

## Investigating the Dimensionality of an Admission Specific Purpose Language Test

*Razieh Rabbani Yekta*

Faculty of Foreign Languages, University of Isfahan, Iran

**Abstract:** The present study examines the dimensionality of an admission language test for Accounting in which general and specific purposes language components are measured in a de-modularized version. The question is that whether items assembled in one subset in such complex tests contributed only to a simple structure (measuring a single trait) or to a complex one (with each item set measuring more than one latent ability) and whether through some modifications in item assembly sub-layer of test architecture, the upgraded item -sets could represent the true dimensionality of the test, hence meaningful part-scores for each set. To this end, two consecutive phases ordered the progress of the research of which the first one is a dimensionality decision phase and the second one is the model recovery. The results showed that the concerned test violates the assumption of unidimensionality from both psychometric and psychological perspectives and therefore, items should be calibrated under some higher-dimensional model conditions. This research has implications for any testing situation where testing agencies in the rush to implement comprehensive accountability systems may be trapped in the chaotic domain of multi-objective, multi-purpose tests.

**Key words:** English for specific purpose test • General-Specific purpose English • Rasch analysis • Test dimensionality

### INTRODUCTION

In Iran's English for Specific Purpose (ESP) context, ESP tests are administered at two high-stakes levels, one as the Master's Degree Entrance Exam and the second for the admission of the graduates into PhD programs. In this latter case, tests were traditionally administered in two stages: 1) EGP as the prerequisite for the second stage which was 2) the specialized module. These two modules were separate from each other in all the phases of design, administration, scoring and reporting. Something that makes the Master's Degree (M.A) counterpart worth investigating and at the same time challenging is the joint nature of these two components at all of the above mentioned phases.

To focus more closely on the M.A. entrance exam, its internal structure is composed of a battery of about 8 different subtests (including one English subtest plus 4 to 7 knowledge subtests of different content courses) where the raw scores of individual subtests are averaged to form an overall test battery mean and participants are ranked based on a composite

percentile. As to the language subtest, after experiencing several years of upgrading, test developers in Iran have finally reached a fixed framework for the language subtest. For almost all of the fields, this subtest consists of two parts: general English and specialized English. The first part starts with 10 vocabulary test items and continues with a cloze test of grammar with a text of non specialized content. The specialized part is composed of 2 field specific reading test-lets, each with about 5 multiple choice (dichotomously scored) items. Regarding the fact that these two parts have been examined in two independent tests for the counterpart exam for entering the PhD program, the present study aimed at detecting whether items assembled in one subset in such complex tests contributed only to a simple structure (measuring a single trait) or to a complex one (with each item set measuring more than one latent ability) and whether through some modifications in item assembly sub-layer of test architecture, the upgraded item -sets could represent the true dimensionality of the test, hence meaningful part-scores for each set.

**Corresponding Author:** Razieh Rabbani Yekta, Faculty of Foreign Languages, Isfahan University,  
Hizar Jarib Street, P.O. Box: 8174673441, Isfahan, Iran.

Table 1: Teacher participants training and teaching background

Groups	Experience Details		
	<i>EFL teaching</i>	<i>Assessment-related Teaching or training</i>	<i>ESP-related Teaching or training</i>
<i>G1. PhD candidate in TEFL (n=4)</i>	2 to 3 years	Three semesters training in language testing	Two semesters training in ESP
<i>G2. Assistant Professors (n=2)</i>	About 8 years	On average for 5 years teaching experience in language assessment	Two semesters training in ESP
<i>G3. EGP / ESP teachers (n=6)</i>	About 8 years	Three semesters training in language assessment	On average for 6 years ESP/EGP teaching experience

**Background of the Study:** Previously, a lot has been done on the nature of the General English and ESP tests and their underlying constructs [1-9]. What can be drawn from recent studies is that specific purpose language tests have the characteristics of both language knowledge and specific purpose content knowledge [1]. This kind of evidence relates only to the substantive aspect of validity indicated by Messick [10]. But, previous research conducted by some of the scholars in the field [11-13] showed that through justifying the dimensional structure of the test (as suggested by the title of the present study), evidence could be provided not only for substantive validity, but also for structural, generalizability (by tracking the invariance of factor structure across different forms of the same test) and consequential aspects of validity (practical usefulness of the test by providing diagnostic subscores). In addition, the dimensionality is also useful in identifying the major threats of construct validity, construct underrepresentation and construct-irrelevant variance [14]. Furthermore, confirmation of the internal structure provided information that can be used to make a decision concerning what scores should be reported or what setting cutscores can be made based on the test structures. This information supplied evidence of the structural aspects of validity. When the dimensions are distinguishable, reporting subscores is appropriate; when there is only one dominant dimension, one total score is reported [15].

The present study aims at investigating the dimensionality of the M.A specific purpose language test which measures both General and Academic purpose language abilities in a de-modularized model. To this end, two consecutive phases ordered the progress of the research of which the first one is a dimensionality decision phase and the second one is the model recovery.

### Methodology

**Participants:** For the first phase of the study, item content analysis was conducted by three groups of experts in language teaching and testing fields. Table 1 shows the teacher participants background and

experiences by group. For the statistical part of the phase, also, the real data was obtained from the administration of language sub- test of the 2010 version of M.A. entrance exams to a group of Accounting students. This group which was composed of 100 respondents was selected from among the students who were passing their last year of the B.A. course of study in Accounting field.

For the second phase, item responses from 1000 simulees to the 30 items were generated by the ConQuest [16] simulate command file whereby item parameters were generated with the Uniform distribution of (-2 : 2) and person parameters were simulated on a multivariate normal distribution of a given mean and variance matrix.

**Instruments:** In addition to the 2010 version of the operational M.A. entrance exam (Appendix A), ConQuest statistical software was used for its two utilities: one, generating the initial value of the parameters and the required design matrix which, then, were imported into software for the simulation purpose and two, calibrating the items to the proposed Rasch models and estimating and evaluating their fit indices.

ConQuest is one of the unique software which has as many different settings as there are psychometric constraints on the items in the test. The multidimensional random coefficients multinomial logit model (MRCMLM) which ConQuest can be fitted for enables the software to be set for the Rasch Sub-dimension model. Monte Carlo Sampler also adds to the program's utilities by augmenting the data obtained from any number of correlated sub-dimensions. Last but not least is the possibility of setting the program for within item multidimensionality, the capability which makes the software distinguished inter alia.

### Procedure

#### Substantive and Statistical Exploration of Test

**Dimensionality:** Many tests thought to be unidimensional are in fact measuring additional latent traits other than the primary trait of interest [17-21]. McKinley and Way [22], for example, in a technical report on TOEFL concluded that there are some important

secondary traits present in the test and that the model fit is enhanced if the items would be calibrated with some multidimensional model of Item Response Theory (IRT) instead of unidimensional models. In the present study, also, since abilities tapped were selected from the continuum of English for General Purposes-English for Specific Purposes (EGP-ESP) [23], at first, something must be known about the nature of the underlying traits measured in different components of the test. The following research question is answered in the first stage of the analysis.

- Is there any correspondence between content specification of the items and dimensional structure of the EGP-ESP test?

Since item content analysis has been introduced in the literature [24] as one of the ways of exploring the secondary abilities in the test, for this stage, participants

were given some kind of worksheets on which they analyzed the items in different sets. The only tools used by the participants in their item analysis were: Nation [25]'s Vocabulary-Profiler for determining the words frequency in vocabulary item sets, Purpura [26]'s coding scheme for grammar subsection, Weir *et al.* [27] parameters of English for Academic purpose (EAP) reading and Brown [28]'s item types for specialized and general reading test-lets. But, for the simplicity or multiplicity of the constructs in each component to be more meaningfully represented, two checklists were also developed and given to the participants in which items were coded in different categories based on the ability dimensions they tapped. Tables 2 and 3 show the categorization of each item which is based on the coding of that item in that category counting to more than 80% of the participants (remaining 20% coding in other categories were ignored).

Table 2: Participants' coding of the items in different categories based on the ability they tapped.

Items	English for Accounting			
	Content Factors			
	Linguistic Factors	Field related	Other fields	Common-core
1			✓	
2		✓		
3			✓	
4			✓	
5		✓		
6		✓		
7			✓	
8			✓	
9		✓		
10			✓	
11	✓			
12	✓			
13	✓			
14	✓			
15	✓			
16		✓		
17		✓		
18		✓		
19		✓		
20		✓		
21		✓		
22	✓			
23				✓
24				✓
25		✓		
26	✓			
27		✓		
28		✓		
29	✓			
30				✓

Table 3: Participants' cross joining of the items across categories

Tests of Participants		Cross-joining to the items across categories					
		English for Accounting					
Test Subsections		*(GE) structure	vocabulary	EGAP vocabulary	reading	ESAP vocabulary	reading
GE	Structure				16,17,18,19,20		22,29
	Vocabulary						
EGAP	Vocabulary				1,5,6		19,23,24,26,30
	Reading						
ESAP	Vocabulary				2,3,4,7,8,9, 10		
	Reading					16,17,18,20,21, 25, 27,28	

\* Note: GE stands for General English

EGAP stands for English for General Academic Purposes

ESAP stands for English for Specific Academic

According to Shealy and Stout [21], when items measure some secondary ability in addition to the primary trait of interest, item responses can be characterized as multidimensional. In the case of the concerned test in this study, participants' coding of the items in Table 3 shows that an overwhelming majority of the items in the Master's Degree EGP-ESP exam are not simple in their substantive structure. But to see whether substantively multidimensional test is also statistically multidimensional, the real data obtained from the administration of M.A exams in accounting field of study were submitted to ConQuest for dimensionality detection. The question that guides the process of this stage is as follows:

- Given multidimensional test data, does the multidimensional Rasch model provide better estimates than the unidimensional Rasch model?

The hypothesis was that upon detecting any misbehavior among items under unidimensional Rasch conditions, the hypothesis that this model of item response fits the data would be rejected and test should be analyzed under multidimensional conditions. 2-dimensional Rasch was also tried out to see whether, upon rejecting the unidimensional hypothesis, tests have had a 2-dimensional simple structure or some higher dimensional structure. Appendix B shows the results of the ConQuest analysis for testing the unidimensionality and two dimensionality of the real data sets. The criteria used for assessing the model-data-fit are Type I error rate, Infit and Outfit mean square (called also weighted and unweighted fit indices) and t statistic. For careful analysis, the chi square and deviance statistics for different models are presented in the following table.

Having in mind the critical values of MNSQ and t statistics, Appendix B shows that Rasch simple logistic model did not fit the data. To be more exact, half of the

Table 4: Rasch unidimensional and 2-dimensional fit indices for accounting ESP test, 2010 version

	1 Dimensional	2 Dimensional
2010		
Chi square	624.32 (df=24)	481.37(df=25)
Deviance	2296.20	2280.95

items in accounting 2010 showed Outfit MNSQ  $t > 2$  and about one third of them showed Infit MNSQ  $t > 2$ . The results also suggested that the Rasch two-dimensional model in which dimensions of the test matched the actual partitioning of the test into EGP and ESP components did not again fit the data. As to degree of deviance and chi square, Table 4 summarizes the results of analysis reported in Appendix B.

In this table, although the amount of deviance reported for unidimensional and 2 dimensional solutions are not significantly different from each other but the reading of about twice higher for chi square indices in 1 dimensional cells shows that the fit of 1 dimensional models is worse than the fit of 2 dimensional model. It follows that the hypothesis that unidimensional model fits these data as well as the two-dimensional model could be rejected.

After examining the indices provided at item and overall levels, the study entered its second phase which was designed to be a simulation study with the following research question and hypothesis (H):

- What is the effect of considering constrained higher dimensional models of IRT as opposed to the unconstrained multidimensional models on the estimate of person and item parameters in the EGP-ESP test?

H) Models which consider secondary abilities tapped along with the primary abilities without any constraints regarding the dependency of the items yield better

Table 5: Summary of the multidimensional IRT test structure for accounting field of study, 2010 version

	English for Accounting			
	3D			3DC4
Test Subsections	C1	C2	C3	Subdimension model
General English (GE)				
●Structure	1	1	1	1,2
●Vocabulary	1	1	1	1,2
English for General Academic Purposes (EGAP)				
●Vocabulary	2	3	2	1,2
●Reading	2	2	1	1,2
English for Specific Academic Purposes (ESAP)				
●Vocabulary	3	3	3	1,3
●Reading	3	3	3	1,3
				1,2,3

estimates of the test of test taker's achievement in each ability dimension than models which consider local item dependency in sub-dimensions as well as the correlation between sub-dimensions.

#### Calibration Phase

**Model and Parameters Recovery for Higher-Dimensional Solution:** As it became clear in the exploratory phase, the test which has been currently administered for the admission of the candidates into Master Degree course of study was neither unidimensional nor 2 dimensional in its underlying traits. So any calibration procedure must be designed in a higher dimensional structure, in this case, 3 dimensional solutions.

To test the research hypothesis (H), two types of dimensional configurations were designed as the framework for Monte Carlo simulation phase of the study: one, in which items were bundled differently according to the participants' cross joining of the abilities (Tables 2 and 3) without taking into account the existing test-let effects or local item dependency between items and, second, a configuration with the local item dependency and inter correlated sub-dimensions as a constraint. Table 5 is a summary of the proposed configurations. Note that Abbreviations of D and C in this table stand for Dimension and Configuration respectively.

In 3DC1, the first dimension is measured by the second test-let plus common core items, second bundle comprises items measuring linguistic factors and the last dimension is measured in remaining items. In 3DC2, ESAP components measure EGAP vocabulary, so item bundles are 1) items which measure linguistic factors, 2) common core items +vocabulary subsection items and 3) remaining items. Configuration 3 in 3D assemble items in the bundles

in a way that EGAP reading component measures also the structure and vocabulary of GE; in 3DC3, therefore, common core items plus items number 22, 26 and 30 comprises the first bundle, cloze items are bundled as a distinct dimension and items measuring content factors comprises the third bundle.

Unlike 3D configurations where not the local item dependency nor the correlation between dimensions was considered, 3DC4 configuration following what Brandt [29, 30] proposed in his Rasch sub-dimension model, accounts for both of the above mentioned constraints by, firstly, making all the items in different components load on one common general factor, i.e. English for accounting and then, allowing a space for each "sub-dimension specific factor", i.e. GE, EGAP, or ESAP to be estimated in correlation with other sub-dimension factors in the test (ibid).

**Evaluating Cluster Solutions:** To evaluate the various MIRT structures, the parameter estimates and data-model fit of the different MIRT solutions were compared to each other. But, because of the existence of over and underfitting items in real data set and automatic removal of these parameters from ConQuest analysis, observed items and ability distribution of real data set could not be relied on any more for designing the simulation study and in the calibration phase, therefore, study should continue with the prior distribution parametrization which was constrained by the researcher [31]. A sample of the ConQuest syntax for generating 3-dimensional item response datasets of 1000 simulees answering 30 multiple choice questions are presented below. Note that for each configuration, a new data set was generated so that the number of items in each item-to-dimension bundling matched the number of items in the original test already suggested by the content analysts for that configuration.

```
simulate !nitems=15:9:6, npersons=1000, maxscore=1,
itemdist=niform(-2:2), abilitydist=mvnormal(0.5:1:1:0:1:0:
0.7:0.8),
method=montecarlo, nodes=2000;
```

For the purpose of the present study, with the exception of 3DC4 (sub-dimension model) where some additional model constraints were applied, scoring matrix was used for defining different item-to-dimension solutions [29, 30]. Appendix C shows the results obtained from ConQuest analysis for each solution. For this stage, the deviance statistics of alternative models were compared to provide a formal statistical test of the relative

Table 6: Comparison of the deviance statistics in different model configurations

	Deviance	Difference
3DC2	35600.27	
3DC1	35528.04	72.23
3DC3	34818.36	709.68
3DC4 (sub-dimension model)	33557.92	1260.44

fit of models. Table 6 shows the deviance statistics of different configurations and the differences between them as follows:

As it is clear from the table, the degree of improvement of the deviance between different configurations increases moving from 3DC2 to 3DC4 with the most significant one reported between 3DC3 and sub-dimension model 3DC4. It can be concluded that among 3D solutions, 3DC4 introduces the best item-to-dimension configuration for optimizing the joint EGP-ESP test both substantively and statistically. But, for the exact differentiation of the proposed configurations and the evaluation of the significance of their effects on the ability estimation of individuals in each dimension, however, a Multivariate Analyses of Variance (MANOVA) was conducted and its results were reported as follows addressing the research hypothesis (H).

**Multivariate ANOVA to Test the Effect of Different Calibration Model on the Ability Estimation:** To test the research hypothesis on the effects of constrained model of IRT as to correlation between sub-dimensions and local item dependency within each dimension as opposed to unconstrained multidimensional models, a Multivariate ANOVA was conducted with four configurations of 3DC1 (abbreviated as *Config1*) to 3DC4 (abbreviated as *Config4*) as grouping variables and different dimensions of ability estimation (D1-D3) as dependent variables.

Box's M test shows that with  $F(18, 5.64) = 1041.49$  and  $p = 0.000$ , the covariance matrix of the dependent variables have been nearly similar and so, the MANOVA can continue with the table of multivariate test.

To see whether the effects of different configurations on the ability estimates have been significant or not, the results of F test are reported in the following table as follows:  $F(9, 9720) = 29.27$ ,  $p < 0.0005$ , effect size = 0.02.

Wilks' Lambada statistics presented in this table show that the effects of different independent variables (Config1-Config4) on the dependent variables (D1-D3) have been significant with each independent variable accounted for 0.02 of the total variance. To identify similarities and differences between the reported effects of different configurations, table of the multiple comparisons (Table 9) gives the mean differences of the dependent variables as well as the significance of this difference for each dependent variable. Asterisk is printed next to differences which are significant at the 0.05 level or better.

Quite consistent with the reported results in the previous section, in all of the categories of dependent variables, the effect of configuration 4 has been more significant than other three configurations. Scheffe Post hoc test shows the same results too.

These results in turn reject the H that predicted a better effect of 3DC1-3DC3 on the ability estimation of the individuals than Rasch sub-dimension model (3DC4).

### Summary of the Results:

Findings of the present study indicated that:

(In the first phase)

- An overwhelming majority of the items in the recently administered Master's Degree EGP-ESP exams are not simple in their structure psychometrically and psychologically.

Table 7: Multivariate Tests<sup>d</sup>

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power <sup>b</sup>
Intercept	Pillai's Trace	.022	29.545 <sup>a</sup>	3.000	3994.000	.000	.022	88.634	1.000
	Wilks' Lambda	.978	29.545 <sup>a</sup>	3.000	3994.000	.000	.022	88.634	1.000
	Hotelling's Trace	.022	29.545 <sup>a</sup>	3.000	3994.000	.000	.022	88.634	1.000
	Roy's Largest Root	.022	29.545 <sup>a</sup>	3.000	3994.000	.000	.022	88.634	1.000
grouping	Pillai's Trace	.063	28.578	9.000	11988.000	.000	.021	257.204	1.000
	Wilks' Lambda	.937	29.274	9.000	9720.497	.000	.021	213.192	1.000
	Hotelling's Trace	.067	29.834	9.000	11978.000	.000	.022	268.504	1.000
	Roy's Largest Root	.067	89.557 <sup>c</sup>	3.000	3996.000	.000	.063	268.672	1.000

a. Exact statistic

b. Computed using alpha = .05

c. The statistic is an upper bound on F that yields a lower bound on the significance level.

d. Design: Intercept + grouping

Table 8: Pair-wise Comparisons

Dependent Variable	(I) grouping	(J) grouping	Mean Difference (I-J)	Std. Error	Sig. <sup>a</sup>	95% Confidence Interval for Difference <sup>a</sup>	
						Lower Bound	Upper Bound
D3	Config1	Config2	.002	.021	.925	-.040	.044
	Config3		-.002	.021	.942	.040	
	Config4		-.049*	.021	.022	-.007	
	Config2	Config1	-.002	.021	.925	-.044	.040
	Config3		-.004	.021	.868	.038	
	Config4		-.051*	.021	.017	-.009	
	Config3	Config1	.002	.021	.942	-.040	.043
	Config2		.004	.021	.868	.045	
	Config4		-.047*	.021	.026	-.006	
	Config4	Config1	.049*	.021	.022	.007	.091
	Config2		.051*	.017	.009	.093	
	Config3		.047*	.021	.006	.089	
D1	Config1	Config2	9.191E-5	.033	.998	-.064	.064
	Config3		-.005	.033	.885	.060	
	Config4		-.225*	.033	.000	-.160	
	Config2	Config1	-9.191E-5	.033	.998	-.064	.064
	Config3		-.005	.033	.882	.060	
	Config4		-.225*	.033	.000	-.160	
	Config3	Config1	.005	.033	.885	-.060	.069
	Config2		.005	.033	.882	.069	
	Config4		-.220*	.033	.000	-.156	
	Config4	Config1	.225*	.033	.000	.160	.289
	Config2		.225*	.033	.000	.289	
	Config3		.220*	.033	.000	.284	
D2	Config1	Config2	.000	.028	.991	-.055	.055
	Config3		-.003	.028	.904	.052	
	Config4		.103*	.028	.000	.158	
	Config2	Config1	.000	.028	.991	-.055	.055
	Config3		-.003	.028	.912	.052	
	Config4		.103*	.028	.000	.158	
	Config3	Config1	.003	.028	.904	-.052	.058
	Config2		.003	.028	.912	.058	
	Config4		.107*	.028	.000	.162	
	Config4	Config1	-.103*	.028	.000	-.158	-.048
	Config2		-.103*	.028	.000	-.048	
	Config3		-.107*	.028	.000	-.052	

Based on estimated marginal means

a. Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).

\*. The mean difference is significant at the .05 level.

Table 9: Multiple Comparisons

						95% Confidence Interval		
Dependent Variable	(I) grouping	(J) grouping	Mean Difference (I-J)	Std. Error	Sig. <sup>a</sup>	Lower Bound	Upper Bound	
D1	Tukey HSD	C1	C2	.0001	.03284	1.000	-.0843	.0845
		C3		-.0048	.03284	.999	.0796	
		c4		-.2247*	.03284	.000	-.1403	
		C2	C1	.0000	.03284	1.000	-.0845	.0843
		C3		-.0049	.03284	.999	.0796	
		c4		-.2248*	.03284	.000	-.1404	
		C3	C1	.0048	.03284	.999	-.0796	.0892
		C2		.0049	.03284	.999	-.0796	.0893
		c4		-.2200*	.03284	.000	-.1356	
		c4	C1	.2247*	.03284	.000	.1403	.3091
		C2		.2248*	.03284	.000	.3092	
		C3		.2200*	.03284	.000	.3044	

Table 9: Continue

						95% Confidence Interval	
Dependent Variable	(I) grouping	(J) grouping	Mean Difference (I-J)	Std. Error	Sig. <sup>a</sup>	Lower Bound	Upper Bound
Scheffe	C1	C2	.0001	.03284	1.000	-.0918	.0919
	C3		-.0048	.03284	.999	-.0966	.0871
	c4		-.2247*	.03284	.000	-.3166	-.1329
	C2	C1	.0000	.03284	1.000	-.0919	.0918
	C3		-.0049	.03284	.999	-.0967	.0870
	c4		-.2248*	.03284	.000	-.3167	-.1330
	C3	C1	.0048	.03284	.999	-.0871	.0966
	C2		.0049	.03284	.999	-.0870	.0967
	c4		-.2200*	.03284	.000	-.3118	-.1281
	c4	C1	.2247*	.03284	.000	.1329	.3166
	C2		.2248*	.03284	.000	.1330	.3167
	C3		.2200*	.03284	.000	.1281	.3118
D2 Tukey HSD	C1	C2	-.0003	.02805	1.000	-.0724	.0718
	C3		-.0034	.02805	.999	-.0755	.0687
	c4		.1032*	.02805	.001	.0311	.1753
	C2	C1	.0003	.02805	1.000	-.0718	.0724
	C3		-.0031	.02805	1.000	-.0752	.0690
	c4		.1035*	.02805	.001	.0314	.1756
	C3	C1	.0034	.02805	.999	-.0687	.0755
	C2		.0031	.02805	1.000	-.0690	.0752
	c4		.1066*	.02805	.001	.0345	.1786
	c4	C1	-.1032*	.02805	.001	-.1753	-.0311
	C2		-.1035*	.02805	.001	-.1756	-.0314
	C3		-.1066*	.02805	.001	-.1786	-.0345
Scheffe	C1	C2	-.0003	.02805	1.000	-.0788	.0781
	C3		-.0034	.02805	1.000	-.0818	.0751
	c4		.1032*	.02805	.004	.0247	.1816
	C2	C1	.0003	.02805	1.000	-.0781	.0788
	C3		-.0031	.02805	1.000	-.0815	.0754
	c4		.1035*	.02805	.004	.0250	.1819
	C3	C1	.0034	.02805	1.000	-.0751	.0818
	C2		.0031	.02805	1.000	-.0754	.0815
	c4		.1066*	.02805	.002	.0281	.1850
	c4	C1	-.1032*	.02805	.004	-.1816	-.0247
	C2		-.1035*	.02805	.004	-.1819	-.0250
	C3		-.1066*	.02805	.002	-.1850	-.0281
D3 Tukey HSD	C1	C2	.0020	.02133	1.000	-.0528	.0568
	C3		-.0015	.02133	1.000	-.0564	.0533
	c4		-.0489	.02133	.100	-.1037	.0059
	C2	C1	-.0020	.02133	1.000	-.0568	.0528
	C3		-.0035	.02133	.998	-.0584	.0513
	c4		-.0509	.02133	.080	-.1057	.0039
	C3	C1	.0015	.02133	1.000	-.0533	.0564
	C2		.0035	.02133	.998	-.0513	.0584
	c4		-.0474	.02133	.118	-.1022	.0075
	c4	C1	.0489	.02133	.100	-.0059	.1037
	C2		.0509	.02133	.080	-.0039	.1057
	C3		.0474	.02133	.118	-.0075	.1022
Scheffe	C1	C2	.0020	.02133	1.000	-.0577	.0617
	C3		-.0015	.02133	1.000	-.0612	.0581
	c4		-.0489	.02133	.154	-.1086	.0108
	C2	C1	-.0020	.02133	1.000	-.0617	.0577
	C3		-.0035	.02133	.999	-.0632	.0561



Table 9: Continue

Dependent Variable	(I) grouping	(J) grouping	Mean Difference (I-J)	Std. Error	Sig. <sup>a</sup>	95% Confidence Interval	
						Lower Bound	Upper Bound
c4		-.0509	.02133	.128	-.1106	.0088	
C3		C1	.0015	.02133	1.000	-.0581	.0612
C2		.0035	.02133	.999	-.0561	.0632	
c4		-.0474	.02133	.177	-.1070	.0123	
c4		C1	.0489	.02133	.154	-.0108	.1086
C2		.0509	.02133	.128	-.0088	.1106	
C3		.0474	.02133	.177	-.0123	.1070	

Based on observed means.

The error term is Mean Square(Error) = .228.

\*. The mean difference is significant at the .05 level.

- These tests violate the assumption of unidimensionality by the inclusion of items which: a) measure some primary and secondary abilities at the same time and b) depend for their answers on the same stimulus.
- Primary and secondary abilities are selected from the specificity spectrum of General-Specific purpose language continuum including: a) specific purpose content knowledge components, b) GE and EGAP ability factors and C) ESAP ability factors.
- In the current versions items are misclassified in terms of the dimension they tapped.
- Items displayed misbehavior under the unidimensional as well as the 2- dimensional conditions of Rasch model.

(In the second phase)

- Overall and item level fit indices indicated that among different 3- dimensional item-to-dimension model structures, Rasch sub-dimension model (3DC4) which constrained the item assembly as to local item dependency and correlation between sub-dimension factors, best fitted the simulated item response patterns.
- Formal test of the hypothesis showed that among four proposed configurations, this model (3DC4) had a significant effect on the ability estimation in all its three dimensions ( $P < 0.001$ ) and therefore its item –to dimension configuration could be used as a framework for upgrading the joint EGAP-ESAP tests item assembly.

## DISCUSSION AND CONCLUSIONS

As described in Tables 2 and 3, almost in all of the sub-components of the concerned test, expert judges pointed to the non-simple structure (more than one latent trait) behind item clusters; this result provided some evidence on the substantive multidimensionality of EGAP-ESAP test data and set the ground for making a

multidimensionality hypothesis which was later tested empirically in a second stage of ConQuest Rasch analysis. On the other hand, a simple two dimensional structure corresponding to the original partitioning of the test into EGAP and ESAP components was rejected in a later Rasch analysis. This stemmed from the overlapping nature of the underlying factors in each of these components and the fact that much of the variance in one component (ESAP) could be accounted for by the other component (EGAP) [28]. The non-zero correlation between different dimensions further complicated the issue and bolstered the idea of cross-joining of the abilities into different sub-dimensions, hence proposing some higher dimensional model of item responses for the test, something that had already been suggested implicitly by the content analysts in their item –to- dimension mapping.

Question b1 was answered after various conditions of 3-dimensionality were simulated with different item-to-dimension mapping patterns. The results of both informal and formal tests of the hypothesis showed that among 4 different 3D configurations, the last model in which items were assembled based on the constraints of Rasch sub-dimension model was the best model for recovering item and persons parameters of the joint EGAP-ESAP test.

Recovering the 3-dimensional sub-dimension model as the optimal model of item responses brings to the fore the constraints (of both quantitative and qualitative nature) and cautions which should be accounted for in assembling a joint EGAP-ESAP test. The constraints are:

- Non-zero correlation between dimension specific factors
- Zero covariance between sub-dimension factors and general factor
- Setting the mean of the ability estimation in general dimension to zero [29].

In upgrading the test assembly sub-layers of concerned test, the first constraint should be met by the selection of dimension specific traits from ELT continuum [23] which runs from General English to very specific ESP components. The second constraint should be met by considering the English for accounting (specific purpose background knowledge) as the general factor and language knowledge factors (general English and common core language elements as the first dimension specific factor and knowledge of the rhetorical functions as the second dimension specific factor). This selection can be justified by the results of a regression analysis conducted by Salmani-Nodoushan on the ESP reading tests [32] which showed that there is a lack of multi-colinearity between specific purpose content knowledge and proficiency. So, the content knowledge is a good option, inter alia, to act as the main dimension in the present situation.

In addition to the above mentioned constraints, one of the primary assumptions of the Rasch sub-dimension model is that the test would comprise up to four test-lets. In upgrading the present ESP test, therefore, the pre-test items can have item assembly sub-layer like this a) a cloze test of general and common core language components, b) a reading test-let which tap pragmatic inference ability and knowledge of rhetorical functions and c) a reading test-let with composite items of both general and specific components; stimulus texts should also be selected in a way that they have no overlapping content, although they must all be from the texts students may encounter in accounting courses with different degrees of difficulty.

**IMPLICATIONS** In this study, Rasch model of analyses was used for joining the two measures which are usually designed, administrated and scored in a modular bases (e.g. general and academic IELTS). This capability of Rasch can be used in any other testing situation where testing agencies in the rush to implement comprehensive accountability systems may be trapped in the chaotic domain of multi-objective, multi-purpose tests.

## REFERENCES

1. Afghari, A. and M. Tavakoli, 2004. Learner variables and ESP test performance. *Iranian Journal of Applied Linguistics*, 7(1): 1-24.
2. Alderson, J.C. and A.H. Urquhart, 1983. *Reading in a foreign language*. Longman.
3. Alderson, J.C. and A.H. Urquhart, 1985a. The effect of students academic discipline on their performance on ESP reading tests. *Language Testing*, 2: 2.
4. Alderson, J.C. and A.H. Urquhart, 1985b. This test is unfair: I'm not an economist. In P.C. Hauptman, R. Le Blanc and M.B. Wesche (Eds.), *Second Language Performance Testing*. Ottawa: University of Ottawa Press.
5. Amiryousefi, M., H. Vahid Dastjerdi and M. Tavakoli, 2012. Iranian EAP students' reading strategy use and their beliefs about reading comprehension components: Toward an EAP reading model. *World Applied Sciences Journal*, 18(8).
6. Clapham, C., 1993. Is ESP testing justified? In D. Douglas & C. Chapelle (Eds.), *A new decade of language testing research* (pp: 257-271). Alexandria, VA: TESOL Publications.
7. Douglas, D., 2001. Language for Specific Purposes assessment criteria: Where do they come from? *Language Testing*, 18(2): 171-185.
8. Hudson, T., 1993. Testing the specificity of ESP reading skills. In D. Douglas & C. A. Chapelle (Eds.), *A new decade of Language testing research* (pp: 58-82). Arlington, VA: Teachers of English to Speakers of Other Languages.
9. Shabani Minabad, M. and F. Fallah Khoshkholgh, 2012. Investigating the effect of genre-based pedagogy on English for Specific Purpose learners' reading comprehension. *World Applied Sciences Journal*, 18(2).
10. Messick, S., 1989. Validity. In R. Linn (Ed.), *Educational measurement*, (3rd ed.). Washington, D.C.: American Council on Education.
11. Hattie, J.A., 1985. Methodology review: Assessing unidimensionality of tests and items. *Applied Psychological Measurement*, 9: 139-164.
12. Nunnally, J.C. and I.H. Bernstein, 1994. *Psychometric theory* (3<sup>rd</sup> ed.) New York: McGraw-Hill.
13. 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). New Jersey: Lawrence Erlbaum.
14. Baghaei, P., 2008. The Rasch model as a construct validation tool. *Rasch Measurement Transactions*, 22(1).
15. Haladyna, T.M., 2004. *Developing and validating multiple-choice test items*. Mahwah, NJ: Lawrence Erlbaum.
16. Wu, M., R. Adams and M. Wilson, 1998. *ACER Conquest: Generalised item response modelling software*. Melbourne: ACER Press.

17. Oshima, T.C. and M.D. Miller, 1992. Multidimensionality and item bias in item response theory. *Applied Psychological Measurement*, 16: 237-248.
18. Roussos, L.A. and W.F. Stout, 1996a. A multidimensionality-based DIF analysis paradigm. *Applied Psychological Measurement*, 20: 355-371.
19. Roussos, L.A. and W.F. Stout, 1996b. Simulation studies of the effects of small sample size and studied item parameters on SIBTEST and Mantel-Haenszel Type I error performance. *Journal of Educational Measurement*, 33: 215-230.
20. Russell, S.S., 2005. Estimates of Type I Error and power for indices of differential bundle and test functioning. *Dissertation Abstracts International*, (UMI No. 3175804).
21. Shealy, R. and W.F. Stout, 1993a. A model-based standardization approach that separates true bias/DIF from group differences and detects test bias/DIF as well as item bias/DIF. *Psychometrika*, 58: 159-194.
22. **Missing**
23. Dudley-Evans, T. and M.J. St John, 1998. *Developments in English for Specific Purposes* Cambridge: Cambridge University Press.
24. Gierl, M.J., X. Tan and C. Wang, 2005. Identifying content and cognitive dimensions on the SAT. (No. 2005-11). New York, NY: The College Board.
25. Nation, I.S.P. (2001, 2003) *Vocab Profile* (software downloaded 4 May 2001), <[http://www.vuow.ac.nz/nation\\_p/software.download](http://www.vuow.ac.nz/nation_p/software.download)>.
26. Purpura, J.E., 2004. *Assessing grammar*. Cambridge, UK: Cambridge University Press.
27. Weir, C.J., H. Yang and Y. Jin, 2000. An empirical investigation of the componentiality of L2 reading in English for academic purposes. Cambridge: Cambridge University Press.
28. Brown, J., 1988. Components of engineering-English reading ability. *System*, 16(2): 193-200.
29. Brandt, S., 2007a. Applications of a Rasch model with subdimensions. Paper presented at the 2007 annual conference of the American Educational Research Association (AERA), Chicago, IL.
30. Brandt, S., 2008. Estimation of a Rasch model including subdimensions. In M. von Davier & D. Haertel (Eds.), *IERI Monograph Series. Issues and Methodologies in Large-Scale Assessments* (Vol. 1, pp: 51-70). Princeton, NJ: IEA-ETS Research Institute.
31. Curtis, D., 2004. Person misfit in attitude surveys: Influences, impacts and implications. *International Education Journal*, 5(2): 125-144.
32. Salmani-Nodoushan, M., 2003. Text familiarity, reading tasks and ESP test performance: A study on Iranian LEP and non-LEP university students. *Reading Matrix: An International Online Journal*, 3(1).

## Appendix A

### Accounting (2010)

#### PART A: Vocabulary

**Directions:** Choose the word or phrase (1), (2), (3), or (4) that best completes each sentence. Then mark the correct choice on your answer sheet.

- 1- In late October, the prime minister..... an emergency meeting.  
1) deduced 2) summoned 3) collaborated 4) associated
- 2- The contract between the two companies will .....at the end of the year.  
1) expire 2) obstruct 3) extinguish 4) surrender
- 3- An elderly man has .....doctors at the hospital by living after he was officially declared dead.  
1) converted 2) conducted 3) corresponded 4) confounded
- 4- These reports are ..... to the many hours of research completed by this committee.  
1) motion 2) testimony 3) submission 4) proximity
- 5- The points the author makes are fine, but the whole essay lacks .....  
1) route 2) profile 3) solidarity 4) coherence

6- The rise in the interest rate had a direct .....on the company's profits.

- 1) domain 2) bearing 3) convergence 4) proportion

7- It is some researchers' ..... that exercise is more important than diet if you want to lose weight.

- 1) exhibition 2) designation 3) contention 4) undertaking

8- There is no ..... on students to take so many subjects in one semester.

- 1) impetus 2) momentum 3) affiliation 4) compulsion

9- The success of the project Is due to the —— amount of work that has gone into It.

- 1) tremendous 2) leading 3) celebrated 4) primary

10- Farmers are still a .....political force In France.

- 1) potent 2) plentiful 3) provisional 4) prognostic

#### **PART B: Cloze Test**

**Directions:** Read the following passage and decide which choice (1), (2), (3), or (4) best fits each space. Then mark the correct choice on your answer sheet.

Nature conservation means the protection of animals and plants in their natural homes. It means the preservation of (11) -----species that live with us on Earth, in the wild places and even in our gardens and homes. (12)----- the sensible use of all the Earth's resources. It requires the development (13) ----- a deep sense of responsibility for the long-term welfare of this planet. (14) ----- man's prehistoric ancestors mastered the use of fire, human beings have influenced the natural environment. In Africa the burning of forests 50,000 years ago contributed to the creation of the great grasslands there. But man's ability to destroy nature (15) ----- in the last 100 years. The pace of technological advance has been startling.

11- 1) great various 2) great variety of 3) the great variety of 4) the great various

12- 1) It involves 2) They involve

3) They are involved in 4) It is involved in

13- 1) in man who 2) of man who 3) which man of 4) in man of

14- 1) When 2) Until 3) Ever since 4) As long as

15- 1) increased enormous 2) had increased enormously

3) has increased enormously 4) had an enormously increase

#### **PART C: Reading Comprehension**

**Directions:** Read the following three passages and choose the best choice (1), (2), (3), or (4). Then mark it on your answer sheet

##### **PASSAGE 1:**

The year 2001 witnessed a series of financial information frauds involving Enron Corporation, auditing firm Arthur Andersen, the telecommunications company WorldCom, Qwest and Sunbeam, among other well-known corporations. These problems highlighted the need to review the effectiveness of accounting standards, auditing regulations and corporate governance principles. In some cases, management manipulated the figures shown in financial reports to indicate a better economic performance. In others, tax and regulatory incentives encouraged over-leveraging of companies and decisions to bear extraordinary and unjustified risk.

The Enron scandal deeply influenced the development of new regulations to improve the reliability of financial reporting and increased public awareness about the importance of having accounting standards that show the financial reality of companies and the objectivity and independence of auditing firms.

In addition to being the largest bankruptcy reorganization in American history, the Enron scandal undoubtedly is the biggest audit failure. The scandal caused the dissolution of Arthur Andersen, which at the time was one of the five largest accounting firms in the world. It involved a financial scandal of Enron Corporation and their auditors Arthur Andersen, which was revealed in late 2001. After a series of revelations involving irregular accounting procedures conducted throughout the 1990s, Enron filed for Chapter 11 bankruptcy protection in December 2001.

One consequence of these events was the passage of Sarbanes-Oxley Act in 2002, as a result of the first admission of fraudulent behavior made by Enron. The act significantly raises criminal penalties for securities fraud, for destroying, altering or fabricating records in federal investigations or any scheme or attempt to defraud shareholders.

16- What does the passage mainly discuss?

- 1) Auditing Regulations 2) Accounting Scandals
- 3) Extraordinary Risk 4) Accounting Standards

17- According to the passage, Enron Corporation, auditing firm Arthur Andersen, the telecommunications company WorldCom, Qwest and Sunbeam in 2001:

- 1) gave wrongful information intended to result in financial gain for them
- 2) involved themselves in economic activities and financial affairs
- 3) emphasized the need to reconsider the efficiency of accounting criteria
- 4) were found guilty and arrested for raising their financial gain

18- What does 'auditing regulations' in line 4 refer to?

- 1) records of a sequence of events in the inspection of an organization's accounts
- 2) rules related to the official inspection of an organization's accounts
- 3) records of a sequence of events from which a history may be reconstructed
- 4) rules related to events from which a history may be reconstructed

19- In some cases, management — -----the figures shown in financial reports to manifest a better economic accomplishment.

- 1) Proficiently adapts 2) skillfully operates
- 3) proficiently influences 4) skillfully controls

20- Which Item *Is* NOT mentioned in the passage among the facts that accounting standards can reveal?

- 1) independence of auditing corporations 2) financial reality of companies
- 3) over-leveraging of companies 4) objectivity of auditing firms

21- Which statement is not true about Enron scandal according to the passage?

- 1) It was the biggest bankruptcy reorganization in American history.
- 2) It brought about the disintegration of Arthur Andersen.
- 3) It was one of the five largest accounting firms in the world.
- 4) It certainly is the largest audit defeat.

22- What does 'It' in line 16 refer to?

- 1) Arthur Andersen dissolution 2) Arthur Andersen
- 3) the largest bankruptcy 4) the scandal

23- After a series of .....concerning unacceptable accounting measures taken throughout the 1990s, Enron filed Chapter 11 bankruptcy protection in December 2001.

- 1) proclamations 2) disclosures 3) prophecies 4) announcements

24- Sarbanes-Oxley Act in 2002 remarkably Induced criminal punishments for the following items EXCEPT for .....in federal Investigations or any scheme or attempt to deceive shareholders.

- 1) denying fraudulent behavior 2) inventing records
- 3) damaging documents 4) changing data

**PASSAGE 2:**

The general definition of an audit is an evaluation of a person, organization, system, process, project or product. Audits are performed to ascertain the validity and reliability of information; also to provide an assessment of systems internal control the goal of an audit is to express an opinion on the person / organization / system in question, under evaluation based on work done on a test basis. Due to practical constraints, an audit seeks to provide only reasonable assurance that the statements are free from material error. Hence, statistical sampling is often adopted in audits. In the case of financial audits, a set of financial statements are said to be true and fair when they are free of material misstatements - a concept influenced by both quantitative and qualitative factors.

Audit is a vital part of Accounting. Traditionally, audits were mainly associated with pining information about financial system and the financial records of a company or a business. However, recent auditing has begun to include other information about the system, such as information about environmental performance. As a result, there are now professions conducting environmental audits.

In financial accounting, an audit is an independent assessment of the fairness by which a company's financial statements are presented by its management. It is performed by competent, independent and objective person(s) known *as* auditors or accountants, who then issue an auditor's report based on the results of the audit.

Such systems must adhere to generally accepted *standards* set by governing bodies regulating businesses; these standards simply provide assurance for third parties or external users that such statements present a company's financial condition sad results of operations fairly.

25- What is the main idea of the above passage?

- 1) Providing an assessment of a system's external management
- 2) Expressing opinions on the person, organization or system in question
- 3) Assessing the validity and reliability of financial information
- 4) Evaluating people, organizations, systems, processes, projects or products

26- According to the passage, .....an audit seeks to provide only reasonable assurance that the statements are free from material error.

- 1) as a result of pragmatic restrictions 2) on account of sensible conditions 3) because of functional qualifications 4) due to realistic reductions

27- Why is statistical sampling often taken on in audits?

- 1) To guarantee that the statements are not liable to relevant mistakes.
- 2) To confirm the evaluation of a system's internal control.
- 3) To provide only reasonable assurances about material errors.
- 4) To question the validity and reliability of information.

28- Which statement is NOT normally coupled with traditional auditing?

- 1) environmental performance of corporation 2) financial systems of a company 3) monetary accomplishment of a firm
- 4) fiscal records of a business

29- What does 'it' in line 16 refer to?

- 1) fairness 2) a company 3) accounting 4) an audit

30- As the author concludes, the systems must generally .....accepted standards .....by governing bodies regulating businesses?

- 1) obey — rejected 2) disregard — regulate
- 3) violate — denied 4) follow - prescribed

## Accounting (2010)

### Unidimensional Analysis

ConQuest: Generalised Item Response Modelling Software Sat Sep 24 22:09 2011

#### SUMMARY OF THE ESTIMATION

Estimation method was: Gauss-Hermite Quadrature with 15 nodes

Assumed population distribution was: Gaussian

Constraint was: ITEMS

The Data File: G:\thesis\Ac89.txt

The format: response 1-30

The regression model:

Grouping Variables:

The item model: item

Sample size: 97

Final Deviance: 2296.20446

Total number of estimated parameters: 26

The number of iterations: 31

Termination criteria: Max iterations=1000, Parameter Change= 0.00010

Deviance Change= 0.00010

Iterations terminated because the deviance convergence criteria was reached

Random number generation seed: 1.00000

Number of nodes used when drawing PVs: 2000

Number of nodes used when computing fit: 200

Number of plausible values to draw: 5

Maximum number of iterations without a deviance improvement: 100

Maximum number of Newton steps in M-step: 10

Value for obtaining finite MLEs for zero/perfects: 0.30000

=>show parameters!table=2;

ConQuest: Generalised Item Response Modelling Software Sat Sep 24 22:09 2011

#### TABLES OF RESPONSE MODEL PARAMETER ESTIMATES

TERM 1: item

VARIABLES				UNWEIGHTED FIT			WEIGHTED FIT		
Item		ESTIMATE	ERROR^	MNSQ	CI	T	MNSQ	CI	T
2	2	-1.079	0.239	1.60	(0.72, 1.28)	3.6	1.45	(0.71, 1.29)	2.7
4	4	0.396	0.218	1.12	(0.72, 1.28)	0.9	1.11	(0.75, 1.25)	0.9
5	5	1.491	0.222	1.50	(0.72, 1.28)	3.1	1.22	(0.76, 1.24)	1.7
6	6	-0.613	0.229	1.73	(0.72, 1.28)	4.2	1.41	(0.73, 1.27)	2.7
7	7	0.073	0.220	0.82	(0.72, 1.28)	-1.3	0.87	(0.75, 1.25)	-1.0
8	8	0.332	0.218	1.02	(0.72, 1.28)	0.2	1.03	(0.75, 1.25)	0.3
9	9	1.835	0.226	1.17	(0.72, 1.28)	1.2	1.20	(0.75, 1.25)	1.5
10	10	0.715	0.217	0.80	(0.72, 1.28)	-1.5	0.90	(0.76, 1.24)	-0.8
11	11	-2.887	0.312	1.06	(0.72, 1.28)	0.4	1.18	(0.37, 1.63)	0.6
12	12	-0.396	0.225	0.99	(0.72, 1.28)	-0.0	1.02	(0.74, 1.26)	0.2
13	13	-1.429	0.250	1.37	(0.72, 1.28)	2.3	1.42	(0.68, 1.32)	2.3
14	14	-3.370	0.337	1.57	(0.72, 1.28)	3.4	1.10	(0.18, 1.82)	0.4

Continued

TERM 1: item									
VARIABLES				UNWEIGHTED FIT			WEIGHTED FIT		
Item		ESTIMATE	ERROR <sup>^</sup>	MNSQ	CI	T	MNSQ	CI	T
15	15	-0.911	0.235	0.60	(0.72, 1.28)	-3.3	0.81	(0.72, 1.28)	-1.4
16	16	0.337	0.218	0.61	(0.72, 1.28)	-3.2	0.75	(0.75, 1.25)	-2.1
17	17	0.910	0.218	0.67	(0.72, 1.28)	-2.6	0.77	(0.76, 1.24)	-2.1
18	18	0.401	0.218	1.14	(0.72, 1.28)	1.0	1.12	(0.75, 1.25)	1.0
20	20	1.429	0.221	0.61	(0.72, 1.28)	-3.1	0.77	(0.76, 1.24)	-2.0
21	21	0.719	0.217	0.71	(0.72, 1.28)	-2.2	0.84	(0.76, 1.24)	-1.4
22	22	0.782	0.217	0.60	(0.72, 1.28)	-3.2	0.74	(0.76, 1.24)	-2.3
23	23	1.428	0.221	0.51	(0.72, 1.28)	-4.1	0.67	(0.76, 1.24)	-3.0
25	25	1.909	0.227	1.00	(0.72, 1.28)	0.0	1.07	(0.75, 1.25)	0.6
26	26	0.972	0.218	0.82	(0.72, 1.28)	-1.3	0.93	(0.76, 1.24)	-0.6
27	27	0.334	0.218	0.98	(0.72, 1.28)	-0.1	1.02	(0.75, 1.25)	0.2
29	29	-0.685	0.230	0.93	(0.72, 1.28)	-0.4	1.07	(0.73, 1.27)	0.5
30	30	-2.693*	1.146	1.41	(0.72, 1.28)	2.6	1.03	(0.43, 1.57)	0.2

An asterisk next to a parameter estimate indicates that it is constrained

Separation Reliability = 0.970

Chi-square test of parameter equality = 624.32, df = 24, Sig Level = 0.000

<sup>^</sup> Quick standard errors have been used

## Two-Dimensional Analysis

ConQuest: Generalised Item Response Modelling Software Sat Sep 24 22:20 2011

### SUMMARY OF THE ESTIMATION

Estimation method was: MonteCarlo with 1000 nodes

Assumed population distribution was: Gaussian

Constraint was: CASES

The Data File: G:\thesis\Ac89.txt

The format: response 1-30

The regression model:

Grouping Variables:

The item model: item

Sample size: 97

Final Deviance: 2280.95611

Total number of estimated parameters: 28

The number of iterations: 147

Termination criteria: Max iterations=1000, Parameter Change= 0.00010

Deviance Change= 0.00010

Iterations terminated because the convergence criteria were reached

Random number generation seed: 1.00000

Number of nodes used when drawing PVs: 2000

Number of nodes used when computing fit: 200

Number of plausible values to draw: 5

Maximum number of iterations without a deviance improvement: 100

Maximum number of Newton steps in M-step: 10

Value for obtaining finite MLEs for zero/perfects: 0.30000

=>show parameters!table=2;



ConQuest: Generalised Item Response Modelling Software Sat Sep 24 22:20 2011

TABLES OF RESPONSE MODEL PARAMETER ESTIMATES

TERM 1: item									
VARIABLES				UNWEIGHTED FIT			WEIGHTED FIT		
Item		ESTIMATE	ERROR^	MNSQ	CI	T	MNSQ	CI	T
2	2	-1.812	0.281	1.38	(0.72, 1.28)	2.4	1.32	(0.72, 1.28)	2.1
4	4	-0.418	0.245	0.97	(0.72, 1.28)	-0.2	1.01	(0.78, 1.22)	0.1
5	5	0.609	0.251	1.34	(0.72, 1.28)	2.2	1.14	(0.77, 1.23)	1.2
6	6	-1.370	0.263	1.40	(0.72, 1.28)	2.5	1.28	(0.75, 1.25)	2.1
7	7	-0.721	0.248	0.74	(0.72, 1.28)	-1.9	0.83	(0.78, 1.22)	-1.5
8	8	-0.478	0.245	0.90	(0.72, 1.28)	-0.6	0.96	(0.78, 1.22)	-0.3
9	9	0.931	0.257	1.07	(0.72, 1.28)	0.5	1.13	(0.76, 1.24)	1.0
10	10	-0.119	0.244	0.73	(0.72, 1.28)	-2.0	0.83	(0.79, 1.21)	-1.6
11	11	-3.541	0.447	0.79	(0.72, 1.28)	-1.5	1.08	(0.36, 1.64)	0.3
12	12	-1.167	0.257	0.86	(0.72, 1.28)	-1.0	0.95	(0.76, 1.24)	-0.4
13	13	-2.149	0.300	1.28	(0.72, 1.28)	1.8	1.28	(0.68, 1.32)	1.6
14	14	-4.015	0.534	0.95	(0.72, 1.28)	-0.3	1.06	(0.18, 1.82)	0.3
15	15	-1.658	0.274	0.61	(0.72, 1.28)	-3.2	0.81	(0.73, 1.27)	-1.5
16	16	-0.607	0.274	0.76	(0.72, 1.28)	-1.8	0.87	(0.72, 1.28)	-0.9
17	17	0.063	0.273	0.78	(0.72, 1.28)	-1.6	0.86	(0.73, 1.27)	-1.1
18	18	-0.532	0.274	2.43	(0.72, 1.28)	7.2	1.30	(0.72, 1.28)	2.0
20	20	0.666	0.277	0.64	(0.72, 1.28)	-2.9	0.86	(0.73, 1.27)	-1.0
21	21	-0.119	0.244	0.68	(0.72, 1.28)	-2.5	0.80	(0.79, 1.21)	-2.0
22	22	-0.086	0.273	0.61	(0.72, 1.28)	-3.1	0.80	(0.73, 1.27)	-1.5
23	23	0.546	0.250	0.54	(0.72, 1.28)	-3.8	0.69	(0.77, 1.23)	-3.0
25	25	0.998	0.259	0.95	(0.72, 1.28)	-0.3	1.01	(0.75, 1.25)	0.1
26	26	0.137	0.273	0.89	(0.72, 1.28)	-0.7	1.03	(0.73, 1.27)	0.2
27	27	-0.607	0.274	1.21	(0.72, 1.28)	1.4	1.14	(0.72, 1.28)	1.0
29	29	-1.440	0.266	0.84	(0.72, 1.28)	-1.1	1.01	(0.75, 1.25)	0.1
30	30	-4.038	0.446	1.66	(0.72, 1.28)	3.9	1.01	(0.46, 1.54)	0.1

An asterisk next to a parameter estimate indicates that it is constrained

Separation Reliability = 0.956

Chi-square test of parameter equality = 481.37, df = 25, Sig Level = 0.000

<sup>^</sup> Quick standard errors have been used

## Appendix C

### ConQuest Analysis for 3- Dimensional Configurations

#### 3-DC1

ConQuest: Generalised Item Response Modelling Software Mon Sep 26 03:17 2011

#### SUMMARY OF THE ESTIMATION

Estimation method was: MonteCarlo with 2000 nodes

Assumed population distribution was: Gaussian

Constraint was: CASES

The Data File: G:\thesis\3DC1\SIMdat.txt

The format: response 1-30

The regression model:

Grouping Variables:

The item model: item

Sample size: 1000

Final Deviance: 35528.04535

Total number of estimated parameters: 36  
 The number of iterations: 171  
 Termination criteria: Max iterations=1000, Parameter Change= 0.00010  
 Deviance Change= 0.00010  
 Iterations terminated because the convergence criteria were reached  
 Random number generation seed: 1.00000  
 Number of nodes used when drawing PVs: 2000  
 Number of nodes used when computing fit: 200  
 Number of plausible values to draw: 5  
 Maximum number of iterations without a deviance improvement: 100  
 Maximum number of Newton steps in M-step: 10  
 Value for obtaining finite MLEs for zero/perfects: 0.30000  
 =>show parameters!table=2;

ConQuest: Generalised Item Response Modelling Software Mon Sep 26 03:18 2011  
 IMPORTED MODEL: G:\thesis\3DC1\conquest analysis\designmatrix.des

Parameter Estimates

VARIABLES	ESTIMATE	ERROR <sup>^</sup>	UNWEIGHTED FIT			WEIGHTED FIT		
			MNSQ	CI	T	MNSQ	CI	T
Parameter 1	-1.48449	0.08119	1.05	(0.91, 1.09)	1.2	1.02	(0.91, 1.09)	0.5
Parameter 2	-0.55246	0.06733	1.06	(0.91, 1.09)	1.3	1.05	(0.96, 1.04)	2.2
Parameter 3	-1.13597	0.07435	1.06	(0.91, 1.09)	1.3	1.03	(0.93, 1.07)	1.0
Parameter 4	-0.14543	0.06538	1.05	(0.91, 1.09)	1.0	1.04	(0.97, 1.03)	2.7
Parameter 5	-1.03855	0.07281	1.05	(0.91, 1.09)	1.1	1.03	(0.93, 1.07)	0.9
Parameter 6	0.66187	0.06828	1.06	(0.91, 1.09)	1.4	1.05	(0.95, 1.05)	2.0
Parameter 7	0.43065	0.06652	1.05	(0.91, 1.09)	1.0	1.04	(0.96, 1.04)	2.0
Parameter 8	-0.55699	0.06737	0.96	(0.91, 1.09)	-0.9	0.97	(0.96, 1.04)	-1.4
Parameter 9	0.23879	0.06563	0.97	(0.91, 1.09)	-0.6	0.98	(0.97, 1.03)	-1.4
Parameter 10	-1.21476	0.07571	0.96	(0.91, 1.09)	-0.9	0.98	(0.92, 1.08)	-0.5
Parameter 11	-1.47309	0.08099	1.01	(0.91, 1.09)	0.3	1.00	(0.91, 1.09)	-0.0
Parameter 12	-2.01695	0.09425	1.02	(0.91, 1.09)	0.4	1.01	(0.88, 1.12)	0.1
Parameter 13	-0.48392	0.06883	1.04	(0.91, 1.09)	1.0	1.03	(0.95, 1.05)	1.2
Parameter 14	0.39164	0.06847	1.00	(0.91, 1.09)	0.0	0.99	(0.95, 1.05)	-0.3
Parameter 15	-1.37693	0.07918	0.98	(0.91, 1.09)	-0.5	0.99	(0.92, 1.08)	-0.2
Parameter 16	-0.80058	0.06968	0.96	(0.91, 1.09)	-0.9	0.97	(0.95, 1.05)	-1.1
Parameter 17	-0.00051	0.06524	0.97	(0.91, 1.09)	-0.6	0.98	(0.97, 1.03)	-1.6
Parameter 18	1.51750	0.08199	0.97	(0.91, 1.09)	-0.7	0.98	(0.90, 1.10)	-0.5
Parameter 19	0.52893	0.06718	0.94	(0.91, 1.09)	-1.5	0.95	(0.96, 1.04)	-2.5
Parameter 20	0.96990	0.07185	0.99	(0.91, 1.09)	-0.2	0.99	(0.94, 1.06)	-0.2
Parameter 21	0.68057	0.06846	0.95	(0.91, 1.09)	-1.2	0.96	(0.95, 1.05)	-1.9
Parameter 22	1.33594	0.07873	0.96	(0.91, 1.09)	-0.9	0.98	(0.92, 1.08)	-0.4
Parameter 23	0.50981	0.07088	1.00	(0.91, 1.09)	0.1	1.00	(0.94, 1.06)	0.2
Parameter 24	-1.77870	0.08754	1.05	(0.91, 1.09)	1.1	1.01	(0.90, 1.10)	0.1
Parameter 25	-1.36954	0.07870	0.95	(0.91, 1.09)	-1.1	0.98	(0.91, 1.09)	-0.4
Parameter 26	-0.54633	0.07098	1.02	(0.91, 1.09)	0.5	1.00	(0.95, 1.05)	-0.0
Parameter 27	-1.12582	0.07629	0.96	(0.91, 1.09)	-0.8	0.97	(0.93, 1.07)	-1.0
Parameter 28	-0.03719	0.06944	1.04	(0.91, 1.09)	0.9	1.03	(0.95, 1.05)	1.3
Parameter 29	-0.02755	0.06943	0.98	(0.91, 1.09)	-0.5	0.99	(0.95, 1.05)	-0.6
Parameter 30	-0.18692	0.06960	1.03	(0.91, 1.09)	0.7	1.02	(0.95, 1.05)	0.7

An asterisk next to a parameter estimate indicates that it is constrained

<sup>^</sup> Quick standard errors have been used

**3-DC2**

ConQuest: Generalised Item Response Modelling Software Mon Sep 26 04:00 2011

**SUMMARY OF THE ESTIMATION**

Estimation method was: MonteCarlo with 2000 nodes

Assumed population distribution was: Gaussian

Constraint was: CASES

The Data File: G:\thesis\3DC2\SIMdat.txt

The format: response 1-30

The regression model:

Grouping Variables:

The item model: item

Sample size: 1000

Final Deviance: 35600.27522

Total number of estimated parameters: 36

The number of iterations: 222

Termination criteria: Max iterations=1000, Parameter Change= 0.00010

Deviance Change= 0.00010

Iterations terminated because the convergence criteria were reached

Random number generation seed: 1.00000

Number of nodes used when drawing PVs: 2000

Number of nodes used when computing fit: 200

Number of plausible values to draw: 5

Maximum number of iterations without a deviance improvement: 100

Maximum number of Newton steps in M-step: 10

Value for obtaining finite MLEs for zero/perfects: 0.30000

=&gt;show parameters!table=2;

ConQuest: Generalised Item Response Modelling Software Mon Sep 26 04:00 2011

IMPORTED MODEL: G:\thesis\3DC2\des.des

**Parameter Estimates**

VARIABLES	ESTIMATE	ERROR <sup>^</sup>	UNWEIGHTED FIT			WEIGHTED FIT		
			MNSQ	CI	T	MNSQ	CI	T
Parameter 1	-0.93304	0.07156	1.00	(0.91, 1.09)	-0.0	1.01	(0.94, 1.06)	0.2
Parameter 2	0.23191	0.06606	0.98	(0.91, 1.09)	-0.4	0.99	(0.97, 1.03)	-0.7
Parameter 3	-0.84728	0.07051	0.98	(0.91, 1.09)	-0.5	0.98	(0.94, 1.06)	-0.6
Parameter 4	0.23191	0.06606	0.99	(0.91, 1.09)	-0.3	0.99	(0.97, 1.03)	-0.7
Parameter 5	-0.57057	0.06784	0.99	(0.91, 1.09)	-0.3	0.99	(0.96, 1.04)	-0.3
Parameter 6	0.01530	0.06569	1.00	(0.91, 1.09)	0.0	1.00	(0.97, 1.03)	0.1
Parameter 7	-0.23144	0.06603	0.98	(0.91, 1.09)	-0.5	0.98	(0.96, 1.04)	-0.9
Parameter 8	-1.38817	0.07910	0.99	(0.91, 1.09)	-0.2	1.00	(0.91, 1.09)	0.0
Parameter 9	-1.65714	0.08524	1.01	(0.91, 1.09)	0.2	1.02	(0.90, 1.10)	0.4
Parameter 10	0.28441	0.06624	1.00	(0.91, 1.09)	0.1	1.00	(0.96, 1.04)	0.2
Parameter 11	-1.95508	0.09238	0.93	(0.91, 1.09)	-1.5	0.97	(0.89, 1.11)	-0.5
Parameter 12	-0.30231	0.06847	0.95	(0.91, 1.09)	-1.1	0.96	(0.95, 1.05)	-1.9
Parameter 13	-1.67932	0.08551	0.92	(0.91, 1.09)	-1.8	0.97	(0.90, 1.10)	-0.6
Parameter 14	-0.79951	0.07170	0.97	(0.91, 1.09)	-0.6	0.99	(0.94, 1.06)	-0.2
Parameter 15	-2.99627	0.13331	0.89	(0.91, 1.09)	-2.5	1.00	(0.79, 1.21)	0.0
Parameter 16	-0.78311	0.07154	0.96	(0.91, 1.09)	-1.0	0.97	(0.94, 1.06)	-1.1
Parameter 17	-1.69245	0.08582	0.98	(0.91, 1.09)	-0.5	0.98	(0.90, 1.10)	-0.5
Parameter 18	-0.12011	0.06801	1.02	(0.91, 1.09)	0.4	1.02	(0.96, 1.04)	0.8

Continued

## Parameter Estimates

VARIABLES	ESTIMATE	ERROR <sup>^</sup>	UNWEIGHTED FIT			WEIGHTED FIT		
			MNSQ	CI	T	MNSQ	CI	T
Parameter 19	0.30608	0.06846	0.99	(0.91, 1.09)	-0.2	1.00	(0.95, 1.05)	-0.1
Parameter 20	0.18487	0.06812	1.00	(0.91, 1.09)	0.1	1.01	(0.96, 1.04)	0.2
Parameter 21	0.19415	0.06814	1.00	(0.91, 1.09)	0.1	1.00	(0.96, 1.04)	0.0
Parameter 22	-0.97884	0.07363	1.04	(0.91, 1.09)	0.8	1.02	(0.94, 1.06)	0.7
Parameter 23	0.34165	0.06648	1.02	(0.91, 1.09)	0.5	1.02	(0.96, 1.04)	1.0
Parameter 24	-0.52022	0.06747	1.02	(0.91, 1.09)	0.5	1.02	(0.96, 1.04)	1.0
Parameter 25	0.13389	0.06803	1.05	(0.91, 1.09)	1.0	1.04	(0.96, 1.04)	1.6
Parameter 26	-0.00072	0.06794	1.06	(0.91, 1.09)	1.3	1.05	(0.96, 1.04)	2.0
Parameter 27	-1.29268	0.07804	1.06	(0.91, 1.09)	1.3	1.02	(0.92, 1.08)	0.5
Parameter 28	0.86027	0.07228	1.02	(0.91, 1.09)	0.4	1.01	(0.94, 1.06)	0.4
Parameter 29	0.28678	0.06841	1.02	(0.91, 1.09)	0.4	1.02	(0.95, 1.05)	0.8
Parameter 30	1.42318	0.07994	1.03	(0.91, 1.09)	0.7	1.02	(0.91, 1.09)	0.4

An asterisk next to a parameter estimate indicates that it is constrained

<sup>^</sup> Quick standard errors have been used**3-DC3**

ConQuest: Generalised Item Response Modelling Software Mon Sep 26 04:32 2011

## SUMMARY OF THE ESTIMATION

Estimation method was: MonteCarlo with 2000 nodes

Assumed population distribution was: Gaussian

Constraint was: CASES

The Data File: G:\thesis\3DC3\SIMdat.txt

The format: response 1-30

The regression model:

Grouping Variables:

The item model: item

Sample size: 1000

Final Deviance: 34818.36246

Total number of estimated parameters: 36

The number of iterations: 236

Termination criteria: Max iterations=1000, Parameter Change= 0.00010

Deviance Change= 0.00010

Iterations terminated because the convergence criteria were reached

Random number generation seed: 1.00000

Number of nodes used when drawing PVs: 2000

Number of nodes used when computing fit: 200

Number of plausible values to draw: 5

Maximum number of iterations without a deviance improvement: 100

Maximum number of Newton steps in M-step: 10

Value for obtaining finite MLEs for zero/perfects: 0.30000

=&gt;show parameters!table=2;

ConQuest: Generalised Item Response Modelling Software Mon Sep 26 04:33 2011

IMPORTED MODEL: G:\thesis\3DC3\des.des

Parameter Estimates

VARIABLES	ESTIMATE	ERROR <sup>^</sup>	UNWEIGHTED FIT			WEIGHTED FIT		
			MNSQ	CI	T	MNSQ	CI	T
Parameter 1	0.32068	0.06618	1.10	(0.91, 1.09)	2.1	1.08	(0.96, 1.04)	4.4
Parameter 2	0.51641	0.06730	1.13	(0.91, 1.09)	2.7	1.10	(0.96, 1.04)	4.6
Parameter 3	-2.01756	0.09598	1.18	(0.91, 1.09)	3.8	1.04	(0.87, 1.13)	0.6
Parameter 4	-0.73679	0.06917	1.12	(0.91, 1.09)	2.5	1.08	(0.95, 1.05)	3.2
Parameter 5	-0.60987	0.06799	1.12	(0.91, 1.09)	2.5	1.09	(0.95, 1.05)	3.7
Parameter 6	-0.70345	0.06884	1.11	(0.91, 1.09)	2.3	1.08	(0.95, 1.05)	3.2
Parameter 7	-0.70345	0.06884	0.99	(0.91, 1.09)	-0.3	0.99	(0.95, 1.05)	-0.4
Parameter 8	-2.22564	0.10342	0.97	(0.91, 1.09)	-0.6	0.99	(0.85, 1.15)	-0.1
Parameter 9	-0.08899	0.06553	1.00	(0.91, 1.09)	-0.1	1.00	(0.97, 1.03)	-0.2
Parameter 10	-0.03321	0.06549	0.99	(0.91, 1.09)	-0.2	0.99	(0.97, 1.03)	-0.6
Parameter 11	-1.53100	0.08247	1.08	(0.91, 1.09)	1.8	1.05	(0.91, 1.09)	1.1
Parameter 12	-0.83044	0.07301	0.95	(0.91, 1.09)	-1.2	0.98	(0.94, 1.06)	-0.8
Parameter 13	-1.13619	0.07634	1.03	(0.91, 1.09)	0.7	1.03	(0.93, 1.07)	0.7
Parameter 14	-1.68602	0.08550	0.99	(0.91, 1.09)	-0.1	0.99	(0.90, 1.10)	-0.1
Parameter 15	0.12694	0.06958	1.00	(0.91, 1.09)	-0.1	0.99	(0.95, 1.05)	-0.2
Parameter 16	-0.48679	0.06707	0.94	(0.91, 1.09)	-1.4	0.95	(0.96, 1.04)	-2.5
Parameter 17	1.07429	0.07351	0.92	(0.91, 1.09)	-1.9	0.95	(0.93, 1.07)	-1.4
Parameter 18	-0.69399	0.06875	0.93	(0.91, 1.09)	-1.5	0.95	(0.95, 1.05)	-2.0
Parameter 19	0.58947	0.06786	0.91	(0.91, 1.09)	-2.1	0.92	(0.96, 1.04)	-3.5
Parameter 20	0.00967	0.06548	0.93	(0.91, 1.09)	-1.5	0.94	(0.97, 1.03)	-3.8
Parameter 21	-1.12519	0.07428	0.94	(0.91, 1.09)	-1.4	0.97	(0.93, 1.07)	-1.0
Parameter 22	-1.03584	0.07499	1.01	(0.91, 1.09)	0.1	1.01	(0.93, 1.07)	0.3
Parameter 23	-0.35359	0.06981	1.02	(0.91, 1.09)	0.6	1.01	(0.95, 1.05)	0.6
Parameter 24	1.32307	0.07928	1.04	(0.91, 1.09)	0.8	1.02	(0.92, 1.08)	0.6
Parameter 25	1.38103	0.07900	0.90	(0.91, 1.09)	-2.4	0.95	(0.91, 1.09)	-1.2
Parameter 26	1.97557	0.09263	1.00	(0.91, 1.09)	0.1	0.99	(0.89, 1.11)	-0.1
Parameter 27	0.84197	0.07036	0.94	(0.91, 1.09)	-1.4	0.95	(0.94, 1.06)	-1.8
Parameter 28	-0.84380	0.07035	0.92	(0.91, 1.09)	-1.8	0.94	(0.94, 1.06)	-2.1
Parameter 29	0.80128	0.07287	0.93	(0.91, 1.09)	-1.6	0.97	(0.94, 1.06)	-1.1
Parameter 30	1.34199	0.07958	0.97	(0.91, 1.09)	-0.6	0.98	(0.92, 1.08)	-0.4

An asterisk next to a parameter estimate indicates that it is constrained

<sup>^</sup> Quick standard errors have been used

### 3-DC4 (Rasch Sub-Dimension Model)

ConQuest: Generalised Item Response Modelling Software Sun Sep 25 22:49 2011

#### SUMMARY OF THE ESTIMATION

Estimation method was: MonteCarlo with 2000 nodes

Assumed population distribution was: Gaussian

Constraint was: CASES

The Data File: G:\thesis\MMM.txt

The format: response 1-30

The regression model:

Grouping Variables:

The item model: item

Sample size: 1000

Final Deviance: 33557.92349

Total number of estimated parameters: 35

The number of iterations: 185

Termination criteria: Max iterations=1000, Parameter Change= 0.00010

Deviance Change= 0.00010

Iterations terminated because the deviance convergence criteria was reached

Random number generation seed: 1.00000

Number of nodes used when drawing PVs: 2000

Number of nodes used when computing fit: 200

Number of plausible values to draw: 5

Maximum number of iterations without a deviance improvement: 100

Maximum number of Newton steps in M-step: 10

Value for obtaining finite MLEs for zero/perfects: 0.30000

=>show parameters!table=2;

ConQuest: Generalised Item Response Modelling Software Sun Sep 25 22:49 2011

IMPORTED MODEL: G:\thesis\NNN.des

Parameter Estimates

VARIABLES	ESTIMATE	ERROR <sup>^</sup>	UNWEIGHTED FIT			WEIGHTED FIT		
			MNSQ	CI	T	MNSQ	CI	T
Parameter 1	-2.03012	0.05950	0.93	(0.91, 1.09)	-1.5	0.95	(0.91, 1.09)	-1.1
Parameter 2	0.49546	0.04974	0.97	(0.91, 1.09)	-0.7	0.97	(0.93, 1.07)	-0.7
Parameter 3	-1.60780	0.05641	0.93	(0.91, 1.09)	-1.5	0.95	(0.91, 1.09)	-1.2
Parameter 4	0.51803	0.04979	1.06	(0.91, 1.09)	1.3	1.06	(0.93, 1.07)	1.7
Parameter 5	0.48203	0.04971	1.03	(0.91, 1.09)	0.7	1.04	(0.93, 1.07)	1.0
Parameter 6	-0.90789	0.05205	1.01	(0.91, 1.09)	0.2	1.01	(0.92, 1.08)	0.2
Parameter 7	0.92314	0.05106	1.05	(0.91, 1.09)	1.1	1.06	(0.93, 1.07)	1.5
Parameter 8	1.32203	0.05303	1.00	(0.91, 1.09)	-0.1	1.01	(0.92, 1.08)	0.3
Parameter 9	-2.07986	0.05987	1.00	(0.91, 1.09)	-0.1	1.01	(0.91, 1.09)	0.3
Parameter 10	0.38869	0.04955	1.06	(0.91, 1.09)	1.3	1.06	(0.93, 1.07)	1.8
Parameter 11	-1.06161	0.05287	0.92	(0.91, 1.09)	-1.9	0.94	(0.92, 1.08)	-1.6
Parameter 12	-0.12103	0.04950	0.95	(0.91, 1.09)	-1.0	0.96	(0.93, 1.07)	-1.2
Parameter 13	-0.62716	0.05080	0.96	(0.91, 1.09)	-0.9	0.98	(0.93, 1.07)	-0.5
Parameter 14	-1.11877	0.05320	0.91	(0.91, 1.09)	-2.1	0.92	(0.92, 1.08)	-2.1
Parameter 15	-0.57849	0.05062	1.00	(0.91, 1.09)	-0.0	1.01	(0.93, 1.07)	0.2
Parameter 16	-0.07109	0.05109	0.98	(0.91, 1.09)	-0.5	0.98	(0.92, 1.08)	-0.5
Parameter 17	1.17870	0.05244	1.02	(0.91, 1.09)	0.5	1.02	(0.92, 1.08)	0.6
Parameter 18	2.16919	0.05727	1.02	(0.91, 1.09)	0.4	1.02	(0.92, 1.08)	0.5
Parameter 19	-0.71547	0.05274	0.93	(0.91, 1.09)	-1.5	0.93	(0.92, 1.08)	-1.8
Parameter 20	2.12894	0.05703	1.01	(0.91, 1.09)	0.2	1.01	(0.92, 1.08)	0.2
Parameter 21	0.09708	0.05092	0.94	(0.91, 1.09)	-1.3	0.95	(0.92, 1.08)	-1.3
Parameter 22	-0.67073	0.05258	0.95	(0.91, 1.09)	-1.1	0.97	(0.92, 1.08)	-0.6
Parameter 23	1.15125	0.05235	1.00	(0.91, 1.09)	0.1	1.01	(0.92, 1.08)	0.2
Parameter 24	-0.56656	0.05222	0.99	(0.91, 1.09)	-0.2	1.01	(0.92, 1.08)	0.2
Parameter 25	1.03048	0.05211	1.01	(0.91, 1.09)	0.3	1.02	(0.93, 1.07)	0.6
Parameter 26	-0.65402	0.05269	1.00	(0.91, 1.09)	0.1	1.00	(0.92, 1.08)	0.1
Parameter 27	-0.72795	0.05295	1.01	(0.91, 1.09)	0.3	1.03	(0.92, 1.08)	0.6
Parameter 28	-1.06633	0.05437	0.97	(0.91, 1.09)	-0.6	1.00	(0.92, 1.08)	0.0
Parameter 29	3.28585	0.06402	1.05	(0.91, 1.09)	1.1	1.05	(0.92, 1.08)	1.1

An asterisk next to a parameter estimate indicates that it is constrained

<sup>^</sup> Quick standard errors have been used