     0.000        2.3137141e-001
     0.500        1.5447659e-001
     1.000        9.8471351e-002
     1.500        7.0105728e-002
     2.000        3.4076680e-002
     2.500        1.0784676e-002
     3.000        2.1359160e-003
     3.500        1.8508957e-004
     4.000        2.1106930e-026
```



After the quadrature distribution, the mean and variance of the ability distribution are printed. These are fixed to 0 and 1 to identify the metric's center point and unit size.

In the table below, “mu” indicates the ability distribution’s mean, “s2” indicates its variance, and “sd” its standard deviation. Because these values are fixed (rather than estimated) to set the scale metric, there are no standard errors and no parameter number (no P#). If there were multiple groups, means and variances would be estimated for every group after the first.

Group Parameter Estimates:

Group	Label	P#	mu	s.e.	P#	s2	s.e.	sd	s.e.
1	Group1		0.00	----		1.00	----	1.00	----

Information and Standard Error of Ability Estimates

The information function for each item at selected ability levels is shown near the end of the output, followed by the overall test information and standard error.

The standard error is 1/(square root of information). Adding up the item information in any column should yield the test information for that ability level. However, you'll notice the sum is one less than the test information. This is because Flexmirt calculates the information for Bayesian ability estimates. This information function is strictly appropriate

This is a supplementary resource to Measurement Theory and Applications for the Social Sciences, by Deborah L. Bandalos.
Copyright © 2018 by The Guilford Press.

for the MAP estimates but is similar for the EAP estimates (Thissen & Orlando, 2001, p. 118). The prior adds one at each ability level because the second derivative of the standard normal function with respect to θ is a constant of one. Adding one has little impact where there is more information from the items but has a bigger impact where there is less information. If you want the information for maximum likelihood estimates, subtract one and re-calculate the standard errors.

This is a supplementary resource to Measurement Theory and Applications for the Social Sciences, by Deborah L. Bandalos.
Copyright © 2018 by The Guilford Press.

```
Item Information Function Values at 15 Values of theta from -2.80 to 2.80 for Group 1: Group1

                                                      Theta:
Item Label -2.8   -2.4   -2.0   -1.6   -1.2   -0.8   -0.4   -0.0    0.4    0.8    1.2    1.6    2.0    2.4    2.8
 1     I1   0.10   0.12   0.14   0.15   0.14   0.13   0.11   0.08   0.06   0.05   0.04   0.03   0.02   0.01   0.01
 2     I2   0.00   0.00   0.00   0.00   0.00   0.00   0.01   0.05   0.15   0.34   0.56   0.63   0.51   0.32   0.18
 3     I3   0.00   0.00   0.01   0.02   0.03   0.06   0.10   0.14   0.18   0.20   0.20   0.18   0.14   0.11   0.08
 4     I4   0.00   0.00   0.00   0.00   0.00   0.01   0.01   0.03   0.07   0.13   0.20   0.27   0.30   0.28   0.22
 5     I5   0.07   0.11   0.14   0.16   0.17   0.17   0.14   0.12   0.09   0.07   0.05   0.03   0.02   0.02   0.01
 6     I6   0.00   0.01   0.02   0.05   0.11   0.19   0.27   0.33   0.33   0.27   0.20   0.14   0.09   0.05   0.03
 7     I7   0.00   0.00   0.00   0.00   0.02   0.09   0.31   0.69   0.92   0.74   0.42   0.20   0.09   0.04   0.01
 8     I8   0.03   0.11   0.32   0.60   0.74   0.60   0.36   0.19   0.09   0.04   0.02   0.01   0.00   0.00   0.00
 9     I9   0.00   0.00   0.00   0.01   0.05   0.15   0.36   0.59   0.66   0.52   0.33   0.18   0.09   0.04   0.02
10     I10  0.00   0.00   0.00   0.00   0.00   0.00   0.00   0.01   0.12   0.86   2.43   2.20   0.85   0.23   0.06
11     I11  0.00   0.00   0.00   0.00   0.00   0.01   0.06   0.21   0.48   0.74   0.72   0.49   0.26   0.13   0.06
12     I12  0.00   0.00   0.00   0.01   0.04   0.17   0.52   0.92   0.94   0.60   0.29   0.13   0.05   0.02   0.01
13     I13  0.01   0.05   0.22   0.67   1.10   0.96   0.52   0.22   0.09   0.03   0.01   0.00   0.00   0.00   0.00
14     I14  0.00   0.00   0.00   0.00   0.00   0.00   0.02   0.07   0.21   0.49   0.73   0.70   0.47   0.26   0.12
15     I15  0.02   0.08   0.24   0.49   0.66   0.60   0.40   0.22   0.11   0.05   0.02   0.01   0.00   0.00   0.00
16     I16  0.00   0.01   0.04   0.15   0.39   0.64   0.66   0.47   0.26   0.13   0.06   0.03   0.01   0.01   0.00
17     I17  0.00   0.02   0.06   0.17   0.36   0.55   0.59   0.46   0.30   0.16   0.08   0.04   0.02   0.01   0.00
18     I18  0.04   0.07   0.12   0.17   0.21   0.23   0.22   0.18   0.14   0.10   0.07   0.05   0.03   0.02   0.01
19     I19  0.10   0.14   0.17   0.19   0.18   0.16   0.13   0.10   0.07   0.05   0.03   0.02   0.02   0.01   0.01
20     I20  0.01   0.01   0.02   0.03   0.05   0.06   0.08   0.10   0.11   0.12   0.11   0.10   0.09   0.07   0.06

Test Info: 1.39   1.74   2.51   3.89   5.26   5.78   5.88   6.18   6.37   6.70   7.58   6.42   4.06   2.62   1.89
ExpSE:      0.85   0.76   0.63   0.51   0.44   0.42   0.41   0.40   0.40   0.39   0.36   0.39   0.50   0.62   0.73

Marginal reliability for response pattern scores: 0.82
```

The marginal reliability is obtained by averaging over the ability distribution. The marginal reliability of .82 is the correlation between estimated and true ability.

For Bayesian scores, one estimation method is reliability = $\dfrac{\sigma^2_{\hat{\theta}_{EAP}}}{\sigma^2_{\hat{\theta}_{EAP}} + \bar{\hat{\sigma}}^2_{err}}$, where $\sigma^2_{\hat{\theta}_{EAP}}$ is the variance in the estimated abilities and $\bar{\hat{\sigma}}^2_{err}$ is the mean of the squared estimated errors. When the scale is identified by setting the ability variance to one, $\sigma^2_{\hat{\theta}_{EAP}} + \bar{\hat{\sigma}}^2_{err} \approx 1$, so another calculation is reliability = $1 - \sigma^2_{\hat{\theta}_{EAP}}$. For ML scores, reliability = $\dfrac{\sigma^2_{\hat{\theta}_{ML}} - \bar{\hat{\sigma}}^2_{err}}{\sigma^2_{\hat{\theta}_{ML}}}$, similar to classical test theory,

This is a supplementary resource to Measurement Theory and Applications for the Social Sciences, by Deborah L. Bandalos.
Copyright © 2018 by The Guilford Press.

because $\sigma^2_{\hat{\theta}_{ML}}$ is the variance of the observed scores. Reliability of the ability estimates will typically be slightly higher than the reliability of the observed summed scores. For these data, coefficient alpha for the summed scores was 0.78, compared to the value of 0.82 for reliability of the EAP scores.

Ability Estimation

During the item parameter estimation, the shape of the ability distribution is estimated, but scores for individual examinees are not.

EAP scores can be requested in the same syntax file as the item parameter estimation, with the line: "**Score=**EAP;" and the line "**SaveSco=**Yes;" as shown on p. 2.

This will produce an output file with the extension "-sco.txt. Because the syntax file was named "ex3PL.flexmirt" the output file will be "ex3PL-sco.txt." The columns in this file will be in the following order: group number, observation number, the scale score, and standard error associated with the scale score.

To obtain the individual ML or MAP scores, the researcher must run a separate syntax file, reading in the item parameter estimates from the previous run. I copied the syntax for ***ex3PL.flexmirt*** and renamed the copy ***ex3PLscore.flexmirt*** and edited a few lines. You may get an error message if you used **emphist=Yes** in the calibration file (as I did on p. 3). If you do, change the last line in your item parameter file (called "ex3PL-prm.txt" in the example below) to read:

0 Group1 1 1 0 0.0000000 1.0000000

Example 1b: ML Scoring using Item Estimates from Previous Calibration

<Project>
Title = "Scoring from previous item estimates";
Description = "3PL"**;**
<Options>
 Mode = Scoring**; //** score using existing item estimates;
 MaxMLscore=5**; //** because the likelihood can be infinitely increasing or decreasing, give a maximum score;
 MinMLscore=-5**;**
 readPRMfile = "ex3PL-prm.txt"; *//read item parameters--these were saved in the calibration run with SaveSco. Default name is the base name of the syntax file, with "-prm.txt";*
 NormalMetric3PL = Yes; *//this needs to be the same as the calibration run, so if you did not use it in the calibration, omit it here;*
 SaveSco=Yes; *//save scores (theta estimates);*
 Score=ML; *//score with ML--could use MAP instead;*
<Groups>
%Group1%
 File = "IRTexample.dat";
 Varnames = ID, I1-I20;

This is a supplementary resource to Measurement Theory and Applications for the Social Sciences, by Deborah L. Bandalos.
Copyright © 2018 by The Guilford Press.

Select = I1-I20;
N = 5000;
NCATS(I1-I20)=2;
Model(I1-I20)=ThreePL;
<Constraints>
//even if you do not specify constraints, you still need the <Constraints> section;

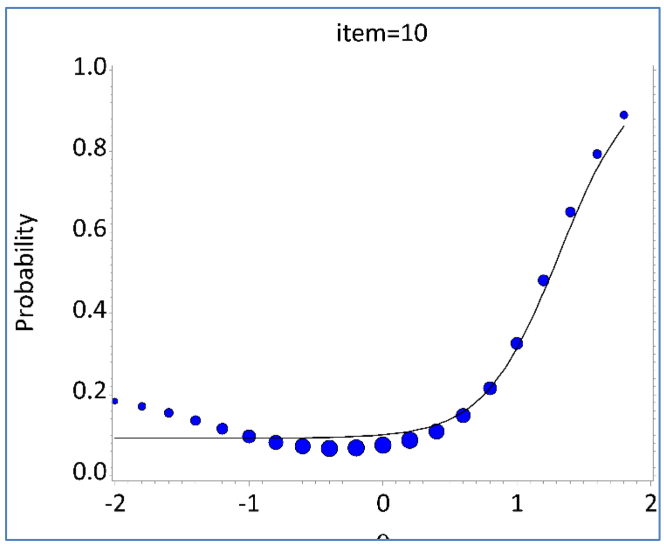
This is a supplementary resource to Measurement Theory and Applications for the Social Sciences, by Deborah L. Bandalos.
Copyright © 2018 by The Guilford Press.

Item Fit

Returning to the calibration run, the line "**GOF=Complete**;" in requests the Orlando-Thissen item fit index discussed in the text (p. 436). The output is shown below.

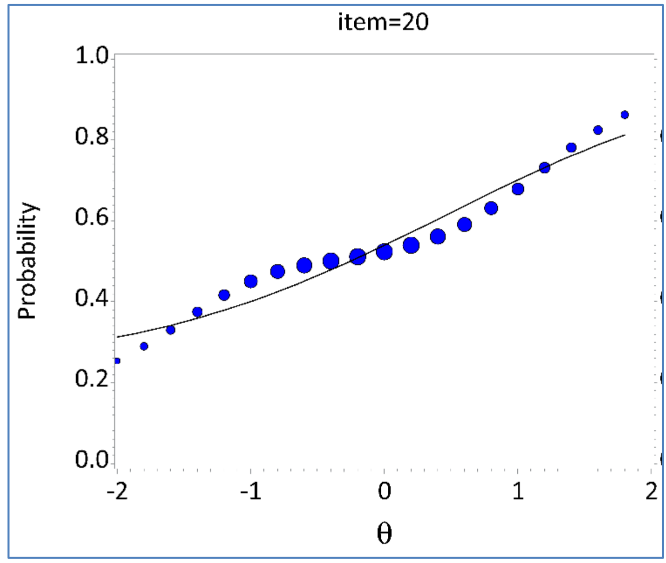
```
Orlando-Thissen-Bjorner Summed-Score Based Item Diagnostic Tables
and χ² values:

  Item 1   S-X2(16) = 17.3, p = 0.3695
  Item 2   S-X2(15) =  5.2, p = 0.9900
  Item 3   S-X2(16) = 20.1, p = 0.2145
  Item 4   S-X2(15) = 13.1, p = 0.5940
```



This is a supplementary resource to Measurement Theory and Applications for the Social Sciences, by Deborah L. Bandalos.
Copyright © 2018 by The Guilford Press.

```
Item 5    S-X2(16) = 14.2, p = 0.5860
Item 6    S-X2(15) = 15.5, p = 0.4168
Item 7    S-X2(15) = 22.0, p = 0.1063
Item 8    S-X2(14) =  7.1, p = 0.9322
Item 9    S-X2(15) =  9.6, p = 0.8417
Item 10   S-X2(15) = 88.2, p < 0.0001
Item 11   S-X2(15) = 10.3, p = 0.8014
Item 12   S-X2(15) = 18.3, p = 0.2487
Item 13   S-X2(13) = 22.8, p = 0.0437
Item 14   S-X2(15) =  6.3, p = 0.9736
Item 15   S-X2(14) = 10.5, p = 0.7241
Item 16   S-X2(15) = 18.7, p = 0.2271
Item 17   S-X2(14) = 11.7, p = 0.6280
Item 18   S-X2(16) = 17.4, p = 0.3615
Item 19   S-X2(16) = 21.5, p = 0.1584
Item 20   S-X2(15) = 73.5, p < 0.0001
```



Items 10 and 20 showed poor fit ($p < 0.0001$), so I graphed these items. In the graphs, examinees were grouped by ability, and the proportion correct was calculated within each group[1]. In the graph, the circles represent proportion correct and the solid line represents the predicted response function. On item 10, a higher proportion of low-ability students than middle-ability students answered correctly. Possibly, the low-ability students guessed randomly and the middle-ability students were drawn to a distractor that reflected partial knowledge or a common misconception.

On item 20, the item discriminated more for low-ability students, flattened somewhat, then discriminated better again at higher ability levels.

[1] Instead of assigning each examinee to a single ability group, I included each examinee in all ability groups, weighting by the examinee's posterior ability distribution (see Sueiro & Abad, 2011). This smoothed the observed proportion-correct function.

This is a supplementary resource to Measurement Theory and Applications for the Social Sciences, by Deborah L. Bandalos.
Copyright © 2018 by The Guilford Press.

2PL Model

There are two ways to run a 2PL model: a) by applying constraints to a 3PL model or b) as a dichotomous graded response model. First, I will illustrate a constrained 3PL model.

The syntax is the same as in the previous 3PL example, so I only provide comments for any new specifications.

Example 2a: 2PL using Constraints on 3PL (**ex2PL.flexmirt**):

<Project>

Title = "Simulated Data";
Description = "2PL as constrained 3PL";

<Options>

Mode = Calibration;
Quadrature = 17, 4.0;
Processors=4;
NormalMetric3PL = Yes;

//Even though I am running a 2PL model, to put the logistic parameters into the normal metric the keyword is NormalMetric3PL, not NormalMetric2PL;

<Groups>

%Group1%
File = "IRTexample.dat";
Varnames = ID, I1-I20;
Select = I1-I20;
N = 5000;
NCATS(I1-I20)=2;
Model(I1-I20)=ThreePL;

//I am running a constrained 3PL model, although the constraints below will change it to a 2PL model;

EmpHist=Yes;
Fix (I1-I20), **Guessing**;

// The **Fix** command allows parameters to be fixed at specific values. Here, I fix the guessing parameters, or lower asymptotes (**Guessing**) to the values indicated in the next line.

Value (I1-I20), **Guessing**, -999;
//The **Value** statement sets the logit of the lower asymptote to -999, which is effectively 0, for all items.

This is a supplementary resource to Measurement Theory and Applications for the Social Sciences, by Deborah L. Bandalos.
Copyright © 2018 by The Guilford Press.

Running these commands will create an output file with the extension "-irt.txt". Because my syntax file was named "ex2PL.flexmirt", the output file is named "ex2PL-irt.txt".

The parameters estimates and their standard errors are shown below. The order of the columns and their content is the same as in the 3PL example. However, note that all the *c* parameter estimates (labeled *g* in Flexmirt) are zero and have no standard error estimates or parameter numbers.

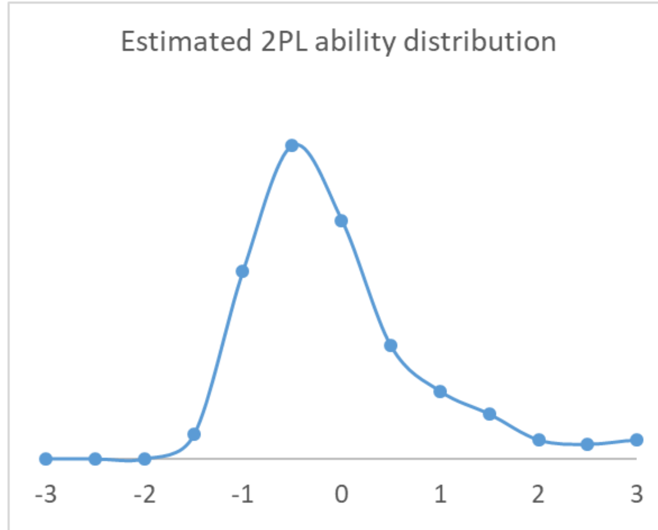
Item	Label	P#	a	s.e.	P#	b	s.e.	P#	g	s.e.
1	I1	2	0.67	0.04	1	-1.84	0.10		0.00	----
2	I2	4	0.46	0.02	3	1.09	0.07		0.00	----
3	I3	6	0.45	0.03	5	0.04	0.04		0.00	----
4	I4	8	0.31	0.02	7	1.64	0.12		0.00	----
5	I5	10	0.73	0.05	9	-1.52	0.07		0.00	----
6	I6	12	0.76	0.04	11	-0.43	0.03		0.00	----
7	I7	14	1.01	0.04	13	-0.14	0.02		0.00	----
8	I8	16	1.91	0.11	15	-1.16	0.03		0.00	----
9	I9	18	0.96	0.04	17	-0.24	0.02		0.00	----
10	I10	20	0.70	0.03	19	1.37	0.07		0.00	----
11	I11	22	0.67	0.03	21	0.37	0.04		0.00	----
12	I12	24	1.18	0.05	23	-0.30	0.02		0.00	----
13	I13	26	2.36	0.13	25	-1.06	0.02		0.00	----
14	I14	28	0.52	0.03	27	0.84	0.06		0.00	----
15	I15	30	1.71	0.09	29	-1.13	0.03		0.00	----
16	I16	32	1.46	0.07	31	-0.91	0.02		0.00	----
17	I17	34	1.32	0.06	33	-0.82	0.02		0.00	----
18	I18	36	0.81	0.04	35	-1.17	0.05		0.00	----
19	I19	38	0.81	0.05	37	-1.68	0.08		0.00	----
20	I20	40	0.40	0.03	39	-0.36	0.05		0.00	----

Notice that the *a*-parameters are smaller than they were for the 3PL model, especially for the most difficult items. This is because the slope flattens to better match the range where the probability is near the lower asymptote, perhaps due to correct guessing. This flattening is greater for the most difficult items because the "guessing" range covers a greater proportion of the examinees.

The shape of the ability distribution is also altered by using the 2PL model. The distribution becomes skewed to absorb some of the model misfit (Woods, 2008). The mean and variance are fixed to 0 and 1, but the distribution is not constrained to be normal. The quadrature points are printed and could be used to calculate skew, kurtosis, etc.

This is a supplementary resource to Measurement Theory and Applications for the Social Sciences, by Deborah L. Bandalos.
Copyright © 2018 by The Guilford Press.

```
 Quadrature
    Point                  Ordinate
    -4.000      8.2788297e-161
    -3.500      7.5758273e-104
    -3.000      1.3932379e-066
    -2.500      2.2683423e-037
    -2.000      1.0206495e-014
    -1.500      2.2623641e-002
    -1.000      1.7779159e-001
    -0.500      2.9614635e-001
     0.000      2.2562086e-001
     0.500      1.0655101e-001
     1.000      6.3775086e-002
     1.500      4.2178489e-002
     2.000      1.7783800e-002
     2.500      1.3799295e-002
     3.000      1.7771370e-002
     3.500      1.5958502e-002
     4.000      1.4265552e-015
```




```
Group Parameter Estimates:

Group  Label    P#    mu    s.e.   P#       s2    s.e.    sd    s.e.
 1     Group1        0.00   ----           1.00  ----    1.00   ---
```

Next, I will run the 2PL model as a 2-category graded response model. As before, I comment only parts of the syntax that differ from the previous.

Example 2b: 2PL as a 2-category graded response model (**ex2Plog.flexmirt**):

<Project>

Title = "Simulated Data";
Description = "2PL as Graded Response";

<Options>

Mode = Calibration;
Quadrature = 17, 4.0;
Processors=4;

//Notice I dropped the **NormalMetric3PL = Yes** command for this graded response model.

<Groups>

This is a supplementary resource to Measurement Theory and Applications for the Social Sciences, by Deborah L. Bandalos.
Copyright © 2018 by The Guilford Press.

%Group1%
File = "IRTexample.dat";
Varnames = ID, I1-I20;
Select = I1-I20;
N = 5000;
NCATS(I1-I20)=2;
Model(I1-I20)=***Graded***;

//Note this change from the ThreePL. **Graded** requests the graded response model.

EmpHist=Yes;

<Constraints>

// This line is still required even though there are no constraints in this model.

Output for this model is saved in file "ex2Plog-irt.txt."

Item	Label	P#	a	s.e.	P#	c	s.e.	b	s.e.
1	I1	2	1.14	0.08	1	2.09	0.06	-1.84	0.10
2	I2	4	0.78	0.04	3	-0.85	0.03	1.09	0.07
3	I3	6	0.77	0.04	5	-0.03	0.03	0.04	0.04
4	I4	8	0.53	0.04	7	-0.87	0.03	1.64	0.12
5	I5	10	1.24	0.08	9	1.89	0.06	-1.52	0.07
6	I6	12	1.29	0.06	11	0.56	0.04	-0.43	0.03
7	I7	14	1.72	0.08	13	0.24	0.04	-0.14	0.02
8	I8	16	3.25	0.18	15	3.76	0.15	-1.16	0.03
9	I9	18	1.63	0.07	17	0.40	0.04	-0.24	0.02
10	I10	20	1.20	0.05	19	-1.64	0.05	1.37	0.07
11	I11	22	1.15	0.05	21	-0.42	0.04	0.37	0.04
12	I12	24	2.01	0.09	23	0.61	0.05	-0.30	0.02
13	I13	26	4.02	0.23	25	4.25	0.19	-1.06	0.02
14	I14	28	0.89	0.04	27	-0.75	0.04	0.84	0.06
15	I15	30	2.91	0.16	29	3.30	0.13	-1.13	0.03
16	I16	32	2.48	0.12	31	2.25	0.09	-0.91	0.02
17	I17	34	2.25	0.10	33	1.85	0.07	-0.82	0.02
18	I18	36	1.37	0.07	35	1.60	0.05	-1.17	0.05
19	I19	38	1.37	0.09	37	2.30	0.07	-1.68	0.08
20	I20	40	0.69	0.05	39	0.25	0.03	-0.36	0.05

Notice that the *a*-parameters are 1.7 times what they were when the 2PL model was run using "**NormalMetric3PL**". Flexmirt uses the logistic metric instead of the normal metric by default. The logistic metric uses D = 1 instead of D = 1.7 (as in Equation 14.3 in the text), so the *a*-parameters are multiplied by 1.7 to compensate. There is also a new parameter, *c* (unrelated to the lower-asymptote in the 3PL model), which equals –*ab*.

This is a supplementary resource to Measurement Theory and Applications for the Social Sciences, by Deborah L. Bandalos.
Copyright © 2018 by The Guilford Press.

1PL and Rasch Models

As with the 2PL model, there are two ways to run a 1PL model: a) by applying constraints to a 3PL model or b) as a dichotomous graded response model. Additionally, one can constrain the discrimination parameters to equal 1 and free the ability variance, yielding the Rasch parameterization. First, I will illustrate a constrained 3PL model.

Example 3a: 1PL using Constraints on 3PL (**exaPL.flexmirt**):

<Project>
Title = "Simulated Data";
Description = "1PL as constrained 3PL";
<Options>
Mode = Calibration;
Quadrature = 17, 4.0;
Processors=4;
NormalMetric3PL = Yes;

// For this model, I use the normal metric, to be consistent with Examples 1 and 2a;

<Groups>
%Group1%
File = "IRTexample.dat";
Varnames = ID, I1-I20;
Select = I1-I20;
N = 5000;
NCATS(I1-I20)=2;

// There are 2 categories, 0 or 1, because I have scored the data;

Model(I1-I20)=***ThreePL***;

//I am running a 3PL model with constraints that yield a 1PL model;

EmpHist=Yes;

//This command requests estimation of the shape of the theta distribution, instead of assuming normal;

<Constraints>
Equal (I1-I20), **Slope**;

// This command sets all slopes equal. This is the only line that differs from the 2PL in Example 2a (*ex2PL.flexmirt*);

Fix (I1-I20), **Guessing**;

This is a supplementary resource to Measurement Theory and Applications for the Social Sciences, by Deborah L. Bandalos.
Copyright © 2018 by The Guilford Press.

// This command fixes the lower asymptotes —the next line indicates what value to fix them to;

Value (I1-I20), **Guessing,** -999; //sets the logit of the lower asymptote to -999, which is effectively 0;

Output in ex1PL-irt.txt:

Item	Label	P#	a	s.e.	P#	b	s.e.	P#	g	s.e.
1	I1	21	0.66	0.01	1	-1.89	0.05		0.00	----
2	I2	21	0.66	0.01	2	0.84	0.04		0.00	----
3	I3	21	0.66	0.01	3	0.03	0.03		0.00	----
4	I4	21	0.66	0.01	4	0.89	0.04		0.00	----
5	I5	21	0.66	0.01	5	-1.66	0.04		0.00	----
6	I6	21	0.66	0.01	6	-0.43	0.03		0.00	----
7	I7	21	0.66	0.01	7	-0.06	0.03		0.00	----
8	I8	21	0.66	0.01	8	-2.12	0.05		0.00	----
9	I9	21	0.66	0.01	9	-0.21	0.03		0.00	----
10	I10	21	0.66	0.01	10	1.48	0.04		0.00	----
11	I11	21	0.66	0.01	11	0.43	0.03		0.00	----
12	I12	21	0.66	0.01	12	-0.29	0.03		0.00	----
13	I13	21	0.66	0.01	13	-2.04	0.05		0.00	----
14	I14	21	0.66	0.01	14	0.74	0.04		0.00	----
15	I15	21	0.66	0.01	15	-1.97	0.05		0.00	----
16	I16	21	0.66	0.01	16	-1.39	0.04		0.00	----
17	I17	21	0.66	0.01	17	-1.17	0.04		0.00	----
18	I18	21	0.66	0.01	18	-1.34	0.04		0.00	----
19	I19	21	0.66	0.01	19	-1.98	0.05		0.00	----
20	I20	21	0.66	0.01	20	-0.26	0.03		0.00	----

Notice the *a*-parameters are equal across items and the *g* (or *c*) parameters are fixed to 0. Only the *b*-parameter is estimated.

If the "**NormalMetric3PL = Yes**" were omitted and the "**Model(I1-I20)=ThreePL** " were changed to "**Model(I1-I20)=Graded**" (as in *ex2Plog.flexmirt*), the model would be on the logistic metric (D = 1). The *b*-parameters would stay the same, but the *a*-parameters would equal 1.12 (1.7 x 0.66 =1.12) to compensate for changing D to 1.

Rasch Parameterization

Now I will show the more traditional Rasch parameterization (although in Flexmirt one cannot set the mean item difficulty to zero, which is relatively common in Rasch scaling).

Example 3b: Rasch Parameterization of 1PL (**exRasch.flexmirt**):

<Project>
Title = "Simulated Data";
Description = "Rasch scaling";
<Options>

This is a supplementary resource to Measurement Theory and Applications for the Social Sciences, by Deborah L. Bandalos.
Copyright © 2018 by The Guilford Press.

Mode = Calibration;
Quadrature = 17, 4.0;
Processors=4;
<Groups>
%Group1%
File = "IRTexample.dat";
Varnames = ID, I1-I20;
Select = I1-I20;
N = 5000;
NCATS(I1-I20)=2;
Model(I1-I20)=***Graded***;

//for the 2PL or 1PL, use ***Graded*** because these models are special cases of the graded response model, unless you want to use the 1.7 in the model, in which case you should run a constrained 3PL;

EmpHist=Yes;
<Constraints>
Fix (I1-I20), **Slope**;

//This command fixes slopes at the values given in the next line;

Value (I1-I20), **Slope**, 1.0;
Free Cov(1,1);
//This frees the variance of theta, which is the only element in the covariance matrix ;

Output in exRasch-irt.txt:

Item	Label	P#	a	s.e.	P#	c	s.e.	b	s.e.
1	I1		1.00	----	1	2.12	0.05	-2.12	0.05
2	I2		1.00	----	2	-0.95	0.04	0.95	0.04
3	I3		1.00	----	3	-0.03	0.04	0.03	0.04
4	I4		1.00	----	4	-0.99	0.04	0.99	0.04
5	I5		1.00	----	5	1.86	0.04	-1.86	0.04
6	I6		1.00	----	6	0.48	0.04	-0.48	0.04
7	I7		1.00	----	7	0.07	0.04	-0.07	0.04
8	I8		1.00	----	8	2.37	0.05	-2.37	0.05
9	I9		1.00	----	9	0.23	0.04	-0.23	0.04
10	I10		1.00	----	10	-1.66	0.05	1.66	0.05
11	I11		1.00	----	11	-0.48	0.04	0.48	0.04
12	I12		1.00	----	12	0.32	0.04	-0.32	0.04
13	I13		1.00	----	13	2.28	0.05	-2.28	0.05
14	I14		1.00	----	14	-0.83	0.04	0.83	0.04
15	I15		1.00	----	15	2.20	0.05	-2.20	0.05
16	I16		1.00	----	16	1.55	0.04	-1.55	0.04
17	I17		1.00	----	17	1.31	0.04	-1.31	0.04
18	I18		1.00	----	18	1.50	0.04	-1.50	0.04
19	I19		1.00	----	19	2.21	0.05	-2.21	0.05

This is a supplementary resource to Measurement Theory and Applications for the Social Sciences, by Deborah L. Bandalos.
Copyright © 2018 by The Guilford Press.

20	I20	1.00	----	20	0.29	0.04	-0.29	0.04

Notice that the *a*-parameters are equal to 1 and they have no standard errors because the values were not estimated. The *b*-parameters = 1.7 x 0.66 x *b* parameters from the 1PL scaling (the 0.66 was the *a*-parameter from the 1PL scaling). The *c*-parameter is equal to -(*a*b).*

Because the *a*-parameters were fixed, the variance of ability was a free parameter. The estimated standard deviation was 1.12, because 1.7 x 0.66 =1.12.

Group	Label	P#	mu	s.e.	P#	s2	s.e.	sd	s.e.
1	Group1		0.00	----	21	1.26	0.03	1.12	0.01

Graded Response Model

Items 21-25, omitted from the preceding analyses, had 3 categories. I will scale these items with the graded response model and scale Items 1-20 with the 3PL. Flexmirt will only use the logistic metric (D = 1) for the graded response model, although Samejima (1969) developed the model in the normal metric. If you want to report the parameters in the normal metric, divide the *a*-parameters by 1.7. For the sake of consistency, I dropped the "**NormalMetric3PL = Yes**;" for Items 1-20 so they would also be on the logistic metric.

Example 4: Graded Response Model (exGraded.flexmirt):

<Project>
Title = "Simulated Data";
Description = "3PL";
<Options>
Mode = Calibration;
Processors=4;
Quadrature = 17, 4.0;
<Groups>
%Group1%
File = "IRTexample.dat";
Varnames = ID, I1-I25;
Select = I1-I25;
N = 5000;
NCATS(I1-I20)=2;

// This command specifies 2 categories, 0 or 1, because I have scored the data;

NCATS(I21-I25)=3;

//This command specifies that items 21-25 have 3 categories, with the first category starting at 0;
Model(I1-I20)=***ThreePL***;
Model(I21-I25)=***Graded(3)***; // specifies 3 ordered categories;

This is a supplementary resource to Measurement Theory and Applications for the Social Sciences, by Deborah L. Bandalos.
Copyright © 2018 by The Guilford Press.

EmpHist=Yes;
<Constraints>
Prior (I1-I25), **Slope**: Normal(1.5,0.5);

//The prior command specifies the mean and standard deviation for the ability distribution in logistic metric. The mean of 1.5 is equal to .88 (1.5/1.7) in the normal metric;

Prior (I1-I20), Guessing: Beta(21,81);

//This will yield a mean of .20 and sd of 0.04.

Output from exGraded-irt.txt:

For Items 1-20, notice that the *a*-parameters are approximately 1.7 times what they were in the 3PL example, but the *b* and *c*-parameters are approximately the same (some small changes due to the presence of other items in the examinees' posterior likelihoods used in the MML estimation).

Item	Label	P#	a	s.e.	P#	c	s.e.	b	s.e.	P#	logit-g	s.e.	g	s.e.
1	I1	3	0.90	0.06	2	1.73	0.08	-1.91	0.14	1	-1.39	0.25	0.20	0.04
2	I2	6	1.91	0.18	5	-2.68	0.26	1.40	0.05	4	-1.46	0.08	0.19	0.01
3	I3	9	1.13	0.10	8	-0.81	0.14	0.71	0.08	7	-1.30	0.15	0.21	0.03
4	I4	12	1.40	0.17	11	-2.51	0.28	1.80	0.08	10	-1.39	0.09	0.20	0.01
5	I5	15	1.00	0.06	14	1.48	0.08	-1.48	0.13	13	-1.35	0.24	0.21	0.04
6	I6	18	1.43	0.09	17	-0.06	0.10	0.05	0.07	16	-1.41	0.17	0.20	0.03
7	I7	21	2.26	0.14	20	-0.72	0.11	0.32	0.04	19	-1.50	0.11	0.18	0.02
8	I8	24	2.08	0.12	23	2.81	0.11	-1.35	0.07	22	-1.39	0.23	0.20	0.04
9	I9	27	1.91	0.12	26	-0.41	0.10	0.21	0.05	25	-1.48	0.13	0.19	0.02
10	I10	30	3.62	0.31	29	-4.69	0.39	1.29	0.03	28	-2.31	0.07	0.09	0.01
11	I11	33	2.10	0.15	32	-1.77	0.17	0.85	0.04	31	-1.46	0.09	0.19	0.01
12	I12	36	2.44	0.14	35	-0.31	0.10	0.12	0.04	34	-1.47	0.11	0.19	0.02
13	I13	39	2.57	0.16	38	3.04	0.13	-1.19	0.06	37	-1.33	0.21	0.21	0.03
14	I14	42	2.21	0.19	41	-2.71	0.26	1.23	0.04	40	-1.41	0.07	0.20	0.01
15	I15	45	2.00	0.12	44	2.51	0.10	-1.25	0.07	43	-1.38	0.23	0.20	0.04
16	I16	48	2.06	0.13	47	1.51	0.09	-0.73	0.06	46	-1.19	0.17	0.23	0.03
17	I17	51	1.84	0.11	50	1.22	0.08	-0.66	0.06	49	-1.51	0.20	0.18	0.03
18	I18	54	1.15	0.07	53	1.15	0.08	-1.00	0.10	52	-1.39	0.23	0.20	0.04
19	I19	57	1.08	0.06	56	1.89	0.09	-1.74	0.12	55	-1.29	0.24	0.22	0.04
20	I20	60	0.82	0.08	59	-0.38	0.14	0.46	0.14	58	-1.24	0.21	0.22	0.04

This is a supplementary resource to Measurement Theory and Applications for the Social Sciences, by Deborah L. Bandalos.
Copyright © 2018 by The Guilford Press.

For Items 21-25, two sets of parameters are displayed.

The ***second set*** is the standard parameterization, Equation 14.5 in the text (with D = 1). As an example, for Item 22, $b_1 = -1.59$ and $b_2 = -0.76$. Thus, 50% of examinees at $\theta = -1.59$ are predicted to score 1 or higher, and 50% of examinees at $\theta = -0.76$ are predicted to score 2.

In the ***first set*** of parameters, the *c* parameters are the negative of the product of the *a* and *b* values in the second set of parameters ($c_1 = -ab_1$ and $c_2 = -ab_2$). For Item 22, for examinees at $\theta = 0$ the log-odds (logits) of scoring 1 or higher $= -((1.73)(-1.59)) = 2.75$ and the log-odds of scoring 2 $= -((1.73)(-.76)) = 1.32$.

Recall that the *c* parameter is labeled *g* in Flexmirt, so do not confuse it with this other *c* parameter.

Graded Items for Group 1: Group1

Item	Label	P#	a	s.e.	P#	c 1	s.e.	P#	c 2	s.e.
21	I21	63	0.99	0.05	61	2.82	0.06	62	1.49	0.04
22	I22	66	1.73	0.06	64	2.75	0.07	65	1.32	0.05
23	I23	69	1.34	0.04	67	1.37	0.04	68	-0.04	0.04
24	I24	72	1.09	0.04	70	0.40	0.04	71	-0.97	0.04
25	I25	75	1.50	0.05	73	0.10	0.04	74	-1.23	0.04

Graded Items for Group 1: Group1

Item	Label	P#	a	s.e.	b 1	s.e.	b 2	s.e.
21	I21	63	0.99	0.05	-2.86	0.12	-1.51	0.07
22	I22	66	1.73	0.06	-1.59	0.05	-0.76	0.03
23	I23	69	1.34	0.04	-1.03	0.04	0.03	0.03
24	I24	72	1.09	0.04	-0.36	0.03	0.90	0.04
25	I25	75	1.50	0.05	-0.07	0.03	0.82	0.03

This is a supplementary resource to Measurement Theory and Applications for the Social Sciences, by Deborah L. Bandalos.
Copyright © 2018 by The Guilford Press.

Partial Credit Models

In this example, I will scale Items 21-25 with the partial credit model, along with scaling Items 1-20 with the Rasch model. I will constrain the *a*-parameters to 1 and free the variance of ability, but alternatively one could constrain the *a*-parameters to equality and fix the variance to 1 (the latter specification is the default parameterization in Flexmirt).

Example 5: Partial Credit and Rasch Model (exPC.flexmirt):

```
<Project>
Title = "Simulated Data";
Description = "Partial Credit";
<Options>
  Mode = Calibration;
  Processors=4;
 Quadrature = 17, 4.0;
 <Groups>
%Group1%
 File = "IRTexample.dat";
 Varnames = ID, I1-I25;
 Select = I1-I25;
 N = 5000;
 NCATS(I1-I20)=2;
 NCATS(I21-I25)=3; //3 categories, starting at 0;
 Model(I1-I20)=Graded; //constrain to Rasch later;
 Model(I21-I25)=GPC(3); //3 categories, specify GPC and then constrain to PC later;
 EmpHist=Yes;
<Constraints>
Fix (I1-I25), Slope; //fixes slopes at the values given in the next line
Value (I1-I25), Slope, 1.0;
Free  Cov(1,1); //the only element in the covariance matrix is the variance of theta
```

This is a supplementary resource to Measurement Theory and Applications for the Social Sciences, by Deborah L. Bandalos.
Copyright © 2018 by The Guilford Press.

Output from exPC-irt.txt:

Item	Label	P#	a	s.e.	P#	c	s.e.	b	s.e.
1	I1		1.00	----	1	2.09	0.05	-2.09	0.05
2	I2		1.00	----	2	-0.93	0.04	0.93	0.04
3	I3		1.00	----	3	-0.04	0.04	0.04	0.04
4	I4		1.00	----	4	-0.98	0.04	0.98	0.04
5	I5		1.00	----	5	1.83	0.04	-1.83	0.04
6	I6		1.00	----	6	0.46	0.04	-0.46	0.04
7	I7		1.00	----	7	0.06	0.04	-0.06	0.04
8	I8		1.00	----	8	2.34	0.05	-2.34	0.05
9	I9		1.00	----	9	0.22	0.04	-0.22	0.04
10	I10		1.00	----	10	-1.63	0.05	1.63	0.05
11	I11		1.00	----	11	-0.48	0.04	0.48	0.04
12	I12		1.00	----	12	0.31	0.04	-0.31	0.04
13	I13		1.00	----	13	2.25	0.05	-2.25	0.05
14	I14		1.00	----	14	-0.82	0.04	0.82	0.04
15	I15		1.00	----	15	2.17	0.05	-2.17	0.05
16	I16		1.00	----	16	1.53	0.04	-1.53	0.04
17	I17		1.00	----	17	1.29	0.04	-1.29	0.04
18	I18		1.00	----	18	1.47	0.04	-1.47	0.04
19	I19		1.00	----	19	2.18	0.05	-2.18	0.05
20	I20		1.00	----	20	0.28	0.04	-0.28	0.04

GPC Items for Group 1: Group1

Item	Label	P#	a	s.e.	b	s.e.	d1	d2	s.e.	d3	s.e.
21	I21		1.00		-1.68	0.03	0	-0.24	0.05	0.24	0.05
22	I22		1.00		-1.28	0.03	0	-0.30	0.04	0.30	0.04
23	I23		1.00		-0.52	0.03	0	-0.06	0.03	0.06	0.03
24	I24		1.00		0.18	0.03	0	0.04	0.03	-0.04	0.03
25	I25		1.00		0.34	0.03	0	-0.18	0.04	0.18	0.04

For Items 21-25, the model is parameterized a bit differently than in the text. In the text, Equation 14.6, the partial credit model was parameterized as

$$P(x_{ij} = k \mid \theta_j, a_i, b_{ik}) = \frac{e^{\sum_{x=0}^{k}(\theta - b_{ix})}}{\sum_{j=0}^{m_i} e^{\sum_{x=0}^{j}(\theta - b_{ix})}}$$

. In Flexmirt, b_{ik} is instead $b_i + d_{i,k+1}$. So for Item 22, for example, the intersection between score 0 and score 1 occurs at $\theta = -1.28 - 0.30 = -1.58$. The intersection between score 1 and score 2 occurs at $\theta = -1.28 + 0.30 = -0.98$.

Later in the output, the estimates of the ability variance/standard deviation are printed.

Group	Label	P#	mu	s.e.	P#	s2	s.e.	sd	s.e.
1	roup1		0.00	----	31	1.14	0.03	1.07	0.01

With the addition of Items 21-25, the estimated standard deviation of the ability distribution is 1.07 instead of the estimate of 1.12 with only Items 1-20. This is why the absolute values of the estimates of the *b*-parameters for Items 1-20 in this calibration are slightly smaller

This is a supplementary resource to Measurement Theory and Applications for the Social Sciences, by Deborah L. Bandalos.
Copyright © 2018 by The Guilford Press.

than what they were when estimated without the polytomous items—the scaling has changed. When the items do not all have the same slope, constraining the slopes to equality yields a metric that is the best compromise given the data in the analysis.

This is a supplementary resource to Measurement Theory and Applications for the Social Sciences, by Deborah L. Bandalos.
Copyright © 2018 by The Guilford Press.

Generalized Partial Credit Models

In this example, I will scale Items 21-25 with the generalized partial credit model, along with scaling Items 1-20 with the 3PL model. In Flexmirt, the GPC is always in the logistic metric (D = 1), so for items 1-20 I will drop the "NormalMetric3PL = Yes;" to put these items in the logistic metric also.

Example 6: Generalized Partial Credit and 3PL Model (exGPC.flexmirt):

```
<Project>
Title = "Simulated Data";
Description = "Generalized Partial Credit";
<Options>
  Mode = Calibration;
  Processors=4;
 Quadrature = 17, 4.0;
 <Groups>
%Group1%
 File = "IRTexample.dat";
 Varnames = ID, I1-I25;
 Select = I1-I25;
 N = 5000;
 NCATS(I1-I20)=2;
 NCATS(I21-I25)=3; //3 categories, starting at 0;
 Model(I1-I20)=ThreePL;
 Model(I21-I25)=GPC(3); //3 ordered categories;
 EmpHist=Yes;
<Constraints>
Prior (I1-I25), Slope: Normal(1.5,0.5);
Prior (I1-I20), Guessing: Beta(21,81); //mean of .20
Prior (I10), Intercept: Normal(0,4);
```

Notice I added a prior to the intercept (−*ab*) for Item 10. I did this because with no prior, the estimated difficulty was unreasonably high. The second parameter for the prior is the standard deviation, **not** the variance, so this is not a very informative prior.

This is a supplementary resource to Measurement Theory and Applications for the Social Sciences, by Deborah L. Bandalos.
Copyright © 2018 by The Guilford Press.

Output from exGPC-irt.txt:

For Items 1-20, notice that the *a*-parameters are approximately 1.7 times what they were in the 3PL example, but the *b* and *c*-parameters are approximately the same (there are some small changes due to the presence of the additional items (I21-25) in the examinees' posterior likelihoods used in the MML estimation).

Item	Label	P#	a	s.e.	P#	c	s.e.	b	s.e.	P#	logit-g	s.e.	g	s.e.
1	I1	3	0.91	0.06	2	1.73	0.08	-1.90	0.14	1	-1.40	0.25	0.20	0.04
2	I2	6	1.84	0.17	5	-2.59	0.25	1.40	0.05	4	-1.47	0.08	0.19	0.01
3	I3	9	1.11	0.10	8	-0.77	0.14	0.70	0.09	7	-1.33	0.16	0.21	0.03
4	I4	12	1.33	0.16	11	-2.43	0.27	1.82	0.09	10	-1.41	0.09	0.20	0.01
5	I5	15	1.01	0.06	14	1.48	0.08	-1.47	0.12	13	-1.36	0.24	0.21	0.04
6	I6	18	1.42	0.09	17	-0.04	0.10	0.03	0.07	16	-1.43	0.17	0.19	0.03
7	I7	21	2.25	0.14	20	-0.69	0.11	0.30	0.04	19	-1.52	0.11	0.18	0.02
8	I8	24	2.11	0.12	23	2.83	0.11	-1.34	0.07	22	-1.40	0.23	0.20	0.04
9	I9	27	1.90	0.12	26	-0.38	0.10	0.20	0.05	25	-1.50	0.13	0.18	0.02
10	I10	30	3.56	0.30	29	-4.60	0.38	1.29	0.03	28	-2.31	0.07	0.09	0.01
11	I11	33	2.07	0.15	32	-1.74	0.17	0.84	0.04	31	-1.46	0.09	0.19	0.01
12	I12	36	2.44	0.14	35	-0.28	0.10	0.11	0.04	34	-1.48	0.11	0.19	0.02
13	I13	39	2.61	0.16	38	3.08	0.13	-1.18	0.06	37	-1.36	0.21	0.20	0.03
14	I14	42	2.15	0.19	41	-2.64	0.25	1.23	0.05	40	-1.42	0.07	0.19	0.01
15	I15	45	2.04	0.12	44	2.54	0.10	-1.25	0.07	43	-1.39	0.23	0.20	0.04
16	I16	48	2.06	0.13	47	1.53	0.09	-0.74	0.06	46	-1.22	0.18	0.23	0.03
17	I17	51	1.86	0.11	50	1.24	0.08	-0.67	0.06	49	-1.52	0.21	0.18	0.03
18	I18	54	1.16	0.07	53	1.16	0.08	-1.00	0.10	52	-1.41	0.24	0.20	0.04
19	I19	57	1.10	0.07	56	1.90	0.09	-1.73	0.12	55	-1.31	0.24	0.21	0.04
20	I20	60	0.83	0.08	59	-0.38	0.14	0.46	0.14	58	-1.22	0.21	0.23	0.04

GPC Items for Group 1: Group1

Item	Label	P#	a	s.e.	b	s.e.	d 1	d 2	s.e.	d 3	s.e.
21	I21	61	0.70	0.03	-2.04	0.08	0	-0.53	0.08	0.53	0.08
22	I22	64	1.23	0.05	-1.15	0.03	0	-0.14	0.04	0.14	0.04
23	I23	67	0.91	0.03	-0.50	0.03	0	-0.11	0.04	0.11	0.04
24	I24	70	0.74	0.03	0.26	0.03	0	-0.14	0.05	0.14	0.05
25	I25	73	1.02	0.04	0.37	0.03	0	-0.19	0.04	0.19	0.04

As described for the partial credit model, b_{ik} is recast as $b_{\mathrm{i}} + d_{\mathrm{i,k+1}}$. So for Item 22, for example, the intersection between score 0 and score 1 occurs at $\theta = -1.15 - 0.14 = -1.29$. The intersection between score 1 and score 2 occurs at $\theta = -1.15 + 0.14 = -1.01$.

This is a supplementary resource to Measurement Theory and Applications for the Social Sciences, by Deborah L. Bandalos.
Copyright © 2018 by The Guilford Press.

References

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplements, 17*.

Suiero, M. J., & Abad, F. J. (2009). Assessing goodness of fit in item response theory with nonparametric models: A comparison of posterior probabilities and kernel-smoothing approaches. *Educational and Psychological Methods, 71* (5), 834-848. doi: 10.1177/0013164408322026

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). Mahwah, NJ: LEA.

Woods, C. M. (2008). Consequences of ignoring guessing when estimating the latent density in item response theory. *Applied Psychological Methods, 32* (5), 371-384. doi: 10.1177/0146621607307691