FedIRT: An R package and shiny app for estimating federated item response theory models

Biying Zhou ©¹, **Yuhao Yuan** ©¹, **Nanyu Luo** ©¹, **Jia Qi** ©¹, and **Feng Ji** ©¹¶

1 Department of Applied Psychology & Human Development, University of Toronto, Toronto, Canada ¶
Corresponding author

Summary

Educational and psychological testing data are often sensitive, making centralized analysis challenging due to privacy concerns. To address this, we developed FedIRT, an R package that estimates Item Response Theory (IRT) models—including 1PL, 2PL, and graded response models, using federated learning. This approach enables multiple institutions (e.g., schools, districts) to collaboratively estimate model parameters without sharing raw data. FedIRT ensures privacy preservation while maintaining estimation accuracy comparable to mainstream IRT packages. Additionally, the package includes a user-friendly Shiny app, making federated IRT analysis accessible to researchers and practitioners without requiring advanced programming skills.

Statement of Need

IRT (Embretson & Reise, 2013) is widely used in educational, social, and behavioral sciences to assess abilities, attitudes, and other latent traits from test or survey responses. For example, IRT is used in standardized testing (e.g., SAT, GRE) to measure student proficiency, in psychological assessments to evaluate mental health indicators, and in large-scale surveys to estimate public opinion trends.

However, traditional IRT estimation requires all individual response data to be collected and processed in a centralized location. This poses significant privacy concerns, particularly when handling sensitive data from schools, educational institutions, and research studies (Lemons, 2014). Many school districts and organizations have strict data-sharing policies that prevent researchers from accessing raw student response data, limiting their ability to conduct large-scale educational assessments.

Federated learning offers a promising solution by enabling parameter estimation in a decentralized manner, ensuring privacy while still allowing researchers to analyze large-scale assessment data. Despite its potential, no existing R package in psychometrics supports federated learning for IRT estimation. Popular packages such as `mirt` (Chalmers, 2012) and `ltm` (Rizopoulos, 2007) require centralized data storage, making them unsuitable for privacy-sensitive applications.

To address this gap, we developed `FedIRT`, the first open-source R package that integrates federated learning with IRT estimation in a fully distributed manner. Unlike most theoretical research on federated learning, which often lacks practical implementations, `FedIRT` provides a working solution that enables multiple institutions (e.g., school districts, research organizations) to collaboratively estimate IRT parameters without sharing raw data. The package also includes a user-friendly Shiny app, making it accessible to education researchers, social scientists, and policymakers who may not have extensive programming experience. By bridging the gap

41 between federated learning research and real-world applications, FedIRT provides a practical
42 tool for privacy-preserving IRT analysis.

43 We estimate IRT models using federated stochastic gradient descent (FedSGD) and federated
44 averaging (FedAvg). This enables decentralized model estimation without sharing raw data,
45 ensuring privacy while maintaining accuracy. For technical details, refer to our methodological
46 discussions on Federated IRT (Zhou et al., 2025; Zhou & Ji, 2023, 2024). We support 1PL,
47 2PL, and graded models.

Comparison with existing packages

49 We demonstrate that our package generates comparable results to established IRT packages,
50 such as mirt (Chalmers, 2012).

51 Figure 1 and Figure 2 show the comparison of the discrimination and difficulty parameters
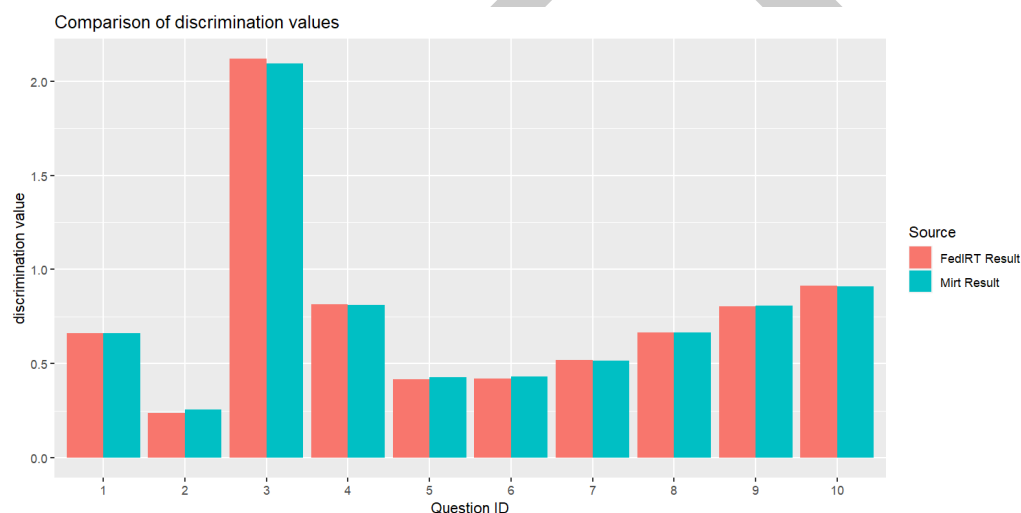52 between mirt and FedIRT based on example_data_2PL in our package.

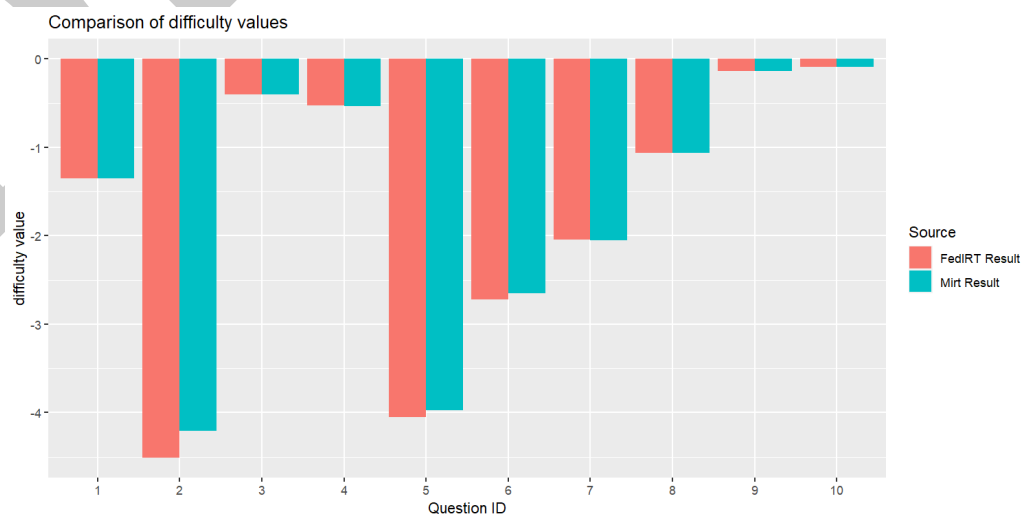**Figure 1:** Discrimination parameter estimates comparison

Figure 2: Difficulty parameter estimates comparison

Availability

The R package `FedIRT` is publicly available on CRAN (stable release) and Github (latest development version):

Github

```
devtools::install_github("Feng-Ji-Lab/FedIRT")
library(FedIRT)
```

CRAN

```
install.packages("FedIRT")
library(FedIRT)
```

Example of the Shiny App

To provide wider access for practitioners in real-world applications, we include the Shiny user interface in our package. A detailed manual was provided in the package. Taking the 2PL as an example, we illustrate how to use the Shiny app below.

In the first step, the server end (e.g., test administer, school board) can be launched by running the Shiny app `runserver()` and the client-end Shiny app can be initialized with `runclient()` with the interface shown below:
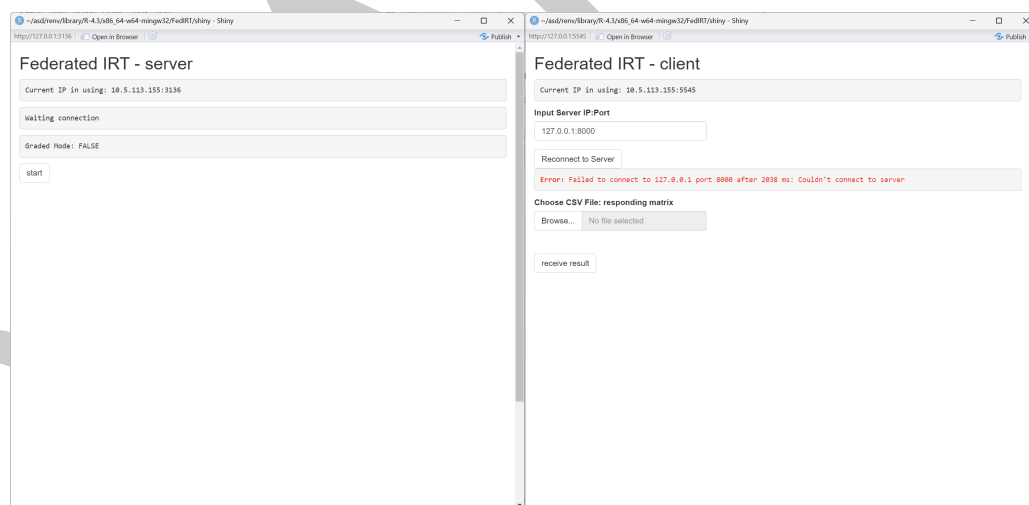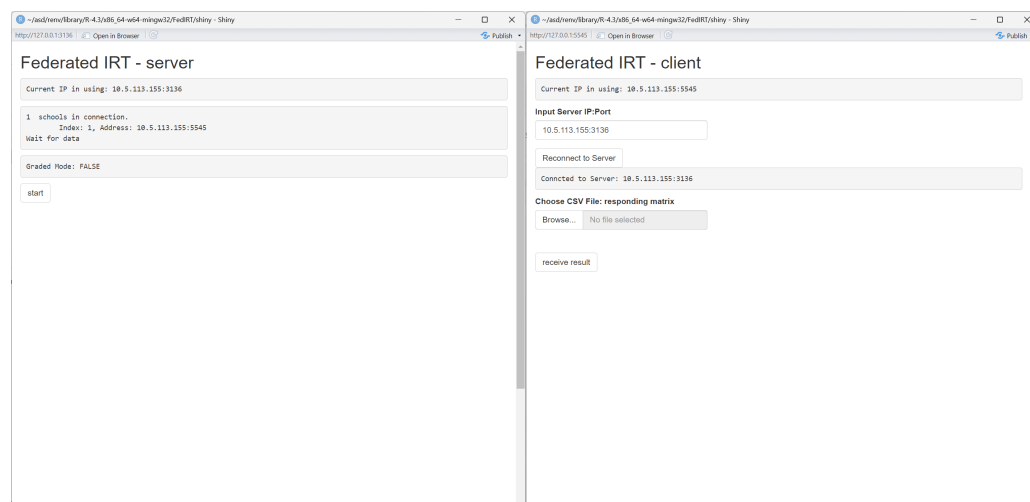

**Figure 3:** The initial server and client interface.

When the client first launches, it will automatically connect to the localhost port 8000 by default.

If the server is deployed on another computer, type the server's IP address and port (which will be displayed on the server's interface), then click "Reconnect". The screenshots of the user interface are shown below.

Figure 4: Server and client interface when one school is connected.

Then, the client should choose a file to upload to the local Shiny app to perform local calculations, without sending it to the server. The file should be a CSV file with either binary or graded responses. All clients should share the same number of items and the same maximum score for each item (if the responses are polytomous); otherwise, an error message will suggest checking the datasets of all clients.
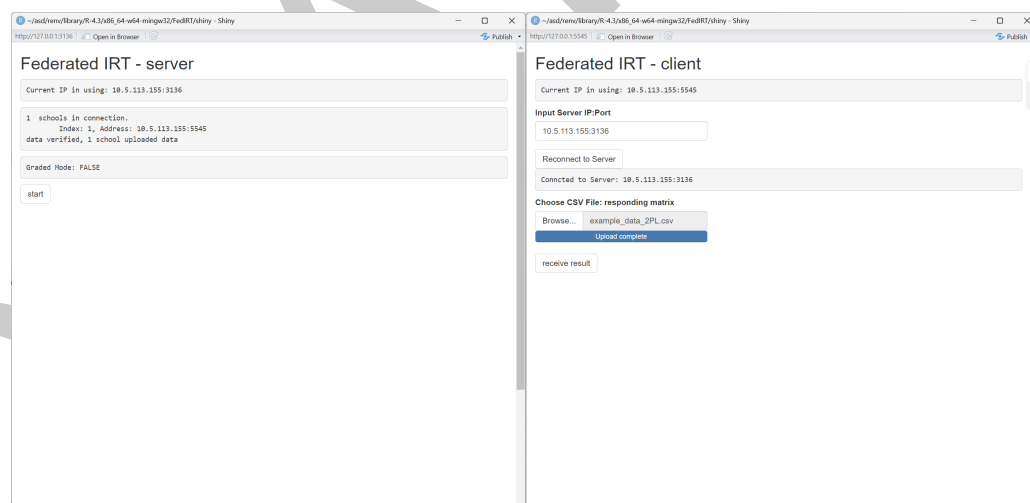
Figure 5: Server interface when one school uploaded dataset and client interface when a dataset is uploaded successfully.

After all the clients upload their data, the server should click "Start" to begin the federated estimation process. After the model converges, the clients should click "Receive Result". The server will display all item parameters, and the clients will display all item parameters and individual ability estimates.
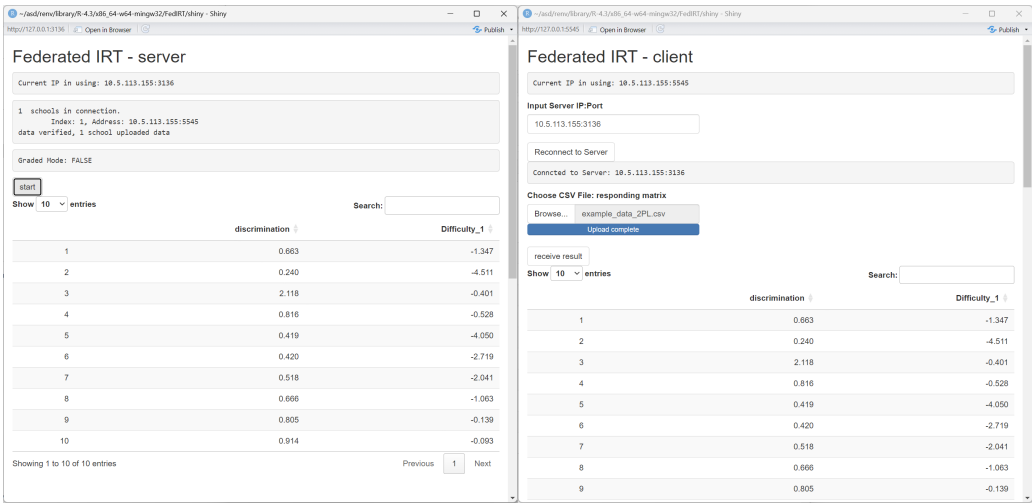
Figure 6: Server interface when estimation is completed and client interface when the results received.

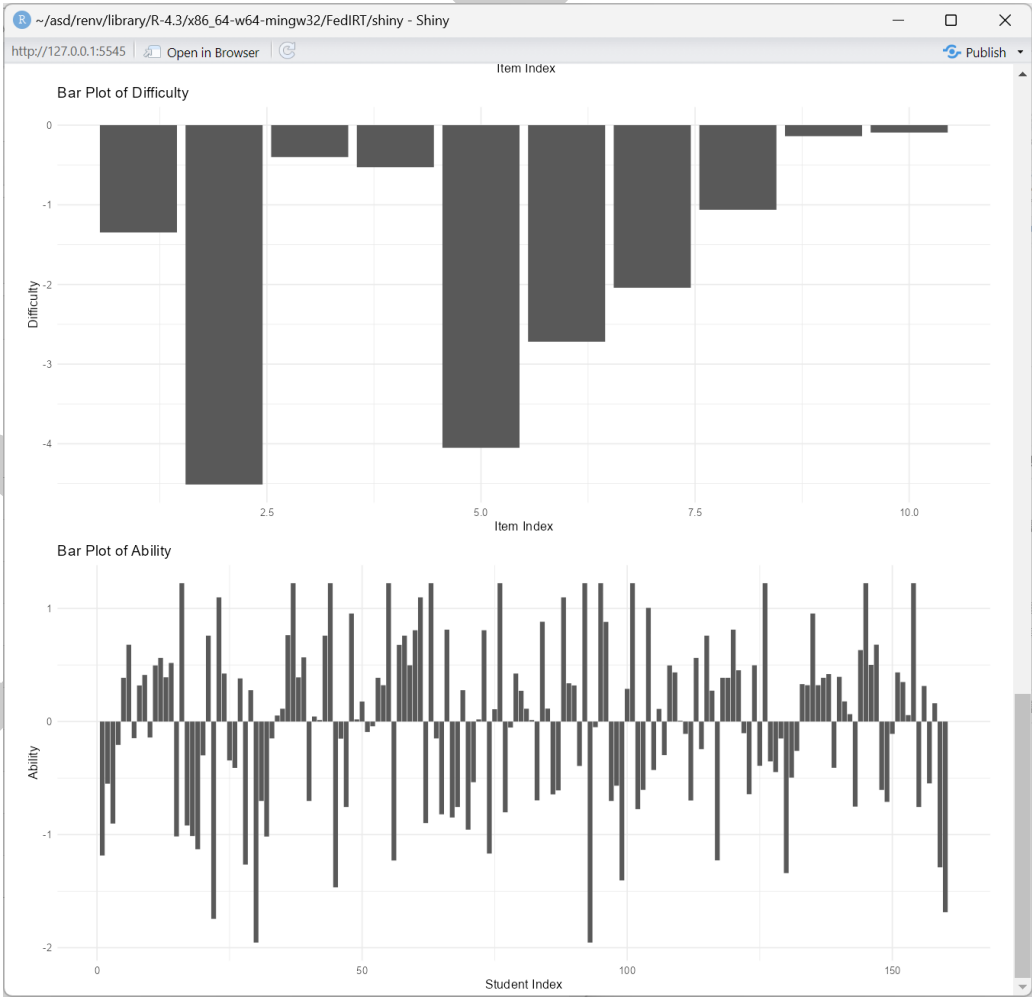The clients will also display bar plots of the ability estimates.

Figure 7: Client interface for displaying results.

Zhou et al. (2025). FedIRT: An R package and shiny app for estimating federated item response theory models. *Journal of Open Source Software*, *0*(0), ¿PAGE? https://doi.org/10.xxxxx/draft.

References

Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, *48*, 1–29. https://doi.org/10.18637/jss.v048.i06

Embretson, S. E., & Reise, S. P. (2013). *Item response theory*. Psychology Press. https://doi.org/10.4324/9781410605269

Lemons, M. Q. (2014). *Predictive modeling of uniform differential item functioning preservation likelihoods after applying disclosure avoidance techniques to protect privacy* [PhD thesis]. Virginia Polytechnic Institute; State University.

Rizopoulos, D. (2007). ltm: An R package for latent variable modeling and item response analysis. *Journal of Statistical Software*, *17*, 1–25. https://doi.org/10.18637/jss.v017.i05

Zhou, B., & Ji, F. (2023). Federated psychometrics: A distributed, privacy-preserving, and efficient IRT estimation algorithm. *APHD Research Gala*.

Zhou, B., & Ji, F. (2024). Federated item response theory: A distributed, privacy-preserving, and efficient IRT estimation algorithm. *DPE MED Research Practicum Poster*.

Zhou, B., Luo, N., & Ji, F. (2025). *Federated item response theory models*. https://arxiv.org/abs/2506.21744