

Classical & Bayesian Instrument Development

University of Kansas Department of Biostatistics

A web application for psychometric analyses

Reliability estimation using a web application

Sponsored by R03NR013236

This GUI references the [Lavaan](#) , [MCMCpack](#) , [OpenBUGS](#) and [psy](#) packages and was built using [Shiny](#).
Authors: Alex Karanevich, Lili Garrard, Marge Bott, Larry Price, Byron Gajewski

2015

Teach yourself CBID

Who should take the tutorial?

You will get the most out of this tutorial if you know the very basics of confirmatory factor analysis and reliability (e.g. Cronbach's alpha).

If confirmatory factor analysis is new to you, you may want to check out the learning resources at <http://lavaan.ugent.be/> before taking this tutorial.

Get started with CBID

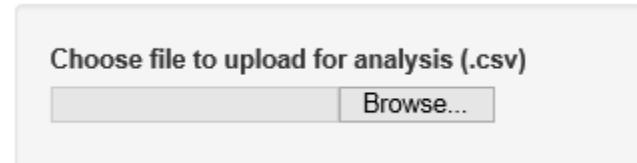
These four tutorials will take you from a confirmatory factor analysis consumer to a user. Also, you will be able to better estimate measures of reliability of your instruments using both classical and Bayesian perspectives.

Click the Tutorial 1 button to get started!

- [Tutorial 1](#) – Data file preparation
- [Tutorial 2](#) – One-factor model

Tutorial 1 – Data file preparation

Before getting started with CBID, please ensure that your data file(s) are properly prepared according to the following **general** requirements:



1. CBID only accepts data files in .csv format for uploading. Most data processing software (e.g., Microsoft Excel, SPSS) will have an option to save the file in .csv format.
2. Data file(s) should be structured in the long format (i.e., subjects as rows and items as columns) and only contain relevant variables (items) used for factor analysis. For example, if data were collected from 50 subjects on 5 items, the data file should have 50 rows and 5 columns, with each column representing one item on the instrument.
3. The first row of the data file should be reserved for item names (e.g., item1, item2) as CBID reads in the first row as variable names used in the analysis. Item names should be descriptive and precise.
4. If your file contains data for multiple groups and separate analysis needs to be conducted for each group, please create a separate data file for each group prior to uploading. CBID currently does not have the function to separate groups or create subsets of data.
5. If your file contains data collected from negatively worded items, please ensure that the negatively worded items have been properly reverse coded prior to uploading. This step is essential to prevent issues from a domain sampling perspective.

Selecting **Bayesian analysis** will provide three options: flat prior, previous data, and expert prior. Both *previous data* and *expert prior* options will require users to upload a separate *prior* data file (i.e., previously collected participant data or subject experts' content data). This is in addition to the currently collected participants' data file. The additional *prior* data file will be used to compute the *priors* needed for Bayesian analysis. The following requirements apply if Bayesian analysis will be performed:

1. The variable or item names in the first row of the *prior* data file must match the variable or item names in the first row of the participants' data file. If the two files do not have matching item names, CBID will output an error message.
2. Both the *prior* data file and the participants' data file must have the same number of columns.
3. If *previous data* is selected, a data file containing previously collected data can be uploaded, as shown in the screen shot below.

How to get the prior distribution?

Previous Data ▾

Choose prior data to upload (.csv)

Browse...

4. If *expert prior* is selected, a data file containing subject experts' content data can be uploaded, as shown in the screen shot below.

How to get the prior distribution?

Expert Prior ▾

Choose expert data to upload (.csv)

Browse...

[Return to Main Menu](#)

Tutorial 2 – One-factor model

CBID offers two approaches to help establish evidence of construct validity: classical and Bayesian. In this tutorial, we will go through a simple example of each to demonstrate how to use CBID.

The screen shot below shows what CBID looks like when first initiated in your selected web browser.

Classical & Bayesian Instrument Development (Beta Version)

This GUI references the [Lavaan](#), [MCMCpack](#), [OpenBUGS](#) and [psyc](#) packages and was built using [Shiny](#).
Authors: Alex Karanovich, Lili Garrard, Marge Bott, Larry Price, Byron Gajewski

[View the Tutorial](#)

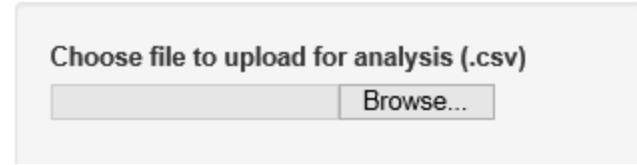
Now, let's consider a simple one-factor model with 6 items. Each item can be rated on a 5-point scale. We will follow the steps below to conduct the analysis.

Step 1: Upload the relevant participants' data file

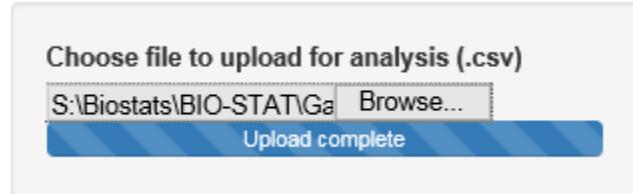
Below is a partial screen shot of the sample participants' data file (*participants.csv*) that we will use. Notice that the first row of this .csv file is reserved for the item names and only the relevant variables are included in the file.

A	B	C	D	E	F
Item1	Item2	Item3	Item4	Item5	Item6
3	2	3	1	5	4
5	4	2	5	2	4
3	1	1	1	1	1
2	3	4	4	5	2
1	2	3	2	1	2
4	1	1	1	2	2
2	1	4	2	1	3

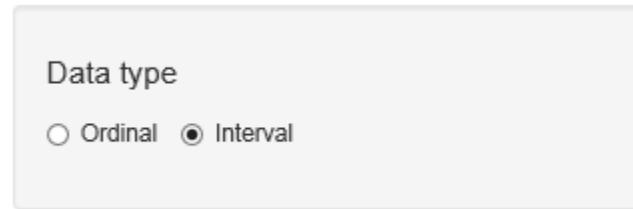
Prior to uploading *participants.csv* to CBID, users will see the following option:



Click on *Browse* and a new window will open for users to locate the data file used for analysis. Once the data file is selected, click *Open* in the new window and CBID will display a blue bar to indicate a successful upload (shown below).



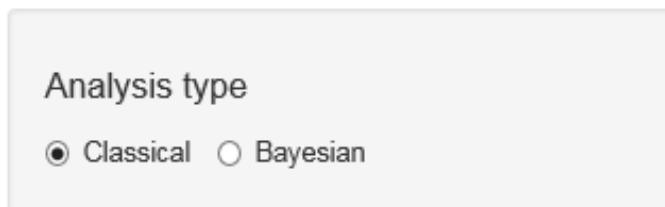
Step 2: Select the appropriate data type



Responses collected from participants can be treated as either interval or ordinal, depending on the number of response categories per item. Traditionally, participant responses are often modeled as interval, even for binary response options. With the advancement in statistical methodology and software, we recommend users to select the *Ordinal* option if the number of response categories for the majority of the items is less than 5. Otherwise, the *Interval* option can be selected. For the classical approach, choosing to model the data as either *Ordinal* or *Interval* will produce different results as different estimators are used to estimate the model parameters. Maximum likelihood (ML) estimator is used for the *Interval* option and weighted least squares means- and variance-adjusted (WLSMV) estimator is used for the *Ordinal* option. For the current example, we will model the data as ordinal.

Step 3: Select analysis type

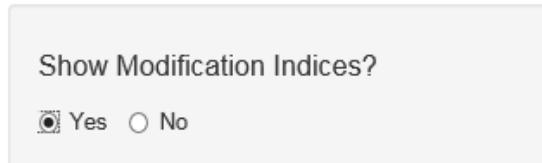
In this step, users will select whether to conduct a classical or Bayesian analysis. For the first demonstration, we will select *Classical*.



Classical Approach

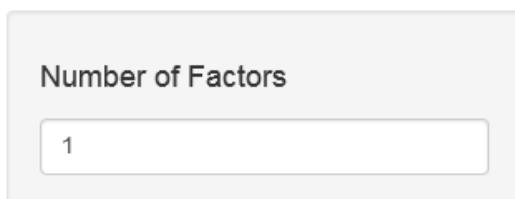
Step 4: Select modification indices

Users may choose to view modification indices if the *Classical* approach is selected, by using the following option. For this example, we want to view the modification indices outputs.



Step 5: Select the number of factors

Users need to specify the number of factors or the number of constructs of interest. For this example, we select a one-factor model.



As previously mentioned, CBID extracts the first row of the participants' data file as the variable names used in analysis. By selecting the number of factors in the model, CBID automatically displays the following option for users to specify the items that load on the single factor. Suppose all 6 items load on the single factor, labeled generically by CBID as *Factor 1*.

Factor 1 items

Item1 Item2 Item3 Item4 Item5 Item6

Step 6: Run the analysis

After all options have been appropriately selected, we can now run the entire analysis by clicking on the *Go* button.



Step 7: Review the outputs

The outputs are organized in 5 sections: [model fit indices](#), [parameter estimates](#), [modification indices](#), [entire reliability](#) and Cronbach's alpha, and potential warning or error messages. Below are screen shots for each of the sections, based on our sample data.

Model fit indices:

Summary:

lavaan (0.5-18) converged normally after 12 iterations

Number of observations	50
------------------------	----

Estimator	DWLS	Robust
Minimum Function Test Statistic	6.989	10.124
Degrees of freedom	9	9
P-value (Chi-square)	0.638	0.341
Scaling correction factor		0.811
Shift parameter		1.509

for simple second-order correction (Mplus variant)

Model test baseline model:

Minimum Function Test Statistic	148.560	120.175
Degrees of freedom	15	15
P-value	0.000	0.000

User model versus baseline model:

Comparative Fit Index (CFI)	1.000	0.989
Tucker-Lewis Index (TLI)	1.025	0.982

Root Mean Square Error of Approximation:

RMSEA	0.000	0.050
90 Percent Confidence Interval	0.000 0.133	0.000 0.173
P-value RMSEA <= 0.05	0.719	0.436

Weighted Root Mean Square Residual:

WRMR	0.423	0.423
------	-------	-------

Parameter estimates (partial outputs):

Parameter estimates:						
	Information		Expected			
	Standard Errors		Robust.sem			
	Estimate	Std.error	Z-value	P(> z)	Std.lv	Std.all
Latent variables:						
F1 =~						
Item1	0.326	0.143	2.276	0.023	0.326	0.326
Item2	0.839	0.097	8.643	0.000	0.839	0.839
Item3	0.717	0.098	7.338	0.000	0.717	0.717
Item4	0.680	0.118	5.758	0.000	0.680	0.680
Item5	0.457	0.133	3.424	0.001	0.457	0.457
Item6	0.546	0.113	4.819	0.000	0.546	0.546
Intercepts:						
Item1	0.000				0.000	0.000
Item2	0.000				0.000	0.000
Item3	0.000				0.000	0.000
Item4	0.000				0.000	0.000
Item5	0.000				0.000	0.000
Item6	0.000				0.000	0.000
F1	0.000				0.000	0.000
Thresholds:						
Item1 t1	-0.772	0.200	-3.864	0.000	-0.772	-0.772
Item1 t2	-0.412	0.185	-2.233	0.026	-0.412	-0.412
Item1 t3	0.100	0.179	0.560	0.576	0.100	0.100
Item1 t4	0.842	0.204	4.123	0.000	0.842	0.842
Item2 t1	-0.772	0.200	-3.864	0.000	-0.772	-0.772
Item2 t2	-0.050	0.179	-0.280	0.779	-0.050	-0.050
Item2 t3	0.412	0.185	2.233	0.026	0.412	0.412
Item2 t4	0.915	0.209	4.376	0.000	0.915	0.915
Item3 t1	-0.583	0.191	-3.059	0.002	-0.583	-0.583
Item3 t2	0.151	0.180	0.840	0.401	0.151	0.151
Item3 t3	0.583	0.191	3.059	0.002	0.583	0.583
Item3 t4	0.842	0.204	4.123	0.000	0.842	0.842
Item4 t1	-0.412	0.185	-2.233	0.026	-0.412	-0.412
Item4 t2	0.202	0.180	1.119	0.263	0.202	0.202
Item4 t3	0.583	0.191	3.059	0.002	0.583	0.583
Item4 t4	0.842	0.204	4.123	0.000	0.842	0.842

Modification indices (partial outputs):

Modification Indices:									
	lhs	op	rhs	mi	mi.scaled	epc	sepc.lv	sepc.all	sepc.nox
1	F1	=~	Item1	0.000	0.000	0.000	0.000	0.000	0.000
2	F1	=~	Item2	0.000	0.000	0.000	0.000	0.000	0.000
3	F1	=~	Item3	0.000	0.000	0.000	0.000	0.000	0.000
4	F1	=~	Item4	0.000	0.000	0.000	0.000	0.000	0.000
5	F1	=~	Item5	0.000	0.000	0.000	0.000	0.000	0.000
6	F1	=~	Item6	0.000	0.000	0.000	0.000	0.000	0.000
7	Item1		t1	0.000	0.000	0.000	0.000	0.000	0.000
8	Item1		t2	0.000	0.000	0.000	0.000	0.000	0.000
9	Item1		t3	0.000	0.000	0.000	0.000	0.000	0.000
10	Item1		t4	0.000	0.000	0.000	0.000	0.000	0.000
11	Item2		t1	0.000	0.000	0.000	0.000	0.000	0.000
12	Item2		t2	0.000	0.000	0.000	0.000	0.000	0.000
13	Item2		t3	0.000	0.000	0.000	0.000	0.000	0.000
14	Item2		t4	0.000	0.000	0.000	0.000	0.000	0.000
15	Item3		t1	0.000	0.000	0.000	0.000	0.000	0.000
16	Item3		t2	0.000	0.000	0.000	0.000	0.000	0.000
17	Item3		t3	0.000	0.000	0.000	0.000	0.000	0.000
18	Item3		t4	0.000	0.000	0.000	0.000	0.000	0.000
19	Item4		t1	0.000	0.000	0.000	0.000	0.000	0.000
20	Item4		t2	0.000	0.000	0.000	0.000	0.000	0.000
21	Item4		t3	0.000	0.000	0.000	0.000	0.000	0.000
22	Item4		t4	0.000	0.000	0.000	0.000	0.000	0.000
23	Item5		t1	0.000	0.000	0.000	0.000	0.000	0.000
24	Item5		t2	0.000	0.000	0.000	0.000	0.000	0.000
25	Item5		t3	0.000	0.000	0.000	0.000	0.000	0.000
26	Item5		t4	0.000	0.000	0.000	0.000	0.000	0.000
27	Item6		t1	0.000	0.000	0.000	0.000	0.000	0.000
28	Item6		t2	0.000	0.000	0.000	0.000	0.000	0.000
29	Item6		t3	0.000	0.000	0.000	0.000	0.000	0.000
30	Item6		t4	0.000	0.000	0.000	0.000	0.000	0.000

Entire reliability and Cronbach's alpha:

```

[[1]]
[1] "Entire Reliability: 0.836"

[[2]]
[1] "Cronbach's Alpha: 0.72"

```

Potential warnings:

For this example, the warnings simply let the users know that there were empty cells in the cross tabs produced by the 6 items.

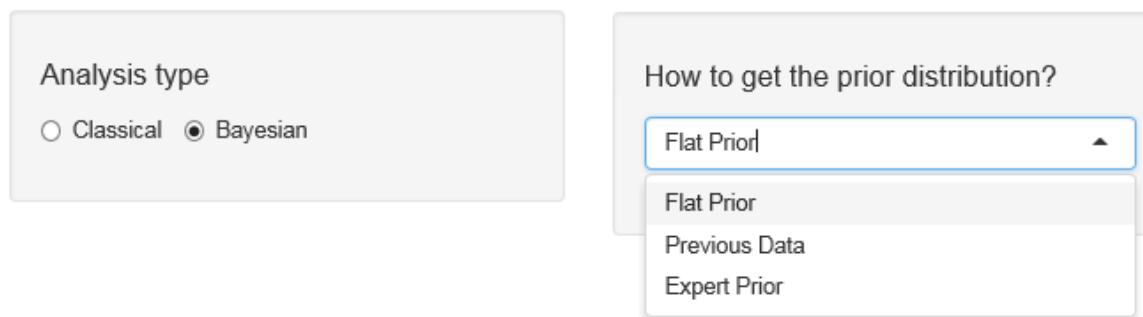
```
[[3]]
[1] "Potential warnings printed below: "

[[4]]
[1] "lavaan WARNING: empty cell(s) in bivariate table of Item2 x Item1"
[2] "lavaan WARNING: empty cell(s) in bivariate table of Item3 x Item1"
[3] "lavaan WARNING: empty cell(s) in bivariate table of Item4 x Item1"
[4] "lavaan WARNING: empty cell(s) in bivariate table of Item5 x Item1"
[5] "lavaan WARNING: empty cell(s) in bivariate table of Item6 x Item1"
[6] "lavaan WARNING: empty cell(s) in bivariate table of Item3 x Item2"
[7] "lavaan WARNING: empty cell(s) in bivariate table of Item4 x Item2"
[8] "lavaan WARNING: empty cell(s) in bivariate table of Item5 x Item2"
[9] "lavaan WARNING: empty cell(s) in bivariate table of Item6 x Item2"
[10] "lavaan WARNING: empty cell(s) in bivariate table of Item4 x Item3"
[11] "lavaan WARNING: empty cell(s) in bivariate table of Item5 x Item3"
[12] "lavaan WARNING: empty cell(s) in bivariate table of Item6 x Item3"
[13] "lavaan WARNING: empty cell(s) in bivariate table of Item5 x Item4"
[14] "lavaan WARNING: empty cell(s) in bivariate table of Item6 x Item4"
[15] "lavaan WARNING: empty cell(s) in bivariate table of Item6 x Item5"
```

Bayesian Approach

Step 4: Select the source of prior

We briefly mentioned in Tutorial 1 that Selecting *Bayesian* approach will provide three options: flat prior, previous data, and expert prior. For this example, we will go through demonstrations of using the *Flat Prior*, *Previous Data*, and *Expert Prior* options.



Flat prior—non-informative normal priors are pre-specified within CBID and users are not required to specify their own flat priors.

Step 5: Select the number of factors

Users need to specify the number of factors or the number of constructs of interest. For this example, we select a one-factor model. Multi-factor models are currently under construction within CBID.



As previously mentioned, CBID extracts the first row of the participants' data file as the variable names used in analysis. By selecting the number of factors in the model, CBID automatically displays the following option for users to specify the items that load on the single factor. Suppose all 6 items load on the single factor, labeled generically by CBID as *Factor 1*.

Factor 1 items

Item1 Item2 Item3 Item4 Item5 Item6

Step 6: Run the analysis

After all options have been appropriately selected, we can now run the entire analysis by clicking on the Go button.



Step 7: Review the outputs

The Bayesian analysis will only output relevant results, such as the Monte Carlo Markov chain (MCMC) procedure acceptance rates, the factor loading estimates and standard errors (labeled as *Estimate* and *Std*), the 95% credible interval for the factor loadings, the entire reliability estimate (along with standard error and 95% credible interval), and any potential warning messages. Desirable MCMC acceptance rates should be between 20-50%.

Summary:

```

Acceptance rates:
  Item1 Item2 Item3 Item4 Item5 Item6
  0.21  0.37  0.28  0.28  0.25  0.23
[[1]]
[[1]][[1]]
  Estimate Std 2.5% CI 97.5% CI
Item1   0.269 0.139  0.026   0.551
Item2   0.883 0.064  0.731   0.968
Item3   0.742 0.102  0.507   0.905
Item4   0.696 0.112  0.440   0.873
Item5   0.461 0.140  0.158   0.705
Item6   0.515 0.136  0.216   0.746

[[1]][[2]]
  Entire Reliability Std 2.5% CI 97.5% CI
[1,]           0.884 0.041  0.793   0.95

[[2]]
[1] "No warnings to report"

```

Previous data—users are required to upload a separate prior data file, which contains data collected from previous participants.

Below is a screen shot of the sample previous data file (*priordata.csv*) that we will use. For example, we have reliable and relevant data collected previously from 100 participants. Notice that the item names of the previously collected data file matches item names from the current participants' data file. The number of columns also matches in both data files.

	A	B	C	D	E	F	
1	Item1	Item2	Item3	Item4	Item5	Item6	
2	3	3	3	3	3	3	
3	5	1	4	4	3	3	
4	5	4	5	5	5	4	
5	3	3	3	3	3	3	
6	5	3	4	4	5	3	
7	5	4	4	4	4	4	

Step 5: Upload previous data file

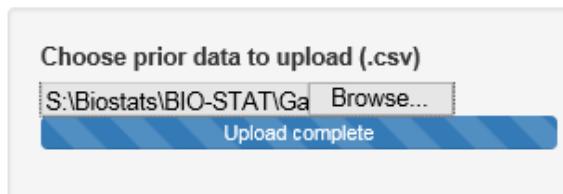
Prior to uploading *priordata.csv* to CBID, users will see the following option:

Analysis type
 Classical Bayesian

How to get the prior distribution?
 Previous Data ▾

Choose prior data to upload (.csv)

Click on *Browse* and a new window will open for users to locate the data file used for analysis. Once the data file is selected, click *Open* in the new window and CBID will display a blue bar to indicate a successful upload (shown below).



Step 6: Select the number of factors

Users need to specify the number of factors or the number of constructs of interest. For this example, we select a one-factor model.

Number of Factors
 1

As previously mentioned, CBID extracts the first row of the participants' data file as the variable names used in analysis. By selecting the number of factors in the model, CBID automatically displays the following option for users to specify the items that load on the single factor. Suppose all 6 items load on the single factor, labeled generically by CBID as *Factor 1*.

Factor 1 items
 Item1 Item2 Item3 Item4 Item5 Item6

Step 7: Run the analysis

After all options have been appropriately selected, we can now run the entire analysis by clicking on the *Go* button.



Step 8: Review the outputs

The Bayesian analysis will only output relevant results, such as the Monte Carlo Markov chain (MCMC) procedure acceptance rates, the factor loading estimates and standard errors (labeled as *Estimate* and *Std*), the 95% credible interval for the factor loadings, the entire reliability estimate (along with standard error and 95% credible interval), and any potential warning messages. Two sets of MCMC acceptance rates will be displayed for the *Previous data* option, one set each for the previously collected participants' data and the current participants' data.

Desirable MCMC acceptance rates should be between 20-50% for each item. However, it is still acceptable if the majority of the items have acceptable rates within the range, with the rest of the items falling slightly outside the range (as shown in our example below).

```
Summary:

Acceptance rates:
Item1 Item2 Item3 Item4 Item5 Item6
0.43 0.35 0.57 0.54 0.46 0.36

Acceptance rates:
Item1 Item2 Item3 Item4 Item5 Item6
0.22 0.27 0.3 0.31 0.25 0.24
[[1]]
[[1]][[1]]
  Estimate Std 2.5% CI 97.5% CI
Item1 0.542 0.112 0.292 0.729
Item2 0.723 0.061 0.584 0.819
Item3 0.846 0.066 0.685 0.936
Item4 0.854 0.063 0.703 0.941
Item5 0.626 0.092 0.419 0.776
Item6 0.616 0.089 0.419 0.762

[[1]][[2]]
  Entire Reliability Std 2.5% CI 97.5% CI
[1,] 0.901 0.022 0.853 0.936

[[2]]
[1] "No warnings to report"
```

Expert prior—users are required to upload a separate prior data file, which contains data collected from subject experts.

Below is a screen shot of the sample experts' data file (*expertdata.csv*) that we will use. For example, we consulted a panel of 6 subject experts and asked the experts to rate the relevancy of each item to the domain of interest, on a 4-point relevancy scale. Notice that the item names of the experts' data file matches item names from the participants' data file. The number of columns also matches in both data files.

	A	B	C	D	E	F
1	Item1	Item2	Item3	Item4	Item5	Item6
2	4	4	3	4	4	4
3	4	4	4	4	3	3
4	2	4	3	3	3	3
5	2	4	4	4	4	4
6	1	4	3	4	4	4
7	3	2	2	3	2	3
8						

Step 5: Upload expert prior data file

Prior to uploading *expertdata.csv* to CBID, users will see the following option:

Analysis type

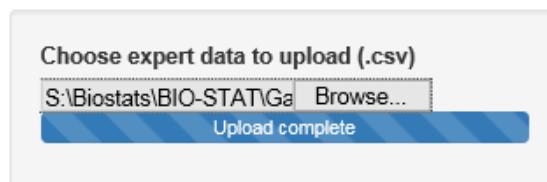
Classical Bayesian

How to get the prior distribution?

Expert Prior

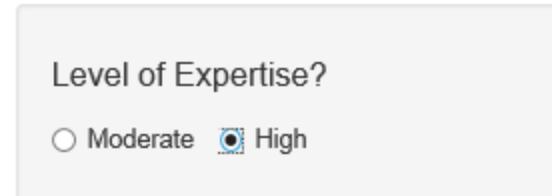
Choose expert data to upload (.csv)

Click on *Browse* and a new window will open for users to locate the data file used for analysis. Once the data file is selected, click *Open* in the new window and CBID will display a blue bar to indicate a successful upload (shown below).



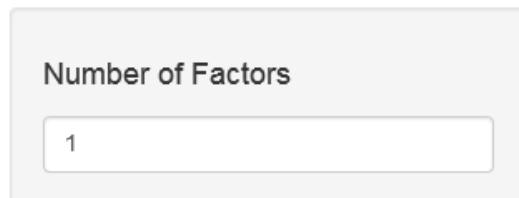
Step 6: Select expert level of expertise

Depending on the expertise level of the experts used for content validity analysis, users can select either *Moderate* or *High* options. If the experts exhibit a high level of expertise in the area of interest, *High* option should be selected. However, for other circumstances, if the panel exhibits a moderate level of expertise in the area, *Moderate* option is more appropriate. For this demonstration, we will assume that the panel consisted of experts that exhibit a high level of expertise in the area.



Step 7: Select the number of factors

Users need to specify the number of factors or the number of constructs of interest. For this example, we select a one-factor model.



As previously mentioned, CBID extracts the first row of the participants' data file as the variable names used in analysis. By selecting the number of factors in the model, CBID automatically displays the following option for users to specify the items that load on the single factor. Suppose all 6 items load on the single factor, labeled generically by CBID as *Factor 1*.

Factor 1 items

- Item1
- Item2
- Item3
- Item4
- Item5
- Item6

Step 8: Run the analysis

After all options have been appropriately selected, we can now run the entire analysis by clicking on the *Go* button.



Step 9: Review the outputs

The outputs are similar to the outputs of using flat prior and previous data, including the Monte Carlo Markov chain (MCMC) procedure acceptance rates, the factor loading estimates and standard errors (labeled as *Estimate* and *Std*), the 95% credible interval for the factor loadings, the entire reliability estimate (along with standard error and 95% credible interval), and any potential warning messages. Desirable MCMC acceptance rates should be between 20-50%.

Summary:

```
Acceptance rates:  
Item1 Item2 Item3 Item4 Item5 Item6  
0.21 0.31 0.26 0.28 0.25 0.25  
[[1]]  
[[1]][[1]]  
Estimate Std 2.5% CI 97.5% CI  
Item1 0.339 0.099 0.130 0.519  
Item2 0.761 0.067 0.607 0.867  
Item3 0.564 0.076 0.400 0.694  
Item4 0.669 0.076 0.497 0.792  
Item5 0.483 0.092 0.286 0.645  
Item6 0.550 0.085 0.363 0.696  
[[1]][[2]]  
Entire Reliability Std 2.5% CI 97.5% CI  
[1,] 0.79 0.034 0.718 0.848  
[[2]]  
[1] "No warnings to report"
```

[Return to Main Menu](#)



Articles

1. Alonso, A., Laenen, A., Molenberghs, G., Geys, H., & Vangeneugden, T. (2010). A Unified Approach to Multi-item Reliability. *Biometrics*, 66(4), 1061-1068.
2. Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. Guilford Publications.
3. Gajewski, B. J., Coffland, V., Boyle, D. K., Bott, M., Price, L. R., Leopold, J., & Dunton, N. (2015). Assessing content validity through correlation and relevance tools. *Methodology*, 8(3), 81-96.
4. Gajewski, B. J., Price, L. R., Coffland, V., Boyle, D. K., & Bott, M. J. (2013). Integrated analysis of content and construct validity of psychometric instruments. *Quality & Quantity*, 47(1), 57-78.
5. Gajewski, B. J., Price, L. R., & Bott, M. J. (2015). Response to “Conceptions of Reliability Revisited, and Practical Recommendations.” *Nursing Research (invited)*, 64(2), 137-139 (PMCID in process).
6. Garrard, L., Price, L. R., Bott, M. J., Gajewski, B. J. (in review). A Novel Method for Expediting the Development of Patient-Reported Outcome Measures and an Evaluation of Its Performance via Simulation.
7. Jiang, Y., Boyle, D. K., Bott, M. J., Wick, J. A., Yu, Q., & Gajewski, B. J. (2014). Expediting clinical and translational research via Bayesian instrument development. *Applied psychological measurement*, 0146621613517165.

[Return to Main Menu](#)