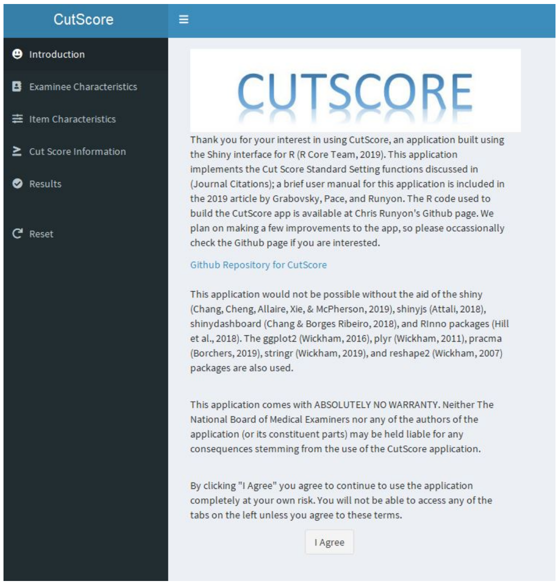
CutScore Manual

Christopher Runyon

October 2020

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, maximum penalized classification error, and weighted classification error function.



Thank you for your interest in using CutScore, an application built using the Shiny interface for R (R Core Team, 2019). This application implements the Cut Score Standard Setting functions discussed in (Journal Citations); a brief user manual for this application is included in the 2019 article by Grabovsky, Pace, and Runyon. The R code used to build the CutScore app is available at Chris Runyon's Github page. We plan on making a few improvements to the app, so please occassionally check the Github page if you are interested.

Github Repository for CutScore

This application would not be possible without the aid of the shiny (Chang, Cheng, Allaire, Xie, & McPherson, 2019), shinyjs (Attali, 2018), shinydashboard (Chang & Borges Ribeiro, 2018), and RInno packages (Hill et al., 2018). The ggplot2 (Wickham, 2016), plyr (Wickham, 2011), pracma (Borchers, 2019), stringr (Wickham, 2019), and reshape2 (Wickham, 2007) packages are also used.

This application comes with ABSOLUTELY NO WARRANTY. Neither The National Board of Medical Examiners nor any of the authors of the application (or its constituent parts) may be held liable for any consequences stemming from the use of the CutScore application.

By clicking "I Agree" you agree to continue to use the application completely at your own risk. You will not be able to access any of the tabs on the left unless you agree to these terms.

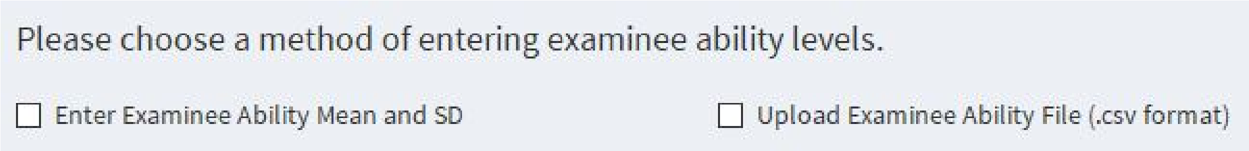
Figure 1: Introductory page.

Upon navigating to the app you will be greeted with an Introduction page that provides information about the app and the R packages that were 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 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.



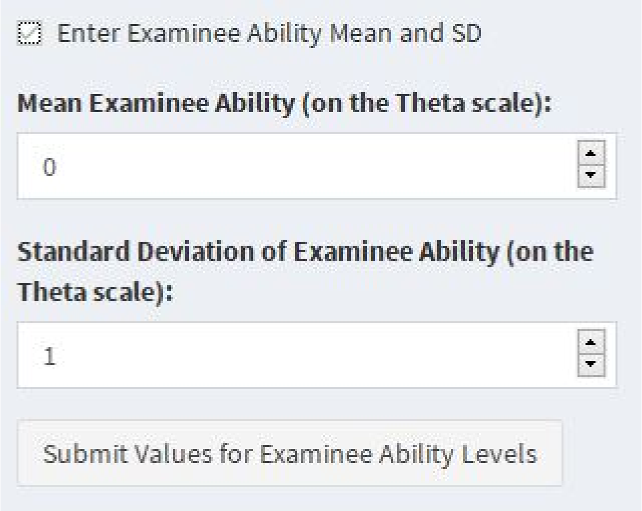
Please choose a method of entering examinee ability levels.

□ Enter Examinee Ability Mean and SD

□ Upload Examinee Ability File (.csv format)

Figure 2: Examinee Characteristics.

There are two methods to enter the examinee ability mean and standard deviation. 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 provide this information for documentation's sake.



☑ Enter Examinee Ability Mean and SD

Mean Examinee Ability (on the Theta scale):

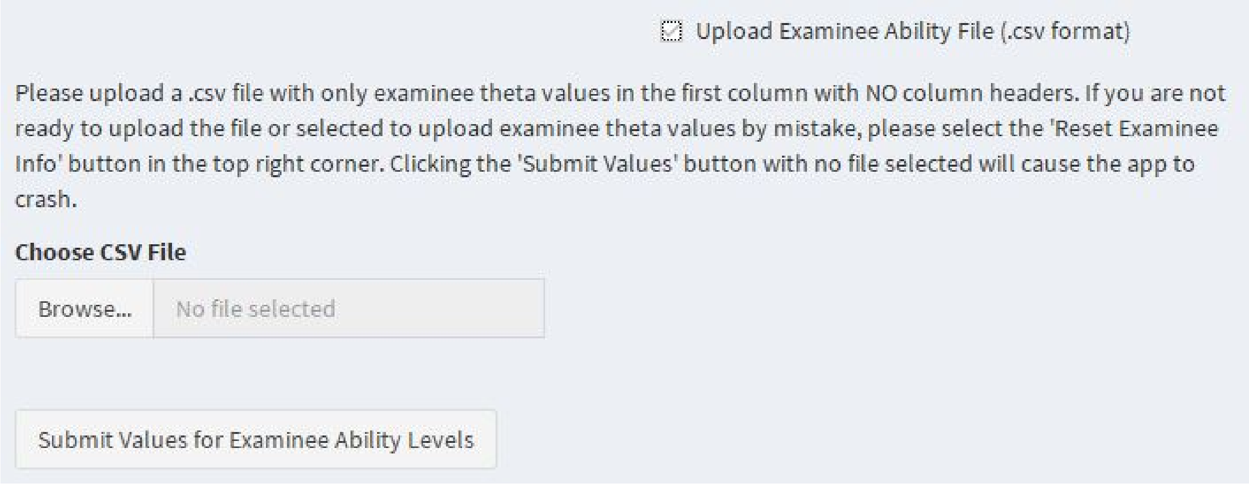
0

Standard Deviation of Examinee Ability (on the Theta scale):

1

Figure 3: Manual entry of examinee ability estimates.

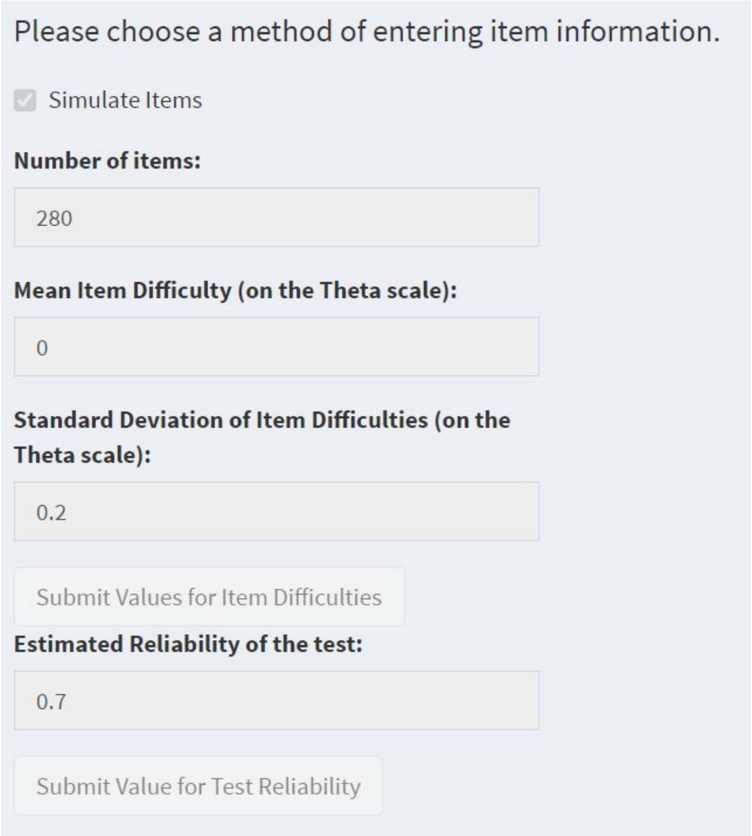
The second method for entering examinee abilities is to upload a .csv file **where the first column is an unlabeled column (i.e., without a column header) of examinee abilities**. That is, the .csv file should have only the examinee ability levels in the first column of the file, with no additional information 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 app to crash and it will need to be restarted in order to continue to use the app. If you have accidentally clicked the “Upload Examinee Ability File (.csv format)” button without uploading a .csv file, you only need to refresh the web page to back out to the previous step.


☑ Upload Examinee Ability File (.csv format)
Choose CSV File
Browse...
No file selected
Submit Values for Examinee Ability Levels

Please upload a .csv file with only examinee theta values in the first column with NO column headers. If you are not ready to upload the file or selected to upload examinee theta values by mistake, please select the 'Reset Examinee Info' button in the top right corner. Clicking the 'Submit Values' button with no file selected will cause the app to crash.

Figure 4: Examinee Characteristics.

Item Characteristics


Please choose a method of entering item information.

☑ Simulate Items

Number of items: 280

Mean Item Difficulty (on the Theta scale): 0

Standard Deviation of Item Difficulties (on the Theta scale): 0.2

Submit Values for Item Difficulties

Estimated Reliability of the test: 0.7

Submit Value for Test Reliability

Figure 5: Manual entry of item parameters and reliability.

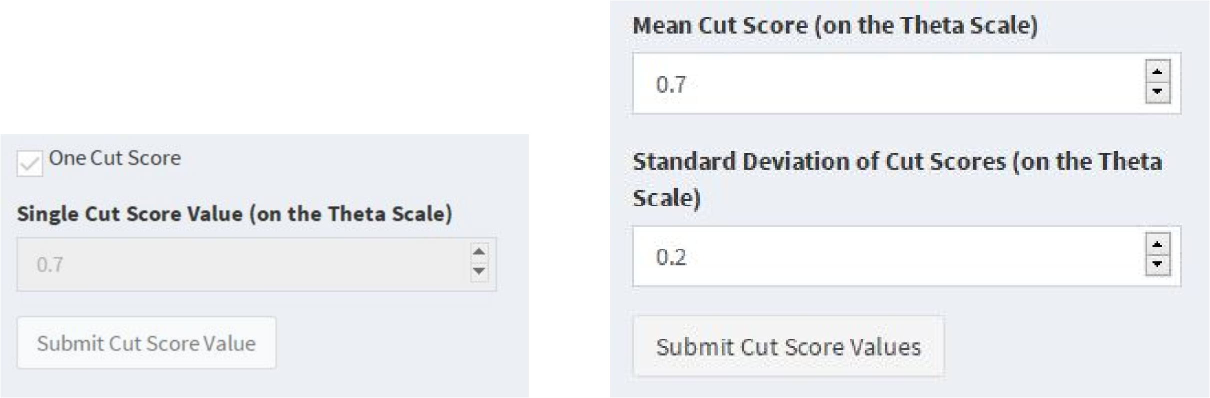
The information on this tab corresponds to the item difficulty and test reliability. After first entering information about the item difficulties, you will next be prompted to enter information about the reliability of the test. As with examinee characteristics, information about the item can be entered 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 deviation of item difficulties is 0.01 to 2. Once you have clicked “Submit Values for Item Difficulties”, you will be prompted to enter the estimated reliability of the test. The range of reliability is restricted to 0.5 to 1. The various functions do not work with low reliabilities.

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 app to crash and the webpage will need to be refreshed. If you have accidentally clicked the “Upload Item Difficulty File (.csv format)” button, you only need refresh the webpage and 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.



☑ One Cut Score

Single Cut Score Value (on the Theta Scale)

0.7

Submit Cut Score Value

(a) One cut score.

Mean Cut Score (on the Theta Scale)

0.7

Standard Deviation of Cut Scores (on the Theta Scale)

0.2

Submit Cut Score Values

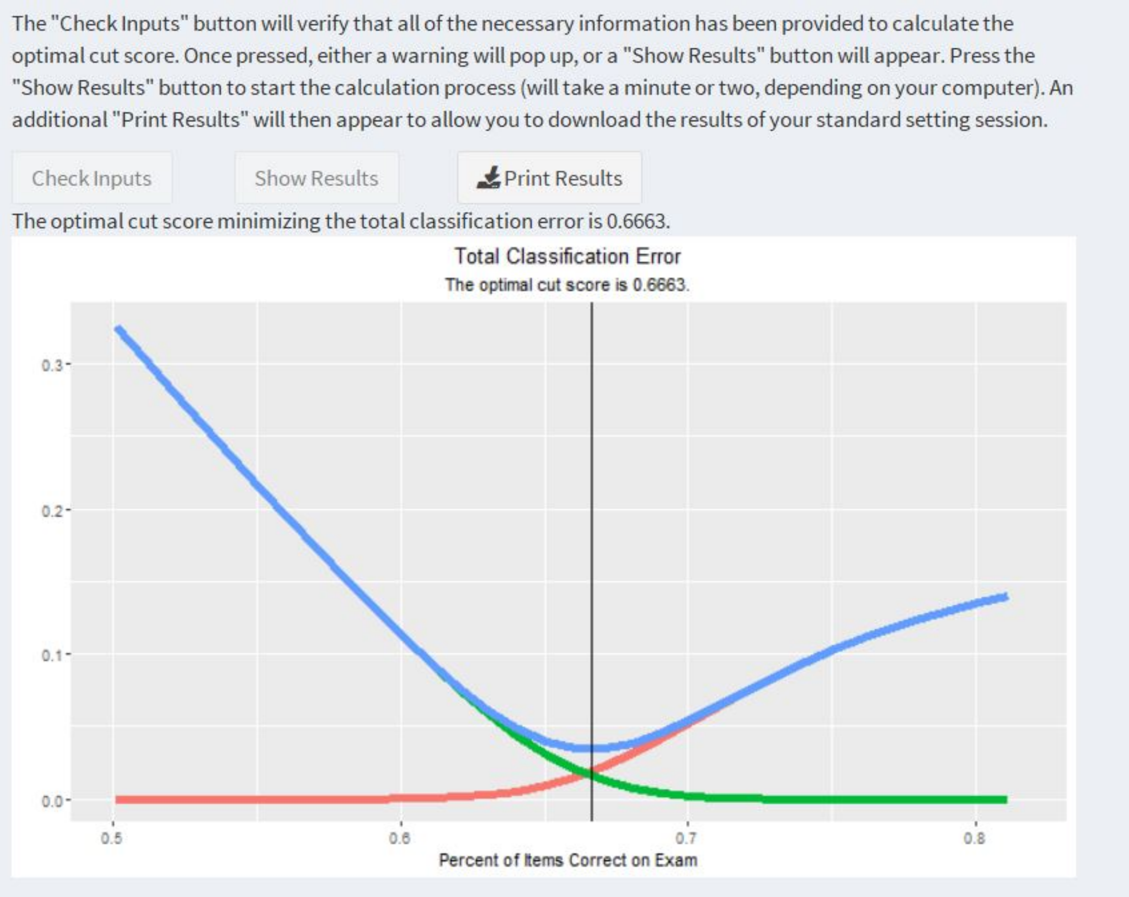
(b) Range of cut scores.

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 calculate the optimal cut scores via the different functions. It is first necessary to click the "Check Inputs" button, which will ensure that all of the necessary information on the Examinee Characteristics, Item Characteristics, and Cut Score Information tabs have been successfully submitted. If the necessary information has been successfully submitted, the "Show Results" tab will appear.

Once the "Show Results" button has been clicked, the app will calculate the optimal cut score via the various functions. 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 page to review the optimal cut score and graph for each method. An example is displayed on the next page.

After the optimal cut score values have been calculated, a "Print Results" button will appear. If you click on this button a pdf will be downloaded to your computer that is a record of your standard-setting session. This pdf includes all of inputs selected by the user, so you may double-check that all of the information was entered correctly. Each subsequent page of the pdf is optimal cut score as per one of the specified functions; this value is displayed along with a large figure of the corresponding function.



The "Check Inputs" button will verify that all of the necessary information has been provided to calculate the optimal cut score. Once pressed, either a warning will pop up, or a "Show Results" button will appear. Press the "Show Results" button to start the calculation process (will take a minute or two, depending on your computer). An additional "Print Results" will then appear to allow you to download the results of your standard setting session.

The optimal cut score minimizing the total classification error is 0.6663.

Total Classification Error
The optimal cut score is 0.6663.
0.3
0.2
0.1
0.0
0.5
0.6
0.7
0.8
Percent of Items Correct on Exam

Figure 7: Example 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, 2020), as well as the shiny (Chang et al., 2020), shinydashboard (Chang and Borges Ribeiro, 2018), and shinyjs (Attali, 2020) 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. (2020). *shinyjs: Easily Improve the User Experience of Your Shiny Apps in Seconds*. R package version 2.0.0.

Borchers, H. W. (2019). *pracma: Practical Numerical Math Functions*. R package version 2.2.9.

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. (2020). *shiny: Web Application Framework for R*. R package version 1.5.0.

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.

R Core Team (2020). *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.