FlexCCT v2.0: Help File

Contents

Copyright

For any issues, please contact Stephen L. France at france@business.msstate.edu.

Introduction

FlexCCT (Flexible Cultural Consensus Theory) provides a flexible, easy to use toolkit for analyzing ratings data. FlexCCT utilizes techniques from the academic discipline of "Cultural Consensus Theory" (CCT). CCT can be thought of as a data pooling or aggregation technique that uses the "wisdom of the crowd" to both create better ratings and to characterize raters with respect to competency and bias. This document will give an overview of the features present in the FlexCCT package, along with an illustrated example.

For technical details of the methodology please see the following papers:

France, S.L., Batchelder, W. H. (2015). A Maximum Likelihood Item Easiness Model for Test Theory Without An Answer Key, Educational and Psychological Measurement, 75(1), 57-77.
<preprint> <reprint>

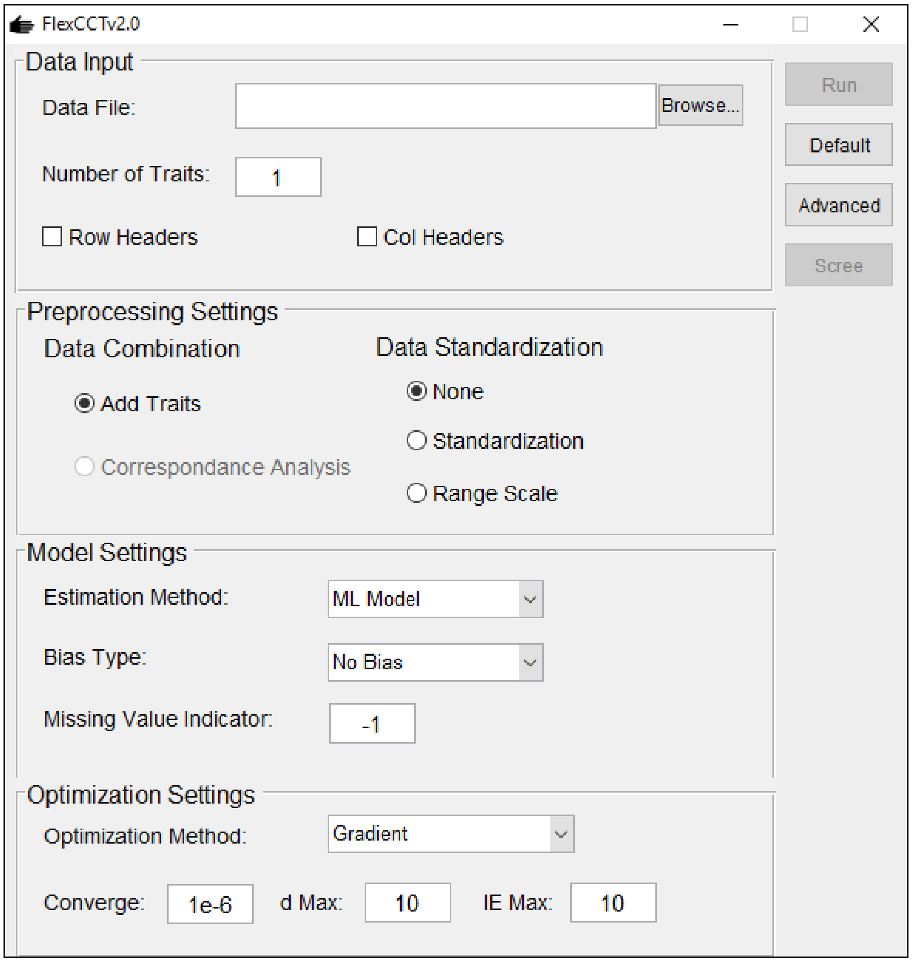
France, S.L., Batchelder, W. H. (2014). Unsupervised Consensus Analysis for On-line Review and Questionnaire Data, Information Sciences, 281, 241-257. (PN016)
<preprint> <reprint>

FlexCCT also implements minimum residual factor analysis, which is described in the following paper:

Comrey, A. L. (1962). The minimum residual method of factor analysis. *Psychological Reports, 11*(1), 15-18.

Main Application

Double clicking on the FlexCCT.exe icon will load the main FlexCCT application. This will bring up the main FlexCCT form, which is shown below.



The form consists of the following sections.

Data Input

This frame defines the input file details. The input file is a csv file with subjects as rows and the items to be rated as columns. The data file can optionally contain row names and column names. Each entry to the file is a rating. FlexCCT allows for missing rating values, which default to -1. Each item can have ratings for multiple traits. For example, an essay grading item could have attributes for grammar, consistency, exposition, and content. Consider a situation with 5 users, 6 items and 2 traits per item. Let R(i,j,a) signify the rating for the i^{th} user rating the j^{th} item on the a^{th} trait. The file configuration is given below.

R(1,1,1),R(1,1,2), R(1,2,1),R(1,2,2),R(1,3,1),R(1,3,2), R(1,4,1),R(1,4,2), R(1,5,1),R(1,5,2), R(1,6,1),R(1,6,2)

R(2,1,1),R(2,1,2), R(2,2,1),R(2,2,2),R(2,3,1),R(2,3,2), R(2,4,1),R(2,4,2), R(2,5,1),R(2,5,2), R(2,6,1),R(2,6,2)

R(3,1,1),R(3,1,2), R(3,2,1),R(3,2,2),R(3,3,1),R(3,3,2), R(3,4,1),R(3,4,2), R(3,5,1),R(3,5,2), R(3,6,1),R(3,6,2)

R(4,1,1),R(4,1,2), R(4,2,1),R(4,2,2),R(4,3,1),R(4,3,2), R(4,4,1),R(4,4,2), R(4,5,1),R(4,5,2), R(4,6,1),R(4,6,2)

R(5,1,1),R(5,1,2), R(5,2,1),R(5,2,2),R(5,3,1),R(5,3,2), R(5,4,1),R(5,4,2), R(5,5,1),R(5,5,2), R(5,6,1),R(5,6,2)

Column/row names are given by adding a column/row before the first ratings. For example, the following graphic gives a file with both raters and items. The raters are analysts rating NFL Teams' drafts. The items are the actual NFL teams.

```
,Arizona,Atlanta,Baltimore,Buffalo,Carolina,Chicago,Cincinnati,Cleveland,Dallas,Denver,Detroit,Green Bay,Hc
ESPN's MEL KIPER,8,6,9,8,8,7,11,8,7,7,9,9,9,11,6,7,6,9,8,5,7,6,5,12,9,10,7,5,8,12,7,7
CBS SPORTS,7,6,10,9,10,6,13,5,10,7,7,9,9,12,8,6,10,10,8,1,9,5,3,10,12,8,10,5,7,9,8,8
FOX SPORTS,10,6,9,9,9,5,12,6,10,8,8,9,8,11,6,10,6,8,12,3,9,6,3,9,13,9,7,9,12,10,6,11
NBC SPORTS,9,5,10,9,6,10,12,9,9,6,10,11,9,13,3,8,9,7,10,3,12,8,3,10,12,10,7,8,10,12,7,11
```

Data File
Click the Browse button and choose the .csv data file using the standard Microsoft Windows file dialog box.

Number of Traits
Define the number of traits used in the file. This defaults to 1. The number of columns must be divisible by the number of traits. For example, 60 columns and 6 traits would indicate a dataset that has 10 items with 6 traits per item

Row Headers
The FlexCCT import routine will attempt to guess whether or not row headers are included in the data file, based upon the data type of the first column. However users can explicitly state that row headers are included. This feature is useful if the row headers are numeric.

Column Headers
The FlexCCT import routine will attempt to guess whether or not column headers are included in the data file, based upon the data type of the first row. However users can explicitly state that column headers are included. This feature is useful if the column headers are numeric.

Preprocessing
The preprocessing frame allows users to specify how the raw imported data is processed. There are two preprocessing options.

Data Combination
The CCT methods implemented in this package assume a single trait. If the data have multiple traits then there are two options for combining the traits into a single meta-trait:

- **Add Traits:** This option utilizes the assumptions of classical test theory. Add the values of the traits together. For example, if an essay scored for 3 out of 6 for grammar, 4 out of 6 for consistency, 3 out of 6 for exposition, and 4 out of 6 content then the overall score for the essay would be 14/24.
- **Correspondence Analysis:** The correspondence analysis option assumes that the individual traits are measured on an ordinal scale. The method of multiple correspondence analysis is then used to create a continuous scale measure that captures as much structure as possible from the ordinal variables.

Data Transformation
CCT assumes that all items are measured on the same scale. If this is not true then this will lead to a violation of the CCT error assumptions and a possibly incorrect analysis. One way of getting around this problem is to rescale the data. The options are as follows:

- **None:** There is no transformation. The raw scores are used. This option should be used if all items are measured on the same scale and no error heterogeneity is suspected.
- **Standardization:** The scores are standardized by subtracting the mean and dividing by standard deviations. This is an option when items are measured on multiple scales.

- **Range Scale:** The scores are range scaled by putting all scores into the range [0,1]. This is achieved by subtracting the minimum score from each score and then dividing by the range (maximum – minimum). This is an option when items are measured on multiple scales.
- **Log Transformation:** The log transformation is utilized when items are measured on the same scale, but there is possible error heterogeneity. An example of this is a situation where raters are estimating distances. All items are measured on the same scale, but based on the Weber-Fechner psychophysical law, large distances are liable to have larger error variances than do smaller differences. This can be rectified by using a log transformation on the data.

Model Settings

The model settings frame determines the CCT model settings. FlexCCT implements a range of models and model features.

Base Model

The Base Model dialog determines the basic CCT model used. Models range from simple aggregation models to complex likelihood models incorporating item difficulty. The options are as follows:

- **Simple Average:** A baseline simple averaging model. It calculates the mean of all of the ratings. Assumes even competencies.
- **Factor Analysis:** Implements a principal axes factor analysis, as per Comrey (1966).
- **ML Model**: The basic maximum likelihood model. Outputs a set of user competencies **d**, where $\sigma_i^2(\varepsilon_{ik}) = 1/d_i$ and a competency weighted item answer key **z**.
- **IE Multiply**: The multiplicative item easiness model. Assumes that for rater *i* and item *k*, the item competencies are affected by multiplicative scalar item easiness values, so that the error variance $\sigma_{ik}^2(\varepsilon_{ik}) = 1/(d_i\beta_k)$.
- **IE Add**: The additive item easiness model. Assumes that for rater *i* and item *k*, the item competencies $\sigma_{ik}^2(\varepsilon_{ik}) = 1/(d_i + \beta_k)$.

Bias Type

Biases can be defined for each model. Both additive and multiplicative biases can be added. The biases are defined, so that for user *i* rating item *k*, the item scores are defined as $x_{ik} = b_{Mi}z_k + b_{Ai}$. Here b_{Mi} is the multiplicative bias for item *i* and b_{Ai} is an additive bias for item *i*. The options are as follows:

- **No Bias:** No biases are included in the model. Thus, for each *i*, $b_{Mi} = 1$ and $b_{Ai} = 0$.
- **Additive Bias:** Only additive biases are included in the model. Thus for each *i*, $b_{Mi} = 1$.
- **Multiplicative Bias:** Only multiplicative biases are included in the model. Thus for each *i*, $b_{Ai} = 0$.
- **Add & Mult Bias:** Both additive and multiplicative biases are included in the model.

All four bias types can be utilized with every model except for the principal axes factor analysis model, which does not allow bias to be modeled.

Missing Value Indicator

The missing value indicator allows the specification of the value used for missing data. The default is "-1". If the ratings values can include negative numbers then this number should be changed to a number outside of the range of the ratings.

Optimization Settings

The optimization settings frame allows the specification of the optimization procedure for the implemented FlexCCT model.

Optimization Method

This dropdown list specifies the method by which the model is optimized (i.e., the likelihood function is optimized). **The default method is the gradient method and this method is to be recommended for non-technical users**.

Each option has advantages and disadvantages. A technical explanation is given as follows:

- **Gradient:** This procedure utilizes first order derivatives of the model likelihood function to maximize the likelihood using a gradient descent procedure. All models specified in the Model Setting frame and all advanced settings can be utilized with this procedure.
- **Derivative Free:** This procedure utilizes a gradient descent procedure to maximize the likelihood function, but does not utilize function derivatives. All models specified in the Model Setting frame and all advanced settings can be utilized with this procedure. This method is significantly slower than the gradient descent procedure.
- **Fixed Point:** This procedure alternatively optimizes the first order conditions (i.e., the derivatives set to 0) for the different variable sets ($\mathbf{d}$, $\mathbf{z}$, $\mathbf{b}_A$, $\mathbf{b}_M$, and $\boldsymbol{\beta}$). This is by far the fastest optimization option, but cannot implement every option. The fixed point procedure cannot implement the "IE Add" model type.

Converge

The convergence criterion for the optimization algorithm. The algorithm completes and is considered to "converge" when at two subsequent iterations the change in objective function is less than this amount. This value defaults to "1e-6" or 1×10^{-6}.

d Max

The maximum value for the user competency. Setting the maximum value for the competency prevents competencies from going to infinity (i.e., to an asymptote of the likelihood function).

IE Max

The maximum value for item easiness. As per "d Max", setting the maximum value for the item easiness prevents the item easiness values from going to infinity.

Form Buttons

The main form has five buttons, which are situated at the top right hand side of the form. The functionality for each button is described below.

Run

This option is only available if a csv file has been specified in the Data Input frame. Load in the data and run the specified FlexCCT model. Run the FlexCCT optimization procedure and output the model parameters and model fit statistics in the CCT Results Output form.

Default

Return all of the form fields to their default settings.

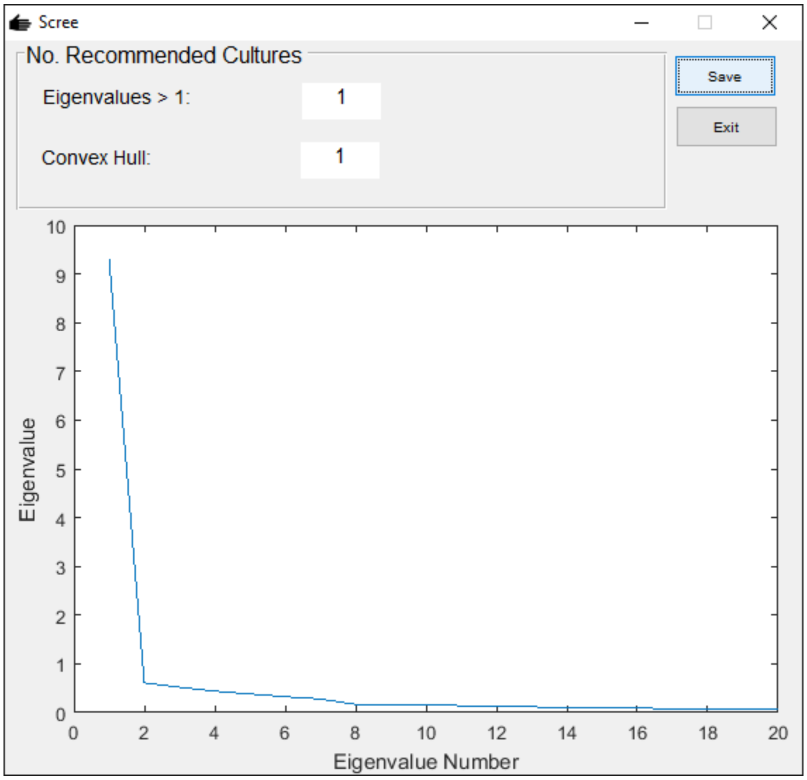
Advanced

Specify advanced features. The advanced features include the ability to specify item identifiability options, define confidence intervals, and set the number of cultures/clusters.

Scree

This option is only available if a csv file has been specified in the Data Input frame. Shows the scree plot form, which can be used to guess the number of clusters/cultures present in the data.

The Scree Plot



The Scree plot form utilizes an eigenvalue decomposition from Comrey's minimum residual factor analysis of the inter-rater correlation matrix to guess the number of clusters or cultures in the data. The rationale behind the use of this technique is that if there is one culture and all the raters have a similar pattern of answers then the correlation matrix should show a one factor structure. Multiple cultures indicates different preferences, e.g., when rating movies, some users like horror movies and hate romance movies, but other users have opposite preferences. In the main graph for the form, each eigenvector number (1^{st}, 2^{nd}, 3^{rd} etc.) is plotted against its actual value. This plot can be used to infer the number of eigenvectors using the criteria in the No. Recommended Cultures criteria:

No. Recommended Cultures

EigenValues > 1

This heuristic states that eigenvalues are taken as significant if they are greater than 1. This is a common criterion in factor analysis. In the preceeding example, only the 1^{st} eigenvalue is greater than 1, so the number of recommended cultures is 1.

Convex Hull

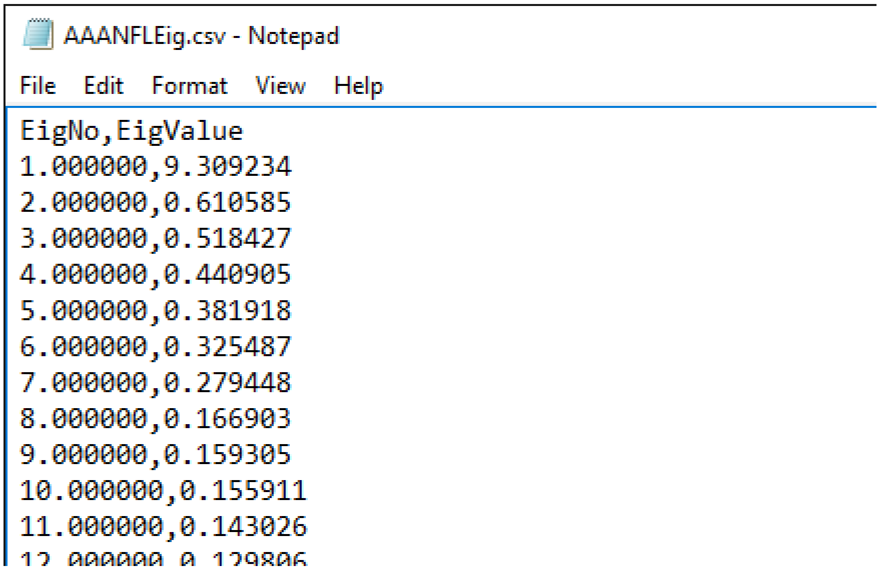
Here, the "convex hull" of the graph is used to determine the number of cultures. The convex hull is the point at which the graph of the eigenvalues goes from making an angle greater than 45° with the origin to making an angle less than 45° with the origin. The chosen eigenvalue number is the one before the transition from >45° to <45°. In the preceding example, the transition occurs at the 2^{nd} eigenvalue, so the number of recommended cultures is 1.

Form Buttons

The main form has two buttons, which are situated at the top right hand side of the form. The functionality for each button is described below.

Save

Allows the saving of the eigenvalue information that is plotted in the graph into a csv file. Brings up a standard Microsoft Windows file dialog box. If a file name is chosen and OK is checked then the file will be saved. An example of the format used is given below.



AAANFLEig.csv - Notepad

File Edit Format View Help

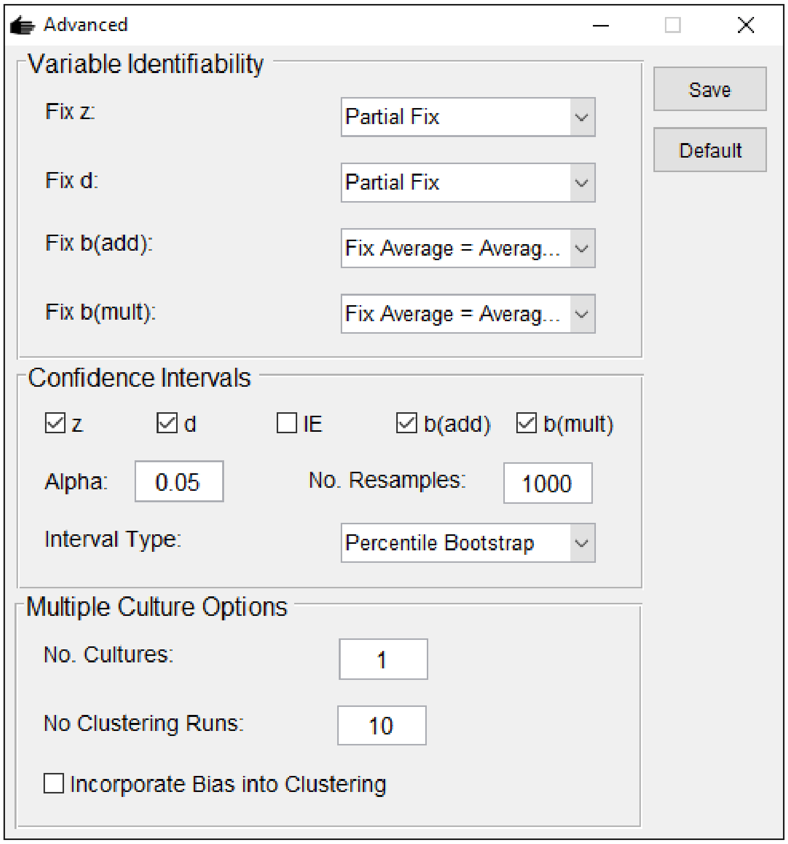
```
EigNo,EigValue
1.000000,9.309234
2.000000,0.610585
3.000000,0.518427
4.000000,0.440905
5.000000,0.381918
6.000000,0.325487
7.000000,0.279448
8.000000,0.166903
9.000000,0.159305
10.000000,0.155911
11.000000,0.143026
12.000000,0.129806
```

Exit

Exits the form and returns to the main GUI form.

Advanced Features

The Advanced Features form gives a range of features to help customize the FlexCCT analysis. Features include the ability to specify item identifiability options, define confidence intervals, and set the number of cultures/clusters.



Variable Identifiability

If all the variables in some of the advanced FlexCCT models are fit simultaneously, then the model may be over-identified, i.e., multiple sets of parameters with different values could give the same solution criteria. To prevent this problem, variables can be restricted, so as to give a grounded model or a model where different types of parameters have different precedencies. Variable identification options are given for the answer key (**z**), the rater competencies (**d**), the additive biases ($\mathbf{b}_A$) and the multiplicative biases ($\mathbf{b}_M$). The options are given as follows:

Fix z:

This variable determines how to fix **z**. The options are given below. The default value is “Partial Fix”.

- **No Fix:** Does not fix the variables. Allows the variables to be optimized freely.
- **Full Fix:** Fixes the **z** values to be the values from a model with no bias or item difficultly parameters. This essentially fits a sub-model, where $\mathbf{b}_A$, $\mathbf{b}_M$, and $\boldsymbol{\beta}$ parameters are then calculated as residual measures of fit.
- **Partial Fix:** Fixes a single value in **z** from the sub model given above. The value of **z** to be chosen is the one with the lowest average error. This grounds the model and prevents **z** values and bias values

from working against each other (e.g., increase each value in **z** by 1 and decrease each value in $\mathbf{b}_A$ by 1).

- **Fix Average = Average Mean:** Uses constrained optimization to fit the average value of **z** to the average of **z** calculated when taking a simple arithmetic mean. Again, this grounds the value of **z**.

Fix d:

This variable determines how to fix **d**. The options are given below. The default value is "Partial Fix".

- **No Fix:** Does not fix the variables. Allows the variables to be optimized freely.
- **Full Fix:** Fixes the **d** values to be the values from a model with no bias or item difficultly parameters. This essentially fits a sub-model, where $\mathbf{b}_A$, $\mathbf{b}_M$, and $\boldsymbol{\beta}$ parameters are residual measures of fit.
- **Partial Fix:** Fixes a single value in **d** from the sub model given above. The value of d_i to be chosen is the one closest to the average. As competency values are only important in a relative sense (e.g., via a competency weighted means), this prevents the competencies becoming arbitrarily low or high.
- **Fix Average = 1:** Uses constrained optimization to fit the average value of **d** to 1, grounding the competencies.

Fix b(add):

Both additive biases and multiplicative biases are assigned to individual raters. Identifiability options can be set to give precedence to fitting one type of bias over the other. This option is available if on the Bias Type dropdown menu, either "Additive Bias" or "Add and Mult Bias" are selected.

- **No Fix:** Does not fix the variables. Allows the variables to be optimized freely.
- **Full Fix:** Fixes the $\mathbf{b}_A$ values to be the values from a model with only additive bias fit. The multiplicative bias variables $\mathbf{b}_M$ are residual measures of fit.
- **Partial Fix:** Fixes a single $\mathbf{b}_A$ value from model given above. The value of b_{Ai} to be chosen is the one closest to the average. This prevents the additive biases from becoming arbitrarily low or high.
- **Fix Average = 0:** Uses constrained optimization to fit the average value of $\mathbf{b}_A$ to 0, grounding the additive bias around 0, i.e., no effective multiplicative bias.

Fix b(mult):

As above. This option is available if for the Bias Type dropdown menu either "Multiplicative Bias" or "Add and Mult Bias" are selected.

- **No Fix:** Does not fix the variables. Allows the variables to be optimized freely.
- **Full Fix:** Fixes the $\mathbf{b}_M$ values to be the values from a model with only multiplicative bias fit. The additive bias variables $\mathbf{b}_A$ are residual measures of fit.
- **Partial Fix:** Fixes a single $\mathbf{b}_M$ value from model given above. The value of $\mathbf{b}_M$ to be chosen is the one closest to the average. This prevents the additive biases from becoming arbitrarily low or high.
- **Fix Average = 1:** Uses constrained optimization to fit the average value of $\mathbf{b}_A$ to 1, grounding the multiplicative bias around 1, i.e. no effective multiplicative bias.

As the partial fix and full fix options for Fix b(add) and Fix b(mult) implicitly define a fitting hierarchy between $\mathbf{b}_A$ and $\mathbf{b}_M$, these options can only be chosen for either Fix b(add) and Fix b(mult) but not for both.

Confidence Intervals

The confidence intervals option allows $(1-\alpha)\times 100\%$ confidence intervals to be defined for each set of parameters included in the CCT model. The confidence intervals are calculated using a bootstrapping or jackknifing procedure. For a detailed overview of the implemented bootstrap procedures see the following link.

Singh, K., & Xie, M. (2008). Bootstrap: a statistical method. *Unpublished manuscript, Rutgers University, USA. Retrieved from* http://www.stat.rutgers.edu/home/mxie/RCPapers/bootstrap.pdf .

Confidence intervals can be added for variable sets **d**, **z**, $\mathbf{b}_A$, $\mathbf{b}_M$, and **β** by selecting the appropriate checkboxes.

- z – CIs for the solution answer key **z**.
- d **–** CIs for the rater competencies **d**.
- b(add) – CIs for the additive biases $\mathbf{b}_A$
- b(mult) – CIs for the multiplicative biases $\mathbf{b}_M$
- IE – CIs for the item easiness values **β**

Alpha

Determines the value of α to create a $(1-\alpha)\times 100\%$ confidence interval. For example, setting $\alpha = 0.05$ creates a 95% confidence interval.

No. Resamples

Bootstrapping takes samples without replacement from the original data sample. This field defines the number of resamples. Jack-knifing works by taking each item out of the original dataset and then using the parameter estimates from these “take one out” data samples to form a confidence interval. If No. Resamples is greater or equal to the number of items in the sample (*n*) then all *n* take one out samples are processed. If No. Resamples is less than *n* then “No. Resamples” are used, with the items to be removed selected randomly.

Interval Type

The Interval Type dropdown menu determines the type of confidence interval that is to be created. The first three options create bootstrap confidence intervals. The last option creates a jackknife confidence interval. The confidence intervals for each parameter are defined below. N is the number of resamples, α defines a $(1-\alpha)\times 100\%$ confidence interval, $\hat{\theta}$ is the point estimate of the parameter, and $\{\theta_1^*, \cdots, \theta_N^*\}$ are the ordered parameter values taken from the *N* resamples.

- **Percentile Bootstrap:** The basic percentile bootstrap, where the confidence interval for parameter θ is calculated as $\left[\theta^*_{N\left(\frac{\alpha}{2}\right)}, \theta^*_{N\left(1-\frac{\alpha}{2}\right)}\right]$.
- **Centered Percentile Bootstrap**: The centered percentile bootstrap, where the confidence interval for parameter θ is calculated as $\left[2\hat{\theta} - \theta^*_{N\left(1-\frac{\alpha}{2}\right)}, 2\hat{\theta} - \theta^*_{N\left(\frac{\alpha}{2}\right)}\right]$.
- **Bootstrap-t:** The *t*-distribution bootstrap, which utilizes a *t*-distribution confidence interval where $t^* = \left(\theta^* - \hat{\theta}\right)/SE(\theta^*)$.
- **Jackknife:** First calculates the standard error for the confidence interval using a series of hold-one out samples and then uses the standard error to create a bias-corrected confidence interval.

Multiple Culture Options

FlexCCT can simultaneously split raters in cultures and estimate within culture rater and item parameters (i.e., **d**, **z**, $\mathbf{b}_A$, $\mathbf{b}_M$, and $\boldsymbol{\beta}$). Raters are split to give the maximum within culture similarity. The number of cultures is specified by the users. The Scree Plot form can be used to give guidance on the number of clusters.

No Cultures

The number of cultures. Defaults to 1. In this case all raters are assigned to the one default cluster.

No. Clustering Runs

The number of clustering runs. FlexCCT utilizes a k-means like procedure and is not guaranteed to find a globally optimal solution. Multiple cluster runs increase the chance of finding globally optimal solutions. Defaults to 10.

Incorporate Bias into Clustering

If this option is checked, bias is accounted for when clustering raters. First, the CCT model is estimated with the assumption of an individual culture. Biases are calculated (depending on the bias settings on the main form) and bias corrected scores (subtracting additive bias and then dividing by multiplicative bias) are calculated for each user. The bias corrected scores are then used to cluster the raters into cultures.

Form Buttons

The main form has two buttons, which are situated at the top right hand side of the form. The functionality for each button is described below.

Save

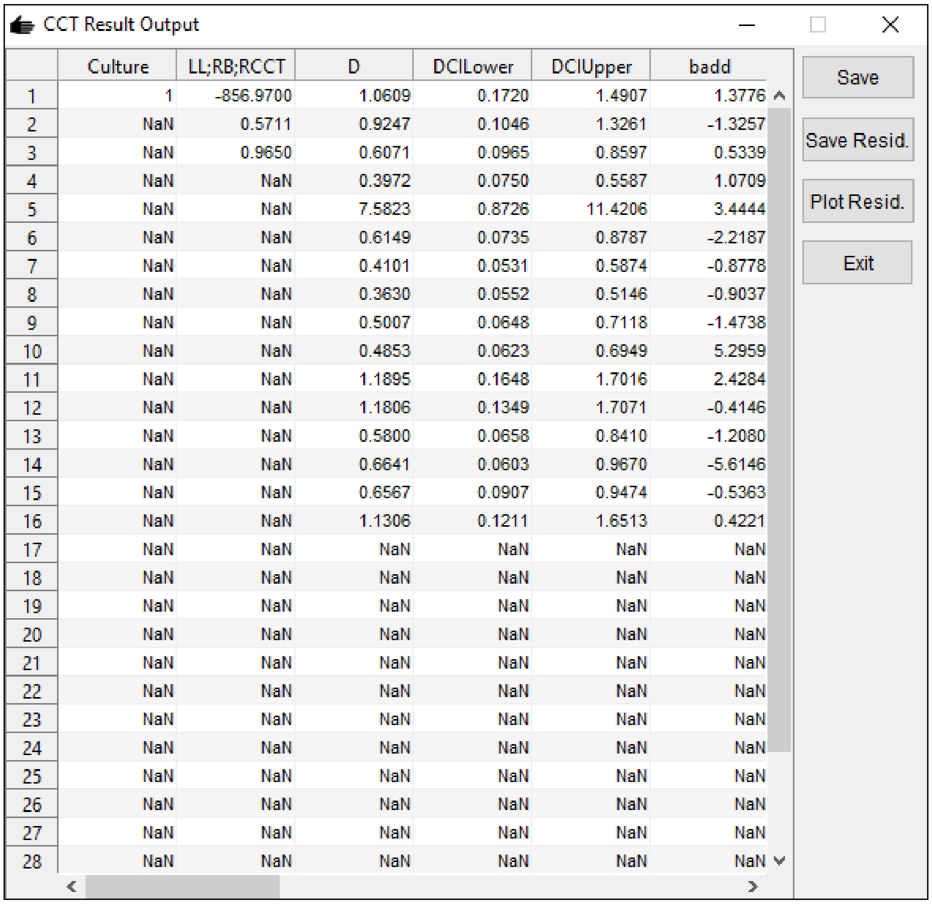
Save the currently selected advanced settings.

Default

Return all of the advanced setting form fields to their default settings.

CCT Results Output

The CCT Results Output form gives a listing of results. The columns returned are based upon the previously described settings.



	Culture	LL;RB;RCCT	D	DCILower	DCIUpper	badd
1	1	-856.9700	1.0609	0.1720	1.4907	1.3776
2	NaN	0.5711	0.9247	0.1046	1.3261	-1.3257
3	NaN	0.9650	0.6071	0.0965	0.8597	0.5339
4	NaN	NaN	0.3972	0.0750	0.5587	1.0709
5	NaN	NaN	7.5823	0.8726	11.4206	3.4444
6	NaN	NaN	0.6149	0.0735	0.8787	-2.2187
7	NaN	NaN	0.4101	0.0531	0.5874	-0.8778
8	NaN	NaN	0.3630	0.0552	0.5146	-0.9037
9	NaN	NaN	0.5007	0.0648	0.7118	-1.4738
10	NaN	NaN	0.4853	0.0623	0.6949	5.2959
11	NaN	NaN	1.1895	0.1648	1.7016	2.4284
12	NaN	NaN	1.1806	0.1349	1.7071	-0.4146
13	NaN	NaN	0.5800	0.0658	0.8410	-1.2080
14	NaN	NaN	0.6641	0.0603	0.9670	-5.6146
15	NaN	NaN	0.6567	0.0907	0.9474	-0.5363
16	NaN	NaN	1.1306	0.1211	1.6513	0.4221
17	NaN	NaN	NaN	NaN	NaN	NaN
18	NaN	NaN	NaN	NaN	NaN	NaN
19	NaN	NaN	NaN	NaN	NaN	NaN
20	NaN	NaN	NaN	NaN	NaN	NaN
21	NaN	NaN	NaN	NaN	NaN	NaN
22	NaN	NaN	NaN	NaN	NaN	NaN
23	NaN	NaN	NaN	NaN	NaN	NaN
24	NaN	NaN	NaN	NaN	NaN	NaN
25	NaN	NaN	NaN	NaN	NaN	NaN
26	NaN	NaN	NaN	NaN	NaN	NaN
27	NaN	NaN	NaN	NaN	NaN	NaN
28	NaN	NaN	NaN	NaN	NaN	NaN

All the results are placed in an exportable numeric matrix. Separate results are given for each separate culture. The columns for each culture are described below:

- LL;RB;RCCT: This column gives the overall measures of fit. The first value (LL) is the optimal log-likelihood value. The second value (RB) is the basic measure of solution reliability described in France and Batchelder (2015) and from Janson and Olson (2001). The third value (RCCT) is a consensus adjusted reliability metric, as given in France and Batchelder (2015).
- d: The values of **d**, the rater competencies.
- dCILower: The lower bound of the confidence interval for **d**.
- dCIUpper: The upper bound of the confidence interval for **d**.
- badd: The rater additive biases $\mathbf{b}_A$.
- baddCILower: The lower bound of the confidence interval for $\mathbf{b}_A$.
- baddCIUpper: The upper bound of the confidence interval for $\mathbf{b}_A$.
- bmult: The rater multiplicative biases $\mathbf{b}_M$.
- bmultCILower: The lower bound of the confidence interval for $\mathbf{b}_M$.
- bmultCIUpper: The upper bound of the confidence interval for $\mathbf{b}_M$.
- z: The rater multiplicative biases **z**.

- zCILower: The lower bound of the confidence interval for **z**
- zCIUpper: The upper bound of the confidence interval for **z**
- IE: The item easiness values **β**.
- IECILower: The lower bound of the confidence interval for **β**.
- IECIUpper: The upper bound of the confidence interval for **β**.
- LLPartial(LB): The value for the partial log-likelihood values of $\lfloor \boldsymbol{z} \rfloor$, i.e., the integer lower bound for **z**.
- LLPartial: The value for the partial log-likelihood values of $\lfloor \boldsymbol{z} \rfloor$, i.e., the integer lower bound for **z**.
- LLPartial(UB): The value for the partial log-likelihood values of $\lfloor \boldsymbol{z} \rfloor$, i.e., the integer lower bound for **z**.

The rationale behind the partial log-likelihoods is to provide results for integer values of **z**, in applications in which integer results are important.

Form Buttons

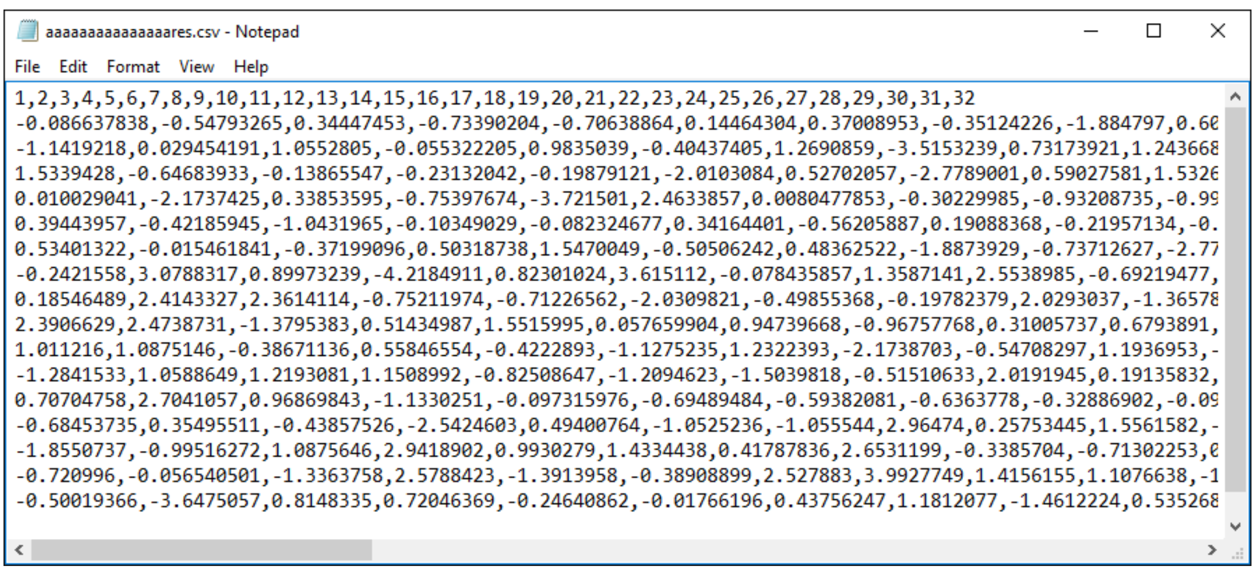
The results form has four buttons, which are situated at the top right hand side of the form. The functionality for each button is described below.

Save

Allows a user save the results in the results matrix into a .csv data file using the standard Microsoft Windows file dialog box.

Save Resid.

Allows a user save the model residuals into a .csv data file. For a rater *i* and an item *k*, the residual is calculated as $\varepsilon_{ik} = x_{ik} - (b_{Mi} z_k + b_{Ai})$. For *n* users and m items, an $n \times m$ matrix of residuals is plotted. Each column is labeled with the number of the item. An example residuals file, for 16 users and 32 items, is given below.




```
1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32
-0.086637838,-0.54793265,0.34447453,-0.73390204,-0.70638864,0.14464304,0.37008953,-0.35124226,-1.884797,0.6
-1.1419218,0.029454191,1.0552805,-0.055322205,0.9835039,-0.40437405,1.2690859,-3.5153239,0.73173921,1.24366
1.5339428,-0.64683933,-0.13865547,-0.23132042,-0.19879121,-2.0103084,0.52702057,-2.7789001,0.59027581,1.532
0.010029041,-2.1737425,0.33853595,-0.75397674,-3.721501,2.4633857,0.0080477853,-0.30229985,-0.93208735,-0.9
0.39443957,-0.42185945,-1.0431965,-0.10349029,-0.082324677,0.34164401,-0.56205887,0.19088368,-0.21957134,-0.
0.53401322,-0.015461841,-0.37199096,0.50318738,1.5470049,-0.50506242,0.48362522,-1.8873929,-0.73712627,-2.7
-0.2421558,3.0788317,0.89973239,-4.2184911,0.82301024,3.615112,-0.078435857,1.3587141,2.5538985,-0.69219477,
0.18546489,2.4143327,2.3614114,-0.75211974,-0.71226562,-2.0309821,-0.49855368,-0.19782379,2.0293037,-1.3657
2.3906629,2.4738731,-1.3795383,0.51434987,1.5515995,0.057659904,0.94739668,-0.96757768,0.31005737,0.6793891,
1.011216,1.0875146,-0.38671136,0.55846554,-0.4222893,-1.1275235,1.2322393,-2.1738703,-0.54708297,1.1936953,-
-1.2841533,1.0588649,1.2193081,1.1508992,-0.82508647,-1.2094623,-1.5039818,-0.51510633,2.0191945,0.19135832,
0.70704758,2.7041057,0.96869843,-1.1330251,-0.097315976,-0.69489484,-0.59382081,-0.6363778,-0.32886902,-0.0
-0.68453735,0.35495511,-0.43857526,-2.5424603,0.49400764,-1.0525236,-1.055544,2.96474,0.25753445,1.5561582,-
-1.8550737,-0.99516272,1.0875646,2.9418902,0.9930279,1.4334438,0.41787836,2.6531199,-0.3385704,-0.71302253,
-0.720996,-0.056540501,-1.3363758,2.5788423,-1.3913958,-0.38908899,2.527883,3.9927749,1.4156155,1.1076638,-1
-0.50019366,-3.6475057,0.8148335,0.72046369,-0.24640862,-0.01766196,0.43756247,1.1812077,-1.4612224,0.53526
```

Plot Resid.

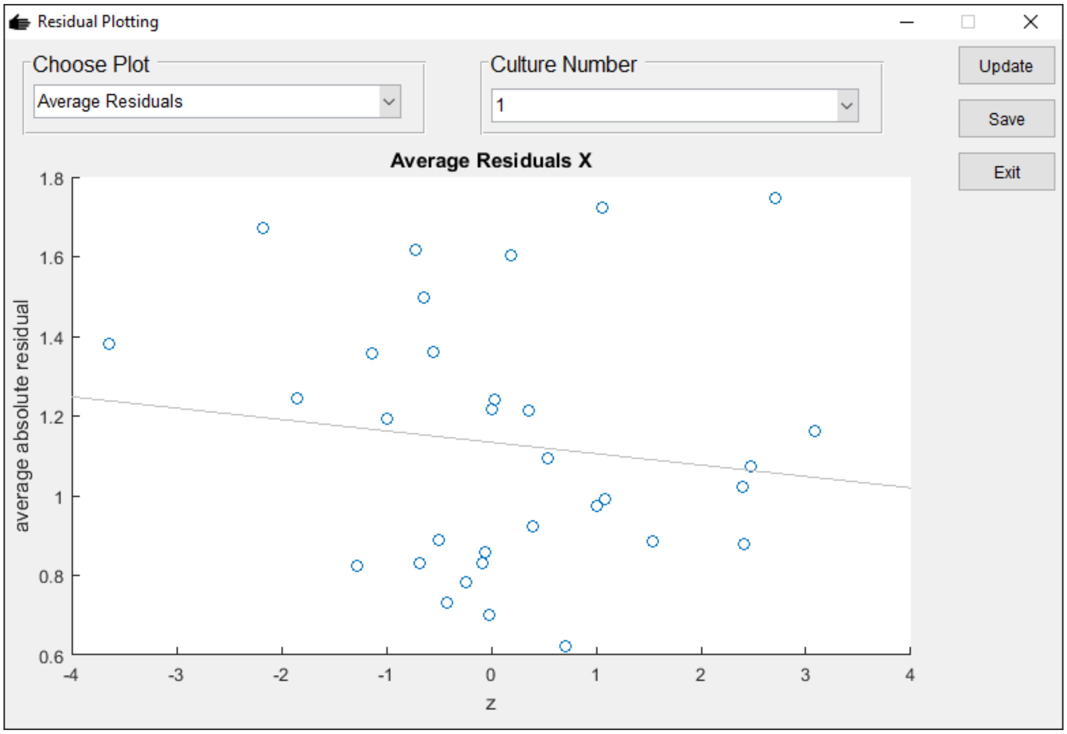
Opens the "Residual Plots" form, which gives a visual analysis of the model residuals.

Exit

Exits from the CCT Results Output form.

Residual Plots

The Residual Plots form gives graphs of model residuals, allowing users to examine residual patterns and check for error heteroscedasticity. The residuals are plotted against the answer key scores contained in **z**.



Choose Plot

The dropdown specifies the current plot type. The plot types are:

Average Residuals: Plots the answer key score z_k against the average absolute residual score $\sum_{i=1}^{n}|\varepsilon_{ik}|/n$ for each answer key item. Plots the optimal least squares fit line. If the line has a steep slope then this indicates error heterogeneity and data transformation (e.g., log transformation) may be recommended for the data.

All Residuals: Plots the answer key score against the residual value ε_{ik} for each combination of rater and item. If the spread of the residuals increases with z_k then this indicates error heterogeneity and data transformation may be recommended for the data.

Culture Number

Residuals are plotted for a single culture. For situations where there are multiple cultures, the Culture Number dropdown selects the currently displayed culture.

Form Buttons

The results form has three buttons, which are situated at the top right hand side of the form. The functionality for each button is described below.

Update

Updates the plot to reflect the current settings.

Save

Allows the saving of the current plot in a graphics (jpg, png, gif, or bmp) file. Brings up a standard Microsoft Windows file dialog box. If a file name is chosen and OK is checked then the file will be saved.

Exit

Exits the form.

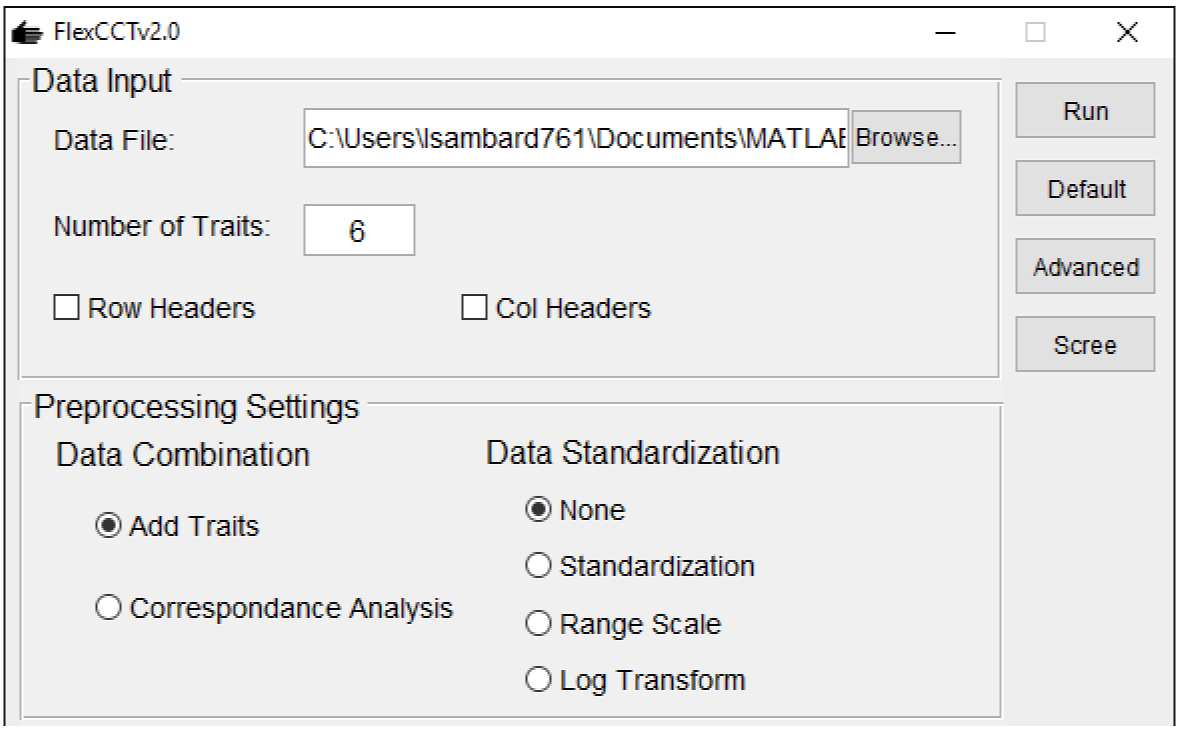
Example: Essay Grading Competency

Overview

This example gives a typical educational evaluation example for the use of CCT. The example is taken from the previously referenced France and Batchelder (2015) paper. A series of essays were graded by two expert graders and 12 peer graders, giving a total of 14 evaluators. Each grader rated 50 different essays. These essays were written by middle school students on the subject of laughter. Each essay was graded using a fixed grading rubric that contained six different attributes. These attributes were “Ideas and Content”, “Organization”, “Voice”, “Word Choice”, “Sentence Fluency”, and “Conventions”. Each essay was graded on each attribute from 1 to 6 by each rater. The attributes were grouped using the format given on p2 in the file GradingSixTraits.csv, which is included in the install directory.

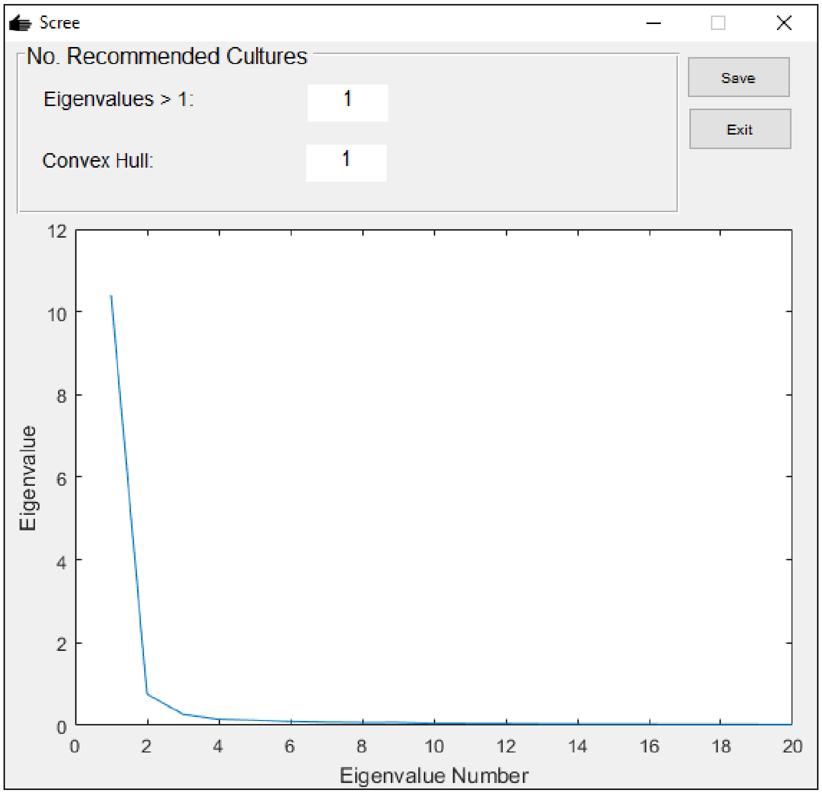
Instructions.

From the main screen select the file, type in 6 into the Number of Traits textbox and then select “Add Traits” in the Data Combination section.



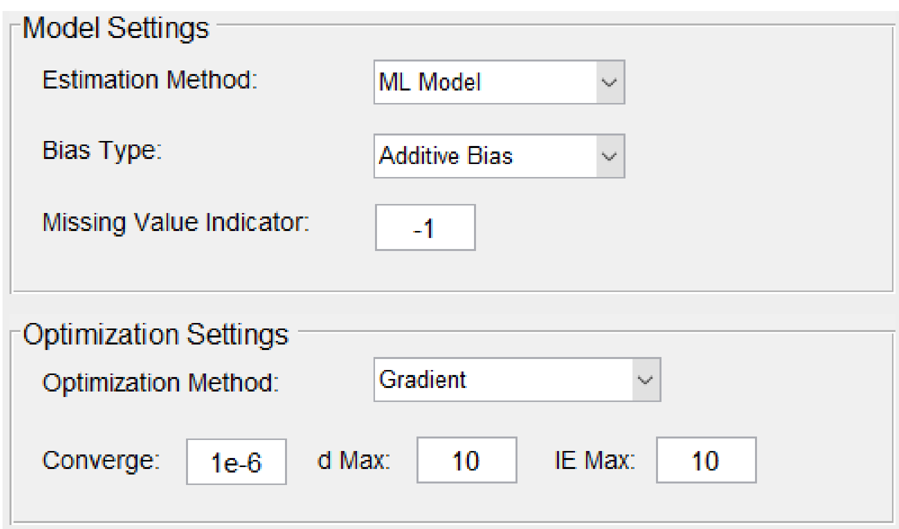
Given that all raters used the same grading rubric, it would be a sensible to assume that the raters belong to a single “rating” culture. However, this is not necessarily always true. Different groups of raters could interpret the rubric in different ways. Thus, it is a sensible precaution to ensure that the number of cultures is one.

Press the Scree button



The scree plot form gives a plot of eigenvalue number versus eigenvalue for a principal axes factor analysis of the inter-rater correlations. Two common criteria for choosing the number of factors in factor analysis are 1) the number of eigenvalues less than one and ii) the convex hull, which is the eigenvalue number where the eigenvalue line plot transitions from making an angle >45° to the origin to making an angle <45° to the origin. In both cases here the number of cultures is equal to 1.

First chose a simple maximum likelihood model, with an additive bias. The input options on the main form are given below. Choose the basic "ML Model" with additive bias. Keep with the default advanced settings.



Click the run button. The results are given below.

CCT Results Output

	ılture	LL;RB;RCCT	d	badd	bmult	z
1	1	-1.8180e+03	0.1100	0.3395	1	6.5091
2	NaN	0.7415	0.2042	1.0995	1	8.0021
3	NaN	0.8267	0.0854	-3.4405	1	12.7414
4	NaN	NaN	0.0809	-3.8005	1	9.5482
5	NaN	NaN	0.1505	1.1795	1	15.9004
6	NaN	NaN	0.1177	-0.4406	1	11.3296
7	NaN	NaN	0.0662	0.6595	1	11.2551
8	NaN	NaN	0.1638	0.6995	1	12.8715
9	NaN	NaN	0.0356	2.3395	1	14.4037
10	NaN	NaN	0.0739	-0.3405	1	14.9339
11	NaN	NaN	0.1471	-2.6205	1	14.8039
12	NaN	NaN	0.0400	-1.6805	1	13.1267
13	NaN	NaN	0.1212	0.9795	1	15.5264
14	NaN	NaN	0.0876	-4.8405	1	16.9955
15	NaN	NaN	NaN	NaN	NaN	14.6560
16	NaN	NaN	NaN	NaN	NaN	16.4982
17	NaN	NaN	NaN	NaN	NaN	17.7806
18	NaN	NaN	NaN	NaN	NaN	17.1014
19	NaN	NaN	NaN	NaN	NaN	16.1095
20	NaN	NaN	NaN	NaN	NaN	19.0084
21	NaN	NaN	NaN	NaN	NaN	18.2676
22	NaN	NaN	NaN	NaN	NaN	20.5942
23	NaN	NaN	NaN	NaN	NaN	21.4487
24	NaN	NaN	NaN	NaN	NaN	19.9335
25	NaN	NaN	NaN	NaN	NaN	23.0985
26	NaN	NaN	NaN	NaN	NaN	22.0664
27	NaN	NaN	NaN	NaN	NaN	21.9587
28	NaN	NaN	NaN	NaN	NaN	23.9433

Save

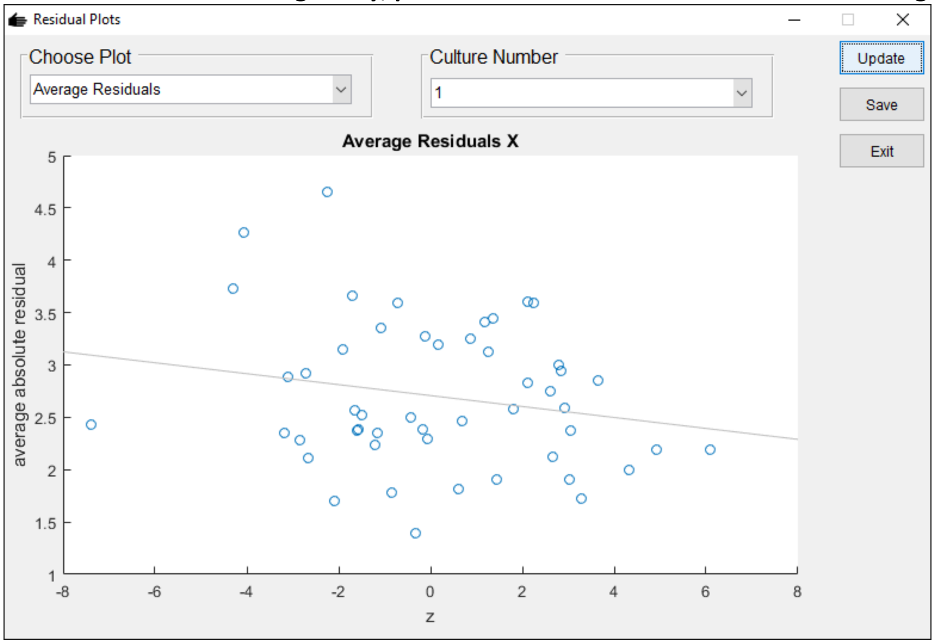
Save Resid.

Plot Resid.

Exit

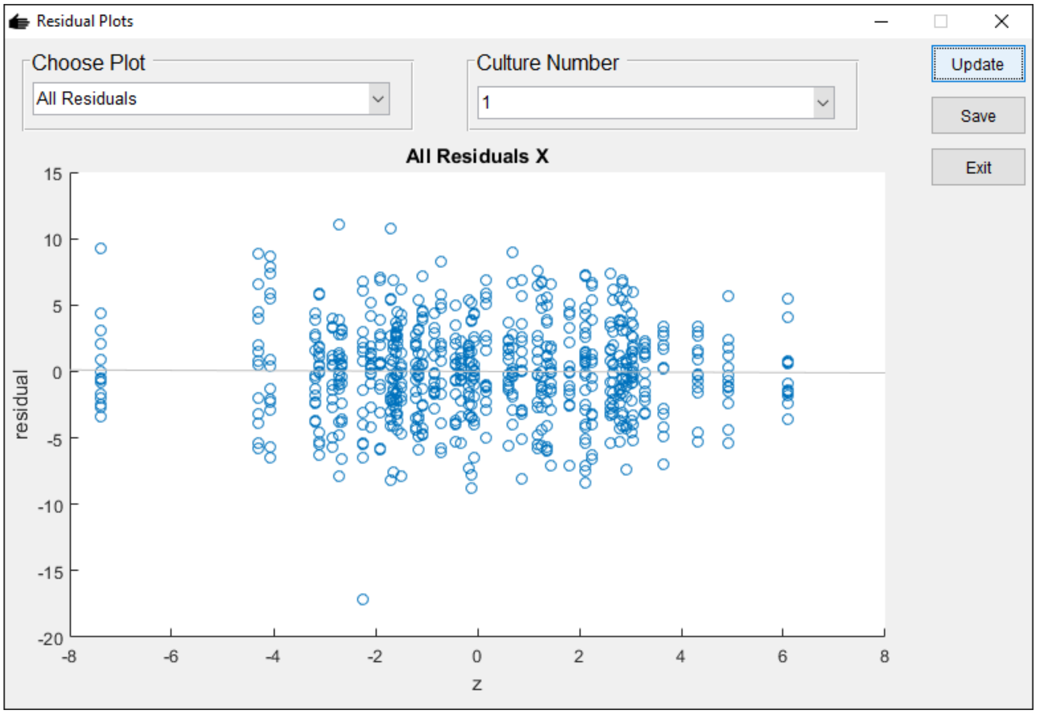
Once can see a range of competencies (d), ranging from 0.049 to 0.2042. The first two raters are the expert raters. Expert rater 2 has the highest competency of all the rates, but expert rater 1 has a more than average competency. Note, that the results have high reliability (0.7415) and even higher competency adjusted reliability (0.8267).

To check for error heterogeneity, press the Plot Resid. Button. The results are given below.



The plot shows the average magnitude of residuals across the values of z. There is a slight downward trend in the residuals, but no clear trend indicating heterogeneity.

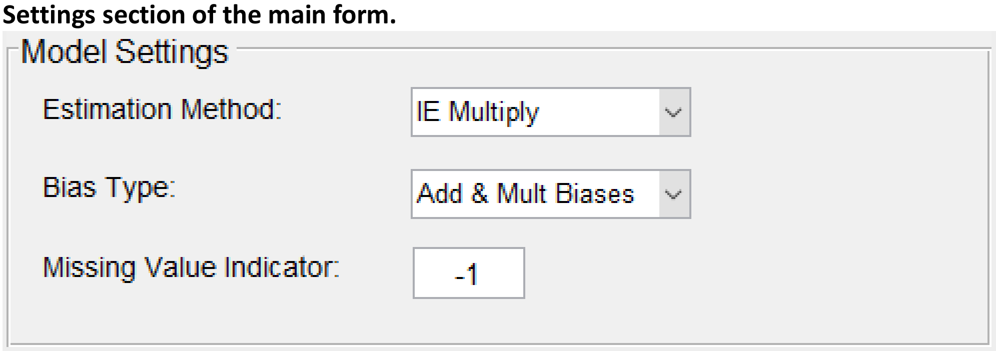
To get a further review of the residuals, click the All Residuals dropdown and then click the Update button.



This further shows that relative to the range of residuals, any trend is almost non-existent, giving further evidence to a lack of error heterogeneity.

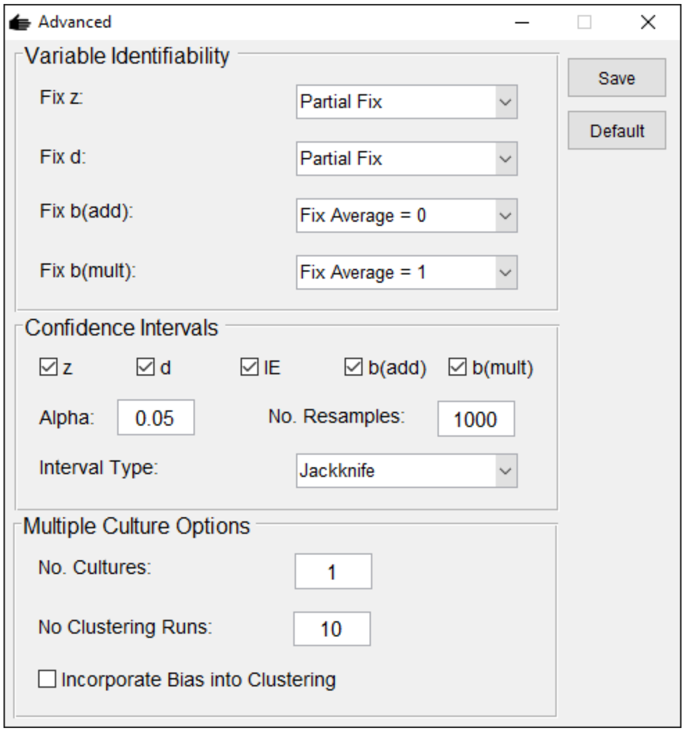
Instructions for Advanced Model

Now select the multiplicative item easiness model with both additive and multiplicative bias in the Model Settings section of the main form.



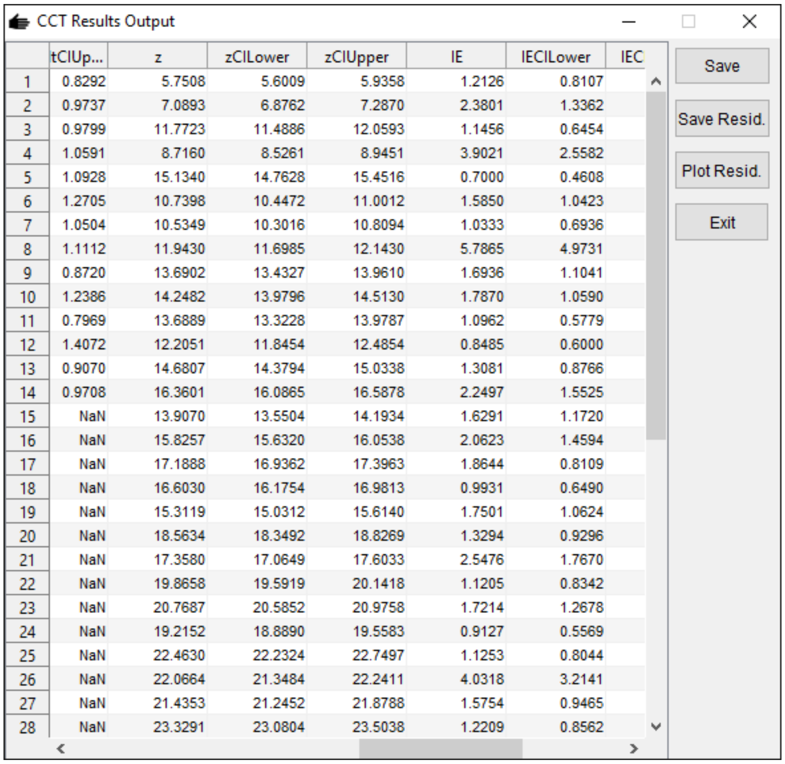
The model is now more flexible and fits item easiness parameters **β**, additive biases $\mathbf{b}_A$, and multiplicative biases $\mathbf{b}_M$.

Click the "Advanced" button and configure the form as follows and click the Save button.



The Variable Identifiability section is used to ground parameters and prevent parameters being traded off against one another. For example, the partial fixes for **z** sets a single value of **z** to be its value for a restricted model for no bias. The partial fix for **d** does likewise. We choose jackknife take-one out sampling for the confidence intervals (hence the number of resamples is not used). The confidence intervals provide a measure of confidence for the underlying parameters. For rater parameters (**d**, $\mathbf{b}_A$, and $\mathbf{b}_M$), the confidence intervals give population estimates of the parameters, under the assumption that the items are sampled randomly from an underlying population of items. For item parameters (**β**, **z**), the assumption that the raters are sampled randomly from an underlying population of raters. The value of Alpha (α), defines a $100\%(1-\alpha)$ confidence interval.

Click the Run Button



	tCIUp...	z	zCILower	zCIUpper	IE	IECILower	IEC
1	0.8292	5.7508	5.6009	5.9358	1.2126	0.8107	
2	0.9737	7.0893	6.8762	7.2870	2.3801	1.3362	
3	0.9799	11.7723	11.4886	12.0593	1.1456	0.6454	
4	1.0591	8.7160	8.5261	8.9451	3.9021	2.5582	
5	1.0928	15.1340	14.7628	15.4516	0.7000	0.4608	
6	1.2705	10.7398	10.4472	11.0012	1.5850	1.0423	
7	1.0504	10.5349	10.3016	10.8094	1.0333	0.6936	
8	1.1112	11.9430	11.6985	12.1430	5.7865	4.9731	
9	0.8720	13.6902	13.4327	13.9610	1.6936	1.1041	
10	1.2386	14.2482	13.9796	14.5130	1.7870	1.0590	
11	0.7969	13.6889	13.3228	13.9787	1.0962	0.5779	
12	1.4072	12.2051	11.8454	12.4854	0.8485	0.6000	
13	0.9070	14.6807	14.3794	15.0338	1.3081	0.8766	
14	0.9708	16.3601	16.0865	16.5878	2.2497	1.5525	
15	NaN	13.9070	13.5504	14.1934	1.6291	1.1720	
16	NaN	15.8257	15.6320	16.0538	2.0623	1.4594	
17	NaN	17.1888	16.9362	17.3963	1.8644	0.8109	
18	NaN	16.6030	16.1754	16.9813	0.9931	0.6490	
19	NaN	15.3119	15.0312	15.6140	1.7501	1.0624	
20	NaN	18.5634	18.3492	18.8269	1.3294	0.9296	
21	NaN	17.3580	17.0649	17.6033	2.5476	1.7670	
22	NaN	19.8658	19.5919	20.1418	1.1205	0.8342	
23	NaN	20.7687	20.5852	20.9758	1.7214	1.2678	
24	NaN	19.2152	18.8890	19.5583	0.9127	0.5569	
25	NaN	22.4630	22.2324	22.7497	1.1253	0.8044	
26	NaN	22.0664	21.3484	22.2411	4.0318	3.2141	
27	NaN	21.4353	21.2452	21.8788	1.5754	0.9465	
28	NaN	23.3291	23.0804	23.5038	1.2209	0.8562	

For each set of parameters with a specified confidence interval, the lower bound and upper bound for the confidence interval are given after the actual value. To save all of the results and perhaps plot/do initial analyses, the Save button allows an output csv file to be exported using a standard Microsoft file dialog box.