

Using Item Response Theory (IRT) to Analyze Employee Perception Survey

Daisy Shi

12/16/2021

Introduction

This study discusses the use of Item Response Theory (IRT), to fit and assess a Graded Response Model (GRM) to analyze the correlations between employee perception and other variables (questions) in a Employee perception survey collected in 2016. Item response theory (IRT) models attempt to specify the relationship between individuals' underlying trait levels and the probability of endorsing an item using item and person characteristics. These models offer several advantages over classical test theory such as the ability to develop shorter, more efficient tests, use more powerful tests of differential item functioning (DIF), and implement computer adaptive testing. The GRM has often been used to analyze ordered categories data [1]. In this analysis, principal components are also used to understand if the analysis should take a Multi-dimensional or Uni-Dimensional approach [2], and examine the factor analysis to understand the underlying structure of the dataset before accessing the IRT and GRM.

Data Description

Public Sector Commission Employee perception survey 2016. In Western Australia, the Public Sector Commission's purpose is to bring leadership and expertise to the public sector to enhance integrity, effectiveness and efficiency. The Commission's vision is of a high performing public sector serving the needs of the State now and into the future. In accordance with section 22D of the Public Sector Management Act 1994, the Public Sector Commissioner reports annually to Parliament on the state of public sector administration and management, and on the extent of compliance with public sector standards and ethical codes. One of the primary information sources used for the State of the sectors 2016 report is the Employee perception survey (EPS). The EPS captures employee views about factors affecting their workplace, such as leadership, communication, work/life balance and performance

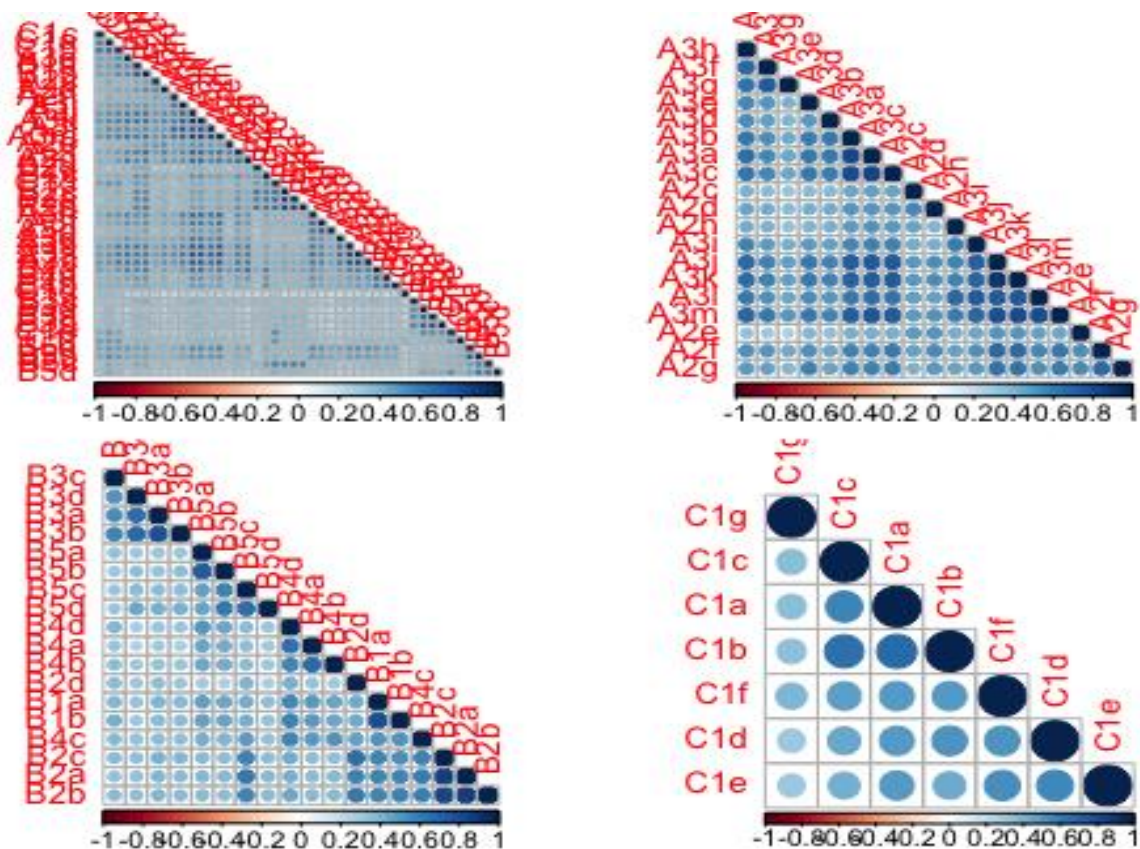
development. This information helps to understand organisational culture across the public sector and is essential in identifying areas for further consideration and acknowledging good practice.

The EPS is typically conducted in each public sector organisation with more than 70 employees once every five years. To improve the representativeness of the sample, organisations are selected across a range of sizes and portfolios. However, care should be taken in interpreting the EPS results because the sample may not be entirely representative of the broader public sector. In 2016, all employees in 11 public sector organisations were asked to complete the survey. The average response rate to the survey was 53%, with 3883 valid responses received. Three files are available for download on data.gov.au: 1) Public Sector Commission EPS 2016.pdf – the survey instrument sent to relevant employees to access online or in hardcopy. 2) Public Sector Commission EPS 2016 data.xls – individual responses to the EPS, with a question key and response key. Please note that some variables have been removed or aggregated to protect the privacy and anonymity of respondents. 3) Public Sector Commission EPS 2016 data.sav – individual responses to the EPS, in .sav format to be used with SPSS. Please note that some variables have been removed or aggregated to protect the privacy and anonymity of respondents. Further information about the EPS sample and summary results for 2016 are available in the State of the sectors - Statistical bulletin 2016.[3]

Method

The data contains three major series(i.e. series A, B, C),and each series contains a different sub-items. We will first examine the frequency of responses and missing value by Exploratory data analysis. Further more, to investigate how the correlations between the items. Using library(mise) package to impute missing values, and principal components are used to visualize and determine if Multi-dimensional or Uni-Dimensional approach shall be implemented in this analysis. And the scree plot which uses a plot of successive eigen vectors. Additionally to access the factor plot to understand the underlying structure of the dataset. Follow up with use of Item Response Theory (IRT), to fit and assess a Graded Response Model(GRM) to analyze the correlations between employee perception and other

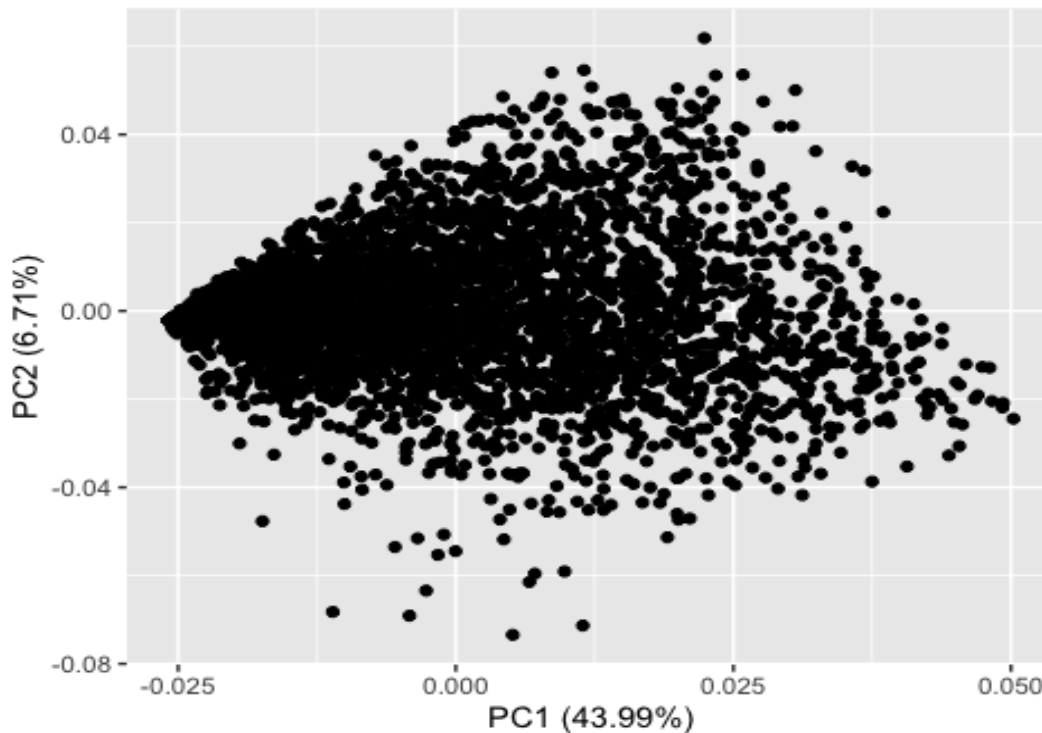
variables(questions) in a Employee perception survey collected in 2016. Graded response model is the recommended model for ordered polytomous response data using a full-information maximum likelihood fitting function. Based on the plot that we can observe that all the plots exhibit strong positive correlations. Hence we can use the whole dataset for our analysis without further cleaning on the grounds of correlation. Before modeling, we need to know whether the analysis should take a multidimensional or one-dimensional approach. This can be visualized using principal components and screen diagrams using continuous eigenvector graphs. Ideally, if the first principal component (PC1) can account for at least 40% of the variation in data, then a one-dimensional approach in IRT should be good enough, but sometimes we do not accept this threshold, which has been assumed by years of analysis and research. Therefore, we can look at the factor model to better understand the underlying structure of our data set.



Main Results

From the principal component analysis, we can see that about 44% of the data is explained by the first principal component. Using the Scree plot to verify that we have the same result. The scree plot suggests Uni-Dimensional approach is better fit. Additionally, by looking at the factor plot can help us to understand the underlying structure of our dataset.

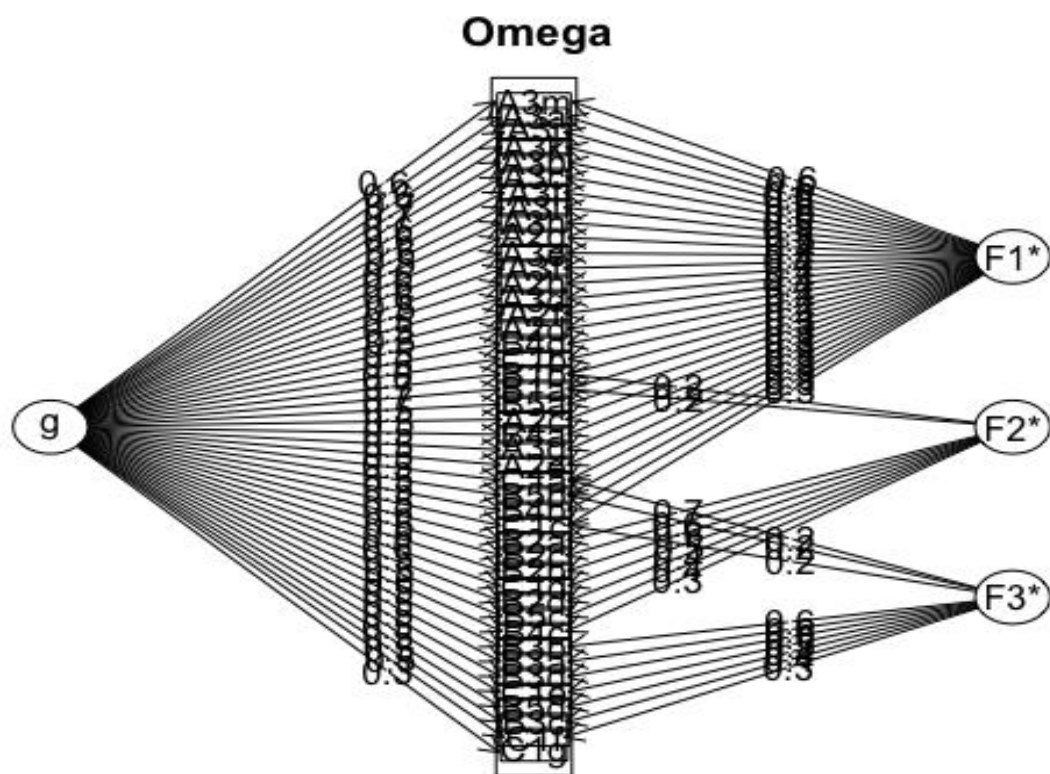
Figure 1: Principal Component PC1-PC2



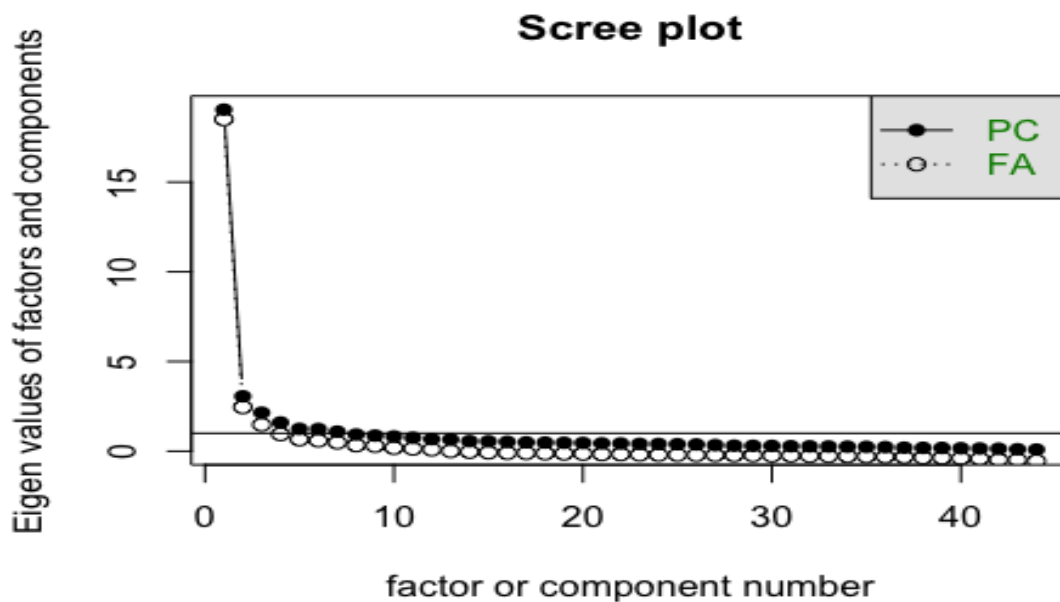
To ensure that our model is contributing and performing better, we fit another model using the same dataset by constraining the discrimination to be the same. Compared the group means between the two models to look into the overall performance of the model. The p-value, when compared to a threshold value of 5%, is significant, suggesting the negation of the null hypothesis that the model is insignificant. We can see that the BIC and AIC of our model is considerably lower than the constrained model, suggesting improved model performance.

Use omega diagram to see the hierarchical structure, we see that T = the degrees of freedom are 817 and the fit is 6.39. The number of observations was 3881 with Chi Square = 24692.39 with $\text{prob} < 0$. The root mean square of the residuals is 0.04. The df corrected root mean

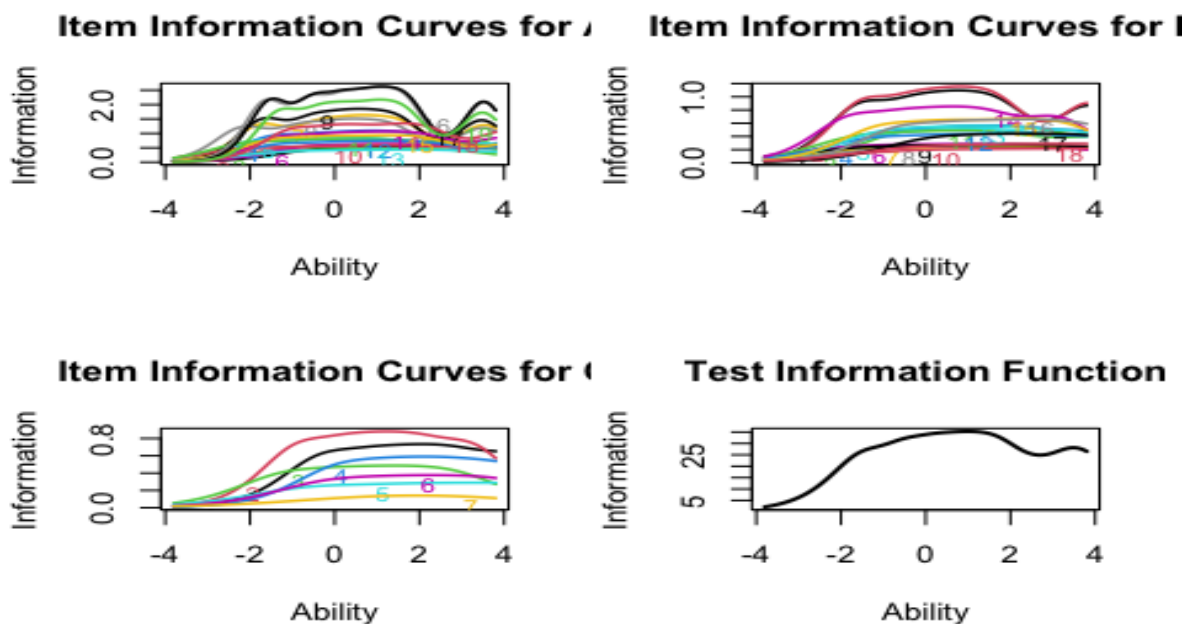
square of the residuals is 0.05. RMSEA index = 0.087 and the 10 % confidence intervals are 0.086, 0.088. BIC = 17940.83. Compare this with the adequacy of just a general factor and no group factors. The degrees of freedom for just the general factor are 902 and the fit is 13.99, The number of observations was 3881 with Chi Square = 54067.38 with prob < 0. The root mean square of the residuals is 0.14, The df corrected root mean square of the residuals is 0.15. RMSEA index = 0.123 and the 10% confidence intervals are 0.122 0.124. BIC = 46613.39.



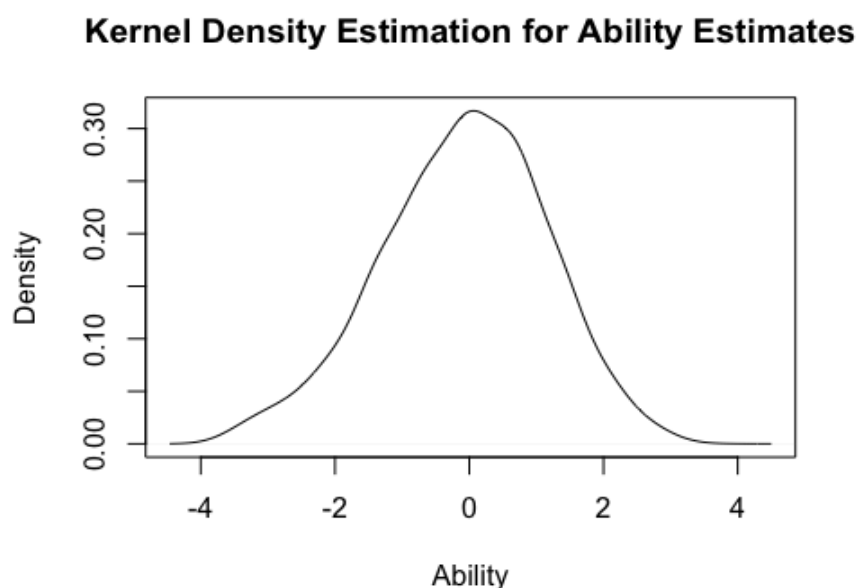
This scree plot explains the underlying structure of the dataset. We can see that the general load factor is distributed among three factors as confirmed by the three series. However, IRT analysis ignores these three factors and tries to formulate a cumulative relationship known as the latent trait which is the job satisfaction. Additionally, it is interesting to note that ITEM-B2a seems to have a higher load than other loadings which is a shift away from the general trend. This could depict that this particular item is too generalized leading to ineffective discrimination and thus rendering the contribution of that item, not much useful. Hence one should try and avoid such items or look into factor treatments.



From the item information curve, we observed a high Information ability with series A, followed by series B and least for series C. The Test Information Function plot suggests the same as well. In general the dataset contributes well enough with respect to understanding the latent trait of job satisfaction.



Graded Response Model results indicated the peak is around zero, and the goal is to shift this peak to the left for maximum job satisfaction. In order to ensure that the model contributes and performs better, hence, need to use the same data to fit the other model by limiting the differentiation of the same. By compared the group means between the two models to understand the overall performance of the model. Compared with the threshold of 5%, the p-value is significant, indicating that the null hypothesis is rejected, that is, the model is not significant. The results show that the BIC and AIC of the model are significantly lower than those of the constrained model, indicating that the performance of the model is improved.



Conclusions and Recommendations

From the results of this study, we were able to evaluate the effectiveness of EPS in measuring job satisfaction. By fitting a hierarchical response model, we can better understand the screening ability of each question item. Based on the test result, we should pay attention to the results of the following question items as they are more indicative of employee satisfaction with their jobs. For example, the agency inspires me to do the best I can with the agency inspires me to help it achieve its goals. I recommend my agency as a great workplace to provide effective leadership to the senior leadership of the agency. I feel that my agency throughout the management of public sector entities should further utilize the following matters

of analysis with low discriminatory results. This suggests that these items are likely to be generalized and may be qualitatively important to the organization. If measures are deemed to be agency-neutral, they may be excluded from further investigation by reducing them. Also, Purchase decisions in my workplace is not present and motivate people in my work group is committed to providing excellent customer service and community have a positive impact on my work team effective use of their time and resources to properly handle the direct supervisor of the employee to perform in the past 12 months, my work group has implemented the process of innovation or policy. By testing the information curve of the results, we determined that this subset of problems was able to distinguish which employees were dissatisfied with their jobs and was suitable for identifying areas of action.

Works Cited

1. LaHuis, D. M., Clark, P., & O'Brien, E. (2011). An Examination of Item Response Theory Item Fit Indices for the Graded Response Model. *Organizational Research Methods*, 14(1), 10–23. <https://doi.org/10.1177/1094428109350930>
2. Walker, C. M., & Beretvas, S. N. (2003). Comparing multidimensional and unidimensional proficiency classifications: Multidimensional IRT as a diagnostic aid. *Journal of Educational Measurement*, 40(3), 255-275.
1. Government, W. A. (2017, 3 30). Public Sector Commission WA Employee Perception Survey 2016. Retrieved 1 2019, from magda: <https://demo.dev.magda.io/dataset/ds-dga-b814c55a-d9d1-4145-af88-0fb78a354f8a/details?q=>

Appendix (R code)

```
`{r}
data<-read.csv("/Users/daisyshi/Desktop/Survey_data.csv")

# Missing Values Treatment

#check for missing values in each series
apply(data, function(x) sum(is.na(x)))
apply(a, function(x) sum(is.na(x)))
apply(b, function(x) sum(is.na(x)))
apply(c, function(x) sum(is.na(x)))

#Impute the missing values, the imputation will take some time
library("mice")
library(randomForest)
library(mice)

set.seed(123)
new<- mice(data, method = "rf", m = 5)

final <- complete(new)
apply(final, function(x) sum(is.na(x)))

#Extract the clean sets
clean_a <- final[,1:19]
clean_b <- final[,20:37]
clean_c <- final[,38:44]

#Check if correct extraction has been done
names(clean_a)
names(clean_b)
names(clean_c)

#Exploratory Data Analysis
library(ltm)
library(corrplot)
library(psych)
#To understand the frequency spread
description <- descript(final)
description$perc *100
#To understand the correlation statistics
relation <- cor(final, method = "spearman")
relationa <- cor(clean_a, method = "spearman")
relationb <- cor(clean_b, method = "spearman")
relationc <- cor(clean_c, method = "spearman")
par(mfrow=c(2,2))
```

```

corrplot(relation, type = "lower", order = "hclust")
corrplot(relationa, type = "lower", order = "hclust")
corrplot(relationb, type = "lower", order = "hclust")
corrplot(relationc, type = "lower", order = "hclust")

#load the libraries
library(ggfortify)
library(ggplot2)

#Check the principal components to see if there is any exhibition of unidimensionality
par(mfrow=c(1,1))
pc <- prcomp(final)
summary(pc)

autoplot(pc, data = final)
#Scree plot
library(psych)
scree(final)

#factor plot
plot(omega(final))
omega(final)

...

##Modelling using GRM
```{r}
#Fit the GRM model and check for convergence
library(ltm)
fit <- grm(final, IRT.param = TRUE)
fit$convergence
fit
...
```{r}
par(mfrow=c(2,2))

#Item Information Curves
plot(fit, type = "IIC", lwd = 1.5, item = 1:19, main = "Item Information Curves for A")
plot(fit, type = "IIC", lwd = 1.5, item = 20:37, main = "Item Information Curves for B")
plot(fit, type = "IIC", lwd = 1.5, item = 38:44, main = "Item Information Curves for C")

#Test Information Plot
plot(fit, type = "IIC", item = 0, lwd = 2)
...

```{r}

```

```

#Item Response Category Characterstics Curves and Item Information Curves for top
performing items
par(mfrow=c(3,2))
plot(fit, lwd = 2, item = c(17,16,19,9,7))
plot(fit, type = "IIC", item = c(17,16,19,9,7), lwd = 2)
```

```

```

```{r}
#Item Response Category Characterstics Curves and Item Information Curves for least
performing items
par(mfrow=c(3,2))
plot(fit, lwd = 2, item = c(44, 29, 26, 25, 28))
plot(fit, type = "IIC", lwd = 2, item = c(44, 29, 26, 25, 28))
```

```

```

```{r}
factor <- ltm::factor.scores(fit, method = "EAP")
plot(factor)
```

```

##Model Evaluation

```

```{r}
#fit the constrained model
fit_test <- grm(final, constrained = TRUE, IRT.param = TRUE)
fit_test$convergence
fit_test

compare the constrained model with the unconstrained model.

#Fit test using comparison of group means
anova(fit_test, fit)
```

```