Pedagogical Interactive Shiny Applications for Estimation, Scoring, and Multi Dimensionality in Item Response Theory

Corresponding Author:
Metin Bulus
Department of Educational Measurement and Evaluation
College of Education, Adiyaman University, Adiyaman, Turkey
Email: bulusmetin@gmail.com

Wes Bonifay
Department of Educational, School and Counseling Psychology
College of Education, University of Missouri at Columbia, Missouri, USA
Email: bonifayw@missouri.edu

Abstract

Comprehension of foundational but fairly complex statistical theories may require assistive interactive tools to understand underlying equations and theory. We provide a collection of interactive shiny applications to demonstrate or explore some of the fundamental yet complex Item Response Theory (IRT) concepts such as estimation, scoring and multidimensionality. Users can explore principles of Maximum Likelihood Estimation such as Newton-Raphson iterations, influence of starting values and extreme scores on the convergence, principles of Expected A Posteriori (EAP) and Maximum A Posteriori (MAP) ability estimation such as likelihood, quadrature points, influence of prior mean, prior standard deviation and prior skewness on EAP and MAP estimates, and multidimensional IRT concepts such as item response surface, item information, compensatory and partially compensatory models for two-, three- and four-parameter logistic or ogive IRT models. We hope that these applications give a head-start to emerging practitioners and researchers interested in advanced measurement topics.

Keywords: *item response theory, maximum likelihood estimation, MIRT, scoring, EAP, MAP*

Pedagogical Interactive Shiny Applications for Estimation, Scoring, and Multi Dimensionality in Item Response Theory

Recent research has demonstrated the effectiveness of computer-based activities in helping students to better comprehend statistical concepts (Braun, White, & Craig, 2014). Further, Garfield and Ben-Zvi (2007) assert that knowledge retention and student enjoyment will increase when traditional statistical instruction is supplemented with interactive applications that allow for exploration of statistical results. We introduce five such applications that are intended to demonstrate and visualize the item response theory (IRT) concepts of estimation, scoring, and multidimensionality. These applications were developed for classroom instruction using the R statistical software program (R Core Team, 2019) and the *shiny* package (Chang, Cheng, Allaire, Xie, & McPherson, 2017). While statistical concepts can be explored by directly manipulating R code and observing the results, this process can be cumbersome and confusing to students, especially those who are new to the R environment. The *shiny* package converts complex R code into a simple interface that allows users to interact with and explore various concepts by varying parameters and visualizing output. Thus, the *shiny* package is a pedagogical tool that has great potential for elevating student knowledge, especially regarding topics as dense and advanced as IRT.

Application 1: Maximum Likelihood Estimation (MLE) of Person Location in a Rasch Model Given Item Difficulties

This application helps users to demonstrate, explore, and visualize the concept of maximum likelihood estimation (MLE) in the Rasch model. This is a topic that is central to IRT but the non-interactive, equation-based explanations found in textbooks are often daunting to students and other users who are new to statistical estimation. By using the intuitive slider controls rather than typing in syntax, the user will be able to: visualize the concepts of convergence and divergence as the estimation algorithm successfully locates or fails to locate the person ability (θ) parameter; and explore the influence of unreasonable starting values and aberrant scores on maximum likelihood estimation.

[Figure 1 about here]

To illustrate MLE in basic IRT (Rasch) models, this application replicates the analysis presented in de Ayala (2009, p. 22), wherein the person location (θ) parameter is estimated given the item locations of a five item math test. Using slider controls, users can manipulate three inputs and observe various estimation concepts as well as the pitfalls inherent in MLE. The first input is the respondent's raw score, the second input is the starting value for the estimation algorithm, and the third input is the number of estimation iterations. The change in θ between iterations, the log-likelihood statistics, the acute angle with the abscissa, and the estimated person ability for that particular iteration are superimposed on the plot to reinforce the aforementioned concepts.

Application 2: Maximum Likelihood Estimation in the 2PL Model

This application is the same as Application 1, though it allows users to investigate a slightly more complex IRT model: the two-parameter logistic (2PL) model. This application will help the user to: grasp the complexity of estimating one additional parameter (i.e., the discrimination parameter) as compared to Rasch model; understand that the 2PL MLEs are found by searching across a 3-dimensional surface rather than a line; and explore the influence of item discrimination on the estimation of item location.

[Figure 2 about here]

The 2PL MLE application simulates 50 responses to an item with discrimination and location parameters determined by the user. Two additional inputs enable the user to change the orientation of the 3D plot to get a better perspective of the estimated item parameters at which the log-likelihood is maximized.

To keep matters simple, in this application we do not go into details of Newton-Raphson equations. However, one may come to realize that Newton-Raphson equations consist of two equations with partial derivatives with respect to δ and α parameters. In the application we only use empirical finite values to draw the 3D plot. Unlike Application 1, the search would have followed a funnel-like zigzag beginning from starting values and narrowing down to the point of maximum.

Application 3: Maximum Likelihood Estimation in the 3PL Model

A third application aids in understanding estimation of the three-parameter logistic (3PL) model (Birnbaum, 1968), one of the most common, yet most complex used in practice today for the analysis of dichotomous response data. In particular, this application will help the user to: comprehend the complexity of estimating an additional parameter – the lower asymptote, or "guessing" parameter – as compared to the 2PL model; and realize that 3PL MLEs are searched through a four-dimensional space.

[Figure 3 about here]

The third application simulates 50 responses to an item with discrimination, location, and lower asymptote parameters determined by the user. As in the 2PL MLE application, the user also has the ability to change the orientation of the perspective plot and thereby gain a better understanding of the estimated item parameters at which the log-likelihood is maximized in the 3PL model.

Application 4: EAP and MAP Scoring

This application demonstrates Expected a Posteriori (EAP, Bock & Mislevy, 1982) and Maximum a Posteriori (MAP, Birnbaum, 1969) scoring techniques in the Rasch model. This is illustrated by replicating the analysis presented in de Ayala (2009, p. 79), in which the person location is estimated given the item locations of the five math items. In particular, this application allows users to: understand how the posterior distribution is obtained; differentiate between EAP and MAP estimation; realize that the ability for extreme responses can be estimated unlike with ML estimation in Application 1; visualize the concept of quadrature as it is used in IRT estimation; observe that MAP scores are more sensitive to characteristics of the prior distribution and the number of quadrature points; understand that increasing the number of quadrature points will result in a smoother posterior and thus more precise EAP and MAP estimates.

[Figure 4 about here]

[Figure 5 about here]

Application 5: Multidimensional Dichotomous IRT Models

This application is not related to estimation or scoring but have been produced to channel a complex concept in the IRT. The final application allows users to explore an item response surface (IRS) by manipulating hypothetical person locations and item discrimination and location parameters for each latent trait in a multidimensional IRT (MIRT, Reckase, 2009) model. In particular, this application will help the user to: identify differences in the IRS of dichotomous MIRT models; differentiate between compensatory and non-compensatory MIRT models; examine discrepancies between logistic and normal ogive MIRT formulations; observe the influence of item location and discrimination parameters on the IRS; and rotate the IRS to gain a better perspective of the precise shape of the 3-dimensional surface.

[Figure 6 about here]

[Figure 7 about here]

[Figure 8 about here]

Summary

Statistics instruction can benefit from interactive R-based applications that supplement the traditional teaching of abstract theorems and equations. The IRT applications described in here will enable students and other users to investigate fundamental estimation and scoring procedures and multidimensional response functions in detail. Not only will this improve comprehension of potentially complicated topics and techniques, it will also prepare students to be insightful applied researchers who are aware of the advantages and disadvantages that exist in various approaches to estimation and scoring.

Table 1

Access to Shiny Applications

Application Number	R	Web
Application 1	`irtDemo("mle")`	https://irtdemo.shinyapps.io/mlest/
Application 2	`irtDemo("est2pl")`	https://irtdemo.shinyapps.io/est2pl/
Application 3	`irtDemo("est3pl")`	https://irtdemo.shinyapps.io/est3pl/
Application 4	`irtDemo("eapmap")`	https://irtdemo.shinyapps.io/eapmap/
Application 5	`irtDemo("mirt")`	https://irtdemo.shinyapps.io/mirt/

References

Birnbaum, A. (1968). *Some latent trait models and their use in inferring an examinee's ability*. In F. M. Lord and M. R. Novick (Eds.), Statistical theories of mental test scores (pp. 397-472), Reading, MA: Addison-Wesley.

Bock, R. D., & Mislevy, R. J. (1982). Adaptive EAP estimation of ability in a microcomputer environment. *Applied Psychological Measurement*, *6*(4), 431-444.

Braun, J. W., White, B. J., & Craig, G. (2014). R tricks for kids. *Teaching Statistics*, *36*, 7–12.

Bulus, M., & Bonifay, W. (2016). irtDemo: Item response theory demo collection. R package version 0.1.2. Available at https://CRAN.R-project.org/package=irtDemo

Chang, W., Cheng, J., Allaire, J., Xie, Y., & McPherson, J. (2017). shiny: Web Application Framework for R. R package version 1.0.3. Available at https://CRAN.R-project.org/package=shiny

De Ayala, R. J. (2013). *The theory and practice of item response theory*. Guilford Publications.

Garfield, J., and Ben-Zvi, D. (2007). How students learn statistics revisited: a current review of research on teaching and learning statistics. *International Statistical Review*, *75*, 372–396.

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org.

Reckase, M. D. (2009*).* Multidimensional Item Response Theory Models. In *Multidimensional Item Response Theory* (pp. 79-112). New York, NY: Springer.

Figures

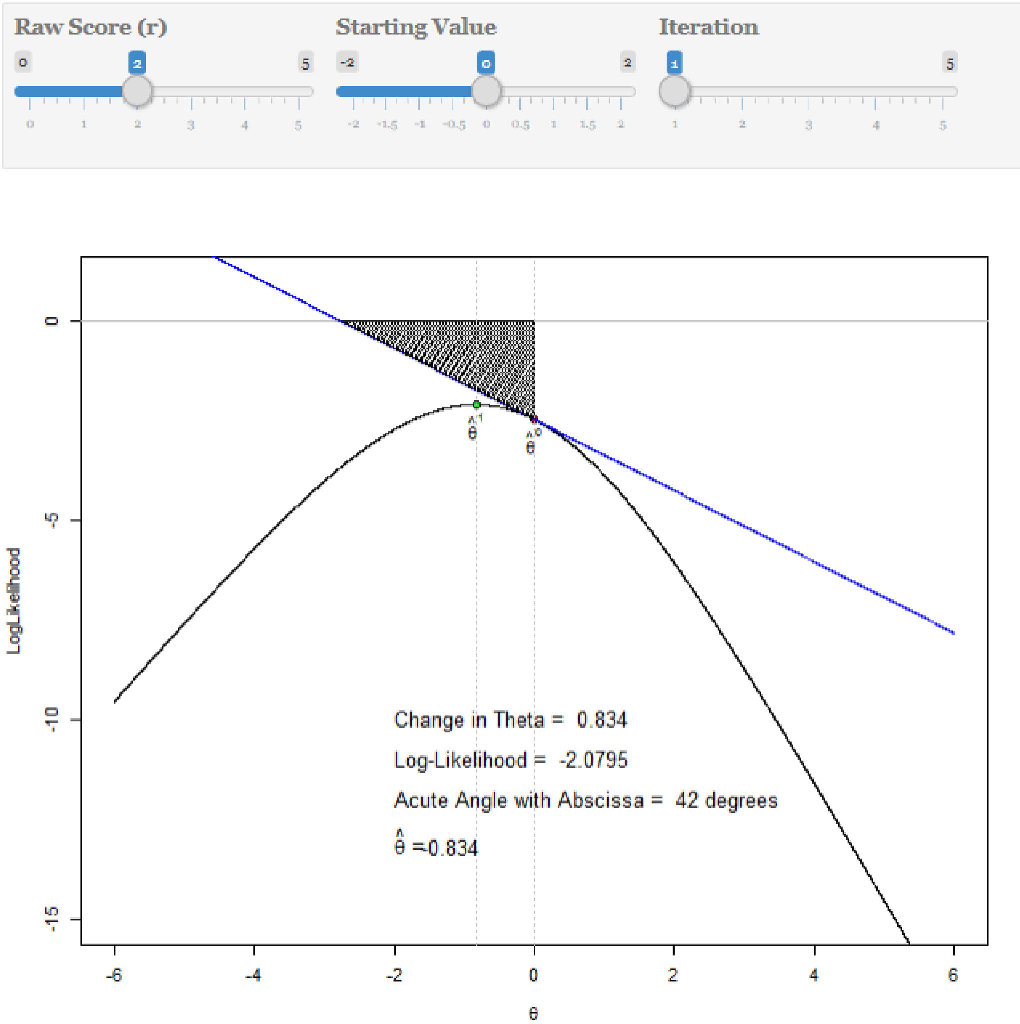

Figure 1: Maximum likelihood estimation in Rasch model

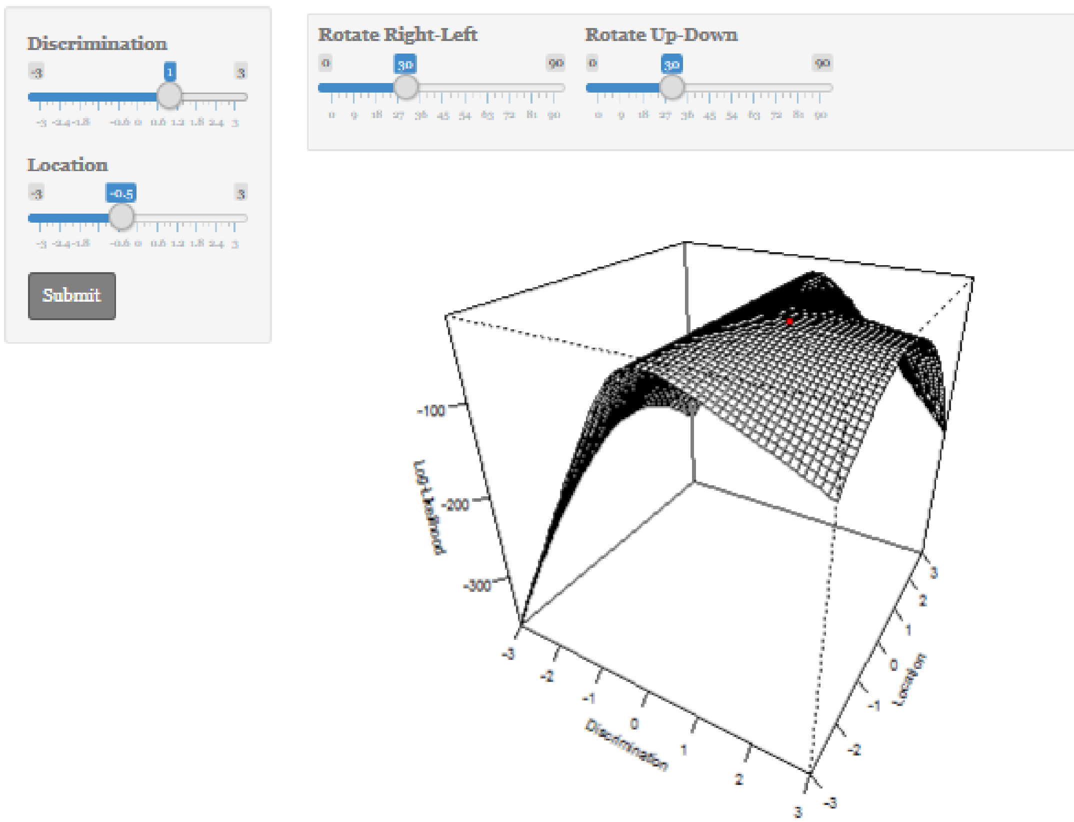

Figure 2: Maximum likelihood estimation in 2PL model

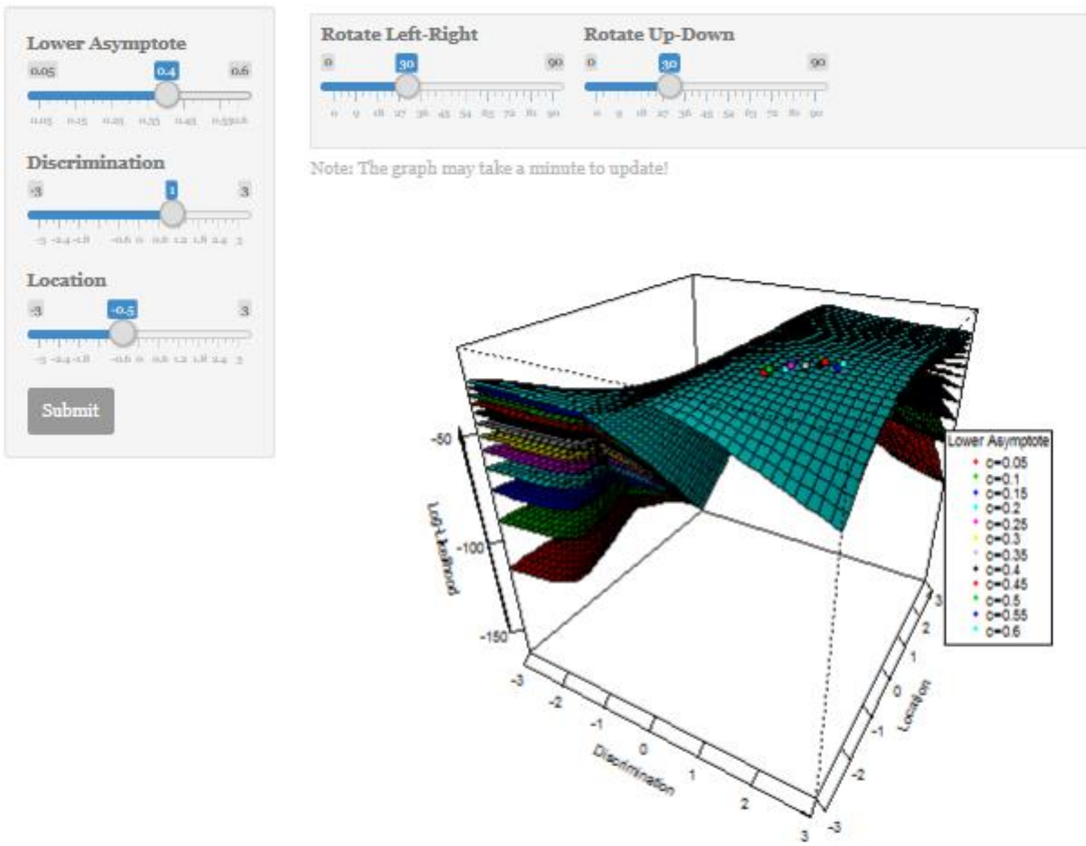

Figure 3: Maximum Likelihood estimation in 3PL model

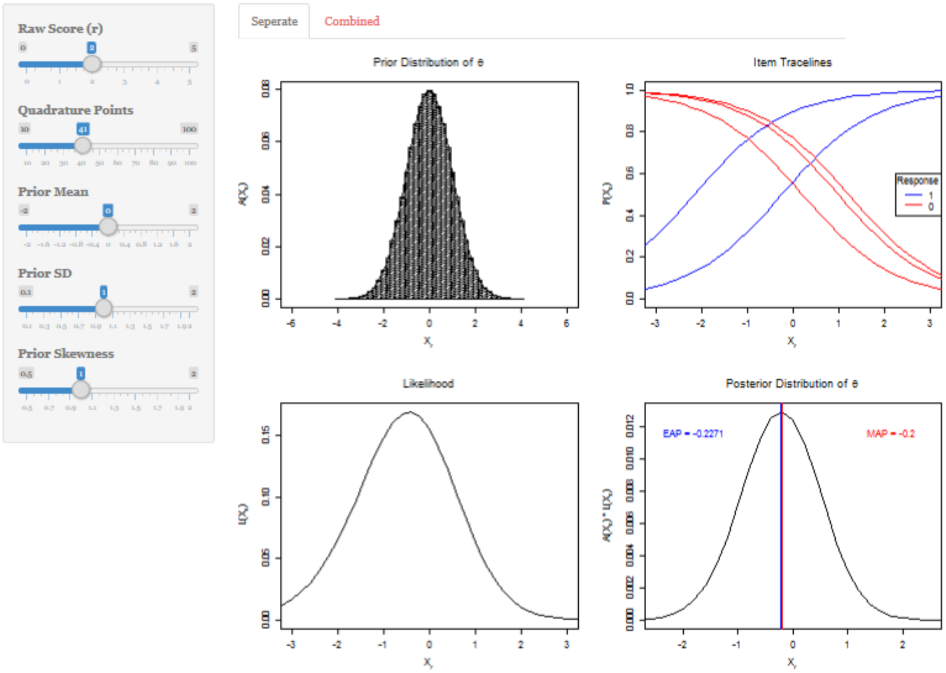

Figure 4: Scoring (with separate plots)

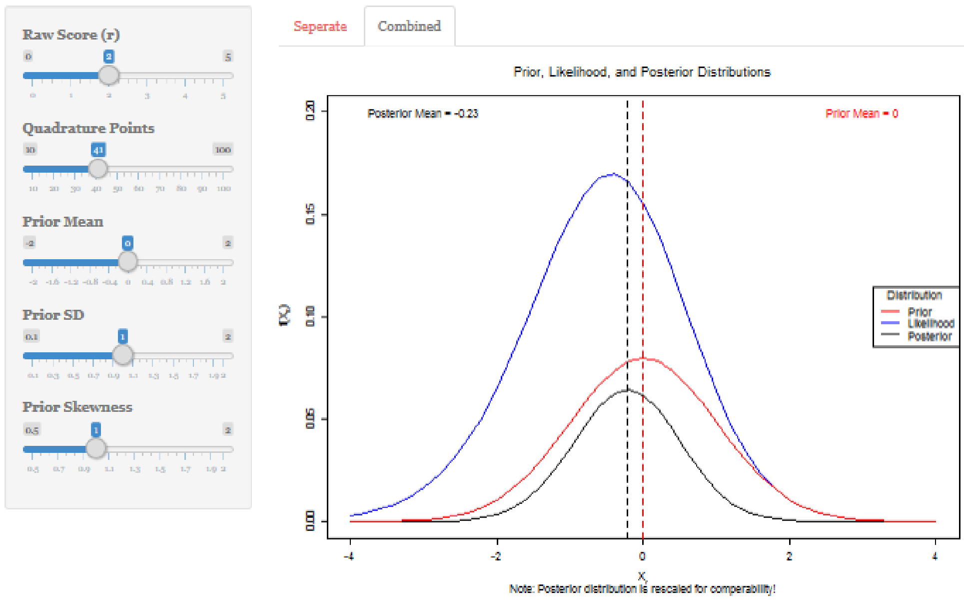

Figure 5: Scoring (with combined plot)

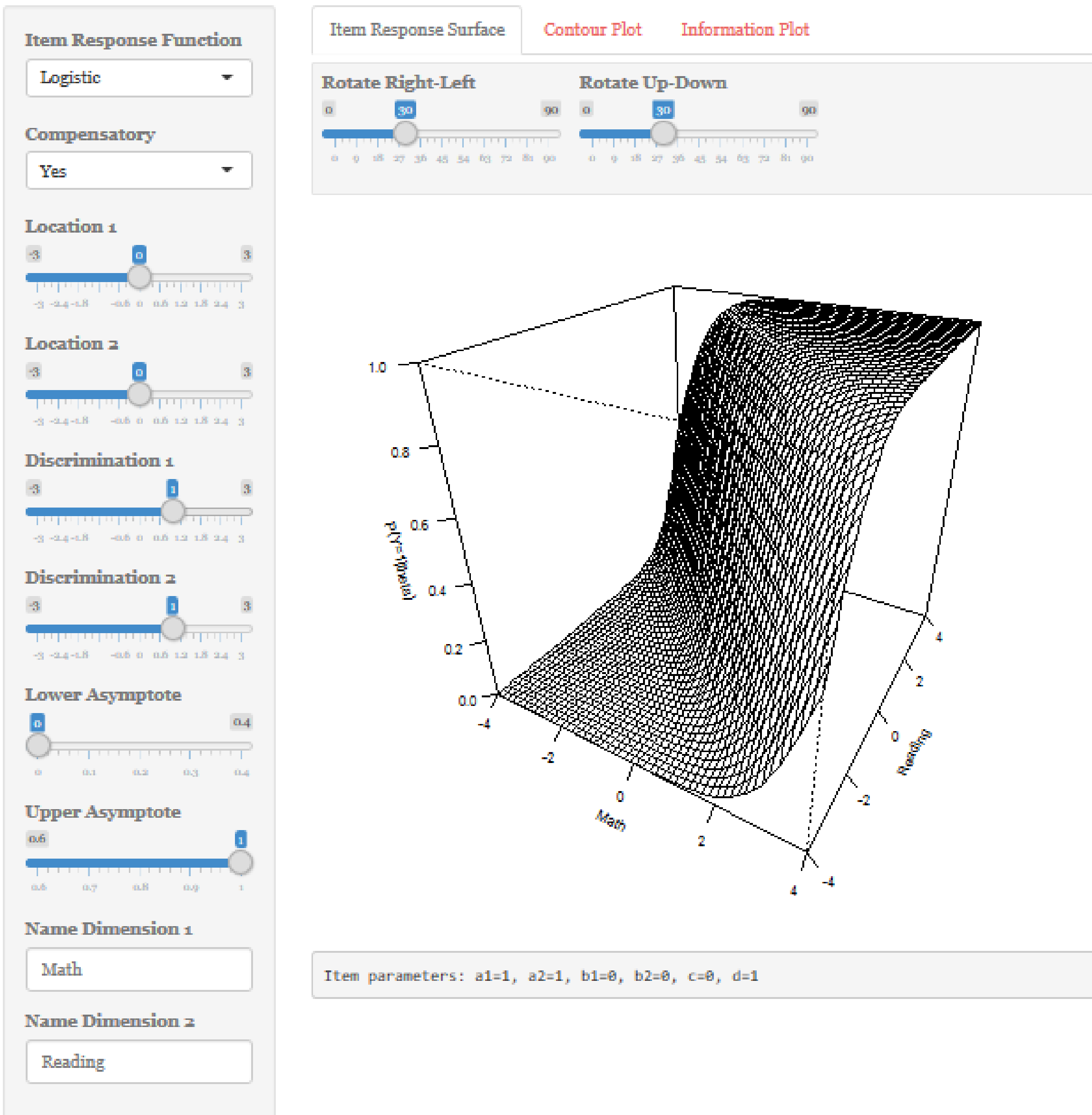

Figure 6: Item response surface

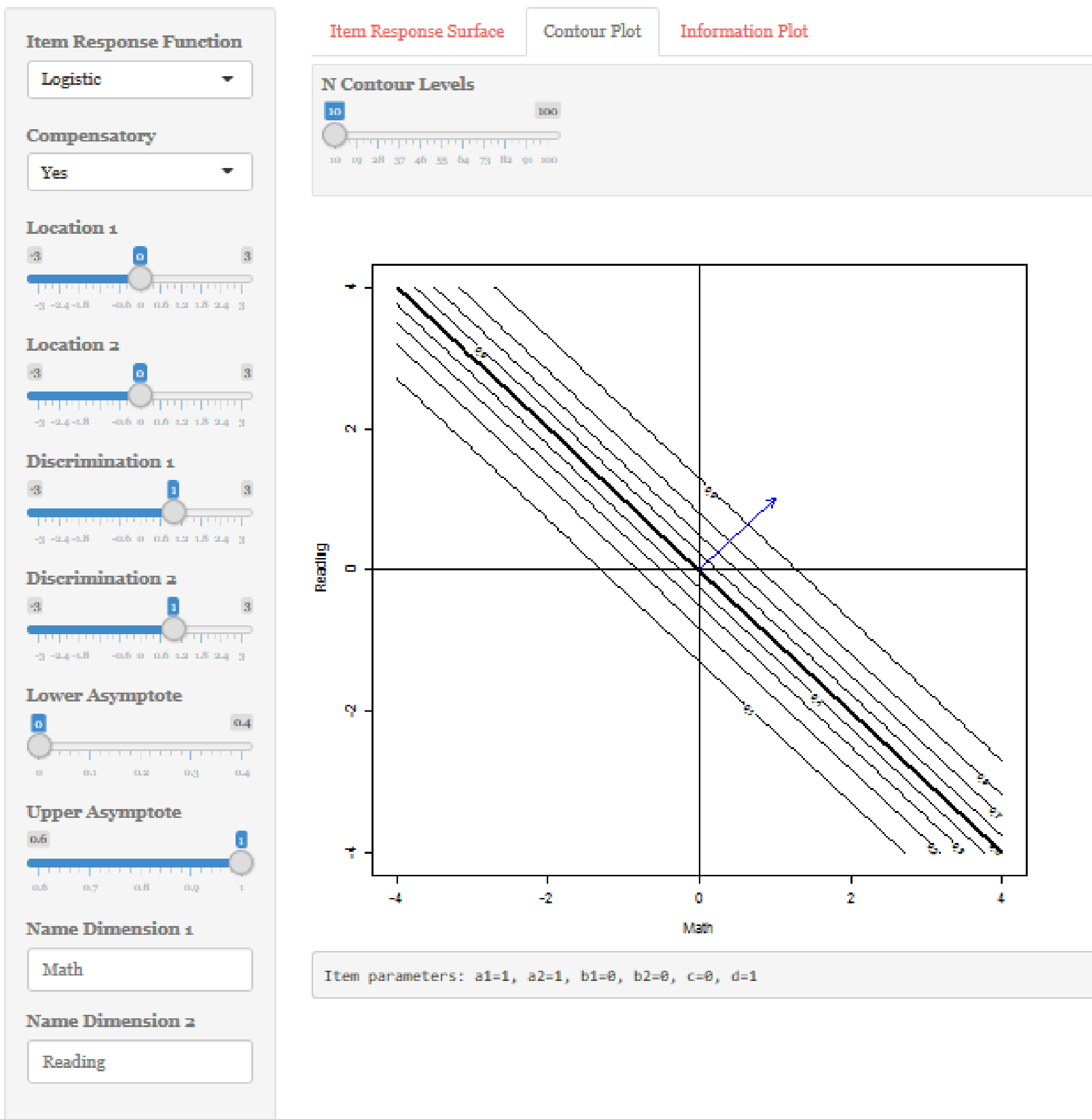

Figure 7: Contour plot tab

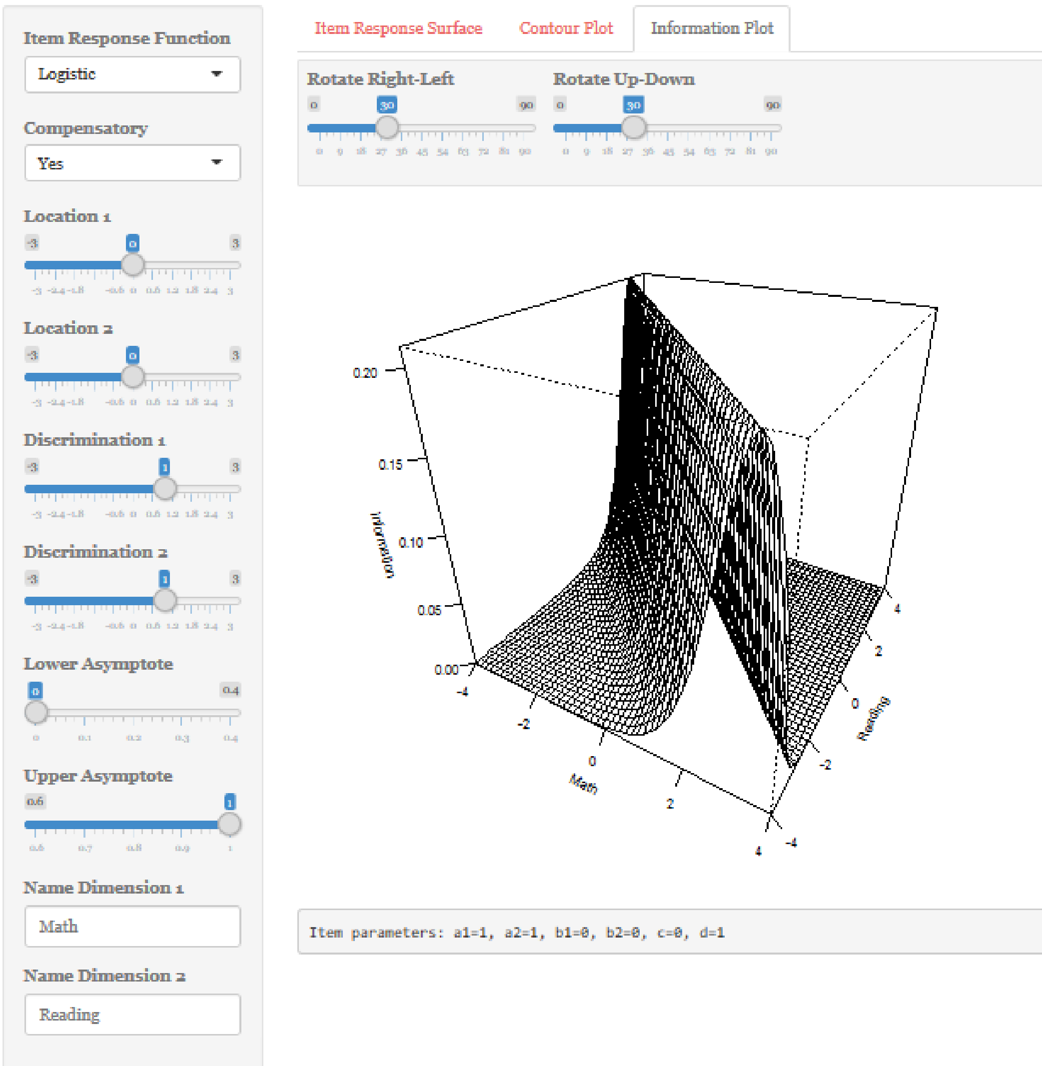

Figure 8: Information plot tab