Software Review

jMetrik item analysis [software application]. Patrick Meyer. Retrieved from http://www.itemanalysis.com/

REVIEWED BY:
Andrea Gotzmann, Ph.D.
Medical Council of Canada

Louise M. Bahry, M.Ed.
University of Massachusetts, Amherst

Technology, and the use of software to enhance or assist with evaluating measurement statistics, is currently a large emphasis for users. Measurement statistics, used in classical test theory (CTT) and item response theory (IRT), have been elusive for some users, as the measurement concepts are complex and investment of time to understand is intensive (Lord, 1980; Lord & Novick, 1968). However, users across many content disciplines are developing their understanding and applying these methodologies to new areas (i.e., medical education, psychology, etc.). As a result, the needs of researchers and applied practitioners have changed, and consequently, require tools to apply psychometrics. Reliance on specialized or esoteric software has been the norm; however, according to Drasgow, Luecht, and Bennett (2006), "Technology offers solutions to many of the challenges faced by testing programs" (p. 471). That is, technology may provide many of the psychometric analyses to be more accessible to broader audiences, so that users of all levels of expertise can take advantage of the advances in educational measurement.

Technology may provide many of the psychometric analyses to be more accessible to broader audiences, so that users of all levels of expertise can take advantage of the advances in educational measurement.

There are many statistics that are calculated and evaluated for both CTT and IRT, and many of them housed in separate programs. Some testing organizations create proprietary software to integrate these analyses, or create user defined solutions with programs such as R, SAS or SPSS. This review describes and evaluates a new software program, called jMetrik, version 2.1.0, that can produce psychometric statistics for both CTT and IRT. jMetrik allows for a more integrated system to conduct psychometric analyses for research and operational purposes without the cost associated with some other psychometric software programs.

Program Description

The jMetrik software was developed by J. Patrick Meyer, Ph.D., with support from the National Council on Measurement in Education (NCME), as a recipient of the Bradley Hanson Award for Contributions to Educational Measurement in 2010. jMetrik is a free and open source software application for classical and modern psychometric analyses. The program is a pure Java application that runs on Windows, Mac, OSX, and Linux platforms, with requirements of 256 MB of available memory, and Java 6 (i.e., JRE 1.6) or higher. The jMetrik graphical user interface (GUI) combines a workspace tree, data view, point-and-click menu, and several dialog boxes. Although the software is currently available not all features are active, or fully functional. Therefore, this review will address the features that currently are available, offering a snapshot of the current version of the software.

Current Available Analyses and Program Interface

The jMetrik software includes psychometric analyses such as CTT, IRT, Differential item functioning (DIF), and Confirmatory Factor Analysis (CFA). All of these analyses are useful in evaluating the psychometric quality of an assessment. In addition, the program offers many graphing features such as histograms and nonparametric IRT characteristic curve estimation.

The data interface to import data is similar to Microsoft Access. Data cannot be entered directly into the program; to import, the data need to be structured as a comma or space delimited text file, where missing values must represented by two consecutive commas or spaces and missing data is automatically scored as zero. The user needs to score the data and indicate the score key for each item as well as missing data; this step is required for both scored and un-scored data. There is a replication tool provided in the point-and-click interface, and an option to write code to simplify this process. There is no option to read in a key file to correspond with importing the data file. In addition, if one wants to change the score process the user must reimport and score the data again. The import step was somewhat difficult to implement. However, once the data was imported and scored, the remaining psychometric analyses were relatively simple.

Classical Test Theory Analyses

The CTT analyses included item and test analysis, and test scaling. The classical item analysis includes options such as the item statistics, reliability analyses, and conditional standard error of measurement (CSEM). The interface to generate this output is relatively easy to navigate. The output can be saved as a text file and includes all of the item statistics, test statistics, and reliability analysis. In the output each of the multiple choice (MC) and constructed response (CR) item options are provided which include the item difficulty, standard deviation, and two types of item correlations. In addition, there are five different methods of calculating reliability provided: Guttman's Lambda, Cronbach's Alpha, Feldt-Gilmer, Feldt-Brennan, and Raju's Beta.

Item analysis procedures also provide decision consistency and accuracy estimates: Huynh's Raw Agreement, Huynh's Kappa, KR-21, Beta-binomial alpha, and Beta-binomial beta. However, details on the method of calculation for some procedures are not clear. For example, the output has the Item-Total Pearson and Correlation Polyserial headers, but it is unclear if either option includes the current item in its correlation calculation. Another example is the CSEM, which is an option, but it is unknown which method is being used to calculate the results. The classical item analyses provided in jMetrik are quite comprehensive, useful to all levels of users both in research and applied settings. Documentation on procedures would be helpful but overall the interface is easy to use and output is complete.

The test scaling options are also easy to use and include many options. The user can quickly convert data to sum, percentile rank, Kelley True, and normalized scores. Users can specify constraints on the minimum, maximum, and precision points, as well as converting to a custom linear transformation. The program allows for renaming the new scored variables; however, only one transformation can occur in one run and once transformed there is no ability to rename variables. The test scaling features in jMetrik are useful, easy to use and understand. In summary, the CTT analyses provided by the program provide a simple point-and-click interface similar to SPSS and EXCEL, but provide the much needed psychometric analyses in one program that currently is unavailable for the psychometric specialist.

jMetrik allows for a more integrated system to conduct psychometric analyses for research and operational purposes without the cost associated with some other psychometric software programs.

Item Response Theory Analyses

The IRT analyses provided by jMetrik include calibration of the Rasch, Partial Credit and Rating Scale model. This program also has options for the calibration, item, and persons, which include the convergence criteria, scoring of missing data, fixing item parameter values, saving theta estimates and person fit statistics. The basic convergence options are available such as number of iterations, and convergence criteria, but the software is missing options such as estimation method, and theta estimate is produced (i.e., Expected a Priori, Maximum a Posteriori). The IRT statistics are saved in the data file, along with the original data sources, which can be exported to a text file.

IRT output includes the final item parameters, standard errors, WMS and UMS (which are assumed to stand for Weighted Mean Square and Unweighted Mean Squares), a score conversion table, and scale quality statistics. The output provided is helpful in evaluating the quality of the IRT analyses, but is missing some statistics that might be useful, such as item fit statistics. In addition, non-parametric item characteristic curves are provided in the program. Figure 1 shows an example item characteristic curve. The item characteristic curves can be printed or saved for each item or graph. There is no option to automatically save or print more than one graphic at a time.

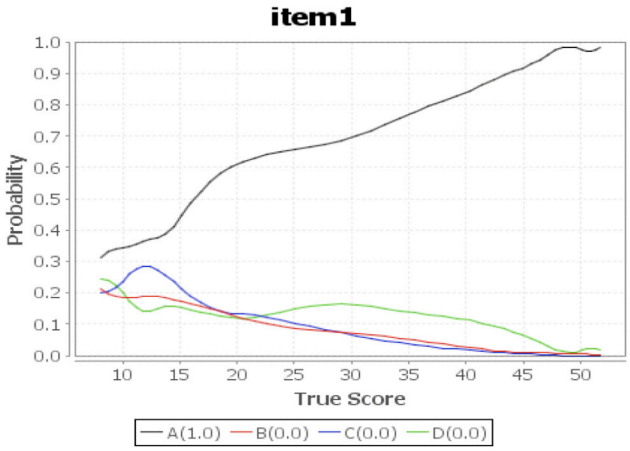

Figure 1: Sample item characteristic curve from jMetrik

In addition to the IRT calibration procedures, jMetrik can produce IRT equating results. The item parameters and thetas for the two sets of results need to be imported separately and in a particular format. Both item parameters and theta values along with weights are required to conduct the IRT equating and the program assumes that the Form Y values are the base form, and the Form X values will have the transformed theta and item parameters. The theta data file needs to have the theta value and weight, and the item parameter data file requires the item name and difficulty parameter.

The IRT equating output provides a robust z test, item summary statistics, and the equating coefficients for the mean/sigma, mean/mean, Haebara, and Stocking-Lord equating procedures. The IRT equating provides quick results and transformed both the item parameters and theta estimates. Test information graphics, or correlations of the item parameters, are not provided for before and after equating and the user needs to re-run each equating method if transformations are required from more than one equating procedure to calculate new theta values.

Differential Item Functioning

The non-parametric DIF analyses available are the Mantel Haentzel (MH) and non-parametric characteristic curves (CC) methods. The data for these analyses can be in a single file with an identifying variable indicating group membership. MH analyses provides two effect size measures and allows for matching either on observed or transformed Kelley scores. The CC analyses have options

to match on normal, true scores as well as several kernel regression options. The output is complete but difficult to interpret without the guidance of the FAQ's from the jMetrik website. There are many acronyms and unfamiliar terms that are not explained in the DIF output.

Documentation

The jMetrik software provides a quick start guide on the website that covers importing data, variable information, scoring item responses, and handling of missing data. There is also a ten-hour training session that is provided in January and August for a cost of $300. Unfortunately, a detailed user-manual was not available at the time of this review. The options provided in the software were fairly expansive, but full details on technical jargon are not outlined, which could be difficult to follow for new or occasional psychometric users. A detailed user-manual would provide much needed guidance and clarification and would enhance understanding of a useful program that users at all levels of expertise could appreciate.

The classical item analyses provided in jMetrik are quite comprehensive, useful to all levels of users both in research and applied settings.

Evaluative Conclusion

jMetrik provides psychometric software consumers with a program that offers a comprehensive assortment of analysis options. Current limitations in functionality of the program do detract from many of the features offered, but with revisions, the package could offer a user-friendly interface with complete psychometric analyses. For example, the CFA analyses offer several of the commonly used fit statistics, but some of the estimation procedures and options are under construction with options greyed out. Once all functionality is available a more thorough evaluation would be possible and the program should be very useful for the novice and expert psychometricians.

The lack of a user-manual makes it difficult to conduct analyses beyond data importing and scoring and interpret output. Additionally, a user-manual would allow for users to trouble-shoot some of the simple software errors as opposed to emailing the software developer for the solutions.

References

Drasgow, F., Luecht, R. M., & Bennett, R. E. (2006). Technology and testing. In R. L. Brennan (Ed.), *Educational measurement* 4th ed. (pp. 471-515). Westport, CT: Praeger Publishers.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores.* Reading MA: Addison-Welsley Publishing Company