Dimensionality Analysis of the Uniform CPA Examination

Emre Gönülateş

August 15, 2014

Contents

Abbreviations

AT Assessment Test. 23, 24

AUD Auditing and Attestation. 11, 28, 30, 32, 35, 38, 42, 50, 51, 53

BEC Business Environment and Concepts. 11, 28, 34, 39, 45, 50

CFA confirmatory factor analysis. 17, 18, 20, 28, 34, 35, 37

CFI comparative fit index. 18, 19, 30, 35–44

CPA Certified Public Accountants. 11–13, 16, 28, 56, 58, 59

DETECT Dimensionality Evaluation to Enumerate Contributing Traits. 16, 22, 25–28, 32, 49–51, 56, 57, 59

DIF differential item functioning. 13

DimPack DIMTEST, DETECT and HCA/CCPROX. 16, 22, 24, 25

EFA exploratory factor analysis. 17–19, 30, 37, 38, 42

FAR Financial Accounting and Reporting. 11, 28, 35, 40, 43, 50, 51

GFI Tanaka's goodness-of-fit index. 21, 32, 33

HCA hierarchical cluster analysis. 16, 25

IRT item response theory. 13, 21, 22

MCQ multiple choice questions. 2, 11, 28–30, 32, 34–36, 38, 45, 46, 49, 50, 53–57

MIRT multidimensional item response theory. 17

ML maximum likelihood. 17

MST multistage test. 11

NOHARM Normal-Ogive Harmonic Analysis Robust Methods. 16, 21, 28, 32, 33, 53, 56, 57, 59

PT Partitioning Test. 23, 24

REG Regulation. 11, 28, 35, 37, 41, 44, 46, 49–51

RMSEA root mean square error of approximation. 18, 19, 30, 35–44, 57

RMSR root mean square residual. 21, 32, 33, 53, 54

SEM structural equation modeling. 16, 17, 20

TBS task based simulations. 2, 11, 28–30, 32, 34, 35, 37, 42, 45–49, 51, 53–55, 57, 58

TLI Tucker−Lewis index. 18, 19, 30, 35–44

ULS unweighted least squares. 21

WLSMV weighted least squares with mean and variance adjusted. 17

List of Figures

List of Tables

Chapter 1

Introduction

1.1 The uniform CPA examination

The uniform Certified Public Accountants (CPA) examination (CPA exam) is one of the requirements for CPA licensure. It consists of four independently scored sections: Auditing and Attestation (AUD), Business Environment and Concepts (BEC), Financial Accounting and Reporting (FAR), and Regulation (REG). Total testing time of these four sections is 14 hours. Examinees should pass all four sections within 18 months.

CPA exam is a computer based multistage test (MST) (Hendrickson, 2007). For security reasons each examinee routed to a different panel (Figure 1.1.1). Each panel consists of four stages. AUD, FAR and REG sections have three stages with MCQ and a fourth stage with TBS. For BEC section, the last stage consists of constructed response (CR) items. At stage 2 and 3, examinees are routed to either medium or a hard difficulty MCQ testlet depending on their performance in the previous stage. In a panel every examinee take same testlets in stage 1 and 4. There are 20-25 items in each MCQ testlet, 5-6 simulations in TBS testlet and 2 writing prompts in CR testlet. In addition to operational items there are pretest items within each testlet that are not scored.

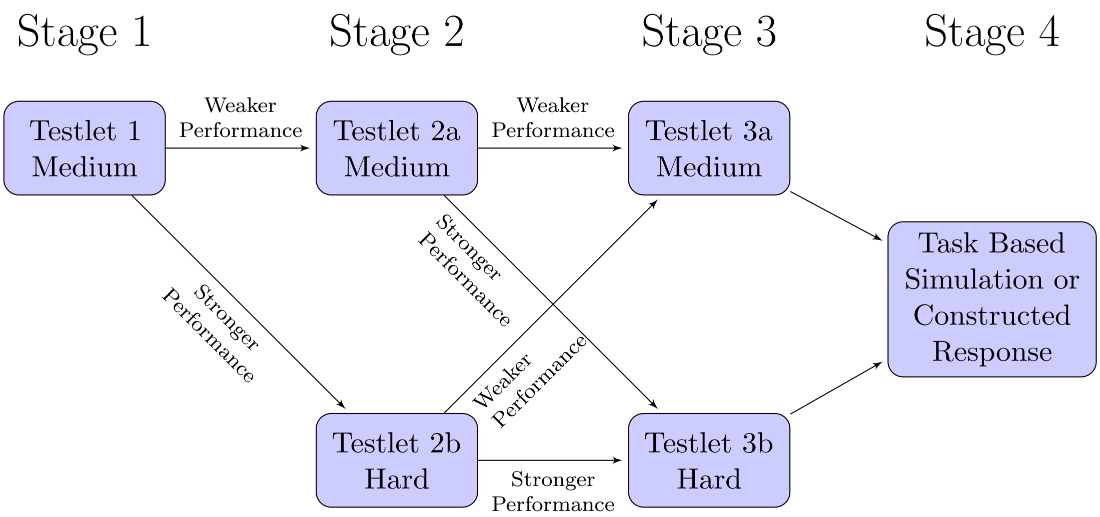

Figure 1.1.1: Structre of the CPA examination. Illustration of a panel.

1.2 Purpose of the Study

The items in the uniform CPA examination (CPA exam henceforth) cover a broad range of content areas (American Institute of Certified Public Accountants [AICPA], 2012). The format of the items are different. Nevertheless, for each section a single score is reported. It is important to investigate the relationship between these various content domains and support the scoring of the test. The examination of the dimensionality of CPA exam will provide a validity evidence for the interpretations and use of the scores (Kane, 2013).

Dimensionality analysis is the assessment of the structure of the test data. It is important to know the dimensional structure of the test data. This allows test developers to determine which content domains are tested, cognitive processes measured by the items and the relationships between them. It supports the claims regarding the assumptions of the models used, such as unidimensionality. Such analysis elicit the construct irrelevant content and substantiate the validity of the test.

The purpose of the dimensionality analysis is to determine the number of dimensions representing test data (Reckase, 2009). If the aim is to understand the underlying relationships and the structure of the test data then overestimating the number of dimensions is recommended. On the other hand, overestimating dimensions means estimating more parameter parameters. This will decrease the accuracy of the parameter estimates. So, if the purpose is to estimate ability parameters, where accuracy of estimates is important, minimum number of dimensions preferred.

Most of the current operational tests use item response theory (IRT) to score examinee responses. IRT is a strong test theory model, it provides very useful information regarding examinees, given the model is fitting the data. To ensure model fit, certain assumptions should be met. Local independence is one of the assumptions. Two item responses should be conditionally independent for a given ability vector. If this ability vector has only one dimension, test is unidimensional, otherwise it is multidimensional. Local independence assumption will not be met if the dimensionality of the test is not correctly specified. Examinee scores will not be fair and rankings will be unjustified.

It is crucial to discover and then eliminate unintended multidimensionality in the test data. Unintentinal multidimensionality can cause a lot of psychometric problems. Reckase (2009) pointed out that representing a multidimensional data with lower number of dimensions may lead to incorrect test score interpretations and wrong decisions regarding examinee abilities. In their study, Walker and Beretvas (2003) found that improperly scoring a multidimensional math test data as unidimensional cause incorrect classifications of examinees.

Another problem that may arise from unintended dimensionality is item calibration errors. Masters (1988) showed that unwanted dimensions might inflate the values of item discrimination parameters. Consequently, this will inflate the test information (Thissen, Steinberg, & Mooney, 1989) which will cause an overestimation of the reliability coefficient (Sireci, Thissen, & Wainer, 1991).

Differential item functioning (DIF) is another consequence of unintended multidimensionality. DIF can be conceptualized as an undesired dimensionality of the test data (Ackerman, 1992). Detection of such extraneous dimensionality might help DIF studies. Even though testing programs are very careful in their endeavor to detect items that are exhibiting DIF, these analyses might sometimes limited to checking item difficulty differences between groups, but the group differences in item discrimination parameters might not be detected (Swaminathan & Rogers, 1990). This will potentially cause item calibration errors.

Finally, unintended dimensionality can cause problems in test equating. Equating practices that relies on precise parameter estimates might be affected. Anchor items are supposed to represent whole test. If test structure is not well known, anchor items might not be the mini version of the test (Tate, 2003).

In sum, current dimensionality analysis of CPA exam will (1) substantiate the validity claims of the exam scores, (2) check whether the unidimensionality assumption required by the currently used three parameter logistic IRT model is tenable and (3) describe the dimensionality of the test data.

Chapter 2

Literature Review

Stout (1990) stated the traditional definition of the test dimensionality as the minimum number of dimensions required for Θ to produce a model which is both locally independent and monotonically increasing, where Θ represents the multidimensional latent variable that is underlying the item responses. If Θ consists of one component, test is unidimensional, otherwise it is multidimensional. McDonald (1981) made a similar definition of dimensionality. A set of n tests has r dimensions if the residuals of the regression of these n tests on r latent traits are uncorrelated. Here a test can be an item or a collection of items.

It is important to note that "dimensionality is the property of data matrix, not the test"(Reckase, 2009, p. 193). A test might have a different dimensional structure in one sample and a different dimensional structure in another sample. Having this in mind, in the rest of this report 'test' and 'test data' might be used interchangeably.

Dimensionality analyses can be run in an exploratory or confirmatory mode. In exploratory mode, there are few or no prior expectations about the dimensional structure of the test data. The aim of the analysis is to find the optimum number of dimensions that explains the relationships between test items. Also, such analyses may complement the confirmatory analyses by exposing unexpected structural features of the test data. But, since exploratory analyses lack substantive framework, it might be difficult to explain and make sense of the dimensions extracted from the test data. In addition, exploratory analyses are not designed to detect hierarchical factor structure in the test, such as bifactor models. Non-simple structures are difficult to detect because default rotation methods are devised to transform the solution to a simple structure (Tate, 2003).

On the other hand, in confirmatory analyses, researcher has a model in mind to explain the test data. Hypotheses are generated using the model and then they are evaluated using the fit statistics. Often the pattern of the relationships between the items and the dimensions are tested. The model to explain the test data is formed by a theory or a substantive framework that guided the development of the test. For example, test specifications or content distribution of the items might guide the formation of the substantive framework.

Many tools has been developed for analyzing the structure of the test data. Hattie (1985) and Tate (2003) gave an overview of the methods that has been used up to their times. The article of Tate (2003) presents a comprehensive list of methods for dimensionality analysis. Also, there are many studies that applied some of these methodologies and compared them. The methods used in this report has been selected in the light of these studies.

Tate (2003) compared the performance of exploratory and confirmatory factor analysis using Mplus, full-information item factor analysis using TESTFACT, non-linear factor analysis using Normal-Ogive Harmonic Analysis Robust Methods (NOHARM), hierarchical cluster analysis (HCA), DIMTEST, Dimensionality Evaluation to Enumerate Contributing Traits (DETECT) and a host of local item dependency tests using real and simulated test data.

Gierl, Tan, and Wang (2005) analyzed the dimensionality and structure of the SAT mathematics test. They used DIMTEST, DETECT and NOHARM in their analysis. In addition, they used Cattells Scree test, the Kaiser rule, and the minimum average partial methods to determine the number of dimensions in the test data. Jasper (2010) also combined the parametric and nonparametric methods in his analysis. He investigated the structure of START-M mathematics test using DIMTEST, DETECT and NOHARM. He also used structural equation modeling (SEM) based confirmatory methods if data showed a complex structure. Stone and Yeh (2006) analyzed the structure of the Multistate Bar Examination using NOHARM, Mplus and TESTFACT. They did not find significant differences between these methods when guessing was not modeled.

Stout et al. (1996) and Douglas, Kim, Roussos, Stout, and Zhang (1999) used DIMTEST, DETECT and HCA/CCPROX (DimPack) to analyze the dimensional structure of the LSAT exam. Jang and Roussos (2007) also used these three methods to investigate the dimensional structure of the TOEFL exam. In a recent study, Seol (2013) used DETECT to evaluate the dimensional structure of the CPA exam.

2.1 Parametric Methods

Parametric methods to evaluate the dimensionality of the test data comes from either a factor analytic framework or multidimensional item response theory (MIRT) framework. These two frameworks can be seen analogous under certain conditions (Takane & de Leeuw, 1987). In parametric methods, models are summarized by model parameters and there are assumptions regarding the data set. In MIRT models, depending on the model, threshold and slope parameters are estimated in addition to the guessing parameters. In factor analysis models, factor loadings and threshold parameters are estimated. Factor analysis models can be divided into two: linear and non-linear factor analysis. This split depends on how the relationship between latent variable and probability of correct response is modeled.

2.1.1 Linear Factor Analysis - Mplus

Mplus (L. K. Muthén & B. O. Muthén, 1998-2012) software is used for both exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). Mplus uses robust weighted least squares method (i.e. weighted least squares with mean and variance adjusted (WLSMV)) for estimation of binary data. The ubiquitous maximum likelihood (ML) estimation requires multivariate normality which might be problematic with binary data. WLSMV proposed by B. O. Muthén, Du Toit, and Spisic (1997) is an alternative for estimating ordinal data which includes binary data. WLSMV gives acceptable solutions even for small sample sizes like 200 (B. O. Muthén et al., 1997). Beauducel and Herzberg (2006) compared ML and WLSMV estimation of CFA models with moderate loading sizes, different numbers of categories and factors. They found that WLSMV performed as good as ML for small and large sample sizes. Especially for categorical variables with two or three categories, authors found that WLSMV is superior to ML estimation.

In SEM literature there are numerous indices to test the model fit (Brown, 2006; Kline, 2010). Mplus prints some of these indices. In the following paragraphs some of them will be explained.

Chi-square fit statistic (χ^2_M) tests the hypothesis stating: observed and expected correlation matrices are not significantly different. If a model is correctly specified chi-square fit statistic is expected to be equal to the degrees of freedom of the model. A p-value smaller than 0.05 implies that the hypothesis test is rejected and the model is not fitting the data. If chi-square fit statistic is not statistically significant, researcher can conclude that the model is consistent with the covariance structure of the data, hence model fits the data. Here it is important to recognize that the correct model is unknown, this statistic merely tells researcher that the covariance structure is correctly estimated by the model.

One limitation of the chi-square fit statistic is that, it rejects the model even for small departures from the perfect fit. Thus, researchers generally look for alternative methods to evaluate the fit of the EFA/CFA models. For example, Tate (2003) checked the model fit using the ratio of the chi-square fit statistic to the degrees of freedom (χ^2_M/df), although he admitted that this is a crude statistic. For the conditions he studied, computer generated data suggested that values of χ^2_M/df that were larger than 1.4 implied a non-trivial model misfit. Kline (2010) does not recommend the usage of such statistics (he called them normed chi-square) due to the arbitrary nature of the cut scores for acceptable model fit, such as 1.4.

Root mean square error of approximation (RMSEA) (Steiger, 1990) is an approximate fit index to evaluate model fit. It can be calculated using the following formula:

$$RMSEA = \sqrt{\frac{\chi^2_M - df_M}{df_M \cdot (N-1)}} \tag{2.1.1}$$

where df_M is the degrees of freedom of the model and N is the sample size. As can be seen from the equation RMSEA corrects the chi-square fit statistic for the complexity of model. In addition, it is not affected by the sample size. Mplus also prints the lower and upper bounds of the 90% confidence interval for the RMSEA. Values of RMSEA less than 0.05 indicates an approximate model fit (Hu & Bentler, 1999). If the upper bound of the confidence interval indicates poor fit, researcher should be suspicious about the fit of the model.

The two other fit indices printed by Mplus are comparative fit index (CFI) and Tucker−Lewis index (TLI). According to Hu and Bentler (1999), values of CFI and TLI lower than 0.95 indicates a poor fit. The formulation of these two indices are:

$$CFI = 1 - \frac{\max[\chi^2_M - df_M, 0]}{\max[\chi^2_M - df_M, \chi^2_B - df_B, 0]} \tag{2.1.2}$$

$$TLI = \frac{\frac{\chi^2_B}{df_B} - \frac{\chi^2_M}{df_M}}{\frac{\chi^2_B}{df_B} - 1} \tag{2.1.3}$$

where χ^2_B and df_B are the chi-square test statistic for the baseline model and the degrees of freedom of the baseline model, respectively. A limitation of Mplus is that it cannot include a guessing parameter in model estimation. This may lead to model misspecification (Tate, 2003). On the other hand, Mplus can handle the missing data, which is not available in most of the methods mentioned above.

2.1.2 Exploratory Factor Analysis

Exploratory analyses are used when researcher does not have a clear idea about the dimensionality structure of the test data or when an unconstrained solution is needed. Exploratory analyses might not detect test structure when there is a complex factor structure (like hierarchical factor structure or bifactor structure) or the dimensionality of the test data is not simple (Tate, 2003).

In Mplus, user defines the number of factors to be extracted from the test data. EFA in Mplus does not provides a number for the count of dimensions, instead researcher determines the number of dimensions using the relevant output. For each solution, Mplus provides a chi-square fit statistic for model fit, RMSEA, CFI and TLI along with other fit indices.

2.1.3 Confirmatory Factor Analysis (Mplus)

In general, it is advised that if there is a prior belief regarding the test structure, confirmatory analyses perform better compared to exploratory methods (Tate, 2003). For instance, if the test structure is rather complex exploratory analyses might not disclose that structure (Tate, 2003).

If test data is believed to be unidimensional, then a one factor model should fit the test data. An example of one-factor (unidimensional) model for four items is in Figure 2.1.1. The λ's in the figure are the factor loadings and ε's are the unique variances. There are no two-way paths between items, which means there is no local dependence between items.

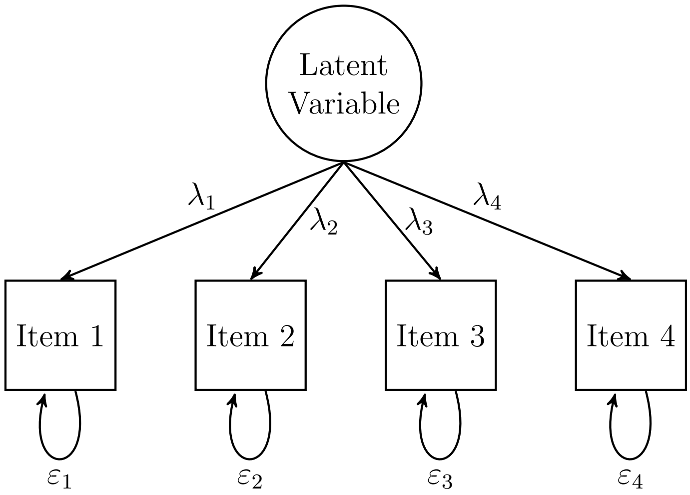

Figure 2.1.1: Path Diagram for One-Factor Model

Mplus prints various fit indices to evaluate model fit. Some of them has been discussed in section 2.1.1 on page 17. Tate (2003) argues that most of the indices printed by Mplus and their rule of thumbs developed and tested for full SEM (Kline, 2010) and may not be appropriate for confirmatory factor analyses. But due to the wide usage of these fit indices in CFA models in literature (Brown, 2006), they were used in this study as well.

Both confirmatory and exploratory factor analysis methods use covariance matrices. If there is guessing in data, the correlations between items might be attenuated compared to their true values. But clearly the effect of guessing depends on the nature of the guessing. A random (or blind) guessing model (Lord & Novick, 1968; Nunnally & Bernstein, 1994) affect the correlations differently compared to a knowledgeable guessing behavior. Mplus does not allow users to enter guessing parameters. This is a big limitation of Mplus (Tate, 2003).

Another limitation of factor analysis models is the assumption of a compensatory model where the lack of ability in one dimension can be compensated by another dimension. In practice this assumption might not reflect the reality.

2.1.4 Nonlinear Factor Analysis (NOHARM)

Classical factor analysis assumes a linear relationship between the latent response variable and multiple latent abilities. On the other hand, non-linear factor analysis (McDonald, 1962, 1981, 1997) assumes a non-linear relationship between latent response variable and the latent abilities of examinees that generate the response patterns. In non-linear factor analysis implemented in NOHARM, the probability of a correct response to an item y_i can be written as:

$$P\left[y_i = 1 \,|\boldsymbol{\theta} = (\theta_1, \ldots \theta_j)\right] = c_i + (1 - c_i) \cdot \phi \left(f_{i0} + f_{i1}\theta_1 + \ldots + f_{ij}\theta_j\right) \quad (2.1.4)$$

where $\boldsymbol{\theta} = (\theta_1, \ldots \theta_j)$ is the latent ability vector, $f_{i1}, \ldots, f_{ij}$ are the factor loadings, f_{i0} is the threshold parameter, c_i is the guessing parameter of ith item and ϕ is the normal distribution function.

Non-linear factor analysis is implemented in NOHARM program (Fraser & McDonald, 2012). Both exploratory and confirmatory models can be fitted using this program.

Parameters in NOHARM are estimated using unweighted least squares (ULS) method. One advantage of ULS is the ease of parameter estimation even for large number of items. But the disadvantage is the lack of direct standard errors of parameter estimates that can aid the assessment of the model fit (Gierl et al., 2005).

NOHARM program offers two indices to evaluate the model fit. One is the root mean square residual (RMSR) and the other is Tanaka's goodness-of-fit index (GFI). GFI can be defined as:

$$GFI = 1 - \frac{\mathrm{Tr}\left(R^2\right)}{\mathrm{Tr}\left(C^2\right)} \quad (2.1.5)$$

where R is the residual covariance matrix, C is the sample covariance matrix and 'Tr' is the trace of a matrix. Various methods developed for the evaluation the goodness-of-fit of the NOHARM solution will be discussed further in Section 3.4 on page 32.

Nonlinear factor analysis has many advantages (De Champlain & Tang, 1997). First of all, it assesses the dimensionality from the same framework as the IRT models. One can fit different complex models. Using guessing parameters one can fit three-parameter models, or by setting guessing parameters to 0, two parameter model can be fitted.

Tate (2003) found that the the performance of the exploratory analysis in NOHARM is very good. NOHARM recovered well to the true model in ten of the twelve simulation cases. The two exceptions were the cases with large discrimination parameters. The results reported by Stone and Yeh (2006) suggested the use of NOHARM when guessing is relevant for the testing situation. These authors did not find a significant difference between NOHARM and TESTFACT when guessing is modeled.

2.2 Nonparametric Methods

Parametric models depends on assumptions to explain data better. In an IRT model these assumptions are monotonicity, undimensionality (or correct specification of dimensionality) and local independence of items (Hambleton & Swaminathan, 1985). In practice, these assumptions might not be met or weakly achieved. In such cases model data fit compromised and more flexible models are needed. Nonparametric methods are handy in those situations (Junker & Sijtsma, 2001).

Tate (2003) pointed out that the significant cost of using parametric models is the strong assumptions regarding the underlying functional form. Violations of this assumption cause biased parameter estimates and incorrect specification of test dimensionality. For example, in his study Tate (2003) found when the assumed model does not allow guessing, results regarding test structure came out distorted. On the other hand, nonparametric methods does not depend on such restrictive models. For nonparametric models, the cost of this freedom from assumptions is the lack of details about the test structure which are provided by the mathematical delineation specified by parametric models. Another advantage of parametric models over nonparametric ones is the sample size requirements. Parametric methods requires large samples sizes to get reliable estimates of parameters and fitting models. Nonparametric models can be used with small sample sizes.

All of the nonparametric analyses used in this report is implemented in DimPack. Non-parametric methods in DimPack treats dimensionality as a whole. HCA/CCPROX searches for clusters of homogenous items using cluster analysis. DIMTEST tests whether the test data is essentially unidimensional and if it is not, DETECT calculates the extent of multidimensionality in the test data. As Stout et al. (1996, p. 351) points out: "Each of the three procedures assesses a different aspect of test multidimensionality, and together they provide an almost complete summary of a tests dimensional characteristics".

In this study two of these non-parametric methods has been used. For the test of essential unidimensionality of items, DIMTEST has been used. If the essential unidimensionality is not tenable, the extent of multidimensionality has been tested using DETECT.

2.2.1 Test of Essential Unidimensionality - DIMTEST

DIMTEST (Stout, 1987; Stout, Froelich, & Gao, 2001) is a statistical procedure that performs the hypothesis test for the existence of essential unidimensionality in test data assuming an infinite number of items and examinees. Assuming that d is the number of dimensions of the test, one can use DIMTEST to test the hypothesis $H_0 : d = 1$ versus $H_A : d > 1$. If DIMTEST test statistic (T), which has an approximate standard normal distribution, significantly different than 0, null hypothesis rejected. Test data is not essentially unidimensional.

DIMTEST accepts only dichotomous items and does not allow missingness in the data. It can be run with 120 items for up to 6000 examinees. As opposed to Mplus, DIMTEST allows user to enter a lower asymptote (guessing) parameter. But this guessing parameter is global, i.e. different items cannot have different guessing parameters.

Initially, a set of items called Assessment Test (AT) selected among the items. Items in AT test should be as homogeneous as possible and dimensionally distinct than the rest of the items, which are called Partitioning Test (PT) items. In previous versions of the DIMTEST, a second AT item subset (AT2) specified to correct for the bias that cause unacceptable levels of Type-I error rates. This happens especially when the AT subtest composed of items with large item discriminations (Nandakumar & Stout, 1993). In the current version, instead of selecting an AT2 item subset, a new bias correction method based on statistical re-sampling introduced by Stout et al. (2001) is used. Authors showed that this new bias correction method produced Type-I error rates at around 0.05 (which is close to nominal error rate $\alpha = 0.05$) except small samples, and has high power.

DIMTEST test statistic can be defined as:

$$T = \frac{T_L - \bar{T}_G}{\sqrt{1 + 1/N}} \tag{2.2.1}$$

In Equation (2.2.1), $\bar{T}_G$ is the mean of the DIMTEST test statistics calculated using the resampling of examinee responses N times (details of this resampling procedure can be found in Stout et al. (2001, p. 369)). T_L in the numerator is based on the covariances of items that has same PT number correct scores:

$$T_L = \frac{\sum\limits_{k=1}^{K} T_{L,k}}{\sqrt{\sum\limits_{k=1}^{K} S_k^2}} \tag{2.2.2}$$

where k is the subgroup of examinees with total number correct score k from PT items, S_k^2 is the asymptotic variance of $T_{L,k}$, and $T_{L,k}$ is defined as

$$T_{L,k} = \hat{\sigma}_k^2 - \hat{\sigma}_{U,k}^2 = 2 \sum_{i<l \in AT} \widehat{\text{cov}} \left(X_i, X_l \,|Z_{PT} = k \right) \tag{2.2.3}$$

where $\hat{\sigma}_k^2$ is the estimate of the variance of examinees in subgroup k. $\hat{\sigma}_k^2$ is defined as:

$$\hat{\sigma}_k^2 = \frac{1}{J_k} \sum_{i=1}^{J_k} \left(Y_j^{(k)} - \bar{Y}^{(k)} \right)^2$$

where J_k is the number of examinees in subgroup k, $Y_j^{(k)}$ is the total correct score of examinee j in subgroup k, and $\bar{Y}^{(k)}$ is the mean correct score of J_k examinees in subgroup k.

In Equation (2.2.3), $\hat{\sigma}_{U,k}^2$ is the estimate for the variance of $Y_j^{(k)}$. It is defined as:

$$\hat{\sigma}_{U,k}^2 = \sum_{i=1}^{m} \hat{p}_i^{(k)} \left(1 - \hat{p}_i^{(k)} \right)$$

where m is the number of AT items and $\hat{p}_i^{(k)}$ is the observed difficulty of item i in subgroup k:

$$\hat{p}_i^{(k)} = \frac{1}{J_k} \sum_{i=1}^{J_k} X_{ij}^{(k)}$$

In this equation, $X_{ij}^{(k)}$ denotes the response of jth examinee in kth subgroup to the ith AT item.

Dimensionality assessment using DIMTEST can be performed using a confirmatory or an exploratory approach. In confirmatory approach, substantive considerations can guide the selection of AT items, like expert opinions or test specifications. For example, if a mathematics test is measuring different sub-domains (such as algebra and geometry), AT test can be chosen from one of the sub-domains and the rest of the items in both sub-domains can form the PT items. AT items can be selected in an exploratory way. For example, using factor analysis, items highly loaded on a factor can be chosen as an AT test. Or, using cluster analysis, HCA/CCPROX within DimPack, one can obtain AT items. Current version of DIMTEST determines the AT items by selecting the items with highest factor loadings in an unrotated principal axes factor analysis of tetrachoric correlation matrix (Gierl et al., 2005). For exploratory analysis, Tate (2003) recommends to split data set into two subsets, select AT items from one subset and calculate T statistic from the other subset.

PT items are used to calculate total scores of examinees. Examinees categorized into different groups based on their total scores in these PT items. For each group, DIMTEST procedure calculates the conditional item covariances between items. These conditional covariances are used to calculate the T statistic. Under the null hypothesis of essential unidimensionality, the test statistic T has an approximate standard normal distribution.

Performance of DIMTEST procedure tested by various researchers. The analysis done by Hattie, Krakowski, Rogers, and Swaminathan (1996) concluded that DIMTEST is reasonably robust in finding the unidimensionality in the data and separating one and multiple dimensions. Tate (2003) found that DIMTEST correctly identified the presence and absence of essential unidimensionality for all cases except weak multidimensionality cases. The results of the study done by Seraphine (2000) concluded that, the increasing mismatch between the ability distribution and the item difficulty distribution reduces the power of DIMTEST. Also, Tate (2003) tabulated an extensive list of studies that used DIMTEST in their analyses.

2.2.2 Test of Multidimensionality - DETECT

DETECT is another nonparametric method to evaluate the extent of multidimensionality in the test data (Zhang & Stout, 1999; Roussos & Ozbek, 2006). The aim of DETECT is to find homogeneous clusters of test items that maximizes the DETECT index. These clusters should be as distinct as possible to maximize the DETECT index.

DETECT program can run 120 dichotomous items and 6000 subjects. It does not support missing data. User specifies the maximum number of dimensions to limit the search of the program. There is no option to specify guessing parameters. Program can be run in either exploratory mode or confirmatory mode. In exploratory mode, program finds the number of dimensions and cluster formation that maximizes the DETECT index. It is recommended to divide data into two parts in exploratory analysis for cross-validation. Program finds the best partition that maximizes the DETECT index using the first cross-validation sample. Using this partition, DETECT index is calculated using second cross-validation sample.

DETECT program first partitions test items into clusters, then calculate the DETECT index for that partition. Clearly there are many possible partitions in a given test. But calculating DETECT index for all possible cluster combinations is not computationally feasible. Douglas et al. (1999) offered two solutions for this: either using HCA (Stout et al., 1996; Roussos, Stout, & Marden, 1998) or genetic algorithm. DETECT analysis implemented in DimPack package uses genetic algorithm to find mutually exclusive and homogeneous item clusters. After finding clusters, DETECT index is calculated using following formula:

$$DETECT = \frac{1}{n\,(n-1)\,/2} \sum_{1 \leq i < j \leq n}^{n} \delta_{ij} \cdot \left(\widehat{\text{cov}}\,(X_i, X_j) - \overline{\text{cov}}\right) \tag{2.2.4}$$

where n is the number of items, δ_{ij} is

$$\delta_{ij} = \begin{cases} 1 & \text{if items } i \text{ and } j \text{ are in the same cluster} \\ -1 & \text{if items } i \text{ and } j \text{ are \underline{not} in the same cluster} \end{cases}$$

In Equation (2.2.4), $\widehat{\text{cov}}\,(X_i, X_j)$ is the covariance of items i and j:

$$\widehat{\text{cov}}\,(X_i, X_j) = \frac{1}{J}\sum_{k=0}^{n-2} J_k \widehat{\text{cov}}\,(X_i, X_j\,|Z_{ij} = k)$$

where Z_{ij} is the observed score on the remaining $(n-2)$ items, J is the total number of examinees, J_k is the number of examinees with score $Z_{ij} = k$. Also in Equation (2.2.4), $\overline{\text{cov}}$ is the average of $\widehat{\text{cov}}\,(X_i, X_j)$ over all examinee subscore groups and item pairs (for details of this formulation check Douglas et al. (1999, p. 6)).

DETECT provides two additional indicators on top of DETECT index. Both of these indices can be used to evaluate the approximate simple structure of the test data. The concept of simple structure is based on the definition of Thurstone (1947). For example, for a three dimensional test, if items lie along only one of the coordinate axes that represents dimensions, then test is said to display simple structure. In a strict sense, factor loading of an item is positive for one dimension but zero for the rest of the dimensions. In practice this definition might not be attained. Instead a test is said to display *approximate* simple structure, if items have one large factor loading and very low factor loadings in the rest of the dimensions.

First indicator DETECT provides to measure simple structure is called IDN index. After the cluster formation, signs of the covariances within the same cluster is expected to be positive and signs from different clusters are expected to be negative. IDN index is the proportion of estimated covariance signs that are in accordance with the signs that are expected from the cluster formation. Values close to 1 indicates that the pattern of covariances matches approximate simple structure. Second indicator DETECT provides is r_{max}. This ratio is also an indicator of simple structure in the test data. Similar to IDN index, a value of 1 indicates an approximate simple structure. In practice, values larger than 0.8 is accepted to indicate an approximate simple structure in the test data .

Studies that use DETECT index generally cite the simulation studies done by Kim (1994)[1] for the interpretation of DETECT index. For example, Jang and Roussos (2007) gave the following guidelines for the interpretation of DETECT index:

[1] I cannot reach this source.

> An index value of 1 or more is considered to indicate large multidimensionality (Kim, 1994). Values of .4 to 1 are considered to indicate moderate to large multidimensionality. Values below .4 are considered to indicate moderate to weak multidimensionality, or even unidimensionality for values below .2. (p. 7)

In addition to these rule of thumbs, Stout et al. (1996) stated that cross-validated DETECT index values smaller than 0.1 indicates essential unidimensionality, and values larger than 1 indicates large multidimensionality. Following these authors recommendations and to be in accordance with the previous study done by Seol (2013), following rule of thumbs used for the interpretation of DETECT index:

D_{max}	**Decision**
0 - 0.1	Almost no multidimensionality
0.1 - 0.4	Weak multidimensionality
0.4 - 1	Moderate multidimensionality
1 and above	Large multidimensionality

Table 2.2.1: Interpretation of D_{max}

The analysis of multidimensionality using DETECT follows three steps. In the first step, the extent of essential multidimensionality investigated by the DETECT index. In the second step, the existence of an approximate simple structure investigated using the r_{max} value. If previous two steps point to a multidimensional simple structure, the number of dimensions and their formation can be evaluated in the third step.

In his simulation study, Tate (2003) found that DETECT successfully detect multidimensionality in 8 of the 12 cases. DETECT failed in (1) test with extreme difficulties, (2) test with two highly correlated abilities, (3) test with two dimensions where all items measure one dominant dimension and a fraction of items measuring a weak dimension, (4) test with diffuse structure (items have two directions that distributed continuously between two dimensions).

Chapter 3

Methodology

The methods applied in this section assumes that test validity and reliability is already established. The flow of the analysis for one section is shown in Figure 3.0.1. This flowchart is very similar to the decision trees proposed by Nandakumar and Ackerman (2004) and Jasper (2010). It combines both approaches and adapts it to the CPA exam.

For each section (AUD, FAR and REG), unidimensionality of the MCQ and TBS as one whole was checked. Since BEC section does not have TBS items, only dimensionality structure of MCQ items for this section investigated.

First, the strict unidimensionality was checked using Mplus. A one factor CFA was fitted to all MCQ and TBS items. In addition to this parametric unidimensionality test, essential unidimensionality was checked using DIMTEST.

If the unidimensionality of the test as a whole rejected, then unidimensionality of each item type was tested separately as described in the previous paragraph. If the unidimensionality still not tenable, then an exploratory dimensionality analysis pursued. Exploratory analysis performed using exploratory factor analysis in Mplus, testing the extent of multidimensionality in DETECT, and non-linear factor analysis using NOHARM.

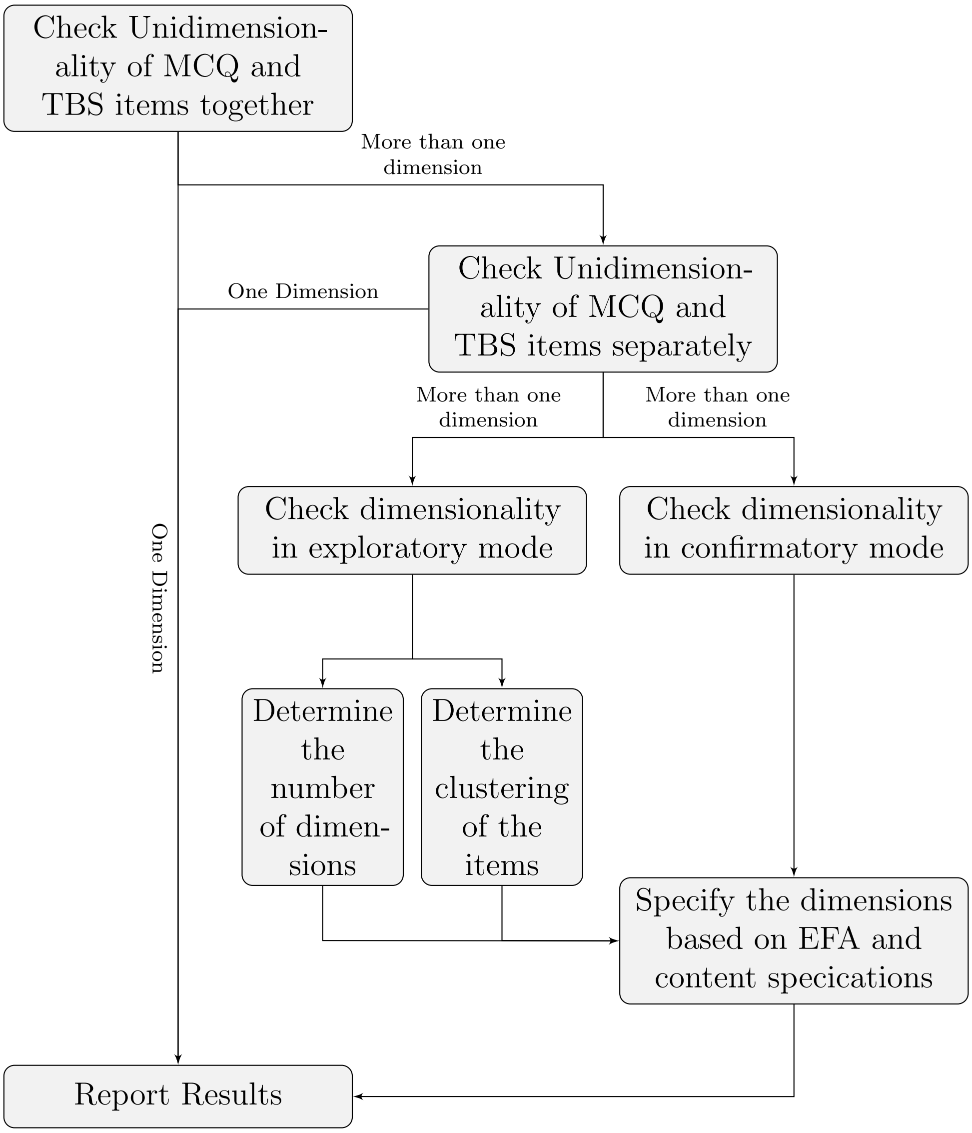

Figure 3.0.1: Structre of Dimensionality Analysis for each Section.

3.1 Mplus

Model fit of linear factor analysis was evaluated using five criteria. The first one is the chi-square model fit statistic (χ^2_M). If the χ^2_M does not show significant departure from hypothesized value, i.e. p-value is smaller than 0.05, this will be accepted as an indication of a good model fit. Second one is the χ^2_M/df ratio. Following the recommendation of Tate (2003), if this value is smaller than 1.4, this will be accepted as an indication of a good model fit. The third one is CFI. A CFI value larger than 0.90 will be accepted as an indication of an approximately good fitting model. Values larger than 0.95 will be accepted as an evidence of good model fit. The fourth one is TLI. The interpretation of TLI is similar to CFI. The fifth one is RMSEA. The values of RMSEA smaller than 0.05 will be accepted as good model fit. Also, it is expected that for a good fitting model, the upper bound of 90% confidence interval of the RMSEA should be smaller than 0.05. As mentioned in the previous chapter, each criteria checks a different aspect of model. In almost all of the cases investigated in this report these five criteria never agreed unanimously, except high dimensional models with many parameters. So, a holistic approach pursued in evaluating the model fit. If four of the five criteria agreed upon model fit, then a model fit is assumed. In EFA, the most parsimonious model satisfying majority of the criteria is accepted as a good fitting model.

3.2 DIMTEST

Essential unidimensionality of the test data within each section has been evaluated using DIMTEST. Essential unidimensionality checked first for combined MCQ and TBS items. Then, essential unidimensionality of MCQ and TBS items checked separately. In addition, unidimensionality of each TBS panel is tested to see the effect of sample size increase. In DIMTEST, the null hypothesis of essential unidimensionality is rejected when the p-value of the test statistic T is smaller than 0.05.

Unlike Mplus, in DIMTEST there is an option to enter a guessing parameter for items. But this guessing parameter cannot be declared for individual items, instead a single guessing parameter for all items should be entered. Even though this poses a limitation, ability to enter a guessing parameter is expected to enhance the power of the test as found by Tate (2003). A single guessing parameter has been calculated for each section. First, the distribution of guessing parameters obtained. The distributions of sections used instead of routes to avoid capitalization on chance. Figure 3.2.1 shows the histogram of the c parameters of all MCQ items in AUD section.

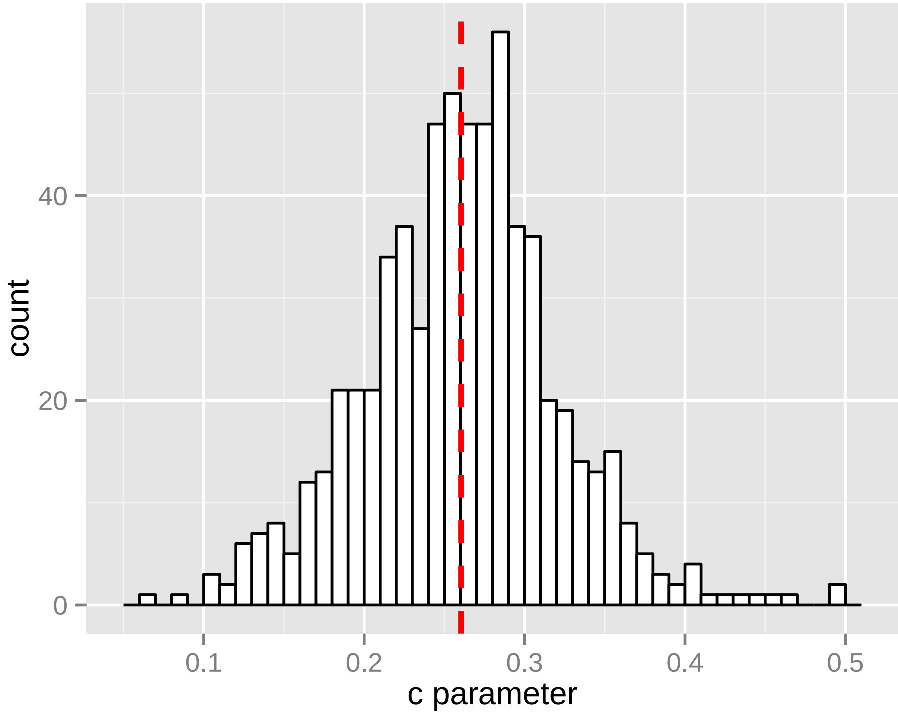

Figure 3.2.1: Histogram of the c parameters of all MCQ items in AUD section

The mean of the c distributions in Figure 3.2.1 is 0.2605. In the manual of DIMTEST, authors suggested users to enter a slightly underestimated value for the guessing parameter:

> It is generally better that this value be slightly underestimated rather than overestimated. (That is, it is better to get as much signal as possible, even if a little extra noise creeps in, than to risk losing some of the signal to avoid a little bit of extra noise.)

Following this recommendation, a slightly underestimated guessing parameter supplied to the program. So, instead of mean values, a slight underestimate of c parameters supplied to the program. For each section, this slightly underestimated value is 40th quantile of the c parameter distribution. For the c distribution in Figure 3.2.1 this corresponds to 0.2485 as can also be seen in Table 4.2.2.

3.3 DETECT

The extent of multidimensionality in the test data has been evaluated by DETECT program. In all DETECT analyses in this report, test data divided into two random parts with equal sample sizes. DETECT program found the cluster formation that maximized the DETECT index from the first part. Then, the DETECT index was calculated from the second part of the test data. The D_{max} columns in the tables in section 4.3 are showing the latter cross-validated DETECT index[1].

As shown in Table 2.2.1, D_{max} values smaller than 0.1 indicate approximate unidimensionality, values between 0.2 and 0.4 indicate weak multidimensionality, values between 0.4 to 1 indicate moderate multidimensionality and values larger than 1 indicate large multidimensionality. The magnitudes of r_{max} values indicates the extent of the approximate simple structure in the test data. If the values of r_{max} are larger than 0.8, this can be deemed as an indication of approximate simple structure.

3.4 NOHARM

Last method to evaluate the extent of multidimensionality in the test data is NOHARM. For this report, NOHARM run in an exploratory mode. Even in the exploratory mode of NOHARM, user needs to specify the number of dimensions to be extracted. One advantage of the NOHARM is the ability to provide guessing parameters for each item. So, instead of entering mean values or 40th quantile for all items, individual guessing parameters for each item provided to NOHARM in every analysis.

While running NOHARM, analysis of two routes out of 65 did not converge due to an error stemming from the guessing parameters that are too large with respect to the sample probabilities. These two routes were combined MCQ and TBS items in AUD section, panel 12, route 'MDD' and TBS items in the same route. Possibly a high guessing parameter in TBS items affected both tests. For these two routes NOHARM converged when all of the guessing parameters reduced to 0.

For model fit, NOHARM prints RMSR, GFI index and model parameters such as factor loadings and unique variances. In the literature there are four methods to evaluate the convergence of model fit for NOHARM solution.

First one is the criteria proposed by Fraser and McDonald (2012). If the value of RMSR is smaller than four times the reciprocal of the square root of

[1] In some texts, the notation of this cross-validated DETECT index is D_{ref}.

sample size, then the model fit is good. Tate (2003) dismissed this criteria by saying:

> In NOHARM, the only provision for the direct assessment of fit offered by the manual is a statement that if the RMSR is of the order of four divided by the square root of the sample size (equal to .09 for the simulation cases considered here), it would be likely that a test of significance would not reject the model. This rule does not work for the conditions of interest here. (p. 166)

The second method to evaluate the model fit in NOHARM is the approximate chi-square test of the fit of a fitted NOHARM model proposed by Gessaroli and De Champlain (1996) and implemented in a computer software called CHIDIM by De Champlain and Tang (1997). The approximate chi-square statistic tests the null hypothesis stating the off-diagonal elements of residual matrix is equal to zero.

The third method is applied by Gierl et al. (2005). These authors checked the dimensionality of SAT math and critical reading tests by observing the improvement of GFI index and RMSR simultaneously and deciding the most parsimonious model. For GFI index there are no guidelines in literature but values close to one are better.

The fourth way to assess model fit in NOHARM was proposed by Tate (2003) and used to assess dimensionality in his paper. Tate defined the test dimensionality as "the highest dimensional model that still produced an approximately 10% or greater decrease in the RMSR over the preceding model, a rule based on previous unpublished analyses by the author." (p. 165). In this research paper this method has been used to evaluate the dimensionality of the test data.

Chapter 4

Results

4.1 Linear Factor Analysis

In this section Mplus software used for both exploratory and confirmatory linear factor analysis. First, unidimensionality of a route has been checked. MCQ and TBS were combined and a one dimensional model fitted to all items. If this one dimensional model did not fit the test data, the unidimensional model fitted to each MCQ and TBS sections separately. If the unidimensionality was not tenable for any sub-section, an exploratory factor analysis pursued to investigate the dimensional structure of the data.

4.1.1 Unidimensionality of TBS and MCQ Items Together

A one factor model fitted to the combined MCQ and TBS items. From each section, except BEC section, five routes selected. These five routes had the largest sample sizes among other routes. CFA analysis results for these selected routes are in Table 4.1.1. The results of two analyses that did not converge has been removed.

Sec.	Pnl	Route	N	Chi Sq.	*df*	χ^2_M/df	*p*-value	CFI	TLI	RMSEA	RMSEA CI
AUD	18	MDD	545	7087.63	5994	1.18	0.0000	0.63	0.62	0.018	0.016 - 0.020
AUD	12	MDD	541	6680.05	5670	1.18	0.0000	0.57	0.56	0.018	0.016 - 0.020
AUD	11	MDD	528	-	-	-	-	-	-	-	-
AUD	19	MDD	522	6665.01	5885	1.13	0.0000	0.68	0.67	0.016	0.014 - 0.018
AUD	21	MDD	519	-	-	-	-	-	-	-	-
FAR	22	MMM	387	7826.12	5885	1.33	0.0000	0.70	0.69	0.029	0.027 - 0.031
FAR	09	MMM	383	8107.96	6440	1.26	0.0000	0.78	0.77	0.026	0.024 - 0.028
FAR	12	MMM	348	7333.24	6104	1.20	0.0000	0.69	0.69	0.024	0.022 - 0.026
FAR	04	MMM	337	7291.81	6104	1.19	0.0000	0.73	0.72	0.024	0.022 - 0.026
FAR	21	MMM	335	7477.41	5994	1.25	0.0000	0.79	0.79	0.027	0.025 - 0.029
REG	07	MMM	561	4612.05	3827	1.21	0.0000	0.87	0.86	0.019	0.017 - 0.021
REG	21	MMM	530	5265.77	3827	1.38	0.0000	0.73	0.72	0.027	0.025 - 0.028
REG	13	MMM	498	5501.50	3569	1.54	0.0000	0.80	0.80	0.033	0.031 - 0.035
REG	20	MMM	481	6859.31	3740	1.83	0.0000	0.68	0.67	0.042	0.040 - 0.043
REG	06	MMM	489	6118.81	4094	1.49	0.0000	0.73	0.73	0.032	0.030 - 0.033

Table 4.1.1: Unidimensionality Analysis Summary for Combined MCQ and TBS Items

Chi-square model fit statistic is significant for all routes. The χ^2_M/df ratio is smaller than 1.4 for AUD and FAR sections. For REG section, it is smaller than 1.4 for two out of five routes. CFI and TLI values are lower than recommended size for good model fit, which is 0.95. All RMSEA values are lower than 0.05. The upper bounds of 90% confidence intervals are also lower than 0.05. RMSEA and χ^2_M/df ratio indicates good fit, other three indicators does not support this claim. Unidimensionality of MCQ and TBS items investigated separately in the next section.

4.1.2 Unidimensionality of MCQ Items

One factor CFA model fitted to MCQ items. Results of these analyses are in Table 4.1.2.

Sec.	Pnl	Route	N	Chi Sq.	*df*	χ^2_M/df	p-value	CFI	TLI	RMSEA	RMSEA CI
AUD	18	MDD	545	-	-	-	-	-	-	-	-
AUD	12	MDD	541	3130.30	2700	1.16	0.0000	0.64	0.63	0.017	0.014 - 0.020
AUD	11	MDD	529	2957.83	2700	1.10	0.0003	0.65	0.64	0.013	0.010 - 0.017
AUD	19	MDD	524	2859.95	2700	1.06	0.0160	0.74	0.73	0.011	0.005 - 0.014
AUD	21	MDD	519	-	-	-	-	-	-	-	-
BEC	02	MMM	484	2147.45	1710	1.26	0.0000	0.83	0.83	0.023	0.020 - 0.026
BEC	15	MMM	468	2091.81	1710	1.22	0.0000	0.86	0.85	0.022	0.018 - 0.025
BEC	22	MMM	465	2162.33	1710	1.26	0.0000	0.77	0.76	0.024	0.021 - 0.027
BEC	04	MMM	457	2207.04	1710	1.29	0.0000	0.85	0.85	0.025	0.022 - 0.028
BEC	18	MMM	441	1981.71	1710	1.16	0.0000	0.90	0.89	0.019	0.015 - 0.023
FAR	22	MMM	397	2893.97	2627	1.10	0.0002	0.90	0.89	0.016	0.012 - 0.020
FAR	09	MMM	395	2931.60	2627	1.12	0.0000	0.85	0.85	0.017	0.013 - 0.021
FAR	12	MMM	363	2914.28	2627	1.11	0.0001	0.85	0.85	0.017	0.013 - 0.021
FAR	04	MMM	355	3021.64	2700	1.12	0.0000	0.81	0.81	0.018	0.014 - 0.022
FAR	21	MMM	353	2876.56	2627	1.09	0.0004	0.86	0.86	0.016	0.012 - 0.020
REG	07	MMM	568	2140.84	1710	1.25	0.0000	0.91	0.91	0.021	0.018 - 0.024
REG	21	MMM	539	2166.55	1710	1.27	0.0000	0.88	0.87	0.022	0.019 - 0.025
REG	13	MMM	503	2225.20	1710	1.30	0.0000	0.87	0.86	0.024	0.021 - 0.027
REG	20	MMM	491	2336.60	1710	1.37	0.0000	0.85	0.84	0.027	0.025 - 0.030
REG	06	MMM	490	2302.91	1710	1.35	0.0000	0.84	0.84	0.027	0.024 - 0.029

Table 4.1.2: Unidimensionality Analysis Summary for MCQ Items

Compared to the results in the previous section, model fit improved for MCQ items. Even though the chi-square test statistic rejected all models, the ratio of chi-square fit statistic to degrees of freedom reduced for all fitted models. This can be observed from the p-values for chi-square model fit statistics. In previous section, all p-values were smaller than 0.00001, when MCQ items fitted alone, some of the p-values increased. This is a sign of better fit.

Same thing is true for CFI and TLI indices. Values of these indices increased and get closer to 1 compared to the results in previous section. But still, some of these values are smaller than 0.90 and all of them are smaller than 0.95, the recommended value for good fit. All RMSEA values are indicating a good model fit. The upper bounds of the 90% confidence interval of RMSEA values are all below 0.05.

Overall, the results show that MCQ items are closer to unidimensionality. Even though most of the indicators shows good fit for all models, some indicators are indicating little deviations from unidimensionality. So, the results warrant for an exploratory analysis of the results.

4.1.3 Unidimensionality of TBS Items

The results of the one-factor CFA solutions for TBS items are in Table 4.1.3.

Sec.	Pnl	Route	N	Chi Sq.	*df*	χ^2_M/df	*p*-value	CFI	TLI	RMSEA	RMSEA CI
AUD	18	MDD	545	2146.19	594	3.61	0.0000	0.65	0.63	0.069	0.066 - 0.072
AUD	12	MDD	541	1132.30	495	2.29	0.0000	0.73	0.72	0.049	0.045 - 0.053
AUD	11	MDD	528	2610.13	629	4.15	0.0000	0.86	0.85	0.077	0.074 - 0.080
AUD	19	MDD	522	-	-	-	-	-	-	-	-
AUD	21	MDD	519	2275.42	560	4.06	0.0000	0.85	0.84	0.077	0.074 - 0.080
FAR	22	MMM	387	2412.74	594	4.06	0.0000	0.73	0.72	0.089	0.085 - 0.093
FAR	09	MMM	383	2720.09	779	3.49	0.0000	0.78	0.77	0.081	0.077 - 0.084
FAR	12	MMM	348	2096.54	665	3.15	0.0000	0.65	0.63	0.079	0.075 - 0.082
FAR	04	MMM	338	1660.01	629	2.64	0.0000	0.80	0.79	0.070	0.066 - 0.074
FAR	21	MMM	335	2374.26	629	3.77	0.0000	0.83	0.82	0.091	0.087 - 0.095
REG	07	MMM	561	945.11	377	2.51	0.0000	0.61	0.58	0.052	0.048 - 0.056
REG	21	MMM	530	1693.83	377	4.49	0.0000	0.57	0.53	0.081	0.077 - 0.085
REG	13	MMM	498	1269.22	299	4.24	0.0000	0.92	0.91	0.081	0.076 - 0.085
REG	20	MMM	481	2631.49	350	7.52	0.0000	0.81	0.80	0.116	0.112 - 0.121
REG	06	MMM	490	1526.10	464	3.29	0.0000	0.85	0.84	0.068	0.065 - 0.072

Table 4.1.3: Unidimensionality Analysis Summary for TBS Items

Chi-square model fit is rejected for all models. The ratios of chi-square fit statistics to degrees of freedoms are all larger than 1.4. CFI and TLI indices indicate poor fit as well. Different than previous two sections, all RMSEA values are larger than 0.05, except one. The upper bounds of the 90% confidence intervals of RMSEA values are all above 0.05. Clearly, unidimensionality does not hold for any of the TBS routes. Results warrant for further EFA for TBS items.

The results in Table 4.1.3 are for the routes. In operational test, all examinees within a panel took same TBS items. The number of examinees that took a set of TBS items is larger. To utilize from larger sample sizes, same one-factor CFA model fitted for TBS items within a panel. Results are in Table 4.1.4. One-factor model still did not fit to any panels investigated. The ratio of chi-square fit statistic to degrees of freedom got even worse. CFI and TLI values slightly increased but nonetheless still shows poor fit. RMSEA values generally increased (except for example REG section, panel 7), showing even worse fit. So, increasing the sample size did not make much difference on the results.

Sec.	Pnl	N	Chi Sq.	*df*	χ^2_M/df	*p*-value	CFI	TLI	RMSEA	RMSEA CI
AUD	19	745	-	-	-	-	-	-	-	-
AUD	01	731	5294.55	527	10.05	0.0000	0.80	0.79	0.111	0.109 - 0.114
AUD	06	719	2784.04	495	5.62	0.0000	0.69	0.67	0.080	0.077 - 0.083
AUD	07	715	2231.50	527	4.23	0.0000	0.63	0.60	0.067	0.064 - 0.070
AUD	02	703	1890.60	405	4.67	0.0000	0.76	0.75	0.072	0.069 - 0.076
FAR	09	807	6230.55	779	8.00	0.0000	0.79	0.78	0.093	0.091 - 0.095
FAR	22	773	5471.60	594	9.21	0.0000	0.78	0.77	0.103	0.101 - 0.106
FAR	10	768	5393.31	702	7.68	0.0000	0.80	0.79	0.093	0.091 - 0.096
FAR	17	760	3389.26	594	5.71	0.0000	0.75	0.73	0.079	0.076 - 0.081
FAR	18	753	5511.98	665	8.29	0.0000	0.81	0.80	0.098	0.096 - 0.101
REG	04	847	3938.23	464	8.49	0.0000	0.74	0.73	0.094	0.091 - 0.097
REG	07	810	1112.35	377	2.95	0.0000	0.72	0.69	0.049	0.046 - 0.052
REG	05	799	2268.51	434	5.23	0.0000	0.84	0.83	0.073	0.070 - 0.076
REG	09	791	4170.75	350	11.92	0.0000	0.85	0.83	0.117	0.114 - 0.121
REG	13	791	2367.55	299	7.92	0.0000	0.92	0.91	0.094	0.090 - 0.097

Table 4.1.4: Unidimensionality Analysis Summary for TBS Items for Panel

4.1.4 Exploratory Factor Analysis of MCQ Items

The results of the selected EFA of MCQ items are in Tables 4.1.5 to 4.1.12. Due to space limitations only two EFA results from each section printed. Rest of the results can be found in Appendix A. Even though the results of the CFA in section 4.1.2 showed that MCQ items are unidimensional, not all of the indicators agreed with this decision. In this section, the dimensional structure of MCQ items further investigated.

For the two routes in AUD section in Tables 4.1.5 and 4.1.6, chi-square model fit statistic become insignificant at the third dimension for the first AUD route (Table 4.1.5), and at the fourth dimension for the second AUD route (Table 4.1.6). All of the χ^2_M/df ratios are smaller than 1.4 for both routes. RMSEA values are indicating good fit for both routes and for any number of dimensions. The upper bounds of the 90% confidence intervals of all RMSEA values are all below 0.05. CFI and TLI indices started to show an approximate good fit at the fourth dimension for the first route and at the sixth dimension for the second route. It can be concluded that three dimensions are needed to explain the test structure for the first AUD route, and four dimensions are needed to explain the test structure for the second AUD route.

Factor	N	Par.	Chi Sq.	*df*	χ^2_M/df	*p*-value	CFI	TLI	RMSEA	RMSEA CI
1	545	75	2870.68	2700	1.06	0.0112	0.77	0.77	0.011	0.006 - 0.014
2	545	149	2756.88	2626	1.05	0.0371	0.83	0.82	0.010	0.003 - 0.013
3	545	222	2652.03	2553	1.04	0.0842	0.87	0.86	0.008	0.000 - 0.013
4	545	294	2546.92	2481	1.03	0.1744	0.91	0.90	0.007	0.000 - 0.012
5	545	365	2455.60	2410	1.02	0.2539	0.94	0.93	0.006	0.000 - 0.011
6	545	435	2370.75	2340	1.01	0.3237	0.96	0.95	0.005	0.000 - 0.011
7	545	504	2287.82	2271	1.01	0.3979	0.98	0.97	0.004	0.000 - 0.010
8	545	572	2205.70	2203	1.00	0.4798	1.00	0.99	0.001	0.000 - 0.010

Table 4.1.5: Exploratory Factor Analysis Summary for MCQ Items of AUD Section, Panel 18, Route MDD

Factor	N	Par.	Chi Sq.	*df*	χ^2_M/df	*p*-value	CFI	TLI	RMSEA	RMSEA CI
1	529	75	2957.83	2700	1.10	0.0003	0.65	0.64	0.013	0.010 - 0.017
2	529	149	2805.91	2626	1.07	0.0074	0.76	0.74	0.011	0.006 - 0.015
3	529	222	2693.44	2553	1.06	0.0262	0.81	0.79	0.010	0.004 - 0.014
4	529	294	2586.43	2481	1.04	0.0687	0.86	0.84	0.009	0.000 - 0.013
5	529	365	2493.47	2410	1.03	0.1154	0.89	0.87	0.008	0.000 - 0.013
6	529	435	2409.16	2340	1.03	0.1560	0.91	0.89	0.007	0.000 - 0.012
7	529	504	2324.70	2271	1.02	0.2117	0.93	0.91	0.007	0.000 - 0.012
8	529	572	2237.87	2203	1.02	0.2971	0.95	0.94	0.005	0.000 - 0.011

Table 4.1.6: Exploratory Factor Analysis Summary for MCQ Items of AUD Section, Panel 11, Route MDD

As for the two routes in BEC section in Tables 4.1.7 and 4.1.8, chi-square statistic become insignificant after the fourth dimension for the first BEC route (Table 4.1.7), and after the third dimension for the second BEC route (Table 4.1.8). For both routes, χ^2_M/df ratios are smaller than 1.4. RMSEA values are indicating good fit for both routes and for any number of dimensions. The upper bounds of the 90% confidence intervals of all RMSEA values are below 0.05. CFI and TLI indices show approximate good fit after the second dimension for both routes. It can be concluded that two dimensions are needed to explain the test structure for both BEC routes.

Factor	N	Par.	Chi Sq.	*df*	χ^2_M/df	*p*-value	CFI	TLI	RMSEA	RMSEA CI
1	484	60	2147.45	1710	1.26	0.0000	0.83	0.83	0.023	0.020 - 0.026
2	484	119	1852.05	1651	1.12	0.0004	0.92	0.92	0.016	0.011 - 0.020
3	484	177	1731.82	1593	1.09	0.0081	0.95	0.94	0.013	0.007 - 0.018
4	484	234	1644.78	1536	1.07	0.0268	0.96	0.95	0.012	0.005 - 0.017
5	484	290	1554.44	1480	1.05	0.0872	0.97	0.97	0.010	0.000 - 0.016
6	484	345	1475.66	1425	1.04	0.1710	0.98	0.98	0.009	0.000 - 0.015
7	484	399	1400.33	1371	1.02	0.2847	0.99	0.99	0.007	0.000 - 0.014
8	484	452	1328.98	1318	1.01	0.4105	1.00	0.99	0.004	0.000 - 0.013

Table 4.1.7: Exploratory Factor Analysis Summary for MCQ Items of BEC Section, Panel 02, Route MMM

Factor	N	Par.	Chi Sq.	*df*	χ^2_M/df	*p*-value	CFI	TLI	RMSEA	RMSEA CI
1	468	60	2091.81	1710	1.22	0.0000	0.86	0.85	0.022	0.018 - 0.025
2	468	119	1851.48	1651	1.12	0.0004	0.92	0.92	0.016	0.011 - 0.020
3	468	177	1712.77	1593	1.08	0.0186	0.95	0.95	0.013	0.006 - 0.017
4	468	234	1616.18	1536	1.05	0.0758	0.97	0.97	0.011	0.000 - 0.016
5	468	290	1523.39	1480	1.03	0.2112	0.98	0.98	0.008	0.000 - 0.014
6	468	345	1441.74	1425	1.01	0.3727	0.99	0.99	0.005	0.000 - 0.013
7	468	399	1359.41	1371	0.99	0.5829	1.00	1.01	0.000	0.000 - 0.011
8	468	452	1284.27	1318	0.97	0.7421	1.00	1.02	0.000	0.000 - 0.009

Table 4.1.8: Exploratory Factor Analysis Summary for MCQ Items of BEC Section, Panel 15, Route MMM

For the two routes in FAR section, chi-square statistic for model fit becomes insignificant after the third dimension for the first FAR route (Table 4.1.9), and after the second dimension for the second FAR route (Table 4.1.10). All of the χ^2_M/df ratios are smaller than 1.4 for both routes. RMSEA values are indicating good fit for both routes and for any number of dimension. The upper bounds of the 90% confidence intervals of all RMSEA values are below 0.05. CFI and TLI indices show approximately good fit for all dimensions for the first route and after the first dimension for the second route. It can be concluded that four dimensions are needed to explain the test structure for the first FAR route , and three dimensions are needed to explain the test structure for the second FAR route.

Factor	N	Par.	Chi Sq.	*df*	χ^2_M/df	*p*-value	CFI	TLI	RMSEA	RMSEA CI
1	397	74	2893.97	2627	1.10	0.0002	0.90	0.89	0.016	0.012 - 0.020
2	397	147	2734.66	2554	1.07	0.0066	0.93	0.92	0.013	0.008 - 0.017
3	397	219	2622.03	2482	1.06	0.0250	0.94	0.94	0.012	0.005 - 0.016
4	397	290	2519.58	2411	1.05	0.0605	0.96	0.95	0.011	0.000 - 0.016
5	397	360	2424.39	2341	1.04	0.1123	0.97	0.96	0.009	0.000 - 0.015
6	397	429	2337.41	2272	1.03	0.1658	0.97	0.97	0.009	0.000 - 0.014
7	397	497	2251.42	2204	1.02	0.2360	0.98	0.98	0.007	0.000 - 0.014
8	397	564	2168.12	2137	1.01	0.3142	0.99	0.98	0.006	0.000 - 0.013

Table 4.1.9: Exploratory Factor Analysis Summary for MCQ Items of FAR Section, Panel 22, Route MMM

Factor	N	Par.	Chi Sq.	*df*	χ^2_M/df	*p*-value	CFI	TLI	RMSEA	RMSEA CI
1	395	74	2931.60	2627	1.12	0.0000	0.85	0.85	0.017	0.013 - 0.021
2	395	147	2689.20	2554	1.05	0.0308	0.94	0.93	0.012	0.004 - 0.016
3	395	219	2587.20	2482	1.04	0.0692	0.95	0.94	0.010	0.000 - 0.015
4	395	290	2489.21	2411	1.03	0.1305	0.96	0.96	0.009	0.000 - 0.015
5	395	360	2400.16	2341	1.03	0.1929	0.97	0.97	0.008	0.000 - 0.014
6	395	429	2318.80	2272	1.02	0.2421	0.98	0.97	0.007	0.000 - 0.014
7	395	497	2238.15	2204	1.02	0.3009	0.98	0.98	0.006	0.000 - 0.013
8	395	564	2157.02	2137	1.01	0.3762	0.99	0.99	0.005	0.000 - 0.013

Table 4.1.10: Exploratory Factor Analysis Summary for MCQ Items of FAR Section, Panel 09, Route MMM

As for the two routes in REG section, chi-square statistic for model fit is significant for all eight dimensions extracted for the first REG route (Table 4.1.11). For the second REG route (Table 4.1.12), chi-square fit statistic becomes insignificant after the third dimension. All of the χ^2_M/df ratios are smaller than 1.4 for both routes. RMSEA values are indicating good fit for both routes and for any number of dimension. The upper bounds of the 90% confidence intervals of all RMSEA values are below 0.05. CFI and TLI indices show approximately good fit for all dimensions for the first route, and after the first dimension for the second route. It can be concluded that three dimensions are needed to explain the test structure for both REG routes.

Factor	N	Par.	Chi Sq.	*df*	χ^2_M/df	p-value	CFI	TLI	RMSEA	RMSEA CI
1	568	60	2140.84	1710	1.25	0.0000	0.91	0.91	0.021	0.018 - 0.024
2	568	119	1977.10	1651	1.20	0.0000	0.93	0.93	0.019	0.015 - 0.022
3	568	177	1841.69	1593	1.16	0.0000	0.95	0.94	0.017	0.013 - 0.020
4	568	234	1743.00	1536	1.13	0.0002	0.96	0.95	0.015	0.011 - 0.019
5	568	290	1655.81	1480	1.12	0.0009	0.96	0.96	0.014	0.010 - 0.018
6	568	345	1566.65	1425	1.10	0.0049	0.97	0.96	0.013	0.008 - 0.017
7	568	399	1484.50	1371	1.08	0.0169	0.98	0.97	0.012	0.006 - 0.016
8	568	452	1407.63	1318	1.07	0.0427	0.98	0.97	0.011	0.002 - 0.016

Table 4.1.11: Exploratory Factor Analysis Summary for MCQ Items of REG Section, Panel 07, Route MMM

Factor	N	Par.	Chi Sq.	*df*	χ^2_M/df	p-value	CFI	TLI	RMSEA	RMSEA CI
1	539	60	2166.55	1710	1.27	0.0000	0.88	0.87	0.022	0.019 - 0.025
2	539	119	1867.73	1651	1.13	0.0001	0.94	0.94	0.016	0.011 - 0.019
3	539	177	1741.67	1593	1.09	0.0051	0.96	0.95	0.013	0.008 - 0.017
4	539	234	1621.78	1536	1.06	0.0628	0.98	0.97	0.010	0.000 - 0.015
5	539	290	1529.00	1480	1.03	0.1832	0.99	0.98	0.008	0.000 - 0.014
6	539	345	1454.27	1425	1.02	0.2888	0.99	0.99	0.006	0.000 - 0.013
7	539	399	1383.93	1371	1.01	0.3979	1.00	0.99	0.004	0.000 - 0.012
8	539	452	1313.42	1318	1.00	0.5304	1.00	1.00	0.000	0.000 - 0.011

Table 4.1.12: Exploratory Factor Analysis Summary for MCQ Items of REG Section, Panel 21, Route MMM

4.1.5 Exploratory Factor Analysis of TBS Items

In section 4.1.3 on page 37 it has been shown that the model fit is comparatively poor for TBS items. In this section the structure of TBS items are investigated in an exploratory mode.

Tables 4.1.13 and 4.1.14 shows the EFA results for AUD section. Chi-square statistic become insignificant after the sixth dimension for the first AUD route (Table 4.1.9), and after the seventh dimension for the second AUD route (Table 4.1.10). The χ^2_M/df ratios are smaller than 1.4 after the fourth dimension for the first route and after the third route for the second dimension. RMSEA values are indicating good fit after the second dimension for the first route, and for all dimensions for the second route. CFI and TLI indices show approximate good fit after the second dimension for the first route and after the third dimension for the second route. It can be concluded that four dimensions are needed to explain the test structure of both routes.

Factor	N	Par.	Chi Sq.	*df*	χ^2_M/df	*p*-value	CFI	TLI	RMSEA	RMSEA CI
1	545	36	2146.19	594	3.61	0.0000	0.65	0.63	0.069	0.066 - 0.072
2	545	71	1336.21	559	2.39	0.0000	0.83	0.80	0.051	0.047 - 0.054
3	545	105	892.28	525	1.70	0.0000	0.92	0.90	0.036	0.032 - 0.040
4	545	138	701.38	492	1.43	0.0000	0.95	0.94	0.028	0.023 - 0.033
5	545	170	587.72	460	1.28	0.0000	0.97	0.96	0.023	0.017 - 0.028
6	545	201	511.09	429	1.19	0.0039	0.98	0.97	0.019	0.011 - 0.025
7	545	231	444.98	399	1.12	0.0557	0.99	0.98	0.015	0.000 - 0.022
8	545	260	385.26	370	1.04	0.2816	1.00	0.99	0.009	0.000 - 0.018

Table 4.1.13: Exploratory Factor Analysis Summary for TBS Items of AUD Section, Panel 18, Route MDD

Factor	N	Par.	Chi Sq.	*df*	χ^2_M/df	*p*-value	CFI	TLI	RMSEA	RMSEA CI
1	541	33	1132.30	495	2.29	0.0000	0.73	0.72	0.049	0.045 - 0.053
2	541	65	818.45	463	1.77	0.0000	0.85	0.83	0.038	0.033 - 0.042
3	541	96	666.76	432	1.54	0.0000	0.90	0.88	0.032	0.027 - 0.036
4	541	126	553.72	402	1.38	0.0000	0.94	0.92	0.026	0.021 - 0.032
5	541	155	476.77	373	1.28	0.0002	0.96	0.94	0.023	0.016 - 0.029
6	541	183	421.04	345	1.22	0.0032	0.97	0.95	0.020	0.012 - 0.027
7	541	210	365.47	318	1.15	0.0342	0.98	0.97	0.017	0.005 - 0.024
8	541	236	314.54	292	1.08	0.1743	0.99	0.98	0.012	0.000 - 0.021

Table 4.1.14: Exploratory Factor Analysis Summary for TBS Items of AUD Section, Panel 12, Route MDD

For the two routes in FAR section in Tables 4.1.15 and 4.1.16, chi-square statistic for model fit never becomes insignificant for the first FAR route (Table 4.1.15), and after the seventh dimension for the second FAR route (Table 4.1.16). The χ^2_M/df ratios are smaller than 1.4 after the seventh dimension for the first route and after the fourth route for the second dimension. RMSEA values are indicating good fit after the fourth dimension for the first route, and after the third dimension for the second route. CFI and TLI indices show approximately good fit after the second dimension for the first route and after the third dimension for the second route. It can be concluded that four dimensions are needed to explain the test structure of both routes.

Factor	N	Par.	Chi Sq.	*df*	χ^2_M/df	*p*-value	CFI	TLI	RMSEA	RMSEA CI
1	383	41	2719.45	779	3.49	0.0000	0.78	0.77	0.081	0.077 - 0.084
2	383	81	1971.48	739	2.67	0.0000	0.86	0.84	0.066	0.062 - 0.070
3	383	120	1581.41	700	2.26	0.0000	0.90	0.88	0.057	0.054 - 0.061
4	383	158	1271.08	662	1.92	0.0000	0.93	0.91	0.049	0.045 - 0.053
5	383	195	1106.36	625	1.77	0.0000	0.94	0.93	0.045	0.040 - 0.049
6	383	231	972.03	589	1.65	0.0000	0.96	0.94	0.041	0.037 - 0.046
7	383	266	810.83	554	1.46	0.0000	0.97	0.96	0.035	0.030 - 0.040
8	383	300	692.49	520	1.33	0.0000	0.98	0.97	0.029	0.023 - 0.035

Table 4.1.15: Exploratory Factor Analysis Summary for TBS Items of FAR Section, Panel 09, Route MMM

Factor	N	Par.	Chi Sq.	*df*	χ^2_M/df	*p*-value	CFI	TLI	RMSEA	RMSEA CI
1	348	38	2096.07	665	3.15	0.0000	0.65	0.63	0.079	0.075 - 0.082
2	348	75	1498.61	628	2.39	0.0000	0.79	0.76	0.063	0.059 - 0.067
3	348	111	1114.86	592	1.88	0.0000	0.87	0.85	0.050	0.046 - 0.055
4	348	146	851.58	557	1.53	0.0000	0.93	0.91	0.039	0.034 - 0.044
5	348	180	728.01	523	1.39	0.0000	0.95	0.93	0.034	0.028 - 0.039
6	348	213	624.65	490	1.27	0.0000	0.97	0.95	0.028	0.021 - 0.034
7	348	245	550.60	458	1.20	0.0019	0.98	0.96	0.024	0.015 - 0.031
8	348	276	475.18	427	1.11	0.0534	0.99	0.98	0.018	0.000 - 0.027

Table 4.1.16: Exploratory Factor Analysis Summary for TBS Items of FAR Section, Panel 12, Route MMM

As for the two routes in REG section in Tables 4.1.17 and 4.1.18, chi-square statistic for model fit becomes insignificant after the fourth dimension for the first REG route (Table 4.1.17), and never becomes insignificant for the second REG route (Table 4.1.18). The χ^2_M/df ratios are smaller than 1.4 after the third dimension for the first route and after the seventh route for the second dimension. RMSEA values are indicating good fit after the first dimension for the first route, and after the fourth dimension for the second route. CFI and TLI indices show good fit after the third dimension for the first route and after the fifth dimension for the second route. It can be concluded that four dimensions are needed to explain the test structure for the first REG route , and seven dimensions are needed to explain the test structure for the second REG route.

Factor	N	Par.	Chi Sq.	*df*	χ^2_M/df	*p*-value	CFI	TLI	RMSEA	RMSEA CI
1	561	29	945.11	377	2.51	0.0000	0.61	0.58	0.052	0.048 - 0.056
2	561	57	658.67	349	1.89	0.0000	0.79	0.75	0.040	0.035 - 0.044
3	561	84	466.54	322	1.45	0.0000	0.90	0.88	0.028	0.022 - 0.034
4	561	110	368.98	296	1.25	0.0025	0.95	0.93	0.021	0.013 - 0.028
5	561	135	309.99	271	1.14	0.0517	0.97	0.96	0.016	0.000 - 0.024
6	561	159	283.41	247	1.15	0.0555	0.97	0.96	0.016	0.000 - 0.024
7	561	182	240.95	224	1.08	0.2082	0.99	0.98	0.012	0.000 - 0.022
8	561	204	215.85	202	1.07	0.2398	0.99	0.98	0.011	0.000 - 0.022

Table 4.1.17: Exploratory Factor Analysis Summary for TBS Items of REG Section, Panel 07, Route MMM

Factor	N	Par.	Chi Sq.	*df*	χ^2_M/df	*p*-value	CFI	TLI	RMSEA	RMSEA CI
1	530	29	1693.83	377	4.49	0.0000	0.57	0.53	0.081	0.077 - 0.085
2	530	57	1184.11	349	3.39	0.0000	0.72	0.68	0.067	0.063 - 0.071
3	530	84	862.61	322	2.68	0.0000	0.82	0.78	0.056	0.052 - 0.061
4	530	110	680.62	296	2.30	0.0000	0.87	0.83	0.050	0.045 - 0.054
5	530	135	531.83	271	1.96	0.0000	0.91	0.87	0.043	0.037 - 0.048
6	530	159	431.50	247	1.75	0.0000	0.94	0.90	0.038	0.032 - 0.043
7	530	182	330.85	224	1.48	0.0000	0.96	0.94	0.030	0.023 - 0.037
8	530	204	239.87	202	1.19	0.0351	0.99	0.97	0.019	0.005 - 0.027

Table 4.1.18: Exploratory Factor Analysis Summary for TBS Items of REG Section, Panel 21, Route MMM

4.2 Test of Essential Unidimensionality - DIMTEST

Essential unidimensionality of each section has been investigated using DIMTEST. As in the previous sections, within each four sections five routes that have the largest number of examinees selected for the analysis. Initially, the essential unidimensionality of combined TBS and MCQ items were checked. This analysis was performed for all sections except BEC section, because this section does not have TBS items. If essential unidimensionality was rejected for any route, the essential unidimensionality of MCQ and TBS items tested separately for that route. In addition, since within a panel each examinee saw the same set of TBS items, DIMTEST was run for TBS items of a panel to increase the sample size for analysis.

4.2.1 Essential Unidimensionality of MCQ and TBS Items

The results for the test of essential unidimensionality of combined MCQ and TBS are in Table 4.2.1. All of the tests were rejected with small p-values except route MMM of panel 20 in REG section. This indicates that data is not essentially unidimensional for almost all of the routes.

Section	Panel	Route	N	n_{item}	Guessing	T	p-value
AUD	18	MDD	539	111	0.23	3.3437	0.0004
AUD	12	MDD	533	108	0.23	4.0872	0.0000
AUD	11	MDD	521	112	0.23	5.7359	0.0000
AUD	19	MDD	516	110	0.23	5.5874	0.0000
AUD	21	MDD	506	110	0.23	4.4168	0.0000
FAR	22	MMM	369	110	0.20	5.5143	0.0000
FAR	09	MMM	361	115	0.20	5.2648	0.0000
FAR	12	MMM	331	112	0.20	5.9432	0.0000
FAR	04	MMM	320	112	0.20	4.6408	0.0000
FAR	21	MMM	329	111	0.20	6.3331	0.0000
REG	07	MMM	552	89	0.19	2.6836	0.0036
REG	21	MMM	525	89	0.19	4.5723	0.0000
REG	13	MMM	491	86	0.19	5.3290	0.0000
REG	20	MMM	474	88	0.19	1.3914	0.0821
REG	06	MMM	478	92	0.19	5.6714	0.0000

Table 4.2.1: Test for Essential Unidimensionality of MCQ and TBS Items

4.2.2 Essential Unidimensionality of MCQ Items

Results in the previous section showed that combined MCQ and TBS items were not essentially unidimensional for any of the routes, except one. For this reason, the unidimensionality of MCQ items and TBS items investigated separately. Table 4.2.2 shows the results of essential unidimensionality test for MCQ items. The p-values of each test is smaller than 0.05, i.e. the null hypothesis that states tests are essentially unidimensional were rejected. None of the routes are essentially unidimensional. But compared to the results of combined tests in the previous section, T values became smaller.

Section	Panel	Route	N	n_{item}	Guessing	T	p-value
AUD	18	MDD	538	75	0.25	2.9944	0.0014
AUD	12	MDD	533	75	0.25	4.5850	0.0000
AUD	11	MDD	516	75	0.25	1.8273	0.0338
AUD	19	MDD	519	75	0.25	3.2392	0.0006
AUD	21	MDD	514	75	0.25	3.2705	0.0005
BEC	02	MMM	475	60	0.21	3.7573	0.0001
BEC	15	MMM	466	60	0.21	3.3906	0.0003
BEC	22	MMM	461	60	0.21	3.1624	0.0008
BEC	04	MMM	454	60	0.21	3.7594	0.0001
BEC	18	MMM	436	60	0.21	3.8075	0.0001
FAR	22	MMM	384	74	0.22	3.3528	0.0004
FAR	09	MMM	384	74	0.22	4.4846	0.0000
FAR	12	MMM	354	74	0.22	2.5717	0.0051
FAR	04	MMM	350	75	0.22	4.0235	0.0000
FAR	21	MMM	348	74	0.22	3.4492	0.0003
REG	07	MMM	562	60	0.21	1.7006	0.0445
REG	21	MMM	527	60	0.21	2.0361	0.0209
REG	13	MMM	500	60	0.21	2.5131	0.0060
REG	20	MMM	486	60	0.21	3.8944	0.0000
REG	06	MMM	486	60	0.21	3.1354	0.0009

Table 4.2.2: Test for Essential Unidimensionality of MCQ Items

4.2.3 Essential Unidimensionality of TBS Items

Results for the TBS section is very similar to the results in the previous sections. Table 4.2.3 shows the results for TBS items in various routes. Essential unidimensionality was also rejected for all routes with small p-values. None of the TBS sections investigated were essentially unidimensional.

Section	Panel	Route	N	n_{item}	Guessing	T	p-value
AUD	18	MDD	533	36	0.15	2.9480	0.0016
AUD	12	MDD	533	33	0.15	2.5150	0.0060
AUD	11	MDD	524	37	0.15	4.9234	0.0000
AUD	19	MDD	521	35	0.15	3.1074	0.0009
AUD	21	MDD	513	35	0.15	3.6912	0.0001
FAR	22	MMM	381	36	0.06	4.6586	0.0000
FAR	09	MMM	379	41	0.06	7.3273	0.0000
FAR	12	MMM	343	38	0.06	4.1116	0.0000
FAR	04	MMM	328	37	0.06	3.2092	0.0007
FAR	21	MMM	335	37	0.06	4.5050	0.0000
REG	07	MMM	556	29	0.10	3.4843	0.0002
REG	21	MMM	510	29	0.10	3.5568	0.0002
REG	13	MMM	485	26	0.10	7.0479	0.0000
REG	20	MMM	475	28	0.10	4.3032	0.0000
REG	06	MMM	481	32	0.10	3.7971	0.0001

Table 4.2.3: Test for Essential Unidimensionality of TBS Route Items

Since examinees within each panel saw the same TBS items, DIMTEST repeated using the examinees within each panel. This increased the sample size for the test. Results are shown in Table 4.2.4. None of the TBS sections were essentially unidimensional. T values increased compared to Table 4.2.3.

Section	Panel	Route	N	n_{item}	Guessing	T	p-value
AUD	19	-	743	35	0.15	4.9191	0.0000
AUD	01	-	720	34	0.15	3.1970	0.0007
AUD	06	-	712	33	0.15	3.3418	0.0004
AUD	07	-	712	34	0.15	4.7539	0.0000
AUD	02	-	698	30	0.15	2.6812	0.0037
FAR	09	-	796	41	0.06	9.2921	0.0000
FAR	22	-	767	36	0.06	6.7792	0.0000
FAR	10	-	743	39	0.06	7.7494	0.0000
FAR	17	-	752	36	0.06	8.2976	0.0000
FAR	18	-	748	38	0.06	5.5799	0.0000
REG	04	-	842	32	0.10	7.1460	0.0000
REG	07	-	804	29	0.10	3.8666	0.0001
REG	05	-	796	31	0.10	4.0396	0.0000
REG	09	-	785	28	0.10	7.0696	0.0000
REG	13	-	774	26	0.10	5.8239	0.0000

Table 4.2.4: Test for Essential Unidimensionality of TBS Panel Items

It is interesting to observe that even though the results in Table 4.2.1

showed that route MMM in panel 20 of REG section is essentially unidimensional, same thing cannot be said for the MCQ or TBS items of the same route. Essential unidimensionality of these two sub-parts rejected with very small p-values.

4.3 Test of Multidimensionality - DETECT

Essential unidimensionality did not hold for any of the routes (except one route). The next step is to investigate the extent of multidimensionality in the test data. In this section, results of the DETECT analysis will be shown. Similar to the previous sections, first the multidimensionality of combined MCQ and TBS items investigated. Following this, the multidimensionality of MCQ and TBS sections within selected routes investigated separately. Multidimensionality of TBS items within a panel investigated to replicate the analysis with a larger sample size.

The results of the exploratory factor analysis indicated that number of dimensions in the test data can go up to eight dimensions. In consideration of this, the maximum number of dimensions to be searched increased to eight from the default value of maximum five dimensions.

4.3.1 Multidimensionality of MCQ and TBS Items

Results of the DETECT analysis for the combined MCQ and TBS items are in Table 4.3.1. All of the results, except one, indicates a weak amount of multidimensionality. DETECT index values are between 0.1 and 0.26. DETECT index value for 'MMM' route of panel 7 in REG section is smaller than 0.1, which indicates an essential unidimensionality.

On the other hand, simple structure is clearly not tenable for any of the routes investigated. Values of r_{max} are far below 0.8. This indicates a complex structure within the test data. Items are possibly load on multiple dimensions. Number of clusters that maximize the DETECT index ranged from three to six.

Section	Panel	Route	N	N_{CV}	n_{item}	$n_{cluster}$	D_{max}	r_{max}
AUD	18	MDD	545	272	111	4	0.1169	0.1565
AUD	12	MDD	541	270	108	3	0.1347	0.1913
AUD	11	MDD	528	264	112	4	0.1617	0.2121
AUD	19	MDD	522	261	110	3	0.1517	0.1861
AUD	21	MDD	519	259	110	3	0.1374	0.1783
FAR	22	MMM	387	193	110	3	0.1827	0.1632
FAR	09	MMM	383	191	115	5	0.2279	0.2067
FAR	12	MMM	348	174	112	3	0.1271	0.1212
FAR	04	MMM	337	168	112	3	0.2611	0.2375
FAR	21	MMM	335	167	111	3	0.1380	0.1124
REG	07	MMM	561	280	89	5	0.0748	0.0766
REG	21	MMM	530	265	89	6	0.1428	0.1449
REG	13	MMM	498	249	86	3	0.1954	0.1951
REG	20	MMM	481	240	88	3	0.2310	0.2048
REG	06	MMM	489	244	92	4	0.1943	0.1899

Table 4.3.1: Test for Multidimensionality of MCQ and TBS Items (Cross-Validated)

4.3.2 Multidimensionality of MCQ Items

Results of the MCQ items are in Table 4.3.2. Values of DETECT index show more variation between sections for MCQ items. For AUD section, all of the values of DETECT index are smaller than 0.1. This indicates almost no multidimensionality for the routes in this section. All of the routes in BEC section have DETECT index values between 0.1 and 0.4. Routes in this section show weak multidimensionality. The routes in FAR section also show weak multidimensionality with one exception. 'MMM' route of panel 9 show almost no multidimensionality. And lastly, all routes in REG section show weak multidimensionality. DETECT index values in this section is between 0.15 and 0.25.

Similar to the combined test in the previous section, simple structure is not tenable for any of the routes. Values of r_{max} are far below 0.8. Number of clusters that maximize the DETECT index ranged from three to six.

Section	Panel	Route	N	N_{CV}	n_{item}	$n_{cluster}$	D_{max}	r_{max}
AUD	18	MDD	545	272	75	3	0.0930	0.1412
AUD	12	MDD	541	270	75	5	0.0939	0.1456
AUD	11	MDD	529	264	75	4	0.0700	0.1070
AUD	19	MDD	524	262	75	5	0.1036	0.1424
AUD	21	MDD	519	259	75	3	0.0725	0.1065
BEC	02	MMM	484	242	60	3	0.1859	0.1762
BEC	15	MMM	468	234	60	3	0.1375	0.1192
BEC	22	MMM	465	232	60	3	0.3484	0.3039
BEC	04	MMM	457	228	60	4	0.2391	0.2070
BEC	18	MMM	441	220	60	3	0.1168	0.1051
FAR	22	MMM	397	198	74	5	0.2312	0.2007
FAR	09	MMM	395	197	74	4	0.0702	0.0604
FAR	12	MMM	363	181	74	6	0.2003	0.1613
FAR	04	MMM	355	177	75	4	0.2092	0.1880
FAR	21	MMM	353	176	74	6	0.2624	0.2112
REG	07	MMM	568	284	60	3	0.1449	0.1506
REG	21	MMM	539	269	60	4	0.1619	0.1645
REG	13	MMM	503	251	60	5	0.2403	0.2232
REG	20	MMM	491	245	60	4	0.2489	0.2212
REG	06	MMM	490	245	60	4	0.1501	0.1443

Table 4.3.2: Test for Multidimensionality of MCQ Items (Cross-Validated)

4.3.3 Multidimensionality of TBS Items

Results of DETECT analysis for the TBS items within a route are in Table 4.3.3, and results for the items within a panel are in Table 4.3.4. On contrary to the results in the previous two sections, the DETECT indexes of the TBS items are larger. The range of DETECT index values is from 0.35 to 0.83. In AUD section, each route showed a weak amount of dimensionality, except route 'MDD' of panel 11. In FAR section, two routes showed moderate multidimensionality and three routes showed weak multidimensionality. In REG section, three routes showed moderate multidimensionality and two routes showed weak multidimensionality. Simple structure is not tenable for any of the routes. Nonetheless, the value of r_{max} increased compared to the previous two sections. The number of clusters that is proposed are either 4 or 5, except two routes.

The results of TBS panel items indicated even more multidimensionality. Overall DETECT index values increased. 9th panel in REG section showed a large amount of multidimensionality. But this panel is also has an approximate simple structure. The r_{max} value for this panel is 0.83.

Section	Panel	Route	N	N_{CV}	n_{item}	$n_{cluster}$	D_{max}	r_{max}
AUD	18	MDD	545	272	36	5	0.4450	0.3619
AUD	12	MDD	541	270	33	4	0.3529	0.3383
AUD	11	MDD	528	264	37	5	0.5829	0.4921
AUD	19	MDD	522	261	35	3	0.4464	0.3442
AUD	21	MDD	519	259	35	4	0.4424	0.4091
FAR	22	MMM	387	193	36	4	0.7893	0.5679
FAR	09	MMM	383	191	41	4	0.6670	0.5121
FAR	12	MMM	348	174	38	5	0.4096	0.3787
FAR	04	MMM	338	169	37	4	0.4558	0.3736
FAR	21	MMM	335	167	37	5	0.4965	0.3577
REG	07	MMM	561	280	29	3	0.4174	0.3264
REG	21	MMM	530	265	29	4	0.5531	0.4306
REG	13	MMM	498	249	26	4	0.5556	0.4463
REG	20	MMM	481	240	28	5	0.8349	0.5546
REG	06	MMM	490	245	32	5	0.4093	0.3633

Table 4.3.3: Test for Multidimensionality of TBS Route Items (Cross-Validated)

Section	Panel	Route	N	N_{CV}	n_{item}	$n_{cluster}$	D_{max}	r_{max}
AUD	19	-	745	372	35	4	0.4172	0.3579
AUD	01	-	731	365	34	3	0.5510	0.5592
AUD	06	-	719	359	33	4	0.5033	0.4641
AUD	07	-	715	357	34	5	0.4548	0.4073
AUD	02	-	703	351	30	3	0.5940	0.5281
FAR	09	-	807	403	41	4	0.7026	0.6170
FAR	22	-	773	386	36	3	0.7374	0.6038
FAR	10	-	768	384	39	5	0.7125	0.6444
FAR	17	-	760	380	36	4	0.5317	0.5370
FAR	18	-	753	376	38	5	0.6359	0.5770
REG	04	-	847	423	32	4	0.9656	0.6902
REG	07	-	810	405	29	5	0.4035	0.3839
REG	05	-	799	399	31	4	0.5851	0.5310
REG	09	-	791	395	28	4	1.3411	0.8260
REG	13	-	791	395	26	3	0.7013	0.5784

Table 4.3.4: Test for Multidimensionality of TBS Panel Items (Cross-Validated)

4.4 Test of Multidimensionality with Non-Linear Factor Analysis - NOHARM

The other exploratory analysis method utilized to investigate the dimensionality of the test data in this report is NOHARM. Due to the large number of analysis and limited space, only one route (or panel) of each item type from AUD section is investigated in this section.

As explained in Section 3.4 on page 32 even though there are four methods found in the literature to evaluate the model fit of a fitted NOHARM output, model fit has been checked using only the fourth method. In fact, the second method proposed by Fraser and McDonald (2012) is easy to apply for this data. But after checking the model fit using that criteria, it was observed that almost all of the unidimensional solutions are fitting very good. For example, for the first route in Table 4.4.1, sample size is 545. Criteria states that a good model fit is achieved if the RMSR value is smaller than $4/\sqrt{545} = 0.17$. All of the RMSR values are smaller than this value, which suggests the most parsimonious model is the unidimensional model. Considering the results from the previous sections, this doesn't seem correct. Consequently, this method did not used to evaluate the model fit.

Table 4.4.1 shows the NOHARM test results for the combined MCQ and TBS items for route 'MDD' of panel 18 of AUD section. From the first dimension to the second dimension RMSR value decreased by 10%. According to our criteria, this is a significant amount of reduction. But the reduction of RMSR from second dimension to the third is 9%. According to our criteria this is not a significant amount of reduction. Consequently, it can be said that two dimensional model fitted to this test data well. And this is the most parsimonious model that fits well.

n_{dim}	N	n_{item}	Tanaka's GoF	RMSR	RMSR % Change
1	545	111	0.8361	0.0080	-
2	545	111	0.8665	0.0072	10%
3	545	111	0.8883	0.0066	9%
4	545	111	0.8986	0.0063	5%
5	545	111	0.9062	0.0061	4%
6	545	111	0.9131	0.0058	4%
7	545	111	0.9195	0.0056	4%
8	545	111	0.9243	0.0054	3%

Table 4.4.1: NOHARM Analysis of MCQ and TBS Items - AUD Section, Panel 18, Route 'MDD'. Mean of guessing parameters for this test is 0.24.

The results for only MCQ items for panel 18 and route 'MDD' is in Table 4.4.2. After first dimension, the percentage reduction of RMSR is only 6%. This is not a significant amount of reduction. So, it can be concluded that the unidimensional model is fitting well to this data and it is the most parsimonious model tested. If the results in Appendix B.2 investigated, it can be seen that most of the tests composed of MCQ items are unidimensional.

n_{dim}	N	n_{item}	Tanaka's GoF	RMSR	RMSR % Change
1	545	75	0.9037	0.0061	-
2	545	75	0.9146	0.0057	6%
3	545	75	0.9252	0.0054	6%
4	545	75	0.9326	0.0051	5%
5	545	75	0.9394	0.0048	5%
6	545	75	0.9449	0.0046	5%
7	545	75	0.9501	0.0044	5%
8	545	75	0.9543	0.0042	4%

Table 4.4.2: NOHARM Analysis of MCQ Items - AUD Section, Panel 18, Route 'MDD'. Mean of guessing parameters for this test is 0.26.

The results for only TBS items for panel 18 and route 'MDD' is in Table 4.4.3. The percentage reduction in RMSR becomes smaller than 10% after the fourth dimension. It can be concluded that a four dimensional model fits the data well. The results for the other TBS routes are similar. All of the results tabulated in Appendix B.3 indicates at least three dimensions.

n_{dim}	N	n_{item}	Tanaka's GoF	RMSR	RMSR % Change
1	545	36	0.8909	0.0137	-
2	545	36	0.9341	0.0106	22%
3	545	36	0.9538	0.0089	16%
4	545	36	0.9659	0.0076	14%
5	545	36	0.9708	0.0071	7%
6	545	36	0.9748	0.0066	7%
7	545	36	0.9783	0.0061	7%
8	545	36	0.9809	0.0057	6%

Table 4.4.3: NOHARM Analysis of TBS Route Items - AUD Section, Panel 18, Route 'MDD'. Mean of guessing parameters for this test is 0.19.

The results for TBS items for all routes in panel 19 is in Table 4.4.4. The percentage reduction in RMSR becomes smaller than 10% after the sixth dimension. It can be concluded that a six dimensional model fits the data well. These results also confirm the results from previous sections. TBS items showed more multidimensionality compared to MCQ items. Rest of the results in Appendix B.4 confirms this conclusion.

n_{dim}	N	n_{item}	Tanaka's GoF	RMSR	RMSR % Change
1	745	35	0.9184	0.0127	-
2	745	35	0.9401	0.0109	14%
3	745	35	0.9549	0.0095	13%
4	745	35	0.9670	0.0081	14%
5	745	35	0.9734	0.0073	10%
6	745	35	0.9808	0.0062	15%
7	745	35	0.9837	0.0057	8%
8	745	35	0.9857	0.0053	6%

Table 4.4.4: NOHARM Analysis of TBS Panel Items - AUD Section, Panel 19. Mean of guessing parameters for this test is 0.21.

Chapter 5

Conclusion and Discussion

Exploring the dimensionality of test data is crucial. A well defined dimensional structure will substantiate the interpretation and uses of the test scores. The purpose of this study was to explore the dimensional structure of the CPA examination.

There are many instruments and methods to evaluate the test structure. Different methods assess different aspects of the test structure, and they may not agree. It is important for the test developer to explore the dimensionality of test data using different methods. Due to the difficulty in the usage of different methods, test developers might want to stick to one method for such analyses. Embretson and Reise (2000) criticizes such practices:

> ...researchers should now be starting to move away from reporting heuristic indices such as 'variance accounted for by the first factor' or 'ratio of first to second eigenvalue' and start implementing the new procedures that tackle these issues in a much more sophisticated manner. [...] we recommend more application of Stout's procedure for determining essential unidimensionality and/or applications of appropriate techniques such as those found in TESTFACT, POLYFACT, NOHARM, and LISCOMP. (p. 245)

In the light of this suggestion, this study used different dimensionality analysis methods from different modelling frameworks to examine the structure of the CPA exam. Linear and non-linear factor analysis methods used from a parametric framework. DIMTEST and DETECT was used as nonparametric methods to evaluate test dimensionality.

Previous dimensionality analysis of the CPA exam has been done by Seol (2013). In that study, author used only DETECT procedure to evaluate the structure of the test data. Seol (2013) found that MCQ items displayed very weak multidimensionality at all sections. TBS items showed moderate multidimensionality. Also, author found that the dimensionality structure differed between sections of the test. Further analysis showed that the items with similar difficulty, similar content and skills, and same response types clustered together.

This study found similar results. Confirmatory factor analysis results showed that MCQ items are unidimensional if the normed-chi square rule proposed by Tate (2003) used to evaluate the model fit. RMSEA values also indicated unidimensionality for MCQ items. On the other hand, TBS items were clearly not unidimensional by any means. Further exploratory factor analysis showed that TBS items can be fitted using three to six dimensions.

The essential unidimensionality test using DIMTEST indicated that none of the routes (except one) investigated were essentially unidimensional. On the other hand, DETECT results were very similar to what Seol (2013) found. MCQ items showed either no multidimensionality or weak multidimensionality. TBS items showed mostly moderate amount of multidimensionality.

Non-linear factor analysis results implemented by NOHARM were in accordance with these findings. Unidimensional models fitted to the most of the MCQ items. But TBS items can only fitted by at least three dimensions. Using panel data instead of route data did not change these conclusions.

Overall the results showed that MCQ items showed very weak amount of multidimensionality and mostly can be fitted by a unidimensional model. On the contrary, TBS items are clearly multidimensional and at least three dimensions were needed for a good fitting model.

A side benefit of this project is the creation of an R package (R Core Team, 2014) to analyze the dimensionality of the test data. This package is developed to help practitioners to perform various methods of dimensionality from one interface, R program. It follows the dimensionality analysis frameworks recommended by Tate (2003), Nandakumar and Ackerman (2004) and Jasper (2010). There are various options available in this package for the assessment of dimensionality of binary test data such as exploratory and confirmatory factor analysis using Mplus software, test for essential unidimensionality using DIMTEST, test for multidimensionality using DETECT and non-linear factor analysis using NOHARM. In addition to dimensionality analysis, estimates of the parameters can be obtained from Mplus and NOHARM. Also it is planned to add more dimensionality analysis methods in the near future. Making available all of these analysis in one place will enable test developers to evaluate the structure of their test data easily. For example, test for essential unidimensionality can be easily performed by merely writing the following code: `dimtest(testData)`. Such ease of use will encourage researchers to use multiple methods, which eventually improves the quality of tests. In addition, this package can also be used for the simulation studies, which are otherwise not feasible with the standalone programs.

5.1 Limitations and Future Directions

One of the limitations of this study is the lack of substantive analysis. Ackerman, Gierl, and Walker (2003) criticized the dependence of current practices of dimensionality analysis solely on the statistical procedures. They argued that "psychometric analyses should always be guided by substantive hypotheses"(p. 38) and dimensionality analysis should take actual test content and cognitive processes into account. One should bear in mind that correct dimensionality is not unique. There might be other statistical models that fits the data equally well. As mentioned by Tate (2002), substantive analyses should guide the choice of final statistical model:

> Statistically, any model of the correct dimensionality is not unique; an infinite number of equivalent models (i.e., models fitting the data just as well) can be obtained with transformations of the dimensions in the initial model (using, e.g., rotations of an initial factor solution). Substantive considerations based on the content and purpose of the test must guide the choice of an appropriate final statistical model. (p. 184)

Unfortunately, due to the time constraints, analyses in this report are based only on statistical methods. But the statistical findings in this report can guide further substantive analysis of the structure of the CPA examinations.

Even though, in this study four different methods used for dimensionality analysis, there are other dimensionality analysis methods that might tap on different aspects of the test data. Methods such as full-information factor analysis as implemented in TESTFACT, parallel analysis and local dependence tests might be used in the future studies. In addition, the fit of complex factor structures such as bifactor model or hierarchical factor model might be tested in the future implementations of the dimensionality analysis.

In this study, the information obtained from the test specifications did not used extensively. Content and skill specifications are available for this data (AICPA, 2012). Since most of the analysis done were exploratory, this information might guide the interpretation of the dimensions formed. For example, for TBS items, the information regarding the type of the item (whether it is tabular entry, free response or journal entry) might provide helpful information regarding the multidimensionality observed in TBS items.

Another limitation is the inability to use complete data structure in dimensionality analyses. By design, CPA exam has missing data. Two examinees that are assigned to the same panel might be directed to two different routes. Even though the routes are different, these two examinees might see overlapping testlets. For example one of them might be directed to the route 'MMM', and the other might be routed to the 'MDM' route. These two routes have two overlapping testlets (see Figure 1.1.1 on page 12). In this dimensionality analysis study, these two routes investigated separately. But in fact, they could be investigated together by acknowledging the missingness by design. This reduced the sample size used for the analysis. Using the complete data matrix would have many benefits. Combining samples from different routes will form a more heterogeneous sample. Increasing the sample size and the variability in sample will increase the power and generalizability of the results. Unfortunately, the methods used in this study does not allow the analysis of the complete data matrix. DIMTEST and DETECT does not allow missingness in the test data. In NOHARM one can run the analysis with missing data by entering the covariance matrix of the test data. Mplus can handle missing data but in these programs missingness assumed to be at random. This was not the case in this data. In the future, new methods might be investigated to cope with this problem. Especially a method that conforms the data structure of multistage testing will add power to the analysis.

References

Ackerman, T. A. (1992). A didactic explanation of item bias, item impact, and item validity from a multidimensional perspective. *Journal of Educational Measurement*, *29*(1), 67–91. Retrieved from http://www.jstor.org/stable/1434777

Ackerman, T. A., Gierl, M. J., & Walker, C. M. (2003). Using multidimensional item response theory to evaluate educational and psychological tests. *Educational Measurement: Issues and Practice*, *22*(3), 37–51. doi:10.1111/j.1745-3992.2003.tb00136.x

American Institute of Certified Public Accountants. (2012). Content and skill specifications for the uniform CPA examination. Generic. Author.

Beauducel, A. & Herzberg, P. Y. (2006). On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in CFA. *Structural Equation Modeling: A Multidisciplinary Journal*, *13*(2), 186–203. doi:10.1207/s15328007sem1302_2

Brown, T. (2006). *Confirmatory factor analysis for applied research*. New York, NY: The Guilford Press.

De Champlain, A. F. & Tang, K. L. (1997). CHIDIM: A FORTRAN program for assessing the dimensionality of binary item responses based on mcdonald's nonlinear factor analytic model. *Educational and Psychological Measurement*, *57*(1), 174–178. doi:10.1177/0013164497057001014

Douglas, J., Kim, H.-R., Roussos, L. A., Stout, W., & Zhang, J. (1999). *Lsat dimensionality analysis for the december 1991, june 1992, and october 1992 administrations* (Report No. LSAC-R-95-05). Law School Admission Council.

Embretson, S. E. & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum.

Fraser, C. & McDonald, R. P. (2012). *NOHARM 4 manual*. Retrieved from http://noharm.niagararesearch.ca/nh4man/nhman.html

Gessaroli, M. E. & De Champlain, A. F. (1996). Using an approximate chi-square statistic to test the number of dimensions underlying the responses to a set of items. *Journal of Educational Measurement*, *33*(2), 157–179. doi:10.2307/1435181

Gierl, M. J., Tan, X., & Wang, C. (2005). *Identifying content and cognitive dimensions on the sat* (Report No. 2005-11). College Board.

Hambleton, R. & Swaminathan, H. (1985). *Item response theory: principles and applications*. Boston, MA: Kluwer-Nijhoff Pub.

Hattie, J. (1985). Methodology review: assessing unidimensionality of tests and items. *Applied Psychological Measurement*, *9*(2), 139–164. doi:10.1177/014662168500900204

Hattie, J., Krakowski, K., Rogers, H. J., & Swaminathan, H. (1996). An assessment of Stout's index of essential unidimensionality. *Applied Psychological Measurement*, *20*(1), 1–14. doi:10.1177/014662169602000101

Hendrickson, A. (2007). An NCME instructional module on multistage testing. *Educational Measurement: Issues and Practice*, *26*(2), 44–52. doi:10.1111/j.1745-3992.2007.00093.x

Hu, L. & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. doi:10.1080/10705519909540118

Jang, E. E. & Roussos, L. A. (2007). An investigation into the dimensionality of TOEFL using conditional covariance-based nonparametric approach. *Journal of Educational Measurement*, *44*(1), 1–21. doi:10.2307/20461840

Jasper, F. (2010). Applied dimensionality and test structure assessment with the START-M mathematics test. *The International Journal of Educational and Psychological Assessment*, *6*(1), 104–125.

Junker, B. W. & Sijtsma, K. (2001). Nonparametric item response theory in action: an overview of the special issue. *Applied Psychological Measurement*, *25*(3), 211–220. doi:10.1177/01466210122032028

Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, *50*(1), 1–73. doi:10.1111/jedm.12000

Kim, H.-R. (1994). *New techniques for the dimensionality assessment of standardized test data* (Thesis). Retrieved from http://hdl.handle.net/2142/19110

Kline, R. B. (2010). *Principles and practice of structural equation modeling*. New York: The Guilford Press.

Lord, F. M. & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

Masters, G. N. (1988). Item discrimination: when more is worse. *Journal of Educational Measurement*, *25*(1), 15–29. doi:10.2307/1435021

McDonald, R. P. (1962). A general approach to nonlinear factor analysis. *Psychometrika*, *27*(4), 397–415. doi:10.1007/BF02289646

McDonald, R. P. (1981). The dimensionality of tests and items. *British Journal of Mathematical and Statistical Psychology*, *34*(1), 100–117. doi:10.1111/j.2044-8317.1981.tb00621.x

McDonald, R. P. (1997). Normal-ogive multidimensional model. In W. J. van der Linden & R. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 257–269). New York, NY: Springer- Verlag.

Muthén, B. O., Du Toit, S. H. C., & Spisic, D. (1997). *Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes.*

Muthén, L. K. & Muthén, B. O. (1998-2012). *Mplus user's guide*. Seventh Edition. Los Angeles, CA: Muthén & Muthén.

Nandakumar, R. & Ackerman, T. A. (2004). Test modeling. In D. Kaplan (Ed.), *The sage handbook of quantitative methodology for the social sciences* (pp. 93–105). Thousand Oaks, CA: SAGE Publications. doi:doi: http://dx.doi.org/10.4135/9781412986311

Nandakumar, R. & Stout, W. (1993). Refinements of stout's procedure for assessing latent trait unidimensionality. *Journal of Educational Statistics*, *18*(1), 41–68. doi:10.2307/1165182

Nunnally, J. C. & Bernstein, I. H. (1994). *Psychometric theory* (3rd). New York: McGraw-Hill.

R Core Team. (2014). *R: a language and environment for statistical computing.* R Foundation for Statistical Computing. Vienna, Austria. Retrieved from http://www.R-project.org

Reckase, M. (2009). *Multidimensional item response theory.* New York: Springer Verlag. doi:10.1007/978-0-387-89976-3

Roussos, L. A. & Ozbek, O. Y. (2006). Formulation of the detect population parameter and evaluation of detect estimator bias. *Journal of Educational Measurement*, *43*(3), 215–243. doi:10.2307/20461825

Roussos, L. A., Stout, W. F., & Marden, J. I. (1998). Using new proximity measures with hierarchical cluster analysis to detect multidimensionality. *Journal of Educational Measurement*, *35*(1), 1–30. doi:10.2307/1435411

Seol, J. (2013). *Dimensionality analysis of AICPA exams for the 2012 administrations.* American Institute of Certified Public Accountants.

Seraphine, A. E. (2000). The performance of dimtest when latent trait and item difficulty distributions differ. *Applied Psychological Measurement*, *24*(1), 82–94. doi:10.1177/01466216000241005

Sireci, S. G., Thissen, D., & Wainer, H. (1991). On the reliability of testlet-based tests. *Journal of Educational Measurement*, *28*(3), 237–247. doi:10.2307/1434845

Steiger, J. H. (1990). Structural model evaluation and modification: an interval estimation approach. *Multivariate Behavioral Research*, *25*(2), 173–180.

Stone, C. A. & Yeh, C.-C. (2006). Assessing the dimensionality and factor structure of multiple-choice exams: an empirical comparison of methods using the multistate bar examination. *Educational and Psychological Measurement*, *66*(2), 193–214. doi:10.1177/0013164405282483

Stout, W. (1987). A nonparametric approach for assessing latent trait unidimensionality. *Psychometrika*, *52*(4), 589–617. doi:10.1007/BF02294821

Stout, W. (1990). A new item response theory modeling approach with applications to unidimensionality assessment and ability estimation. *Psychometrika*, *55*(2), 293–325. doi:10.1007/BF02295289

Stout, W., Froelich, A. G., & Gao, F. (2001). Using resampling methods to produce an improved DIMTEST procedure. In A. Boomsma, M. A. van Duijn, & T. A. Snijders (Eds.), *Essays on item response theory* (pp. 357–375). New York: Springer.

Stout, W., Habing, B., Douglas, J., Kim, H. R., Roussos, L. A., & Zhang, J. (1996). Conditional covariance-based nonparametric multidimensionality assessment. *Applied Psychological Measurement*, *20*(4), 331–354. doi:10.1177/014662169602000403

Swaminathan, H. & Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement*, *27*(4), 361–370. Retrieved from http://www.jstor.org/stable/1434855

Takane, Y. & de Leeuw, J. (1987). On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika*, *52*(3), 393–408. doi:10.1007/bf02294363

Tate, R. (2002). Test dimensionality. In G. Tindal & T. M. Haladyna (Eds.), *Large-scale assessment programs for all students: validity, technical adequacy, and implementation* (pp. 181–211). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Tate, R. (2003). A comparison of selected empirical methods for assessing the structure of responses to test items. *Applied Psychological Measurement*, *27*(3), 159–203. doi:10.1177/0146621603027003001

Thissen, D., Steinberg, L., & Mooney, J. A. (1989). Trace lines for testlets: a use of multiple-categorical-response models. *Journal of Educational Measurement*, *26*(3), 247–260. doi:10.2307/1434990

Thurstone, L. L. (1947). *Multiple factor analysis*. Chicago: University of Chicago Press.

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. doi:10.2307/1435130

Zhang, J. & Stout, W. (1999). The theoretical detect index of dimensionality and its application to approximate simple structure. *Psychometrika*, *64*(2), 213–249. doi:10.1007/BF02294536

Appendices

Appendix A

Exploratory Factor Analysis Results

In this section, the analyses that were not presented in Section 4.1.4 on page 38 and Section 4.1.5 on page 42 are presented.

A.1 Exploratory Factor Analysis Results for MCQ items

Factor	N	Par.	Chi Sq.	*df*	χ^2_M/df	*p*-value	CFI	TLI	RMSEA	RMSEA CI
1	524	75	2859.95	2700	1.06	0.0160	0.74	0.73	0.011	0.005 - 0.014
2	524	149	2731.59	2626	1.04	0.0739	0.83	0.82	0.009	0.000 - 0.013
3	524	222	2624.34	2553	1.03	0.1590	0.88	0.87	0.007	0.000 - 0.012
4	524	294	2521.20	2481	1.02	0.2819	0.94	0.93	0.006	0.000 - 0.011
5	524	365	2418.02	2410	1.00	0.4502	0.99	0.98	0.003	0.000 - 0.010
6	524	435	2332.73	2340	1.00	0.5398	1.00	1.01	0.000	0.000 - 0.009
7	524	504	2251.93	2271	0.99	0.6085	1.00	1.04	0.000	0.000 - 0.009
8	524	572	2170.75	2203	0.99	0.6840	1.00	1.07	0.000	0.000 - 0.008

Table A.1.1: Exploratory Factor Analysis Summary for MCQ Items of AUD Section, Panel 19, Route MDD

Factor	N	Par.	Chi Sq.	*df*	χ^2_M/df	*p*-value	CFI	TLI	RMSEA	RMSEA CI
1	465	60	2162.33	1710	1.26	0.0000	0.77	0.76	0.024	0.021 - 0.027
2	465	119	1859.55	1651	1.13	0.0002	0.89	0.89	0.016	0.012 - 0.020
3	465	177	1707.64	1593	1.07	0.0230	0.94	0.94	0.012	0.005 - 0.017
4	465	234	1576.67	1536	1.03	0.2298	0.98	0.98	0.008	0.000 - 0.014
5	465	290	1503.87	1480	1.02	0.3268	0.99	0.99	0.006	0.000 - 0.013
6	465	345	1433.05	1425	1.01	0.4352	1.00	0.99	0.003	0.000 - 0.012
7	465	399	1367.17	1371	1.00	0.5242	1.00	1.00	0.000	0.000 - 0.012
8	465	452	1299.78	1318	0.99	0.6344	1.00	1.01	0.000	0.000 - 0.011

Table A.1.2: Exploratory Factor Analysis Summary for MCQ Items of BEC Section, Panel 22, Route MMM

Factor	N	Par.	Chi Sq.	*df*	χ^2_M/df	*p*-value	CFI	TLI	RMSEA	RMSEA CI
1	457	60	2207.04	1710	1.29	0.0000	0.85	0.85	0.025	0.022 - 0.028
2	457	119	1882.81	1651	1.14	0.0001	0.93	0.93	0.018	0.013 - 0.021
3	457	177	1776.81	1593	1.12	0.0008	0.94	0.94	0.016	0.011 - 0.020
4	457	234	1666.78	1536	1.09	0.0105	0.96	0.95	0.014	0.007 - 0.018
5	457	290	1572.11	1480	1.06	0.0473	0.97	0.97	0.012	0.001 - 0.017
6	457	345	1485.51	1425	1.04	0.1292	0.98	0.98	0.010	0.000 - 0.015
7	457	399	1410.65	1371	1.03	0.2228	0.99	0.98	0.008	0.000 - 0.015
8	457	452	1337.93	1318	1.02	0.3449	0.99	0.99	0.006	0.000 - 0.014

Table A.1.3: Exploratory Factor Analysis Summary for MCQ Items of BEC Section, Panel 04, Route MMM

Factor	N	Par.	Chi Sq.	*df*	χ^2_M/df	*p*-value	CFI	TLI	RMSEA	RMSEA CI
1	441	60	1981.71	1710	1.16	0.0000	0.90	0.89	0.019	0.015 - 0.023
2	441	119	1746.67	1651	1.06	0.0500	0.96	0.96	0.011	0.000 - 0.017
3	441	177	1640.84	1593	1.03	0.1974	0.98	0.98	0.008	0.000 - 0.014
4	441	234	1550.72	1536	1.01	0.3910	0.99	0.99	0.005	0.000 - 0.013
5	441	290	1474.94	1480	1.00	0.5323	1.00	1.00	0.000	0.000 - 0.012
6	441	345	1407.79	1425	0.99	0.6222	1.00	1.01	0.000	0.000 - 0.011
7	441	399	1340.30	1371	0.98	0.7183	1.00	1.01	0.000	0.000 - 0.010
8	441	452	1276.57	1318	0.97	0.7889	1.00	1.02	0.000	0.000 - 0.009

Table A.1.4: Exploratory Factor Analysis Summary for MCQ Items of BEC Section, Panel 18, Route MMM

Factor	N	Par.	Chi Sq.	*df*	χ^2_M/df	*p*-value	CFI	TLI	RMSEA	RMSEA CI
1	363	74	2914.28	2627	1.11	0.0001	0.85	0.85	0.017	0.013 - 0.021
2	363	147	2682.73	2554	1.05	0.0375	0.94	0.93	0.012	0.003 - 0.017
3	363	219	2573.18	2482	1.04	0.0989	0.95	0.95	0.010	0.000 - 0.015
4	363	290	2479.52	2411	1.03	0.1618	0.96	0.96	0.009	0.000 - 0.015
5	363	360	2387.45	2341	1.02	0.2469	0.98	0.97	0.007	0.000 - 0.014
6	363	429	2300.84	2272	1.01	0.3314	0.98	0.98	0.006	0.000 - 0.013
7	363	497	2223.24	2204	1.01	0.3825	0.99	0.99	0.005	0.000 - 0.013
8	363	564	2143.41	2137	1.00	0.4569	1.00	1.00	0.003	0.000 - 0.012

Table A.1.5: Exploratory Factor Analysis Summary for MCQ Items of FAR Section, Panel 12, Route MMM

Factor	N	Par.	Chi Sq.	*df*	χ^2_M/df	*p*-value	CFI	TLI	RMSEA	RMSEA CI
1	503	60	2225.20	1710	1.30	0.0000	0.87	0.86	0.024	0.021 - 0.027
2	503	119	2014.91	1651	1.22	0.0000	0.91	0.90	0.021	0.018 - 0.024
3	503	177	1842.36	1593	1.16	0.0000	0.94	0.93	0.018	0.014 - 0.021
4	503	234	1682.55	1536	1.10	0.0050	0.96	0.96	0.014	0.008 - 0.018
5	503	290	1581.65	1480	1.07	0.0329	0.97	0.97	0.012	0.004 - 0.016
6	503	345	1493.80	1425	1.05	0.1001	0.98	0.98	0.010	0.000 - 0.015
7	503	399	1416.72	1371	1.03	0.1904	0.99	0.98	0.008	0.000 - 0.014
8	503	452	1342.47	1318	1.02	0.3132	0.99	0.99	0.006	0.000 - 0.013

Table A.1.6: Exploratory Factor Analysis Summary for MCQ Items of REG Section, Panel 13, Route MMM

A.2 Exploratory Factor Analysis Results for TBS items

Factor	N	Par.	Chi Sq.	*df*	χ^2_M/df	*p*-value	CFI	TLI	RMSEA	RMSEA CI
1	528	37	2607.35	629	4.15	0.0000	0.86	0.85	0.077	0.074 - 0.080
2	528	73	1853.41	593	3.13	0.0000	0.91	0.90	0.063	0.060 - 0.067
3	528	108	1466.09	558	2.63	0.0000	0.94	0.92	0.056	0.052 - 0.059
4	528	142	1215.80	524	2.32	0.0000	0.95	0.94	0.050	0.046 - 0.054
5	528	175	1042.41	491	2.12	0.0000	0.96	0.95	0.046	0.042 - 0.050
6	528	207	895.74	459	1.95	0.0000	0.97	0.96	0.042	0.038 - 0.047
7	528	238	772.24	428	1.80	0.0000	0.98	0.96	0.039	0.035 - 0.043
8	528	268	658.49	398	1.65	0.0000	0.98	0.97	0.035	0.030 - 0.040

Table A.2.1: Exploratory Factor Analysis Summary for TBS Items of AUD Section, Panel 11, Route MDD

Factor	N	Par.	Chi Sq.	*df*	χ^2_M/df	*p*-value	CFI	TLI	RMSEA	RMSEA CI
1	522	35	1677.57	560	3.00	0.0000	0.70	0.68	0.062	0.058 - 0.065
2	522	69	1235.25	526	2.35	0.0000	0.81	0.78	0.051	0.047 - 0.055
3	522	102	1010.93	493	2.05	0.0000	0.86	0.83	0.045	0.041 - 0.049
4	522	134	805.27	461	1.75	0.0000	0.91	0.88	0.038	0.033 - 0.042
5	522	165	662.55	430	1.54	0.0000	0.94	0.91	0.032	0.027 - 0.037
6	522	195	543.24	400	1.36	0.0000	0.96	0.94	0.026	0.020 - 0.032
7	522	224	468.89	371	1.26	0.0004	0.97	0.96	0.022	0.016 - 0.029
8	522	252	399.56	343	1.16	0.0189	0.98	0.97	0.018	0.008 - 0.025

Table A.2.2: Exploratory Factor Analysis Summary for TBS Items of AUD Section, Panel 19, Route MDD

Factor	N	Par.	Chi Sq.	*df*	χ^2_M/df	*p*-value	CFI	TLI	RMSEA	RMSEA CI
1	519	35	2275.42	560	4.06	0.0000	0.85	0.84	0.077	0.074 - 0.080
2	519	69	1367.34	526	2.60	0.0000	0.93	0.92	0.056	0.052 - 0.059
3	519	102	1077.61	493	2.19	0.0000	0.95	0.94	0.048	0.044 - 0.052
4	519	134	821.15	461	1.78	0.0000	0.97	0.96	0.039	0.034 - 0.043
5	519	165	666.85	430	1.55	0.0000	0.98	0.97	0.033	0.028 - 0.037
6	519	195	584.25	400	1.46	0.0000	0.98	0.98	0.030	0.024 - 0.035
7	519	224	523.11	371	1.41	0.0000	0.99	0.98	0.028	0.022 - 0.034
8	519	252	457.28	343	1.33	0.0000	0.99	0.98	0.025	0.019 - 0.031

Table A.2.3: Exploratory Factor Analysis Summary for TBS Items of AUD Section, Panel 21, Route MDD

Factor	N	Par.	Chi Sq.	*df*	χ^2_M/df	*p*-value	CFI	TLI	RMSEA	RMSEA CI
1	338	37	1658.99	629	2.64	0.0000	0.80	0.79	0.070	0.066 - 0.074
2	338	73	1048.63	593	1.77	0.0000	0.91	0.90	0.048	0.043 - 0.052
3	338	108	873.24	558	1.56	0.0000	0.94	0.93	0.041	0.036 - 0.046
4	338	142	754.98	524	1.44	0.0000	0.96	0.94	0.036	0.030 - 0.042
5	338	175	650.34	491	1.32	0.0000	0.97	0.96	0.031	0.024 - 0.037
6	338	207	579.80	459	1.26	0.0001	0.98	0.97	0.028	0.020 - 0.035
7	338	238	499.33	428	1.17	0.0097	0.99	0.98	0.022	0.012 - 0.030
8	338	268	445.30	398	1.12	0.0507	0.99	0.98	0.019	0.000 - 0.028

Table A.2.4: Exploratory Factor Analysis Summary for TBS Items of FAR Section, Panel 04, Route MMM

Factor	N	Par.	Chi Sq.	*df*	χ^2_M/df	*p*-value	CFI	TLI	RMSEA	RMSEA CI
1	498	26	1269.22	299	4.24	0.0000	0.92	0.91	0.081	0.076 - 0.085
2	498	51	773.13	274	2.82	0.0000	0.96	0.95	0.060	0.055 - 0.066
3	498	75	443.84	250	1.78	0.0000	0.98	0.98	0.039	0.033 - 0.045
4	498	98	313.53	227	1.38	0.0001	0.99	0.99	0.028	0.020 - 0.035
5	498	120	246.24	205	1.20	0.0258	1.00	0.99	0.020	0.008 - 0.029
6	498	141	204.21	184	1.11	0.1465	1.00	1.00	0.015	0.000 - 0.025
7	498	161	168.93	164	1.03	0.3797	1.00	1.00	0.008	0.000 - 0.022
8	498	180	133.56	145	0.92	0.7425	1.00	1.00	0.000	0.000 - 0.016

Table A.2.5: Exploratory Factor Analysis Summary for TBS Items of REG Section, Panel 13, Route MMM

Factor	N	Par.	Chi Sq.	*df*	χ^2_M/df	*p*-value	CFI	TLI	RMSEA	RMSEA CI
1	481	28	2631.49	350	7.52	0.0000	0.81	0.80	0.116	0.112 - 0.121
2	481	55	1768.03	323	5.47	0.0000	0.88	0.86	0.096	0.092 - 0.101
3	481	81	1236.71	297	4.16	0.0000	0.92	0.90	0.081	0.076 - 0.086
4	481	106	912.47	272	3.35	0.0000	0.95	0.93	0.070	0.065 - 0.075
5	481	130	677.65	248	2.73	0.0000	0.96	0.95	0.060	0.055 - 0.065
6	481	153	479.73	225	2.13	0.0000	0.98	0.96	0.049	0.043 - 0.055
7	481	175	320.63	203	1.58	0.0000	0.99	0.98	0.035	0.027 - 0.042
8	481	196	270.70	182	1.49	0.0000	0.99	0.98	0.032	0.024 - 0.040

Table A.2.6: Exploratory Factor Analysis Summary for TBS Items of REG Section, Panel 20, Route MMM

Factor	N	Par.	Chi Sq.	*df*	χ^2_M/df	*p*-value	CFI	TLI	RMSEA	RMSEA CI
1	490	32	1526.10	464	3.29	0.0000	0.85	0.84	0.068	0.065 - 0.072
2	490	63	1014.00	433	2.34	0.0000	0.92	0.90	0.052	0.048 - 0.057
3	490	93	733.53	403	1.82	0.0000	0.95	0.94	0.041	0.036 - 0.046
4	490	122	542.70	374	1.45	0.0000	0.98	0.97	0.030	0.025 - 0.036
5	490	150	358.88	346	1.04	0.3054	1.00	1.00	0.009	0.000 - 0.019
6	490	177	310.10	319	0.97	0.6290	1.00	1.00	0.000	0.000 - 0.015
7	490	203	279.85	293	0.96	0.6999	1.00	1.00	0.000	0.000 - 0.014
8	490	228	246.81	268	0.92	0.8191	1.00	1.01	0.000	0.000 - 0.012

Table A.2.7: Exploratory Factor Analysis Summary for TBS Items of REG Section, Panel 06, Route MMM

Appendix B

Extended NOHARM Output

B.1 NOHARM Analysis of MCQ and TBS Items

n_{dim}	N	n_{item}	Tanaka's GoF	RMSR	RMSR % Change
1	541	108	0.8496	0.0072	-
2	541	108	0.8791	0.0065	10%
3	541	108	0.8939	0.0061	6%
4	541	108	0.9037	0.0058	5%
5	541	108	0.9132	0.0055	5%
6	541	108	0.9195	0.0053	4%
7	541	108	0.9252	0.0051	4%
8	541	108	0.9302	0.0049	3%

Table B.1.1: NOHARM Analysis of MCQ and TBS Items - AUD Section, Panel 12, Route 'MDD'. Mean of guessing parameters for this test is 0.

n_{dim}	N	n_{item}	Tanaka's GoF	RMSR	RMSR % Change
1	528	112	0.8027	0.0085	-
2	528	112	0.8272	0.0079	6%
3	528	112	0.8487	0.0074	6%
4	528	112	0.8683	0.0069	7%
5	528	112	0.8830	0.0065	6%
6	528	112	0.8946	0.0062	5%
7	528	112	0.9034	0.0059	4%
8	528	112	0.9108	0.0057	4%

Table B.1.2: NOHARM Analysis of MCQ and TBS Items - AUD Section, Panel 11, Route 'MDD'. Mean of guessing parameters for this test is 0.22.

n_{dim}	N	n_{item}	Tanaka's GoF	RMSR	RMSR % Change
1	522	110	0.8341	0.0079	-
2	522	110	0.8532	0.0075	6%
3	522	110	0.8712	0.0070	6%
4	522	110	0.8848	0.0066	5%
5	522	110	0.8935	0.0064	4%
6	522	110	0.9019	0.0061	4%
7	522	110	0.9095	0.0059	4%
8	522	110	0.9162	0.0056	4%

Table B.1.3: NOHARM Analysis of MCQ and TBS Items - AUD Section, Panel 19, Route 'MDD'. Mean of guessing parameters for this test is 0.24.

n_{dim}	N	n_{item}	Tanaka's GoF	RMSR	RMSR % Change
1	519	110	0.8304	0.0079	-
2	519	110	0.8537	0.0073	7%
3	519	110	0.8710	0.0069	6%
4	519	110	0.8837	0.0065	5%
5	519	110	0.8937	0.0062	4%
6	519	110	0.9029	0.0060	4%
7	519	110	0.9110	0.0057	4%
8	519	110	0.9169	0.0055	3%

Table B.1.4: NOHARM Analysis of MCQ and TBS Items - AUD Section, Panel 21, Route 'MDD'. Mean of guessing parameters for this test is 0.24.

n_{dim}	N	n_{item}	Tanaka's GoF	RMSR	RMSR % Change
1	387	110	0.8099	0.0127	-
2	387	110	0.8521	0.0112	12%
3	387	110	0.8704	0.0105	6%
4	387	110	0.8856	0.0098	6%
5	387	110	0.8950	0.0094	4%
6	387	110	0.9029	0.0091	4%
7	387	110	0.9108	0.0087	4%
8	387	110	0.9164	0.0084	3%

Table B.1.5: NOHARM Analysis of MCQ and TBS Items - FAR Section, Panel 22, Route 'MMM'. Mean of guessing parameters for this test is 0.21.

n_{dim}	N	n_{item}	Tanaka's GoF	RMSR	RMSR % Change
1	383	115	0.8098	0.0127	-
2	383	115	0.8490	0.0113	11%
3	383	115	0.8708	0.0104	7%
4	383	115	0.8860	0.0098	6%
5	383	115	0.8961	0.0094	5%
6	383	115	0.9035	0.0090	4%
7	383	115	0.9110	0.0087	4%
8	383	115	0.9166	0.0084	3%

Table B.1.6: NOHARM Analysis of MCQ and TBS Items - FAR Section, Panel 09, Route 'MMM'. Mean of guessing parameters for this test is 0.2.

n_{dim}	N	n_{item}	Tanaka's GoF	RMSR	RMSR % Change
1	348	112	0.7958	0.0117	-
2	348	112	0.8305	0.0107	9%
3	348	112	0.8477	0.0101	5%
4	348	112	0.8625	0.0096	5%
5	348	112	0.8751	0.0092	5%
6	348	112	0.8843	0.0088	4%
7	348	112	0.8917	0.0086	3%
8	348	112	0.8992	0.0083	4%

Table B.1.7: NOHARM Analysis of MCQ and TBS Items - FAR Section, Panel 12, Route 'MMM'. Mean of guessing parameters for this test is 0.21.

n_{dim}	N	n_{item}	Tanaka's GoF	RMSR	RMSR % Change
1	337	112	0.7744	0.0122	-
2	337	112	0.8262	0.0107	12%
3	337	112	0.8479	0.0100	6%
4	337	112	0.8647	0.0095	6%
5	337	112	0.8782	0.0090	5%
6	337	112	0.8863	0.0087	3%
7	337	112	0.8937	0.0084	3%
8	337	112	0.9002	0.0081	3%

Table B.1.8: NOHARM Analysis of MCQ and TBS Items - FAR Section, Panel 04, Route 'MMM'. Mean of guessing parameters for this test is 0.2.

n_{dim}	N	n_{item}	Tanaka's GoF	RMSR	RMSR % Change
1	335	111	0.8076	0.0131	-
2	335	111	0.8541	0.0114	13%
3	335	111	0.8670	0.0109	5%
4	335	111	0.8773	0.0104	4%
5	335	111	0.8863	0.0100	4%
6	335	111	0.8947	0.0097	4%
7	335	111	0.9021	0.0093	4%
8	335	111	0.9082	0.0090	3%

Table B.1.9: NOHARM Analysis of MCQ and TBS Items - FAR Section, Panel 21, Route 'MMM'. Mean of guessing parameters for this test is 0.21.

n_{dim}	N	n_{item}	Tanaka's GoF	RMSR	RMSR % Change
1	561	89	0.9033	0.0094	-
2	561	89	0.9146	0.0089	6%
3	561	89	0.9232	0.0084	5%
4	561	89	0.9287	0.0081	4%
5	561	89	0.9333	0.0078	3%
6	561	89	0.9376	0.0076	3%
7	561	89	0.9419	0.0073	3%
8	561	89	0.9456	0.0071	3%

Table B.1.10: NOHARM Analysis of MCQ and TBS Items - REG Section, Panel 07, Route 'MMM'. Mean of guessing parameters for this test is 0.23.

n_{dim}	N	n_{item}	Tanaka's GoF	RMSR	RMSR % Change
1	530	89	0.8679	0.0105	-
2	530	89	0.8926	0.0095	10%
3	530	89	0.9048	0.0089	6%
4	530	89	0.9143	0.0084	5%
5	530	89	0.9211	0.0081	4%
6	530	89	0.9264	0.0078	3%
7	530	89	0.9314	0.0076	3%
8	530	89	0.9355	0.0073	3%

Table B.1.11: NOHARM Analysis of MCQ and TBS Items - REG Section, Panel 21, Route 'MMM'. Mean of guessing parameters for this test is 0.2.

n_{dim}	N	n_{item}	Tanaka's GoF	RMSR	RMSR % Change
1	498	86	0.8584	0.0109	-
2	498	86	0.8858	0.0098	10%
3	498	86	0.9045	0.0090	9%
4	498	86	0.9178	0.0083	7%
5	498	86	0.9242	0.0080	4%
6	498	86	0.9295	0.0077	4%
7	498	86	0.9350	0.0074	4%
8	498	86	0.9394	0.0071	3%

Table B.1.12: NOHARM Analysis of MCQ and TBS Items - REG Section, Panel 13, Route 'MMM'. Mean of guessing parameters for this test is 0.2.

n_{dim}	N	n_{item}	Tanaka's GoF	RMSR	RMSR % Change
1	481	88	0.8215	0.0128	-
2	481	88	0.8704	0.0109	15%
3	481	88	0.8908	0.0100	8%
4	481	88	0.9026	0.0095	6%
5	481	88	0.9105	0.0091	4%
6	481	88	0.9191	0.0087	5%
7	481	88	0.9244	0.0084	3%
8	481	88	0.9291	0.0081	3%

Table B.1.13: NOHARM Analysis of MCQ and TBS Items - REG Section, Panel 20, Route 'MMM'. Mean of guessing parameters for this test is 0.19.

n_{dim}	N	n_{item}	Tanaka's GoF	RMSR	RMSR % Change
1	489	92	0.8471	0.0121	-
2	489	92	0.8971	0.0099	18%
3	489	92	0.9067	0.0094	5%
4	489	92	0.9159	0.0089	5%
5	489	92	0.9246	0.0085	5%
6	489	92	0.9326	0.0080	5%
7	489	92	0.9386	0.0076	5%
8	489	92	0.9436	0.0073	4%

Table B.1.14: NOHARM Analysis of MCQ and TBS Items - REG Section, Panel 06, Route 'MMM'. Mean of guessing parameters for this test is 0.21.

B.2 NOHARM Analysis of MCQ Items

n_{dim}	N	n_{item}	Tanaka's GoF	RMSR	RMSR % Change
1	541	75	0.9031	0.0058	-
2	541	75	0.9150	0.0055	6%
3	541	75	0.9250	0.0051	6%
4	541	75	0.9314	0.0049	4%
5	541	75	0.9377	0.0047	5%
6	541	75	0.9429	0.0045	4%
7	541	75	0.9477	0.0043	4%
8	541	75	0.9520	0.0041	4%

Table B.2.1: NOHARM Analysis of MCQ Items - AUD Section, Panel 12, Route 'MDD'. Mean of guessing parameters for this test is 0.26.

n_{dim}	N	n_{item}	Tanaka's GoF	RMSR	RMSR % Change
1	529	75	0.8914	0.0062	-
2	529	75	0.9050	0.0058	6%
3	529	75	0.9150	0.0055	5%
4	529	75	0.9231	0.0052	5%
5	529	75	0.9309	0.0049	5%
6	529	75	0.9362	0.0047	4%
7	529	75	0.9414	0.0045	4%
8	529	75	0.9460	0.0044	4%

Table B.2.2: NOHARM Analysis of MCQ Items - AUD Section, Panel 11, Route 'MDD'. Mean of guessing parameters for this test is 0.24.

n_{dim}	N	n_{item}	Tanaka's GoF	RMSR	RMSR % Change
1	524	75	0.8994	0.0059	-
2	524	75	0.9097	0.0056	5%
3	524	75	0.9176	0.0054	4%
4	524	75	0.9247	0.0051	4%
5	524	75	0.9319	0.0049	5%
6	524	75	0.9383	0.0047	5%
7	524	75	0.9437	0.0044	4%
8	524	75	0.9483	0.0043	4%

Table B.2.3: NOHARM Analysis of MCQ Items - AUD Section, Panel 19, Route 'MDD'. Mean of guessing parameters for this test is 0.26.

n_{dim}	N	n_{item}	Tanaka's GoF	RMSR	RMSR % Change
1	519	75	0.8878	0.0064	-
2	519	75	0.9036	0.0060	7%
3	519	75	0.9175	0.0055	8%
4	519	75	0.9258	0.0052	5%
5	519	75	0.9323	0.0050	4%
6	519	75	0.9376	0.0048	4%
7	519	75	0.9430	0.0046	4%
8	519	75	0.9477	0.0044	4%

Table B.2.4: NOHARM Analysis of MCQ Items - AUD Section, Panel 21, Route 'MDD'. Mean of guessing parameters for this test is 0.26.

n_{dim}	N	n_{item}	Tanaka's GoF	RMSR	RMSR % Change
1	484	60	0.9102	0.0098	-
2	484	60	0.9244	0.0090	8%
3	484	60	0.9339	0.0084	6%
4	484	60	0.9407	0.0080	5%
5	484	60	0.9464	0.0076	5%
6	484	60	0.9517	0.0072	5%
7	484	60	0.9561	0.0068	5%
8	484	60	0.9600	0.0065	5%

Table B.2.5: NOHARM Analysis of MCQ Items - BEC Section, Panel 02, Route 'MMM'. Mean of guessing parameters for this test is 0.24.

n_{dim}	N	n_{item}	Tanaka's GoF	RMSR	RMSR % Change
1	468	60	0.9112	0.0097	-
2	468	60	0.9245	0.0090	8%
3	468	60	0.9368	0.0082	9%
4	468	60	0.9437	0.0077	6%
5	468	60	0.9493	0.0073	5%
6	468	60	0.9539	0.0070	5%
7	468	60	0.9581	0.0067	5%
8	468	60	0.9620	0.0064	5%

Table B.2.6: NOHARM Analysis of MCQ Items - BEC Section, Panel 15, Route 'MMM'. Mean of guessing parameters for this test is 0.24.

n_{dim}	N	n_{item}	Tanaka's GoF	RMSR	RMSR % Change
1	465	60	0.8896	0.0109	-
2	465	60	0.9224	0.0091	16%
3	465	60	0.9348	0.0083	8%
4	465	60	0.9451	0.0077	8%
5	465	60	0.9498	0.0073	4%
6	465	60	0.9539	0.0070	4%
7	465	60	0.9577	0.0067	4%
8	465	60	0.9614	0.0064	5%

Table B.2.7: NOHARM Analysis of MCQ Items - BEC Section, Panel 22, Route 'MMM'. Mean of guessing parameters for this test is 0.24.

n_{dim}	N	n_{item}	Tanaka's GoF	RMSR	RMSR % Change
1	457	60	0.9099	0.0103	-
2	457	60	0.9330	0.0089	14%
3	457	60	0.9409	0.0083	6%
4	457	60	0.9472	0.0079	5%
5	457	60	0.9523	0.0075	5%
6	457	60	0.9567	0.0071	5%
7	457	60	0.9603	0.0068	4%
8	457	60	0.9638	0.0065	4%

Table B.2.8: NOHARM Analysis of MCQ Items - BEC Section, Panel 04, Route 'MMM'. Mean of guessing parameters for this test is 0.24.

n_{dim}	N	n_{item}	Tanaka's GoF	RMSR	RMSR % Change
1	441	60	0.9122	0.0102	-
2	441	60	0.9349	0.0088	14%
3	441	60	0.9425	0.0083	6%
4	441	60	0.9486	0.0078	5%
5	441	60	0.9529	0.0075	4%
6	441	60	0.9569	0.0072	4%
7	441	60	0.9605	0.0069	4%
8	441	60	0.9637	0.0066	4%

Table B.2.9: NOHARM Analysis of MCQ Items - BEC Section, Panel 18, Route 'MMM'. Mean of guessing parameters for this test is 0.23.

n_{dim}	N	n_{item}	Tanaka's GoF	RMSR	RMSR % Change
1	397	74	0.8877	0.0110	-
2	397	74	0.9007	0.0103	6%
3	397	74	0.9098	0.0098	5%
4	397	74	0.9175	0.0094	4%
5	397	74	0.9242	0.0090	4%
6	397	74	0.9297	0.0087	4%
7	397	74	0.9350	0.0084	4%
8	397	74	0.9400	0.0080	4%

Table B.2.10: NOHARM Analysis of MCQ Items - FAR Section, Panel 22, Route 'MMM'. Mean of guessing parameters for this test is 0.25.

n_{dim}	N	n_{item}	Tanaka's GoF	RMSR	RMSR % Change
1	395	74	0.8771	0.0109	-
2	395	74	0.8971	0.0099	9%
3	395	74	0.9052	0.0096	4%
4	395	74	0.9130	0.0091	4%
5	395	74	0.9193	0.0088	4%
6	395	74	0.9247	0.0085	3%
7	395	74	0.9298	0.0082	3%
8	395	74	0.9344	0.0079	3%

Table B.2.11: NOHARM Analysis of MCQ Items - FAR Section, Panel 09, Route 'MMM'. Mean of guessing parameters for this test is 0.25.

n_{dim}	N	n_{item}	Tanaka's GoF	RMSR	RMSR % Change
1	363	74	0.8667	0.0117	-
2	363	74	0.8935	0.0105	11%
3	363	74	0.9029	0.0100	5%
4	363	74	0.9110	0.0096	4%
5	363	74	0.9183	0.0092	4%
6	363	74	0.9247	0.0088	4%
7	363	74	0.9302	0.0085	4%
8	363	74	0.9354	0.0081	4%

Table B.2.12: NOHARM Analysis of MCQ Items - FAR Section, Panel 12, Route 'MMM'. Mean of guessing parameters for this test is 0.25.

n_{dim}	N	n_{item}	Tanaka's GoF	RMSR	RMSR % Change
1	355	75	0.8520	0.0116	-
2	355	75	0.8897	0.0101	14%
3	355	75	0.8994	0.0096	5%
4	355	75	0.9072	0.0092	4%
5	355	75	0.9145	0.0089	4%
6	355	75	0.9215	0.0085	4%
7	355	75	0.9277	0.0081	4%
8	355	75	0.9333	0.0078	4%

Table B.2.13: NOHARM Analysis of MCQ Items - FAR Section, Panel 04, Route 'MMM'. Mean of guessing parameters for this test is 0.25.

n_{dim}	N	n_{item}	Tanaka's GoF	RMSR	RMSR % Change
1	353	74	0.8677	0.0114	-
2	353	74	0.8891	0.0105	8%
3	353	74	0.9001	0.0099	5%
4	353	74	0.9083	0.0095	4%
5	353	74	0.9156	0.0091	4%
6	353	74	0.9225	0.0087	4%
7	353	74	0.9278	0.0084	3%
8	353	74	0.9331	0.0081	4%

Table B.2.14: NOHARM Analysis of MCQ Items - FAR Section, Panel 21, Route 'MMM'. Mean of guessing parameters for this test is 0.25.

n_{dim}	N	n_{item}	Tanaka's GoF	RMSR	RMSR % Change
1	568	60	0.9347	0.0092	-
2	568	60	0.9415	0.0087	5%
3	568	60	0.9476	0.0082	5%
4	568	60	0.9531	0.0078	5%
5	568	60	0.9574	0.0074	5%
6	568	60	0.9614	0.0070	5%
7	568	60	0.9651	0.0067	5%
8	568	60	0.9681	0.0064	4%

Table B.2.15: NOHARM Analysis of MCQ Items - REG Section, Panel 07, Route 'MMM'. Mean of guessing parameters for this test is 0.23.

n_{dim}	N	n_{item}	Tanaka's GoF	RMSR	RMSR % Change
1	539	60	0.9266	0.0090	-
2	539	60	0.9361	0.0084	7%
3	539	60	0.9435	0.0079	6%
4	539	60	0.9500	0.0074	6%
5	539	60	0.9544	0.0071	4%
6	539	60	0.9581	0.0068	4%
7	539	60	0.9610	0.0065	4%
8	539	60	0.9639	0.0063	4%

Table B.2.16: NOHARM Analysis of MCQ Items - REG Section, Panel 21, Route 'MMM'. Mean of guessing parameters for this test is 0.24.

n_{dim}	N	n_{item}	Tanaka's GoF	RMSR	RMSR % Change
1	503	60	0.9145	0.0101	-
2	503	60	0.9301	0.0091	10%
3	503	60	0.9414	0.0083	8%
4	503	60	0.9493	0.0078	7%
5	503	60	0.9546	0.0073	5%
6	503	60	0.9587	0.0070	5%
7	503	60	0.9625	0.0067	5%
8	503	60	0.9659	0.0064	5%

Table B.2.17: NOHARM Analysis of MCQ Items - REG Section, Panel 13, Route 'MMM'. Mean of guessing parameters for this test is 0.24.

n_{dim}	N	n_{item}	Tanaka's GoF	RMSR	RMSR % Change
1	491	60	0.9207	0.0102	-
2	491	60	0.9315	0.0095	7%
3	491	60	0.9385	0.0090	5%
4	491	60	0.9454	0.0084	6%
5	491	60	0.9514	0.0080	6%
6	491	60	0.9559	0.0076	5%
7	491	60	0.9603	0.0072	5%
8	491	60	0.9637	0.0069	4%

Table B.2.18: NOHARM Analysis of MCQ Items - REG Section, Panel 20, Route 'MMM'. Mean of guessing parameters for this test is 0.21.

n_{dim}	N	n_{item}	Tanaka's GoF	RMSR	RMSR % Change
1	490	60	0.9128	0.0102	-
2	490	60	0.9251	0.0094	7%
3	490	60	0.9361	0.0087	8%
4	490	60	0.9434	0.0082	6%
5	490	60	0.9499	0.0077	6%
6	490	60	0.9550	0.0073	5%
7	490	60	0.9594	0.0069	5%
8	490	60	0.9628	0.0066	4%

Table B.2.19: NOHARM Analysis of MCQ Items - REG Section, Panel 06, Route 'MMM'. Mean of guessing parameters for this test is 0.23.

B.3 NOHARM Analysis of TBS Route Items

n_{dim}	N	n_{item}	Tanaka's GoF	RMSR	RMSR % Change
1	541	33	0.9233	0.0111	-
2	541	33	0.9465	0.0093	17%
3	541	33	0.9632	0.0077	17%
4	541	33	0.9731	0.0066	15%
5	541	33	0.9789	0.0058	11%
6	541	33	0.9834	0.0052	11%
7	541	33	0.9863	0.0047	9%
8	541	33	0.9888	0.0042	10%

Table B.3.1: NOHARM Analysis of TBS Route Items - AUD Section, Panel 12, Route 'MDD'. Mean of guessing parameters for this test is 0.

n_{dim}	N	n_{item}	Tanaka's GoF	RMSR	RMSR % Change
1	528	37	0.8329	0.0166	-
2	528	37	0.8715	0.0145	12%
3	528	37	0.9008	0.0128	12%
4	528	37	0.9243	0.0111	13%
5	528	37	0.9416	0.0098	12%
6	528	37	0.9515	0.0089	9%
7	528	37	0.9608	0.0080	10%
8	528	37	0.9654	0.0075	6%

Table B.3.2: NOHARM Analysis of TBS Route Items - AUD Section, Panel 11, Route 'MDD'. Mean of guessing parameters for this test is 0.16.

n_{dim}	N	n_{item}	Tanaka's GoF	RMSR	RMSR % Change
1	522	35	0.8887	0.0142	-
2	522	35	0.9204	0.0120	15%
3	522	35	0.9412	0.0103	14%
4	522	35	0.9524	0.0093	10%
5	522	35	0.9615	0.0084	10%
6	522	35	0.9699	0.0074	12%
7	522	35	0.9754	0.0067	10%
8	522	35	0.9773	0.0064	4%

Table B.3.3: NOHARM Analysis of TBS Route Items - AUD Section, Panel 19, Route 'MDD'. Mean of guessing parameters for this test is 0.21.

n_{dim}	N	n_{item}	Tanaka's GoF	RMSR	RMSR % Change
1	519	35	0.8815	0.0136	-
2	519	35	0.9202	0.0112	18%
3	519	35	0.9403	0.0097	13%
4	519	35	0.9550	0.0084	13%
5	519	35	0.9638	0.0075	10%
6	519	35	0.9678	0.0071	6%
7	519	35	0.9720	0.0066	7%
8	519	35	0.9755	0.0062	6%

Table B.3.4: NOHARM Analysis of TBS Route Items - AUD Section, Panel 21, Route 'MDD'. Mean of guessing parameters for this test is 0.2.

n_{dim}	N	n_{item}	Tanaka's GoF	RMSR	RMSR % Change
1	387	36	0.8525	0.0179	-
2	387	36	0.9045	0.0144	20%
3	387	36	0.9406	0.0114	21%
4	387	36	0.9565	0.0097	14%
5	387	36	0.9647	0.0088	10%
6	387	36	0.9710	0.0079	9%
7	387	36	0.9748	0.0074	7%
8	387	36	0.9772	0.0070	5%

Table B.3.5: NOHARM Analysis of TBS Route Items - FAR Section, Panel 22, Route 'MMM'. Mean of guessing parameters for this test is 0.14.

n_{dim}	N	n_{item}	Tanaka's GoF	RMSR	RMSR % Change
1	383	41	0.8545	0.0190	-
2	383	41	0.9083	0.0151	21%
3	383	41	0.9303	0.0132	13%
4	383	41	0.9486	0.0113	14%
5	383	41	0.9615	0.0098	13%
6	383	41	0.9692	0.0088	11%
7	383	41	0.9749	0.0079	10%
8	383	41	0.9777	0.0074	6%

Table B.3.6: NOHARM Analysis of TBS Route Items - FAR Section, Panel 09, Route 'MMM'. Mean of guessing parameters for this test is 0.11.

n_{dim}	N	n_{item}	Tanaka's GoF	RMSR	RMSR % Change
1	348	38	0.8061	0.0165	-
2	348	38	0.8793	0.0130	21%
3	348	38	0.9217	0.0105	19%
4	348	38	0.9392	0.0092	12%
5	348	38	0.9498	0.0084	9%
6	348	38	0.9563	0.0078	7%
7	348	38	0.9603	0.0075	5%
8	348	38	0.9650	0.0070	6%

Table B.3.7: NOHARM Analysis of TBS Route Items - FAR Section, Panel 12, Route 'MMM'. Mean of guessing parameters for this test is 0.13.

n_{dim}	N	n_{item}	Tanaka's GoF	RMSR	RMSR % Change
1	338	37	0.8241	0.0164	-
2	338	37	0.8810	0.0135	18%
3	338	37	0.9240	0.0108	20%
4	338	37	0.9461	0.0091	16%
5	338	37	0.9578	0.0080	12%
6	338	37	0.9654	0.0073	9%
7	338	37	0.9696	0.0068	6%
8	338	37	0.9737	0.0064	7%

Table B.3.8: NOHARM Analysis of TBS Route Items - FAR Section, Panel 04, Route 'MMM'. Mean of guessing parameters for this test is 0.11.

n_{dim}	N	n_{item}	Tanaka's GoF	RMSR	RMSR % Change
1	335	37	0.8988	0.0170	-
2	335	37	0.9218	0.0150	12%
3	335	37	0.9380	0.0133	11%
4	335	37	0.9506	0.0119	11%
5	335	37	0.9594	0.0108	9%
6	335	37	0.9670	0.0097	10%
7	335	37	0.9706	0.0092	6%
8	335	37	0.9743	0.0086	7%

Table B.3.9: NOHARM Analysis of TBS Route Items - FAR Section, Panel 21, Route 'MMM'. Mean of guessing parameters for this test is 0.12.

n_{dim}	N	n_{item}	Tanaka's GoF	RMSR	RMSR % Change
1	561	29	0.9285	0.0121	-
2	561	29	0.9444	0.0106	12%
3	561	29	0.9566	0.0094	12%
4	561	29	0.9671	0.0082	13%
5	561	29	0.9725	0.0075	9%
6	561	29	0.9746	0.0072	4%
7	561	29	0.9779	0.0067	7%
8	561	29	0.9795	0.0065	4%

Table B.3.10: NOHARM Analysis of TBS Route Items - REG Section, Panel 07, Route 'MMM'. Mean of guessing parameters for this test is 0.24.

n_{dim}	N	n_{item}	Tanaka's GoF	RMSR	RMSR % Change
1	530	29	0.8837	0.0165	-
2	530	29	0.9128	0.0143	13%
3	530	29	0.9318	0.0126	12%
4	530	29	0.9437	0.0115	9%
5	530	29	0.9534	0.0104	9%
6	530	29	0.9598	0.0097	7%
7	530	29	0.9637	0.0092	5%
8	530	29	0.9651	0.0090	2%

Table B.3.11: NOHARM Analysis of TBS Route Items - REG Section, Panel 21, Route 'MMM'. Mean of guessing parameters for this test is 0.14.

n_{dim}	N	n_{item}	Tanaka's GoF	RMSR	RMSR % Change
1	498	26	0.8712	0.0168	-
2	498	26	0.9240	0.0129	23%
3	498	26	0.9474	0.0108	17%
4	498	26	0.9577	0.0096	10%
5	498	26	0.9608	0.0093	4%
6	498	26	0.9648	0.0088	5%
7	498	26	0.9665	0.0086	3%
8	498	26	0.9682	0.0084	2%

Table B.3.12: NOHARM Analysis of TBS Route Items - REG Section, Panel 13, Route 'MMM'. Mean of guessing parameters for this test is 0.12.

n_{dim}	N	n_{item}	Tanaka's GoF	RMSR	RMSR % Change
1	481	28	0.8185	0.0215	-
2	481	28	0.8761	0.0178	17%
3	481	28	0.9086	0.0153	14%
4	481	28	0.9231	0.0140	8%
5	481	28	0.9349	0.0129	8%
6	481	28	0.9423	0.0121	6%
7	481	28	0.9490	0.0114	6%
8	481	28	0.9495	0.0114	1%

Table B.3.13: NOHARM Analysis of TBS Route Items - REG Section, Panel 20, Route 'MMM'. Mean of guessing parameters for this test is 0.13.

n_{dim}	N	n_{item}	Tanaka's GoF	RMSR	RMSR % Change
1	490	32	0.9271	0.0139	-
2	490	32	0.9455	0.0120	14%
3	490	32	0.9619	0.0100	16%
4	490	32	0.9741	0.0083	17%
5	490	32	0.9823	0.0068	17%
6	490	32	0.9851	0.0063	8%
7	490	32	0.9866	0.0059	5%
8	490	32	0.9876	0.0057	4%

Table B.3.14: NOHARM Analysis of TBS Route Items - REG Section, Panel 06, Route 'MMM'. Mean of guessing parameters for this test is 0.16.

B.4 NOHARM Analysis of TBS Panel Items

n_{dim}	N	n_{item}	Tanaka's GoF	RMSR	RMSR % Change
1	731	34	0.8685	0.0163	-
2	731	34	0.9181	0.0129	21%
3	731	34	0.9353	0.0114	11%
4	731	34	0.9483	0.0102	11%
5	731	34	0.9560	0.0094	8%
6	731	34	0.9626	0.0087	8%
7	731	34	0.9663	0.0083	5%
8	731	34	0.9695	0.0079	5%

Table B.4.1: NOHARM Analysis of TBS Panel Items - AUD Section, Panel 01. Mean of guessing parameters for this test is 0.2.

n_{dim}	N	n_{item}	Tanaka's GoF	RMSR	RMSR % Change
1	719	33	0.8783	0.0139	-
2	719	33	0.9155	0.0116	17%
3	719	33	0.9449	0.0094	19%
4	719	33	0.9556	0.0084	10%
5	719	33	0.9604	0.0079	6%
6	719	33	0.9640	0.0076	5%
7	719	33	0.9660	0.0074	3%
8	719	33	0.9674	0.0072	2%

Table B.4.2: NOHARM Analysis of TBS Panel Items - AUD Section, Panel 06. Mean of guessing parameters for this test is 0.24.

n_{dim}	N	n_{item}	Tanaka's GoF	RMSR	RMSR % Change
1	715	34	0.8932	0.0141	-
2	715	34	0.9242	0.0119	16%
3	715	34	0.9396	0.0106	11%
4	715	34	0.9481	0.0099	7%
5	715	34	0.9545	0.0092	6%
6	715	34	0.9586	0.0088	5%
7	715	34	0.9622	0.0084	4%
8	715	34	0.9641	0.0082	3%

Table B.4.3: NOHARM Analysis of TBS Panel Items - AUD Section, Panel 07. Mean of guessing parameters for this test is 0.22.

n_{dim}	N	n_{item}	Tanaka's GoF	RMSR	RMSR % Change
1	703	30	0.9170	0.0134	-
2	703	30	0.9476	0.0106	20%
3	703	30	0.9619	0.0091	15%
4	703	30	0.9730	0.0076	16%
5	703	30	0.9813	0.0064	17%
6	703	30	0.9881	0.0051	20%
7	703	30	0.9899	0.0047	8%
8	703	30	0.9920	0.0041	11%

Table B.4.4: NOHARM Analysis of TBS Panel Items - AUD Section, Panel 02. Mean of guessing parameters for this test is 0.17.

n_{dim}	N	n_{item}	Tanaka's GoF	RMSR	RMSR % Change
1	807	41	0.8967	0.0175	-
2	807	41	0.9346	0.0139	20%
3	807	41	0.9499	0.0122	12%
4	807	41	0.9638	0.0103	15%
5	807	41	0.9725	0.0090	13%
6	807	41	0.9785	0.0080	11%
7	807	41	0.9829	0.0071	11%
8	807	41	0.9863	0.0064	11%

Table B.4.5: NOHARM Analysis of TBS Panel Items - FAR Section, Panel 09. Mean of guessing parameters for this test is 0.11.

n_{dim}	N	n_{item}	Tanaka's GoF	RMSR	RMSR % Change
1	773	36	0.8984	0.0173	-
2	773	36	0.9364	0.0137	21%
3	773	36	0.9600	0.0108	21%
4	773	36	0.9696	0.0094	13%
5	773	36	0.9783	0.0080	15%
6	773	36	0.9841	0.0068	15%
7	773	36	0.9870	0.0062	10%
8	773	36	0.9882	0.0059	5%

Table B.4.6: NOHARM Analysis of TBS Panel Items - FAR Section, Panel 22. Mean of guessing parameters for this test is 0.14.

n_{dim}	N	n_{item}	Tanaka's GoF	RMSR	RMSR % Change
1	768	39	0.9126	0.0159	-
2	768	39	0.9408	0.0131	18%
3	768	39	0.9529	0.0117	11%
4	768	39	0.9649	0.0101	14%
5	768	39	0.9750	0.0085	16%
6	768	39	0.9826	0.0071	16%
7	768	39	0.9862	0.0063	11%
8	768	39	0.9893	0.0056	12%

Table B.4.7: NOHARM Analysis of TBS Panel Items - FAR Section, Panel 10. Mean of guessing parameters for this test is 0.12.

n_{dim}	N	n_{item}	Tanaka's GoF	RMSR	RMSR % Change
1	760	36	0.8957	0.0139	-
2	760	36	0.9400	0.0105	24%
3	760	36	0.9579	0.0088	16%
4	760	36	0.9712	0.0073	17%
5	760	36	0.9753	0.0068	7%
6	760	36	0.9796	0.0062	9%
7	760	36	0.9819	0.0058	6%
8	760	36	0.9838	0.0055	6%

Table B.4.8: NOHARM Analysis of TBS Panel Items - FAR Section, Panel 17. Mean of guessing parameters for this test is 0.15.

n_{dim}	N	n_{item}	Tanaka's GoF	RMSR	RMSR % Change
1	753	38	0.8556	0.0159	-
2	753	38	0.9010	0.0132	17%
3	753	38	0.9333	0.0108	18%
4	753	38	0.9465	0.0097	10%
5	753	38	0.9581	0.0086	11%
6	753	38	0.9659	0.0077	10%
7	753	38	0.9724	0.0070	10%
8	753	38	0.9757	0.0065	6%

Table B.4.9: NOHARM Analysis of TBS Panel Items - FAR Section, Panel 18. Mean of guessing parameters for this test is 0.13.

n_{dim}	N	n_{item}	Tanaka's GoF	RMSR	RMSR % Change
1	847	32	0.8897	0.0179	-
2	847	32	0.9322	0.0140	22%
3	847	32	0.9526	0.0117	16%
4	847	32	0.9661	0.0099	15%
5	847	32	0.9771	0.0082	18%
6	847	32	0.9832	0.0070	14%
7	847	32	0.9878	0.0060	15%
8	847	32	0.9893	0.0056	6%

Table B.4.10: NOHARM Analysis of TBS Panel Items - REG Section, Panel 04. Mean of guessing parameters for this test is 0.12.

n_{dim}	N	n_{item}	Tanaka's GoF	RMSR	RMSR % Change
1	810	29	0.9457	0.0106	-
2	810	29	0.9581	0.0093	12%
3	810	29	0.9687	0.0080	14%
4	810	29	0.9771	0.0069	14%
5	810	29	0.9806	0.0063	8%
6	810	29	0.9839	0.0058	9%
7	810	29	0.9851	0.0055	4%
8	810	29	0.9872	0.0051	7%

Table B.4.11: NOHARM Analysis of TBS Panel Items - REG Section, Panel 07. Mean of guessing parameters for this test is 0.24.

n_{dim}	N	n_{item}	Tanaka's GoF	RMSR	RMSR % Change
1	799	31	0.9306	0.0133	-
2	799	31	0.9511	0.0112	16%
3	799	31	0.9663	0.0093	17%
4	799	31	0.9779	0.0075	19%
5	799	31	0.9863	0.0059	21%
6	799	31	0.9902	0.0050	16%
7	799	31	0.9919	0.0046	9%
8	799	31	0.9930	0.0042	7%

Table B.4.12: NOHARM Analysis of TBS Panel Items - REG Section, Panel 05. Mean of guessing parameters for this test is 0.12.

n_{dim}	N	n_{item}	Tanaka's GoF	RMSR	RMSR % Change
1	791	28	0.8463	0.0235	-
2	791	28	0.9129	0.0177	25%
3	791	28	0.9549	0.0127	28%
4	791	28	0.9694	0.0105	18%
5	791	28	0.9749	0.0095	9%
6	791	28	0.9774	0.0090	5%
7	791	28	0.9782	0.0088	2%
8	791	28	0.9793	0.0086	3%

Table B.4.13: NOHARM Analysis of TBS Panel Items - REG Section, Panel 09. Mean of guessing parameters for this test is 0.13.

n_{dim}	N	n_{item}	Tanaka's GoF	RMSR	RMSR % Change
1	791	26	0.8727	0.0180	-
2	791	26	0.9317	0.0132	27%
3	791	26	0.9540	0.0108	18%
4	791	26	0.9659	0.0093	14%
5	791	26	0.9702	0.0087	7%
6	791	26	0.9731	0.0083	5%
7	791	26	0.9741	0.0081	2%
8	791	26	0.9753	0.0079	2%

Table B.4.14: NOHARM Analysis of TBS Panel Items - REG Section, Panel 13. Mean of guessing parameters for this test is 0.12.