

The School of Mathematics



THE UNIVERSITY  
*of* EDINBURGH

# Finding the Golden Rule in Cognitive Estimation Tests

by

LI LU, s1775712

July 2019

Supervised by  
Dr Mike Allerhand  
Dr Vanda Inacio de Carvalho

## Own Work Declaration

I assure that the content of the report is my own work excluding where otherwise indicated

LL LU

## Acknowledgements

I would like to thank Dr Mike Allerhand and Dr Vanda Inacio de Carvalho for supervising the project and giving me a lot of useful advice. I am also grateful to Dr Sarah E. MacPherson for providing the background information and data for this project.

## Executive summary

### Background

The Cognitive Estimation Test (CET) is used to determine the ability of a person to estimate the items that people may not know the exact answer.

### Research Question

The project aims at improving the existing CET by analyzing performance of different CET items, examining the relationships between different CET items and different aspects of cognition, determining whether different items distinguish between distinct patient groups, and finding a new gold-standard version of the CET.

### Data

We assess 379 neurological patients with brain damage in which 198 performed CET-A completely, 75 performed CET-B completely, and 14 patients entirely performed both CET-A and CET-B.

### Method

We conduct correlation analysis using Polychoric correlations and employ Principal Component Analysis (PCA) and Parallel Analysis. In addition, we use multiple-group (MG) IRT fixed-parameter estimation (FPE) method to detect the Differential Item Functioning (DIF).

### Result

Under Polychoric correlations, we notice that Question 3 and 4, Question 4 and 5, Question 2 and 6, Question 2 and 7, and Question 3 and 8 have significant relationships in CET-A and there are significant associations between Question 3 and 5, Question 3 and 6, Question 4 and 7, and Question 2 and 9 in CET-B. PC1 and PC2 in CET-A explain 36.4% of the variation and those explain 44% in CET-B under Principal Component Analysis. After Parallel Analysis, we find that one principal component is retained in CET-A and two are remained in CET-B. Age, years of education and gender have significant influences on the participants' performances on both CET, where such associations are detected using Kruskal Wallis H Test and DIF. Finally, we select Question 5,6 and 9 in CET-A and Question 4 in CET-B to build a golden rule.

## Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
1.1	Background and Motivation . . . . .	3
1.2	Data Source . . . . .	3
<b>2</b>	<b>Literature review</b>	<b>3</b>
<b>3</b>	<b>Exploratory Data Analysis</b>	<b>4</b>
3.1	Polychoric Correlations . . . . .	5
3.2	CET-A . . . . .	5
3.3	CET-B . . . . .	6
<b>4</b>	<b>Methods</b>	<b>6</b>
4.1	Principal Component Analysis (PCA) . . . . .	6
4.1.1	Bi-plots . . . . .	8
4.2	Parallel Analysis . . . . .	9
4.3	Association between item scores and possible confounds . . . . .	10
<b>5</b>	<b>Models and Results</b>	<b>11</b>
5.1	Differential Item Functioning (DIF) . . . . .	11
5.2	Multiple Group Estimation . . . . .	11
5.2.1	Age . . . . .	12
5.2.2	Years of education . . . . .	12
5.2.3	Gender . . . . .	12
5.3	The Golden Rule . . . . .	15
5.3.1	Model checking . . . . .	16
<b>6</b>	<b>Conclusion</b>	<b>17</b>
6.1	Limitations . . . . .	17
6.2	Future Directions . . . . .	17
	<b>Reference</b>	<b>20</b>
	<b>Appendices</b>	<b>21</b>
	<b>A Data: further information</b>	<b>21</b>
	<b>B Methods for model checking</b>	<b>21</b>
	<b>C R code and packages used</b>	<b>21</b>

# 1 Introduction

## 1.1 Background and Motivation

The Cognitive Estimation Test (CET) is one of the most frequently used methods to determine the ability to give estimations of the items that people may not know the correct answer. For example, these questions include “How long is the average tie ?” and “What is the length of the average newborn baby?”[9]. Reasoning and numeracy, strategies’ applications, response plausibility checking and general knowledge are of great significance in the performance of CET[13]. Executive functions such as understanding the question, choosing the suitable plan, using cognitive skills to carry out the plan and checking the final answer are processes to conduct the CET and they are related to the frontal lobes of the brain.

Since the original versions of CET are out of date, we devise two new versions of CET-A and CET-B. It is notable that frontal lobe damage patients have poorer performance in both CET-A and CET-B, and the performance in CET-B is also hindered for patients with non-frontal lobe impairment. Moreover, fluid intelligence can not fully explain the impairment on CET-A[4]. Therefore, we can conclude that CET-A and CET-B are not equivalent.

The goal of this article is to figure out the relationships between the distinct CET items and to determine whether different items distinguish between distinct patient groups. Finally, we produce one final “gold standard” after choosing items from CET-A and CET-B and check the effectiveness of the new CET on differentiating patient groups.

## 1.2 Data Source

The data set used in this project contains information of 379 patients and six background variables (age, years of education, gender, aetiology, severity in small vessels diseases and generalised atrophy). There are six kinds of aetiologies: vascular, brain tumour, neurodegenerative, other, neuropsychiatry and mixed. We have nine items for each of CET version A and version B. The responses to each item are marked with error scores ranging from 0 to 3 based on the normality of the patients’ performance (eg. 0 for “normal” and 3 for the most “abnormal”). In addition, raw score in each version represents the sum of the coded responses to the nine items and adjusted score represents item responses adjusted for age, education, and gender.

For each individual, several measures of general cognitive ability such as score on national adult reading test, scores for Wechsler adult intelligence scale are provided. Subcategories of WAIS scores such as Verbal IQ and Performance IQ, measures of verbal comprehension, working memory, and perceptual organisation, measures of executive ability with frontal lobes as well as measures of mood known to affect task performance like anxiety and depression are also included.

It is noteworthy that there are a variety of missing values in the data set. Some participants have not completed all the CET items, some finished only CET-A and some finished only CET-B, and some completed neither CET-A nor CET-B. We delete all the participants with missing values since it is meaningless to fill in the missing data with random values as the data will become inaccurate and will lead to devastating errors in data analysis afterwards.

# 2 Literature review

We have found three useful articles that conducted studies on both versions of the Cognitive Estimation Test (CET). The author of “Cognitive estimation: Performance of patients with focal frontal and posterior lesions”[4] developed two versions of CET (CET-A and CET-B) and conducted an error analysis based on the participants’ performance on CET-A, where the group of people was composed of patients with focal, unilateral and frontal ( $n = 38$ ) or posterior ( $n =$

22) lesions and healthy controls ( $n = 39$ ). It was found that the patients with focal, unilateral and frontal lesions not only performed worse than the healthy group, but also had a poorer performance than patients with posterior lesions. It also suggested that extreme and very extreme responses were impaired by frontal lobe damage. In addition, only very extreme responses were more hindered by frontal lobe than posterior damage.

The second study (“Cognitive estimation abilities in healthy and clinical populations: the use of the Cognitive Estimation Test”)[13] aimed to review the CET and other cognitive estimation’s tests as well as analyzed CET and other ways of executive functions. The article showed that it was still unclear if there was a specific area in the frontal lobes answering the estimation questions that required numerical estimations. It was found that when comparing distinct patients, etiology and localization were significant. On the other hand, when analyzing the patients, we could not ignore the importance of neuro-psychological processes tapped by the CET.

The third article, “Bringing the Cognitive Estimation Task into the 21st Century: Normative Data on Two New Parallel Forms”[9], used data of 184 healthy male and female participants aged 18–79 years with 9–22 years of education and devised two new forms of CET with 9 items separately. The authors found that successful CET performances were positively related to age and years of education but were negatively correlated to intellect, naming, arithmetic and semantic memory abilities. Male participants had advantage over female on CET while women performed better than men on the FAS phonemic fluency task. In addition, the correlation analysis presented that the CET was multidimensional in nature with distinct cognitive functions. However, lacking a posterior control group was one of the limitations of the study.

The goal of our study is to provide two contributions to the existing literature. The first contribution is that we use parallel analysis and principal component analysis (PCA) to determine the relationships between different CET items and to decide whether there exists distinct constructions of CET items.

The second contribution is that we give one final gold standard version of the CET. The literature reviewed above only devised CET-A and CET-B but has not provided a combined version of CET items.

### 3 Exploratory Data Analysis

To begin with, we perform data analysis on two versions of CET and build two correlation matrix to determine the relationships between items in CET separately.

Pearson correlation is the most widely used correlation statistic to measure the relationships between continuous variables. For the Pearson correlation, both variables should satisfy three conditions: normally distributed, linearity and homoscedasticity[3]. Variables have a bell-shaped curve if it is normal distribution. Linearity assumes that each of the two variables has a straight linear relationship and homoscedasticity means that variables are equally distributed about the regression line. However, in this case we have to adopt another method since the data are discrete and ordinal (the numbers only include 0,1,2 and 3), which do not meet the three assumptions of the Pearson correlation.

Although Spearman correlation can deal with data, it is found that one of the assumptions of the Spearman correlation is that the two variables should follow a monotonic relationship (If the value of one variable increases, the other variable should increase or decrease at the same time)[3]. Apparently, our data do not follow this relationship. Therefore, we look for a method that has even more lenient constraints for evaluating the correlation, the polychoric correlation.

### 3.1 Polychoric Correlations

Specifically, our data is not simple categorical values, but underlying continuum, in which the values represent the degree of normality(eg. 0 for “normal” and 3 for the most “abnormal”). This is quite similar as likert scale items such as “1=Strongly Disagree” and “5=Strongly Agree”. The polychoric correlation assumes that the values are underlying continuum and are not truly continuous[8]. It measures the strength and direction of the relationship between two variables and this is what we want in our report.

Suppose  $Z_1$  and  $Z_2$  are two ordinal items with  $m_1$  and  $m_2$  categories, we assume that there exist normally distributed variables  $Z_1^*$  and  $Z_2^*$  underlying these items and their combined distribution is a normal bivariate distribution with a correlation  $\rho$ . The correlation  $\rho$  in the bivariate normal distribution  $N(0,0,1,1,\rho)$  is the polychoric correlation of the latent variables  $Z_1^*$  and  $Z_2^*$  which is formulated as[7]:

$$P[X = i, Y = j] = p_{ij} = \int_{a_{i-1}}^{a_i} \int_{b_{j-1}}^{b_j} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp^{-\frac{1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)} dx dy$$

This can be calculated by maximizing the maximum likelihood of the distribution:

$$\ln L = \sum_{i=1}^{m_1} \sum_{j=1}^{m_2} n_{ij} \log p_{ij}$$

After deleting the rows with missing values, we implement Polychoric correlation and consider CET-A and CET-B separately.

### 3.2 CET-A

The correlation matrix of CET-A is shown in Figure 1. The correlations between Question 1 and 7, Question 2 and 7, Question 5 and 7, Question 5 and 9, and Question 7 and 8 are negative while other correlations are positive.

We calculate p-value for each item in the correlation matrix and set the threshold of p-value as 0.05. After barring the non significant coefficients in CET-A which is shown in Figure 1(b), it is noticeable that all other associations are not important except Question 3 and 4, Question 4 and 5, Question 2 and 6, Question 2 and 7, and Question 3 and 8.

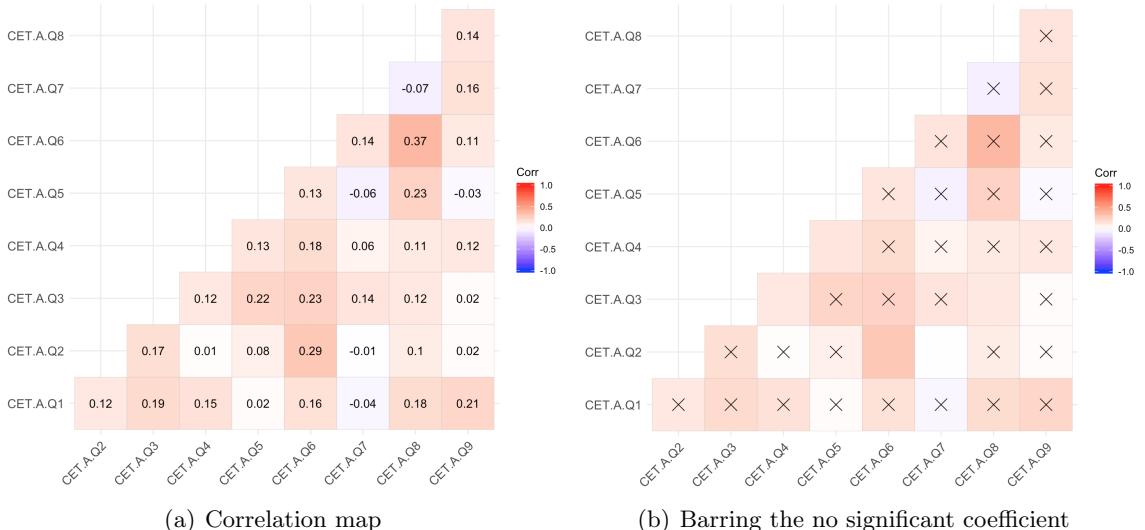


Figure 1: Correlation matrix for CET-A

### 3.3 CET-B

For CET-B the Polychoric correlation matrix is shown in Figure 2. It is clear that Question 1 and 5, Question 3 and 5, Question 8 and 5, Question 6 and 7, Question 6 and 3, and Question 6 and 8 are negatively correlated and others are positively correlated.

In Figure 2(b), we can see that the significant relationships in CET-B is Question 3 and 5, Question 3 and 6, Question 4 and 7, and Question 2 and 9 while the associations between other items are not significant.

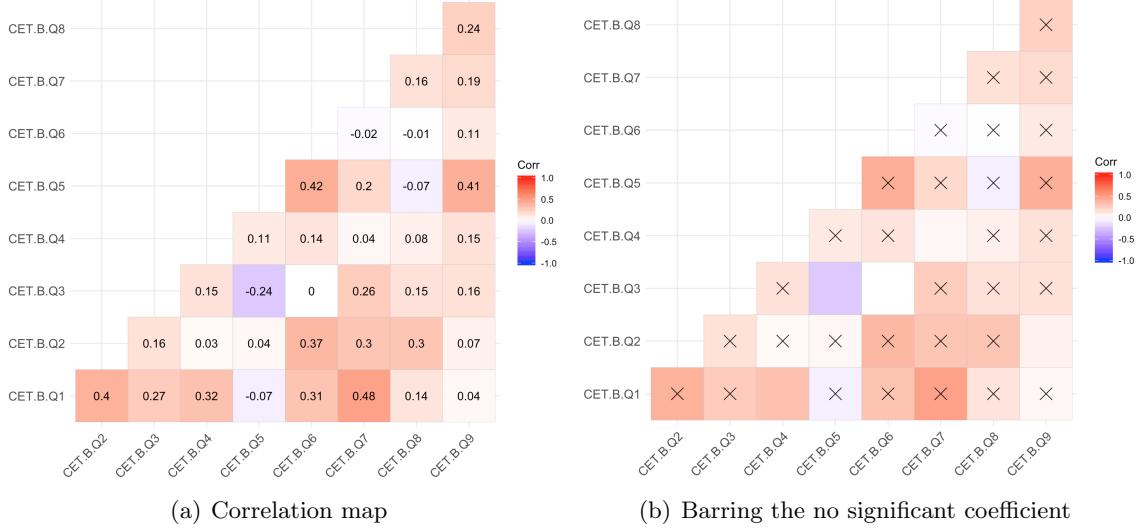


Figure 2: Correlation matrix for CET-B

However, as suggested in the previous literature review, there may exist some confounds such as age, years of education and gender, which will affect the pairwise relationships among the items. Therefore, such calculations of correlations and associations may not accurately represent their relationships and can only be considered as a reference.

## 4 Methods

In our data-set, a wide range of variables exist and it is hard to determine which samples are similar and which samples are different from each other. We have to use data reduction techniques to capture the variance in the independent variables in a smaller set and combine similar samples into one group.

### 4.1 Principal Component Analysis (PCA)

There are two methods for us to choose: Exploratory Factor Analysis (EFA) and Principal Component Analysis (PCA). EFA is that one variable drives the differences in all the independent variables, while PCA is that we take weighted sum of variables and aggregate the effects contributed by the variables. The comparison of PCA and EFA is presented in Figure 3[6].

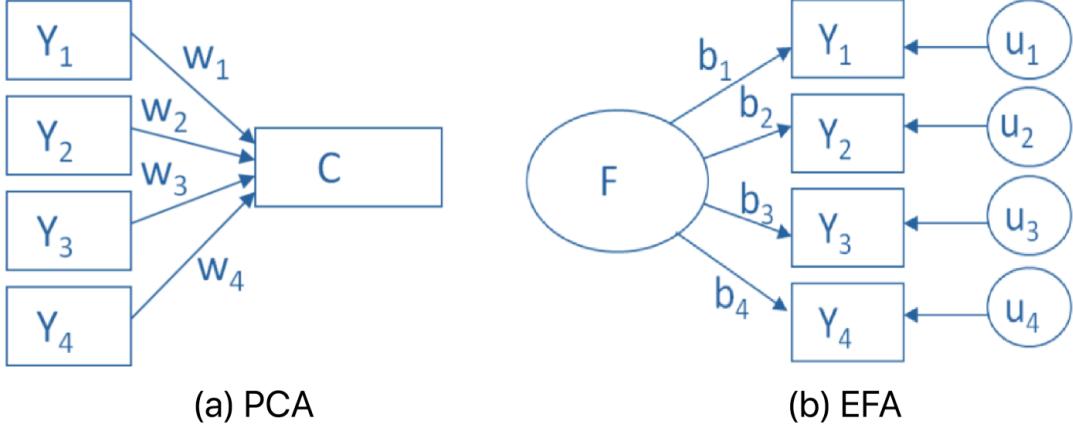


Figure 3: Comparison of PCA and EFA

We conduct PCA when the index is an outcome of the indicators, but EFA is used when the variables are latent and can not be measured directly (eg. intelligence and prejudice). While we cannot guarantee the orthogonality in EFA, we can always get orthogonal eigenvectors as long as there exists an eigendecomposition of the correlation matrix for PCA[12]. In this case, PCA is preferred since it is nice and simple compared to EFA. We can then use the eigenvectors to compute the corresponding PC in full scores.

In PCA, we simplify the data set by combining our original values to principal components, which represent the directions of the data that explains the maximum amount of variance. To analyze the PCA mathematically, we should use the concept of eigenvalues and eigenvectors. An eigenvector is a direction and an eigenvalue is a number representing how much variance explained in such a direction. In addition, the eigenvector which has the largest eigenvalue is the first principal component.

We obtain nine principal components (PC) and each PC represents some proportion of variance. The variances explained in each CET item is shown in Figure 4. It is obvious that the first component explains the most variance and the second explains the second most variance. It is found that in CET-A, PC1 explains 22.8% and PC2 explains 13.6%; therefore, they explain 36.4% of the variation in total. In CET-B, PC1 explains 26.4% and PC2 explains 17.6% of the variation and the result is 44% of variance explained after summing them up.

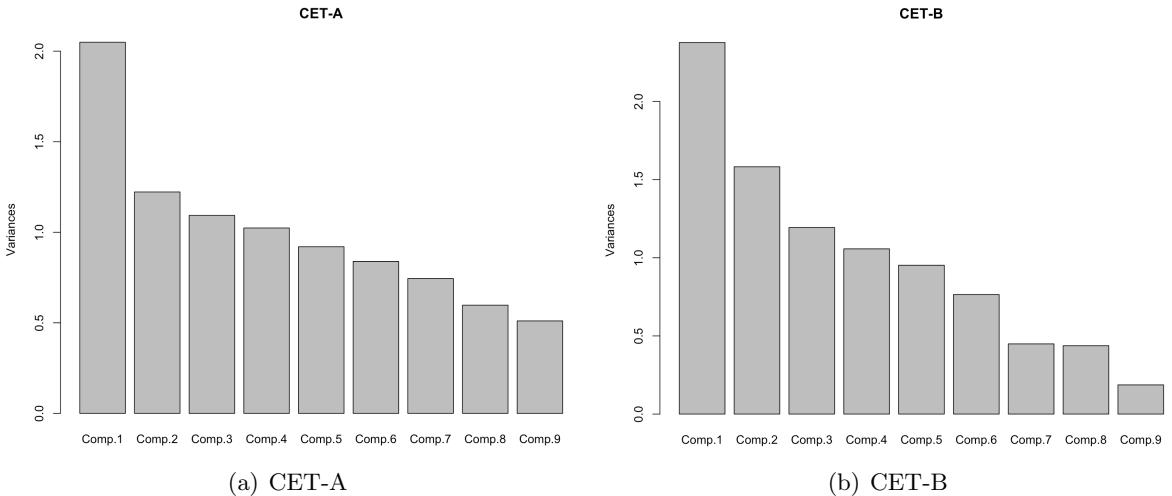


Figure 4: Variances explained in each CET item

#### 4.1.1 Bi-plots

Bi-plots are the plotting of PCA and we make bi-plots for CET-A and CET-B respectively, which are shown in Figure 4 and Figure 5 with label 1-6 representing six kinds of aetiologies separately. Bi-plot is the combination of PCA score plot and loading plot. For Figure 5, a PCA plot shows clusters of samples based on their similarity and we can determine that the cluster of each disease does not seem quite similar.

In addition, a loading plot shows how strongly each characteristic influences a principal component. We can see from the figure that for CET-A, Question 1,3,4 and 6 strongly influence PC1 while Question 5 and 7 have significant influences on PC2.

Moreover, Question 1 and 4, Question 3 and 6, Question 2 and 8 are positively correlated since each pair of the two vectors form a small angle. For Figure 6, the biplot of CET-B is another story. Aetiology 2 and 6 are different and can be classified easily. Question 2, 4 and 7 strongly influence PC1 and Question 5 has important influence on PC2. Furthermore, Question 1, 2 and 7 and Question 6 and 9 are positively related.

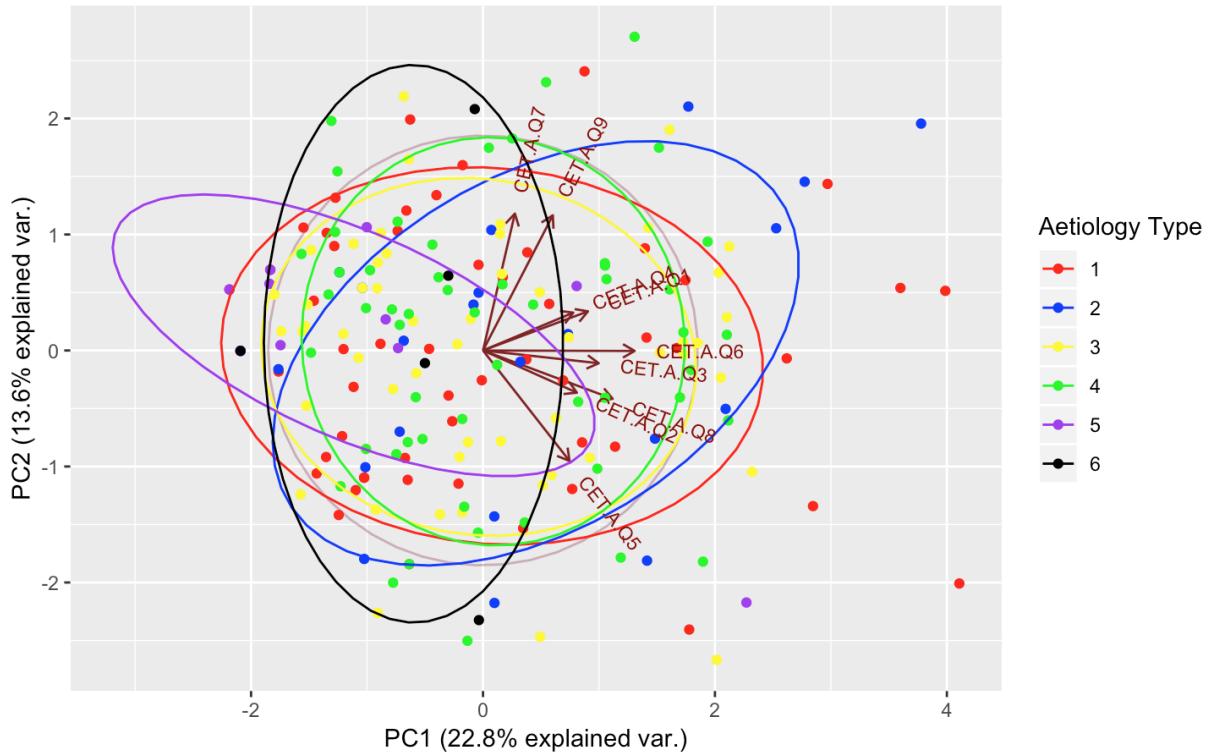


Figure 5: Bi-plots for CET-A

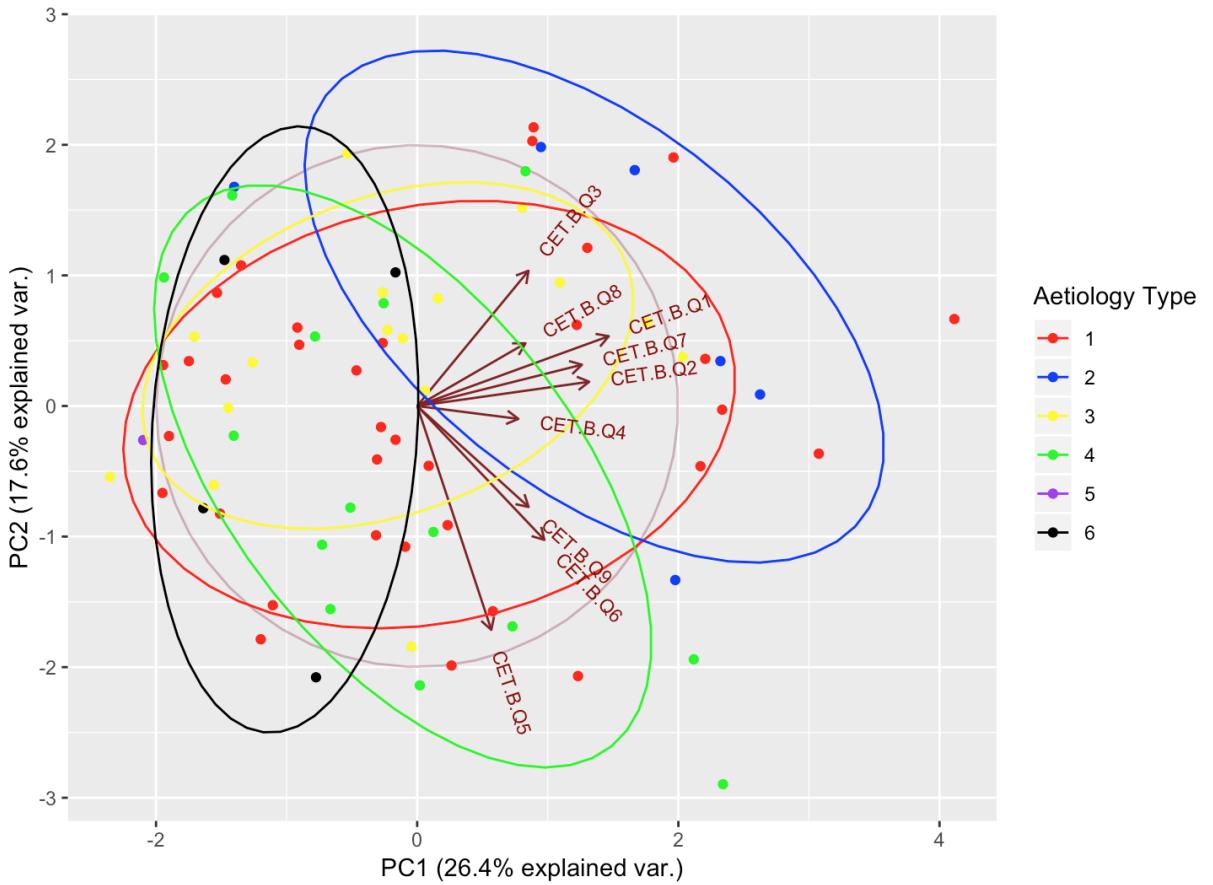


Figure 6: Bi-plots for CET-B

In practice, we usually take the first few principal components. However, there is a more mathematical and statistical way to examine the number of PCs to be retained.

#### 4.2 Parallel Analysis

Parallel Analysis is a method for determining how many components should be retained from PCA. Parallel Analysis, according to random data generation paralleling to the actual data set, conducts the Monte Carlo Simulation Technique to decide the number of retained components by comparing the eigenvalues of the two data sets.

If the associated eigenvalue is bigger than the 95th of the distribution of eigenvalues derived from the random data, then the component is retained. The eigenvalue found is significantly larger than eigenvalues generated by random matrix, which is similar as permutation test[5].

Furthermore, we perform Parallel Analysis and the plots for CET-A and CET-B are shown in Figure 7 separately. To decide how many components to retain, there are two ways. The first method is to see how many eigenvalues in the PCA column are greater than the average eigenvalues in the PA column and the second approach is to see the dashed line for Parallel Analysis crosses the solid line before reaching which component. We conclude that there is one component retained in CET-A and two components retained in CET-B. The formula for CET-A component 1 is

$$0.341Q1 + 0.305Q2 + 0.376Q3 + 0.294Q4 + 0.281Q5 + 0.493Q6 + 0.104Q7 + 0.422Q8 + 0.227Q9$$

and that for CET-B component 1 is

$$0.476Q1 + 0.428Q2 + 0.277Q3 + 0.252Q4 + 0.184Q5 + 0.316Q6 + 0.409Q7 + 0.269Q8 + 0.277Q9$$

For CET-B component 2, the formula is

$$0.213Q1 + 0.412Q3 - 0.682Q5 - 0.408Q6 + 0.126Q7 + 0.191Q8 - 0.307Q9$$

Since the principal components are constructed as linear combinations of the original variables, they are hard to interpret and have no real meaning.

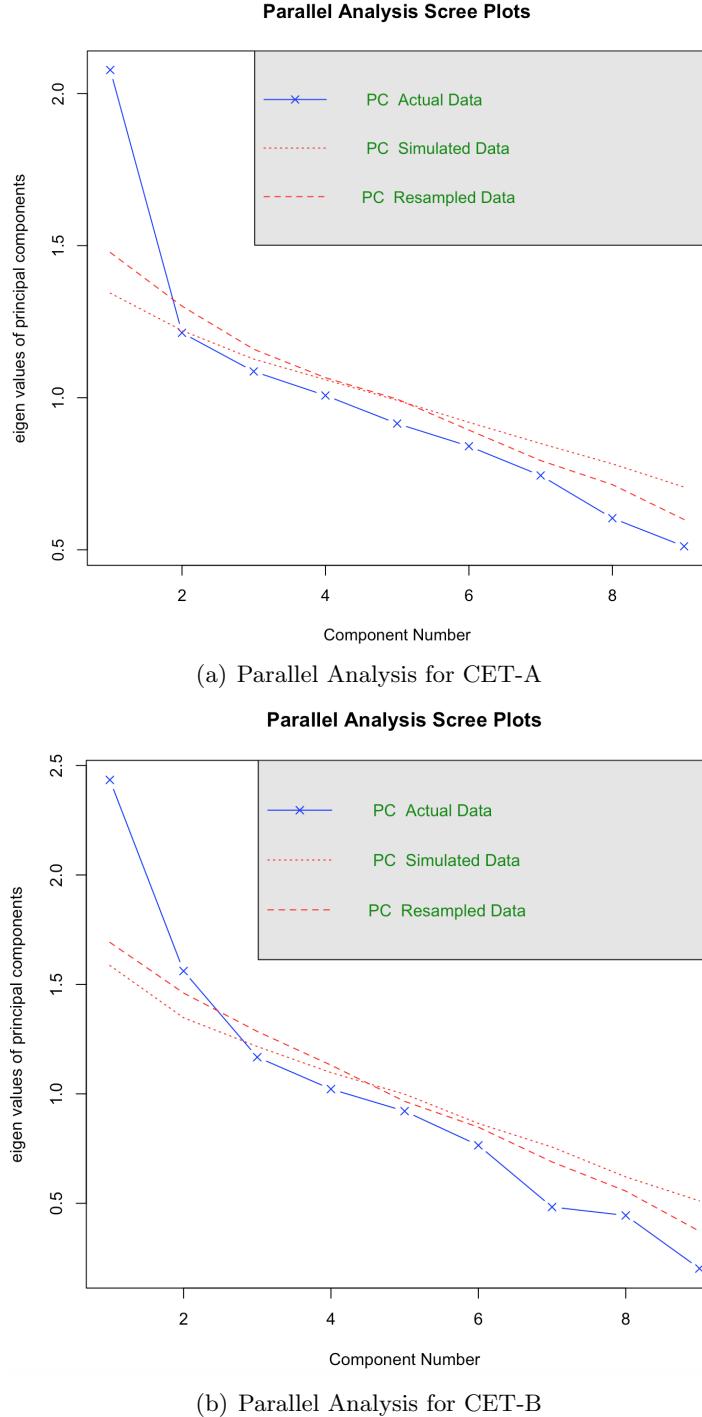


Figure 7: Parallel Analysis for CET-A and CET-B

### 4.3 Association between item scores and possible confounds

To find the associations between item scores and possible confounds, we consider Kruskal-Wallis test and one-way ANOVA Test. Both one-way ANOVA test and Kruskal-Wallis test can examine the significant difference on a continuous dependent variable. One-way ANOVA test assumes

that the dependent variable to be normally distributed and scores across groups have approximately equal variance while Kruskal-Wallis test do not have these assumptions and can be used for both continuous and ordinal data[10]. Hence, we choose Kruskal-Wallis test since it is a more general method.

Furthermore, Kruskal-Wallis test is a non-parametric test whose null hypothesis assumes that samples are from identical populations. Kruskal-Wallis test's distribution is similar to a Chi-square Test. If the p-value is less than the critical chi-square value, we cannot reject the null hypothesis; otherwise, we reject the null hypothesis and conclude that at least one of the samples comes from a distinct population.

## 5 Models and Results

Table 1 presents the p-values of scores in nine questions and different confounds such as age, years of education, and gender in CET-A and CET-B respectively. In CET-A, it is obvious that there are significant associations between Q1, Q3, Q6 and gender, Q4 and years of education, and Q9 and all of the confounds. In CET-B, gender and Q1 as well as Q2 have significant associations.

Table 1: P values of CET-A and CET-B

	CET-A			CET-B		
	Age	Education	Gender	Age	Education	Gender
Q1	0.486	0.523	<b>0.003</b>	0.727	0.520	<b>0.049</b>
Q2	0.238	0.102	0.495	0.186	0.284	<b>0.006</b>
Q3	0.287	0.580	<b>0.033</b>	1	0.975	0.122
Q4	0.124	<b>0.014</b>	0.33	0.991	0.709	0.247
Q5	0.911	0.655	0.886	0.652	0.412	0.371
Q6	0.070	0.681	<b>0.032</b>	0.869	0.981	0.353
Q7	0.153	0.711	0.848	0.806	0.262	0.256
Q8	0.366	0.136	0.293	0.306	0.800	0.400
Q9	<b><math>4.763 \times 10^{-5}</math></b>	<b>0.004</b>	<b>0.0167</b>	0.294	0.327	0.150

### 5.1 Differential Item Functioning (DIF)

Differential Item Functioning (DIF) analysis assumes that if minority examinees are not familiar with some test items, then these examinees have higher probability to give incorrect answer to these items although their overall cognitive ability are similar to the majority[14]. This assumption will lead to the lower score of the minority examinees compared to the majority and do not show their real cognition. Hence, DIF analysis is to identify these items which can be edited or deleted from the final version of the new test.

### 5.2 Multiple Group Estimation

A multiple-group (MG) IRT fixed-parameter estimation (FPE) method can be used to detect DIF and is based on item response theory (IRT)[11]. This theory is not similar to other simpler method. It creates scales and assesses the responses of questionnaire and it does not assume each item is of the same difficulty level.

In IRT, the scale for measuring ability (denoted as  $\theta$ ) is determined through a linear combination. The presumption of IRT is that the probability of a correct response is a function of an underlying ability. When a new test is used, we should establish scale to the examinee's ability distribution and the item parameters. If the new test contains items from old test and their IRT parameters are established pretty well, we should use the scale of the old items to estimate the

parameters for all other new items.

Under these assumptions, the fixed-parameter estimation (FPE) method using the expectation-maximization (EM) algorithm is employed iteratively to determine the ability distribution's probabilities and the new items' parameters[11].

Figure 8 and Figure 9 show the multiple group estimations of age, years of education and gender in CET-A and CET-B. We divide age group as young and old, education group as high education and low education, and gender as male and female. We can compare which group performs better based on the expected total score and higher score means more abnormality.

### 5.2.1 Age

The expected total scores and expected item scoring functions against ability for age are shown in Figure 8(a) and Figure 9(a) separately for CET-A and CET-B. In CET-A, the performance of young and old group is similar. Specifically, if the ability is higher than average, the performance is nearly the same while if the ability is lower than average, old group performs worse than young group. Moreover, young group and old group perform similarly in Question 3, 6, and 8.

In CET-B, if the ability is lower than average, the performance is more or less the same and if the ability is higher than average, old group performs better than young group. Young group and old group have hige different performance in Question 3, 4, 5 and 8.

### 5.2.2 Years of education

For education, the expected total scores and expected item scoring functions against ability are presented in Figure 8(b) and Figure 9(b) separately for two versions of CET. In CET-A if the ability is lower than average, low education group performs worse than high education group. However, if the ability is higher than average, high education group performs worse than low education group. Participants perform distinctively in Question 7 and 9.

In CET-B, there is a different result in education. If the ability is lower than average, high education group performed worse than low education group, but if the ability is higher than average, the result is on the contrary. Both education groups perform similarly in Question 3 and 7.

### 5.2.3 Gender

As for gender, we provide the expected total scores and expected item scoring functions against ability in Figure 8(c) and Figure 9(c) respectively for CET-A and CET-B. Male perform better than female generally speaking. And both gender groups perform quite differently in Question 4, 5 and 7 in CET-A but Question 3, 4, 6, 7 and 9 in CET-B. In CET-A, all three groups have considerable different performance in Question 7 while in CET-B, the results are different in Question 4.

To sum up, age, years of education and gender have more or less influences on the performance of both CET-A and CET-B; thus, we should leave out the effects of these three elements when building a model to choose the golden rule.

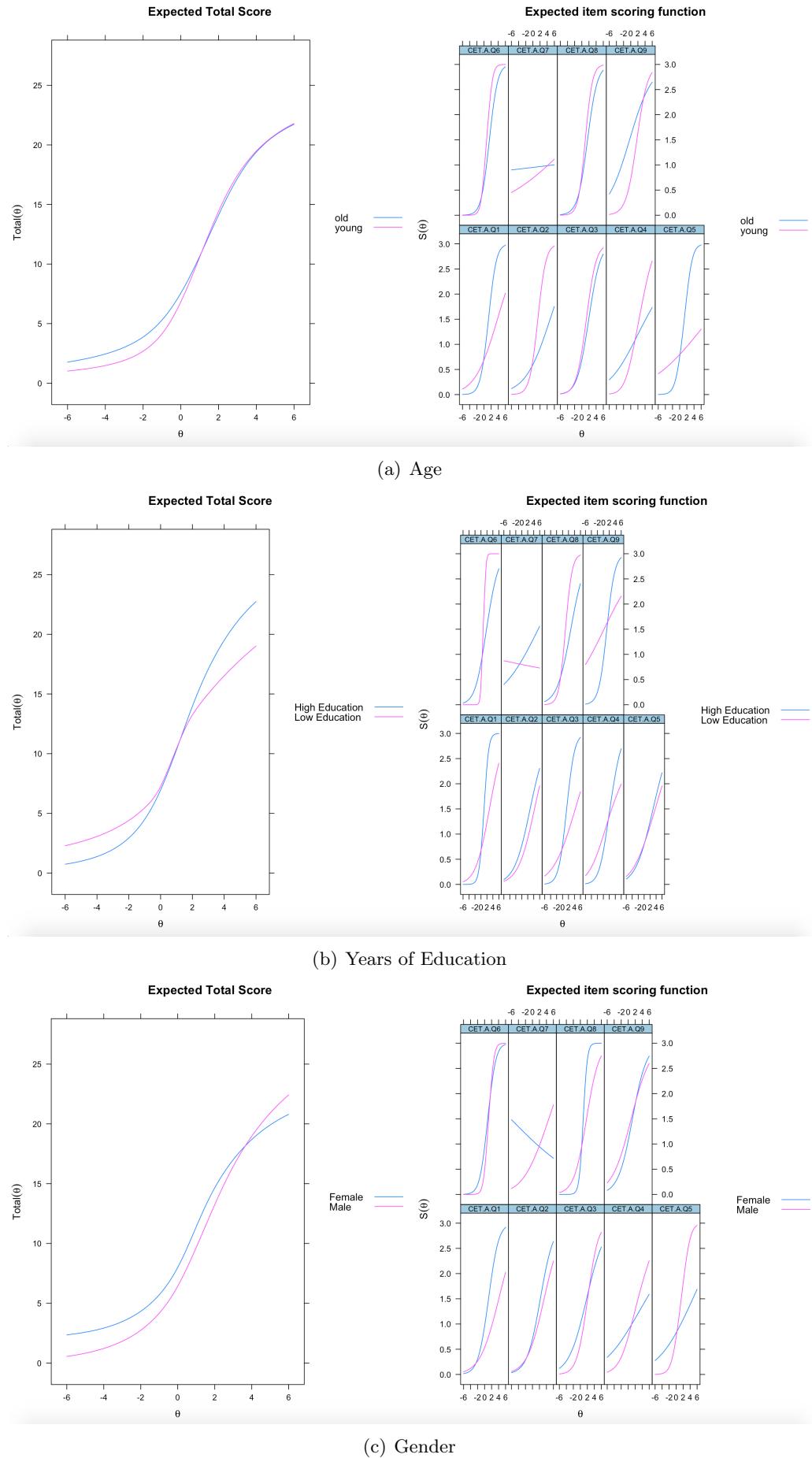


Figure 8: Association with Age, Years of Education and Gender in CET-A.

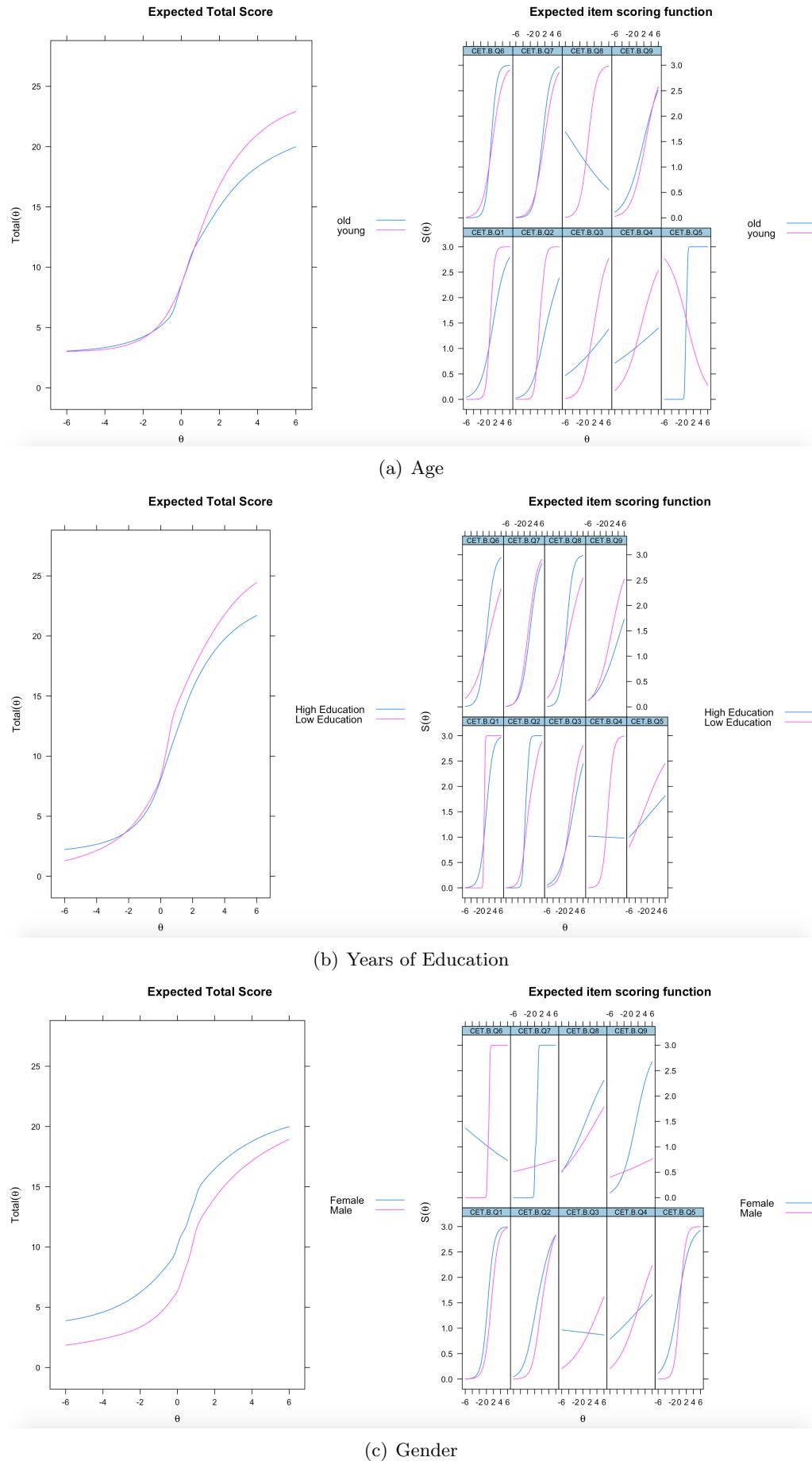


Figure 9: Association with Age, Years of Education and Gender in CET-B

### 5.3 The Golden Rule

Many variables often simultaneously have an effect on the dependent variable. Thus, the relationship between a dependent variable which we focus on and an explanatory variable can be confounded by other variables[2]. As mentioned before, some of the literature reviews compute an adjusted score that controls for gender, education and age.

We build a linear regression  $fit$  and take out the effects for individual items before the Kruskal-Wallis test. The response variable is CET items and the covariates are age, years of education and gender. We get residual of  $fit$  as the new items. Instead of the adjusted total score, we have adjusted item scores and calculate the p-values for the two versions of CET. The model is as following:

$$fit = lm(\text{Items} \sim \text{Age} + \text{Years of education} + \text{Gender})$$

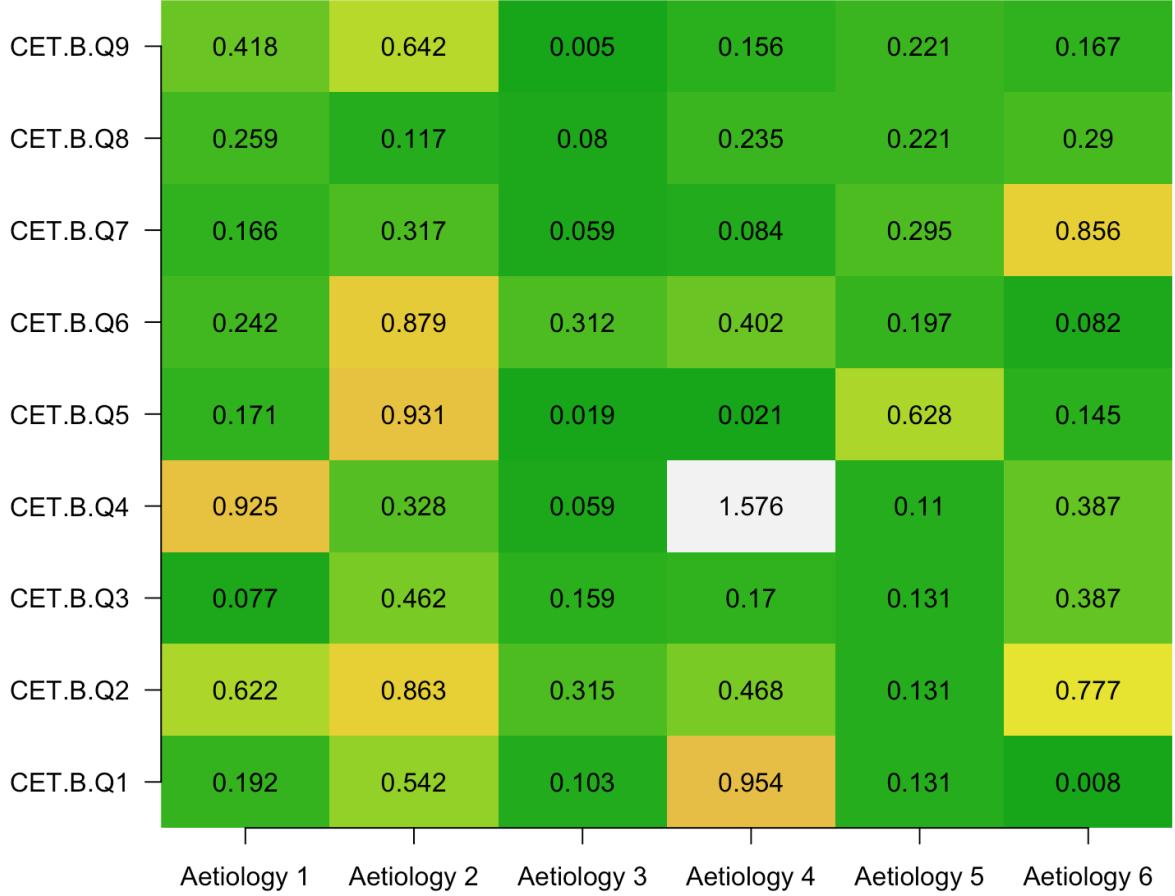
$$\text{New Item} = \text{residual}(fit)$$

In order to visualize the results more clearly, we calculate  $-\log_{10}\cdot p$  values of CET-A and CET-B and they are shown in Figure 10 and Figure 11 separately.

P-values of Question 5, 6 and 9 in CET-A for aetiology 5 and p-value of Question 4 are less than 0.05; therefore, the relationships between these questions and the aetiology are of great significance. Hence, Question 5, 6 and 9 in CET-A can distinguish Aetiology 5 and Question 4 in CET-B can distinguish Aetiology 4. We select Question 5, 6 and 9 from CET-A and Question 4 from CET-B to build a new golden rule.



Figure 10:  $-\log_{10}\cdot p$  values of CET-A

Figure 11:  $-\log_{10} \cdot p$  values of CET-B

### 5.3.1 Model checking

Our golden rule is used to classify Aetiology 4 and 5, so we build four logistic regression models to find whether our new model is better or not. Let  $C$  denote the participants in each group and let  $Y_i$  be the random variable which indicates whether patient  $i$  was diagnosed with some type of aetiology (eg. Aetiology 4 or 5) for each  $i$  in  $C$ . The distribution of the  $Y_i$  is in the following:

$$Y_i \sim \text{Bernoulli}(p_i) \quad \forall i \in C$$

Then we use a logit link function and the probabilities for  $p_i$  are determined by:

$$\text{logit}(p_i) = \log\left(\frac{p_i}{1-p_i}\right) = \mathbf{x}_i^T \boldsymbol{\beta}$$

Thus, we can yield:

$$p_i = \frac{\exp(\mathbf{x}_i^T)}{1 + \exp(\mathbf{x}_i^T)} = \frac{1}{1 + \exp(-\mathbf{x}_i^T)}$$

Using R code, we can easily get the coefficient  $\beta_j$  for every variable. If the value of  $\beta_j$  is positive, then the covariate and  $p_i$  has a positive relationship while negative coefficient means the opposite.

The first logistic regression is Aetiology 5 with respect to age, education, gender and CET items we choose. The second regression is Aetiology 5 with respect to all CET items as well as age, education, gender. The summaries of model 1 and 2 are given in Table 2 and those of model 3 and 4 are provided in Table 3. It is clear that the Akaike information criterion (AIC) for model 1 is 16 and for model 2 is 26. Since less value of AIC means better performance, the

first model is better than the second one and we conclude that the new model is better.

Similarly, the third regression is Aetiology 4 with respect to age, education, gender and CET items in the golden standard; the AIC is 16. The second regression is Aetiology 4 with respect to all CET items adding age, education, gender; the AIC is 26. Also, the new model is better than the old one.

## 6 Conclusion

Through the exploratory data analysis, we find that there are many latent variables which make the relationships between each CET item not obvious. We conduct the Principal Component analysis (PCA) and obtain nine principal components. After the Parallel Analysis, we conclude that PC1 is retained in CET-A, and PC1 as well as PC2 are retained in CET-B. We find some significant associations in age, years of education and gender with individual items of CET-A and CET-B. In order to account for the confounding effect, we build a linear regression for individual items with confounding covariates and take the residuals as new items for further analysis. We leave out the effects of age, years of education and gender before the Kruskal Wallis test and choose Question 5, 6 and 9 in CET-A and Question 4 in CET-B to build a new golden rule.

### 6.1 Limitations

One limitation of the project is that the amount of data used is small and it is hard to perform a model check with the limited data. In detail, only 14 people completed both CET-A and CET-B. If we use the data, it is not accurate for the model checking. There are other model checking methods other than AIC if more data were given and the number of cases and controls are balanced.

The second limitation is that there is no healthy-control group and we can not compare the performance of healthy people and those who suffered from neurological disease. Also, we cannot understand the influences of the neurological disease to the patients.

The third limitation is that we always choose threshold of p-value as 0.05 but in some cases in the project, the threshold could be chosen to be bigger since the number of items we select is small. However, we are not sure which p-value should we use to make the result more accurate.

### 6.2 Future Directions

The data set contains a lot of other information such as anxiety and depression. We can conduct another study and come up with a CET that will distinguish people anxiety/depression from normal people.

Moreover, if we are interested in exploring the predictive power of the items with respect to different machine learning methods in classification between controls and people with brain lesion, we can extend or conduct analysis from logistic regression to decision tree and supported vector machine or even artificial neural network. In addition, we can perform k-fold cross validation for different machine learning classification methods to assess the accuracy and sensitivity of the predictions.

Table 2: Logistic regression for model 1 and 2

	<i>Dependent variable:</i>	
	Aetiology for category 5	
	(1)	(2)
Age	0.870 (13, 122.030)	1.258 (11, 153.490)
Years.of.Edu	-29.639 (107, 429.800)	-9.292 (52, 548.100)
Gender	-12.174 (47, 461, 104.000)	-0.890 (166, 254.100)
CET.A.Q1		36.005 (202, 099.800)
CET.A.Q2		-0.589 (107, 456.500)
CET.A.Q3		32.510 (198, 607.700)
CET.A.Q4		14.661 (133, 340.700)
CET.A.Q5	-3.721 (1, 039, 448.000)	-42.280 (244, 273.300)
CET.A.Q6	-4.016 (463, 112.700)	-12.056 (64, 145.280)
CET.A.Q7		-28.363 (122, 079.600)
CET.A.Q8		-15.644 (76, 996.010)
CET.A.Q9	-25.529 (136, 525.700)	-7.615 (62, 832.440)
CET.B.Q1		
CET.B.Q2		
CET.B.Q3		
CET.B.Q4	-26.400 (47, 461, 704.000)	
CET.B.Q5		
CET.B.Q6		
CET.B.Q7		
CET.B.Q8		
CET.B.Q9		
Constant	305.206 (47, 464, 140.000)	41.779 (961, 313.200)
Observations	13	13
Log Likelihood	-0.000	-0.000
Akaike Inf. Crit.	16.000	26.000

Note:

\*p&lt;0.1; \*\*p&lt;0.05; \*\*\*p&lt;0.01

Table 3: Logistic regression for model 3 and 4

	<i>Dependent variable:</i>	
	Aetiology for category 4	
	(3)	(4)
Age	-0.913 (3, 917.609)	-0.102 (11, 153.490)
Years.of.Edu	-5.272 (77, 248.920)	3.999 (52, 548.090)
Gender	14.673 (47, 453, 355.000)	-14.485 (166, 254.100)
CET.A.Q1		-21.035 (202, 099.700)
CET.A.Q2		1.166 (107, 456.500)
CET.A.Q3		8.099 (198, 607.700)
CET.A.Q4		4.403 (133, 340.700)
CET.A.Q5	8.982 (155, 865.400)	14.172 (244, 273.300)
CET.A.Q6	-1.160 (73, 672.830)	2.214 (64, 145.270)
CET.A.Q7		6.629 (122, 079.600)
CET.A.Q8		1.528 (76, 995.990)
CET.A.Q9	-0.596 (64, 551.020)	2.973 (62, 832.440)
CET.B.Q1		
CET.B.Q2		
CET.B.Q3		
CET.B.Q4	-15.555 (47, 453, 306.000)	
CET.B.Q5		
CET.B.Q6		
CET.B.Q7		
CET.B.Q8		
CET.B.Q9		
Constant	77.249 (47, 463, 139.000)	-62.171 (961, 313.200)
Observations	13	13
Log Likelihood	-0.000	-0.000
Akaike Inf. Crit.	16.000	26.000

Note:

\*p&lt;0.1; \*\*p&lt;0.05; \*\*\*p&lt;0.01

## References

- [1] Ken Aho, DeWayne Derryberry, and Teri Peterson. “Model selection for ecologists: the worldviews of AIC and BIC”. In: *Ecology* 95.3 (2014), pp. 631–636.
- [2] Heiko Becher. “The concept of residual confounding in regression models and some applications”. In: *Statistics in medicine* 11.13 (1992), pp. 1747–1758.
- [3] Sorana-Daniela Bolboaca and Lorentz Jäntschi. “Pearson versus Spearman, Kendall’s tau correlation analysis on structure-activity relationships of biologic active compounds”. In: *Leonardo Journal of Sciences* 5.9 (2006), pp. 179–200.
- [4] Lisa Cipolotti et al. “Cognitive estimation: Performance of patients with focal frontal and posterior lesions”. In: *Neuropsychologia* 115 (2018), pp. 70–77.
- [5] Ömay Çokluk and Duygu Koçak. “Using Horn’s Parallel Analysis Method in Exploratory Factor Analysis for Determining the Number of Factors.” In: *Educational Sciences: Theory and Practice* 16.2 (2016), pp. 537–551.
- [6] KAREN GRACE-MARTIN. *The Fundamental Difference Between Principal Component Analysis and Factor Analysis*. July 2019. URL: <https://www.theanalysisfactor.com/the-fundamental-difference-between-principal-component-analysis-and-factor-analysis/>.
- [7] Francisco Pablo Holgado-Tello et al. “Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables”. In: *Quality & Quantity* 44.1 (2010), p. 153.
- [8] Karl G Jöreskog. “On the estimation of polychoric correlations and their asymptotic covariance matrix”. In: *Psychometrika* 59.3 (1994), pp. 381–389.
- [9] Sarah E MacPherson et al. “Bringing the cognitive estimation task into the 21st century: normative data on two new parallel forms”. In: *PLoS One* 9.3 (2014), e92554.
- [10] Patrick E McKnight and Julius Najab. “Kruskal-wallis test”. In: *The corsini encyclopedia of psychology* (2010), pp. 1–1.
- [11] Established Ability Scale, Seonghoon Kim, and Michael J Kolen. “Center for Advanced Studies in Measurement and Assessment CASMA Research Report”. In: (2016).
- [12] BG Tabachnik and LS Fidell. “Principal Components Analysis and Factor Analysis”. In: *Using Multivariate Statistics* (), pp. 372–445.
- [13] Gabriela Peretti Wagner et al. “Cognitive estimation abilities in healthy and clinical populations: the use of the Cognitive Estimation Test”. In: *Neurological sciences* 32.2 (2011), pp. 203–210.
- [14] Cindy M Walker. “What’s the DIF? Why differential item functioning analyses are an important part of instrument development and validation”. In: *Journal of Psychoeducational Assessment* 29.4 (2011), pp. 364–376.

## Appendices

### A Data: further information

We obtained the data from CET-A and CET-B respectively and deleted all rows with the missing values. For CET-A, the data contains 198 participants and for CET-B, the data includes 75 patientst. We performed Polychoric correlations, PCA and Parallel Analysis on two versions of CET separately based on the two data sets.

In order to perform Kruskal-Wallis H test on nine items in each CET against age, education and gender respectively, we obtained and cleared six data sets:

- 1) Age and nine CET-A items
- 2) Years of education and nine CET-A items
- 3) Gender and nine nine CET-A items
- 4) Age and nine CET-B items
- 5) Years of education and nine CET-B items
- 6) Gender and nine nine CET-B items

We also conducted Multiple Group Estimations on these six data sets. Moreover, the numbers of people in these data sets are 198, 194, 198, 75, 73 and 75 respectively.

We built another two data sets to estimate p-values for nine items against six aetiologies and found the final golden-standard rule. In the first data set for CET-A, we chose age, years of education, gender and nine CET-A items and left out the rows with all missing values. We built the second data set with age, years of education, gender and nine CET-B items. The numbers of participants are 194 and 73. For model checking, we obtained and cleared data of age, education, gender and all items in CET-A and CET-B. There are only 14 patients in the data set and the result may be less accurate due to the limited data set.

### B Methods for model checking

#### Akaike Information Criterion

AIC measures the quality of each model with respect to other models. Let k denotes the number of estimated parameters in the model and let  $\hat{L}$  denote the maximum likelihood of the model. The formula of AIC is in the following[1]:

$$AIC = 2k - 2\ln(\hat{L})$$

We prefer the model which has the minimum AIC value in all of the models. AIC rewards goodness of fit and penalties functions which have increasing number of estimated parameters. In addition, the model doses not encourage over-fitting.

### C R code and packages used

We listed the R packages used in our data analysis and gave the complete version of R code in the link below.

#### R packages

To calculate the Polychoric correlation between two ordinal variables, we used the **polycor** package. We installed **ggcorrplot** to plot the heatmap for polychoric correlation and calculated p-values after barring the non-significant coefficient. The Bi-plots were produced using the **ggbiplot** package. In addition, Parallel Analysis was performed with the package **psych** and Multiple Group Estimation was conducted with the package **mirt**. Moreover, we plotted the

result of the Multiple Group Estimation using the **gridExtra** package and generated regression tables into LATEX format using the **stargazer** package.

### R code and files

It would make the report messy if we include the R code directly as a text. Therefore, the complete code can be downloaded through the link:

<https://www.dropbox.com/sh/cwbxusynxbmxsl1/AAAURXrHFjsEPHfLPXv-JSWua?dl=0>

The data file can be downloaded in the link:

<https://www.dropbox.com/s/nuu198hgwl8f8e/CET.csv?dl=0>