CutScore Manual

Christopher Runyon

October 2019

Thank you for your interest in using CutScore, an app that allows users to implement the cut-score operating functions discussed in Grabovsky and Wainer (2017a), Grabovsky and Wainer (2017b), and Grabovsky, Pace, and Runyon (forthcoming). The current version of the program (0.0) provides users with the optimal cut score with respect to total classification error, maximum classification error, conditional classification error, total penalized classification error, and maximum penalized classification error. Future versions of the program will implement the weighted classification error function, and the user interface and other features will be included.

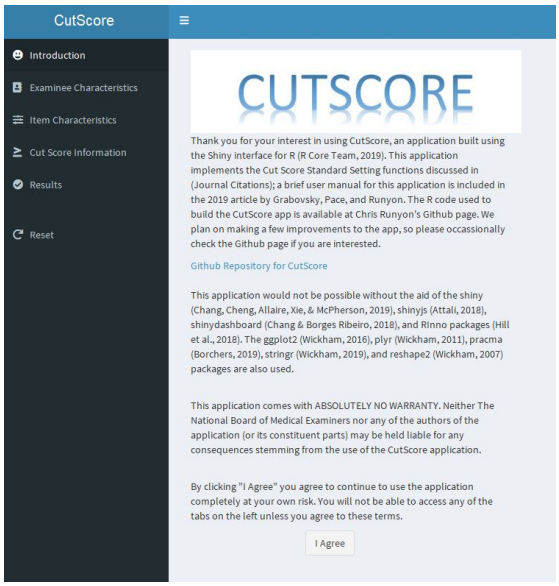


Figure 1: Introductory page.

Upon opening the app you will be greeted with an Introduction page that provides information about the program and the software that was used for its creation. At the bottom of that introductory page is a box labeled “I Agree”. Clicking on that box indicates that you agree to use the program at your own risk. **You must click on "I Agree" to use rest of the program.** The other sections of the app (accessible by clicking on the tabs on the left panel) will be blank pages until this box is clicked.

The calculation of the optimal cut-score depends on knowing the distribution of examinee abilities, the item difficulties for the items to be used on the test, the estimated reliability of the test, and a hypothesized best cut score (or range of cut scores). The tabs on the left panel of the app and the sections in this document correspond to these requisite sources of information.

Examinee Characteristics

The cut-score operating function (and thus the app) conceptualize examinee ability as a latent variable measured on the theta (θ) scale, as is common in item response theory. To calculate the optimal cut score the mean and standard deviation of the examinee abilities is required. As per the formulae in Grabovsky and Wainer (2017a), we assume that this ability distribution is approximately normal.

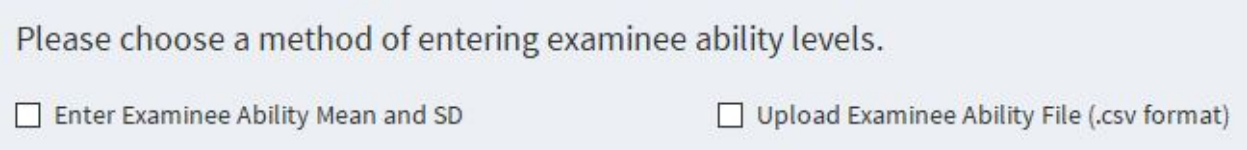


Figure 2: Examinee Characteristics.

The examinee ability mean and standard deviation may be entered two ways. The first method is to simply enter this information manually by simply typing the mean and sd in the appropriate boxes. The limits of the mean ability range from -4 to 4, and the limits of the standard deviation are 0.1 to 2. This is not meant to imply that the cut score function will work with extreme values; we want to provide this information for documentation's sake.

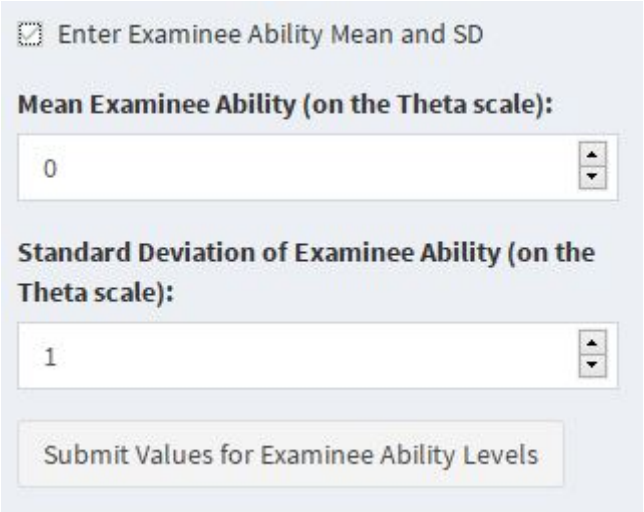


Figure 3: Manual entry of examinee ability estimates.

The second method for entering examinee abilities is to upload a .csv file **where the first column is an unlabeled column (i.e., without a column header) of examinee abilities**. That is, the .csv file should have only the examinee ability levels in the first column of the file, with no additional information or column headers. **A known glitch of the program is clicking the "Submit Values for Examinee Ability Levels" button without first uploading a .csv file.** This causes the program to crash and it will need to be restarted prior to its use. If you have accidentally clicked the "Upload Examinee Ability File (.csv format)" button, you only need to click the red "Reset Examinee Info" button at the top right corner of the page to back out to the previous step.

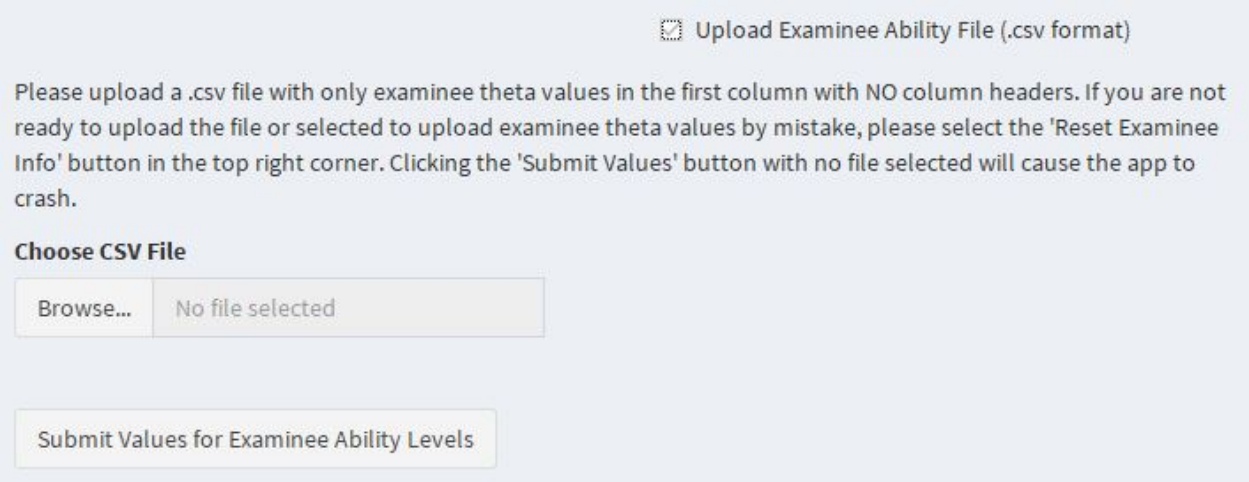


Figure 4: Examinee Characteristics.

After you have finished entering or uploading the examinee ability estimates, you should see the green "Examinee Info Complete!" box. If you do not see the green box displayed behow, you have not yet completed the entry of examinee ability levels.

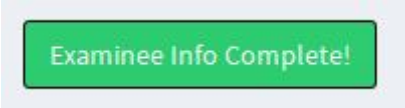


Figure 5: Confirmation of examinee ability entry.

Item Characteristics

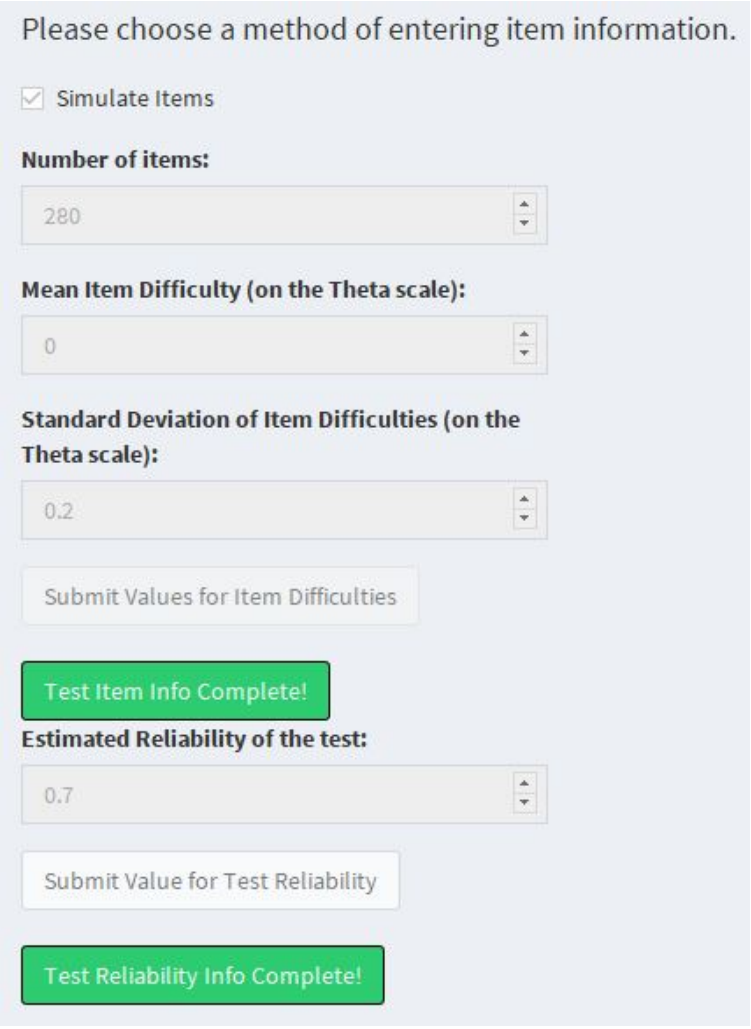


Figure 6: Manual entry of item parameters and reliability.

The information on this tab corresponds to the item difficulty and test reliability. After first entering information about the item difficulties, you will next be prompted to enter information about the reliability of the test. As with examinee characteristics, information about the item can be entered two different ways: they can be entered manually (e.g., item difficulties are simulated) or by uploading a .csv file.

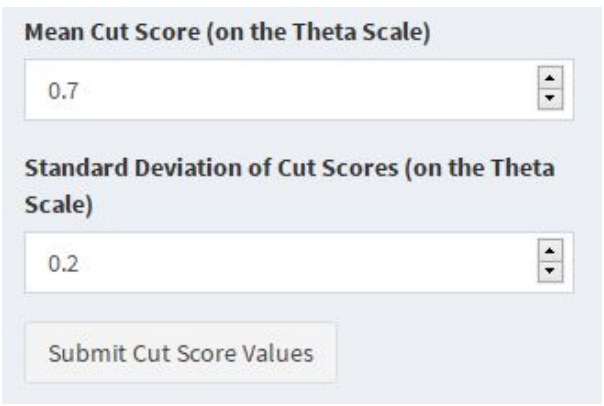
When simulating items, the program randomly draws the specified number of items from a normal distribution with the entered mean and standard deviation. Because the items are simulated, the optimal cut scores will likely change slightly if the calculations are redone with all of the same values. The limits on the mean of the Mean Item Difficulty are -4 to 4, and the limits on the standard deviaiton of item difficulties is 0.01 to 2. Once you have clicked "Submit Values for Item Difficulties", you will receive a notice that the test item information has been entered and you will be prompted to enter the estimated reliability of the test. The range of reliability is restricted to be 0.5 to 1.

Entering item information by uploading a .csv file follows the same process as uploading a .csv file for examinee abilities. A .csv file with a single column should be uploaded, **where the first column is an unlabeled column (i.e., without a column header) of item difficulties**. **A known glitch of the program is clicking the "Submit Values for Item Difficulties" button without first uploading a .csv file.** This causes the program to crash and it will need to be restarted prior to its use. If you have accidentally clicked the "Upload Item Difficulty File (.csv format)" button, you only need to click the red "Reset Item Info" button at the top right corner of the page to start over with entering item characteristics.

Cut Score Information

The cut score operating function requires that either an initial cut score or a range of cut scores be supplied to determine the optimal cut score. These cut score(s) are on also on the theta scale, so thus should be conceptualized as the minimum ability level for an examinee that has demonstrated the requisite competency. Examples for the selection of this initial cut score may be a previously-used cut score or the proposed cut score arising from a standard setting panel. In addition, the mean and standard deviation of cut scores may be submitted. This distribution of cut scores may reflect a standard-setting session where there is disagreement upon what single cut score is appropriate. Again, a green box will appear once the appropriate cut score information has been successfully submitted.

(a) One cut score.



(b) Range of cut scores.

Results

Once the information has been succesfully entered on the Examinee Characteristics, Item Characteristics, and Cut Score Information tabs, you can navigate to the Results tab and click the "Show Results" button to have the program calculate the optimal cutscore via the different methods. At the top of the page are three boxes that correspond to what information has been submitted. These boxes are reminders to help ensure that all of the required information has been submitted. As the the required information is submitted on each tab, the corresponding red box will turn into a green box for that tab's information. No results will be displayed when the "Show Results" button has been clicked unless all three boxes are green at the top of this page.

Figure 8: Submission confirmation boxes.

Once all information has been entered and the "Show Results" button has been clicked, the program will calculate the optimal cut score via the various methods. This may take a little time depending on the processing speed of your computer, although in our experience the results and figures are often generated in less than 5 seconds. The results for each of the cut score methods are displayed vertically, so users will need to use the scroll bar on the right side of the app to review the optimal cut score and graph for each method. An example is displayed on the next page.

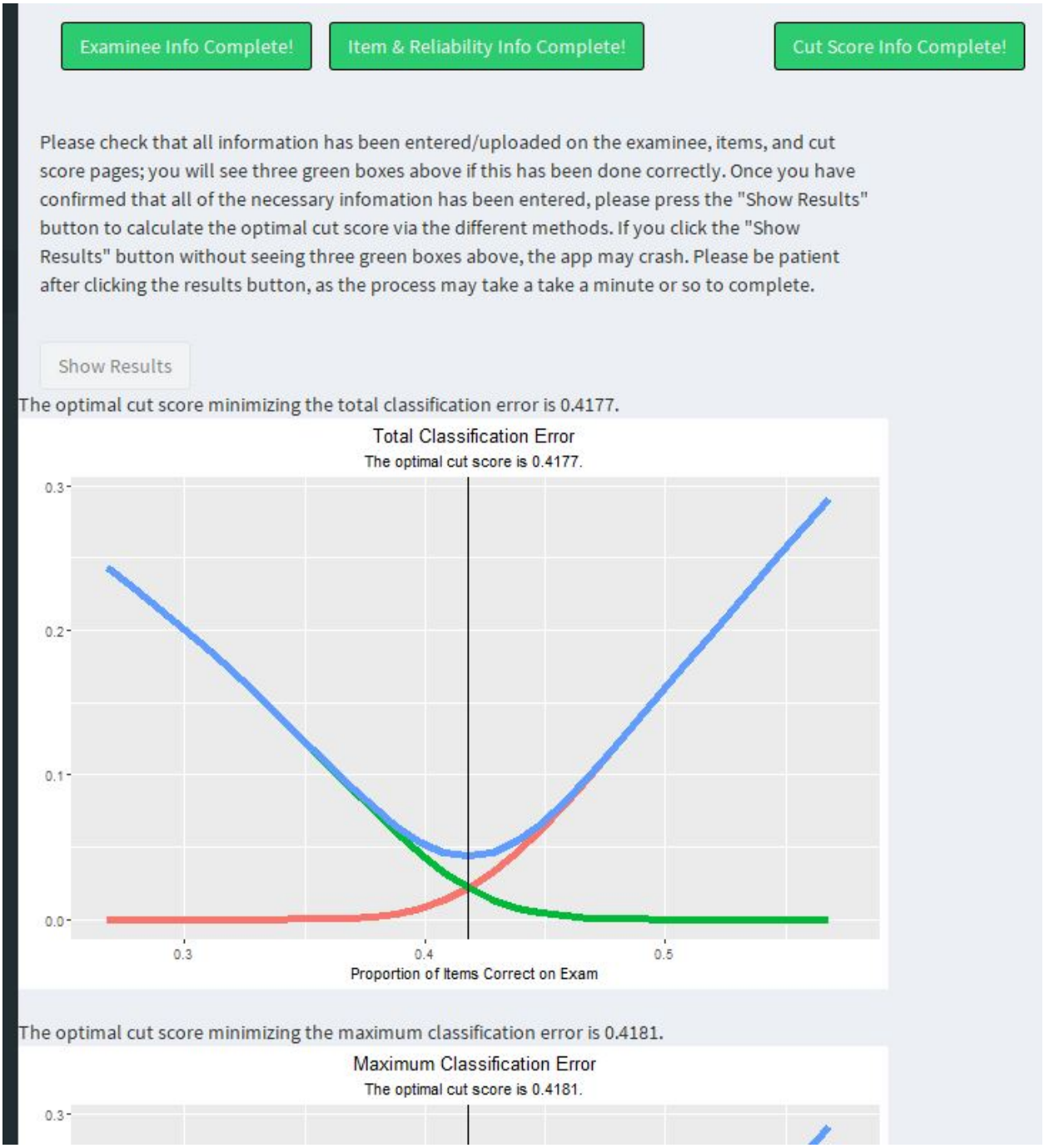


Figure 9: Example results of the cut score functions.

Reset

The reset page contains buttons to either reset the entire app or only specific tabs of the app. **Clicking the "Reset" tab header on the left navigation sidebar will not automatically reset the app; you must select the appropriate button on the page once the "Reset" tab has been selected.** The "Reset All" button resets all of the information that has been entered in the program and will virtually restart the program (you will *not* need to revisit the Introduction tab and re-agree to the terms of using the app). There are also reset buttons to reset the content on a specific tab of the program. While these buttons are also available on the tabs the reset, we also have put them on this page for convience.

Figure 10: Results of the cut score functions.

Acknowledgements

The creation of this app would not be possible without the R Project for Statistical Computing (R Core Team, 2019), as well as the shiny (Chang et al., 2019), shinydashboard (Chang and Borges Ribeiro, 2018), shinyjs (Attali, 2018), and RInno Hill et al. (2018) packages. In addition, the ggplot2 (Wickham, 2016), plyr (Wickham, 2011), pracma (Borchers, 2019), reshape2 (Wickham, 2007), and stringr (Wickham, 2019) apps were also used in some way to either calculate the cut score or aid in displaying the results.

We also graciously thank Daniel Jurich for his extensive testing and feedback on the app.

References

Attali, D. (2018). *shinyjs: Easily Improve the User Experience of Your Shiny Apps in Seconds*. R package version 1.0.

Borchers, H. W. (2019). *pracma: Practical Numerical Math Functions*. R package version 2.2.5.

Chang, W. and Borges Ribeiro, B. (2018). *shinydashboard: Create Dashboards with 'Shiny'*. R package version 0.7.1.

Chang, W., Cheng, J., Allaire, J., Xie, Y., and McPherson, J. (2019). *shiny: Web Application Framework for R*. R package version 1.3.2.

Grabovsky, I. and Wainer, H. (2017a). The cut-score operating function: A new tool to aid in standard setting. *Journal of Educational and Behavioral Statistics*, 42(3):251–263.

Grabovsky, I. and Wainer, H. (2017b). A guide for setting the cut-scores to minimize weighted classification errors in test batteries. *Journal of Educational and Behavioral Statistics*, 42(3):264–281.

Hill, J., Pang, W. L., Csárdi, G., Wickham, H., Chang, W., Hester, J., Morgan, M., and Tenenbaum, D. (2018). *RInno: An Installation Framework for Shiny Apps*. R package version 1.0.1.

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Wickham, H. (2007). Reshaping data with the reshape package. *Journal of Statistical Software*, 21(12):1–20.

Wickham, H. (2011). The split-apply-combine strategy for data analysis. *Journal of Statistical Software*, 40(1):1–29.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Wickham, H. (2019). *stringr: Simple, Consistent Wrappers for Common String Operations*. R package version 1.4.0.