

Using ACER ConQuest program to examine multidimensional and many-facet models

Mahmut Sami Koyuncu^{1,*}, Mehmet Şata²

¹Afyon Kocatepe University, Faculty of Education, Department of Educational Sciences, Afyonkarahisar, Türkiye

²Ağrı İbrahim Çeçen University, Faculty of Education, Department of Educational Sciences, Ağrı, Türkiye

ARTICLE HISTORY

Received: Jan. 17, 2023

Revised: May 25, 2023

Accepted: Jan. 12, 2023

Keywords:

ConQuest,
Multidimensionality,
Many-facet Rasch,
Validity.

Abstract: The main aim of this study was to introduce the ConQuest program, which is used in the analysis of multivariate and multidimensional data structures, and to show its applications on example data structures. To achieve this goal, a basic research approach was applied. Thus, how to use the ConQuest program and how to prepare the data set for analysis were explained step by step. Then, two example applications were made considering the multidimensional structures. Finally, different sources of variability (e.g., item, student, rater, gender), which are both multidimensional and independent of each other, were performed by considering different sources of variability together. According to the analyses, the dimensionality of the data structures must be examined in the analysis process. If the data structure is multidimensional, appropriate multidimensional IRT analyses should be performed.

1. INTRODUCTION

1.1. What is ACER ConQuest?

The ACER ConQuest (Adams et al., 2020) program is developed at the Australian Council for Educational Research (ACER) and the University of California, Berkeley. It is a paid statistical package program that can examine the fit of item response and latent regression models, including multidimensional item response models, in a single program. It also provides the integration of item responses and regression analysis (Adams et al., 2022).

The ConQuest can run the analyses of the following models (Adams et al., 2022): Rasch Simple Logistic Model (Rasch, 1980), Rating Scale Model (Andrich, 1978), Partial Credit Model (Masters, 1982), Ordered Partition Model (Wilson, 1992), Linear Logistic Test Model (Fischer, 1983), Many-facet Models (Linacre, 1994), Generalized Unidimensional Models, Multidimensional Item Response Models (Adams et al., 1997; Wang, 1995), and Latent Regression Models (Adams et al., 1997).

Joint maximum likelihood (JML) or marginal maximum likelihood (MML) estimates can be generated by ACER ConQuest generates for the parameters of the specified models. The MML estimation algorithms used are;

*CONTACT: Mahmut Sami Koyuncu ✉ koyuncumsami@gmail.com 📍 Afyon Kocatepe University, Faculty of Education, Department of Educational Sciences, Afyonkarahisar, Türkiye

- Gauss-Hermite quadrature (Volodin & Adams, 1995).
- Bock/Aitken quadrature (Bock & Aitkin, 1981)
- Monte Carlo (Volodin & Adams, 1995).
- Markov chain Monte Carlo (Patz & Junker, 1999).

The Gauss method is generally used for three or less-dimensional problems, while the Monte Carlo method is preferred for more than three-dimensional problems. Moreover, the Gauss method cannot be used when the distribution is discrete or there is no independent variable to estimate the dependent variable in the regression model. Thus, when there is a regression variable in the Conquest program, the Quadrature method is used as the default method. Otherwise, the Gauss method is used. If there is missing data in all items in a dimension, the JML method cannot be used. In addition, estimating item parameters is not possible when the JML method is used (Adams et al., 2022).

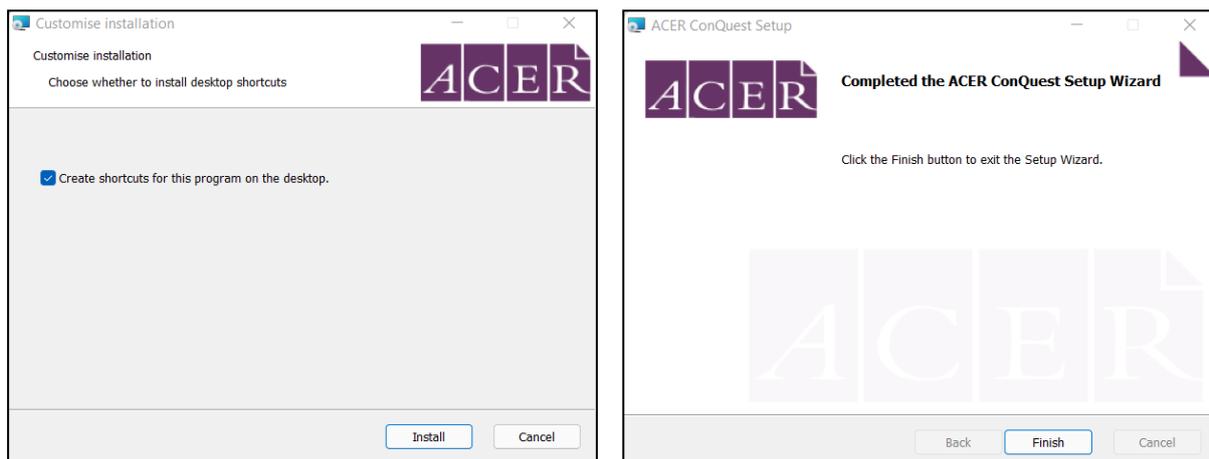
Through the ConQuest program, the following applications can be performed: item analysis (IRT and Traditional), DIF, Exploring Rater Effects, Latent correlation estimation and Estimating Latent Correlations and Testing Dimensionality, and Drawing Plausible Values (Adams et al., 2022). ACER ConQuest can model up to 50 different facets and analyze item clusters designed to produce measurements of up to 30 latent dimensions.

1.2. Installation and User Interfaces of ACER ConQuest Program

The program can be used in both Windows and Mac OS operating systems. For the Windows operating system, the program has both GUI (graphical user interface) and CMD (console interface) console versions. However, for the Mac OS system, only the console version is available to use. The GUI version is more user-friendly and has drawing functions that the console version does not have. However, especially for larger and more complex analyses, the console version, which works faster than the GUI version, may be preferred.

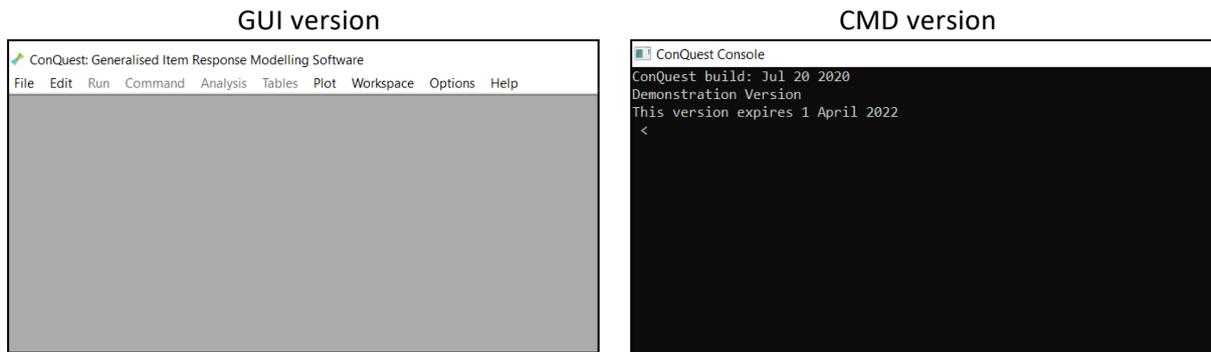
The ConQuest program has a 1-month free DEMO trial for users to experience. However, in the demo version, the sample size is limited to 3000 and the number of items is limited to 100. The installation of the program is simple. Figure 1 shows the program setup screens for the Windows operating system.

Figure 1. *ConQuest Windows setup screen.*



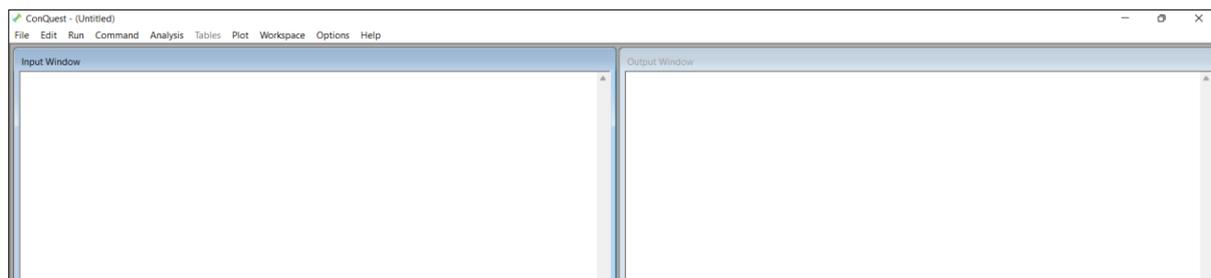
Once the program is installed on the PC, both GUI and CMD versions become ready to use. Figure 2 shows the start screens of the GUI and CMD versions of the ConQuest program.

Figure 2. Start screens of GUI and CMD versions.



The GUI version includes menus such as *File*, *Edit*, *Run*, *Command*, *Analysis*, *Tables*, *Plot*, *Workspace*, *Options*, and *Help*. With the *New* command in the *File* menu, a new working screen including both input and output windows is opened (Figure 3). Analysis using a command line (i.e., CMD version) can be performed in one step. when running from a command-line interpreter is to provide the command file as a command-line argument. In this demonstration, the GUI version was used.

Figure 3. ConQuest Input and Output screen.



The codes (i.e., syntax) required for analysis are entered in the *Input Window*. The analysis is performed by running the codes (i.e., syntax) with the help of the *Run* menu. The analyses performed are displayed in the *Output Window* section. Besides, the analysis results specially requested by the researcher can be saved as *.txt* files with the help of the syntaxes entered in the code file. For more detailed information, the ACER ConQuest Manual can be applied (Adams et al., 2022).

In educational and psychological research, measurement tools are the main data collection sources. In educational sciences, the measurements are done indirectly, so it is important to provide evidence regarding the validity and reliability of the measures (Köse, 2012). The selection of the analysis method and package program appropriate for the nature of the data are the factors that contribute to the reliability and validity of the scores obtained from the measurement tools.

Regarding the nature of the data used for educational purposes, the concept of dimensionality is important (Finch & Habing, 2003; Mroch & Bolt, 2006; Özbek-Baştuğ, 2012; Özer-Özkan & Acar-Güvendir, 2014). While providing evidence for the reliability and validity of the scores obtained from the measurement tools, determining the dimensionality of the data or the number of dimensions/factors will contribute to the reliability and validity. According to Messick (1995), there are two threats to validity: construct underrepresentation and construct-irrelevant variance. An accurate definition of the studied data set in terms of dimensionality will directly contribute to the validity of the measurements. This is due to the concern of underrepresenting the structure intended to be measured is eliminated (Messick, 1995).

Regarding the historical development of measurement theories, the classical test theory, which was founded on the assumption of total score or unidimensionality, was first put forward. Then, the one-dimensional item response theory emerged. Therefore, these models investigate one-dimensional constructs and variables. This means that the unidimensionality of the structures planned to be measured must be tested (Özbek-Baştuğ, 2012).

Various methods determine the dimensionality of the constructs or the number of dimensions. These methods are either parametric or non-parametric. Research has compared these methods with each other (Mroch & Bolt, 2006) to identify the most effective method (Stout et al., 2001). In addition to, some studies have focused only on dimensionality analysis (Jang & Roussos, 2007).

In studies on dimensionality, item (individual) and ability parameters were negatively affected because multidimensional structures were analyzed as one-dimensional structures (Özer-Özkan, 2012). Regarding the structures of the measurement tools used in the measurement of cognitive skills, it is difficult to provide the unidimensionality assumption. Considering that many skills are used together in the measurement of high-level cognitive skills, it confronts us with the fact that the unidimensionality assumption will not be met. This situation requires multidimensional analysis or modeling of the measurement tool (Ackerman, 1994).

Multidimensional modeling and analysis methods attract more and more attention day by day, as they eliminate the limitations of one-dimensional measurement models and offer models that are more suitable for real-life situations. Due to the increasing need for measuring multidimensional structures, many statistical package programs have been developed recently (Köse, 2012). These programs include IRTPRO, MULTILOG, BILOG, MIRTE, TESTFACT, PARSCALE, Xcalibre, and R package programs (i.e., eRm, pl. rasch), flexMIRT, BMIRT, and NOHARM. Almost all these programs can analyze both one-dimensional and multidimensional measurement models. Besides, the ConQuest package program can analyze both multidimensional measurement models and multivariate measurements at the same time and also allows the interactions between variables to be examined at the same time.

ACER Conquest program is frequently used in studies in many different fields recently. Its more widespread use is preferred especially in IRT model analysis, mostly in Rasch model estimations (Brnic & Greefrath, 2021; Hahn & Kähler, 2022; Jolin & Wilson, 2022; Jüttler & Schumann, 2022; Krell et al., 2022; Koch et al., 2022; Lou et al., 2022; Mischo et al., 2022; Oko, 2022; Osterhaus et al., 2022, Spink et al., 2022; Unfried et al., 2022; Wall et al., 2022). Besides, the use of the Conquest program is preferred in studies where multidimensional structures are examined or many-facet models are used (Bartolomé & Garaizar, 2022; Mendoza et al., 2022; Wang et al., 2022; Zhai, 2022).

2. METHOD

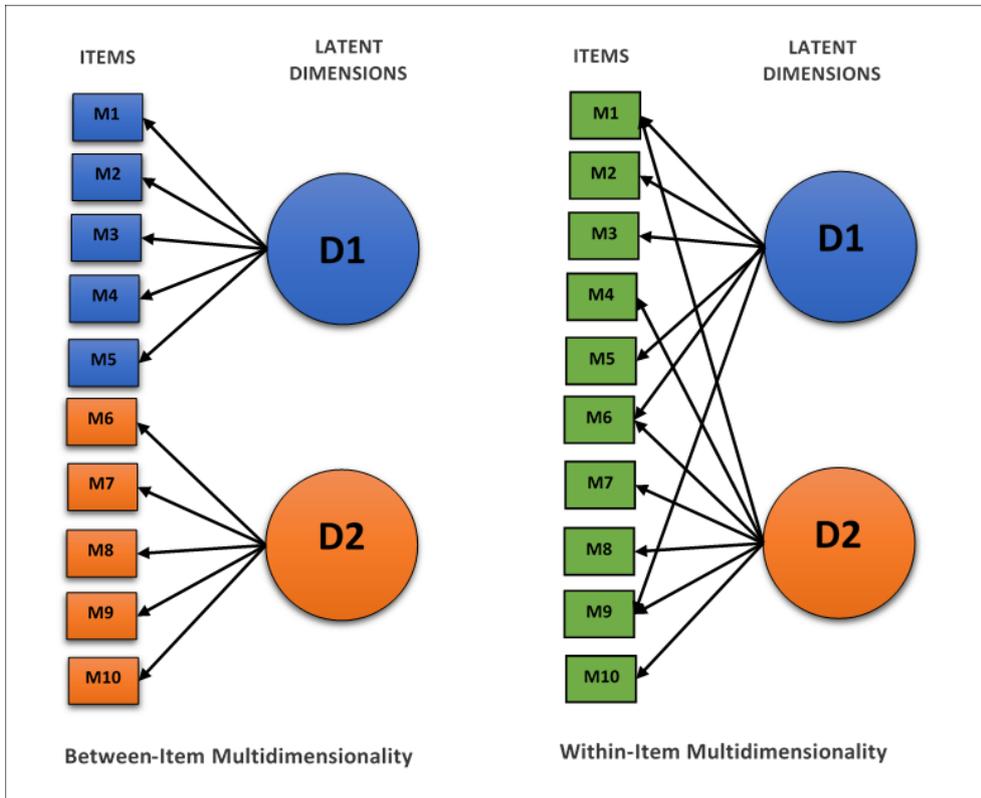
2.1. Application of The Conquest Program

This research aimed to introduce the ConQuest program, applying multidimensional models on example data sets. Therefore, firstly, the general features of the ConQuest program, its installation, and the analysis process were explained. Then, multidimensional model applications were carried out on example data sets. In addition, example syntaxes appropriate for multidimensional model analysis were created to benefit the researchers. Especially considering that many structures are multidimensional by nature, this research is important in terms of eliminating the lack of multidimensional models in the literature.

By creating synthetic data on multidimensional models, three different examples scenarios were presented within the scope of the study in order to guide the researchers. Analyses were performed via the ConQuest GUI Demo version (5.12.3). The first example application belongs to between-item multidimensional models, and the second example application belongs to

within-item multidimensional models. If a test consists of several one-dimensional subscales, it is Between-Item Multidimensionality. If any of the items are related to more than one latent dimension, this test is considered as Within-Item Multidimensionality (Adams et al., 1997; Wang, 1995). The structure of the synthetic data of the Between-Item and Within-Item multidimensional model created within the scope of the study was presented in Figure 4.

Figure 4. Within-Item and Between-Item Multidimensionality.



Also, an example of many-facet multidimensional models was presented as a third example application within the scope of the study.

2.2. Example 1: Between-Item Multidimensional Model

It is assumed that the data structure created for the Between-Item Multidimensional Model example consists of 10 Likert-type items scored from 1 to 5. As shown in Figure 4, items from 1 to 5 represent the first dimension of the scale and items from 6 to 10 represent the second dimension. The data used in the research consists of hypothetical data. The main reason for this is that the program is intended to be implemented and to guide researchers. This situation was taken into consideration as the limitation of the research. The data structure of 50 individuals was generated and the data file with the .txt extension required for analysis was prepared. Then, the command with the .cqc extension was created for analysis. It is important to prepare the data for analysis, and the structure in the data file must be defined in the script with the necessary syntaxes. In addition, the labels of the variables in the data file can be created in a separate file with a .txt extension to make the analysis outputs more understandable. Table 1 contains the command, data, and tag file examples created by the researchers for the Between-Item Multidimensional Model.

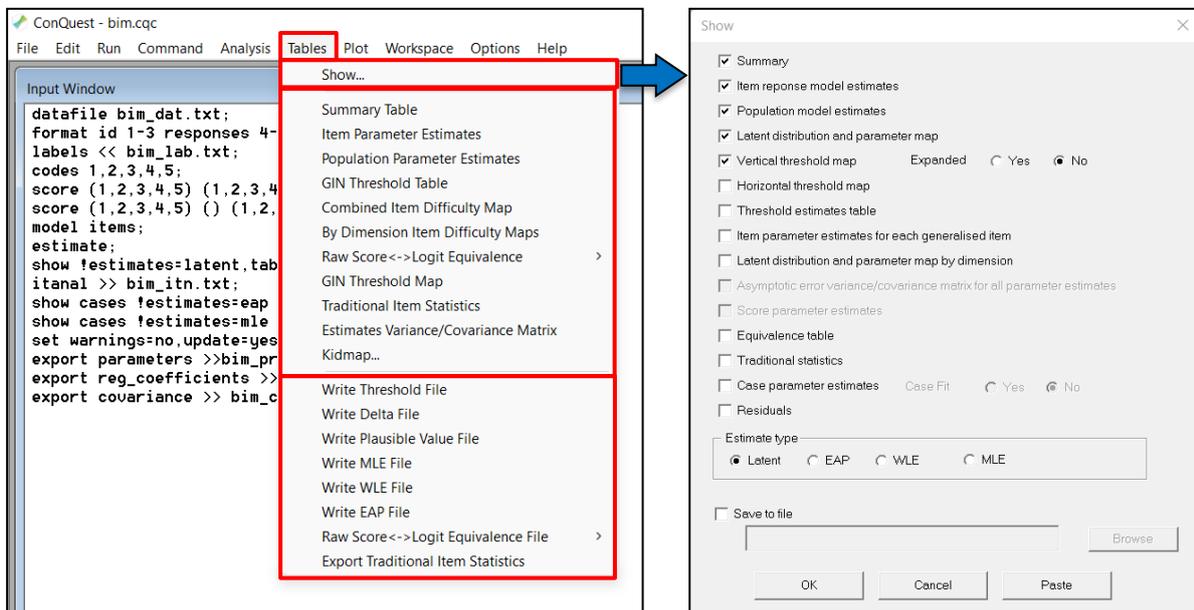
Table 1. Example script, data, and tag files of the Between-Item Multidimensional Model.

bim.cqc	bim_dat.txt	bim_lab.txt
Command statements (required)	Dataset (required)	Dataset label (optional)
<pre>datafile bim_dat.txt; format id 1-3 responses 4-13; labels << bim_lab.txt; codes 1,2,3,4,5; score (1,2,3,4,5) (1,2,3,4,5) () ! items(1-5); score (1,2,3,4,5) () (1,2,3,4,5) ! items(6-10); model items; estimate; show !estimates=latent,tables=1:2:3:9>> bim_shw.txt; itanal >> bim_itn.txt; show cases !estimates=eap >> bim_eap.txt; show cases !estimates=mle >> bim_mle.txt;</pre>	<pre>0014555555551 0025555515555 0032553141311 0045555555555 0055555455455 0065455511451 0075555445555 0085555553551 0095555555555 0105555455555 0115555354551</pre>	<pre>===> item 1 M1 2 M2 3 M3 4 M4 5 M5 6 M6 7 M7 8 M8 9 M9 10 M10</pre>

Note. The first set of parentheses contains a set of codes (the codes list). The second set of parentheses contains a set of scores on dimension one for each of those codes (a score list). The third set contains a set of scores on dimension two (a second score list) and so on. The number of separate codes in the codes list indicates the number of response categories that will be modeled for each item. The number of score lists indicate the number of dimensions in the model. The codes and scores in the lists can be comma-delimited or space-delimited.

To perform the analysis, the command file is opened and *run* in the ConQuest program. The necessary analysis results can be added to the command file as *.txt* in the ConQuest program or created with the help of the *Tables* menus in the program after the command is run. Figure 5 displays the statistics and output files that can be created with the help of the *Show...* tab in the *Tables* menu.

Figure 5. ConQuest Tables menu and Show tab.



Rasch Analysis results for application example-1 (*bim_shw.txt*) and traditional item analysis results (*bim_itn.txt*) are presented below.

Rasch Analysis results (*bim_shw.txt*) file includes a summary of the estimation, item parameter estimates, regression coefficients, item parameter estimates for each term in the model (in this example there is only one term: item), covariance/correlation matrix, reliability coefficients, map of latent distributions, and response model parameter estimates, respectively. Table 2 presents the summary of the estimation and item parameter estimates output of the example

application. Table 3 contains the universe model parameter estimations including the output of regression coefficients, covariance/correlation matrix, reliability coefficients, and brief explanations.

Table 2. Summary of the estimation and item parameter estimates output.

Summary of the estimation	Explanation
<pre> Estimation method was: Gauss-Hermite Quadrature with 225 nodes No node filtering Xsi increment max: 1.00000 FacOldXsi: 0.00000 Assumed population distribution was: Gaussian Location constraint was: DEFAULT Scale constraint was: Not applicable The Data File: bim_dat.txt The format: id 1-3 responses 4-13 No case weights The regression model: Grouping Variables: The item model: items Slopes are fixed Cases in file: 51 Cases in estimation: 50 Final Deviance: 680.44644 Akaike Information Criterion (AIC): 706.44644 Akaike Information Criterion Corrected (AICc): 701.41419 Bayesian Information Criterion (BIC): 731.30274 Total number of estimated parameters: 13 The number of iterations: 52 Termination criteria: Max iterations=1000, Parameter Change= 0.00010 Deviance Change= 0.00010 Iterations terminated because the deviance convergence criteria was reached Random number generation seed: 1.00000 Number of nodes used when drawing PVs: 2000 Number of nodes used when computing fit: 200 Number of plausible values to draw: 5 Maximum number of iterations without a deviance improvement: 100 Maximum number of Newton steps for each parameter in M-step: 10 Value for obtaining finite MLEs for zero/perfects: 0.30000 </pre>	<p>In the summary of the estimation results, there is information such as the analysed data file, the format in the data file, the desired model, and example size. Besides, Deviance, AIC, AICc and BIC values used to evaluate model fit are also included. Regarding the relative fit, (for example, which model fits better compared to more than one model), the smaller value fits the data better (Chen et al., 2013; De Ayala, 2009, p. 41).</p>
<pre> TERM 1: items ----- VARIABLES ----- item ESTIMATE ERROR^ MNSQ CI T MNSQ CI T ----- 1 M1 0.207 0.190 0.79 (0.61, 1.39) -1.1 0.85 (0.36, 1.64) -0.5 2 M2 -0.349 0.232 0.68 (0.61, 1.39) -1.8 0.97 (0.17, 1.83) -0.0 3 M3 -0.284 0.226 0.45 (0.61, 1.39) -3.4 0.73 (0.20, 1.80) -0.7 4 M4 -0.395 0.287 0.85 (0.61, 1.39) -0.7 1.22 (0.30, 1.70) 0.7 5 M5 -0.820* 0.173 1.08 (0.61, 1.39) 0.5 0.72 (0.50, 1.50) -1.2 6 M6 0.105 0.209 2.06 (0.61, 1.39) 4.1 1.77 (0.27, 1.73) 1.8 7 M7 0.105 0.209 0.85 (0.61, 1.39) -0.7 0.99 (0.27, 1.73) 0.1 8 M8 0.201 0.204 0.97 (0.61, 1.39) -0.1 1.52 (0.29, 1.71) 1.3 9 M9 -1.286 0.349 0.30 (0.61, 1.39) -4.9 2.13 (0.00, 2.29) 1.4 10 M10 0.075* 0.189 0.95 (0.61, 1.39) -0.2 1.15 (0.47, 1.53) 0.5 </pre> <p>An asterisk next to a parameter estimate indicates that it is constrained</p> <p>Separation Reliability = 0.764</p> <p>Chi-square test of parameter equality = 21.98, df = 8, Sig Level = 0.005</p> <p>* Empirical standard errors have been used</p> <p>Term is a fixed effect</p>	<p>There are item parameter estimations for each term in the model. The example presents only one term (<i>item</i>). Since the example has 10 items in total, there are some concordance statistics including the estimations and standard errors for each item.</p>

According to the item parameter estimations of the Rasch model in Table 2, the Mean-Square (MNSQ) value and the confidence interval of this value are included as the fit index. If the calculated MNSQ fit value is outside the expected confidence interval, the T statistic corresponding to the MNSQ value will exceed the |2.0|, meaning that the item does not fit the model well (Adams et al., 2022).

The parameter estimates in this table are for the difficulties of each of the items. For the purposes of model identification, ACER ConQuest constrains the difficulty estimate for the last item to ensure an average difficulty of zero. This constraint has been achieved by setting the difficulty of the last item to be the negative sum of the previous items. The fact that this item is constrained is indicated by the asterisk (*) placed next to the parameter estimate.

At the bottom of the item parameter output of the Rasch model, there are Separation Reliability and Chi-square values. Separation reliability (Wright & Stone, 1979) takes a value between 0 and 1 and is an index of equality of parameters. It provides information on how well its parameters are separated. However, it may not be useful to examine the chi-square value in all cases; it will be more useful to examine the significance of parameter equality, especially in cases such as rater severity (Adams et al., 2022).

Table 3. Population Model Parameter Estimates.

Regression coefficients	Explanation
<pre> REGRESSION COEFFICIENTS ----- Dimension Regression Variable Dimension_1 Dimension_2 CONSTANT 1.931 (0.233) 2.132 (0.361) An asterisk next to a parameter estimate indicates that it is constrained </pre>	<p>These are average ability estimates in each dimension.</p>
<pre> UNCONDITIONAL COVARIANCE/CORRELATION MATRIX ----- Dimension Dimension 1 2 Dimension_1 Dimension_2 0.665 0.920 Variance 1.147 (0.543) 1.671 (0.584) An asterisk next to a parameter estimate indicates that it is constrained Values below the diagonal are correlations and values above are covariances </pre>	<p>Correlation values (0.665) and covariance values (0.920) between dimensions are included. Also, there are the estimated variance values for the two dimensions</p>
<pre> RELIABILITY COEFFICIENTS ----- Dimension: (Dimension_1) MLE Person separation RELIABILITY: Unavailable WLE Person separation RELIABILITY: Unavailable EAP/PV RELIABILITY: 0.651 Dimension: (Dimension_2) MLE Person separation RELIABILITY: Unavailable WLE Person separation RELIABILITY: Unavailable EAP/PV RELIABILITY: 0.667 </pre>	<p>There are reliability values for each dimension. There are three different reliability values. Since the calculation is not made according to the maximum likelihood estimation, only the EAP/PV value is displayed, while the other values are not.</p>

Besides, item difficulty map output, which provides the opportunity to examine latent ability estimation and item difficulty on the same scale, and which can be obtained separately or in combination according to the dimensions, can also be created. Figure 6 displays the output of the latent distribution map and response model parameter estimations, and Figure 7 shows the output of generalized item threshold tables and maps.

Figure 6. Map of latent distributions and response model parameter estimates output.

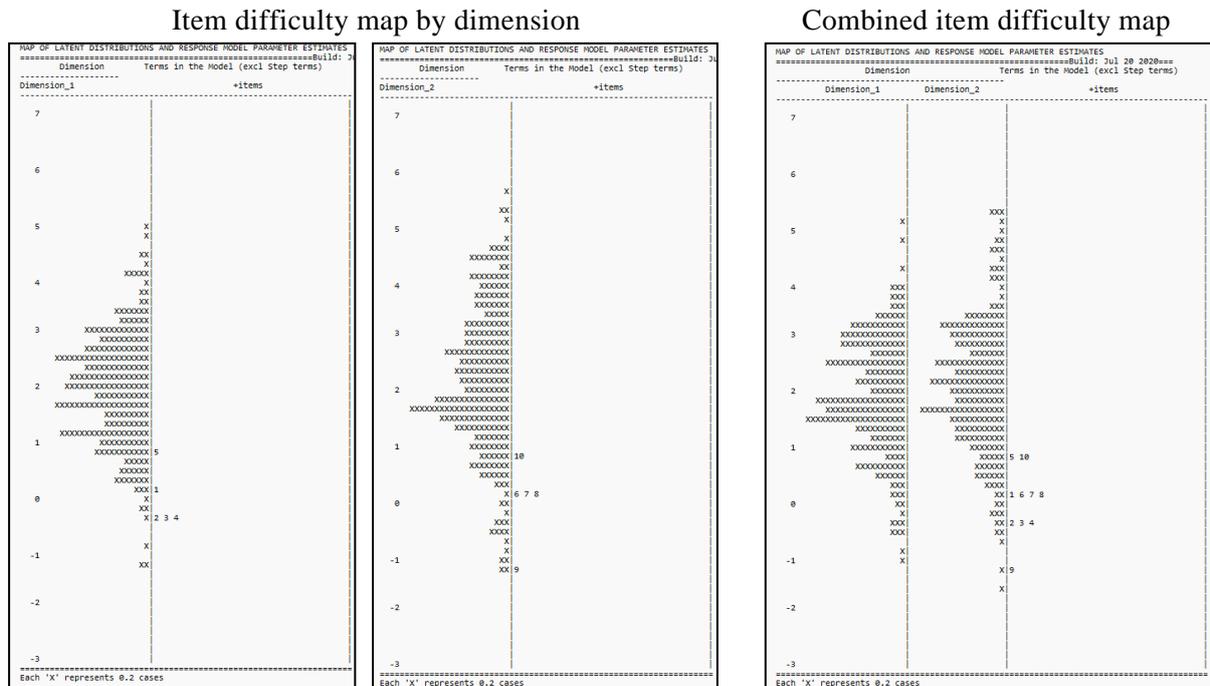
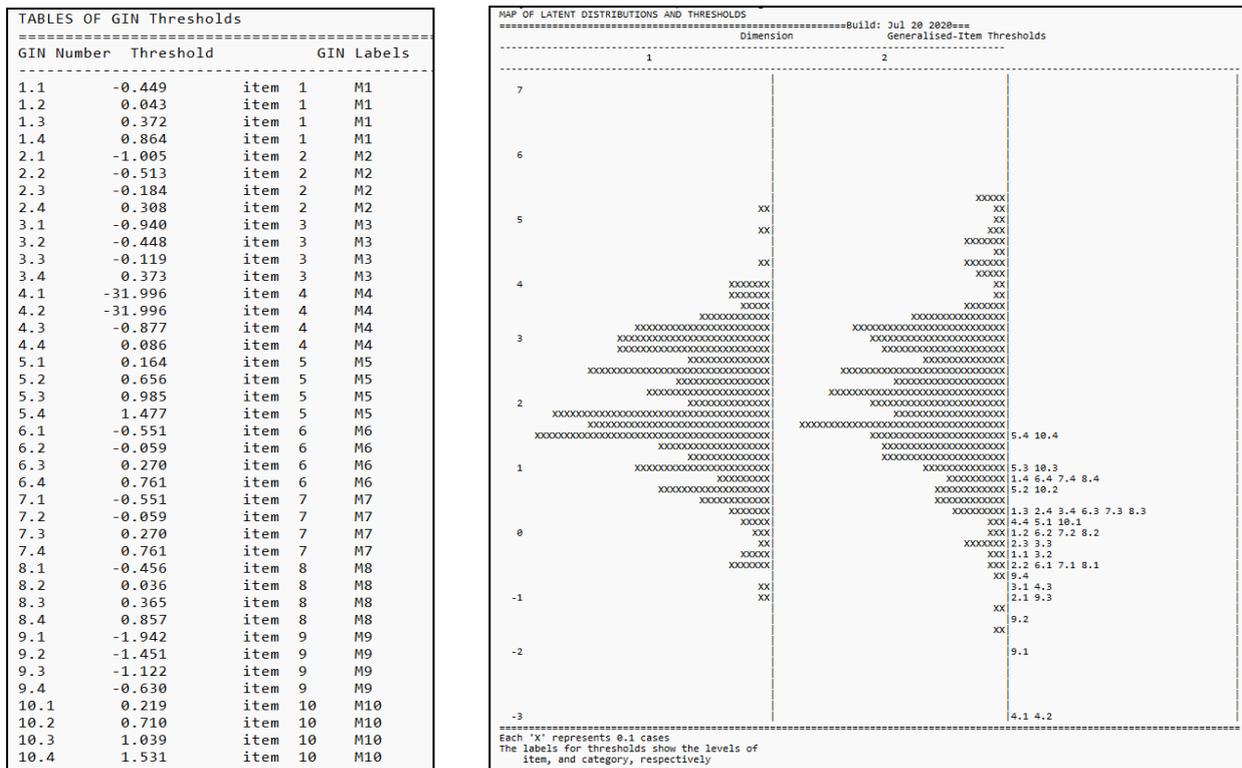


Figure 6 shows that item difficulty maps can be created separately or combined as a single output according to dimensions. In the first column, the latent ability estimation distribution of the individuals, and in the second column, the difficulty estimations of the items are located. This distribution map allows for each independent variable to be interpreted as a dependent variable by placing many sources of variability such as student, item, size, rater, and time on the same scale (logit scale) (Esfandiari, 2015). The ability estimations of the individuals in the example data set vary between -1 and +5 logit in the first dimension, they vary between -1 and +6 logit in the second dimension, and the group generally has a high ability in both dimensions. Regarding the figure in which both dimensions are combined, individuals exhibit a slightly higher ability in the second dimension. On the other hand, the difficulty levels of the items indicate that the most difficult items are item number five in the first sub-dimension and item number 10 in the second sub-dimension. According to the variable map, in which both dimensions were combined, was examined, item 9 was found the easiest. In addition, items 2, 3, and 4 had similar difficulty levels, and items 1, 6, 7, and 8 were similar to each other.

The combined item difficulty map of the first and second dimensions does not mean that they are on the same scale and comparable. The researcher should pay attention to this issue while reporting his results.

Figure 7. Generalized item thresholds table and map.



In Figure 7, there is a map that includes the thresholds (Thurstonian) estimation values for each item and the latent talent estimations of these values along with the dimensions. The example application has 4 thresholds since the Likert-type items scored from 1 to 5. Among the values in the rightmost column, the first term indicates the item number, and the second term indicates the threshold value. For example, 5.4 represents the 4th threshold of the 5th item.

The traditional item analysis result (*bim_itn.txt*) file involves statistics and test statistics for each item. In Figure 8, there are only results for item 1, and Figure 9 shows the test statistics results.

Figure 8. Statistics related to item 1.

```

=====
Item 1
-----
item:1 (M1)
Cases for this item      50  Item-Rest Cor.  0.70  Item-Total Cor.  0.78
Item Threshold(s):     -0.45  0.04  0.37  0.86  Weighted MNSQ   0.85
Item Delta(s):         0.21  0.21  0.21  0.21
-----
Label   Score   Count  % of tot  Pt Bis    t (p)   PV1Avg:1 PV1 SD:1  PV1Avg:2 PV1 SD:2
-----
  1     1.00     2     4.00    -0.62   -5.41(.000)  0.295  0.701    0.074  0.733
  2     2.00     1     2.00    -0.46   -3.57(.001)  0.067  0      -0.637  0
  3     3.00     4     8.00    -0.06   -0.41(.680)  1.006  0.342    1.831  0.512
  4     4.00     5    10.00     0.01    0.05(.959)  1.596  0.827    1.293  0.811
  5     5.00    38    76.00     0.47    3.64(.001)  2.406  0.86     2.57  1.301
=====
    
```

Figure 8 shows the distribution of the score categories of item 1, the point biserial correlation coefficient and t value, and the MNSQ fit value. In addition, since it is a multi-category item, item threshold and delta values are also included.

Figure 9. Conventional test statistics.

```

=====
The following traditional statistics are only meaningful for complete
designs and when the amount of missing data is minimal.
In this analysis 0.00% of the data are missing.

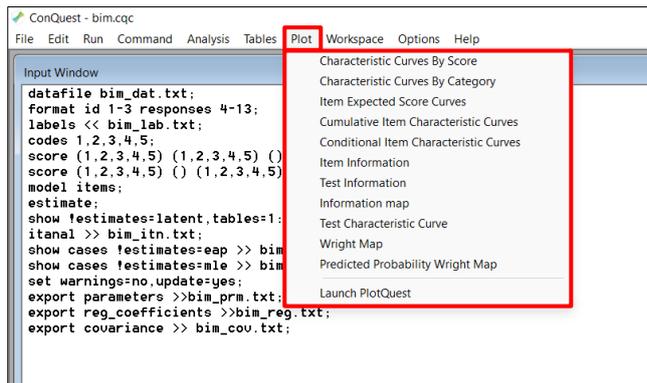
The following results are scaled to assume that a single response
was provided for each item.

N                50
Mean             45.60
Standard Deviation  6.14
Variance         37.71
Skewness         -2.55
Kurtosis         7.14
Standard error of mean  0.87
Standard error of measurement  2.37
Coefficient Alpha  0.85
=====
    
```

Figure 9 presents mean, standard deviation, variance, skewness, kurtosis, standard error of measurement, and reliability values. Since the items in the example are Likert-type, the reliability coefficient is the Cronbach Alpha value. However, when there is a test consisting of double-scored (0-1) items, the reliability value will express the KR-20 value.

After the analyses are performed, the Plot menu becomes active. With the help of the tabs in Figure 10, visual outputs such as characteristic curves, item expected score curves, cumulative and conditional item characteristic curves, item information function, test information function, information function, test characteristic curve, Wright map, and Predicted Probability Wright Map can be created. However, Test infographics and Test characteristic curves cannot be created in multidimensional models. Some example graphic outputs of example application-1 are presented in Appendix 1.

Figure 10. ConQuest Plot menu and sub-tabs.



2.3. Example 2: Within-Item Multidimensional Model

The data created for the Within-Item Multidimensional Model consists of 10 Likert-type items scored from 1 to 5. Assuming that items 1, 6, and 9 are related to both the first and second dimensions of the scale, the within-item multidimensional model in Figure 4 is defined. Items 2, 3, and 5 are only in the first dimension, while items 4, 7, 8, and 10 are only in the second dimension. After creating the data of 50 individuals, the data file with the .txt extension was prepared for analysis. Table 4 involves the command, data, and tag file created by the researchers for the Within-item multidimensional model.

Table 4. An example command, data, and tag file in Within-item Multidimensional model.

wim.cqc	wim_dat.txt	wim_lab.txt
Command statements (required)	Dataset (required)	Dataset label
<pre> datafile wim_dat.txt; format id 1-3 responses 4-13; labels << wim_lab.txt; codes 1,2,3,4,5; score (1,2,3,4,5) (1,2,3,4,5) (1,2,3,4,5) ! items(1); score (1,2,3,4,5) (1,2,3,4,5) () ! items(2); score (1,2,3,4,5) (1,2,3,4,5) () ! items(3); score (1,2,3,4,5) () (1,2,3,4,5) ! items(4); score (1,2,3,4,5) (1,2,3,4,5) () ! items(5); score (1,2,3,4,5) (1,2,3,4,5) (1,2,3,4,5) ! items(6); score (1,2,3,4,5) () (1,2,3,4,5) ! items(7); score (1,2,3,4,5) () (1,2,3,4,5) ! items(8); score (1,2,3,4,5) (1,2,3,4,5) (1,2,3,4,5) ! items(9); score (1,2,3,4,5) () (1,2,3,4,5) ! items(10); model items; estimate; show festimates=latent,tables=1:2:3:9>> wim_shw.txt; itanal >> wim_itn.txt; show cases festimates=eap >> wim_eap.txt; show cases festimates=mle >> wim_mle.txt; </pre>	<pre> 00145555555551 0025555515555 0032553141311 0045555555555 0055555455455 0065455511451 0075555445555 008555553551 0095555555555 010555455555 011555354551 </pre>	<pre> ===> item 1 M1 2 M2 3 M3 4 M4 5 M5 6 M6 7 M7 8 M8 9 M9 10 M10 </pre>

To perform the analysis, the command is opened and run. Since Rasch Analysis results (*wim_shw.txt*) and traditional item analysis results (*wim_itn.txt*) are similar to application example-1, they are presented in Appendix 2. However, a generalized item thresholds table and map cannot be created for Within-Item Multidimensional models. This situation will be better understood when item-1 and item-2 in the traditional item statistics part of Appendix 2 are examined. For example, since item-1 is in both dimensions, threshold values for item-1 cannot be calculated, but these values are calculated for item-2 in only one dimension. Also, some example graphic outputs of example application-2 are presented in Appendix 3.

2.4. Example 3: Many-facet Multidimensional Model

The data created for the Many-facet Multidimensional Model consists of 10 Likert-type items scored from 1 to 5. Items from 1 to 5 are assumed to represent the first dimension, while items from 6 to 10 are assumed to represent the second dimension of the scale. Data from 50

individuals in total were created. It was assumed that each individual's response to each item was scored by three different raters (MSK, FE, MS). Therefore, this example application includes two different facets (item and rater). Table 5 contains the command, data, and tag file created by the researchers for the Many-facet multidimensional model.

Table 5. An example command, data, and tag file in Many-facet Multidimensional Model.

mfm.cqc	mfm_dat.txt	mfm_lab.txt
Command statements (required)	Dataset (required)	Dataset label
<pre>datafile mfm_dat.txt; format id 1-3 rater 4-5 rater 6-7 rater 8-9 responses 10-19 responses 20-29 responses 30-39; labels << mfm_lab.txt; codes 1,2,3,4,5; score (1,2,3,4,5) (1,2,3,4,5) () ! items(1-5); score (1,2,3,4,5) () (1,2,3,4,5) ! items(6-10); model item+rater+item*rater; estimate ! nodes=30;; show !estimates=latent,tables=1:2:3:9>> mfm_shw.txt; itanal >> mfm_itn.txt; show cases !estimates=eap >> mfm_eap.txt; show cases !estimates=mle >> mfm_mle.txt; export parameters >>mfm_prm.txt; export reg_coefficients >>mfm_reg.txt;</pre>	<pre>00101020345555555134555554423555455551 00201020355555155554455515444555515555 0030102032555314131113531412232553141211 00401020355555555345455544455555555555 00501020355554555534555544455555555554 00601020354555114514455511242555511441 0070102035554455534555444555555555555 0080102035555355514455552442555553551 00901020355555554455554445555555555555 01001020355554555544555445555555555555 01101020355534551445534432555534551 01201020355555554455554435555555555555 013010203154514325344551432311555143353 014010203544535555444352425545355444 015010203443355554133335534232355554 016010203555555543354545342555545554 0170102035444455543244445432534445554 0180102035555555434555453425555455555 01901020355555555345555343555555555555 020010203555553554445555325555552454 021010203555555544555553255555555554 022010203555555544455555345555555554 0230102035555555445555534455555555555 0240102035555534544455553343555553454</pre>	<pre>====> item 1 M1 2 M2 3 M3 4 M4 5 M5 6 M6 7 M7 8 M8 9 M9 10 M10 ====> rater 01 MSK 02 FE 03 MS</pre>

To perform the analysis, the command file is opened and run. Rasch Analysis results (*mfm_shw.txt*) and traditional item analysis results (*mfm_itn.txt*) for application example-3 are presented below.

Appendix 4 displays the summary of the estimation and Population Model Parameter Estimates (regression coefficients, covariance/correlation matrix, reliability coefficients), which are similar to the previous application example. On the other hand, Appendix 5 presents example graphic printouts. However, unlike the other two applications, the number of terms in the model has changed. Since a two-facet model is created, there are three terms: *item*, *rater*, and *item*rater*. Therefore, the item parameters of each term in the model were estimated. Table 6 presents the predicted item parameters for each term in the model.

Table 6. Item parameter estimates for each term in the model.

Explanation									
TERM 1: item									
VARIABLES		UNWEIGHTED FIT					WEIGHTED FIT		
item	ESTIMATE	ERROR [^]	MNSQ	CI	T	MNSQ	CI	T	
1 M1	0.501	0.102	0.91 (0.61, 1.39)	-0.4	0.98 (0.59, 1.41)	-0.1			
2 M2	0.048	0.124	1.06 (0.61, 1.39)	0.4	1.39 (0.58, 1.42)	1.7			
3 M3	-0.504	0.132	1.15 (0.61, 1.39)	0.8	2.62 (0.29, 1.71)	3.2			
4 M4	-0.603	0.163	1.20 (0.61, 1.39)	1.0	2.16 (0.42, 1.58)	3.1			
5 M5	0.558*	0.095	1.47 (0.61, 1.39)	2.1	1.55 (0.52, 1.48)	1.9			
6 M6	-0.144	0.107	2.80 (0.61, 1.39)	6.2	3.80 (0.42, 1.58)	5.7			
7 M7	-0.228	0.110	1.53 (0.61, 1.39)	2.4	2.40 (0.40, 1.60)	3.4			
8 M8	0.244	0.105	1.84 (0.61, 1.39)	3.5	2.41 (0.51, 1.49)	4.1			
9 M9	-0.582	0.148	1.57 (0.61, 1.39)	2.5	1.59 (0.59, 1.41)	2.4			
10 M10	0.710*	0.089	1.19 (0.61, 1.39)	1.0	1.47 (0.57, 1.43)	1.9			

An asterisk next to a parameter estimate indicates that it is constrained
 Separation Reliability = 0.904
 Chi-square test of parameter equality = 79.90, df = 8, Sig Level = 0.000
[^] Empirical standard errors have been used
 Term is a fixed effect

Term 1: item
 It includes parameter estimations of ten items and some fit values.

TERM 2: rater

VARIABLES		UNWEIGHTED FIT			WEIGHTED FIT				
rater		ESTIMATE	ERROR^	MNSQ	CI	T	MNSQ	CI	T
1	MSK	-0.267	0.062	0.84	(0.61, 1.39)	-0.8	1.00	(0.51, 1.49)	0.0
2	FE	0.524	0.056	1.10	(0.61, 1.39)	0.6	0.96	(0.60, 1.40)	-0.2
3	MS	-0.257*	0.062	0.88	(0.61, 1.39)	-0.5	1.01	(0.50, 1.50)	0.1

An asterisk next to a parameter estimate indicates that it is constrained
 Separation Reliability = 0.989
 Chi-square test of parameter equality = 105.55, df = 2, Sig Level = 0.000
 ^ Empirical standard errors have been used
 Term is a fixed effect

Term 2: rater

Parameter estimates and some fit values of three different raters are included. A negative estimate indicates that the rater is leniency, and a positive estimate indicates that the rater is severity.

TERM 3: item*rater

VARIABLES		UNWEIGHTED FIT			WEIGHTED FIT						
item	rater	ESTIMATE	ERROR^	MNSQ	CI	T	MNSQ	CI	T		
1	M1	1	MSK	-0.262	0.148	0.67	(0.61, 1.39)	-1.8	0.90	(0.35, 1.65)	-0.2
2	M2	1	MSK	-0.389	0.185	0.56	(0.61, 1.39)	-2.5	1.05	(0.19, 1.81)	0.2
3	M3	1	MSK	0.232	0.186	0.42	(0.61, 1.39)	-3.7	0.89	(0.22, 1.78)	-0.2
4	M4	1	MSK	0.133	0.229	0.69	(0.61, 1.39)	-1.7	1.08	(0.26, 1.74)	0.3
5	M5	1	MSK	0.287*	0.128	0.51	(0.61, 1.39)	-3.0	0.54	(0.50, 1.50)	-2.2
6	M6	1	MSK	0.236	0.153	0.98	(0.61, 1.39)	-0.9	1.33	(0.31, 1.69)	-0.3
7	M7	1	MSK	0.321	0.155	0.51	(0.61, 1.39)	-2.9	0.75	(0.32, 1.68)	-0.7
8	M8	1	MSK	-0.071	0.150	1.02	(0.61, 1.39)	0.1	1.35	(0.34, 1.66)	1.1
9	M9	1	MSK	-0.511	0.223	0.23	(0.61, 1.39)	-5.8	1.50	(0.00, 2.17)	0.9
10	M10	1	MSK	0.025*	0.121	0.69	(0.61, 1.39)	-1.7	0.91	(0.51, 1.49)	-0.3
1	M1	2	FE	0.502	0.126	2.05	(0.61, 1.39)	4.1	0.88	(0.65, 1.35)	-0.7
2	M2	2	FE	0.874	0.154	1.03	(0.61, 1.39)	0.2	0.86	(0.67, 1.33)	-0.9
3	M3	2	FE	-0.509	0.180	0.39	(0.61, 1.39)	-4.0	0.89	(0.24, 1.76)	-0.2
4	M4	2	FE	-0.356	0.214	0.66	(0.61, 1.39)	-1.9	0.87	(0.36, 1.64)	-0.3
5	M5	2	FE	-0.511*	0.125	0.56	(0.61, 1.39)	-2.6	0.58	(0.50, 1.50)	-1.9
6	M6	2	FE	-0.587	0.150	1.14	(0.61, 1.39)	0.7	1.44	(0.31, 1.69)	1.2
7	M7	2	FE	-0.549	0.153	0.57	(0.61, 1.39)	-2.5	0.85	(0.30, 1.70)	-0.3
8	M8	2	FE	0.147	0.146	0.71	(0.61, 1.39)	-1.6	0.68	(0.68, 1.32)	-2.2
9	M9	2	FE	1.009	0.170	2.52	(0.61, 1.39)	5.5	1.30	(0.61, 1.39)	1.4
10	M10	2	FE	-0.021*	0.109	1.20	(0.61, 1.39)	1.0	0.75	(0.64, 1.36)	-1.5
1	M1	3	MS	-0.240*	0.147	0.80	(0.61, 1.39)	-1.0	0.92	(0.35, 1.65)	-0.2
2	M2	3	MS	-0.485*	0.188	0.47	(0.61, 1.39)	-3.3	1.09	(0.15, 1.85)	0.3
3	M3	3	MS	0.277*	0.183	0.42	(0.61, 1.39)	-3.7	1.05	(0.24, 1.76)	0.2
4	M4	3	MS	0.223*	0.225	0.41	(0.61, 1.39)	-3.7	0.77	(0.29, 1.71)	-0.6
5	M5	3	MS	0.224*	0.128	0.58	(0.61, 1.39)	-2.4	0.60	(0.49, 1.51)	-1.8
6	M6	3	MS	0.351*	0.150	0.99	(0.61, 1.39)	0.0	1.24	(0.35, 1.65)	0.8
7	M7	3	MS	0.229*	0.157	0.51	(0.61, 1.39)	-2.9	0.83	(0.30, 1.70)	-0.4
8	M8	3	MS	-0.076*	0.149	0.92	(0.61, 1.39)	-0.4	1.21	(0.34, 1.66)	0.7
9	M9	3	MS	-0.498*	0.221	0.42	(0.61, 1.39)	-3.7	1.70	(0.00, 2.16)	1.2
10	M10	3	MS	-0.005*	0.121	0.74	(0.61, 1.39)	-1.4	0.97	(0.50, 1.50)	-0.1

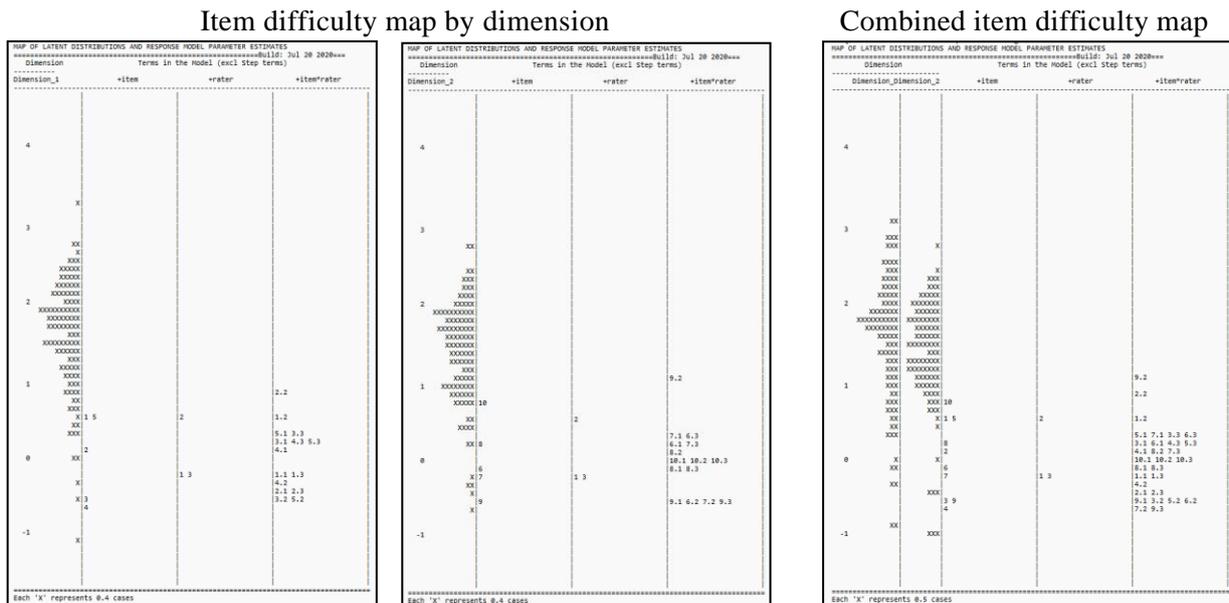
An asterisk next to a parameter estimate indicates that it is constrained
 Separation Reliability = 0.882
 Chi-square test of parameter equality = 144.87, df = 16, Sig Level = 0.000
 ^ Empirical standard errors have been used
 Term is a fixed effect

Term 3: item*rater

Since it is the interaction of the item and rater terms, there are 3x10 (a total of 30) parameter estimations and some fit values. The term interaction is used to determine the joint effect between the variables rather than the main effects.

Similarly, due to the increase in the number of terms in the model, there were some changes in the item difficulty map outputs created according to the latent ability distribution and item difficulty. In Figure 11, there is an item difficulty map created for each dimension separately or by combining the dimensions.

Figure 11. The item difficulty map formed separately according to the dimension and the item difficulty map formed by combining the dimensions.



As can be seen in Figure 11, each dimension was calculated separately and combined, and individual, item, rater, and item*rater interactions were given together. Results for each

dimension can be interpreted separately or in combination. The most difficult item is the number 10 followed by 1 and 5. In the example, the most difficult items were items 5 and 10. Regarding the logit values of the raters, raters numbered one and three were more generous, while rater numbered two was stricter. For item*rater interactions, rater number two exhibited more rigid behaviors in items numbered one, two, and nine.

The Item difficulty map obtained for the multidimensional and many-facet model provides that both the items, the dimensions, the raters, and the item and rater interactions are given on the same scale. In addition, it transforms all sources of variability into logit scales and makes them dependent variables. For example, when the combined item difficulty map is examined, it is seen that the first dimension has a wider range than the second dimension. In addition, it provides the opportunity to see information such as which is the most difficult question or who is the severity rater.

Considering the traditional item analysis results, since there are two terms in the model, item statistics were calculated as much as the interaction number of these two terms. As it is assumed that each item is evaluated by three different raters, traditional item statistics for each item were calculated for each rater. Figure 12 presents the traditional item statistical outputs of three different raters for item-1

Figure 12. Statistics of three different raters for item-1.

GENERALISED ITEM ANALYSIS										
Group All Students										

Item 1										

rater:1 (MSK) item:1 (M1)										
Cases for this item		50	Item-Rest Cor.		0.73	Item-Total Cor.		0.76		
Item Threshold(s):		-0.69	-0.19	0.14	0.63					
Item Delta(s):		-0.03	-0.03	-0.03	-0.03					

Label	Score	Count	% of tot	Pt Bis	t (p)	PV1Avg:1	PV1 SD:1	PV1Avg:2	PV1 SD:2	

1	1.00	2	4.00	-0.60	-5.24(.000)	-0.152	0.595	-0.002	0.628	
2	2.00	1	2.00	-0.49	-3.87(.000)	-0.011	0	-0.541	0	
3	3.00	4	8.00	-0.09	-0.61(.544)	0.721	0.601	1.679	0.58	
4	4.00	5	10.00	-0.04	-0.29(.773)	1.362	0.338	1.154	0.504	
5	5.00	38	76.00	0.52	4.24(.000)	1.822	0.555	1.529	0.65	

Item 2										

rater:2 (FE) item:1 (M1)										
Cases for this item		50	Item-Rest Cor.		0.10	Item-Total Cor.		0.16		
Item Threshold(s):		0.87	1.36	1.69	2.18					
Item Delta(s):		1.53	1.53	1.53	1.53					

Label	Score	Count	% of tot	Pt Bis	t (p)	PV1Avg:1	PV1 SD:1	PV1Avg:2	PV1 SD:2	

1	1.00	6	12.00	-0.27	-1.93(.060)	0.616	0.546	1.085	0.839	
2	2.00	4	8.00	0.06	0.39(.701)	1.55	0.294	1.361	0.674	
3	3.00	20	40.00	0.23	1.66(.103)	1.643	0.503	1.504	0.641	
4	4.00	19	38.00	-0.06	-0.40(.694)	1.852	0.842	1.408	0.865	
5	5.00	1	2.00	-0.11	-0.73(.467)	0.656	0	1.249	0	

Item 3										

rater:3 (MS) item:1 (M1)										
Cases for this item		50	Item-Rest Cor.		0.72	Item-Total Cor.		0.75		
Item Threshold(s):		-0.65	-0.16	0.17	0.66					
Item Delta(s):		0.00	0.00	0.00	0.00					

Label	Score	Count	% of tot	Pt Bis	t (p)	PV1Avg:1	PV1 SD:1	PV1Avg:2	PV1 SD:2	

1	1.00	2	4.00	-0.60	-5.24(.000)	-0.152	0.595	-0.002	0.628	
2	2.00	1	2.00	-0.49	-3.87(.000)	-0.011	0	-0.541	0	
3	3.00	5	10.00	-0.08	-0.53(.597)	0.912	0.673	1.497	0.646	
4	4.00	3	6.00	-0.07	-0.49(.623)	0.919	0.32	1.468	0.415	
5	5.00	39	78.00	0.55	4.52(.000)	1.836	0.519	1.505	0.657	

Figure 12 shows that the statistical values of the items change according to the rater's attitude. While the item-rest correlation values for the item are high for raters 1 and 3 ($r=0.72-0.73$), the value calculated for rater 2 is low ($r=0.10$).

In addition to this Many-facet Multidimensional Model, hypothetical, some research examples in the field of education can also be given. For instance, research such as scoring language skills with a multidimensional structure, which is frequently used in classroom measurement and assessment practices, can be designed by peer assessment. As another example, research can be conducted on scoring students' presentation skills by more than one rater. As a different example, the socio-economic levels and genders of the students can be included in the model as a variable, and analyses can be carried out through the Conquest program by using the Many-facet Multidimensional Model in evaluating academic achievements with a multidimensional structure.

3. DISCUSSION, CONCLUSION and SUGGESTIONS

Please This research aimed to introduce ConQuest, the statistical package program used in the analysis of multidimensional and many-facet data structures, and to show its applications using an example data set. Thus, the installation of the program and the steps of the analysis process were explained.

Conquest is a user-friendly program because of its simple interface. When real-life situations are examined, it is often observed that data sets are complex and multidimensional. Thus, the meaning of performing analyses with only one-dimensional data sets actually means that many data sets cannot be analyzed in real terms. This situation (construct under-representation) creates a negative situation on the validity of direct measurements (Messick, 1995). In this context, analyzing data sets representing real-life situations (e.g., ConQuest) will contribute to the validity of the measurements. Besides, an important feature that distinguishes ConQuest from other programs that perform multidimensional IRT analyses is that it can simultaneously include many different sources of variability in the analysis and show all variables on a single scale. All sources of variability can be interpreted in an interdependent manner. Also, the ConQuest program can provide outputs for the main effects of the variables as well as their common interactions. Thus, rater biases and item biases, which are frequently used in validity studies, can also be analyzed. The third example is an application that takes this situation into account and has not been tested in previous studies. It can be stated that Many-Facet can be applied in many cases that require multidimensionality analysis.

In a nutshell, the ConQuest program is a suitable and user-friendly package program for many-facet and multidimensional data analysis. Many-facet multidimensional analyses can be easily run via the ConQuest package program in situations where there is more than one construct such as higher-order mental skills and in cases where decisions are made by jury evaluations. In addition to its advantages such as easy use, simple interface, and fast analysis, it also has disadvantages such as being a paid program and limited demo version.

Acknowledgments

The authors would like to thank the blind reviewers for their useful comments and insightful suggestions.

Declaration of Conflicting Interests and Ethics

The authors declare no conflict of interest. This research study complies with research publishing ethics. The scientific and legal responsibility for manuscripts published in IJATE belongs to the authors.

Authorship Contribution Statement

Mahmut Sami Koyuncu: Investigation, Resources, Visualization, Software, Formal Analysis, and Writing-original draft. **Mehmet Şata:** Methodology, Supervision, and Validation.

Orcid

Mahmut Sami Koyuncu  <https://orcid.org/0000-0002-6651-4851>

Mehmet Şata  <https://orcid.org/0000-0003-2683-4997>

REFERENCES

- Ackerman, T.A. (1994). Using multidimensional item response theory to understand what items and tests are measuring. *Applied Measurement In Education*, 7(4), 255–278. https://doi.org/10.1207/s15324818ame0704_1
- Adams, R.J., Wilson, M.R., & Wang, W. (1997). The Multidimensional Random Coefficients Multinomial Logit Model. *Applied Psychological Measurement*, 21, 1–24. <https://doi.org/10.1177%2F0146621697211001>
- Adams, R.J., Wilson, M.R., & Wu, M.L. (1997). Multilevel Item Response Models: An Approach to Errors in Variables Regression. *Journal of Educational and Behavioural Statistics*, 22, 46–75. <https://doi.org/10.2307/1165238>
- Adams, R., Cloney, D., Wu, M., Osses, A., Schwantner, V., & Vista, A. (2022). ACER ConQuest Manual. <https://conquestmanual.acer.org/>
- Adams, R.J, Wu, M.L, Cloney, D., and Wilson, M.R. (2020). *ACER ConQuest: Generalised Item Response Modelling Software* [Computer software]. Version 5. Camberwell, Victoria: Australian Council for Educational Research.
- Andrich, D. (1978). A Rating Formulation for Ordered Response Categories. *Psychometrika*, 43, 561–573. <https://doi.org/10.1007/BF02293814>
- Bartolomé, J., & Garaizar, P. (2022). Design and Validation of a Novel Tool to Assess Citizens' Netiquette and Information and Data Literacy Using Interactive Simulations. *Sustainability*, 14(6), 3392. <https://doi.org/10.3390/su14063392>
- Bock, D.R., & Aitkin, M. (1981). Marginal Maximum Likelihood Estimation of Item Parameters: An Application of the EM Algorithm. *Psychometrika*, 46, 443–459. <https://doi.org/10.1007/BF02293801>
- Bronic, M., & Greefrath, G. (2021, September 13–16). *Does The Gender Matter? The Use of A Digital Textbook Compared To Printed Materials*. 15th International Conference on Technology In Mathematics Teaching (ICTMT 15), Copenhagen, Denmark.
- Chen, J., de la Torre, J., & Zhang, Z. (2013). Relative and absolute fit evaluation in cognitive diagnosis modeling. *Journal of Educational Measurement*, 50(2), 123-140. <https://doi.org/10.1111/j.1745-3984.2012.00185.x>
- De Ayala, R.J. (2009). *The theory and practice of item response theory. Methodology in the Social Sciences*. New York: Guildford.
- Finch, H., & Habing, B. (2003, April). *Comparison of NOHARM and DETECT in item cluster recovery: Counting dimensions and allocating items*. Paper presented at the annual meeting of the National Council on Measurement, Chicago.
- Fischer, G.H. (1983). Logistic Latent Trait Models with Linear Constraints. *Psychometrika*, 48, 3–26. <https://doi.org/10.1007/BF02314674>
- Hahn, I. & Kähler, J. (2022). *NEPS Technical Report for Science: Scaling Results of Starting Cohort 3 for Grade 11* (NEPS Survey Paper No. 93). Leibniz Institute for Educational Trajectories, National Educational Panel Study. <https://doi.org/10.5157/NEPS:SP93:1.0>
- Jang, E.E., & Roussos, L.A. (2007). An investigation into the dimensionality of TOEFL using conditional covariance-based non-parametric approach. *Journal of Educational Measurement*, 44(1), 1-21. <https://doi.org/10.1111/j.1745-3984.2007.00024.x>

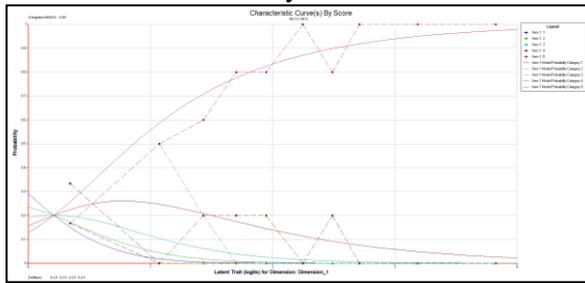
- Jolin, J., & Wilson, M. (2022). Developing a Theory of Two Latent Soft Skills Progress Variables using the BEAR Assessment System: Validity Evidence for the Internal Structure of the Social Evaluative in the Workplace Instrument. *Journal of Psychoeducational Assessment*, 40(3), 381-399. <https://doi.org/10.1177/07342829211057641>
- Jüttler, M., & Schumann, S. (2022). The long-term effects of students' economic competencies on the transition from school to university in the international context. *Research in Comparative and International Education*, 17(2), 196-224. <https://doi.org/10.1177/17454999221086191>
- Krell, M., Khan, S., Vergara, C., Cofré, H., Mathesius, S., & Krüger, D. (2022). Pre-Service Science Teachers' Scientific Reasoning Competencies: Analysing the Impact of Contributing Factors. *Research in Science Education*, 1-21. <https://doi.org/10.1007/s1165-022-10045-x>
- Koch, A., Wißhak, S., Spener, C., Naumann, A., & Hochholdinger, S. (2022). Transfer knowledge of trainers in continuing vocational education and training: Construction and piloting of a test instrument. *Journal for Research on Adult Education*, 1-17. <https://doi.org/10.1007/s40955-022-00210-0>
- Köse, İ.A. (2012). Çok boyutlu madde tepki kuramı [Multidimensional Item Response Theory]. *Journal of Measurement and Evaluation in Education and Psychology*, 3(1), 221-229.
- Linacre, J.M. (1994). *Many-Facet Rasch Measurement*. MESA Press.
- Lou, J., Chen, H., & Li, R. (2022). Emotional Intelligence Scale for Male Nursing Students and Its Latent Regression on Gender and Background Variables. *Healthcare*, 10(5), 814. <https://doi.org/10.3390/healthcare10050814>
- Masters, G.N. (1982). A Rasch Model for Partial Credit Scoring. *Psychometrika*, 47, 149–174. <https://doi.org/10.1007/BF02296272>
- Mendoza, N.B., Cheng, E.C., & Yan, Z. (2022). Assessing teachers' collaborative lesson planning practices: Instrument development and validation using the SECI knowledge-creation model. *Studies in Educational Evaluation*, 73, 101139. <https://doi.org/10.1016/j.stueduc.2022.101139>
- Messick, S. (1995). Validity of psychological assessment. *American Psychologist*, 50(9), 741-749. <https://doi.org/10.1037/0003-066X.50.9.741>
- Mischo, C., Wolstein, K., & Peters, S. (2022). Professional vision of early childhood teachers: relations to knowledge, work experience and teacher child-interaction. *Early Years*, 1-17. <https://doi.org/10.1080/09575146.2022.2028129>
- Mroch, A.A., & Bolt, D.M. (2006). A simulation comparison of parametric and nonparametric dimensionality detection procedures. *Applied Measurement in Education*, 19(1), 67-91. https://doi.org/10.1207/s15324818ame1901_4
- Okon, J. (2022). Creating a motivation scale for secondary school students in Papua New Guinea. *Journal of Applied Learning and Teaching*, 5(1), 1-10. <https://doi.org/10.37074/jalt.2022.5.1.4>
- Osterhaus, C., Kristen-Antonow, S., Kloo, D., & Sodian, B. (2022). Advanced scaling and modeling of children's theory of mind competencies: Longitudinal findings in 4-to 6-year-olds. *International Journal of Behavioral Development*, 46(3), 251-259. <https://doi.org/10.1177/01650254221077334>
- Özbek-Baştuğ, O.Y. (2012). Assessment of dimensionality in social science subtest. *Educational Sciences: Theory and Practice*, 12(1), 375-385.
- Özer-Özkan, Y. (2012). Öğrenci başarılarının belirlenmesi sınavından (ÖBBS) klasik test kuramı, tek boyutlu ve çok boyutlu madde tepki kuramı modelleri ile kestirilen başarı puanlarının karşılaştırılması. [A Comparison of Estimated Achievement Scores Obtained From Student Achievement Assessment Test Utilizing Classical Test Theory,

- Unidimensional And Multidimensional Item Response Theory Models] [Doctoral dissertation, Ankara University]. National Thesis Center of Higher Education Board. <https://tez.yok.gov.tr/UlusalTezMerkezi/>
- Özer-Özkan, Y., & Acar-Güvendir, M. (2014). The analysis of large-scale tests applied in Turkey in terms of their multidimensionality. *Mehmet Akif Ersoy University Journal of Education Faculty*, 1(29), 31-47.
- Patz, R.J., & Junker, B.W. (1999). A straightforward approach to Markov chain Monte Carlo methods for item response models. *Journal of Educational and Behavioral Statistics*, 24(2), 146–178. <https://doi.org/10.2307/1165199>
- Rasch, G. (1980). *Probabilistic Models for Some Intelligence and Attainment Test*. University of Chicago Press.
- Spink, J., Cloney, D., & Berry, A. (2022, January 01). Beyond letters and numbers: the COVID-19 pandemic and foundational literacy and numeracy in Indonesia. *International Education Research*. https://research.acer.edu.au/int_research/7
- Stout, W., Froelich, A.G., & Gao, F. (2001). Using resampling methods to produce an improved DIMTEST procedure. In A. Boomsma, M.A.J. van Duijn, & T.A.B. Snijders (Eds.), *Essay on item response theory* (pp. 357-375). Springer. https://doi.org/10.1007/978-1-4613-0169-1_19
- Unfried, A., Rachmatullah, A., Alexander, A., & Wiebe, E. (2022). An alternative to STEBI-A: validation of the T-STEM science scale. *International Journal of STEM Education*, 9(1), 1-14. <https://doi.org/10.1186/s40594-022-00339-x>
- Volodin, N., & Adams, R.J. (1995). Identifying and estimating a d-dimensional item response model. *International Objective Measurement Workshop*, University of California.
- Wall, S.P., Castillo, P., Shuchat-Shaw, F., Norman, E., Brown, D., Martinez-López, N., & Ravenell, J. E. (2022). Targeting versus Tailoring Educational Videos for Encouraging Deceased Organ Donor Registration in Black-Owned Barbershops. *Journal of Health Communication*, 27(1), 37-48. <https://doi.org/10.1080/10810730.2022.2035021>
- Wang, W. (1995). *Implementation and application of the multidimensional random coefficients multinomial logit*. [Unpublished Doctoral dissertation]. University of California.
- Wang, X., Yan, Z., Huang, Y., Tang, A., & Chen, J. (2022). Re-Developing the Adversity Response Profile for Chinese University Students. *International Journal of Environmental Research and Public Health*, 19, 6389. <https://doi.org/10.3390/ijerph19116389>.
- Wilson, M.R. (1992). The ordered partition model: an extension of the partial credit model. *Applied Psychological Measurement*, 16, 309-325. <https://doi.org/10.1177/014662169201600401>
- Wright, B.D., & Stone, M.H. (1979). *Best test design: Rasch measurement*. MESA Press.

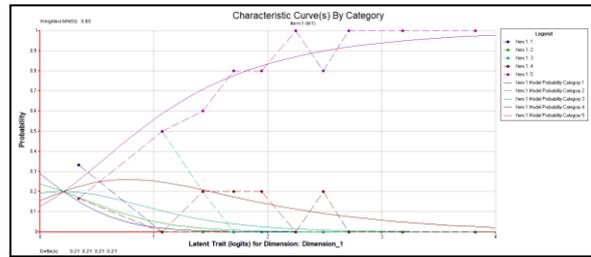
APPENDIX

Appendix 1. Between-Item Multidimensional Model Sample PlotQuest Outputs.

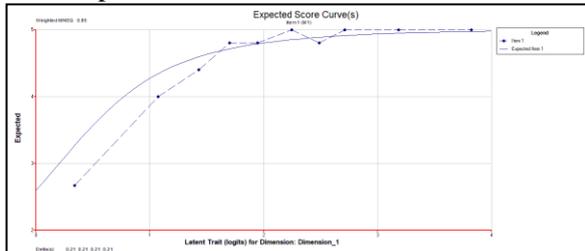
Characteristic Curves By Score



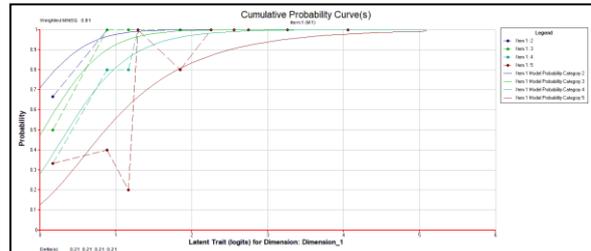
Characteristic Curves By Category



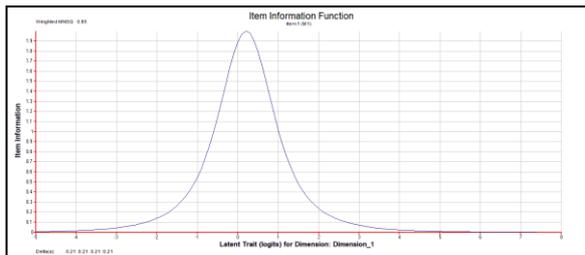
Item Expected Score Curves



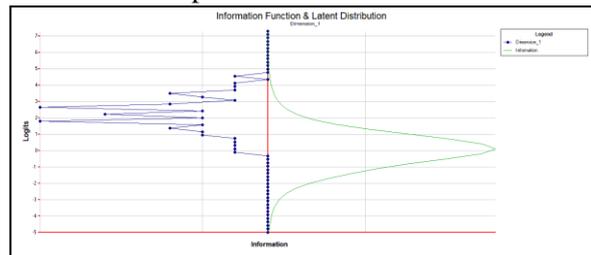
Cumulative Item Characteristic Curves



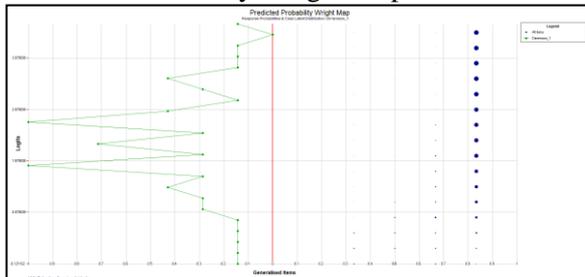
Item Information



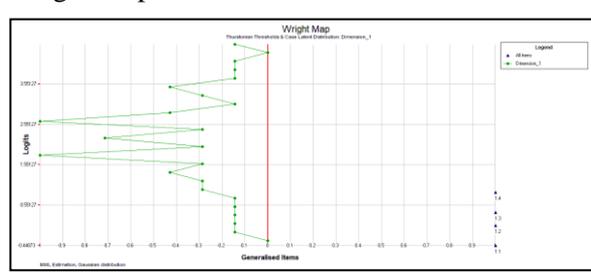
Information map



Predicted Probability Wright Map



Wright Map



Appendix 2. Within - Item Multidimensionality Rasch Analysis results (wim_shw.txt), traditional item analysis results (wim_itn.txt).

Summary of the estimation

```

=====
ConQuest: Generalised Item Response Modelling Software   Sun Mar 06 23:11 2022
SUMMARY OF THE ESTIMATION
=====Build: Jul 20 2020=====
Estimation method was: Gauss-Hermite Quadrature with 225 nodes
No node filtering
Xsi increment max: 1.00000
FacOldXsi: 0.00000
Assumed population distribution was: Gaussian
Location constraint was: DEFAULT
Scale constraint was: Not applicable
The Data File: wim_dat.txt
The format: id 1-3 responses 4-13
No case weights
The regression model:
Grouping Variables:
The item model: items
Slopes are fixed
Cases in file: 51 Cases in estimation: 50
Final Deviance: 694.83877
Akaike Information Criterion (AIC): 720.83877
Akaike Information Criterion Corrected (AICC): 715.80652
Bayesian Information Criterion (BIC): 745.69507
Total number of estimated parameters: 13
The number of iterations: 57
Termination criteria: Max iterations=1000, Parameter Change= 0.00010
Deviance Change= 0.00010
Iterations terminated because the deviance convergence criteria was reached
Random number generation seed: 1.00000
Number of nodes used when drawing PVs: 2000
Number of nodes used when computing fit: 200
Number of plausible values to draw: 5
Maximum number of iterations without a deviance improvement: 100
Maximum number of Newton steps for each parameter in M-step: 10
Value for obtaining finite MLEs for zero/perfects: 0.30000
=====
    
```

Population Model Parametre Estimates

```

=====
TABLES OF POPULATION MODEL PARAMETER ESTIMATES
=====Build: Jul 20 2020=====
REGRESSION COEFFICIENTS
-----
Dimension
-----
Regression Variable  Dimension_1  Dimension_2
CONSTANT            1.538 ( 0.200)  1.501 ( 0.204)
An asterisk next to a parameter estimate indicates that it is constrained
UNCONDITIONAL COVARIANCE/CORRELATION MATRIX
-----
Dimension
-----
Dimension 1  Dimension 2
Dimension_1  0.642
Dimension_2  0.814
-----
Variance  0.604 ( 0.294)  1.030 ( 0.338)
An asterisk next to a parameter estimate indicates that it is constrained
Values below the diagonal are correlations and values above are covariances
-----
RELIABILITY COEFFICIENTS
-----
Dimension: (Dimension_1)
MLE Person separation RELIABILITY: Unavailable
MLE Person separation RELIABILITY: Unavailable
EAP/PV RELIABILITY: 0.622
Dimension: (Dimension_2)
MLE Person separation RELIABILITY: Unavailable
MLE Person separation RELIABILITY: Unavailable
EAP/PV RELIABILITY: 0.678
=====
    
```

Rasch Model Item parameter estimates

```

=====
TABLES OF RESPONSE MODEL PARAMETER ESTIMATES
=====Build: Jul 20 2020=====
TERM 1: items
-----
TERM 1: items
-----
VARIABLES
-----
item  ESTIMATE  ERROR^  UNSWEIGHTED FIT  WEIGHTED FIT
-----
item  ESTIMATE  ERROR^  MNSQ  CI  T  MNSQ  CI  T
-----
1 M1  0.398  0.114  1.19 ( 0.61, 1.39)  1.0  1.30 ( 0.24, 1.76)  0.8
2 M2  -0.426  0.215  0.52 ( 0.61, 1.39)  -2.9  0.90 ( 0.19, 1.81)  -0.2
3 M3  -0.366  0.210  0.44 ( 0.61, 1.39)  -3.5  0.74 ( 0.21, 1.79)  -0.6
4 M4  -0.757  0.276  1.10 ( 0.61, 1.39)  0.5  1.17 ( 0.30, 1.70)  0.6
5 M5  0.608  0.163  0.82 ( 0.61, 1.39)  -0.9  0.73 ( 0.54, 1.46)  -1.2
6 M6  0.344  0.118  3.02 ( 0.61, 1.39)  6.7  2.16 ( 0.21, 1.79)  2.2
7 M7  -0.239  0.189  1.11 ( 0.61, 1.39)  0.6  0.94 ( 0.33, 1.67)  -0.1
8 M8  -0.158  0.184  0.90 ( 0.61, 1.39)  -0.5  1.30 ( 0.34, 1.66)  0.9
9 M9  -0.558*  0.223  0.31 ( 0.61, 1.39)  -4.8  2.37 ( 0.00, 2.35)  1.5
10 M10  1.155*  0.222  1.08 ( 0.61, 1.39)  0.4  1.21 ( 0.50, 1.50)  0.8
An asterisk next to a parameter estimate indicates that it is constrained
Separation Reliability = 0.839
Chi-square Test of parameter equality = 51.41, df = 8, Sig Level = 0.000
^ Empirical standard errors have been used
Term is a fixed effect
=====
    
```

Traditional Item (item-1 and item-2) and Test statistics

```

=====
Item 1
-----
Item:1 (M1)
Cases for this item 50 Item-Rest Cor. 0.68 Item-Total Cor. 0.80
Item Threshold(s): NOT AVAILABLE Weighted MNSQ 1.30
Item Delta(s): 0.80 0.80 0.80 0.80
-----
Label  Score  Count  % of tot  Pt Bis  t (p)  PVI Avg:1  PVI SD:1  PVI Avg:2  PVI SD:2
-----
1  2.00  2  4.00  -0.60  -5.16(.000)  0.003  0.042  -0.246  0.229
2  4.00  1  2.00  -0.49  -3.85(.000)  -0.047  0  -0.661  0
3  6.00  4  8.00  -0.03  -0.17(.862)  1.104  0.543  1.856  0.613
4  8.00  5  10.00  0.01  0.09(.925)  1.376  0.225  1.129  0.232
5  10.00  38  76.00  0.44  3.39(.001)  1.876  0.706  1.884  0.847
-----
Item 2
-----
Item:2 (M2)
Cases for this item 50 Item-Rest Cor. 0.57 Item-Total Cor. 0.63
Item Threshold(s): -1.00 -0.59 -0.26 0.21 Weighted MNSQ 0.99
Item Delta(s): -0.43 -0.43 -0.43 -0.43
-----
Label  Score  Count  % of tot  Pt Bis  t (p)  PVI Avg:1  PVI SD:1  PVI Avg:2  PVI SD:2
-----
1  1.00  1  2.00  -0.59  -5.10(.000)  0.033  0  -0.084  0
3  3.00  1  2.00  -0.02  -0.11(.911)  0.797  0  0.909  0
4  4.00  7  14.00  -0.15  -1.08(.285)  1.208  0.474  0.961  0.691
5  5.00  41  82.00  0.36  2.68(.010)  1.786  0.783  1.699  0.928
=====
    
```

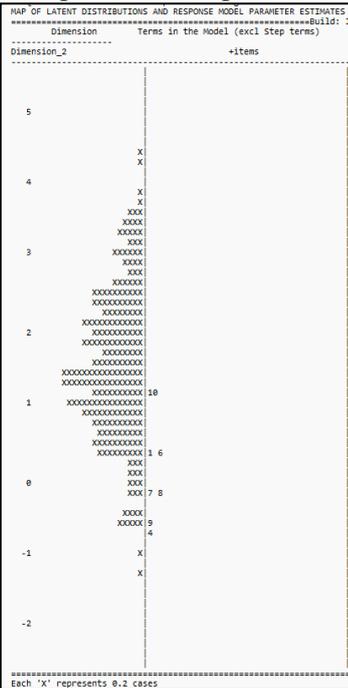
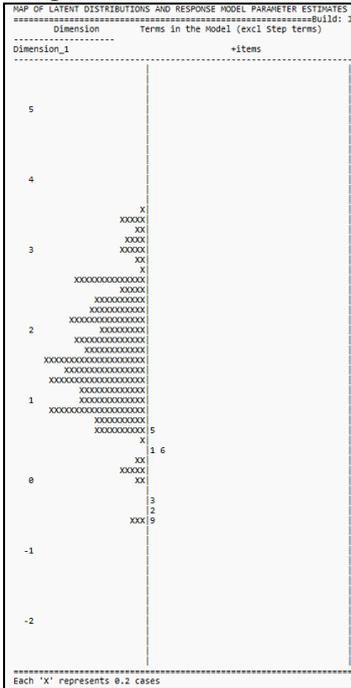
```

=====
The following traditional statistics are only meaningful for complete designs and when the amount of missing data is minimal.
In this analysis 0.00% of the data are missing.

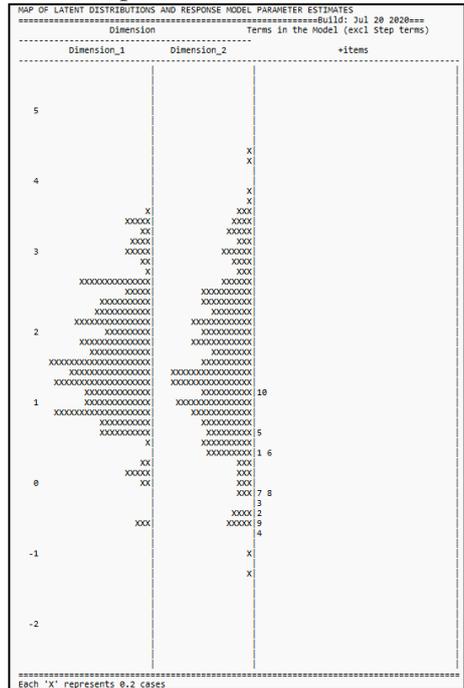
The following results are scaled to assume that a single response was provided for each item.

N 50
Mean 59.58
Standard Deviation 7.92
Variance 62.70
Skewness -2.69
Kurtosis 7.85
Standard error of mean 1.12
Standard error of measurement 3.29
Coefficient Alpha 0.83
=====
    
```

Map of latent distributions and response model parameter estimates outputs



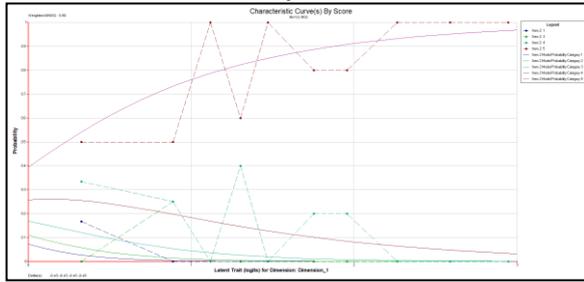
By dimension item difficulty maps



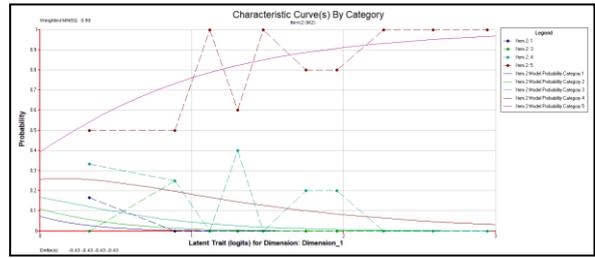
Combined item difficulty map

Appendix 3. Within-Item Multidimensional Model Sample PlotQuest Outputs.

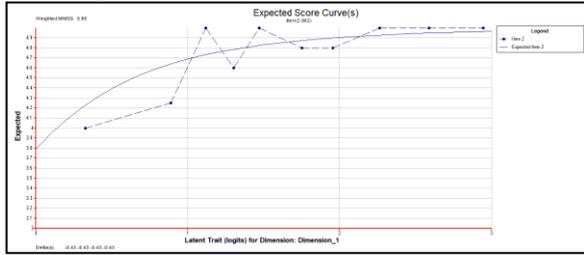
Characteristic Curves By Score



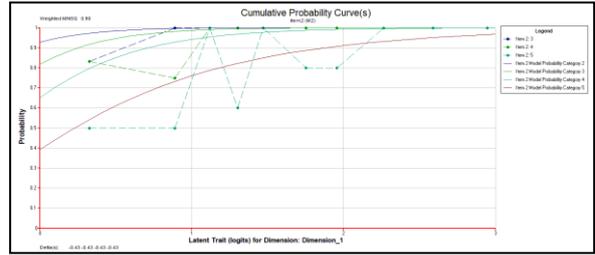
Characteristic Curves By Category



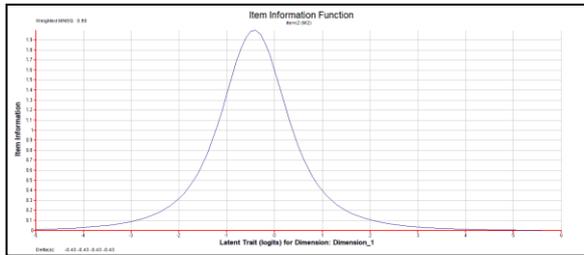
Item Expected Score Curves



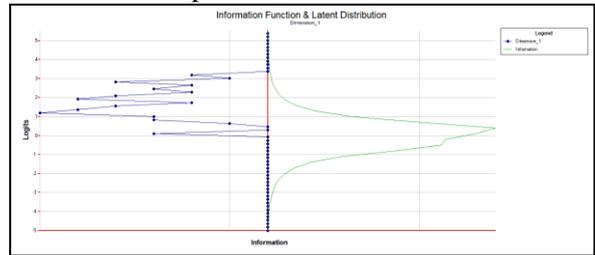
Cumulative Item Characteristic Curves



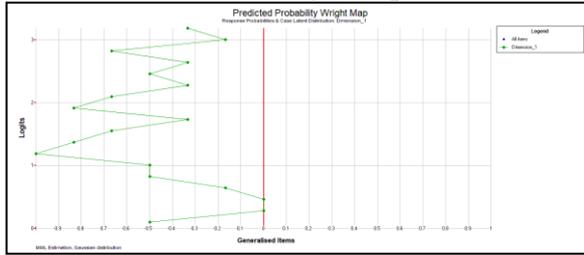
Item Information



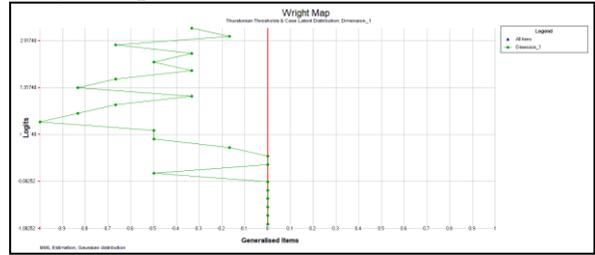
Information map



Predicted Probability Wright Map



Wright Map



Appendix 4. Many-facet Multidimensional Model Rasch Analysis results (wim_shw.txt), traditional item analysis results (wim_itn.txt).

Summary of the estimation

```

SUMMARY OF THE ESTIMATION
=====Build: Jul 20 2020====
Estimation method was: Gauss-Hermite Quadrature with 900 nodes
No node filtering
Xsi increment max: 1.00000
FacOldxsi: 0.00000
Assumed population distribution was: Gaussian
Location constraint was: DEFAULT
Scale constraint was: Not applicable
The Data File: wfm_dat.txt
The format: id 1-3 rater 4-5 rater 6-7 rater 8-9 responses 10-19 responses 20-29 responses 30-39
No case weights
The regression model:
Grouping Variables:
The item model: item+rater+item*rater
Slopes are fixed
Cases in file: 50 Cases in estimation: 50
Final Deviance: 2388.17613
Akaike Information Criterion (AIC): 2450.17613
Akaike Information Criterion Corrected (AICC): 2426.92613
Bayesian Information Criterion (BIC): 2509.44884
Total number of estimated parameters: 31
The number of iterations: 31
Termination criteria: Max iterations=1000, Parameter Change= 0.00010
Deviance Change= 0.00010
Iterations terminated because the deviance convergence criteria was reached
Random number generation seed: 1.00000
Number of nodes used when drawing PVS: 2000
Number of nodes used when computing fit: 200
Number of plausible values to draw: 5
Maximum number of iterations without a deviance improvement: 100
Maximum number of Newton steps for each parameter in N-step: 10
Value for obtaining finite MLEs for zero/perfects: 0.30000
=====
    
```

Traditional test statistics

```

=====
The following traditional statistics are only meaningful for complete
designs and when the amount of missing data is minimal.
In this analysis 0.00% of the data are missing.

The following results are scaled to assume that a single response
was provided for each item.

N          50
Mean       130.84
Standard Deviation 16.25
Variance   264.01
Skewness   -2.38
Kurtosis    6.25
Standard error of mean 2.30
Standard error of measurement 4.19
Coefficient Alpha 0.93
=====
    
```

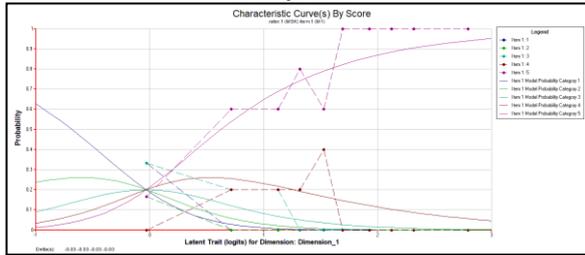
Population Model Parametre Estimates

```

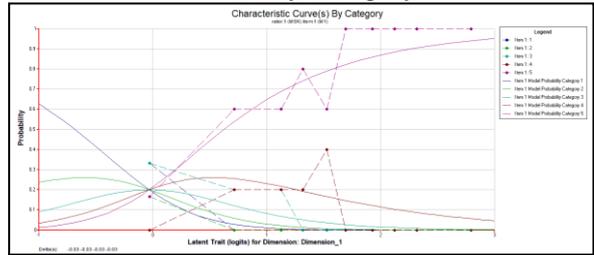
=====Build: Jul 20 2020====
TABLES OF POPULATION MODEL PARAMETER ESTIMATES
REGRESSION COEFFICIENTS
-----
Dimension
-----
Regression Variable  Dimension_1  Dimension_2
CONSTANT            1.539 ( 0.135)  1.367 ( 0.127)
-----
An asterisk next to a parameter estimate indicates that it is constrained
=====
UNCONDITIONAL COVARIANCE/CORRELATION MATRIX
-----
Dimension
-----
Dimension  1  2
Dimension_1  1  0.320
Dimension_2  0.529  1
Variance    0.643 ( 0.149)  0.569 ( 0.130)
-----
An asterisk next to a parameter estimate indicates that it is constrained
Values below the diagonal are correlations and values above are covariances
=====
RELIABILITY COEFFICIENTS
-----
Dimension: (Dimension_1)
MLE Person separation RELIABILITY: Unavailable
WLE Person separation RELIABILITY: Unavailable
EAP/PV RELIABILITY: 0.809
Dimension: (Dimension_2)
MLE Person separation RELIABILITY: Unavailable
WLE Person separation RELIABILITY: Unavailable
EAP/PV RELIABILITY: 0.773
=====
    
```

Appendix 5. Many-facet Multidimensional Model Sample PlotQuest Outputs).

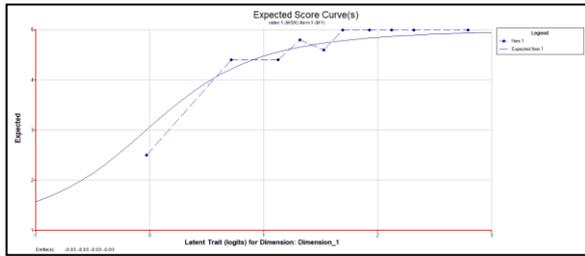
Characteristic Curves By Score



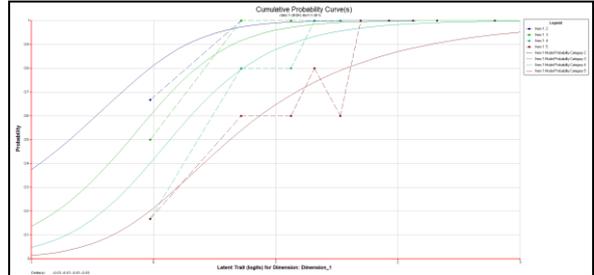
Characteristic Curves By Category



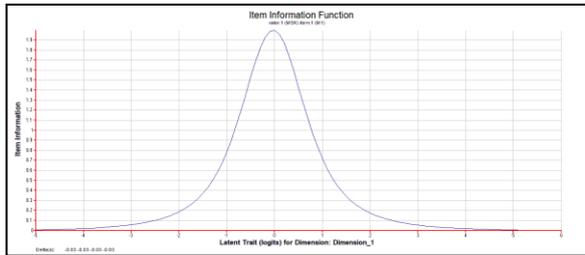
Item Expected Score Curves



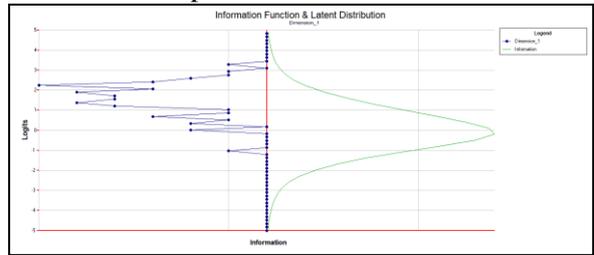
Cumulative Item Characteristic Curves



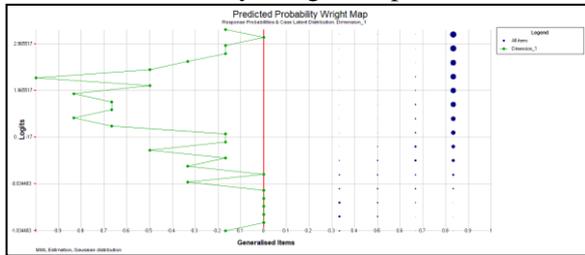
Item Information



Information map



Predicted Probability Wright Map



Wright Map

