ISSN: 1309 – 6575
Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi
Journal of Measurement and Evaluation in Education and Psychology
2019; 10(2);165-178

JMETRIK: Classical Test Theory and Item Response Theory Data Analysis Software

Gökhan AKSU*   Cem Oktay GÜZELLER**   Mehmet Taha ESER***

Abstract

The aim of this study is to introduce the jMetric program which is one of the open source programs that can be used in the context of Item Response Theory and Classical Test Theory. In this context, the interface of the program, importing data to the program, a sample analysis, installing the jmetrik and support for the program are discussed. In sample analysis, the answers given by a total of 500 students from state and private schools, to a 10-item math test were analyzed to see whether they shows differentiating item functioning according to the type of school they attend. As a result of the analysis, it was found that two items were showing medium-level Differential Item Functioning (DIF). As a result of the study, it was found that the jMetric program, which is capable of performing Item Response Theory (IRT) analysis for two-category and multi-category items, is open to innovations, especially because it is open-source, and that researchers can easily add the suggested codes to the program and thus the program can be improved. In addition, an advantage of the program is producing visual results related to the analysis through the item characteristic curves.

Keywords: jMetrik, item response theory, classical test theory, differential item functioning.

INTRODUCTION

For researchers nowadays, technology has almost the same meaning as the software that they use every day. Software products offer solutions for many challenges faced by the users. Technology extended the usage of software analysis by accessing to a wider audience and by this means researchers at each specialization level may develop themselves and experience different software products relevant to their field. The use of software, which has a great importance in all fields in scientific terms, is of great importance in terms of calculation, evaluation and development of statistics on measurement results in the field of measurement and evaluation and psychometry. The statistics calculated and used in the context of classical test theory (CTT) and item response theory (IRT), which have a very important role in psychometrics, are complex, difficult in terms of manual calculation and time consuming, which encourages researchers to use software. At this point, the needs of the researchers may change over time and the need towards a software that is easy to use, cheaper or free, fast, open-to-development increases day-by-day.

Classical Test Theory, which is also called as true score model occasionally, covers the mathematical computations laying in the background of measurement tool development process. CTT is almost 100 years old and it is still widely used. The statistics, such as correlation among items, covariance, difficulty index, discrimination power index, reliability coefficient, variance/standard deviation of the sample, measurement error, etc. are calculated by CTT, which is mainly used for the purpose of developing and improving the reliability and validity of measurement tools (Crocker and Algina, 1986; Mcdonald, 1999). Most of the statistics covered in CTT are based on mean, ratio and correlation. The

* Dr., Adnan Menderes University, Aydın Vocational School, Aydın-Turkey, e-mail: gokhanaksu1983@hotmail.com ORCID ID: 0000-0003-2563-6112
** Prof. Dr., Akdeniz University, Tourism Faculty, e-mail: cguzeller@gmail.com ORCID ID: 0000-0002-2700-3565
*** Öğr. Gör., Akdeniz University, Statistical Research Center, e-mail: tahaeser@gmail.com ORCID ID: 0000-0001-7031-1953

To cite this article:
Last name, A., & Last name, B. (201?). JMETRIK: Classical test theory and item response theory data analysis software. *Journal of Measurement and Evaluation in Education and Psychology, 10*(2), 165-178. DOI: 10.21031/epod.483396
Received: 15.11.2018
Accepted: 12.03.2019

theory has a constant perspective to deal with important problems related to measurement. The need of seeking for another test theory emerged due to several weaknesses of CTT, including: item and test statistics depend on the test and on the group, which the test was applied; a single error estimation is obtained for all ranges of skill level; and the weaknesses on test linking/equating. These weaknesses led to the development of IRT that is seen as a significant innovation in the field of Psychometry (Hambleton and Swaminathan, 1985; Embretson and Reise, 2000; Meyer, 2010). Individuals often get low scores from difficult tests and higher scores from easy ones whereas their skill level stays constant. This caused the development of another test theory, which is IRT, originally presented in the manuscript of Lord and Novick (1968). Compared to CTT, IRT is stronger regarding the applications of linking/equating, differential item functioning (DIF) and individualized computer test. Since the statistics of IRT have more complex structure in terms of both computation and interpretation, compared to the statistics of CTT, various software products were developed to facilitate the tasks of the researchers in every sense. There are software products performing the computations of both CTT's and IRT's statistics, as well as software products solely performing the calculations according to CTT or IRT. CITAS, ITEMAN, Lertap, and TAP are the packages that are widely used by researchers, using which only the analysis of CTT applications can be performed; BILOG-MG, flexMIRT, ICL, MULTILOG, PARSCALE, PARAM-3PL, Winsteps and Xcalibre, IRT PRO, NOHARM, TESTFACT, flexMIRT are the packages using which the analysis of IRT applications can be performed; whereas jMetrik, R and Mplus are the popular software products can compute statistics for CTT and IRT. Regarding the software packages, which considerably facilitate the computations of IRT, researchers may only use the complete edition of jMetrik, PARAM-3PL, NOHARM and R free of charge.

This study aims to provide information about the functionality of jMetrik software, which has been developed to help statistical and psychometric procedures related to both CTT and IRT, and to indicate the differences between jMetrik and the other software products. For this purpose, the readers are informed about the functionality, installation, interface, strength and support of the software, and the outputs of an analysis performed by the software were illustrated as an example.

jMetrik is a free, open source psychometric software. It can be run on any Windows, Mac, OSX or Linux-based platform with a current Java version. The first version of the software has been released in 2009, then the second version with two major revisions was released, followed by the third release to which some statistical methods and interface changes were added. The current version of jMetrik is jMetrik 4.1.1. Dr. Meyer, the developer and copy right owner of jMetrik, continues his work at the University of Virginia.

jMetrik is a user-friendly software, it is designed to facilitate working in a production environment and to enable each researcher to use advanced psychometric procedure . Compared to similar software products, it provides a more integrated system in terms of carrying out psychometric analysis for research and operational purposes free of cost, unlike some other psychometric software. jMetrik provides comprehensive statistical and psychometric procedures such as descriptive statistics, IRT parameter estimation, linking scales and score equalization. Moreover, jMetrik helps to create various graphs and tables for the visualization of the data. The structure of software's graphical user interface is intuitive and easy to learn. In addition, it scales according to the experience of the user. New users can execute psychometric procedures via pop-up menus with marks, whereas experienced users can use jMetrik commands to automate the analysis. Another significant feature of jMetrik is being an integrated database that allows users to easily organize and manage data. Results obtained from an analysis can be saved in the database and they can be used as input for another analysis. There is no need to manipulate or reshape the data between each psychometric procedure, which significantly reduces the time required for a complete and comprehensive psychometric analysis as well as the efforts made for analysis. jMetrik can perform many statistical and psychometric methods. The most important of these are undoubtedly the analytic and psychometric methods that are related to IRT.

Although the frequency of use of jMetrik in international studies increased day by day, its national use has not reached the desired level yet. jMetrik software is not known sufficiently, which may explain the reason of this fact. In the literature, there is only one national study concerning the introduction of jMetrik (Aksu, Reyhanlıoğlu and Eser, 2017) which provides information about the functionality and strengths of jMetrik, thus this study is considered to be important. It is believed that the introduction of this software, which is easy to use and free of charge, to the researchers who want to perform analysis within the scope of both IRT and CTT, will make significant contributions to future studies

FUNCTIONALITY of the SOFTWARE, STRENGTHS of the SOFTWARE, SOFTWARE INTERFACE, SAMPLE ANALYSIS and SOFTWARE SUPPORT

Functionality of the Software

jMetrik is used for the calculation of statistics, reliability estimation, test scaling, DIF, nonparametric IRT applications, Rasch measurement models, IRT models (3PLM, 4PLM, GPCM vb.), IRT linking and equalization. jMetrik 4 has a great importance in the use of parametric and non-parametric IRT applications. Ramsay (1991, 2000) has used Kernel Regression to directly estimate the item characteristic curves for two-category and multi-category items. Kernel regression is a method used not only to predict characteristic curves of the items but also to estimate curves for both groups (DIF). In jMetrik, non-parametric IRT procedures can be easily saved in color, as .jpg or .png files. Nonparametric characteristic curves provide an easy and fast tool to examine the data and analyze the relationship between latent traits and correct responses. The only limitation of nonparametric characteristic curves is the actual difficulty of each item and the subjective interpretation of discrimination. Parametric IRT makes it easier to quantify these properties, compare the items, or compare two different groups of the same item.

jMetrik offers two estimation options in terms of parametric IRT. Software uses maximum likelihood estimation for Rasch, partial credit and rating scale models (Wright and Masters, 1982). Partial credit model is formulated by the item difficulty parameter and two or more threshold parameters. Regarding rating scale model, it can be said that it is a special case of partial credit model with threshold parameters. jMetrik uses Proportional Curve Fitting Algorithm (Meyer and Hailey, 2012) instead of Newton-Raphson Method for individual, item and threshold parameters. The software computes goodness of fit statistics for the items and the individuals in addition to parameter estimations. In addition, scale quality statistics such as separation and reliability can be calculated within the scope of the software.

jMetrik, uses marginal maximum likelihood estimation (MMLE) for two-category and multi-category IRT models, including 3-Parameter Logistic Model (3PLM), 4-Parameter Logistic Model (4PLM), and generalized partial credit model (GPCM). In addition to MMLE, the software offers Bayes Model Estimation and normal, lognormal, four-parameter a priori beta distribution options for each item parameter. Generalized S-X2 Statistics are used in terms of item fit of these models (Kang and Chen, 2007; Orlando and Thissen, 2000). jMetrik has three options for scoring the individual characteristics, which are maximum likelihood, maximum a posteriori (MAP) and expected a posteriori (EAP). Software options allow the creation of output tables with analysis results, which can be used as inputs in the procedures such as linking the scales, etc. (Meyer, 2018).

jMetrik offers two options for depicting the analysis results of IRT. The first method provides item characteristic curves, information functions and standard error functions for all items separately and for the whole test. The software uses the information contained in the output tables to automatically select the appropriate IRT model and produce these graphics quickly. The second method provide item maps in the analysis results. Item mapping method is quite common within the scope of the Rasch measurement model and it illustrates the distribution of individual's skill estimates and the distribution of item parameter's estimates in the form of two histograms with a common axis. The method is useful in terms of assessing the quality of match between individuals and items, and to determine whether more (or less) items are needed to obtain a more precise (or more effective) estimate of individual's skill. In other words, the method is a tool that guides the selection of items for the test development process and the test.

Psychometry experts usually need to link the items in different test formats in a common scale or equalize the scores . jMetrik offers several options for linking under IRT and the equalization of non-equivalent group design in the data collected according to a common item. These options are simultaneous calibration, constant common item parameter and conversion coefficient methods. Within the scope of jMetrik, simultaneous calibration and constant common item parameter methods are limited to Rasch model family at the moment. The conversion coefficient methods covered in jMetrik include a wider range of IRT models, including 3PLM, GPCM and graded response model (PRM) (2PLM, Rasch and partial credit model options, which are special cases of these models are included within the scope of the software). Linking the scales can be executed in a combination with one of these models or with any model (mixed test model). Conversion coefficient methods covered in jMetrik includes mean/average (Loyd and Hoover, 1980), mean/sigma (Marco, 1977), Haebara (Haebara, 1980) and Stocking-Lord (Stocking and Lord, 1983) procedures. For the characteristic curve methods, the type of distribution that minimizes the criterion function can be selected. Evenly spaced points can be used from a normal or uniform distribution, four points and weights, or a histogram of estimated individual skills' values. jMetric has the option of minimizing these distributions by forward, backward or symmetrical moves of criterion function (Kim and Kolen, 2007).

Linking scales places the parameters on a common scale. Linking scales is sufficient for achieving comparable scores at the point where the conversion of the participant's skill occurs in the metric to be reported. On the other hand, an additional score equalization step is needed if the reporting metric is a conversion of the observed score (Cook and Eignor, 1991; Meyer, 2018). In such cases, jMetrik allows users to perform IRT real-time score equalization procedure through single-format or mixed-format tests

Regarding DIF, Mantel-Haenszel, Joint Probability Effect Size, Standardized p-DIF effect size and ETS DIF classification levels can be obtained within the scope of jMetrik. These statistics help to evaluate the statistical and practical importance of DIF.

Advantages of the Software

jMetrik is a Java application and it works on Windows, Max OSX, or Linux operating systems with Java 7 or a higher version. jMetrik does not require more than 512 megabytes of available memory. This memory allocation is sufficient for large samples up to 1,000,000, but it can be increased when needed.

Another advantage of jMetrik is that it uses a single frame to combine psychometric methods that require multiple software, which allows a researcher to quickly switch from one analysis method to another (For example, the output of parameter estimation of jMetrik is input for linking scales). This tight integration contradicts other software. A researcher who has not used jMetrik may need a maximum of three software to re-shape and manage the data, to estimate item parameters, and to establish a scale linking. Even with a software like R, it is important to be able to operate functions of a package efficiently with the functions of another package. jMetrik is designed to avoid this hassle by integrating the workflow for various psychometric procedures.

jMetrik has a user-friendly interface that is easy to use. Analysis can be performed from point-and-click menus and dialog boxes. This feature allows new users to learn the software quickly and also makes teaching a lot easier. With conventional software, the time devoted to the course is consumed by the time required to debug old archaic syntax and Fortran format expressions. The point-and-click interface of jMetrik prevents these struggles and allows trainers to regain their time for teaching the theory.

The point-and-click interface is the most obvious way to perform an analysis in jMetrik, but this is not the only way. Each analysis can also be executed and automated through syntax. The task of analysis windows is to generate code in the background. All codes executed by the software are saved in the log. A user can save the log and scripts for later use and edit them. In jMetrik, the command log is separated from the error log. The command log keeps a record of all methods executed by a user. It can be saved and used to run the analysis again with a few changes (for example, changing the names of the data tables).

With jMetrik, a more transparent approach was adopted to psychometric calculation. The use of personal software that are closed to improvement is seen as a statistical norm in the development of large-scale tests but this limitation makes it difficult for stakeholders to check and verify the integrity of the software. There are two publicly available Java libraries for jMetrik: jMetrik library contains all interface and database codes; whereas psychometry library contains all measurement and psychometric methods. All codes of the software are available online at www.github.com, a storage space integrated with git software. Anyone can browse and install source codes. Programmers who know Java programming language can make changes in the codes, they can add patches and new features (Code changes are added to the library after review and approval). Psychometry library provides royalty-free use in special software without licensing and any conditions, allowing the institutions to use the psychometry library to create registered systems at institutional level using public tools.

www.ItemAnalysis.com is the official website of the software. The website also includes sample data files, quick procedures for the software, and answers to frequently asked questions. Questions about the software are answered very quickly by the software developer himself. jMetrik is an open source application distributed under General Public License version 3 or higher. Source code of psychometric procedures covered by jMetrik is also open source and it is distributed under Apache License version 2 and it can be installed from https://github.com/meyerjp3/psychometrics address.

At the same time, factor analysis with various rotation method options, polychoric and polyserial correlation analysis can be performed within the scope of jMetrik.

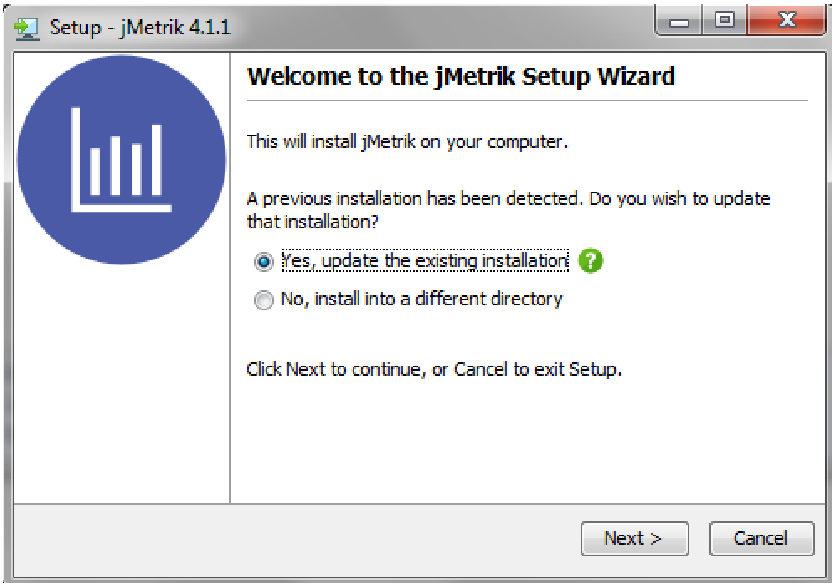
CTT analysis within the scope of JMetrik includes item and test analysis and test scaling. Classical item analysis includes the options such as item statistics, reliability analysis and conditional standard error of measurement. The analysis output of CTT, which includes item statistics, test statistics, and reliability analysis, can be saved as a text file. The output contains item difficulty, standard deviation, and two different item correlations (biserial correlation and point-biserial correlation) for each of the multiple choice and structured open-ended items. In addition, five different reliability calculation methods are available, namely Guttman’s Lambda, Cronbach’s Alfa, Feldt-Gilmer, Feldt-Brennan and Raju’s Beta. Decision consistency and accuracy estimates are provided for item analysis: Huynh’s Raw Agreement, Huynh’s Kappa, KR-21,Beta-binamial alpha and Beta-binomial beta. The classical item analyzes offered by jMetrik are very comprehensive for all user levels in both research and practical environments.

Test scaling options of CTT are very easy to use in jMetrik and many options are available. Users can quickly convert the data to overall, classifying percentage, Kelley’s true score and normalized score. At the same time, users can determine the constraints for minimum, maximum and precision points, in addition to these they can also perform an optional linear conversion. CTT analysis provided by the software offers a point-and-click type interface similar to SPSS and Excel.

SPSS files (.sav) can be directly imported to jMetrik and the software can convert the data set to a file with .sav extension and export it.

Installing the Software

Researchers may install 4.1.1 version of jMetrik software, released in February 2018 from the address https://itemanalysis.com/jmetrik-download/after determining the appropriate operating system for their computers. During the installation of the software, if the java application on the computer is an older version, the software will direct you to the latest java application. The minimum java version required for jMetrik software is 1.8, otherwise the software will not function properly. To install the software, install it here first link under the heading “install jMetrik version 4.1.1” should be clicked. After installing the setup file to your computer, the install process is continued by clicking the Run button. Then, the installation is continued by clicking Next button on the screen shown in Figure 1.



Setup - jMetrik 4.1.1

Welcome to the jMetrik Setup Wizard

This will install jMetrik on your computer.

A previous installation has been detected. Do you wish to update that installation?

● Yes, update the existing installation

○ No, install into a different directory

Click Next to continue, or Cancel to exit Setup.

Figure 1. jMetrik Software Setup Screen

After this operation, click the option to accept the terms of the license agreement as in the installation of other software and click Next button. The shortcut of jMetrik software will be added on your desktop after completing the install process as described above.

Software Interface

The main interface of jMetrik software is shown in Figure 2. This interface consists of a) the main menu and toolbar area, b) a list of database tables, c) a tabbed pane showing a view of the data and analysis output, and d) a status bar providing feedback to the user. The main menu allows you to access software procedures. For example, data management features such as import and export can be accessed through the manage menu; whereas psychometric procedures can be accessed through transform, analyze or graph menus.

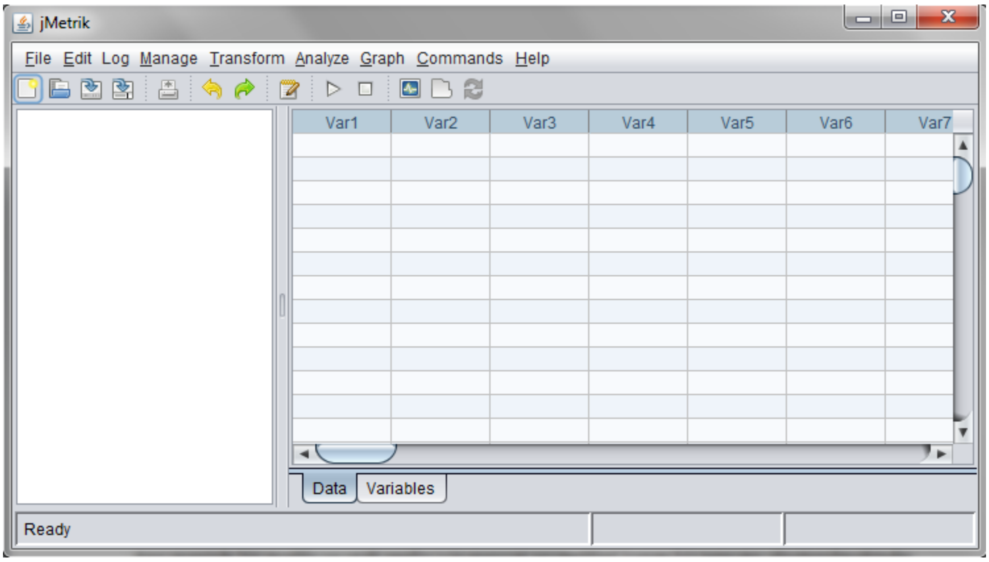


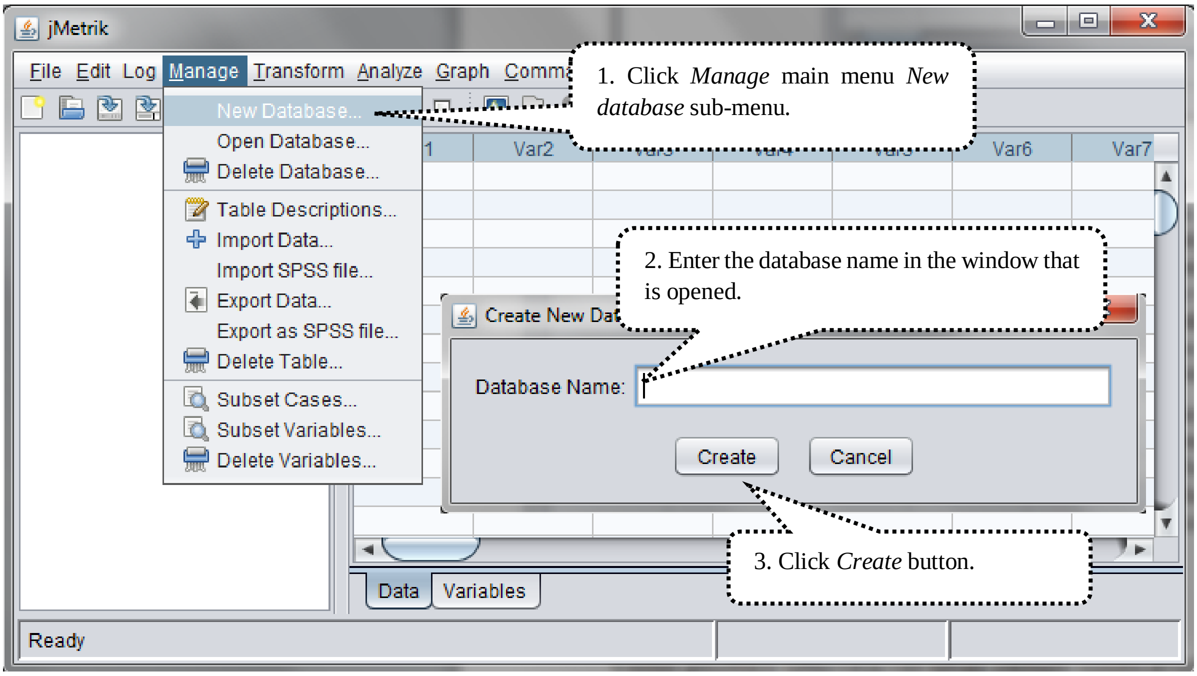
Figure 2. jMetrik Software Main Screen

Selecting an item in the main menu creates a window with options available for analysis. The components in the windows (radio buttons, checkboxes, etc.) directly correspond to various jMetrik commands. The user has to select the options she/he needs for her/his analysis and then press the run button. When the run button is pressed, the instruction for the analysis is automatically recorded into the log file of the software and the analysis is performed. To view the command (and all other commands that run during a session), log> script log path in the main menu is followed. Errors and other problems are recorded in a separate area that can be displayed by following log> view log path.

Any analysis in jMetrik can be performed by typing the commands or by copying and pasting them into command index. A new text file is opened by following File> New path and the command is entered into the new text file. The file directory can be saved by following File> Save As path. To perform the analysis, Commands> Run Commands path is followed in the main menu after entering the command (or multiple commands).

Preparing Data for Analysis

Data is usually presented in the form of tables or databases. However, the data storage method of jMetrik software is the input of data that you have previously obtained from different databases. Therefore, when performing analysis with jMetrik software, the database must be defined first. Figure 3 shows the procedures to be followed to create a database.



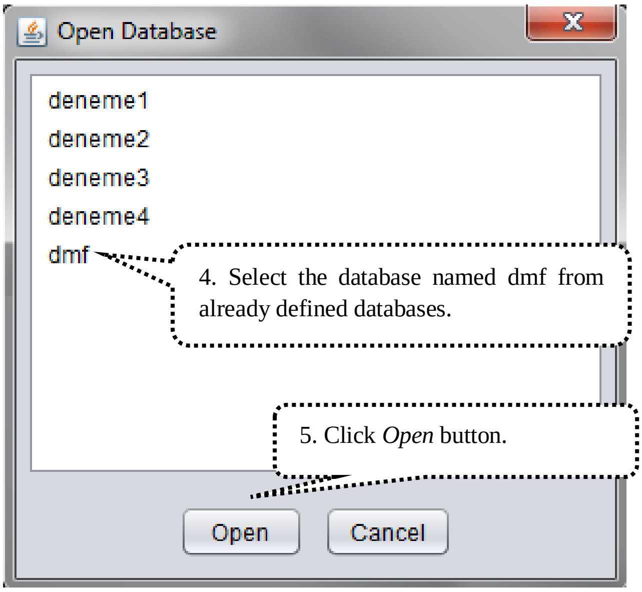
1. Click *Manage* main menu *New database* sub-menu.

2. Enter the database name in the window that is opened.

3. Click *Create* button.

Figure 3. JMetrik Software Database Creation Window

After this process, the creation of the database defined as dmf in the example application is completed by clicking *Create* button. The appearance of *ready* sign at the bottom left of the screen means that the database is created. The next step is to open this database by following Manage >> Open Database steps. Figure 4 shows how to open already defined database.

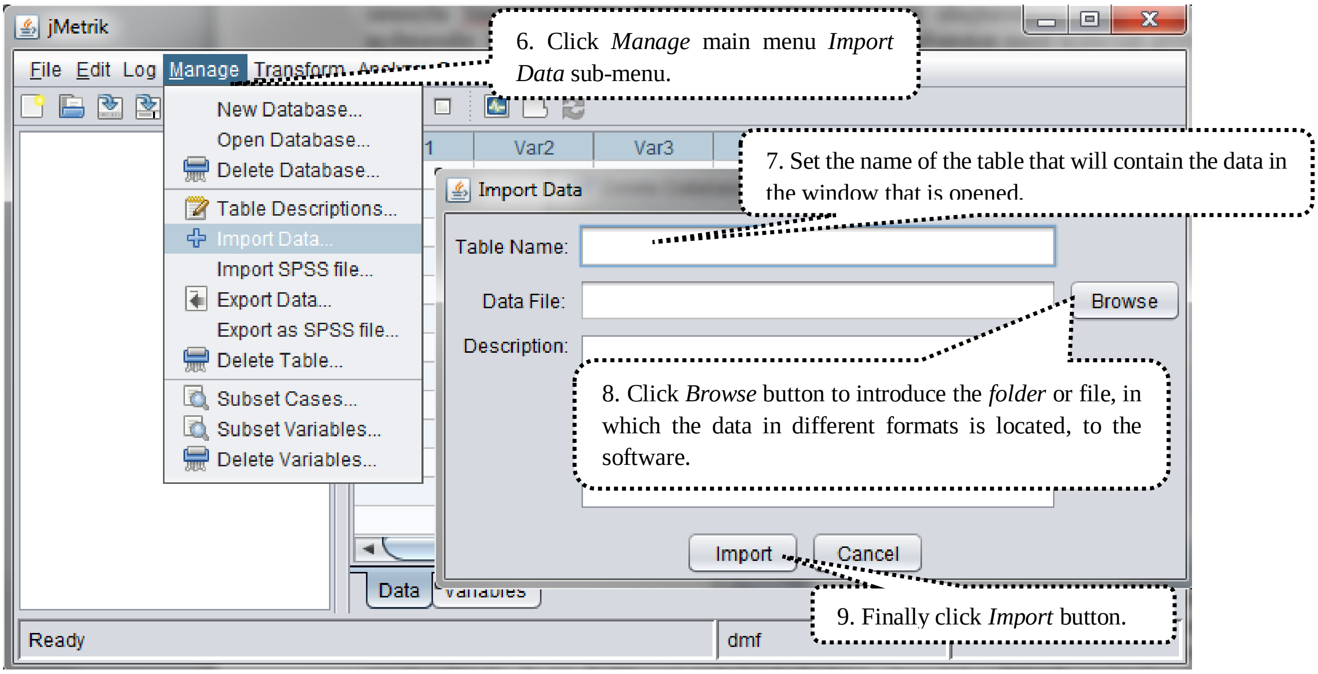


4. Select the database named dmf from already defined databases.

5. Click *Open* button.

Figure 4. jMetrik Software Database Opening Window

After this process, the name of the created database will appear in the open database window as in the Figure 4. After opening the database, the data file that will be used for the analysis should be imported into the software. jMetrik software can process different data types easily, in order to transfer data to jMetric, select Manage >> Import data. After this operation, the main screen shown in Figure 5 will appear.



6. Click *Manage* main menu *Import Data* sub-menu.

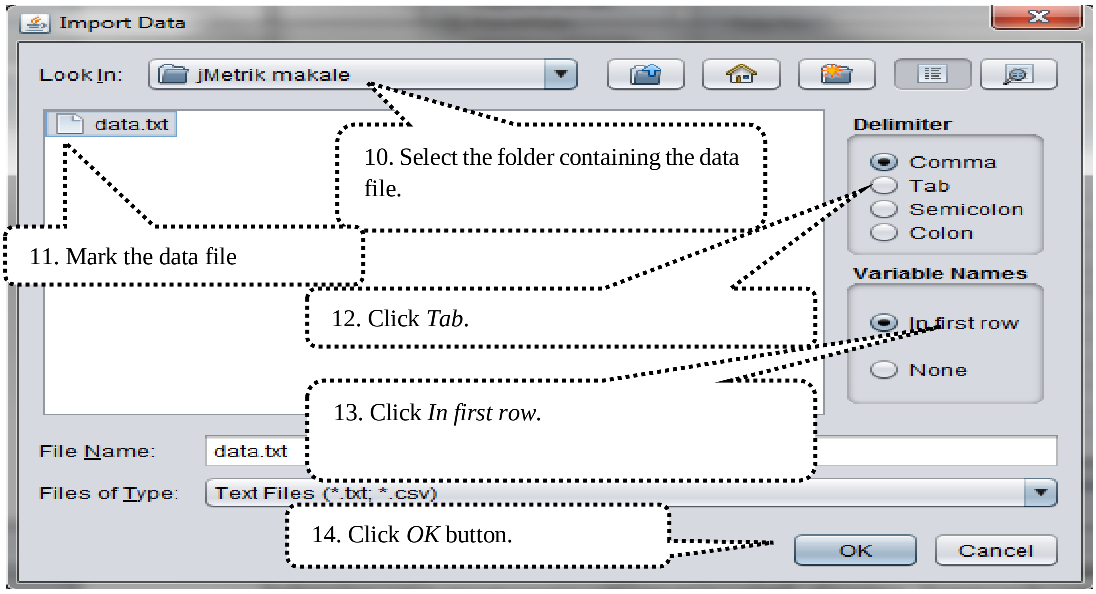
7. Set the name of the table that will contain the data in the window that is opened.

8. Click *Browse* button to introduce the *folder* or file, in which the data in different formats is located, to the software.

9. Finally click *Import* button.

Figure 5. jMetrik Software Data Transfer Window

In order to transfer already prepared data file to the software, click the Browse button shown in Figure 4 and make the definitions in the window that is opened. In the data definition window shown in Figure 6, enter the information such as type of data file and how to define the data and if first row contains data names.



10. Select the folder containing the data file.

11. Mark the data file

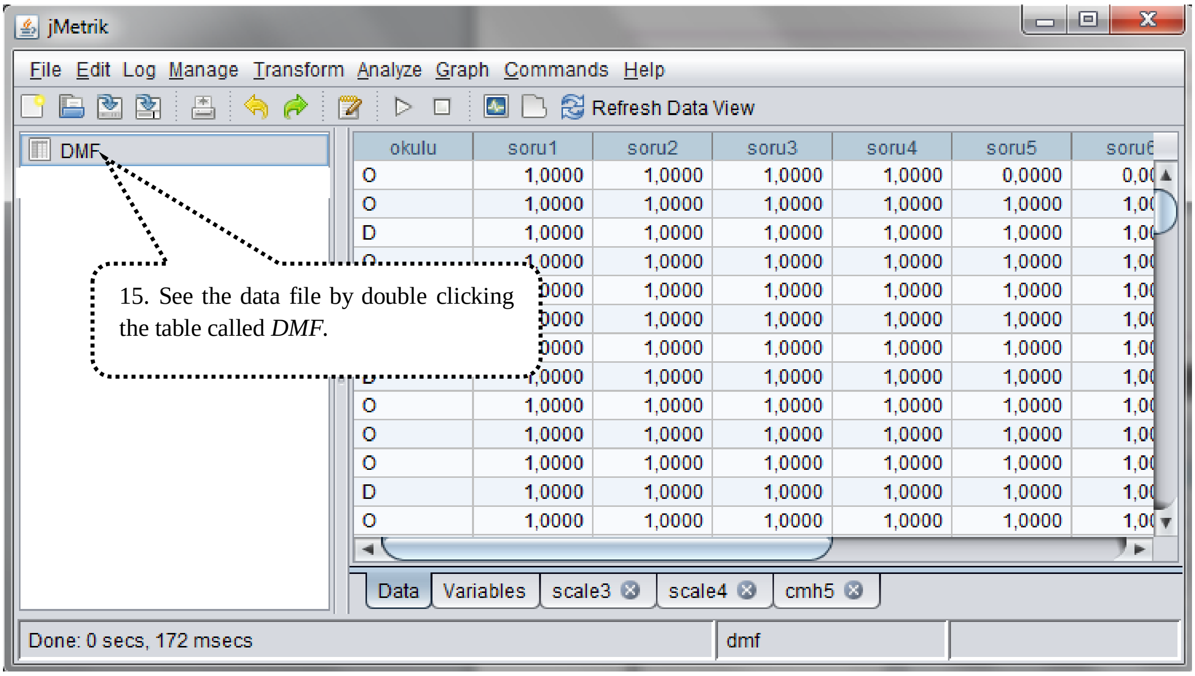
12. Click *Tab*.

13. Click *In first row*.

14. Click *OK* button.

Figure 6. jMetrik Software Data Definition Window – I

After the necessary definitions, click the OK button and the table containing the data will be imported to the software. If the import operation is carried out correctly, the table named “dmf” will be shown in the window on the left side of the jMetrik main screen. Sample data file contains the answers given to a 10-question math test by the students from two different schools, defined as state and private school. The data file can be seen by double-clicking the dmf table in the window shown in Figure 7.



okulu	soru1	soru2	soru3	soru4	soru5	soru6
O	1,0000	1,0000	1,0000	1,0000	0,0000	0,0
O	1,0000	1,0000	1,0000	1,0000	1,0000	1,0
D	1,0000	1,0000	1,0000	1,0000	1,0000	1,0
O	1,0000	1,0000	1,0000	1,0000	1,0000	1,0
[illegible]	0000	1,0000	1,0000	1,0000	1,0000	1,0
[illegible]	0000	1,0000	1,0000	1,0000	1,0000	1,0
[illegible]	0000	1,0000	1,0000	1,0000	1,0000	1,0
D	1,0000	1,0000	1,0000	1,0000	1,0000	1,0
O	1,0000	1,0000	1,0000	1,0000	1,0000	1,0
O	1,0000	1,0000	1,0000	1,0000	1,0000	1,0
O	1,0000	1,0000	1,0000	1,0000	1,0000	1,0
D	1,0000	1,0000	1,0000	1,0000	1,0000	1,0
O	1,0000	1,0000	1,0000	1,0000	1,0000	1,0

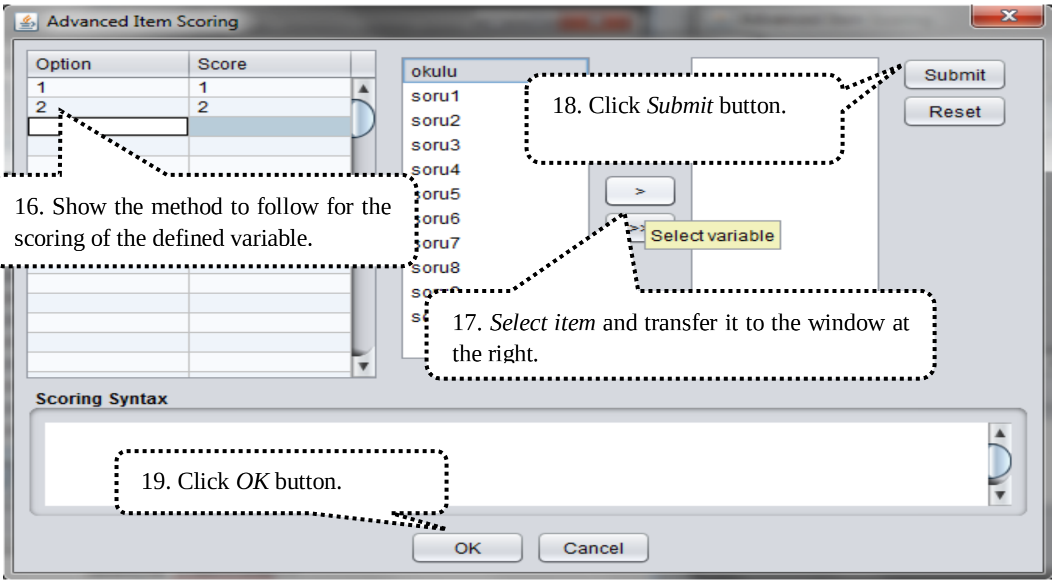
15. See the data file by double clicking the table called *DMF*.

Done: 0 secs, 172 msecs

dmf

Figure 7. jMetrik Software Data Definition Window – II

Researchers will encounter a window with which they are familiar from different data analysis software. In this window there are two tabs; the one called Data contains the data and the other called Variables contains the variables. If required, researchers may easily correct missing or incorrect data from this window. The most important process to be done after the introduction of the data to jMetrik software is defining how to score this data. Otherwise, jMetrik software will not perform any analysis. For this reason, Transform >> Advance Item Scoring steps should be followed and the way of scoring the data should be defined. At this stage, meaning of the numbers in the options must be defined in this step for each variable, as shown in the main screen illustrated in Figure 8.



16. Show the method to follow for the scoring of the defined variable.

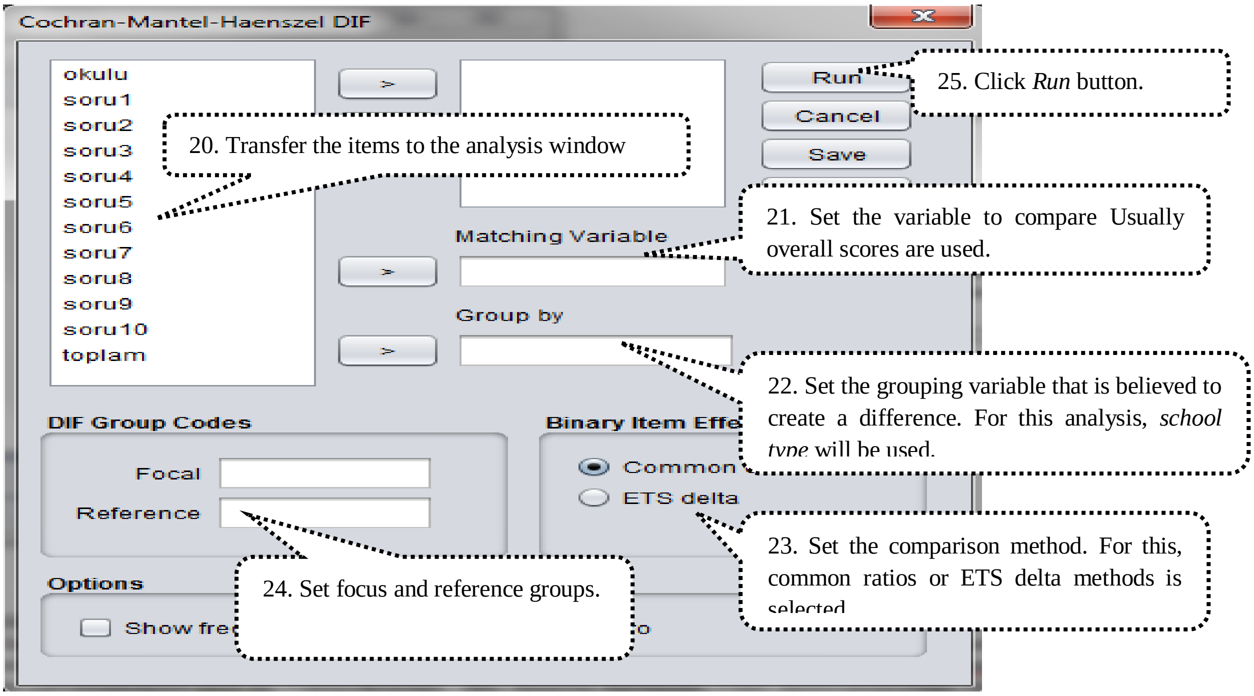
17. *Select item* and transfer it to the window at the right.

18. Click *Submit* button.

19. Click *OK* button.

Figure 8. jMetrik Software Item Scoring Window

Similarly, the items included in the test are scored as 1-0 in terms of being right and wrong, and OK button is clicked. Differential Item Functioning (DIF) analysis performed as an example in the study addressing whether the responses given to 10 items differ significantly according to students' school type (state or private) or not. The steps to be followed for DIF analysis referred in the context of the bias study, which is considered to be one of the evidences to be presented regarding the item validity, are Analyze>>DIF: Mantel-Haenszel steps. You will then see the main screen shown in Figure 9.



20. Transfer the items to the analysis window

21. Set the variable to compare Usually overall scores are used.

22. Set the grouping variable that is believed to create a difference. For this analysis, *school type* will be used.

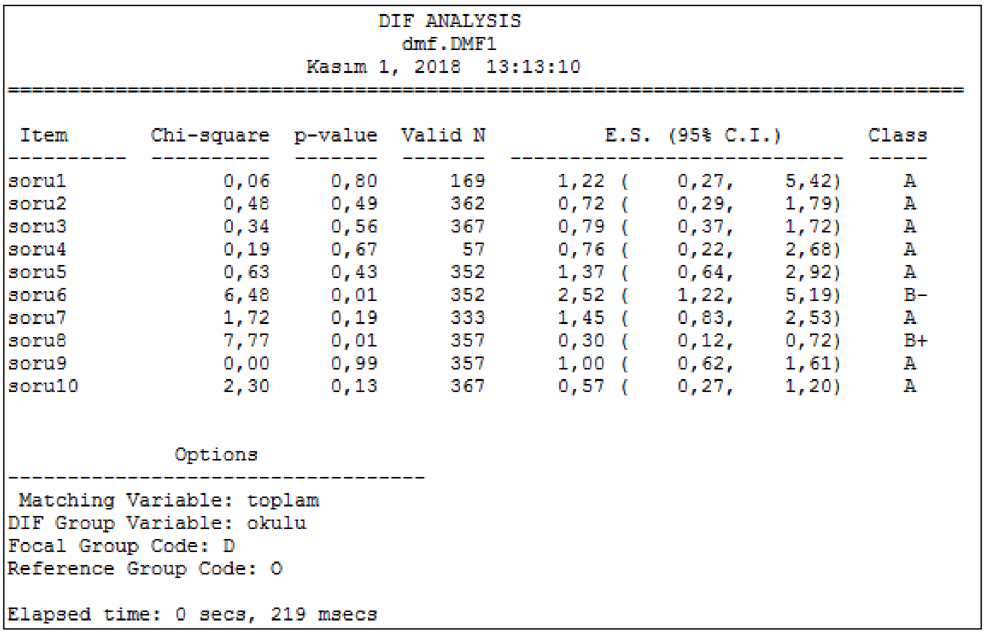
23. Set the comparison method. For this, common ratios or ETS delta methods is selected

24. Set focus and reference groups.

25. Click *Run* button.

Figure 9. jMetrik Software DIF Analysis Window

Test scaling menu is used to create the total scores in jMetrik software. In DIF analysis, the total scores obtained from all items will be used as the comparison variable; after the necessary descriptions, RUN button is clicked to complete the analysis process. The result of the analysis is the Output file shown in Figure 10.



```
                               DIF ANALYSIS
                                 dmf.DMF1
                         Kasım 1, 2018  13:13:10
================================================================================

 Item       Chi-square  p-value  Valid N        E.S. (95% C.I.)        Class
----------  ----------  -------  -------  ----------------------------  -----
soru1             0,06     0,80      169      1,22 (    0,27,    5,42)     A
soru2             0,48     0,49      362      0,72 (    0,29,    1,79)     A
soru3             0,34     0,56      367      0,79 (    0,37,    1,72)     A
soru4             0,19     0,67       57      0,76 (    0,22,    2,68)     A
soru5             0,63     0,43      352      1,37 (    0,64,    2,92)     A
soru6             6,48     0,01      352      2,52 (    1,22,    5,19)     B-
soru7             1,72     0,19      333      1,45 (    0,83,    2,53)     A
soru8             7,77     0,01      357      0,30 (    0,12,    0,72)     B+
soru9             0,00     0,99      357      1,00 (    0,62,    1,61)     A
soru10            2,30     0,13      367      0,57 (    0,27,    1,20)     A


              Options
-----------------------------------
 Matching Variable: toplam
DIF Group Variable: okulu
Focal Group Code: D
Reference Group Code: O

Elapsed time: 0 secs, 219 msecs
```

Figure 10. jMetrik Software DIF Analysis Results Window

As a result of DIF analysis, it was found that question 6 has a moderate differential item functioning in favor of the focus group, whereas question 8 has a moderate differential item functioning in favor of the reference group. Other questions in the test were found to have negligible DIF (A).

Support for the Software

To get an insight about the software, new users can read quick start guide that can be found in https://itemanalysis.com/jMetrik-quick-start/ address. For reaching frequently asked questions about the software and the answers https://itemanalysis.com/jMetrik-faq/ can be visited, whereas https://groups.google.com/forum/#!forum/jMetrik-user-group can be visited for more detailed question. *Applied Measurement with jMetrik*, which was written by Dr. Meyer, the developer of the software, is the source book containing the theoretical information about CTT and IRT and the sample analysis of the software. The book can be used as a guide containing general information about CTT and IRT and the use of jMetrik. The book consists of 10 chapters, namely *Data Management*, *Item Scoring*, *Test Scaling*, *Item Analysis*, *Reliability*, *Differential Item Functioning*, *Rasch Model*, *Multi-Category Rasch Models*, *Graphical Representation of Item and Test Properties*, *IRT-Based Scale Linking and Score Equalization.*

RESULTS and DISCUSSION

jMetrik is a software by which CTT and IRT based data analysis can be performed under a single roof without requiring any other software. IRT analysis of two-category and multi-category items can be carried out by jMetrik. Differential item function analysis can be performed based on IRT. Analyzes that can be performed using pop-up windows help researchers to perform analyzes very easily. In addition, each analysis can be carried out through commands. Being an open source project, jMetrik is a software that can be installed and used at no cost. Owing of being open source, either the codes created by people contributing to jMetrik can be used or the researchers can create their own codes. In this way, the researchers can contribute both to themselves and to other researchers who use jMetrik across the world. It is thought that being free of charge puts the software one step ahead of the other software packages that can perform the same analyzes but can only be used by paying a fee. *Applied Measurement With jMetrik*, the book written by the developer of the software can be used as a reference guide for the software.

In addition to these advantages, some aspects of jMetrik should be developed. Currently, the calculations of Multidimensional Item Response Theory (MIRT) cannot be performed within scope of the software. But given that the software is open source and therefore the researchers have the opportunity to contribute to the development of the software, this limitation is thought to be overcome easily in the future. Although the reference book *Applied Measurement With jMetrik,* which has been written by the developer of the software, contains information about how to use the software, the version used while writing the article is different from the current version, therefore, a very comprehensive resource, including information on additional analyzes and software features in the current version, will be helpful for users. The web address created for reaching frequently asked questions and answers about the software and the web address created for answering more detailed questions about the software need further improvements. Consequently, considering the advantages and disadvantages of jMetrik it is thought that the software will help researchers in answering many problems related to CTT and IRT and the execution of the application; it will reduce the workload of the researchers considering that it is free and open source and the analysis can be performed easily and quickly; having knowledge about the formulas working in background algorithms in account of being open source and the quality of the outputs in terms of readability will also contribute to ease researchers' workload. Psychometry and measurement software review studies in the future, can be carried out considering the strength and weaknesses of the parameters related to the analysis that can be performed by the software to be compared.

In addition, the jMetrik program reports the results of the analysis on its own interface compared to the other IRT analysis programs. As with other software, researchers can easily move these outputs to their work areas. In addition, the parameters obtained in jMetrik are very similar to the programs such as IRTPRO, BILOG and PARSCALE. Since the theoretical foundations of the program are inspired by the most commonly used IRT programs in the literature, it provides a great advantage to the users in interpreting and reporting the results of the analysis. In relation to that Aksu, Reyhanlıoğlu and Eser (2017) found that the results obtained from BILOG, IRT PRO and JMETRİK programs have correlation values of .99 and above in terms of both item parameters and ability estimations and it is an indication of how the results of the program are consistent with other programs used in the field. jMetrik program can perform analyzes in a very short time compared to other programs. The only negative feature related to the program is that the analysis cannot be performed without performing the database creation process which is not defined in other IRT programs. As a matter of fact, since jMetrik is Java based, researchers should first create a database where they can perform analyzes and then transfer their data to this database.

REFERENCES

Aksu, G., Reyhanlıoğlu, Ç., Eser M. T. (2017). Examining the two categorical datas by jMetrik, Bilog-MG and IRTPRO with application of mathematics exam. *European Scientific Journal*, 13 (33), 20-43. doi: dx.doi.org/10.19044/esj.2017.v13n33p20

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Orlando, FL: Holt, Rinehart & Winston.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ, US: Lawrence Erlbaum Associate, Inc.

Haebara, T. (1980). Equating logistic ability scales by a weighted least squares method. *Japanese Psychological Research*, 22(3), 144–149.

Hambleton, R. K., & Swaminathan, H. (1985). Item response theory principles and applications. Boston-USA: Kluwer-Nijhoff Publishing.

Kim, S. & Kolen, M. J. (2007). Effects of scale linking on different definitions of criterion functions for the IRT characteristic curve methods. *Journal of Educational and Behavioral Statistics*, *32*(4), 371–397.

Lord, F. M., Novick, M. R., & Birnbaum, A. (1968). *Statistical theories of mental test scores*. Oxford, England: Addison-Wesley.

Loyd, B. H. & Hoover, H. D. (1980). Vertical equating using the Rasch model. *Journal of Educational Measurement*, *17*(3), 179–193. doi: http://dx.doi.org/10.1111/j.1745-3984.1980.tb00825.x

Marco, G. L. (1977). Item characteristic curve solutions to three intractable testing problems. *Journal of Educational Measurement*, *14*(2), 139–160. doi: http://dx.doi.org/10.1111/j.1745-3984.1977.tb00033.x

McDonald, R. P. (1999). *Test theory: A unified treatment.* Mahwah, NJ: Lawrence Erlbaum Associates.

Meyer, J. P. (2010). Understanding measurement: Reliability. New York: Oxford University Press.

Meyer, J. P. & Hailey, E. (2012). A study of Rasch partial credit, and rating scale model parameter recovery in WINSTEPS and jMetrik. *Journal of Applied Measurement*, *13*(3), 248–258.

Meyer, J. P. (2014). *Applied Measurement with jMetrik*. New York: Routledge.

Meyer, J. P. (2018). jMetrik. In W. van der Linden (Ed.). *Handbook of Item Response Theory* (pp.557-567). Boca Raton, FL: Taylor & Francis.

Stocking, M. L. & Lord, F. M. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement*, *7*(2), 201–210. doi: http://dx.doi.org/10.1177/014662168300700208

Wright, B. D. & Masters, G. N. (1982). *Rating scale analysis*. Chicago, IL: MESA Press.