

# Cluster verification for listening learner strategies

by

Stacey Kean Mei Dobbie

A thesis

submitted to the Victoria University of Wellington

in fulfilment of the requirements for the degree of

Master of Science in Statistics

Victoria University of Wellington

2015

## **Abstract**

This thesis presents a comparison of statistical methodologies for cluster verification on ordinal response variables. Methodologies will be applied to a Listening Strategy dataset collected by the Language Learner Strategies research team at the National Institute of Education in Singapore. From this listening dataset, eight clusters suggested by Linguistics theory require verification. The methodologies undertaken is to find which listening strategies have been formed well.

Methods used includes the proportional odds model, confirmatory factor analysis and ordinal agreement model. The proportional odds model is used to establish how well each cluster of questions is built. This is established by checking how similar questions within clusters are. The confirmatory factor analysis is used to verify how well the overall listening clusters have been built. This will be compared to clusters proposed by a statistical method. Lastly, the ordinal agreement model is applied to see how much agreement there is within each of the listening clusters. This will be able to show us which clusters is built better than the other clusters for this listening questionnaire.

Results show that the prediction listening strategy has the highest level of agreement as well as no difference between questions within this cluster. The Socio-affective listening strategy has the lowest level of agreement and very strong evidence of a difference between questions within the cluster. This suggests that the prediction cluster has been formed better than the Socio-affective cluster.

## **Acknowledgement**

I would like to express my gratitude to my supervisor, Dr Ivy Liu for her invaluable guidance, advice and suggestions throughout my Masters study at Victoria University of Wellington.

Special thanks to go to my parents Stephen and Evelyn Dobbie for their support and encouragement throughout the many years of my school and adult life.

Thanks goes to my colleagues as well, especially my Manager Stephen Atkinson for his support and wisdom over the last two years.

Last but not least, special thanks to go to Ian Collins for the comfort, support and encouragement he has given me throughout my two years of doing my Masters.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Background . . . . .	1
1.2	Structure of thesis . . . . .	8
1.3	Research objectives . . . . .	10
<b>2</b>	<b>Data description</b>	<b>11</b>
2.1	Data collection . . . . .	11
2.2	Data cleaning . . . . .	12
2.3	Data characteristics . . . . .	14
<b>3</b>	<b>Proportional odds model</b>	<b>29</b>
3.1	Proportional odds property . . . . .	29
3.2	Meaning of $\beta$ . . . . .	30
3.3	Motivation of proportional odds models . . . . .	32
3.4	Various models for repeated measurements . . . . .	33
3.4.1	Marginal approach . . . . .	34
3.4.2	Subject-specific approach . . . . .	36
3.5	Fuzzy biclustering method . . . . .	39
3.6	Wald test . . . . .	42
3.7	Analysis . . . . .	47
3.7.1	Data coding . . . . .	47
3.7.2	Linguistics listening strategies . . . . .	48

<b>4</b>	<b>Structural equation models</b>	<b>61</b>
4.1	Structural equation models . . . . .	61
4.2	Basic concepts used in structural equation models . . . . .	62
4.2.1	Latent and observed variables . . . . .	62
4.2.2	Exogenous and endogenous latent variables . . . . .	63
4.2.3	The factor analytic model . . . . .	63
4.2.4	Ordinal variables . . . . .	64
4.2.5	Polychoric correlation . . . . .	65
4.3	Confirmatory factor analysis with ordinal variables . . . . .	67
4.3.1	CFA model represented diagrammatically . . . . .	68
4.3.2	CFA model represented mathematically . . . . .	71
4.3.3	Estimation methods . . . . .	73
4.4	Goodness-of-fit . . . . .	76
4.4.1	Comparative fit index and Tucker-Lewis fit index . . . . .	77
4.4.2	Root mean square error of approximation . . . . .	78
4.5	Confirmatory factor analysis application . . . . .	79
4.5.1	Eight strategy structure based on linguistics theory . . . . .	79
4.5.2	Five cluster structure based on the fuzzy biclustering method . . . . .	82
<b>5</b>	<b>Ordinal agreement model</b>	<b>86</b>
5.1	Agreement between questions . . . . .	86
5.2	Ordinal agreement model . . . . .	88
5.3	Agreement among multiple questions . . . . .	90
5.4	Bootstrap method with the ordinal agreement model . . . . .	91
5.4.1	General description of the bootstrap method . . . . .	91
5.4.2	Bootstrapping the ordinal agreement model . . . . .	92
5.4.3	Summary of the overall agreement $\hat{\beta}$ . . . . .	93
5.4.4	Estimation of pairwise agreement . . . . .	95
5.5	Ordinal agreement model application . . . . .	98
5.5.1	Data coding . . . . .	98

5.5.2	Results . . . . .	99
<b>6</b>	<b>Discussion</b>	<b>106</b>
<b>A</b>	<b>Listening questionnaire</b>	<b>116</b>
<b>B</b>	<b>Proportional odds models</b>	<b>119</b>
B.1	Approaches to proportional odds model . . . . .	123
B.1.1	GEE approach . . . . .	123
B.1.2	Subject-specific approach . . . . .	127
<b>C</b>	<b>Confirmatory factor analysis</b>	<b>132</b>
C.1	Linguistic strategies . . . . .	132
C.2	Fuzzy biclustering clusters . . . . .	135
<b>D</b>	<b>Ordinal agreement model</b>	<b>138</b>
D.1	Ordinal agreement model function . . . . .	138
D.2	Bootstrapping of ordinal agreement model . . . . .	139
D.3	Summary of overall agreement $\hat{\beta}$ . . . . .	144
D.4	Estimation of pairwise agreement . . . . .	145

# List of Figures

2.1	Distribution of Standardised Exam Scores for grade 6 students . . .	18
3.1	Cumulative probabilities in a proportional odds model . . . . .	31
4.1	A general structure of a CFA model . . . . .	70
4.2	Linguistics hypothesised eight factor model . . . . .	80
4.3	Statistical hypothesised five factor model . . . . .	83
5.1	Distribution of $\{\hat{\beta}^i, i = 1, \dots, s\}$ for Self-initiation strategy . . . . .	94
5.2	Distribution of $\hat{\beta}_{j,j'}$ for all pairwise combinations for Self-initiation strategy . . . . .	96

# List of Tables

1.1	Listening Strategies . . . . .	7
2.1	Number of students removed from data cleaning . . . . .	14
2.2	Number of students by School and Gender . . . . .	15
2.3	Number of students by Ethnicity and Gender . . . . .	16
2.4	Number of students by Ethnicity and School . . . . .	16
2.5	Column proportion of students by Ethnicity and School . . . . .	17
2.6	Row proportion of students by Ethnicity and School . . . . .	17
2.7	Self-initiation - number and proportion of students by question and rating . . . . .	19
2.8	Planning - number and proportion of students by question and rating	20
2.9	Monitoring and Evaluation - number and proportion of students by question and rating . . . . .	21
2.10	Perceptual Processing - number and proportion of students by question and rating . . . . .	22
2.11	Inferencing - number and proportion of students by question and rating	23
2.12	Prediction - number and proportion of students by question and rating	24
2.13	Utilisation - number and proportion of students by question and rating	25
2.14	Socio-affective - number and proportion of students by question and rating . . . . .	26
3.1	Stages of model selection for Self-initiation strategy using the GEE approach . . . . .	45



3.2	Self-initiation strategy . . . . .	49
3.3	Planning strategy . . . . .	49
3.4	Monitoring and Evaluation strategy . . . . .	50
3.5	Perceptual Processing strategy . . . . .	51
3.6	Inferencing strategy . . . . .	52
3.7	Prediction strategy . . . . .	52
3.8	Utilisation strategy . . . . .	53
3.9	Socio-affective strategy . . . . .	54
3.10	Self-initiation strategy question effects . . . . .	55
3.11	Utilisation strategy question effects . . . . .	56
3.12	Monitoring and Evaluation strategy question effects for Chinese students . . . . .	57
3.13	Monitoring and Evaluation strategy question effects for Other students	58
3.14	Perceptual Processing strategy question effects . . . . .	59
3.15	Socio-affective strategy question effects . . . . .	60
4.1	Chi-square estimation for Linguistics hypothesised model . . . . .	81
4.2	Chi-square estimation for Linguistics baseline model . . . . .	81
4.3	Fit indices for Linguistics hypothesised model . . . . .	81
4.4	Chi-square estimation for five factor hypothesised model . . . . .	84
4.5	Chi-square estimation for five factor baseline model . . . . .	84
4.6	Fit indices for five cluster hypothesised model . . . . .	85
5.1	Question 11 versus Question 12 . . . . .	87
5.2	Standardised Pearson Residuals . . . . .	88
5.3	Bootstrap sample structure for Self-initiation strategy . . . . .	93
5.4	Scale (1, 2, 3, 4, 5) . . . . .	99
5.5	Scale (1, 2, 4, 5, 6) . . . . .	100
5.6	Scale (1, 3, 4, 5, 7) . . . . .	100
5.7	Agreement between questions for Prediction strategy . . . . .	101

5.8	Standardised Pearson residuals for Prediction strategy . . . . .	102
5.9	Agreement between questions for Socio-affective strategy . . . . .	103
5.10	Standardised Pearson residuals for Socio-affective strategy - Question 10 versus Question 25 . . . . .	104
5.11	Standardised Pearson residuals for Socio-affective strategy - Question 10 versus Question 38 . . . . .	105

# Chapter 1

## Introduction

### 1.1 Background

The Language Learner Strategy Research Team (2009) at the National Institute of Education (NIE) designed and administered a Listening Strategy Questionnaire. The questionnaire was administered at six local schools in Singapore in January 2005 amongst students from Grades 4-6. The purpose of the questionnaire undertaken by NIE was to empower learners, help teachers, and inform policy makers with detailed information about language learning strategies. Each strategy is formed from several listening questions from the questionnaire. The aim of this research is to find out which listening language learner strategies has been formed well.

In the Listening Strategy Questionnaire, there are two sections. The first section contains demographic questions about the student. This includes gender, age, grade, school and ethnicity. Section two contains thirty eight listening strategy questions. For these listening strategy questions, students can choose an answer from the scale (*never, rarely, sometimes, often, always*). This scale has a neutral middle category and is often called a Likert scale. This type of scale is an ordered categorical scale and so is called ordinal.

Ordinal scales are highly used in social sciences for measuring attitudes and opinions. For example, a question administered in the listening questionnaire asked each student “When I’m free, I find interesting things to listen to in English” using the categories (*never, rarely, sometimes, often, always*). It is often that for each observation, the choice of a category is subjective. To lessen subjectivity, it is helpful to provide guidance about what the categories represent.

For ordinal scales, there is a clear ordering of the levels. However, the absolute distances among them are unknown. Listening questions measured with categories (*never, rarely, sometimes, often, always*) is ordinal because, a student who chooses “often” would have attempted listening to English more than a student who chose “never”. However, no numerical measure is given to the difference between these levels. An ordinal variable is quantitative rather than qualitative as each label on its scale refers to a greater or smaller magnitude of a certain characteristic than another level.

Researchers often treat ordinal categorical response variables as continuous response variables. The categorical structure of the response variable is ignored and standard parametric methods for continuous response variables are used instead. This approach would include assigning numerical scores to the ordered categories and then using the method of ordinary least squares (OLS) such as linear regression and analysis of variance (Agresti 2010).

Linear regression is a simplistic method that is generally used to model ordinal response scores by assigning numerical scores. Although this approach can be useful at identifying variables that clearly affect a response variables, there are limitations.

Agresti (2010, Ch. 1) listed limitations of treating ordinal categorical response variables as continuous response variables as follows:

1. There is usually not a clear-cut choice for the scores. This is again due to the unknown distances between levels in the scale.
2. A particular response outcome is likely to be consistent with a range of values for some underlying latent variable. An ordinary regression analysis does not allow for the measurement error that results from replacing such a range by a single numerical value.
3. Does not yield estimated probabilities for the response categories at fixed settings of the explanatory variables.
4. Can yield predicted values above the highest category score or below the lowest category score.
5. Ignores the facts that the variability of the responses is naturally non-constant for categorical data. For ordinal response variables, there is little variability of predictor values for which the observations fall mainly in the highest category or mainly in the lowest category. However, there is considerable variability of predictor values for which observations tend to be spread among the categories.

Limitations 2, 4 and 5 described above does not account for “ceiling effects” and “floor effects” (Agresti 2010, Ch. 1). These occur because of the upper and lower limits of the ordinal response variable. Such effects can cause ordinary regression modelling to give misleading results.

For example, Hastie et al. (1989) presented a study of women in South Africa that modelled an ordinal measurement  $y$  of osteoporosis in terms of the explanatory variable  $x$  for age and an indicator variable  $z$  for whether women had osteoarthritis. Using OLS, for each osteoarthritis group, the line relating age to the predicted

osteoporosis score took value at the lowest ordinal level near a relatively low age level. However, the line for the group with positive osteoporosis had a significantly greater slope as age increased. When an ordinal model was used on the data instead, they found no evidence of this interaction.

There is still widespread practice of treating ordinal categorical variables as continuous variables. In the field of Linguistics, researchers analysing data of learning English surveys, including Zheng (2013), Wu (2014), Souriyavongsa et al. (2013) and others, have used Likert scales in their surveys and continue to treat the ordinal categorical variables as continuous variables. They have done this by assigning numerical scores to the ordinal categorical variables. As part of their analysis, they calculate the mean and standard deviation of these scores. Calculating the mean relies on the assumption that the distance between scores are known. For example, the response “never”, “rarely”, “sometimes”, “often” and “always” are assigned scores “1”, “2”, “3”, “4” and “5”. The distance between “rarely” and “sometimes” is then the same as the distance between “sometimes” and “often”. However, in reality, these distances may be different and unknown.

On the other hand, some researchers are treating ordinal categorical variables as nominal variables. This is done by ignoring their ordering nature. Advantages, suggested by Agresti (2010, Ch. 1), of treating an ordinal categorical variables as ordinal rather than nominal are as follows:

- Ordinal data description can use measures that are similar to those used in ordinary regression and analysis of variance for quantitative variables, such as correlations and means.
- Ordinal analysis can use a greater variety of models. Those models are more parsimonious and have simpler interpretations than the standard models for non-ordered categorical variables.

- Ordinal methods have greater power for detecting relevant trends or alternatives to the null hypothesis of no effect of an explanatory variable on the response variable.
- Ordinal analysis can give more powerful results than an analysis that ignores ordinality.

### **Clustering**

The main purpose of this research is to verify that the English listening learner strategies formed by the questionnaire are appropriate. There might be several ways to check the adequacy, such as using Linguistics theory. This thesis focuses on checking whether the questions within each cluster (strategy) are similar or not. In this section, a discussion on where the clusters came about are described.

Listening strategies are grouped into three categories. They are metacognitive, cognitive and socioaffective strategies. Metacognitive strategies involve thinking about the mental processes used in the learning process, monitoring learning while it takes place, and evaluating learning after it has taken place (O'Malley & Chamot 1990). Overall strategies comprising of metacognitive strategies can include the following:

- Advance organisers
- Directed attention
- Selective attention
- Self management
- Self monitoring
- Delayed production
- Self evaluation

Cognitive strategies refer to processes and behaviours used by learners to help them improve their ability and learn or remember something. They involve active manipulation of the listening task and can include the following:

- Repetition
- Inference
- Elaboration
- Deduction
- Transfer
- Recombination

Social and affective strategies play important roles in instructional systems designed for second language learners. These strategies entail cooperative teaming, questioning for clarification and affective control over listening experiences. They are described as the following:

- Cooperation - working with one or more peers to obtain feedback
- Question for clarification - asking a teacher for repetition, paraphrasing explanation or examples.



Based on these three types of strategies, the following table shows the eight strategies used in the questionnaire that will be tested and verified.

Table 1.1: Listening Strategies

Strategy Type	Strategy	Example
Metacognitive	Self-initiation	I look for opportunities to listen in English.
	Planning	Before I start listening, I decide if I need to pay attention to details or to the main ideas.
	Monitoring and Evaluating	If I have a problem in understanding, I quickly decide whether I should continue or listen again.
Cognitive	Perceptual Processing	If I can't understand a word or phrase, I repeat it to myself.
	Inferencing	When I don't understand something, I use my general knowledge to make a guess.
	Predicting	Before I listen to something, I think about the main idea of what I am going to hear.
	Utilisation	I try to connect what I heard with my own experiences.
Social and Affective	Socio-affective strategies	When I have a problem in listening, I ask my teachers for help.

The thirty-eight questions given in Appendix A are split into each of the above eight strategy clusters as follows:

- Self-initiation: Questions 11, 12, 14 and 18
- Planning: Questions 3, 15 and 23
- Monitoring and evaluation: Questions 5, 9, 21, 32 and 36
- Perceptual Processing: Questions 8, 24, 34, and 37
- Inferencing: Questions 1, 4, 17, 26, 27, 31 and 33
- Predicting: Questions 6, 7, 13, 29 and 35
- Utilisation: Questions 16, 19, 20, 22 and 30
- Socio-affective: Questions 2, 10, 25, 28 and 38.

These eight groupings of the listening survey questions will be used in all methodologies in this thesis to check that the strategies have been formed well.

## 1.2 Structure of thesis

Chapter two describes the data, including how it was collected, possible issues and how the data was cleaned. Chapter two also describes the demographic and question characteristics of the sample of students surveyed.

Chapter three introduces the concept of the proportional odds model (McCullagh 1980) with repeated measurements applied to each strategy individually. A clustering method based on proportional odds models (Matechou et al. 2011), called the fuzzy biclustering method will be described as well. It uses the expectation-maximisation algorithm to cluster both the rows and columns of the data. Parts of this clustering will be incorporated in the various approaches. Lastly, the results

to the application of these approaches will be described.

Chapter four describes the concepts on structural equation models and specifically the confirmatory factor analysis. The purpose of this method is to see how well formed the eight listening strategies are. The results to this model will be compared to the results of the five clusters suggested by the fuzzy biclustering method described in Chapter three. This will show us how well both types of clustering is formed and whether the listening strategies are better formed.

Chapter five describes the concepts of the ordinal agreement model. The bootstrapping method, will be incorporated to make inferences for the ordinal agreement model. This chapter will show the application of this analysis to the data and see how much agreement there is between questions within a strategy.

Chapter six discusses the findings of this research and makes the conclusions.

### 1.3 Research objectives

The aim of this research is to verify eight strategy clusters suggested by Linguistic theory and compare their clusters with the ones suggested by the fuzzy clustering method. We are able to find whether a strategy has been designed effectively, in the sense that all questions within the strategy are similar.

There are three key objectives to this research:

1. Establish how well each cluster of questions is built.
2. Verify how well the overall listening clusters have been built.
3. How much agreement there is within each of the listening clusters.

# Chapter 2

## Data description

This chapter describes the data collection procedure, demographic characteristics of the students surveyed, characteristics of the strategies, coding of the data and quality issues for the English listening dataset.

### 2.1 Data collection

At the six Singapore schools where the Language Learner Strategy Research Team from NIE undertook the English listening questionnaire, guidelines were given to each teacher involved Gu et al. (2005). The guidelines included what materials were involved, the procedures to follow and answers to potential questions that students may have.

The material guidelines included what each teacher should have to successfully administer the questionnaire, consent forms for themselves and the students, and a description of what is in the guidelines document.

The procedures guidelines describe in detail what the teachers must do. This included a detailed timeline on how to administer the questionnaire with the students.

This would allow consistency between the schools and classrooms with how long each student spends on answering the questions in the questionnaire.

The potential questions and answers section had examples of questions that students may ask and how the teacher could answer. When the questionnaire was administered, an NIE researcher was present. This was so that if students had further questions that the teachers were not sure about, the researcher could answer for them.

## 2.2 Data cleaning

In this section we describe the data cleaning that was required to prepare the data for three statistical analyses which are presented in the next few chapters.

The listening questionnaire was completed by grade 4, 5 and 6 students. For our analysis, we are going to look at data completed by grade 6 students only, suggested by the data owner, Dr Peter Gu (School of Linguistics and Applied Language Studies, Victoria University of Wellington). This is because grade 6 students are older and would have a better sense and understanding of the questions being asked. Research has indicated that by 11 years old, children's ability to remember is not so different to adults. That at this age, most children are fully able to articulate their perceptions, opinions and beliefs (Scott 2000). Responses would therefore be more reliable from grade 6 students.

For data analysis, missing data are often problematic. The methods used in this thesis to analyse the data are maximum likelihood (ML) and generalised estimating equations (GEE) (Liang & Zeger 1986) methods. The ML inference has the advantage of being applicable under weaker assumptions about the missing data mechanism than GEE inference. In particular, the GEE method requires the data

to be missing completely at random (MCAR) whereas the ML method only requires missing at random (MAR).

If data is MCAR then the missing data are a random sample of all observations. That is, the probability of any observation being missing is independent of the value of the data, missing or observed (Little & Rubin 2002, p. 12). In modelling ordinal listening responses as a function of questions, if the probability that the listening responses is missing is the same for all students regardless of the question and the response, the data are MCAR.

If data are MAR, then the missingness of the data does not depend on the components that are missing (Little & Rubin 2002, p. 12). If the probability that the listening responses are missing varies between the questions but does not vary according to the listening response of the students with the same questions, then the data are MAR.

Due to the simplicity of demonstrating various methods on a clean dataset, students with missing values are excluded from the dataset for all analyses. Although deleting the missing cases is very strong, it is valid under MCAR (Little & Rubin 2002, Little 1988). For this listening dataset, a small proportion of grade 6 students (14%) were removed.

The following points outline the data cleaning that was applied to the listening learner questionnaire dataset:

1. Select Grade 6 students only.
2. Remove all rows with missing values.

Table 2.1 shows the number of students (each record is a student) in the data clean for each step described above.

Table 2.1: Number of students removed from data cleaning

Step Description	Difference	Number of Records
Total number of records		3,618
Grade 6 students	2,407	1,211
Remove records with missing values	172	1,039

The remaining 1,039 students were used for all analyses described in later chapters.

## 2.3 Data characteristics

In this section the population of interest is defined and explored. The characteristics of the students such as their school, gender, ethnicity and exam scores are described. The relationship between the demographic characteristics are explored and the count and proportion of students for each question in each strategy are then explored. This will establish our expectations about the relationships among the students' characteristics and the questions.

### Student characteristics

Table 2.2 shows the number and proportion of grade 6 students from each school that participated in the survey by gender. The largest proportion of students, at 24%, are students of Fuchun Primary School, followed by Junyuan Primary School. The majority of students are male with 56.5% overall. This is partly due to one of the schools being an all boys school. Amongst boy students, the highest proportion of boys are from Catholic High School (27.4%). For all schools, excluding Catholic High School, the split of boys and girls are approximately fifty-fifty. Jurong West



Primary School has a slightly higher proportion of girls of 56.6%.

Table 2.2: Number of students by School and Gender

School	Student Count			Column %			Row %	
	Boy	Girl	Total	Boy	Girl	Overall	Boy	Girl
Catholic High School	161	0	161	27.4	0	15.5	100	0
Clementi Primary School	50	49	99	8.5	10.8	9.5	50.5	49.5
Fuchun Primary School	124	127	251	21.1	28.1	24.2	49.4	50.6
Junyuan Primary School	109	114	223	18.6	25.2	21.5	48.9	51.1
Jurong West Primary School	56	73	129	9.5	16.2	12.4	43.4	56.6
Lakeside Primary School	87	89	176	14.8	19.7	16.9	49.4	50.6
Total	587	452	1,039	100	100	100	56.5	43.5

Table 2.3 shows the number and proportion of grade 6 students split by ethnicity and gender. The majority of students, at 68%, are of Chinese ethnicity, followed by Malay at 25%. Indian, Eurasian and Other ethnicity make up 6.5% of the students. Chinese and Malay ethnicities being the largest proportion of students surveyed is not unexpected as the majority of the Singapore population is 74% Chinese followed by 13% Malays, as at the end of June 2013 (Department of Statistics Singapore). The gender split by ethnicity is not as evenly distributed as seen in the school split. All ethnicities except Malays, has a higher proportion of boys than girls.

Table 2.3: Number of students by Ethnicity and Gender

Ethnicity	Student Count			Column %			Row %	
	Boy	Girl	Total	Boy	Girl	Overall	Boy	Girl
Chinese	421	290	711	71.7	64.2	68.4	59.2	40.8
Eurasian	4	2	6	0.7	0.4	0.6	66.7	33.3
Indian	26	21	47	4.4	4.6	4.5	55.3	44.7
Malay	126	134	260	21.5	29.6	25.0	48.5	51.5
Other	10	5	15	1.7	1.1	1.4	66.7	33.3
Total	587	452	1,039	100	100	100	56.5	43.5

Table 2.4, 2.5 and 2.6 shows the number and proportion of grade 6 students by school and ethnicity. The counts for Eurasian, Indian and Other ethnicities are very small by school. For all schools, the majority of students (over 55%) are of Chinese ethnicity. At Catholic High School, 98% of students surveyed are of Chinese ethnicity. This could be due to the school being a Catholic religion school. The remaining schools all have Malay as their second most common ethnicity.

Table 2.4: Number of students by Ethnicity and School

School	Student Count					
	Chinese	Eurasians	Indian	Malay	Others	Total
Catholic High School	157	1	2	1	0	161
Clementi Primary School	63	2	9	23	2	99
Fuchun Primary School	158	1	19	69	4	251
Junyuan Primary School	137	2	3	79	2	223
Jurong West Primary School	98	0	5	25	1	129
Lakeside Primary School	98	0	9	63	6	176
Total	711	6	47	260	15	1,039

Table 2.5: Column proportion of students by Ethnicity and School

School	Column %				
	Chinese	Eurasians	Indian	Malay	Others
Catholic High School	22.1	16.7	4.3	0.4	0
Clementi Primary School	8.9	33.3	19.1	8.8	13.3
Fuchun Primary School	22.2	16.7	40.4	26.5	26.7
Junyuan Primary School	19.3	33.3	6.4	30.4	13.3
Jurong West Primary School	13.8	0	10.6	9.6	6.7
Lakeside Primary School	13.8	0	19.1	24.2	40.0
Total	100	100	100	100	100

Table 2.6: Row proportion of students by Ethnicity and School

School	Row %					Total
	Chinese	Eurasians	Indian	Malay	Others	
Catholic High School	97.5	0.6	1.2	0.6	0	100
Clementi Primary School	63.6	2.0	9.1	23.2	2.0	100
Fuchun Primary School	62.9	0.4	7.6	27.5	1.6	100
Junyuan Primary School	61.4	0.9	1.3	35.4	0.9	100
Jurong West Primary School	76.0	0	3.9	19.4	0.8	100
Lakeside Primary School	55.7	0	5.1	35.8	3.4	100

Figure 2.1 shows the distribution of grade 6 students exam scores. Note that the students' exam scores have been standardised separately by each school before putting the scores together. This is because each school have different tests and so the exam scores are not comparable between schools.

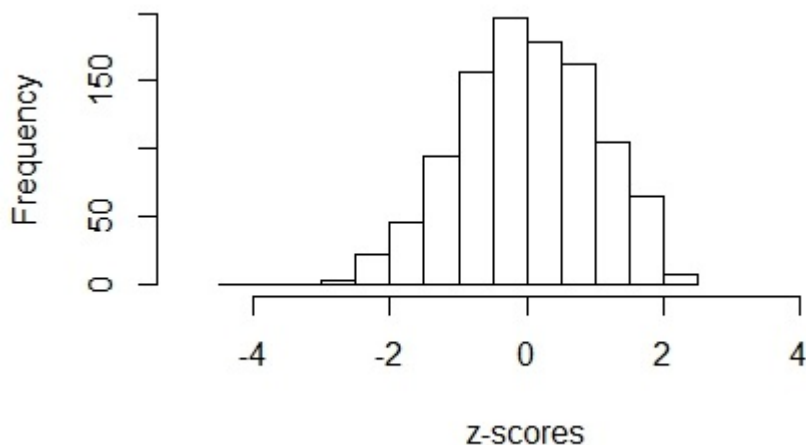


Figure 2.1: Distribution of Standardised Exam Scores for grade 6 students

The distribution of scores shows a bell-shaped curve around zero, which suggests the scores may be normally distributed.

### Each strategy's characteristics

For each strategy, the count and proportion of students' responses for each question and rating scale are displayed and described. The tables show a rating scale of 1 to 5 where 1 represents "rarely" and 5 represents "always".

Table 2.7 shows the number and proportion of students' responses for each question in the Self-initiation strategy by rating scale. Ratings 1 and 4 have similar proportions of students for each question being approximately 6-9% and 22-25% respectively. The proportion for rating 2 however varies quite a bit by question. Questions

11 and 12 have similar proportions of 17% and 19% consecutively, whereas questions 14 and 18 have similar proportions of 12% and 14% consecutively. Between these two pairs, the proportions appear different. For rating 3, question 18 has a lower proportion of 26% compared to 32-36% for the other three questions. A similar pattern has occurred for rating 5 but reversed, where question 18 has a higher proportion of 29% compared to 16-21% for the other three questions.

Table 2.7: Self-initiation - number and proportion of students by question and rating

Question	Student Count					%				
	1	2	3	4	5	1	2	3	4	5
11	89	180	331	251	188	9%	17%	32%	24%	18%
12	65	196	378	229	171	6%	19%	36%	22%	16%
14	66	127	368	258	220	6%	12%	35%	25%	21%
18	73	145	275	241	305	7%	14%	26%	23%	29%

Note the questions for table 2.7 are as follows:

- Q11: To improve my listening in English, I watch English TV programmes.
- Q12: I look for opportunities to listen in English.
- Q14: I try to find out how to improve my listening in English.
- Q18: When I'm free, I find interesting things to listen to in English (for example, TV, radio, etc).

Table 2.8 shows the number and proportion of students' responses for each question in the Planning strategy by the rating scale. The counts of students across the rating scale appears similar between the questions. The proportion of counts make it easier to see that the proportion of students in each category of the scale

is similar between the questions.

Table 2.8: Planning - number and proportion of students by question and rating

Question	Student Count					%				
	1	2	3	4	5	1	2	3	4	5
13	115	181	357	249	137	11%	17%	34%	24%	13%
15	123	180	310	262	164	12%	17%	30%	25%	16%
23	110	188	342	227	172	11%	18%	33%	22%	17%

Note the questions for table 2.8 are as follows:

- Q13: When I listen, I use what I have already heard to think about what I will hear next.
- Q15: Before I listen to something, I ask myself whether it is important to me.
- Q23: Before I start listening, I decide if I need to pay attention to details or to the main idea.

Table 2.9 shows the number and proportion of students' responses for each question in the Monitoring and Evaluation strategy by the rating scale. Questions 9 and 36 has the highest proportion of students scoring a rating 2 and 3 compared to the other questions. This is offset by these same questions having the lowest proportion of students scoring a rating 4 and 5 compared to the other questions.

Table 2.9: Monitoring and Evaluation - number and proportion of students by question and rating

Question	Student Count					%				
	1	2	3	4	5	1	2	3	4	5
5	48	114	336	321	220	5%	11%	32%	31%	21%
9	85	161	350	267	176	8%	15%	34%	26%	17%
21	59	108	287	313	272	6%	10%	28%	30%	26%
32	83	121	324	297	214	8%	12%	31%	29%	21%
36	71	166	371	255	176	7%	16%	36%	25%	17%

Note the questions for table 2.9 are as follows:

- Q5: I compare what I am hearing with what I have already heard to make sure I understand correctly.
- Q9: During or after listening, I ask myself whether the information is the same as what I already know.
- Q21: When I have a problem in listening, I decide whether I should pay more attention to it.
- Q32: If I have a problem in understanding, I quickly decide whether I should continue or listen again.
- Q36: During or after listening, I check how much I have understood.

Table 2.10 shows the number and proportion of students' responses for each question in the Perceptual Processing strategy by the rating scale. The proportion of students in each rating category is similar between questions 8, 24 and 34. Question 37 appears to have a different pattern to the other questions. The lower end of the rating scale (rating 1 and 2) has a low proportion of students (13%) for question 37 compared to the other questions (26%-33%). This has been offset by the upper

end of the rating scale where 30% of students selected rating 5 compared to the other questions (12-17%).

Table 2.10: Perceptual Processing - number and proportion of students by question and rating

Question	Student Count					%				
	1	2	3	4	5	1	2	3	4	5
8	128	223	328	225	135	12%	21%	32%	22%	13%
24	93	173	328	272	173	9%	17%	32%	26%	17%
34	78	187	375	270	129	8%	18%	36%	26%	12%
37	26	166	288	309	310	3%	10%	28%	30%	30%

Note the questions for table 2.10 are as follows:

- Q8: When I listen, I repeat the words or phrases I can understand.
- Q24: If I can't understand a word or phrase, I repeat it to myself.
- Q34: When I listen, I repeat the pronunciation of the words I have heard.
- Q37: When I listen, I pay attention to every word that is said.

Table 2.11 shows the number and proportion of students' responses for each question in the Inferencing strategy by the rating scale. For question 1 over half of students (52%) selected rating 3. The other questions on the other hand, had between 32% and 40% of students selected rating 3. Question 17 and 26 both had much higher proportions of students with rating 5 of 18% and 21% respectively compared to the other questions being between 10% and 13%. Question 27 had a higher proportion of students (42%) with ratings 1 and 2, offset by a lower proportion of students (26%) with ratings 4 and 5. The proportions between ratings for questions 31 and 33 are similar.



Table 2.11: Inferencing - number and proportion of students by question and rating

Question	Student Count					%				
	1	2	3	4	5	1	2	3	4	5
1	55	121	539	197	127	5%	12%	52%	19%	12%
4	63	171	395	276	134	6%	16%	38%	27%	13%
17	49	164	380	260	186	5%	16%	37%	25%	18%
26	66	131	333	296	213	6%	13%	32%	28%	21%
27	139	287	341	164	108	13%	28%	33%	16%	10%
31	81	225	415	217	101	8%	22%	40%	21%	10%
33	84	187	410	234	124	8%	18%	39%	23%	12%

Note the questions for table 2.11 are as follows:

- Q1: When I don't understand something, I use my knowledge about the English language to guess.
- Q4: When I don't understand something, I use what I have already heard to guess it.
- Q17: When I don't understand something, I make several guesses.
- Q26: When I don't understand something, I use my general knowledge to make a guess.
- Q27: When I listen, I use my knowledge about the English language to think about what I will hear next.
- Q31: When I don't understand something, I quickly decide whether I should continue or listen again.
- Q33: When I don't understand something, I use my knowledge about the topic to guess.

Table 2.12 shows the number and proportion of students' responses for each question in the Prediction strategy by the rating scale. For all questions, rating 1 and 5 had very similar proportions of students ranging between 9% and 13%. Ratings 2 and 4 had similar proportion of students between all questions (between 19% and 23%) as well. Question 29 had the highest proportion of students scoring a rating of 3 at 41%. This is slightly higher than the proportions in the other questions of this rating.

Table 2.12: Prediction - number and proportion of students by question and rating

Question	Student Count					%				
	1	2	3	4	5	1	2	3	4	5
6	125	206	359	236	113	12%	20%	35%	23%	11%
7	108	218	347	231	135	10%	21%	33%	22%	13%
13	124	197	365	236	117	12%	19%	35%	23%	11%
29	114	195	428	212	90	11%	19%	41%	20%	9%
35	106	244	379	198	112	10%	23%	36%	19%	11%

Note the questions for table 2.12 are as follows:

- Q6: When I listen, I use my knowledge about text type (for example, story, report, etc) to think about what I will hear next.
- Q7: Before I listen to something, I think about the main idea of what I am going to hear.
- Q13: When I listen, I use what I have already heard to think about what I will hear next.
- Q29: When I listen, I use my knowledge about the topic to think about what I will hear next.
- Q35: When I listen, I use my general knowledge to think about what I will hear next.

Table 2.13 shows the number and proportion of students' responses for the Utilisation strategy by rating scale. Question 19 had the lowest proportion of students with rating 1 and 2 at 4% and 12% respectively. The other questions had proportions ranging between 7% and 11% for rating 1 and between 19% and 21% for rating 2. Question 19 has the highest proportion of students responding with a rating 4 and 5 at 31% and 20% respectively compared to the other questions. Question 20 has the highest proportion of students responding with a rating 3 at 40% compared to other questions with between 31% and 34%.

Table 2.13: Utilisation - number and proportion of students by question and rating

Question	Student Count					%				
	1	2	3	4	5	1	2	3	4	5
16	113	200	355	240	131	11%	19%	34%	23%	13%
19	45	126	338	323	207	4%	12%	33%	31%	20%
20	70	202	420	238	109	7%	19%	40%	23%	10%
22	101	215	317	251	155	10%	21%	31%	24%	15%
30	79	214	350	278	118	8%	21%	34%	27%	11%

Note the questions for table 2.13 are as follows:

- Q16: I try to find problems with how the speaker presented an idea so that I can avoid the same problems.
- Q19: When I listen, I try to remember useful words and phrases so that I can use them.
- Q20: I try to connect what I heard with my own experiences.
- Q22: After I finish listening, I make a summary in my mind about what I heard.
- Q30: After I finish listening, I use my own words to retell what I heard in my mind.

Table 2.14 shows the number and proportion of students' responses for each question in the Socio-affective strategy by rating scale. The proportions between all the questions within the Socio-affective strategy vary significantly. Question 25 and 38 have the lowest proportion of ratings 1 and 2 of 22% and 19%, respectively, while they have the highest rating 5 of 23% and 33%, respectively. Question 28 appears different to the other questions, where it has a higher proportion of students with a rating of 2 and 3 than the other questions offset by a lower rating 4 and 5 compared to the other questions.

Table 2.14: Socio-affective - number and proportion of students by question and rating

Question	Student Count					%				
	1	2	3	4	5	1	2	3	4	5
2	132	238	325	224	120	13%	23%	31%	22%	12%
10	112	207	314	237	169	11%	20%	30%	23%	16%
25	86	143	303	272	235	8%	14%	29%	26%	23%
28	139	287	341	164	108	13%	28%	33%	16%	10%
38	81	113	272	234	339	8%	11%	26%	23%	33%

Note the questions for table 2.14 are as follows:

- Q2: When I have a problem in listening, I ask my family members for help.
- Q10: When I have a problem in listening, I ask my friends for help.
- Q25: I tell myself to enjoy listening to English.
- Q28: When I have a problem in listening, I ask my teachers for help.
- Q38: I tell myself not to worry when I listen in English.

The eight strategies can be summarised as follows:

- Self-initiation - 3 out of the 4 questions in this strategy show similar proportions for each rating.
- Planning - for all 3 questions in this strategy, the proportions for each rating is similar between questions.
- Monitoring and Evaluation - questions 9 and 36 show similar proportions for each rating between the two questions. Same with questions 5, 21 and 32. Proportions between these two groups of questions appear different.
- Perceptual Processing - questions 8, 24 and 34 shows similar rating proportions between the questions. Proportions across ratings for question 37 appear different to the other questions in this strategy.
- Inferencing - proportions between questions for each rating appears different. This is especially the case for question 27 where proportions for ratings 1, 2 and 4 are different from other questions for the same rating.
- Prediction - question 29 shows different rating proportions to the other questions within this strategy. The remaining four questions have similar rating proportions between the questions.
- Utilisation - rating proportions for the questions in this strategy are somewhat different between questions. Question 19 has the lowest proportion in rating 1 and 2, offset by the highest proportion in rating 4 and 5. Question 20 has the highest proportion in rating 3 compared to other questions. This is offset by having the lowest proportion in rating 5.
- Socio-affective - rating proportions between questions are different between all questions.

In the next three chapters, we discuss three statistical methodologies of ordinal data. These methods will be applied to the listening strategies dataset to investigate further how well the strategies have been formed.

# Chapter 3

## Proportional odds model

This chapter discusses proportional odds models. Section 3.1 describes the proportional odds property. Section 3.2 describes the meaning behind the coefficients of model predictors. Section 3.3 describes the motivation behind the use of the proportional odds model. Section 3.4 describes various methods for repeated measurements. This includes the marginal approach (e.g., GEE) and subject-specific approach (e.g., random effect models). In section 3.5, we explore a fuzzy clustering method. Section 3.6 describes the Wald test and how it can be used for model selection and model inferencing. Section 3.7 presents the results of the analysis based on the different approaches of the proportional odds model.

### 3.1 Proportional odds property

The proportional odds model suggested by McCullagh (1980) is a class of multivariate generalised linear models used for the analysis of ordinal data. Let  $Y$  be denoted as the response variable, where it is treated as a ordered categorical variable.

The general form of the model can be written as the following:

$$\log \left( \frac{P(Y \leq j)}{1 - P(Y \leq j)} \right) = \text{logit}[P(Y \leq j)] = \alpha_j - \boldsymbol{\beta}'\mathbf{x} \quad j = 1, \dots, c - 1 \quad (3.1)$$

where the response variable  $Y$  has  $c$  ordered categories and  $\mathbf{x}$  is a set of explanatory variables. The parameters  $\alpha_j$  are known as cutpoints, where  $\alpha_1 < \alpha_2 < \dots < \alpha_{c-1}$ . This model is called the proportional odds model because, for each of the  $c - 1$  logistic regression models, they simultaneously have the same parameter  $\boldsymbol{\beta}$ . That is, the effect  $\boldsymbol{\beta}$  is the same for all  $j$ .

## 3.2 Meaning of $\boldsymbol{\beta}$

In this section, the meaning of the effect  $\boldsymbol{\beta}$  is discussed where the predictor variable  $\mathbf{x}$  is continuous or discrete.

For a continuous predictor variable  $x$  and a fixed  $j$ , the response curve is a logistic regression curve for a binary response with outcomes  $Y \leq j$  and  $Y > j$  (Agresti 2012, p. 275). When  $c$  is, for example, equal to 4, the response curves for  $j = 1, 2$ , and 3 have the same shape, where they share the same rate of increase or decrease. This is because the common effect  $\boldsymbol{\beta}$  is included in all three cumulative logits. The size of  $|\boldsymbol{\beta}|$  determines the gradient of the curve, where the curve would be flatter for small  $|\boldsymbol{\beta}|$  and steep for large  $|\boldsymbol{\beta}|$ . The only difference is that the curves are horizontally shifted from each other. Figure 3.1 shows cumulative probabilities in a proportional odds model when  $c = 4$ .



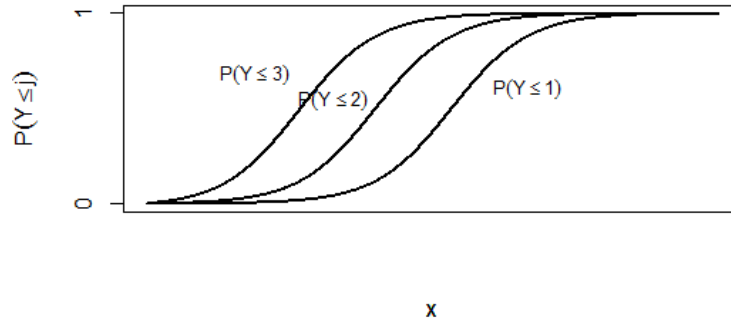


Figure 3.1: Cumulative probabilities in a proportional odds model

The proportional odds model then satisfies

$$\begin{aligned} \text{logit}[P(Y \leq j|x_1)] - \text{logit}[P(Y \leq j|x_2)] \\ &= (\alpha_j - \beta x_1) - (\alpha_j - \beta x_2) \\ &= \beta(x_2 - x_1) \end{aligned}$$

The odds of making response  $\leq j$  at  $x = x_1$  are  $\exp(\beta(x_2 - x_1))$  times the odds at  $x = x_2$ . The log cumulative odds ratio is proportional to the distance between  $x_1$  and  $x_2$ . The same proportionality constant applies to each logit.

When the explanatory variable is a qualitative indicator variable, for example a categorical variable with  $r$  levels, the predictor is then a factor in the model. This model has the form

$$\text{logit}[P(Y \leq j)] = \alpha_j - (\tau_1 z_1 + \dots + \tau_{r-1} z_{r-1}) \quad j = 1, 2, \dots, c - 1,$$

where  $z_i = 1$  for an observation in level  $i$  and  $z_i = 0$  otherwise. As a result, it would be redundant to include an indicator variable for the final category. The effects terms take the form of row effects. We can therefore express the model in terms of the effect for an observation in level  $i$  as

$$\text{logit}[P(Y \leq j)] = \alpha_j - \tau_i \quad j = 1, 2, \dots, c - 1, \quad i = 1, \dots, r - 1.$$

For comparing levels  $a$  and  $b$ , the cumulative log odds ratio is

$$\begin{aligned} \text{logit}[P(Y \leq j|X = a)] - \text{logit}[P(Y \leq j|X = b)] \\ &= (\alpha_j - \tau_a) - (\alpha_j - \tau_b) \\ &= \tau_b - \tau_a \end{aligned}$$

The cumulative log odds ratio is the same for the  $c - 1$  possible collapsings of the response to binary,  $Y \leq j$  and  $Y > j$ .

### 3.3 Motivation of proportional odds models

A proportional odds model can be motivated by a latent continuous variable, of which will explain why distributions of  $Y$  at different settings of explanatory variables are stochastically ordered (Agresti 2010, chp. 3).

Let  $Y^*$  denote an underlying latent variable, which is an unobserved continuous variable assumed to underlie  $Y$  (Anderson & Philips 1981). For fixed values of explanatory variables  $\mathbf{x}$ , suppose  $Y^*$  has cumulative distribution function (cdf)  $G(y^* - \eta)$  where values of  $y^*$  varies around a location parameter  $\eta$ , such as a mean (Agresti 2012, chp. 7). The location parameter  $\eta$  depends on  $\mathbf{x}$  through  $\eta(\mathbf{x}) = \boldsymbol{\beta}'\mathbf{x}$  (Agresti 2010, chp. 3). Suppose that  $-\infty = \alpha_0 < \alpha_1 < \dots < \alpha_j = \infty$  are thresholds of the continuous scale such that the observed response  $Y$  satisfies

$$Y = j \quad \text{if} \quad \alpha_{j-1} < Y^* \leq \alpha_j$$

This would suggest that  $Y$  falls in category  $j$  when the latent variable falls in the  $j$ th interval of values and as such has the form

$$P(Y \leq j|\mathbf{x}) = P(Y^* \leq \alpha_j|\mathbf{x}) = G(\alpha_j - \boldsymbol{\beta}'\mathbf{x}).$$

The link function,  $G^{-1}$  the inverse of the cdf of  $Y^*$ , can be applied to  $P(Y \leq j|\mathbf{x})$  to obtain a linear predictor (Agresti 2010, chp. 3). That is,

$$G^{-1}[P(Y \leq j|\mathbf{x})] = \alpha_j - \boldsymbol{\beta}'\mathbf{x}$$

When  $G$  is the cdf of the standard logistic distribution, then  $G^{-1}$  is the logit link function, where  $G(\epsilon) = \exp^{\epsilon} / (1 + \exp^{\epsilon})$ . This shows that the logistic latent variable model implies the model for the observed response is,

$$\text{logit}[P(Y \leq j|\mathbf{x})] = \alpha_j - \boldsymbol{\beta}'\mathbf{x}$$

The proportional odds model is the result of this, where each cumulative probability has the same effects.

The latent variable motivation for the proportional odds model explains why distributions of  $Y$  at different settings of explanatory variables are stochastically ordered (Agresti 2010, chp. 3). This model is sensitive to location effects but not to the effects of the variability of  $Y$  as the explanatory variables changes (Agresti 2010, chp. 3).

### 3.4 Various models for repeated measurements

Studies often observe many responses within a subject/cluster. In practice, like the listening questionnaire, there are more than one observation for each subject. The responses to thirty-eight questions will be treated as repeated observations on a subject forming a cluster. Analyses of this questionnaire should take the dependency among responses into account, as responses within a student are likely to be more similar than two responses from different students.

Mainly, there are two types of approaches to model repeated ordinal categorical data – the marginal and the subject-specific approaches (Agresti 2010). For the marginal approach, we emphasise on the generalised estimating equations method (Liang & Zeger 1986). These two types of approaches will be used in an analysis of the listening clusters in a later section of this chapter.

### 3.4.1 Marginal approach

At observation  $Y_t$  in  $Y_1, Y_2, \dots, Y_T$  on a  $c$ -category scale, the marginal response distribution can be modelled using a  $c-1$  cumulative logits. Let  $Y_{it}$  be the multinomial response for subject  $i$  at time  $t$ . A model for  $Y_{it}$  has the form

$$\text{logit}[P(Y_{it} \leq j)] = \alpha_j + \boldsymbol{\beta}'_j \mathbf{x}_{it} \quad j = 1, \dots, c-1, t = 1, \dots, T, i = 1, \dots, n \quad (3.2)$$

When we replace  $\boldsymbol{\beta}_j$  with  $\boldsymbol{\beta}$ , this model takes the proportional odds form. The marginal model now has the form

$$\text{logit}[P(Y_{it} \leq j)] = \alpha_j - \boldsymbol{\beta}' \mathbf{x}_{it} \quad (3.3)$$

For marginal models, maximum likelihood fitting is computationally cumbersome and difficult for many datasets. This is especially the case for models with explanatory variables. An alternative to this is to use a multivariate generalisation of quasi-likelihood (Agresti 2012, chp. 11).

For a univariate response, the quasi-likelihood method specifies a model for  $\mu = E(Y)$  and specifies a variance function  $v(\mu)$  where  $\text{var}(Y)$  depends on  $\mu$ . This model applies to the marginal distribution for each  $Y_t$ .

This method requires a working correlation matrix to be selected for the responses. The estimates for each parameter via this approach are solutions to equations that take into account the mean and variance structure (Agresti 2010, chp. 9). These equations are called generalised estimating equations and this method is known as the GEE method. When the marginal model holds, the GEE estimates of the model parameters are valid even if the working correlation structure is misspecified.

A GEE approach for proportional odds models with repeated ordinal responses was proposed by Lipsitz et al. (1994) and Touloumis et al. (2013). For each pair of outcome categories, a working correlation matrix is selected for the pairs of repeated

observations (Agresti 2010, chp. 9).

Assume that MCAR holds (Rubin 1976). For each subject  $i$ , each observation  $Y_{it}$  can be denoted by a set of  $c - 1$  indicator variables. The  $j^{\text{th}}$  indicator variable shows whether  $Y_{it}$  falls in category  $j$ , where  $j = 1, 2, \dots, c - 1$ . This is defined as  $Y_{ijt} = I(Y_{it} = j)$  for  $j = 1, 2, \dots, c - 1$ , where  $I(A)$  denotes the indicator function of the event  $A$  Touloumis et al. (2013). Now let  $y_{ijt} = 1$  if observation  $t$  in subject  $i$  has outcome  $j$ , with  $y_{ijt} = 0$  otherwise. Let  $\mathbf{y}_i$  be the  $T(c - 1)$  binary indicator for subject  $i$  for the  $T$  observations. The covariance matrix  $\mathbf{V}_i$  for  $\mathbf{y}_i$  is a  $T(c - 1) \times T(c - 1)$  matrix. The covariance matrix for  $\mathbf{V}_{it}$  for the  $c - 1$  indicators for each  $Y_{it}$  is a  $(c - 1) \times (c - 1)$  matrix block on the main diagonal of  $\mathbf{V}_i$ . This means that the covariance matrix  $\mathbf{V}_{it}$  for  $y_{i1t}, y_{i,c-1,t}$  has entry  $v_{ijt} = P(Y_{ijt} = 1) \times [1 - P(Y_{ijt} = 1)]$  for the cell on the main diagonal in row  $j$  and column  $j$  and  $-P(Y_{ijh} = 1)P(Y_{ijt} = 1)$  for the cell in row  $h$  and column  $j$  where  $h \neq j$  (Agresti 2010, chp. 9).

The elements of  $\mathbf{V}_i$  that are remaining are  $Cov(Y_{iht}, Y_{ijs})$ , where  $s \neq t$ . These elements are not determined by the marginal multinomial covariances. For each pair  $(h, j)$  of outcome categories, the GEE approach uses a working correlation matrix for the pairs  $(Y_{is}, Y_{it})$  of subject observations. The working covariance matrix  $\mathbf{V}_i$  for  $\mathbf{y}_i$  specifies a pattern for  $Corr(Y_{ijt}, Y_{ih_s})$  for each pair of outcome categories  $(h, j)$  and each pair  $(s, t)$  of observations for a subject (Agresti 2010, chp. 9).

Let  $\boldsymbol{\mu}_i = E(\mathbf{y}_i)$  where this is a function of the model parameters  $\boldsymbol{\beta}$ . The generalised estimating equations of  $\boldsymbol{\beta}$  are

$$\mathbf{u}(\boldsymbol{\beta}) = \sum_{i=1}^n \mathbf{D}'_i \mathbf{V}_i^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i) = \mathbf{0}$$

where  $\mathbf{D}'_i = \partial \boldsymbol{\mu}_i / \partial \boldsymbol{\beta}$ .

For the covariance matrix, standard errors based on the assumed marginal and joint structure may not be valid. This can be the case when the actual structure is quite

different from the working structure. An example of this is when standard errors based on independence working correlations are strongly biased when the responses are strongly correlated (Agresti 2010, chp. 9). A more appropriate standard error results from an adjustment made by the GEE approach when using the empirical dependencies exhibited by the data. The standard errors based on the working correlation structure are updated using this empirical dependencies to provide more robust standard errors. These use a sandwich covariance matrix that is based on the product of three matrices.

By substituting  $\hat{\boldsymbol{\mu}}_i$  from the model fit in  $\mathbf{D}_i$  and  $\mathbf{V}_i$ , and replacing  $Cov(\mathbf{Y}_i)$  by the empirical covariance matrix of  $\mathbf{y}_i$ , the empirical adjusted sandwich covariance matrix for the GEE estimator  $\hat{\boldsymbol{\beta}}$  can be estimated. This is estimated by

$$\left[ \sum_{i=1}^n \mathbf{D}'_i \mathbf{V}_i^{-1} \mathbf{D}_i \right]^{-1} \left[ \sum_{i=1}^n \mathbf{D}'_i \mathbf{V}_i^{-1} Cov(\mathbf{Y}_i) \mathbf{V}_i^{-1} \mathbf{D}_i \right] \left[ \sum_{i=1}^n \mathbf{D}'_i \mathbf{V}_i^{-1} \mathbf{D}_i \right]^{-1}$$

Due to the computational simplicity of the GEE method compared to the maximum likelihood, it is often a more appealing method to use. However, as this approach does not completely specify a joint distribution, it does not have a likelihood function (Agresti 2012, chp. 11). As likelihood based methods are not available for inferencing, a Wald statistic can be used instead. A Wald statistic is constructed with the asymptotic normality of the estimators as well as their estimated covariance matrix (Agresti 2012, chp. 11). The Wald statistic is used in the Wald test for both model selection and model inferencing. This will be described in a later section of this chapter.

### 3.4.2 Subject-specific approach

An alternative approach to the marginal approach for clustered ordinal data is a subject-specific approach. For this approach, the linear predictor of a model contains subject effects terms.

Let  $Y_{it}$  denote the response for subject  $i$  at time  $t$ . The extension of equation (3.3) then has the form

$$\text{logit}[P(Y_{it} \leq j|u_i)] = u_i - \alpha_j - \boldsymbol{\beta}' \mathbf{x}_{it} \quad (3.4)$$

where  $u_i$  is the subject effects term. The subject effects term is unobserved and so is unknown. It takes the same value for each observation in a subject/cluster, but takes a different value for different clusters (Agresti 2012, chp. 12). Usually the assumption for  $u_i$  is that it varies from cluster to cluster according to a normal distribution with variance  $\sigma_u^2$ . The model is a type of random effect models.

When response variables are continuous (assumed to be normal distributed), we can treat the subject effect terms as either fixed or random in a general linear model. For binary responses, if the subject effect terms are treated as fixed in a logistic regression model, the unconditional maximum likelihood estimators for  $\boldsymbol{\beta}$  are not consistent (Ghosh 1995). It happens whenever the number of parameters grows at the same rate as the sample size (Neyman & Scott 1948). A similar problem occurs for proportional odds models (Liu & Agresti 1996). Because of it, we only consider random effect models in the subject-specific approach.

The overall intercept term is of the form  $u_i - \alpha_j$ . This allows the subjectivity of the choices of subjects for outcome categories by allowing the ordinal scale cutpoints to vary amongst the subjects (Wolfe & Firth 2002).

Parameter interpretations for a subject-specific model is different to a marginal model (Agresti 2012, chp. 12). A subject-specific model has conditional interpretations. There are two types of conditional interpretations, within-cluster effect and between-cluster effect.

The effect of an explanatory variable is a within-cluster effect when the coefficient of

the explanatory variable refers to the effect on the response from a subject (Agresti 2010, chp. 10). An example of this is in a study when comparing  $T$  treatments, treatment varies from observation to observation in the  $T$  many clusters of observations.

A between-cluster effect is when the coefficient of an explanatory variable is a consistent value among the observations in a cluster (Agresti 2010, chp. 10). For example, when each subject is a cluster, and the subjects gender may be the explanatory variable. The gender coefficient is the same for all observations for each subject, however, between clusters, the gender coefficient varies.

It is due to these two cluster effects that the random effects model are conditional models. Both the within-clusters and the between-clusters effects apply conditionally on the random effect value. This is different to the interpretation of the marginal model described previously, where the effects are averaged over all clusters.



### 3.5 Fuzzy biclustering method

Suppose we are not aware of the eight clusters described by Linguistics theory and just have the thirty-eight questions. We could use a fuzzy biclustering method based on proportional odds models to find the clusters for questions and for students simultaneously. That is, we can identify the groups of students or questions with similar responses.

Many cluster analysis methods for ordinal data use matrix decomposition and eigenvalues for dimension reduction which incorrectly treats the ordinal data as a continuous measurement (Kaufman & Rousseeuw 1990, Lewis et al. 2003). There are model based clustering methods as well. Majority of model based clustering methods deal with one-dimensional groupings and focus on responses that are either continuous or categorical with no ordering nature (Melnykov 2003).

To group the thirty eight listening questions and students, a cluster method proposed by Matechou et al. (2011) and Pledger & Arnold (2013) is used. We call the method the *fuzzy biclustering method*. This method is purely based on model fitting and uses the proportional odds model to simultaneously cluster the rows (e.g., students) and columns (e.g., questions) of ordered categorical response data.

Suppose that the rows come from a finite mixture with  $R$  row groups and the columns come from a finite mixture with  $G$  column groups. If cell  $i, k$  belongs to row group  $r$  and column group  $g$ , then under the proportional odds model

$$\text{logit}[P(Y_{ik} \leq j)] = \alpha_j - \mu_i - \beta_k$$

where  $\alpha_j$  is the  $j^{\text{th}}$  cut-off point, with  $\alpha_1 < \alpha_2 < \dots < \alpha_{c-1}$  and  $\mu_i, \beta_k$  are the effect of row  $i$  and column  $k$ , respectively, the model becomes

$$\text{logit}[P(Y_{ik} \leq j)] = \alpha_j - \mu_r - \beta_g$$

However, row and column group memberships are latent unobserved variables. We therefore define  $Z_{ir}$  and  $X_{kg}$ , indicator variables for group membership of row  $i$  in row group  $r$  are column  $k$  in column group  $g$ , respectively. The posterior probability that row  $i$  belongs to row group  $r$  is

$$E(Z_{ir}) = z_{ir}, \quad \sum_{r=1}^R z_{ir} = 1 \quad \forall i$$

The posterior probability that column  $k$  belongs to column group  $g$  is

$$E(X_{kg}) = x_{kg}, \quad \sum_{g=1}^G x_{kg} = 1 \quad \forall k$$

$z_{ir}$  and  $x_{kg}$  are obtained during the E-step of the Expectation-Maximisation (EM) algorithm as follows

$$\hat{z}_{ir} = \frac{\hat{\pi}_r \prod_{k=1}^p \sum_{g=1}^G \hat{\kappa}_g \sum_{j=1}^q \hat{\theta}_{rgj}^{I(y_{ik}=j)}}{\sum_{a=1}^R \hat{\pi}_a \prod_{k=1}^p \sum_{b=1}^G \hat{\kappa}_b \sum_{j=1}^q \hat{\theta}_{abj}^{I(y_{ik}=j)}}$$

and

$$\hat{x}_{kg} = \frac{\hat{\kappa}_g \prod_{i=1}^n \sum_{r=1}^R \hat{\pi}_r \sum_{j=1}^q \hat{\theta}_{rgj}^{I(y_{ik}=j)}}{\sum_{b=1}^G \hat{\kappa}_b \prod_{i=1}^n \sum_{a=1}^R \hat{\pi}_a \sum_{j=1}^q \hat{\theta}_{abj}^{I(y_{ik}=j)}}$$

where  $\hat{\pi}_r$  and  $\hat{\kappa}_g$  correspond to the estimated proportion of rows that belong to row group  $r$  and the estimated proportion of columns that belong to column group  $g$ , respectively,  $\sum_{r=1}^R \hat{\pi}_r = \sum_{g=1}^G \hat{\kappa}_g = 1$ . These proportions are estimated during the M-step of the algorithm by

$$\hat{\pi}_r = \sum_{i=1}^n \hat{z}_{ir} / n$$

and

$$\hat{\kappa}_g = \sum_{k=1}^p \hat{x}_{kg} / p$$

The fuzzy biclustering method was fitted to the listening questionnaire data with responses from 1,039 students and thirty eight questions. Each student represented a row and each question a column. We fitted various models with  $R = 3, 4, 5$  and

$G = 4, 5, 6$  using the R programme provided by Matechou et al. (2011). Using the model selection criterion AIC (Akaike information criterion), the model with  $R = 4$  and  $G = 5$  is the best. Because the detail of the analysis is beyond the focus of this thesis, we only report the result from the model with  $R = 4$  and  $G = 5$ . This means that 1,039 students can be divided into 4 row clusters, where the students within each row cluster are more similar than students in other row clusters. The same goes for the five column clusters, where the questions within each column cluster are more similar than the questions in the other column clusters.

The columns are split into five clusters and they consist of the following questions:

1. Questions 5, 18 and 19
2. Questions 21, 37 and 38
3. Questions 14, 17, 25, 26 and 32
4. Questions 2, 6, 7, 8, 13, 16, 27, 28, 29, 31 and 35
5. Questions 1, 3, 4, 9, 10, 11, 12, 15, 20, 22, 23, 24, 30, 33, 34 and 36.

The rows are split into four clusters where cluster one consists of 105 students, cluster two consists of 381 students, cluster three consists of 104 students and cluster four consists of 449 students.

### 3.6 Wald test

As the GEE method is not a likelihood-based method, we cannot use the likelihood ratio test for the model effects. In this section, we will review and describe the Wald test that will be used for model selection and to make inferences about the questions in each strategy. The Wald test is used to test the significance and difference of an explanatory variable in a statistical model.

If for a particular explanatory variable, or a group of explanatory variables, the Wald test is significant, then we would conclude that the parameter associated with these variables is not zero (Agresti 2012). This would mean that the variable should be included in the model. If the Wald test is not significant, this suggests that the parameter associated with these variables is zero. We can conclude from this that the explanatory variable can be omitted from the model.

We will first illustrate the Wald test with a single parameter. Let  $\gamma$  denote an arbitrary parameter, that has null hypothesis of  $H_0 : \gamma = \gamma_0$ . Let  $\hat{\gamma}$  denote the estimator, and  $SE$  be the standard error of  $\hat{\gamma}$ . When  $H_0$  is true, the test statistic

$$z = (\hat{\gamma} - \gamma_0)/SE$$

has approximately a standard normal distribution and  $z^2$  has approximately a chi-squared distribution with  $df = 1$  (Agresti 2007, p. 11). This test statistic is called the Wald statistic.

The multivariate extension for the Wald test, which will be used for model selection and inferencing, has test statistic

$$Z^2 = (\hat{\gamma} - \gamma_0)^T [cov(\hat{\gamma})]^{-1} (\hat{\gamma} - \gamma_0) \quad (3.5)$$

Since  $\hat{\gamma}$  has asymptotic multivariate normal distribution, this implies an asymptotic chi-squared distribution for  $Z$ . The  $df$  equals the rank of  $cov(\hat{\gamma})$ , which is the

number of non-redundant parameters in  $\gamma$  (Agresti 2012, p. 11).

The Wald test will be used for the proportional odds repeated measures models in the backward elimination model selection.

### Model Selection

Model selection methods aim at determining submodels with parameters that fit the data adequately. We can do the Backward elimination model. This method starts with the full model, where it contains all explanatory variables, and then gets tested against a submodel. The submodel is obtained by removing parameters associated with certain variables. At each stage of the selection, it selects the parameter for which its removal has the least damaging effect on the model. This procedure is stopped when any further removal of a term leads to significantly poorer fit (Agresti 2012, p. 214).

Statistical significance should not be the only criterion for including a term in a model. It is reasonable to include a variable that is the main focus of the study and report its estimated effects, even if it is not statistically significant (Agresti 2012, p. 214).

For each of the eight strategies, the main focus is on testing whether the questions within the strategy are significantly different by controlling other factors. For each model, the question, the primary school, gender and ethnicity of the students are used as explanatory variables. As all three factor interactions for our data are not significant, we demonstrate the backward elimination model selection starting from a model with two factor interactions containing the question term. Consider the model with  $Q + C + G + H + Q : C + Q : G + Q : H$ , where  $Q$  is the question term,  $C$  is the ethnicity term,  $G$  is the gender term and  $H$  is the school term.

For model selection, the highest order terms for each variable will be tested first. It would not be appropriate to remove a main effect term if the model has interactions with that term (Agresti 2012, p. 214).

For all our models, we first test to remove a two factor interaction from the complex model. Each removal of a term is tested until further deletion of a term leads to a significantly poorer fit or the question term is the last term in the model to test. Because the question effects within each strategy are the main focus of this research, if they are not statistically significant, they will still remain a part of the final model.

We will use the Self-initiation strategy as an example of model selection for the model  $Q + C + G + H + Q : C + Q : G + Q : H$ .

The full model for the GEE method to the proportional odds model has the form

$$\text{logit}[P(Y_{it} \leq j)] = \alpha_j - \beta'^Q \mathbf{x}_i^Q - \beta'^C \mathbf{x}_i^C - \beta'^G \mathbf{x}_i^G + \beta'^H \mathbf{x}_i^H - \beta'^{QC} \mathbf{x}_{it}^{QC} - \beta'^{QG} \mathbf{x}_{it}^{QG} - \beta'^{QH} \mathbf{x}_{it}^{QH}$$

where  $\alpha_j$  are the cutpoints and  $j = 1, 2, 3, 4$ . Let  $\mathbf{x}_i^Q$  denote the questions within the Self-initiation strategy,  $\mathbf{x}_i^C$  denote the ethnicity for student  $i$ ,  $\mathbf{x}_i^G$  denote the gender for student  $i$  and  $\mathbf{x}_i^H$  denote the school for student  $i$ . The full model includes the interaction terms between the question and ethnicity, question and gender, and question and school. When this model is fitted, the Wald test is used to see which of the explanatory variables are necessary for the model to fit well.

Table 3.1: Stages of model selection for Self-initiation strategy using the GEE approach

Stage	Terms	Wald Statistic	df	P-value
1	$Q \times G$	3.8	3	0.29
	$Q \times C$	1.1	3	0.78
	$Q \times H$	0.82	15	1.0
2	$Q \times G$	6.1	3	0.11
	$Q \times C$	2.6	3	0.46
	H	5.3	5	0.38
3	$Q \times G$	6.8	3	0.08
	C	6.8	1	0.009
	H	5.3	5	0.38
4	$Q \times G$	6.6	3	0.09
	C	16.3	1	< 0.0001
5	G	2.1	1	0.15
	C	16.3	1	< 0.0001
6	C	17.5	1	< 0.0001

Table 3.1 shows the explanatory terms at each stage of model selection for the GEE approach. At stage 1, the interaction term between the questions and school,  $Q : H$ , show the highest level of no evidence against the null hypothesis. This suggests the interaction term may be removed from the model. At each stage, a term is removed with the least evidence (highest p-value), until all terms are significant at the 5% level. Exception is given to the question term  $Q$ . At the end of this process, the ethnicity term is the only term that shows significant evidence against the null hypothesis of a difference. The model best fitted to the data for the Self-initiation strategy has the form

$$\text{logit}[P(Y_{it} \leq j)] = \alpha_j - \beta^Q \mathbf{x}_t^Q - \beta^C \mathbf{x}_i^C$$

where the gender, school and interaction terms are not included.

The same method of model selection is used for the subject-specific (random effects) approach to the proportional odds model. The full model for this approach has the form

$$\text{logit}[P(Y_{it} \leq j)|u_i] = u_i - \alpha_j - \boldsymbol{\beta}'^Q \mathbf{x}_t^Q - \boldsymbol{\beta}'^C \mathbf{x}_i^C - \boldsymbol{\beta}'^G \mathbf{x}_i^G + \boldsymbol{\beta}'^H \mathbf{x}_i^H - \boldsymbol{\beta}'^{QC} \mathbf{x}_{it}^{QC} - \boldsymbol{\beta}'^{QG} \mathbf{x}_{it}^{QG} - \boldsymbol{\beta}'^{QH} \mathbf{x}_{it}^{QH}$$

where  $i$  is the  $i^{\text{th}}$  student and  $u_i$  is the student random effects term.



## 3.7 Analysis

Techniques and methods have been discussed in Section 3.4 to check whether the questions within strategies suggested by Linguistics theory are similar. If they are similar, there should be no differences between questions by controlling on other possible factors. The results in applying these techniques are discussed in this section. Subsection 3.7.1 describes how the dataset was coded to be able to apply the approaches described in section 3.4 to the data. Subsection 3.7.2 looks at differences between questions within strategies using the marginal approach (GEE method) and the subject-specific approach (random effect models) applied to proportional odds models.

These approaches discussed in this section are fitted using the statistical software, R. Appendix B shows how to use the `ordLORgee` function (Touloumis 2013, Touloumis et al. 2013) for the GEE method in the marginal approach and `clmm` function (Christensen 2013) for the subject-specific approach.

### 3.7.1 Data coding

The following steps describe how the data were coded to allow the application of the different approaches of the proportional odds model to the data.

1. Every student is allocated an ID number (1 to 1,039).
2. Each ID number is replicated thirty eight times. This is because there are thirty eight questions in the questionnaire. This will give a total of 39,482 rows.
3. The response variables are computed where each student's responses to the questions are transposed. This is so all of the answers are in one column. Each ID number will therefore have thirty eight responses.

4. Eight columns are computed to represent the questions for the eight listening strategies. For example, a column was created for the Self-initiation strategy questions. This strategy has four questions. For each student, responses for questions 11, 12, 14 and 18 are allocated a value 1, 2, 3 and 4 respectively. All other questions are allocated a zero. This is done for the remaining seven strategies.
5. The gender, ethnicity and school of the student is then added to the dataset. Due to few observations for Malay and Indian students, we only consider two ethnicity groups. The first group includes Chinese and Eurasian ethnicity, and will be called ‘Chinese’ as majority of the students in this group are of Chinese ethnicity. The second group includes Malay, Indian and Other ethnicity, and will be called ‘Other’. The school has 8 levels, representing 8 schools.

### 3.7.2 Linguistics listening strategies

The Self-initiation strategy has four questions:

Q11: To improve my listening in English, I watch English TV programmes.

Q12: I look for opportunities to listen in English.

Q14: I try to find out how to improve my listening in English.

Q18: When I’m free, I find interesting things to listen to in English (for example, TV, radio, etc).

The best fit model for the Self-initiation strategy using the backward elimination model selection, includes the question, gender and ethnicity terms. There are no significant interactions between question and gender ( $Q : G$ ); question and school ( $Q : H$ ); and question and ethnicity ( $Q : C$ ). Table 3.2 shows how similar the questions are within the strategy for two approaches. There is strong evidence against

the null hypothesis of no difference. This means there are strong differences between questions within this strategy.

Table 3.2: Self-initiation strategy

Model	Wald Statistic	Degree of Freedom	P-value
GEE approach	14.31	3	0.0025
Subject-specific approach	13.62	3	0.0035

The Planning strategy has three questions:

Q13: When I listen, I use what I have already heard to think about what I will hear next.

Q15: Before I listen to something, I ask myself whether it is important to me.

Q23: Before I start listening, I decide if I need to pay attention to details or to the main idea.

Table 3.3 shows the question effects for the Planning strategy. Both the GEE approach and subject-specific approach shows no evidence to reject the null hypothesis of no difference between questions. This suggests there are no significant differences between the questions within this strategy.

Table 3.3: Planning strategy

Model	Wald Statistic	Degree of Freedom	P-value
GEE approach	1.62	2	0.4439
Subject-specific approach	1.41	2	0.4944

The Monitoring and Evaluation strategy has five questions:

Q5: I compare what I am hearing with what I have already heard to make sure I understand correctly.

Q9: During or after listening, I ask myself whether the information is the same as what I already know.

Q21: When I have a problem in listening, I decide whether I should pay more attention to it.

Q32: If I have a problem in understanding, I quickly decide whether I should continue or listen again.

Q36: During or after listening, I check how much I have understood.

Table 3.4: Monitoring and Evaluation strategy

Model	Interaction	Wald Statistic	Degree of Freedom	P-value
GEE approach	Chinese	15.98	4	0.0031
	Other	42.71	4	< 0.0001
Subject-specific approach	Chinese	15.50	4	0.0038
	Other	44.38	4	< 0.0001

Table 3.4 shows the question effects for the Monitoring and Evaluation strategy. The GEE approach and subject-specific approach had the  $Q : E$  interaction term significant and so there is a difference between the two ethnicity groups for the question effects. Both approaches shows significant evidence to reject the null hypothesis of no difference between questions for both ethnicity groups. This suggests there are significant differences between the questions within the strategy.

The Perceptual Processing strategy has four questions:

Q8: When I listen, I repeat the words or phrases I can understand.

Q24: If I can't understand a word or a phrase, I repeat it to myself.

Q34: When I listen, I repeat the pronunciation of the words I have heard.

Q37: When I listen, I pay attention to every word that is said.

Table 3.5: Perceptual Processing strategy

Model	Wald Statistic	Degree of Freedom	P-value
GEE approach	110.13	3	< 0.0001
Subject-specific approach	109.07	3	< 0.0001

Table 3.5 shows the question effects for the Perceptual Processing strategy. Both the GEE approach and subject-specific approach shows significant evidence to reject the null hypothesis of no difference between questions. This suggests there are significant differences between the questions within the strategy.

The Inferencing strategy has seven questions:

Q1: When I don't understand something, I use my knowledge about the English language to guess.

Q4: When I don't understand something, I use what I have already heard to guess it.

Q17: When I don't understand something, I make several guesses.

Q26: When I don't understand something, I use my general knowledge to make a guess.

Q27: When I listen, I use my knowledge about the English language to think about what I will hear next.

Q31: When I don't understand something, I quickly decide whether I should continue or listen again.

Q33: When I don't understand something, I use my knowledge about the topic to guess.

Table 3.6: Inferencing strategy

Model	Wald Statistic	Degree of Freedom	P-value
GEE approach	1.31	6	0.9709
Subject-specific approach	1.51	6	0.9587

Table 3.6 shows the question effects for the Inferencing strategy. Both the GEE approach and subject-specific approach shows no evidence to reject the null hypothesis of no difference between questions. This suggests there are no significant differences between the questions within this strategy.

The Prediction strategy has five questions:

Q6: When I listen, I use my knowledge about text type (for example, story, report, etc) to think about what I will hear next.

Q7: Before I listen to something, I think about the main idea of what I am going to hear.

Q13: When I listen, I use what I have already heard to think about what I will hear next.

Q29: When I listen, I use my knowledge about the topic to think about what I will hear next.

Q35: When I listen, I use my general knowledge to think about what I will hear next.

Table 3.7: Prediction strategy

Model	Wald Statistic	Degree of Freedom	P-value
GEE approach	0.01	4	1.0
Subject-specific approach	0.04	4	0.9998

Table 3.7 shows the question effects for the Prediction strategy. Both the GEE approach and subject-specific approach shows no evidence to reject the null hypothesis of no difference between questions. This suggests there are no significant differences between the questions within this strategy.

The Utilisation strategy has five questions:

Q16: I try to find problems with how the speaker presented an idea so that I can avoid the same problems.

Q19: When I listen, I try to remember useful words and phrases so that I can use them.

Q20: I try to connect what I heard with my own experience.

Q22: After I finish listening, I make a summary in my mind about what I heard.

Q30: After I finish listening, I use my own words to retell what I heard in my mind.

Table 3.8: Utilisation strategy

Model	Wald Statistic	Degree of Freedom	P-value
GEE approach	17.89	4	0.0013
Subject-specific approach	19.87	4	0.0005

Table 3.8 shows the question effects for the Utilisation strategy. Both approaches shows significant evidence to reject the null hypothesis of no difference between questions. This suggests there are significant differences between the questions within the strategy.

Lastly, the Socio-affective strategy has five questions:

Q2: When I have a problem in listening, I ask my family members for help.

Q10: When I have a problem in listening, I ask my friends for help.

Q25: I tell myself to enjoy listening to English.

Q28: When I have a problem in listening, I ask my teachers for help.

Q38: I tell myself not to worry when I listen to English.

Table 3.9: Socio-affective strategy

Model	Wald Statistic	Degree of Freedom	P-value
GEE approach	63.66	4	< 0.0001
Subject-specific approach	56.66	4	< 0.0001

Table 3.9 shows the question effects for the Socio-affective strategy. Both approaches shows significant evidence to reject the null hypothesis of no difference between questions. This suggests there are significant differences between the questions within the strategy.

For Self-initiation, Utilisation, Monitoring and Evaluation, Perceptual Processing and Socio-affective strategies, there is significant evidence of a difference between questions. The following tables (Tables 3.10 - 3.15) will explore further showing which questions, if any, amongst these strategies is similar to one another.



Table 3.10 shows the pairwise question effects within the Self-initiation strategy for the different approaches to the proportional odds model. Following the notations in Model (3.6), the table shows  $\beta_k^Q - \beta_{k'}^Q$  for  $k \neq k' = 11, 12, 14, 18$ . Notice that  $\beta_{11}^Q = 0$ , because level 1 is the reference level in the R output (shown in Appendix B). For both approaches there appears to be no evidence of a difference between questions 11 and 12. Question 11 and 12 ask students about the opportunities taken to improve listening in English. It is reasonable for there to be no evidence of a difference between these two questions, as both are asking a similar question. All other pairwise questions have significant evidence that the pairwise questions are different.

Table 3.10: Self-initiation strategy question effects  
GEE approach

	11	12	14	18
11	-	0.0527	-0.2356(***)	-0.4409(***)
12	-	-	-0.2883(***)	-0.4936(***)
14	-	-	-	-0.2053(***)
18	-	-	-	-
Subject-specific approach				
	11	12	14	18
11	-	0.0718	-0.2954(***)	-0.5145(***)
12	-	-	-0.3672(***)	-0.5863(***)
14	-	-	-	-0.2191(***)
18	-	-	-	-

1

---

<sup>1</sup>\*\*\* : significant at the 0.1% level

Table 3.11 shows the pairwise question effects for the Utilisation strategy from the proportional odds model with random effects and GEE model. Following the notations in Model (3.6), the table shows  $\beta_k^Q - \beta_{k'}^Q$  for  $k \neq k' = 6, 7, 13, 29, 35$ .

Table 3.11: Utilisation strategy question effects  
Subject-specific approach

	6	7	13	29	35
6	-	0.8524(***)	0.0570	0.1260	0.1281
7	-	-	-0.7954(***)	-0.7264(***)	-0.7243(***)
13	-	-	-	0.0690	0.0711
29	-	-	-	-	0.0021
35	-	-	-	-	-

GEE approach

	6	7	13	29	35
6	-	-0.6804(***)	-0.0462	-0.1004	-0.0981
7	-	-	0.6342(***)	0.5800(***)	0.5823(***)
13	-	-	-	-0.0542	-0.0519
29	-	-	-	-	0.0023
35	-	-	-	-	-

Both approaches show there is no evidence of a difference between questions:

- 6 and 13,
- 6 and 29,
- 6 and 35,
- 13 and 29,
- 13 and 35, and
- 29 and 35.

The remaining pairwise questions show extremely strong evidence that there is a difference between them.

Table 3.12: Monitoring and Evaluation strategy question effects for Chinese students

		GEE approach				
	5	9	21	32	36	
5	-	0.1821(*)	0.0129	0.3778(***)	0.4569(***)	
9	-	-	-0.1692(*)	0.1957(*)	0.2748(***)	
21	-	-	-	0.3649(***)	0.4440(***)	
32	-	-	-	-	0.0791	
36	-	-	-	-	-	
		Subject-specific approach				
	5	9	21	32	36	
5	-	-0.2392(*)	-0.0045	-0.4476(***)	-0.5332(***)	
9	-	-	0.2347(*)	-0.2084(*)	-0.2940(***)	
21	-	-	-	-0.4431(***)	-0.5287(***)	
32	-	-	-	-	-0.0856	
36	-	-	-	-	-	

2

For the Monitoring and Evaluation strategy, there are significant interaction effects between the question and ethnicity. That is, for the Chinese group, the difference between questions is not the same as the one for the Other group.

Table 3.12 shows the pairwise question effects for the Monitoring and Evaluation strategy from the GEE approach and subject-specific approach to the proportional odds model for Chinese group. Following the notations in Model (3.6), the table shows  $(\beta_k^Q - \beta_{k'}^Q)$  with  $k \neq k' = 5, 9, 21, 32, 36$  for the Chinese group. Notice that

---

<sup>2</sup>\* : significant at the 5% level

the level 1 (Chinese) is the reference level. Similarly,  $\beta_5^Q = 0$ . Both models show no evidence of a difference between pairwise questions 5 and 21, and 32 and 36. All other pairwise questions show evidence of a difference between them.

Table 3.13: Monitoring and Evaluation strategy question effects for Other students  
GEE approach

	5	9	21	32	36
5	-	0.5638(***)	0.4504(***)	0.9141(***)	0.7153(***)
9	-	-	-0.1134	0.3503(***)	0.1515
21	-	-	-	0.4637(***)	0.2649(***)
32	-	-	-	-	-0.1988(**)
36	-	-	-	-	-
Subject-specific approach					
	5	9	21	32	36
5	-	-0.6968(***)	-0.5321(***)	-1.0935(***)	-0.8600(***)
9	-	-	0.1647(*)	-0.3967(***)	-0.1632(*)
21	-	-	-	-0.5614(***)	-0.3279(***)
32	-	-	-	-	0.2335(**)
36	-	-	-	-	-

3

Table 3.13 shows the pairwise question effects for the Monitoring and Evaluation strategy for the Other ethnicity group from the GEE approach and subject-specific approach to the proportional odds model. Following the notations in Model (3.6), the table shows  $(\beta_k^Q - \beta_{k'}^Q) + (\beta_k^{QC} - \beta_{k'}^{QC})$  with  $k \neq k' = 5, 9, 21, 32, 36$  for the Other group. These two models show different results. The subject-specific approach shows that all pairwise question effects have evidence of a difference between the questions. There are no pairwise questions that are similar to each other. The GEE approach shows that there is no evidence of a difference between pairwise questions

---

<sup>3</sup>\*\* : significant at the 1% level

9 and 21, and 9 and 36. The remaining pairwise questions all show strong evidence of a difference between the questions.

Table 3.14: Perceptual Processing strategy question effects  
GEE approach

	8	24	34	37
8	-	-0.3683(***)	-0.2520(***)	-1.1427(***)
24	-	-	0.1163	-0.7744(***)
34	-	-	-	-0.8907(***)
37	-	-	-	-
Subject-specific approach				
	8	24	34	37
8	-	0.4371(***)	0.2988(***)	1.3786(***)
24	-	-	-0.1383	0.9415(***)
34	-	-	-	1.0798(***)
37	-	-	-	-

Table 3.14 shows the pairwise question effects for the Perceptual Processing strategy for both approaches to the proportional odds model. Following the notations in Model (3.6), the table shows  $\beta_k^Q - \beta_{k'}^Q$  for  $k \neq k' = 8, 24, 34, 37$ . They show the same result where there is no significant evidence of a difference between pairwise questions 24 and 34. All other pairwise question effects show extremely strong evidence of a difference between questions.

Table 3.15: Socio-affective strategy question effects  
GEE approach

	2	10	25	28	38
2	-	-0.2492(***)	0.2141(***)	-0.6648(***)	-0.9856(***)
10	-	-	0.4633(***)	-0.4156(***)	-0.7364(***)
25	-	-	-	-0.8789(***)	-1.1997(***)
28	-	-	-	-	-0.3208(***)
38	-	-	-	-	-
Subject-specific approach					
	2	10	25	28	38
2	-	0.2762(***)	-0.2550(***)	0.7502(***)	1.1269(***)
10	-	-	-0.5312(***)	0.4740(***)	0.8507(***)
25	-	-	-	1.0052(***)	1.3819(***)
28	-	-	-	-	0.3767(***)
38	-	-	-	-	-

Table 3.15 shows the pairwise question effects for the Socio-affective strategy for both approaches to the proportional odds model. Following the notations in Model (3.6), the table shows  $\beta_k^Q - \beta_{k'}^Q$  for  $k \neq k' = 2, 10, 25, 28, 38$ . All pairwise question effects have extremely strong evidence of a difference. No pairwise questions appear similar.

# Chapter 4

## Structural equation models

### 4.1 Structural equation models

Structural equation models (SEM) is a method that takes a hypothesis testing approach to the analysis of a structural theory. The structural theory is generally based on some principle and represents “causal” processes that generate observations on multiple variables (Bentler 1988). Two important attributes of this method is that (a) the causal processes under study are represented by a series of structural equations and (b) that these structural relations can be modelled as a diagram to allow a clearer concept of the theory that is being studied (Byrne 2006, chp. 1).

To determine to what extent the hypothesised model is consistent with the data, it can be tested in a simultaneous analysis of the system of variables. If the goodness-of-fit of the model is adequate, it suggests that the assumed relations among the variables is plausible. Otherwise, if it is not adequate, plausibility of such relationships is rejected (Byrne 2006, chp. 1).

Key aspects of SEM can be summarised as the following (Byrne 2006, chp .1):

- Takes a confirmatory (hypothesis driven) approach to data analysis
- Provides explicit estimates of error variance parameters
- Incorporates both latent and observed variables
- Allows the modelling of multivariate relations, estimating point and/or interval indirect effects.

Before describing SEM models, concepts associated with the method are explained in more detail.

## 4.2 Basic concepts used in structural equation models

In this section, concepts associated with the SEM model are described. Concepts include latent and observed variables, exogenous and endogenous latent variables, factor analytic models, ordinal variables and polychoric correlation.

### 4.2.1 Latent and observed variables

Researchers are often interested in studying theoretical concepts that cannot be observed directly. These concepts are known as latent variables or factors. An example of latent variables are the eight listening strategies being analysed in this thesis.

As latent variables are not observed directly, they cannot be measured directly. As a result, the latent variables must be defined in terms of the behaviour the researcher believes represents it. As such an observed variable is linked to the latent variable. They serve as indicators of the underlying concepts that they are



presumed to represent. This enables the measurement of the unobserved variable possible (Byrne 2006, chp. 1).

In the data that we are analysing, the thirty eight questions answered in the survey are the observed variables and the eight listening strategies are the unobserved latent variables.

### **4.2.2 Exogenous and endogenous latent variables**

Here we distinguish between latent variables that are exogenous and endogenous. Exogenous latent variables are the same as independent variables. They can cause variation in the values of the other latent variables in the model. These variations are not explained by the model. Rather, they are considered to be influenced by other factors external to the model (Byrne 2006, chp.1). Variables such as gender and age are examples of these external factors.

Endogenous latent variables are the same as dependent variables. They are influenced by the exogenous latent variables in the model, either directly or indirectly. The variations in the values of endogenous variables are said to be explained by the model. This is because all of the latent variables that influence the endogenous variables are included in the model specification (Byrne 2006, chp. 1).

### **4.2.3 The factor analytic model**

Factor analysis is a well known statistical procedure for investigating relationships between sets of observed and unobserved variables. This procedure allows the examination of the covariance among a set of observed variables. This enables information to be gathered on their underlying latent concepts.

There are two basic types of factor analytic models: exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). Both these models concentrate on

how and to what extent the observed variables are linked to their underlying latent variables. In this thesis we will concentrate on the CFA model.

#### 4.2.4 Ordinal variables

Let  $y_1, y_2, \dots, y_p$  be  $p$  ordinal variables. It is assumed that there is a continuous variable  $y_i^*$  *underlying* the ordinal variable  $y_i$  (Muthén 1984, Lee et al. 1990, Jöreskog 1990). This continuous variable  $y_i^*$  represents the attitude underlying the ordered responses to  $y_i$  and is assumed to have a range from  $-\infty$  to  $+\infty$  (Yang-Wallentin et al. 2010). Notice that the notations used in this section are slightly different from the notations in the proportional odds model. Because this chapter illustrates a different approach, we decided to use the notations consistent with Yang-Wallentin et al. (2010).

The underlying variable  $y_i^*$  is unobservable and the ordinal variable  $y_i$  is observed. For an ordinal variable  $y_i$  with  $m_i$  categories, the association between the ordinal variable  $y_i$  and the underlying variable  $y_i^*$  is

$$y_i = c \quad \iff \quad \tau_{c-1}^{(i)} < y_i^* < \tau_c^{(i)} \quad c = 1, 2, \dots, m_i$$

where

$$\tau_0^{(i)} = -\infty, \quad \tau_1^{(i)} < \tau_2^{(i)} < \dots < \tau_{m_i-1}^{(i)}, \quad \tau_{m_i}^{(i)} = +\infty$$

are the threshold parameters. For variable  $y_i$  with  $m_i$  categories, there are  $m_i - 1$  strictly increasing threshold parameters  $\tau_1^{(i)}, \tau_2^{(i)}, \dots, \tau_{m_i-1}^{(i)}$  (Yang-Wallentin et al. 2010).

As the only ordinal information available is from  $y_i$ , the distribution of  $y_i^*$  is determined only up to a monotonic transformation. Let  $y_i^*$  have a standard normal distribution with density function  $\psi(\cdot)$  and distribution function  $\Psi(\cdot)$ . Then the

probability,  $\pi_c^{(i)}$ , of the response,  $y_i$ , being in category  $c$  is

$$\pi_c^{(i)} = Pr[y_i = c] = Pr[\tau_{c-1}^{(i)} < y_i^* < \tau_c^{(i)}] = \int_{\tau_{c-1}^{(i)}}^{\tau_c^{(i)}} \psi(u) du = \Psi(\tau_c^{(i)}) - \Psi(\tau_{c-1}^{(i)})$$

for  $c = 1, 2, \dots, m_{i-1}$ , so that

$$\tau_c^{(i)} = \Psi^{-1}(\pi_1^{(i)} + \pi_2^{(i)} + \dots + \pi_c^{(i)})$$

where  $\Psi^{-1}$  is the inverse of the standard normal distribution function. The quantity  $(\pi_1^{(i)} + \pi_2^{(i)} + \dots + \pi_c^{(i)})$  is the probability of a response being in category  $c$  or lower (Yang-Wallentin et al. 2010).

The probabilities  $\pi_c^{(i)}$  are unknown population quantities, but can be estimated by the corresponding proportions  $p_c^{(i)}$  of responses in category  $c$  on variable  $y_i$ . Thus, estimates of the thresholds can be obtained as

$$\tau_c^{(i)} = \Psi^{-1}(p_1^{(i)} + p_2^{(i)} + \dots + p_c^{(i)}) \quad c = 1, 2, \dots, m_{i-1} \quad (4.1)$$

The quantity  $(p_1^{(i)} + p_2^{(i)} + \dots + p_c^{(i)})$  is the proportion of observations in the data responding in category  $c$  or lower on variable  $y_i$  (Yang-Wallentin et al. 2010).

### 4.2.5 Polychoric correlation

The polychoric correlation coefficient matrix is largely used to replace the covariance matrix. This helps with the estimation of the structural equation model parameters. The matrix estimates the linear relationship between two unobserved continuous variables underlying two ordinal variables (Flora & Curran 2004).

Let  $y_i$  and  $y_j$  be two ordinal variables with  $m_i$  and  $m_j$  categories respectively. Their marginal distribution in the data is represented by a contingency table

$$\begin{pmatrix} n_{11}^{(ij)} & n_{12}^{(ij)} & \dots & n_{1m_j}^{(ij)} \\ n_{21}^{(ij)} & n_{22}^{(ij)} & \dots & n_{2m_j}^{(ij)} \\ \vdots & \vdots & \ddots & \vdots \\ n_{m_i1}^{(ij)} & n_{m_i2}^{(ij)} & \dots & n_{m_i m_j}^{(ij)} \end{pmatrix}$$

where  $n_{ab}^{ij}$  is the number of observations in the data in category  $a$  on variable  $y_i$  and in category  $b$  on variable  $y_j$ . The underlying variable  $y_i^*$  and  $y_j^*$  are assumed to be bivariate normal with zero means, unit variance and with correlation  $\rho_{ij}$ , the polychoric correlation (Yang-Wallentin et al. 2010).

Let  $\tau_1^{(i)}, \tau_2^{(i)}, \dots, \tau_{m_i-1}^{(i)}$  be the thresholds (as described in section 4.2.4) for variable  $y_i^*$  and let  $\tau_1^{(j)}, \tau_2^{(j)}, \dots, \tau_{m_j-1}^{(j)}$  be the thresholds for variable  $y_j^*$ . The polychoric correlations can be estimated by maximising the log-likelihood of the multinomial distribution (see Olsson (1979) for more detail). Maximising the log-likelihood gives the sample polychoric correlation denoted as  $r_{ij}$ .

The polychoric correlation can be estimated by a two step procedure (Olsson 1979). In the first step, the thresholds are estimated from the univariate marginal distributions by equation 4.1. The second step involves the estimation of the polychoric correlations from the bivariate marginal distributions by maximising log-likelihood for given thresholds (Yang-Wallentin et al. 2010).

### 4.3 Confirmatory factor analysis with ordinal variables

Confirmatory factor analysis (CFA) is a theory or hypothesis driven method. It is often used when there is some knowledge of the underlying latent variable structure. This can be based on knowledge of theory or empirical research. Using this knowledge, relationships between the observed measures and the underlying factors can be tested. As this method is solely based on the link between factors (latent variables) and their observed variables, it represents what is termed a *measurement model* (Byrne 2006, chp .1).

The main task in model testing, is to establish the goodness-of-fit between the hypothesised model and the sample data. As such, the structure of the hypothesised model is fitted to the sample data and then tested to see how well the observed data fit the constrained structure. The likelihood that the hypothesised structure will be a perfect fit to the observed data is highly unlikely. As a result, there is a differential between the two that is known as the *residual* (Byrne 2006, chp. 1). The model fitting process is summarised as follows:

$$Data = Model + Residual$$

where:

- *Data* is the score measurements related to the observed variables.
- *Model* is the hypothesised structure linking the observed variables to the latent unobserved variables.
- *Residual* is the difference between the hypothesised model and the observed data.

The model can be expressed diagrammatically and mathematically through a set of equations. Both shows how the observed and unobserved latent variables are

related to each other. Once the structure of the model is specified, the plausability of the model is tested based on the sample data that contains all observed variables in the model (Byrne 2006, chp. 1).

In the next two subsections, we will describe the structure of the hypothesised model pictorially (path diagrams) and describe the structure of the hypothesised model mathematically (matrix form).

### 4.3.1 CFA model represented diagrammatically

CFA models can be represented diagrammatically by using path diagrams. Path diagrams allow the structure of the hypothesised model to be displayed visually. It allows the relationship between the observed and unobserved latent variables being reviewed to be portrayed pictorially.

Before we display the structure of a hypothesised model, we will first describe the different symbols used in the diagram followed by the different components of the diagram.

#### Symbols

The hypothesised model is schematically represented using four particular geometric symbols. This includes:

- circles (or ellipses)
- squares (or rectangles)
- single headed arrows
- double headed arrows

The circles (or ellipses) represent the unobserved latent variables. The squares represent the observed variables. The single headed arrows represent the impact of one variable on another and the double headed arrows represent the covariances or correlations between pairs of variables (Byrne 2006, chp. 1).

### **Components**

In addition to the four symbols, there are also four important components in this process. Figure 4.1 shows the general structure of a path diagram for a CFA model with three unobserved latent variables and nine observed variables, which depicts the four components. The letters on this figure represent these important components of the process. They are described as the following (Byrne 2006, chp. 1):

- a. single headed arrow from a circle to a square is a path coefficient for regression of an observed variable onto an unobserved latent variable.
- b. single or double headed arrow from a circle to a circle depicts a path coefficient for regression of one unobserved variable onto another unobserved variable.
- c. single headed arrow to a square is the measurement error associated with the observed variable.
- d. single headed arrow to a circle is the residual error in the prediction of an unobserved latent variable.

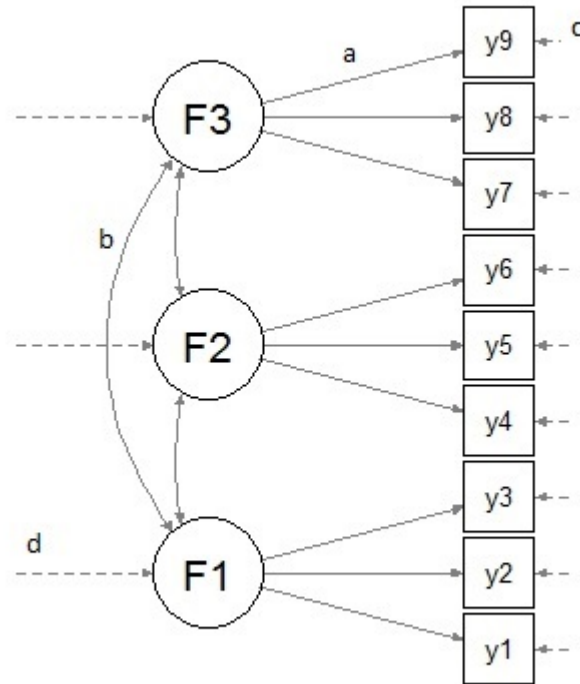


Figure 4.1: A general structure of a CFA model

This model (figure 4.1) shows that there are three *unobserved latent variables* F1, F2 and F3, and nine *observed variables* where three are considered to measure each latent variable. These nine observed variables function as indicators of their respective latent unobserved variables (Byrne 2006, chp. 1).

Associated with each observed variable is an error term (label c), and with each latent variable a residual term (label d). There is a distinction between these two error terms. An error associated with the observed variables represents measurement error which reflects on their adequacy in measuring the related underlying variables (F1, F2 and F3) (Byrne 2006, chp.1). Residual terms represent error in the prediction of dependent variables from independent variables, for example, the error in the prediction of F1, F2 and F3 (Byrne 2006, chp.1).



The one-way arrows (label a) represent structural regression coefficients where they indicate the impact of one variable to another. In figure 4.1, for example, the one-way arrow leading from F1 to each of the three observed variables (y1, y2 and y3) suggest that each of these are influenced by their respective underlying latent variables. As such, these path coefficients represent the magnitude of expected change in the observed variables for every change in the related latent variable.

The one-way arrows pointing from the nine error terms (label c) indicates the impact of measurement error on the observed variables, the residual and the impact of the error in the prediction of each latent unobserved variable (F1, F2 and F3) (Byrne 2006, chp. 1).

A path diagram, such as figure 4.1, depicting a particular CFA model is the graphical equivalent of its mathematical representation. This is where a set of equations relates dependent variables to their explanatory variables.

### 4.3.2 CFA model represented mathematically

CFA models can also be represented mathematically by a series of regression equations. This is because:

- (a) Regression equations can represent the influence of one or more variables.
- (b) Each equation summarises the impact of all relevant variables in the model (observed and unobserved) on one specific variable (observed or unobserved).

Therefore one relatively simple approach to express these equations is to note each variable that has one or more arrows (from path diagrams) pointing towards it and then record the total of all influences for each of those dependent variables.

A series of regression equations can be displayed in a matrix. So let  $y$  be a vector of order  $p \times 1$  observed variables. Let the vector  $\xi$  of order  $k \times 1$  represent the latent unobserved variables and  $\delta$  be a vector of residuals of order  $p \times 1$ . Therefore, a CFA model has the form:

$$y = \Lambda\xi + \delta \quad (4.2)$$

where  $\Lambda$  of order  $p \times k$  contains the factor loadings  $\lambda_{ij}$  (these are associated with the latent unobserved variables). To make sure the model is identified, some elements of  $\Lambda$  may be fixed at zero (Yang-Wallentin et al. 2010).

For the path diagram (shown in figure 4.1), the CFA model has matrix form as follows:

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \\ y_9 \end{bmatrix} = \begin{bmatrix} \lambda_{11} & 0 & 0 \\ \lambda_{21} & 0 & 0 \\ \lambda_{31} & 0 & 0 \\ 0 & \lambda_{42} & 0 \\ 0 & \lambda_{52} & 0 \\ 0 & \lambda_{62} & 0 \\ 0 & 0 & \lambda_{73} \\ 0 & 0 & \lambda_{83} \\ 0 & 0 & \lambda_{93} \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{bmatrix} + \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \\ \delta_6 \\ \delta_7 \\ \delta_8 \\ \delta_9 \end{bmatrix}$$

For the special case of ordinal variables, the CFA model is estimated by:

$$y^* = \Lambda\xi + \delta \quad (4.3)$$

where  $y^*$  is a vector of  $p \times 1$  of underlying variables corresponding to the  $p \times 1$  vector of the observed ordinal variable  $y$ , as defined above.

For the unknown values in the regression equations that form the CFA model to be estimated, there are several estimation methods that can be used.

### 4.3.3 Estimation methods

For continuous variables, various least square methods are available for estimating CFA models. These methods cannot be used directly with ordinal data, however they can be used in a modified form.

The hypothesis for the equation system is that the covariance matrix of the observed variables is a function of a set of parameters that if the model was correct and the parameters are known, the population covariance matrix would be exactly produced. This is represented formally as

$$\Sigma(\theta) = \Sigma \quad (4.4)$$

where  $\Sigma$  denotes the population covariance matrix of a set of observed variables and  $\Sigma(\theta)$  denotes the population covariance matrix implied by  $\theta$ , a vector of model parameters.

From equation 4.3, let  $\Psi$  and  $\Theta_\epsilon$  be the covariance matrices of  $\xi$  and  $\delta$  respectively. We assume the latent unobserved variables are uncorrelated, which means that  $\Theta_\epsilon$  is a diagonal matrix. The covariance matrix of  $y^*$  is then

$$\Sigma(\Lambda, \Psi) = \Lambda\Psi\Lambda' + \Theta_\epsilon \quad (4.5)$$

where  $\Lambda$  is a matrix of factor loadings  $\lambda_{ij}$ .

For the special case of ordinal variables, let us assume that  $\Psi$  is a correlation matrix with ones on the diagonal. As the underlying variable  $y_i^*$  have variances equal to one, it follows from equation 4.5 that

$$\Theta_\epsilon = I - \text{diag}(\Lambda\Psi\Lambda') \quad (4.6)$$

so that

$$\Sigma(\Lambda, \Psi) = \Lambda\Psi\Lambda' + I - \text{diag}(\Lambda\Psi\Lambda') \quad (4.7)$$

This is the correlation matrix implied by the model that is to be fitted to the matrix of polychoric correlations  $\mathbf{R}$  (Yang-Wallentin et al. 2010).

To estimate the model, there are several methods that can be applied. Popular methods used include maximum likelihood, unweighted least squares (ULS), diagonally weighted least squares (DWLS) and weighted least squares (WLS).

When the observed variables are continuous data, the maximum likelihood and least squares methods are used (Luo 2011). The maximum likelihood estimator does not perform well when the observed variables are ordinal, especially when the number of observed variables is small (for example five or fewer) (Flora & Curran 2004). The chi-square statistic tends to be inflated (Babakus et al. 1987, Green et al. 1997, Hutchinson & Olmos 1998, Muthen & Kaplan 1992), parameters underestimated (Babakus et al. 1987, Muthen & Kaplan 1992) and standard error estimates tend to be biased (Muthén & Kaplan 1985, Muthen & Kaplan 1992). When the observed variables are ordinal data, least squares methods can be used for estimation after being slightly adjusted (Luo 2011).

All three least squares methods are carried out in two steps. In the first step, the polychoric correlations  $\mathbf{r}$  and their asymptotic covariance matrix  $\mathbf{W}$  are estimated. Note that  $\mathbf{r} = (r_{21}, r_{31}, r_{32}, \dots, r_{p,p-1})'$  is a vector of the polychoric correlations below the diagonal of the polychoric correlation matrix  $\mathbf{R}$ . Both  $\mathbf{r}$  and  $\mathbf{W}$  are estimated using the sample data without the use of the model. The matrix  $\mathbf{W}$  contains the elements of the estimated  $N \times ACov(r_{gh}, r_{ij})$  arranged to correspond to  $\mathbf{r}$  (Yang-Wallentin et al. 2010).

In the second step, a least squares function can be used based on  $\mathbf{r}$  by minimising the fit function  $F$ . The minimisation procedure of the least squares function  $F$  depends

on the choice of the weight matrix  $\hat{V}$ . The weight matrix can be represented as

$$WLS : \hat{V} = \hat{W}^{-1} \quad (4.8)$$

$$DWLS : \hat{V} = (\text{diag}(\hat{W}))^{-1} \quad (4.9)$$

$$ULS : \hat{V} = I \quad (4.10)$$

A number of studies have addressed the performance of these estimators. It is well established that WLS performs poorly unless the sample size is large and the model size is small (Flora & Curran 2004, Muthen & Kaplan 1992). Both DWLS and ULS perform well. DWLS generally outperforms ULS in convergence rates but is outperformed by ULS in estimation accuracy (Forero et al. 2009). Even though the WLS estimator performs poorer than the other two least squares estimators, we will still include the results WLS in the output.

## 4.4 Goodness-of-fit

In this section, we describe tests that are used for measuring how well the CFA model fits the observed data.

A chi-squared goodness-of-fit test can be constructed to determine the overall model fit where the null hypothesis is the population covariance matrix  $\Sigma$  satisfies the model

$$H_0 : \text{there exists a } \theta_0 \text{ such that } \Sigma = \Sigma(\theta_0)$$

(Browne & Cudeck 1993). A large  $\chi^2$  and rejection of the null hypothesis means that the model estimates do not sufficiently produce sample covariance. On the other hand, a small  $\chi^2$  and failure to reject the null hypothesis means a good model fit (Albright & Park 2009). However, the  $\chi^2$  test is widely recognised to be problematic (Jöreskog 1967).

As there are drawbacks of the  $\chi^2$  statistic, alternative fit statistics have been developed, each with their own advantages and disadvantages. There are two groups of model fit indices: comparative and absolute. In this section, two comparative indices of fit and one absolute index of fit will be described. Comparative indices of fit that will be described includes the Comparative Fit Index (CFI) and Tucker-Lewis Fit Index (TLI). The absolute index of fit that will be described is the Root Mean Square Error of Approximation (RMSEA).

### 4.4.1 Comparative fit index and Tucker-Lewis fit index

Both the CFI and TLI measure the proportionate improvement in model fit. This is done by comparing the hypothesised model, which imposes the structure, to the baseline model. They both focus on the comparison of nested models.

#### CFI

With CFI, the inputs are standardised so the test statistic will range from zero to one. A test statistic close to one is indicative of a well fitting model. A rule-of-thumb for this value, is that values  $> 0.95$  indicates a well fitting model (Hu & Bentler 1999). CFI is calculated as follows (Byrne 2012)

$$CFI = 1 - \frac{\chi_H^2 - df_H}{\chi_B^2 - df_B}$$

where  $\chi_H^2$  and  $\chi_B^2$  are chi-square values for the hypothesised model and baseline model respectively, and  $df_H$  and  $df_B$  are the degrees of freedom for the hypothesised model and baseline model respectively.

#### TLI

TLI on the otherhand has inputs that are not standardised and so the test statistic is not restricted to being between zero and one. The interpretation of TLI is however the same as for CFI where the values close to one indicates a well fitting model. TLI is calculated as follows (Byrne 2012)

$$TLI = \left( \frac{\chi_B^2}{df_B} - \frac{\chi_H^2}{df_H} \right) - \left( \frac{\chi_B^2}{df_B} - 1 \right)$$

### 4.4.2 Root mean square error of approximation

The RMSEA depends only on how well the hypothesised model fits the sample data. The closer the index is to zero, the better the model fits. In other words, this type of absolute fit index decrease as goodness-of-fit improves.

Cutpoints for measuring goodness-of-fit include (Browne & Cudeck 1993, Steiger 1989, Browne & Mels 1990, MacCallum et al. 1996)

- $< 0.05$  indicates good model fit
- values up to 0.08 indicate reasonable errors of approximation in the population
- 0.08 – 0.10 indicates mediocre fit and
- $> 0.10$  indicates poor fit

RMSEA is calculated as follows (Byrne 2012)

$$RMSEA = \sqrt{\frac{\delta_H}{df_H}}$$

where

$$\delta_H = (\chi_H^2 - df_H)/N$$

To assess the precision of the RMSEA estimates, confidence intervals are used. A small RMSEA, but a wide confidence interval would indicate that the estimated value is quite imprecise. This disproves any possibility to determine accurately the degree of fit in the population. On the other hand, if the confidence interval was narrow, the RMSEA value would be argued as good precision in reflecting the model's fit to the population (MacCallum et al. 1996).



## 4.5 Confirmatory factor analysis application

CFA will be used to test how well the eight strategy structure, based on linguistics theory, fits. The five cluster structure suggested by the fuzzy biclustering method will also be tested using CFA. The goodness-of-fit of both structures will be compared.

The two structures are fitted using the statistical software R. Appendix C shows how to use the `cfa` function (Rosseel 2012) to find the CFA result for the two hypothesised structures.

### 4.5.1 Eight strategy structure based on linguistics theory

The first model, based on Linguistics theory, postulates a priori that English listening is an eight factor structure composed of Self-initiation, Planning, Monitoring and Evaluation, Perceptual Processing, Inferencing, Predicting, Utilisation and Socio-affective. The CFA model hypothesizes a priori that

- English listening can be explained by eight factors: Self-initiation, Planning, Monitoring and Evaluation, Perceptual Processing, Inferencing, Predicting, Utilisation and Socio-affective;
- each item-pair measure has a non-zero loading on the English listening factor that it was designed to measure and a zero loading on all other factors;
- the eight English listening factors are correlated; and
- residual errors associated with each measure are uncorrelated.

As indicated in Figure 4.2, we have a model that is hypothesised to have eight factors.<sup>1</sup> As indicated by the configuration of the thirty eight rectangles, each factor (latent variable) is measured by four to eight observed variables. All variables, latent or observed, are either dependent or independent. Observed variables, in this

---

<sup>1</sup>Please note this figure has been simplified for illustrative purposes.

case the thirty eight questions, are dependent variables and the eight strategies are independent variables.

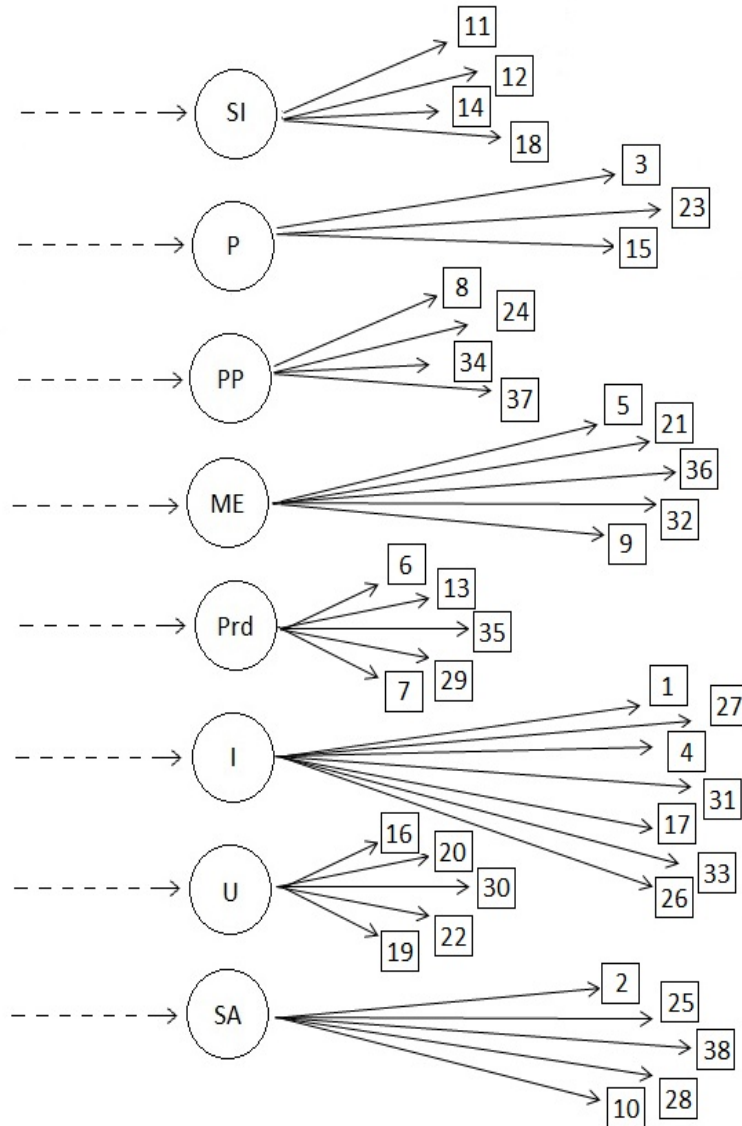


Figure 4.2: Linguistics hypothesised eight factor model

Table 4.1 and 4.2 shows the chi-square estimation of the hypothesised model and baseline model, respectively, for all three least squares methods. All three estimators show significant evidence against the null hypothesis that the population covariance matrix is equivalent to the observed sample covariance matrix.

Table 4.1: Chi-square estimation for Linguistics hypothesised model

Fit Estimate	WLS	DWLS	ULS
$\chi^2$	6228.22	2318.36	1818.79
degrees of freedom	637	637	637
P-value (chi-square)	0	0	0

Table 4.2: Chi-square estimation for Linguistics baseline model

Fit Estimate	WLS	DWLS	ULS
$\chi^2$	26533.94	56232.06	72954.87
degrees of freedom	703	703	703
P-value (chi-square)	0	0	0

Table 4.3 shows the fit indices for all three least squares methods for Linguistics hypothesised model.

Table 4.3: Fit indices for Linguistics hypothesised model

Fit Indices	WLS	DWLS	ULS
CFI	0.783	0.977	0.979
TFI	0.761	0.974	0.977
RMSEA	0.092	0.050	0.042
(90 Percent CI)	(0.090, 0.094)	(0.048, 0.053)	(0.040, 0.045)

For both DWLS and ULS, the CFI and TFI are greater than 0.95. This suggests that the model fits well. RMSEA for the same two least squares methods is approximately 0.05 or less. This again indicates the model fits well. The 90% confidence

interval for DWLS contains the 0.05 threshold for a good fit model, but only slightly. The WLS for all three of these fit indices shows a different story, where CFI is less than 0.95, TFI is further away from 1.0 and RMSEA is closer to 0.10. Using WLS indicates a poorer model fit than when using DWLS and ULS. This is expected as described earlier.

### 4.5.2 Five cluster structure based on the fuzzy biclustering method

In the listening questionnaire data, each student represented a row and each question a column. The fuzzy biclustering method (Chapter 3.5) suggested that the 1,039 students are grouped into  $R = 4$  clusters and the thirty eight questions are grouped into  $C = 5$  clusters. The five column clusters are

1. Questions 5, 18 and 19
2. Questions 21, 37 and 38
3. Questions 14, 17, 25, 26 and 32
4. Questions 2, 6, 7, 8, 13, 16, 27, 28, 29, 31 and 35
5. Questions 1, 3, 4, 9, 10, 11, 12, 15, 20, 22, 23, 24, 30, 33, 34 and 36.

This model postulates a priori that English listening is a five factor structure. This CFA model, presented in Figure 4.3, hypothesizes a priori that

- English listening can be explained by five factors;
- each item-pair measure has a non-zero loading on the English listening factor that it was designed to measure and a zero loading on all other factors;
- the five English listening factors are correlated; and
- residual errors associated with each measure are uncorrelated.

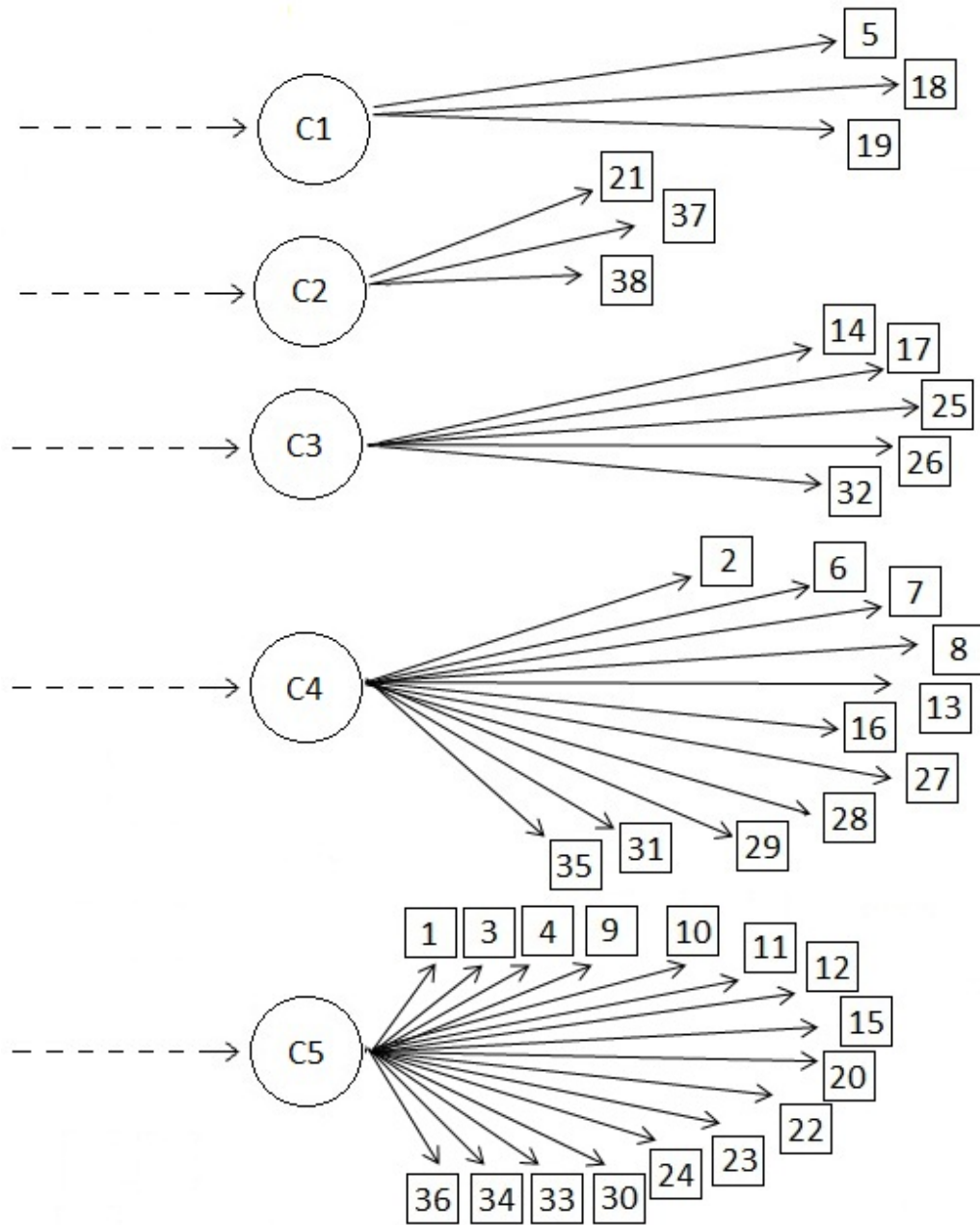


Figure 4.3: Statistical hypothesised five factor model

Table 4.4 and 4.5 shows the chi-square estimation for the hypothesised model and baseline model, respectively, for all three least squares methods. All three estimators show significant evidence against the null hypothesis that the population covariance matrix is equivalent to the observed sample covariance matrix.

Table 4.4: Chi-square estimation for five factor hypothesised model

Fit Estimate	WLS	DWLS	ULS
$\chi^2$	6854.63	3188.64	2443.05
degrees of freedom	655	655	655
P-value (chi-square)	0	0	0

Table 4.5: Chi-square estimation for five factor baseline model

Fit Estimate	WLS	DWLS	ULS
$\chi^2$	26523.94	72954.89	56232.06
degrees of freedom	703	703	703
P-value (chi-square)	0	0	0

Table 4.6 shows the fit indices for all three least squares measures. For both DWLS and ULS methods, the fit indices CFI and TFI, are greater than 0.95. This suggests the model fits well. The RMSEA for the same two least square methods are 0.061 and 0.051 respectively. These values suggest that the fit of the model is reasonable to well fit. With the ULS method, it appears to have a better model fit because RMSEA 90% confidence interval contains 0.05. In other words, the RMSEA for ULS method suggests a good model fit, where as DWLS measure suggests a reasonable model fit.

Table 4.6: Fit indices for five cluster hypothesised model

Fit Indices	WLS	DWLS	ULS
CFI	0.760	0.965	0.968
TLI	0.742	0.962	0.965
RMSEA	0.095	0.061	0.051
90 Percent CI	(0.095, 0.097)	(0.059, 0.063)	(0.049, 0.053)

Overall the fit indices, CFI and TFI suggest both the Linguistics eight cluster model and the five cluster model suggest a well fit. The CFI and TFI indices are slightly lower for the five cluster model than the ones for the Linguistics eight cluster model. This would suggest that the Linguistics model performs slightly better than the the five cluster model. However, there is a small gain by adding three additional clusters to describe the thirty eight questions. Although the overall fitting for the five cluster model might be adequate, the grouping information does not have theoretical meanings in the Linguistics area.

# Chapter 5

## Ordinal agreement model

In this chapter we will evaluate the amount of agreement between each of the questions within the listening strategies. We start with providing a description of what agreement between questions is, followed by a description of the ordinal agreement model and agreement between multiple questions. We then describe the bootstrap method and the incorporation of the ordinal agreement model to make inferences. The application of this model is then presented and described at the end of this chapter.

### 5.1 Agreement between questions

Consider two questions, A and B. Let  $\pi_{ij}$  denote the probability that question A is answered with category  $i$  and question B with category  $j$ . Their ratings of a particular question ‘agree’ if their classifications are in the same category. Therefore, in a square table, the main diagonal  $i = j$  represents question agreement. Hence,  $\sum_i \pi_{ii}$  is the total probability of agreement. Perfect agreement occurs when  $\sum_i \pi_{ii} = 1$ .

Let’s consider two questions only from the listening data. In a squared contingency table, the rows are ordinal response categories for question one and the columns



are ordinal response categories for question two. For our example, there are five response categories, and so we have a  $5 \times 5$  contingency table. We illustrate this in table 5.1 for questions 11 and 12 from the Self-initiation strategy. The categories in the rating scales are (1) never, (2) rarely, (3) sometimes, (4) often and (5) always.

Table 5.1: Question 11 versus Question 12

Question 11	Question 12					Total
	1	2	3	4	5	
1	17	22	28	10	12	89
2	24	53	61	27	15	180
3	11	69	136	86	29	331
4	8	31	99	59	54	251
5	5	21	54	47	61	188
Total	65	196	378	229	171	1,039

This square contingency table can be used to display joint ratings of the two questions. With this table, one can analyse differences in the marginal distributions and the frequency of main-diagonal occurrence within the joint distribution of the ratings (Agresti 2010, p. 248). A cell-by-cell comparison of the observed and estimated expected frequencies can help show this. Haberman (1973) suggested a standardised Pearson residual that is asymptotically standard normal. This is calculated as

$$\frac{n_{ij} - \hat{\mu}_{ij}}{[\hat{\mu}_{ij}(1 - p_{i+})(1 - p_{+j})]^{1/2}}$$

where  $n_{ij}$  is the observed frequency,  $\hat{\mu}_{ij} = \frac{n_{i+}n_{+j}}{n}$  is the expected frequency,  $p_{i+}$  is the row proportion and  $p_{+j}$  is the column proportion.

Table 5.2: Standardised Pearson Residuals

Question 11	Question 12				
	1	2	3	4	5
1	5.2	1.5	-1.0	-2.6	-0.8
2	4.3	4.0	-0.8	-2.5	-3.2
3	-2.7	1.1	2.2	2.1	-4.6
4	-2.3	-3.0	1.2	0.6	2.5
5	-2.2	-3.0	-2.4	1.1	6.5

Table 5.2 shows the standardised Pearson residuals for the same two questions in table 5.1. The main diagonal has larger positive residuals, indicating agreement for each rating is greater than expected by chance, especially for the first and last rating. The off-diagonal is primarily negative residuals indicating disagreement has occurred less than expected under independence, although evidence of this is weaker for ratings closer together. The most common disagreement are if students choose rating 3 on question 11 then they are unlikely to choose rating 5 on question 12.

## 5.2 Ordinal agreement model

The ordinal agreement model is used to describe agreement between questions which classifies a sample on a subjective ordinal scale. Many ordinal scales are quite subjective, such as the scale used in this study (never, rarely, sometimes, often and always). There rarely is perfect agreement that occurs between questions for such scales. This is partly because of differing perceptions about the meanings of the category labels, and partly because of factors such as intra-question variability (Agresti 1988).

For the Listening Strategy Questionnaire, each strategy contains several questions.

If all questions within strategies are similar, the responses from these questions should have a high level of agreement. The agreement model was originally used to describe agreement between raters. We will be using the same type of model to describe agreement between questions for each strategies.

Suppose each of the  $n$  students assigns a response rating separately for two questions, A and B. Let  $m_{ij} = n\pi_{ij}$  denote the expected frequency for the  $i^{\text{th}}$  row and  $j^{\text{th}}$  column in the contingency table. When each question has the same ordered ratings, the ordinal agreement model has the form:

$$\log m_{ij} = \lambda + \lambda_i^A + \lambda_j^B + \beta\mu_i v_j + \delta I(i = j) \quad (5.1)$$

where  $\mu_i$  is the score for the  $i$ th row and  $v_j$  is the score for the  $j$ th column. In the model,  $\delta$  is used to describe beyond-chance agreement on the main diagonal and the parameter  $\beta$  to describe association and correlation off that diagonal for the chosen scores (Agresti 2010, chp. 8).

When equation (5.1) holds, the null hypothesis of independence between the ratings  $H_0 : \beta = \delta = 0$  is true. The null hypothesis of no extra agreement beyond that due to the baseline association between ratings is  $H_0 : \delta = 0$  ( $H_1 : \delta \geq 0$ ). The null hypothesis of no extra association beyond that due to exact agreement is  $H_0 : \beta = 0$  ( $H_1 : \beta > 0$ ) (Agresti 1988).

The independence model ( $\beta = \delta = 0$ ), implies that all ratings,  $i$  and  $j$ , are indistinguishable. When  $\beta > 0$  and  $\delta \geq 0$ , the degree of distinguishability increases as the distance between the ratings increases (Agresti 1988).

The ordinal agreement model has two key components,  $\delta$  and  $\beta$ . The relative influences of these two components is the variation of the response categories (ratings). Equal interval scores, for example (1, 2, 3, 4, 5), for a five category variable, give

the simplest interpretations and are a common reasonable choice in scores. This is unless there is a more natural scoring. When the model fits well, the distances between the estimated scores can be used in describing distinguishability of ratings.

Overall, the ordinal agreement model has several positive features, as described by Agresti (1988). These include:

- It utilises the ordering of the response categories.
- Given that questions disagree, it does not assume that the ratings are independent.
- It is unsaturated on the main diagonal.
- The baseline association and extra agreement parameters are easily interpreted.

### 5.3 Agreement among multiple questions

With several questions, the description of agreement and association between two questions is conditional on the ratings by the other students. It is usually more relevant to study agreement marginally, without conditioning on the other ratings. Furthermore, simultaneously modelling the pairwise agreement structure requires modelling all the two-way marginal tables of the joining table for all the questions (Becker & Agresti 1992).

An alternative simpler approach is to take the average pairwise measure of agreement over all possible pairs of questions (Agresti 2010, chp. 8).

## 5.4 Bootstrap method with the ordinal agreement model

Bootstrapping methods are used to make inferences for comparing multiple pairs of questions in ordinal agreement models. We investigate how much agreement there is between questions within each strategy. This is so we can test whether questions within each strategy are more similar to each other than questions between strategies.

Bootstrapping is a resampling method to estimate standard errors for parameter estimators and to construct confidence intervals for parameters.

In this section we will first provide a general description of the bootstrap method and then describe the procedure taken to bootstrap the ordinal agreement model. Finally, a description on how the bootstrap method is used to summarise the agreement question combinations within strategies.

### 5.4.1 General description of the bootstrap method

Suppose there is a sample of size  $n$  and we want to estimate a parameter, determine the standard errors or a confidence interval for the parameter. To do this, we look at the sample and consider the empirical distribution. The empirical distribution is the probability distribution that has probability  $1/n$  assigned to each sample value. The bootstrap idea is to replace the unknown population distribution with the known empirical distribution.

From the bootstrap sampling, a Monte Carlo approximation of the bootstrap estimate is obtained. The procedure is as follows

1. Generate a sample with replacement from the empirical distribution

2. Compute  $\theta$  to the values of  $\hat{\theta}$  obtained by using the bootstrap sample in place of the original sample
3. Repeat steps 1 and 2  $s$  many times.

By repeating steps 1 and 2  $s$  many times, a Monte Carlo approximation to the distribution of  $\hat{\theta}$  is obtained.

### 5.4.2 Bootstrapping the ordinal agreement model

For the dataset of grade 6 students, steps are undertaken to get an estimate of the parameter  $\beta$  from the ordinal agreement model for each bootstrap sample. We then summarise the pairwise information to estimate the overall agreement parameter and its associated confidence interval within a strategy.

The following steps describe how bootstrap samples  $\hat{\beta}$  of the ordinal agreement model is undertaken:

1. Sample 1,039 students of the grade 6 dataset with replacement. A sample with replacement means that the results for a student can occur more than once. A sample of 1,039 is taken from the population of 1,039 students.
2. Build the ordinal agreement model for each pairwise combination of the questions within each strategy. For example, the Self-initiation strategy has four questions. This means that there are six pairwise combinations and so six ordinal agreement models are built.
3. From each original agreement model built, the Maximum Likelihood estimate  $\hat{\beta}$  is extracted .
4. Steps 1 to 3 are repeated  $s$  ( $= 1,000$ ) times to get 1,000  $\hat{\beta}$  for every pairwise combination within each strategy.

Table 5.3 shows the bootstrap sample structure of parameter  $\beta$  created for the Self-initiation strategy. It has six rows, each representing a pairwise combination of the questions in the strategy, and  $s = 1,000$  columns representing the 1,000 bootstrap samples taken from the original grade 6 data set.

Table 5.3: Bootstrap sample structure for Self-initiation strategy

Pairwise combinations	Sample					
	1	2	3	...	$s - 1$	$s$
Q11 vs Q12	$\beta_{11,12}^1$	$\beta_{11,12}^2$	$\beta_{11,12}^3$	...	$\beta_{11,12}^{s-1}$	$\beta_{11,12}^s$
Q11 vs Q14	$\beta_{11,14}^1$	$\beta_{11,14}^2$	$\beta_{11,14}^3$	...	$\beta_{11,14}^{s-1}$	$\beta_{11,14}^s$
Q11 vs Q18	$\beta_{11,18}^1$	$\beta_{11,18}^2$	$\beta_{11,18}^3$	...	$\beta_{11,18}^{s-1}$	$\beta_{11,18}^s$
Q12 vs Q14	$\beta_{12,14}^1$	$\beta_{12,14}^2$	$\beta_{12,14}^3$	...	$\beta_{12,14}^{s-1}$	$\beta_{12,14}^s$
Q12 vs Q18	$\beta_{12,18}^1$	$\beta_{12,18}^2$	$\beta_{12,18}^3$	...	$\beta_{12,18}^{s-1}$	$\beta_{12,18}^s$
Q14 vs Q18	$\beta_{14,18}^1$	$\beta_{14,18}^2$	$\beta_{14,18}^3$	...	$\beta_{14,18}^{s-1}$	$\beta_{14,18}^s$

The R-code used to implement the bootstrap method for the ordinal agreement model can be found in the Appendix D.

### 5.4.3 Summary of the overall agreement $\hat{\beta}$

After the bootstrap samples of the grade 6 data are completed, all pairwise combinations of questions and their parameter estimates can be found for each strategy.

Several steps are undertaken to calculate the overall agreement  $\hat{\beta}$  across different pairwise combinations of questions within strategies. The steps are illustrated using the Self-initiation strategy questions where there are six pairwise combinations. The steps are as follows:

1. Given the estimates for each pairwise combinations, take the mean across

pairwise combinations:

$$\hat{\beta}^i = \frac{\hat{\beta}_{11,12}^i + \hat{\beta}_{11,14}^i + \hat{\beta}_{11,18}^i + \hat{\beta}_{12,14}^i + \hat{\beta}_{12,18}^i + \hat{\beta}_{14,18}^i}{q}$$

where  $i = 1, 2, \dots, s$  and  $q$  is the number of pairwise combinations. Figure 5.1 shows the distribution of  $\{\hat{\beta}^i, i = 1, \dots, s\}$ . The distribution appears normally distributed.

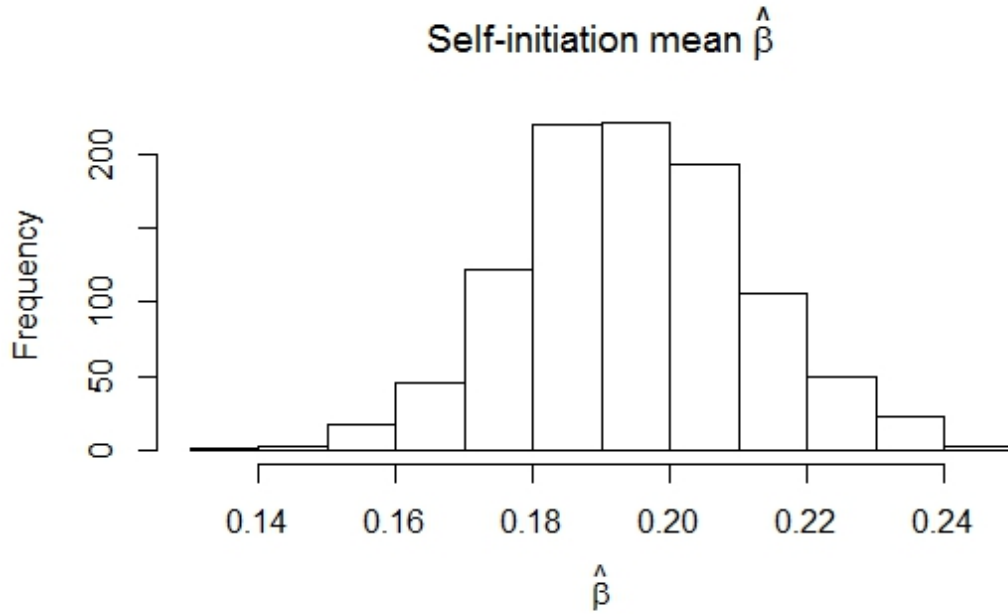


Figure 5.1: Distribution of  $\{\hat{\beta}^i, i = 1, \dots, s\}$  for Self-initiation strategy

2. The next step is to calculate the mean of  $\{\hat{\beta}^i, i = 1, \dots, s\}$ . This gives us the summary of agreement parameter estimate for the strategy Self-initiation.

$$\hat{\beta} = \frac{\sum_{i=1}^s \hat{\beta}^i}{s}$$

3. The standard error of  $\hat{\beta}$  of the strategy is then calculated using the formula:

$$SE = \sqrt{\frac{\sum_{i=1}^s (\hat{\beta}^i - \hat{\beta})^2}{s - 1}}$$



4. With the standard error, the 95% confidence interval can be calculated using the formula:  $\hat{\beta} \pm 1.96 \times SE$ .
5. Repeat steps 1 to 4 for the other seven strategies.

The bootstrap inference can also be applied to check whether the agreement parameters are the same for two strategies. For example, we can conduct a 95% bootstrap confidence interval of  $\beta^{Prediction} - \beta^{Utilisation}$ . If the interval contains 0, it implies that the agreement level for the strategy Prediction is not significantly different from that for the strategy Utilisation.

#### 5.4.4 Estimation of pairwise agreement

This section describes the bootstrap inference for the pairwise agreement between two questions. Self-imitation strategy is again used as an example.

The Self-imitation strategy has four questions and so six pairwise combinations. Figure 5.2 shows the distribution of  $\hat{\beta}_{j,j'}$  over 1,000 ( $= s$ ) bootstrap samples for the six pairwise  $(j, j')$  combinations for the Self-imitation strategy, where  $j < j' = 11, 12, 14, 18$ .

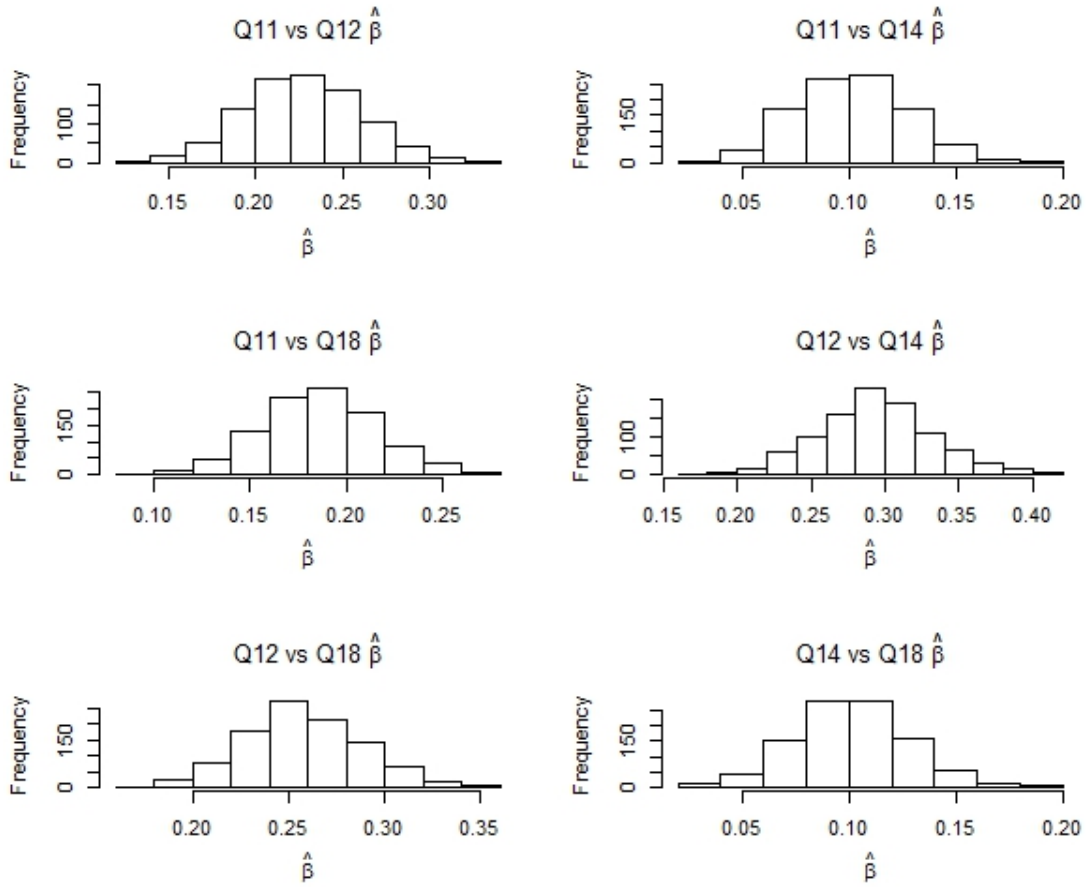


Figure 5.2: Distribution of  $\hat{\beta}_{j,j'}$  for all pairwise combinations for Self-initiation strategy

The steps of making bootstrap inferences for pairwise agreement parameters  $\{\beta_{j,j'}, j < j' = 11, 12, 14, 18\}$  are as follows:

1. The  $\hat{\beta}_{j,j'}$  is calculated by taking the mean of each pairwise combination as follows:

$$\hat{\beta}_{j,j'} = \frac{\sum_{i=1}^s \hat{\beta}_{j,j'}^i}{s} \quad j < j' = 11, 12, 14, 18,$$

where  $j$  and  $j'$  are the questions and  $s$  is the number of bootstrap samples (1,000).

2. The standard error for  $\hat{\beta}_{j,j'}$  (six in the case of the Self-initiation strategy) is

#### 5.4. BOOTSTRAP METHOD WITH THE ORDINAL AGREEMENT MODEL97

then calculated as follows:

$$SE_{j,j'} = \sqrt{\frac{\sum_{i=1}^s (\hat{\beta}_{j,j'}^i - \hat{\beta}_{j,j'})^2}{s-1}} \quad j < j' = 11, 12, 14, 18.$$

3. The 95% confidence interval using the mean  $\hat{\beta}_{j,j'}$  and standard error can now be calculated for every pairwise combination.

## 5.5 Ordinal agreement model application

This section describes the coding of the data and the results of overall agreement within strategies and pairwise agreement based on ordinal agreement models using the bootstrap method. This model is fitted using the the statistical software R. Appendix D shows how to use the `glm` function (R Core Team (2013)) to fit the agreement model.

### 5.5.1 Data coding

This subsection describes how the data were coded to allow us to apply the ordinal agreement model.

The dataset has 1,039 rows (as described earlier in Chapter 2) where each row is a student and their answers to the thirty eight questions. There are three ordinal agreement models applied to each pairwise combination. The difference between each model is the 5-point scale that is used to represent the ordered responses of the questions. These are:

1. (1, 2, 3, 4, 5)
2. (1, 2, 4, 5, 6)
3. (1, 3, 4, 5, 7)

These three 5-point scales are used as each one has different spacing between each point in the scale. The first scale has equal spacing between all points. The second scale has two spacings between “rarely” and “sometimes” and finally the third scale is where the extreme points in the scale are further away from the centre points.

### 5.5.2 Results

The results for the three scales (1, 2, 3, 4, 5), (1, 2, 4, 5, 6), and (1, 3, 4, 5, 7) are shown in tables 5.4, 5.5 and 5.6 respectively.

Table 5.4: Scale (1, 2, 3, 4, 5)

Strategy	$\hat{\beta}$	Confidence Interval of $\hat{\beta}$
Self-initiation	0.19	(0.16, 0.23)
Planning	0.19	(0.15, 0.23)
Monitoring and Evaluation	0.18	(0.15, 0.21)
Perceptual Processing	0.15	(0.12, 0.18)
Inferencing	0.22	(0.19, 0.25)
Prediction	0.31	(0.27, 0.35)
Utilisation	0.24	(0.21, 0.27)
Socio-affective	0.12	(0.10, 0.14)

For scale (1, 2, 3, 4, 5), the Prediction strategy has the largest amount of agreement between questions. This strategy's level of agreement appears to be significantly different to the other seven strategies. The Inferencing and Utilisation strategies have the next highest level of agreement. The strategy with the least amount of agreement between questions is the Socio-affective strategy and Perceptual Processing strategy.

Table 5.5: Scale (1, 2, 4, 5, 6)

Strategy	$\hat{\beta}$	Confidence Interval of $\hat{\beta}$
Self-initiation	0.1096	(0.0912, 0.1281)
Planning	0.1027	(0.0820, 0.1235)
Monitoring and Evaluation	0.0991	(0.0824, 0.1235)
Perceptual Processing	0.0991	(0.0824, 0.1235)
Inferencing	0.1151	(0.0994, 0.1308)
Prediction	0.1635	(0.1445, 0.1825)
Utilisation	0.1255	(0.1085, 0.1426)
Socio-affective	0.0609	(0.0478, 0.0739)

The same pattern can be seen in table 5.5 as in table 5.4. The Prediction strategy has the highest amount of agreement of all the strategies and the Socio-affective strategy has the least amount of agreement of all the strategies.

Table 5.6: Scale (1, 3, 4, 5, 7)

Strategy	$\hat{\beta}$	Confidence Interval of $\hat{\beta}$
Self-initiation	0.0974	(0.0796, 0.1151)
Planning	0.0900	(0.0669, 0.1130)
Monitoring and Evaluation	0.0871	(0.0704, 0.1038)
Perceptual Processing	0.0746	(0.5773, 0.0915)
Inferencing	0.1108	(0.0934, 0.1282)
Prediction	0.1567	(0.1345, 0.1790)
Utilisation	0.1223	(0.1036, 0.1411)
Socio-affective	0.0625	(0.0490, 0.0759)

As in the previous two tables, the Prediction strategy has the highest level of agreement and the Socio-affective strategy has the lowest level of agreement. Tables 5.7 and 5.9 shows the amount of agreement between questions within the Prediction

strategy and Socio-affective strategy along with the corresponding 95% confidence intervals, respectively, for scale (1, 2, 3, 4, 5). Because the three scales have similar results, we only report the result from the scale (1, 2, 3, 4, 5).

Table 5.7: Agreement between questions for Prediction strategy

	6	7	13	29	35
6	-	0.31 (0.24, 0.38)	0.28 (0.21, 0.35)	0.31 (0.23, 0.39)	0.32 (0.25, 0.40)
7	-	-	0.26 (0.19, 0.33)	0.25 (0.18, 0.32)	0.29 (0.22, 0.36)
13	-	-	-	0.35 (0.26, 0.44)	0.34 (0.26, 0.41)
29	-	-	-	-	0.41 (0.31, 0.51)
35	-	-	-	-	-

Amongst the questions within the Prediction strategy, the level of agreement between questions 29 and 35 is the highest at an average of 0.41. However, this is not significantly different from the rest of the pairwise level of agreement within the Prediction strategy. Questions 29 and 35 were:

- Q29: When I listen, I use my knowledge about the topic to think about what I will hear next.
- Q35: When I listen, I use my general knowledge to think about what I will hear next.

Both these questions are very similar in what they are asking the student. It appears sensible that the answers to these two questions by the students would highly agree with each other.

If we now look at the standardised Pearson residuals (described in section 5.1) between question 29 and 35 using the results of the questionnaire (no bootstrapping), we can see that on the main diagonal there are large positive residuals, with the end points being the largest of 16.2 and 10.4 (shown in table 5.8). This indicates for each of these categories, agreement on the diagonal is greater than expected. The off-diagonal is primarily negative residuals, where the most negative is -5.6. This indicates that disagreement occurred less than expected under independence, although evidence of this is weaker for categories close to ratings 4 and 5. With disagreement occurring much less than expected, this would generally give us a higher  $\beta$  and so a higher level of agreement.

Table 5.8: Standardised Pearson residuals for Prediction strategy

Question 29	Question 35				
	1	2	3	4	5
1	16.2	-1.1	-3.6	-4.0	-3.6
2	-1.0	10.7	-2.5	-5.5	-2.8
3	-5.6	-0.7	8.0	-1.4	-4.3
4	-3.7	-5.0	-2.5	8.5	3.5
5	-1.5	-5.2	-2.9	2.2	10.4



Now looking at the Socio-affective strategy, the highest pairwise level of agreement is between questions 2 and 28 of 0.20. This is significantly lower than the highest pairwise agreement in the Prediction strategy.

Table 5.9: Agreement between questions for Socio-affective strategy

	2	10	25	28	38
2	-	0.12 (0.07, 0.17)	0.13 (0.08, 0.18)	0.20 (0.14, 0.25)	0.10 (0.05, 0.14)
10	-	-	0.05 (-0.001, 0.09)	0.18 (0.12, 0.23)	0.02 (-0.02, 0.07)
25	-	-	-	0.10 (0.05, 0.15)	0.18 (0.12, 0.23)
28	-	-	-	-	0.12 (0.07, 0.17)
38	-	-	-	-	-

Within the Socio-affective strategy, there are two pairwise questions where their 95% confidence interval includes zero. This suggests that there is little evidence against the null hypothesis of there being no agreement between the questions. The pairwise questions are questions 10 and 25, and questions 10 and 38. These questions in the questionnaire were:

- Q10: When I have a problem in listening, I ask my friends for help.
- Q25: I tell myself to enjoy listening to English.
- Q38: I tell myself not to worry when I listen in English.

What these questions are asking is not very similar. Question 10 asks about how students would ask for help; question 25 is about a students enjoyment of listening to English and; question 38 is about how students cope when listening to English. The topics of these questions appear different.

We will now look at the standardised Pearson residuals for the Socio-affective strategy using the results of the questionnaire. Table 5.10 shows the standardised Pearson residuals for question 10 versus question 25, and table 5.11 question 10 versus question 38. These two pairwise questions had the lowest agreement between questions.

Table 5.10: Standardised Pearson residuals for Socio-affective strategy - Question 10 versus Question 25

Question 10	Question 25				
	1	2	3	4	5
1	4.3	-0.4	-1.0	-1.4	0.2
2	0.2	1.5	-0.7	0.7	-1.3
3	-1.2	0.3	2.1	-0.2	-1.6
4	-1.5	-1.6	1.3	0.7	0.2
5	-0.6	0.2	-2.5	0.0	3.0

In table 5.10, the residuals on the main diagonal are all positive, where the highest is on the far ends of the diagonal. These residuals are much smaller than those shown in table 5.8 for one pairwise question of the Prediction strategy. The off diagonal is primarily negative residuals, however 8 of the possible 20 off diagonal residuals are positive. Disagreement occurred less than expected for majority of the ratings, but disagreement on 8 residuals occurred more than expected. This would give a lower  $\beta$  in the modelling.

Table 5.11: Standardised Pearson residuals for Socio-affective strategy - Question 10 versus Question 38

Question 10	Question 38				
	1	2	3	4	5
1	3.1	0.9	-1.4	-1.5	0.3
2	0.5	-0.6	0.0	1.0	-0.8
3	-2.1	1.9	1.7	-0.8	-0.9
4	-1.0	-0.9	0.3	1.7	-0.7
5	0.6	-1.5	-1.2	-0.8	2.5

In table 5.11, the residuals on the main diagonal are mostly positive. Where both ratings is 2, we have a residual of -0.6. This suggests that agreement for this category is less than expected. The highest main diagonal residual is at the far ends again with residuals of 3.1 and 2.5 (much smaller residuals than the same ratings in the Prediction strategy). The off diagonal is primarily negative residuals, however again 8 of the possible 20 off diagonal residuals are positive. Disagreement occurred less than expected for majority of the ratings, but disagreement on 8 residuals occurred more than expected. This would give a lower  $\beta$  in the modelling.

# Chapter 6

## Discussion

In this thesis, several methods have been applied to analyse ordered categorical data from the listening questionnaire. A range of categorical ordinal methods are used to demonstrate how well the eight Linguistic clusters perform. This analysis has established a picture of which clusters contain questions similar to each other as well as which pairwise questions within a cluster are similar. It has also shown how well the overall eight cluster model performs. The work itself is useful as the focus is on treating the categorical ordinal variables as ordinal rather than continuous.

Ordered categorical data are often treated as continuous data because of the simplicity of the analysis, but there are many drawbacks using this naive approach mentioned in Chapter 1. By treating ordered responses as ranking data, correlation coefficients can show the associations between questions. These preliminary results are consistent with the findings from proportional odds models and ordinal agreement models. For example, the Prediction strategy had the highest range of correlations between questions, ranging between 0.36 and 0.54. The Socio-affective strategy had the lowest range of correlations between questions, ranging between 0.07 and 0.34. This suggests the questions within the Prediction strategy have a strong relationship and the questions within the Socio-affective strategy have a weak relationship. However, it is difficult to make proper inferences for the pairwise

correlation coefficients by taking all possible variables into account.

The different approaches to the proportional odds model showed that the Planning, Inferencing and Prediction strategies had no difference between questions. This suggests that student responses for these strategies are similar. The Self-initiation, Utilisation, Monitoring and Evaluation, Preceptual Processing and Socio-affective strategies showed a difference between questions. That the student responses to the questions within each strategy are not similar to each other. There are however pairwise questions within these strategies that are similar to each other.

The Self-initiation strategy had only one pairwise questions that showed no difference between them, questions 11 and 12. The Utilisation strategy had six pairwise questions that showed no difference. The remaining four pairwise questions showed very strong significance of a difference, where the common question in each pair was question 7. For the Monitoring and Evaluation strategy, there were significant ethnicity and question interactions. For the Chinese students, there was no difference between questions 32 and 36, and questions 5 and 21. For the other ethnicity students, the results were different. The subject-specific approach showed that there was a difference between all pairwise questions, while the GEE approach showed no difference between questions 9 and 21, and questions 9 and 36. The Perceptual Processing strategy showed no difference between questions 24 and 34 only. Finally, in the Socio-affective strategy, there were no pairwise questions that were similar to each other.

In the Ordinal Agreement model, the Prediction strategy had the highest level of agreement followed by the Utilisation strategy. On the other hand, the Socio-affective strategy had the lowest level of agreement between questions in the cluster followed by Perceptual Processing strategy. In the Socio-affective strategy, the lowest pairwise questions were questions 10 and 25, and questions 10 and 38. This

is different to the proportional odds model, where the pairwise questions with the largest difference were questions 25 and 28, questions 2 and 38, and questions 25 and 38.

The Confirmatory Factor analysis method showed how well the eight linguistic clusters fit the ordinal data. This method showed that the eight clusters fit the data well, where all the fit indices that were measured were better than the threshold of a good fit. The fuzzy biclustering method grouped the thirty eight questions into five clusters that had a higher likelihood than the eight linguistic clusters. The CFA showed a reasonable model fit for the five clusters where all fit indices were slightly lower than that of the eight linguistic clusters. The main disadvantage of the five cluster structure is that the meaning of each clusters is not clear in Linguistic field.

There are similarities between the three models described in this thesis. This includes:

- The proportional odds model and CFA both have an underlying latent variable of an unobserved continuous variable.
- All methods take the ordering nature into account.
- Both proportional odds and ordinal agreement models allow to include many explanatory variables to control possible confounding variables.

There are differences between the three models described as well, which include:

- CFA evaluates the model structure considering all eight listening strategies together, while the other two can find the differences between questions and strategies.
- For the proportional odds model, the marginal approach focuses on the marginal distribution of the responses for the various observations. The effect  $\beta$  is population-averaged. The subject-specific approach allows probabilities to

vary by subject. The effect  $\beta$  is defined conditional on the subject. The magnitude of  $\beta$  might be different between these two approaches, but often the significance of an effect is similar (Agresti 2012, p. 497).

- The proportional odds model and the ordinal agreement model have different focuses. The proportional odds model focuses on the underlying response distribution of each questions. The agreement parameter in an ordinal agreement model describes whether the response on Question A goes up as the response on Question B goes up. For example, suppose the response on Question B is always one level higher than the response on Question A for all students. It would result in a large value for the agreement parameter in an ordinal agreement model. Therefore, the two questions are very similar. However, the underlying response distribution for Question A would be different from that for Question B. Thus, the proportional odds model might suggest that there is a difference between these two questions.

Future work based on this research could involve an improvement of the methods used for the full consideration of the missing data. A small proportion of students were removed for the GEE approach issue where the data is missing completely at random. Future research will incorporate the GEE method with missing data (Twisk & de Vente 2002) for proportional odds models.

# Bibliography

- Agresti, A. (1988), ‘A model for agreement between ratings on an ordinal scale’, *Biometrics* **44**, 539–548.
- Agresti, A. (2007), *An introduction to categorical data analysis*, 2nd edn, John Wiley & Sons, Inc.
- Agresti, A. (2010), *Analysis of Ordinal Categorical Data*, 2nd edn, John Wiley & Sons, Inc.
- Agresti, A. (2012), *Categorical Data Analysis*, 3rd edn, John Wiley & Sons, Inc.
- Albright, J. J. & Park, H. M. (2009), ‘Confirmatory factor analysis using amos, lisrel, mplus, and sas/stat calis’, *The Trustees of Indiana University* **1**, 1–85.
- Anderson, J. & Philips, P. (1981), ‘Regression, discrimination, and measurement models for ordered categorical variables’, *Applied Statistics* **30**, 22–31.
- Babakus, E., Ferguson Jr, C. E. & Jöreskog, K. G. (1987), ‘The sensitivity of confirmatory maximum likelihood factor analysis to violations of measurement scale and distributional assumptions’, *Journal of Marketing Research* pp. 222–228.
- Becker, M. & Agresti, A. (1992), ‘Log-linear modelling of pairwise interobserver agreement on a categorical scale’, *Statistics in Medicine* **11**, 101–114.
- Bentler, P. M. (1988), Causal modeling via structural equation systems, *in* ‘Handbook of multivariate experimental psychology’, Springer, pp. 317–335.



- Browne, M. & Cudeck, R. (1993), *Alternative ways of assessing model fit*. In K.A. Bollen & J.S. Long (Eds.), Newbury Park, CA: Sage.
- Browne, M. & Mels, G. (1990), 'Ramona user's guide, columbus: Department of psychology', *Ohio State University* .
- Byrne, B. (2006), *Structural Equation Modeling with EQS: Basic Concepts, Applications, and Programming*, Multivariate applications book series, Lawrence Erlbaum Associates.  
**URL:** <http://books.google.co.nz/books?id=0U9DkYfu-xcC>
- Byrne, B. M. (2012), *Structural equation modeling with Mplus: Basic concepts, applications, and programming*, Routledge.
- Christensen, R. H. B. (2013), 'ordinal—regression models for ordinal data'. R package version 2013.9-30 <http://www.cran.r-project.org/package=ordinal/>.
- Flora, D. B. & Curran, P. J. (2004), 'An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data.', *Psychological methods* **9**(4), 466.
- Forero, C. G., Maydeu-Olivares, A. & Gallardo-Pujol, D. (2009), 'Factor analysis with ordinal indicators: A monte carlo study comparing dwls and uls estimation', *Structural Equation Modeling* **16**(4), 625–641.
- Ghosh, M. (1995), 'Inconsistent maximum likelihood estimators for the rasch model', *Statistics & Probability Letters* **23**(2), 165–170.
- Green, S. B., Akey, T. M., Fleming, K. K., Hershberger, S. L. & Marquis, J. G. (1997), 'Effect of the number of scale points on chi-square fit indices in confirmatory factor analysis', *Structural Equation Modeling: A Multidisciplinary Journal* **4**(2), 108–120.
- Gu, P., Hu, G. & Zhang, L. (2005), 'Investigating language learner strategies among lower primary school pupils in singapore', *Language and Education* .

- Haberman, S. J. (1973), 'The analysis of residuals in cross-classification tables', *Biometrics* **25**, 205–220.
- Hastie, T., Botha, J. & Schnitzler, C. (1989), 'Regression with an ordered categorical response', *Statistics in Medicine* **8**, 785–794.
- Hu, L.-t. & 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.
- Hutchinson, S. R. & Olmos, A. (1998), 'Behavior of descriptive fit indexes in confirmatory factor analysis using ordered categorical data', *Structural Equation Modeling: A Multidisciplinary Journal* **5**(4), 344–364.
- Jöreskog, K. G. (1967), 'A general approach to confirmatory maximum likelihood factor analysis', *ETS Research Bulletin Series* **1967**(2), 183–202.
- Jöreskog, K. G. (1990), 'New developments in lisrel: Analysis of ordinal variables using polychoric correlations and weighted least squares', *Quality & Quantity* **24**(4), 387–404.
- Kaufman, L. & Rousseeuw, P. (1990), *Finding groups in data: an introduction to cluster analysis*, John Wiley & Sons, Inc.
- Language Learner Strategy Research Team (2009), 'English language learning in singapore schools'.  
**URL:** <http://www.nie.edu.sg/research-projects/english-language-learning-strategies>
- Lee, S.-Y., Poon, W.-Y. & Bentler, P. (1990), 'Full maximum likelihood analysis of structural equation models with polytomous variables', *Statistics & probability letters* **9**(1), 91–97.

- Lewis, S., Foltynie, T., Blackwell, A., Robbins, T., Owen, A. & Barker, R. (2003), 'Heterogeneity of parkinson's disease in the early clinical stages using a data driven approach', *Journal Neurology, Neurosurgery and Psychiatric* **76**, 343–348.
- Liang, K. Y. & Zeger, S. L. (1986), 'Longitudinal data-analysis using generalized linear-models', *Biometrika* **73**(1), 13–22.
- Lipsitz, S. R., Kim, K. & Zhao, L. (1994), 'Analysis of repeated categorical data using generalized estimating equations', *Statistics in medicine* **13**(11), 1149–1163.
- Little, R. (1988), 'Missing data adjustments in large surveys', *Journal of Business & Economic Statistics* **6**, 287–296.
- Little, R. & Rubin, D. (2002), *Statistical analysis with missing data*, 2nd edn, John Wiley & Sons, Inc.
- Liu, I.-M. & Agresti, A. (1996), 'Mantel-haenszel-type inference for cumulative odds ratios with a stratified ordinal response', *Biometrics* pp. 1223–1234.
- Luo, H. (2011), 'Some aspects on confirmatory factor analysis of ordinal variables and generating non-normal data'.
- MacCallum, R. C., Browne, M. W. & Sugawara, H. M. (1996), 'Power analysis and determination of sample size for covariance structure modeling.', *Psychological methods* **1**(2), 130.
- Matechou, E., Liu, I., Pledger, S. & Arnold, R. (2011), 'Biclustering models for ordinal data'. Presentation at the NZ Statistical Association Annual Conference, University of Auckland, 28-31 August 2011.
- McCullagh, P. (1980), 'Regression models for ordinal data (with discussion)', *Journal of the Royal Statistical Society* **42**, 109–142.
- Melnykov, V. (2003), 'Challenges in model-based clustering', *WIRE Computational Statistics* **5**, 135–148.

- Muthén, B. (1984), ‘A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators’, *Psychometrika* **49**(1), 115–132.
- Muthén, B. & Kaplan, D. (1985), ‘A comparison of some methodologies for the factor analysis of non-normal likert variables’, *British Journal of Mathematical and Statistical Psychology* **38**(2), 171–189.
- Muthen, B. & Kaplan, D. (1992), ‘A comparison of some methodologies for the factor analysis of non-normal likert variables: A note on the size of the model’, *British Journal of Mathematical and Statistical Psychology* **45**(1), 19–30.
- Neyman, J. & Scott, E. (1948), ‘Consistent estimates based on partially consistent observations’, *Econometrika* **16**(1), 1–22.
- Olsson, U. (1979), ‘Maximum likelihood estimation of the polychoric correlation coefficient’, *Psychometrika* **44**(4), 443–460.
- O’Malley, J. M. & Chamot, A. U. (1990), *Learning strategies in second language acquisition*, Cambridge University Press.
- Pledger, S. & Arnold, R. (2013), ‘Multivariate methods using mixtures: correspondence analysis, scaling and pattern detection’, *Computational Statistics and Data Analysis* .
- R Core Team (2013), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.  
**URL:** <http://www.R-project.org/>
- Rosseel, Y. (2012), ‘lavaan: An R package for structural equation modeling’, *Journal of Statistical Software* **48**(2), 1–36.  
**URL:** <http://www.jstatsoft.org/v48/i02/>
- Rubin, D. (1976), ‘Inference and missing data’, *Biometrika* **63**, 581–592.

- Scott, J. (2000), 'The challenge for quantitative methods', *Research with children: Perspectives and practices* pp. 98–119.
- Souriyavongsa, T., Abidin, M., Sam, R., Mei, L. & Aloysius, I. (2013), 'Investigating learning english strategies and english needs of undergraduate students at the national university of laos', *English Language Teaching* **6**, 57–71.
- Steiger, J. H. (1989), *EzPATH: A supplementary Module for SYSTAT and SYGRAPH*, Evanston, IL: Systat.
- Touloumis, A. (2013), 'multgee: a GEE solver for correlated nominal and ordinal multinomial responses'.  
**URL:** <http://CRAN.R-project.org/package=multgee/>
- Touloumis, A., Agresti, A. & Kateri, M. (2013), 'GEE for multinomial responses using a local odds ratios parametrization', *Biometrics* **69**, 633–640.
- Twisk, J. & de Vente, W. (2002), 'Attrition in longitudinal studies: how to deal with missing data', *Journal of clinical epidemiology* **55**(4), 329–337.
- Wolfe, R. & Firth, D. (2002), 'Modelling subjective use of an ordinal response scale in a many period cross over experiment', *Journal of the Royal Statistical Society* **51**, 245–255.
- Wu, J. (2014), 'Characteristics of chinese primary school students' efl learning strategies', *Theory and Practices in Language* **4**, 1266–1272.
- Yang-Wallentin, F., Jöreskog, K. G. & Luo, H. (2010), 'Confirmatory factor analysis of ordinal variables with misspecified models', *Structural Equation Modeling* **17**(3), 392–423.
- Zheng, J. (2013), 'A survey of tuition-free english major students' use of language learner studies', *International Education Studies* **6**, 9–17.

# Appendix A

## Listening questionnaire

In this part, all thirty eight questions used in the analyses in this thesis will be listed out. The answers for all questions has the scale (never, rarely, sometimes, often, always).

1. When I don't understand something, I use my knowledge about the English language to guess.
2. When I have a problem in listening, I ask my family members for help.
3. Before I listen to something important, I ask myself what I already know about the topic.
4. When I don't understand something, I use what I have already heard to guess it.
5. I compare what I am hearing with what I have already heard to make sure I understand correctly.
6. When I listen, I use my knowledge about text type (for example, story, report, etc) to think about what I will hear next.
7. Before I listen to something, I think about the main idea of what I am going to hear.

8. When I listen, I repeat the words or phrases I can understand.
9. During or after listening, I ask myself whether the information is the same as what I already know.
10. When I have a problem in listening, I ask my friends for help.
11. To improve my listening in English, I watch English TV programmes.
12. I look for opportunities to listen in English.
13. When I listen, I use what I have already heard to think about what I will hear next.
14. I try to find out how to improve my listening in English.
15. Before I listen to something, I ask myself whether it is important to me.
16. I try to find problems with how the speaker presented an idea so that I can avoid the same problems.
17. When I don't understand something, I make several guesses.
18. When I'm free, I find interesting things to listen to in English (for example, TV, radio, etc).
19. When I listen, I try to remember useful words and phrases so that I can use them.
20. I try to connect what I heard with my own experiences.
21. When I have a problem in listening, I decide whether I should pay more attention to it.
22. After I finish listening, I make a summary in my mind about what I heard.
23. Before I start listening, I decide if I need to pay attention to details or to the main idea.

24. If I can't understand a word or phrase, I repeat it to myself.
25. I tell myself to enjoy listening to English.
26. When I don't understand something, I use my general knowledge to make a guess.
27. When I listen, I use my knowledge about the English language to think about what I will hear next.
28. When I have a problem in listening, I ask my teachers for help.
29. When I listen, I use my knowledge about the topic to think about what I will next.
30. After I finish listening, I use my own words to retell what I heard in my mind.
31. When I don't understand something, I use my knowledge about text type (for example, story, report, etc.) to help me understand.
32. If I have a problem in understanding, I quickly decide whether I should continue or listen again.
33. When I don't understand something, I use my knowledge about the topic to guess.
34. When I listen, I repeat the pronunciation of the words I have heard.
35. When I listen, I use my general knowledge to think about what I will hear next.
36. During or after listening, I check how much I have understood.
37. When I listen, I pay attention to every word that is said.
38. I tell myself not to worry when I listen in English.



# Appendix B

## Proportional odds models

The R-code below goes through how the data were coded so that the proportional odds model with the GEE method could be applied. Comments about the R-code starts with `##`.

```
library(MASS)
library(car)
library(cluster)
library(AER)

LinguisticsCluster = factor(rep(c(rep(1,4),rep(2,3),rep(3,5),rep(4,4),rep(5,7),
    rep(6,5),rep(7,5),rep(8,5)),dim(Year6)[1]))

## LinguisticsCluster - A column of eight strategies where each strategy is repeated
by the number of questions in the strategy. For example, Self-initiation
is strategy 1. It has four questions in the strategy and so "1" is repeated four
times. The same goes with all other strategies.

Student = rep(1:dim(Year6)[1],each=38)

## Student - a column of students each repeated thirty eight times.
```

```

S1 = rep(c(rep(1:4,each=1),rep(0,38-4)),dim(Year6)[1])
S2 = rep(c(rep(0,4),rep(1:3,each=1),rep(0,38-7)),dim(Year6)[1])
S3 = rep(c(rep(0,7),rep(1:5,each=1),rep(0,38-12)),dim(Year6)[1])
S4 = rep(c(rep(0,12),rep(1:4,each=1),rep(0,38-16)),dim(Year6)[1])
S5 = rep(c(rep(0,16),rep(1:7,each=1),rep(0,38-23)),dim(Year6)[1])
S6 = rep(c(rep(0,23),rep(1:5,each=1),rep(0,38-28)),dim(Year6)[1])
S7 = rep(c(rep(0,28),rep(1:5,each=1),rep(0,38-33)),dim(Year6)[1])
S8 = rep(c(rep(0,33),rep(1:5,each=1)),dim(Year6)[1])

## S1-S8 : a column of 1's and 0's where 1 indicates the row/question is part of
the strategy and 0 indicates the row/question is not part of the strategy.

Questions = Year6[,7:44]
QuestionsNewOrder = cbind(Questions[,11],Questions[,12],Questions[,14],
  Questions[,18],Questions[,3],Questions[,15],Questions[,23],
  Questions[,21],Questions[,32],Questions[,5],Questions[,9],Questions[,36],
  Questions[,8],Questions[,24],Questions[,34],Questions[,37],Questions[,1],
  Questions[,4],Questions[,17],Questions[,26],Questions[,31],Questions[,27],
  Questions[,33],Questions[,6],Questions[,7],Questions[,13],Questions[,29],
  Questions[,35],Questions[,16],Questions[,19],Questions[,20],Questions[,22],
  Questions[,30],Questions[,2],Questions[,10],Questions[,28],Questions[,25],
  Questions[,38])

## Questions - the columns of questions in the original dataset is re-ordered into
strategies. Self-initiation questions are first followed by Planning questions etc.

Response = as.vector(append(t(QuestionsNewOrder[1,]),t(QuestionsNewOrder[2,])))
for(i in 3:dim(QuestionsNewOrder)[1])
{
  Response = append(Response,t(QuestionsNewOrder[i,]))
}

```

```
}

```

```
## Transposes the first two questions in the Questions matrix and appends them
together. Each column is appended to the previous column until all columns are
combined into one column.
```

```
Response.fac = factor(Response)
Questions.fac = factor(Question)
```

```
##Response.fac and Questions.fac - changes the numeric column "Response" and
"Questions" into factors.
```

```
Gender = rep(Year6$V4,each=38)
Chinese = rep(Year6$V5,each=38)
School = rep(Year6$V46,each=38)
```

```
## Gender/Chinese/School - repeat the Gender/Ethnicity/School column from
the dataset thirty eight times.
```

```
levels(Chinese) = list("Chinese"=c("Chinese","Eurasian"),
  "Other"=c("Malay","Others","Indian"))
```

```
## This changes the different levels of the Ethnicity into two groups, Chinese
and Other.
```

```
newdata = as.data.frame(cbind(Response,Response.fac,Student,Gender,Chinese,
  School,RowCluster,Questions.fac,LinguisticsCluster,
  S1,S2,S3,S4,S5,S6,S7,S8))
```

```
## newdata - combine each column created above into one dataset.
```

```
newdata$$S1.fac = factor(newdata$$S1)
newdata$$S2.fac = factor(newdata$$S2)
newdata$$S3.fac = factor(newdata$$S3)
newdata$$S4.fac = factor(newdata$$S4)
newdata$$S5.fac = factor(newdata$$S5)
newdata$$S6.fac = factor(newdata$$S6)
newdata$$S7.fac = factor(newdata$$S7)
newdata$$S8.fac = factor(newdata$$S8)
newdata$$Student.fac = factor(Student)
newdata$$Gender.fac = factor(Gender)
newdata$$Chinese.fac = factor(Chinese)
newdata$$School.fac = factor(School)

## The "factor" function used above changes all the above columns from
numeric to factors.
```

This dataset created can now be used for the different approaches of the proportional odds model.



```

Wald.S1.gee = t(L%%S1.gee$coef)%%solve(L%%S1.gee$robust.variance%%t(L))
              %% (L%%S1.gee$coef)
Wald.S1.gee
1 - pchisq(Wald.S1.gee,3)

S1.gee.1 = ordLORgee(ordered(Response.fac)~ factor(S1) + Gender.fac
                   + Chinese.fac + School.fac + factor(S1)*Gender.fac
                   + factor(S1)*Chinese.fac, id=Student.fac, repeated=S1,
                   data=subset(gee_table,LinguisticsCluster==1))
summary(S1.gee.1)$coefficients
L=matrix(c(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,0,0,0),nrow=1)
Wald.S1.gee = t(L%%S1.gee.1$coef)%%solve(L%%S1.gee.1$robust.variance%%t(L))
              %% (L%%S1.gee.1$coef)
Wald.S1.gee
1 - pchisq(Wald.S1.gee,3)

S1.gee.2 = ordLORgee(ordered(Response.fac)~ factor(S1) + Gender.fac
                   + Chinese.fac + School.fac + factor(S1)*Gender.fac, id=Student.fac,
                   repeated=S1, data=subset(gee_table,LinguisticsCluster==1))
summary(S1.gee.2)$coefficients
L=matrix(c(0,0,0,0,0,0,0,0,0,0,0,1,1,1,1,1,0,0,0),nrow=1)
Wald.S1.gee = t(L%%S1.gee.2$coef)%%solve(L%%S1.gee.2$robust.variance%%t(L))
              %% (L%%S1.gee.2$coef)
Wald.S1.gee
1 - pchisq(Wald.S1.gee,5)

S1.gee.3 = ordLORgee(ordered(Response.fac)~ factor(S1) + Gender.fac
                   + Chinese.fac + factor(S1)*Gender.fac, id=Student.fac,
                   repeated=S1, data=subset(gee_table,LinguisticsCluster==1))
summary(S1.gee.3)$coefficients
L=matrix(c(0,0,0,0,0,0,0,0,0,0,0,1,1,1),nrow=1)

```

```

Wald.S1.gee = t(L%%S1.gee.3$coef)%%solve(L%%S1.gee.3$robust.variance%%t(L))
              %% (L%%S1.gee.3$coef)
Wald.S1.gee
1 - pchisq(Wald.S1.gee,3)

S1.gee.4 = ordLORgee(ordered(Response.fac)~ factor(S1) + Gender.fac
                   + Chinese.fac, id=Student.fac, repeated=S1,
                   data=subset(gee_table,LinguisticsCluster==1))
summary(S1.gee.4)$coefficients
L=matrix(c(0,0,0,0,0,0,0,1,0),nrow=1)
Wald.S1.gee = t(L%%S1.gee.4$coef)%%solve(L%%S1.gee.4$robust.variance%%t(L))
              %% (L%%S1.gee.4$coef)
Wald.S1.gee
1 - pchisq(Wald.S1.gee,1)

S1.gee.5 = ordLORgee(ordered(Response.fac)~ factor(S1) + Chinese.fac,
                   id=Student.fac, repeated=S1, data=subset(gee_table,LinguisticsCluster==1))
summary(S1.gee.5)$coefficients
L=matrix(c(0,0,0,0,0,0,0,1),nrow=1)
Wald.S1.gee = t(L%%S1.gee.5$coef)%%solve(L%%S1.gee.5$robust.variance%%t(L))
              %% (L%%S1.gee.5$coef)
Wald.S1.gee
1 - pchisq(Wald.S1.gee,1)

#Model inference for Self-initiation strategy
## Final model where only the Chinese variable is significant.
## Inference of the strategy variables is calculated and the results of the
Wald test is displayed.

S1.gee.5 = ordLORgee(ordered(Response.fac)~ factor(S1) + Chinese.fac,
                   id=Student.fac, repeated=S1, data=subset(gee_table,LinguisticsCluster==1))

```

```

summary(S1.gee.5)$coefficients
L=matrix(c(0,0,0,0,1,1,1,0),nrow=1)
Wald.S1.gee = t(L%%S1.gee.5$coef)%%solve(L%%S1.gee.5$robust.variance)%%t(L)
          %%(L%%S1.gee.5$coef)
Wald.S1.gee
1 - pchisq(Wald.S1.gee,3)

#Model inference output from R
> summary(S1.gee.5)$coefficients
              Estimate  san.se    san.z Pr(>|san.z|)
beta01          -2.35038 0.09213 -25.51238   0.00000
beta02          -0.99220 0.06865 -14.45248   0.00000
beta03           0.46133 0.06333   7.28460   0.00000
beta04           1.57719 0.07040  22.40196   0.00000
factor(S1)2      0.05267 0.06461   0.81529   0.41491
factor(S1)3     -0.23558 0.07082  -3.32640   0.00088
factor(S1)4     -0.44089 0.06466  -6.81910   0.00000
Chinese.facOther -0.33209 0.07940  -4.18276   0.00003
> L=matrix(c(0,0,0,0,1,1,1,0),nrow=1)
> Wald.S1.gee = t(L%%S1.gee.5$coef)%%solve(L%%S1.gee.5$robust.variance
          %%%t(L))%%(L%%S1.gee.5$coef)
> Wald.S1.gee
          [,1]
[1,] 14.31322
> 1 - pchisq(Wald.S1.gee,3)
          [,1]
[1,] 0.002508379

##This value at the end is the p-value for the Self-initiation strategy variable.
It suggests there is very strong evidence against the null hypothesis of no
difference.

```





```

L1=matrix(c(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,0,0,0,0),nrow=1)
Wald.S1.SS = t(L%%as.vector(summary(S1.SS.1)$coefficients[,1]))
            %%solve(L1%%vcov(S1.SS.1)%%t(L1))%%
            (L%%as.vector(summary(S1.SS.1)$coefficients[,1]))
Wald.S1.SS
1 - pchisq(Wald.S1.SS,3)

S1.SS.2=clmm(as.factor(Response.fac)~ S1.fac + (1|Student.fac) + Gender.fac
            + Chinese.fac + S1.fac*Gender.fac + S1.fac*Chinese.fac,
            data=subset(standard.newdata,LinguisticsCluster==1),link="logit",Hess=TRUE)
summary(S1.SS.2)
L=matrix(c(0,0,0,0,0,0,0,0,0,0,1,1,1,0,0,0),nrow=1)
L1=matrix(c(0,0,0,0,0,0,0,0,0,0,1,1,1,0,0,0,0),nrow=1)
Wald.S1.SS = t(L%%as.vector(summary(S1.SS.2)$coefficients[,1]))
            %%solve(L1%%vcov(S1.SS.2)%%t(L1))%%
            (L%%as.vector(summary(S1.SS.2)$coefficients[,1]))
Wald.S1.SS
1 - pchisq(Wald.S1.SS,3)

S1.SS.3=clmm(as.factor(Response.fac)~ S1.fac + (1|Student.fac) + Gender.fac
            + Chinese.fac + S1.fac*Gender.fac,
            data=subset(standard.newdata,LinguisticsCluster==1),link="logit",Hess=TRUE)
summary(S1.SS.3)
L=matrix(c(0,0,0,0,0,0,0,0,0,0,1,1,1),nrow=1)
L1=matrix(c(0,0,0,0,0,0,0,0,0,0,1,1,1,0),nrow=1)
Wald.S1.SS = t(L%%as.vector(summary(S1.SS.3)$coefficients[,1]))
            %%solve(L1%%vcov(S1.SS.3)%%t(L1))%%
            (L%%as.vector(summary(S1.SS.3)$coefficients[,1]))
Wald.S1.SS
1 - pchisq(Wald.S1.SS,3)

```

```
S1.SS.4=clmm(as.factor(Response.fac)~ S1.fac + (1|Student.fac) + Gender.fac
            + Chinese.fac, data=subset(standard.newdata,LinguisticsCluster==1),
            link="logit",Hess=TRUE)
```

```
summary(S1.SS.4)
```

```
L=matrix(c(0,0,0,0,0,0,0,0,1),nrow=1)
```

```
L1=matrix(c(0,0,0,0,0,0,0,0,1,0),nrow=1)
```

```
Wald.S1.SS = t(L%%as.vector(summary(S1.SS.4)$coefficients[,1]))
            %%solve(L1%%vcov(S1.SS.4)%%t(L1))%%
            (L%%as.vector(summary(S1.SS.4)$coefficients[,1]))
```

```
Wald.S1.SS
```

```
1 - pchisq(Wald.S1.SS,3)
```

```
S1.SS.5=clmm(as.factor(Response.fac)~ S1.fac + (1|Student.fac) + Gender.fac
            + Chinese.fac, data=subset(standard.newdata,LinguisticsCluster==1),
            link="logit",Hess=TRUE)
```

```
summary(S1.SS.5)
```

```
L=matrix(c(0,0,0,0,0,0,0,0,1),nrow=1)
```

```
L1=matrix(c(0,0,0,0,0,0,0,0,1,0),nrow=1)
```

```
Wald.S1.SS = t(L%%as.vector(summary(S1.SS.5)$coefficients[,1]))
            %%solve(L1%%vcov(S1.SS.5)%%t(L1))%%
            (L%%as.vector(summary(S1.SS.5)$coefficients[,1]))
```

```
Wald.S1.SS
```

```
1 - pchisq(Wald.S1.SS,1)
```

```
S1.SS.6=clmm(as.factor(Response.fac)~ S1.fac + (1|Student.fac) + Chinese.fac,
            data=subset(standard.newdata,LinguisticsCluster==1),link="logit",Hess=TRUE)
```

```
summary(S1.SS.6)
```

```
L=matrix(c(0,0,0,0,0,0,0,1),nrow=1)
```

```
L1=matrix(c(0,0,0,0,0,0,0,1,0),nrow=1)
```

```
Wald.S1.SS = t(L%%as.vector(summary(S1.SS.6)$coefficients[,1]))
            %%solve(L1%%vcov(S1.SS.6)%%t(L1))%%
```

```

(L%%as.vector(summary(S1.SS.6)$coefficients[,1]))
Wald.S1.SS
1 - pchisq(Wald.S1.SS,1)

##Model inference of Self-initiation strategy
## Final model where only the Chinese variable is significant.
### Inference of the strategy variable is calculated and results of the Wald
test is displayed.

S1.SS.final=clmm(as.factor(Response.fac)~ S1.fac + (1|Student.fac) + Chinese.fac,
  data=subset(standard.newdata,LinguisticsCluster==1),
  link="logit",Hess=TRUE)
summary(S1.SS.final)
L=matrix(c(0,0,0,0,1,1,1,0),nrow=1)
L1=matrix(c(0,0,0,0,1,1,1,0,0),nrow=1)
Wald.S1.SS = t(L%%as.vector(summary(S1.SS.final)$coefficients[,1]))
  %%solve(L1%%vcov(S1.SS.final)%%t(L1))%%
  (L%%as.vector(summary(S1.SS.final)$coefficients[,1]))
Wald.S1.SS
1 - pchisq(Wald.S1.SS,3)

#Model inference output
> summary(S1.SS.final)
Cumulative Link Mixed Model fitted with the Laplace approximation

formula: as.factor(Response.fac) ~ S1.fac + (1 | Student.fac) + Chinese.fac
data:    subset(standard.newdata, LinguisticsCluster == 1)

link threshold nobs logLik AIC niter max.grad cond.H
logit flexible 4156 -6025.25 12068.50 771(3067) 4.83e-03 4.8e+01

```

Random effects:

Groups	Name	Variance	Std.Dev.
	Student.fac (Intercept)	1.505	1.227

Number of groups: Student.fac 1039

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
S1.fac2	-0.07177	0.08057	-0.891	0.373024
S1.fac3	0.29540	0.08175	3.613	0.000302 ***
S1.fac4	0.51449	0.08288	6.208	5.37e-10 ***
Chinese.facOther	0.42638	0.10399	4.100	4.13e-05 ***

---

Signif. codes: 0 \*\*\* 0.001 \*\* 0.01 \* 0.05 . 0.1 1

Threshold coefficients:

	Estimate	Std. Error	z value
1 2	-2.90268	0.09990	-29.055
2 3	-1.26993	0.08233	-15.425
3 4	0.59441	0.07947	7.479
4 5	2.00660	0.08605	23.319

> Wald.S1.SS

[,1]

[1,] 13.62346

> 1 - pchisq(Wald.S1.SS,3)

[,1]

[1,] 0.003465181

##This value at the end is the p-value for the Self-initiation strategy variable.  
It suggests there is very strong evidence against the null hypothesis of no difference.

# Appendix C

## Confirmatory factor analysis

### C.1 Linguistic strategies

The R-code in this appendix goes through how CFA is run. Comments about the R-code starts with `##`.

```
library(sem)
library(semPlot)
library(lavaan)
library(tikzDevice)

## Structure of the eight Linguistics strategies (hypothesised model structure) to be
used in through the CFA function.

v.model = ' selfinitiation =~ Q11 + Q12 + Q14 + Q18
planning =~ Q3 + Q15 + Q23
monitoring =~ Q21 + Q32 + Q5 + Q9 + Q36
perceptualrproc =~ Q8 + Q24 + Q34 + Q37
inferencing =~ Q1 + Q4 + Q17 + Q26 + Q31 + Q27 + Q33
prediction =~ Q6 + Q7 + Q13 + Q29 + Q35
utilisation =~ Q16 + Q19 + Q20 + Q22 + Q30
```

```
socioaffective =~ Q2 + Q10 + Q28 + Q25 + Q38 '
```

```
## The next three cfa functions fits the hypothesised model structure using the  
three least squares methods - weighted least squares, diagonal weighted  
least squares and uniform least squares.
```

```
## Important note for this function is to make sure the ratings for each  
question is ordered. This is by including the 'ordered' statement in the function.
```

```
fit.WLS = cfa(v.model,data = data_Year6,std.lv=TRUE,estimator="WLS",  
             mimic="Mplus",ordered=c("Q1","Q2","Q3","Q4","Q5","Q6","Q7","Q8","Q9",  
             "Q10","Q11","Q12","Q13","Q14","Q15","Q16","Q17","Q18","Q19","Q20",  
             "Q21","Q22","Q23","Q24","Q25","Q26","Q27","Q28","Q29","Q30","Q31",  
             "Q32","Q33","Q34","Q35","Q36","Q37","Q38"))
```

```
fit.ULS = cfa(v.model,data = data_Year6,std.lv=TRUE,estimator="ULS",  
             mimic="Mplus",ordered=c("Q1","Q2","Q3","Q4","Q5","Q6","Q7","Q8","Q9",  
             "Q10","Q11","Q12","Q13","Q14","Q15","Q16","Q17","Q18","Q19","Q20",  
             "Q21","Q22","Q23","Q24","Q25","Q26","Q27","Q28","Q29","Q30","Q31",  
             "Q32","Q33","Q34","Q35","Q36","Q37","Q38"))
```

```
fit.DWLS = cfa(v.model,data = data_Year6,std.lv=TRUE,estimator="DWLS",  
              mimic="Mplus",ordered=c("Q1","Q2","Q3","Q4","Q5","Q6","Q7","Q8","Q9",  
              "Q10","Q11","Q12","Q13","Q14","Q15","Q16","Q17","Q18","Q19","Q20",  
              "Q21","Q22","Q23","Q24","Q25","Q26","Q27","Q28","Q29","Q30","Q31",  
              "Q32","Q33","Q34","Q35","Q36","Q37","Q38"))
```

```
## Summary output for DWLS estimator.
```

```
> summary(fit.DWLS, fit.measures = T)
```

```
lavaan (0.5-16) converged normally after 46 iterations
```

Number of observations	1039
Estimator	DWLS
Minimum Function Test Statistic	2318.352
Degrees of freedom	637
P-value (Chi-square)	0.000

```
Model test baseline model:
```

Minimum Function Test Statistic	72954.886
Degrees of freedom	703
P-value	0.000

```
User model versus baseline model:
```

Comparative Fit Index (CFI)	0.977
Tucker-Lewis Index (TLI)	0.974

```
Root Mean Square Error of Approximation:
```

RMSEA	0.050
90 Percent Confidence Interval	0.048 0.053
P-value RMSEA <= 0.05	0.371



## C.2 Fuzzy biclustering clusters

## Structure of the five fuzzy biclustering clusters (hypothesised model structure) to be put through the CFA function.

```
ideal.model = ' F1 =~ Q5 + Q18 + Q19
F2 =~ Q21 + Q37 + Q38
F3 =~ Q14 + Q17 + Q25 + Q26 + Q32
F4 =~ Q2 + Q6 + Q7 + Q8 + Q13 + Q16 + Q27 + Q28 + Q29 + Q31 + Q35
F5 =~ Q1 + Q3 + Q4 + Q9 + Q10 + Q11 + Q12 + Q15 + Q20 + Q22 + Q23 + Q24
      + Q30 + Q33 + Q34 + Q36 '
```

## The next three cfa functions fits the hypothesised model using the three least squares methods - weighted least squares, diagonal weighted least squares and uniform least squares.

## Important note for this function is to make sure the ratings for each question is ordered. This is by including the 'ordered' statement in the function.

```
ideal.WLS = cfa(ideal.model,data = data_Year6,std.lv=TRUE,estimator="WLS",
  mimic="Mplus",ordered=c("Q1","Q2","Q3","Q4","Q5","Q6","Q7","Q8","Q9",
  "Q10","Q11","Q12","Q13","Q14","Q15","Q16","Q17","Q18","Q19","Q20",
  "Q21","Q22","Q23","Q24","Q25","Q26","Q27","Q28","Q29","Q30","Q31",
  "Q32","Q33","Q34","Q35","Q36","Q37","Q38"))
```

```
ideal.ULS = cfa(ideal.model,data = data_Year6,std.lv=TRUE,estimator="ULS",
  mimic="Mplus",ordered=c("Q1","Q2","Q3","Q4","Q5","Q6","Q7","Q8","Q9",
  "Q10","Q11","Q12","Q13","Q14","Q15","Q16","Q17","Q18","Q19","Q20",
  "Q21","Q22","Q23","Q24","Q25","Q26","Q27","Q28","Q29","Q30","Q31",
  "Q32","Q33","Q34","Q35","Q36","Q37","Q38"))
```

```
ideal.DWLS = lavaan(ideal.model,data = data_Year6,std.lv=TRUE,estimator="DWLS",
  mimic="Mplus",ordered=c("Q1","Q2","Q3","Q4","Q5","Q6","Q7","Q8","Q9",
  "Q10","Q11","Q12","Q13","Q14","Q15","Q16","Q17","Q18","Q19","Q20",
  "Q21","Q22","Q23","Q24","Q25","Q26","Q27","Q28","Q29","Q30","Q31",
  "Q32","Q33","Q34","Q35","Q36","Q37","Q38"))
```

```
## Summary output for DWLS estimators.
```

```
> summary(ideal.DWLS, fit.measures = T)
```

```
lavaan (0.5-16) converged normally after 48 iterations
```

```
Number of observations                1039
```

```
Estimator                            DWLS
```

```
Minimum Function Test Statistic      3189.063
```

```
Degrees of freedom                   655
```

```
P-value (Chi-square)                 0.000
```

```
Model test baseline model:
```

```
Minimum Function Test Statistic      72954.886
```

```
Degrees of freedom                   703
```

```
P-value                               0.000
```

```
User model versus baseline model:
```

```
Comparative Fit Index (CFI)         0.965
```

```
Tucker-Lewis Index (TLI)            0.962
```

Root Mean Square Error of Approximation:

RMSEA		0.061
90 Percent Confidence Interval	0.059	0.063
P-value RMSEA $\leq$ 0.05		0.000

# Appendix D

## Ordinal agreement model

### D.1 Ordinal agreement model function

The following is R-code to build an ordinal agreement model. The pieces of code with `#` in front, is used when the scale of the ratings is changed. Comments about the R-code starts with `##`.

`##` A function called 'ordinal' is created here. This is used to create an ordinal agreement model between two questions. The output of this function is the term  $\beta$ .

`##` Columns  $u$  and  $v$  (as described in Chapter 5, section 5.2) are created from the rating values (1-5).

`##` Column  $uv$  is also created by multiplying the  $u$  and  $v$  columns together.

`##` The column 'delta' is a column of 0's and 1's, where 1 is where  $u = v$  and 0's else where.

```
ordinal = function(Q1,Q2)
{
  table.survey=data.frame(table(Q1,Q2))
  colnames(table.survey)=c("Var1","Var2","count")
  table.survey$Var1.fac=factor(table.survey$Var1)
```

```

    table.survey$Var2.fac=factor(table.survey$Var2)
# table.survey$Var1.u = ifelse(table.survey$Var1==1,1,
    ifelse(table.survey$Var1==2,3,ifelse(table.survey$Var1==3,4,
    ifelse(table.survey$Var1==4,5,7))))
# table.survey$Var2.v = ifelse(table.survey$Var2==1,1,
    ifelse(table.survey$Var2==2,3,ifelse(table.survey$Var2==3,4,
    ifelse(table.survey$Var2==4,5,7))))
# table.survey$Var1.u.fac = factor(table.survey$Var1.u)
# table.survey$Var2.v.fac = factor(table.survey$Var2.v)
    table.survey$uv = as.numeric(table.survey$Var1)*as.numeric(table.survey$Var2)
# table.survey$uv.2 = as.numeric(table.survey$Var1.u)*as.numeric(table.survey$Var2.v)
    table.survey$delta = c(1,0,0,0,0,0,1,0,0,0,0,0,1,0,0,0,0,0,1,0,0,0,0,0,1)
    linear.log = glm(count~Var1.fac+Var2.fac+uv+delta,family=poisson,data=table.survey)
# linear.log = glm(count~Var1.u.fac+Var2.v.fac+uv.2+delta,family=poisson,
    data=table.survey)
    beta=summary(linear.log)$coef[10,1]
    output = data.frame(beta)
    output
}

```

## D.2 Bootstrapping of ordinal agreement model

```
## Number of repetitions to be used for bootstrap method.
```

```
n = 1000
```

```
## Creates a matrix for each strategy where it has q columns and n rows,
where q is the number of pairwise questions and n is the number of repetitions.
```

```
self_initiation = matrix(NA,nrow=n,ncol=6)
```

```
planning = matrix(NA,nrow=n,ncol=3)
```

```
MandE = matrix(NA,nrow=n,ncol=10)
```

```

pp = matrix(NA,nrow=n,ncol=6)
inferencing = matrix(NA,nrow=n,ncol=21)
prediction = matrix(NA,nrow=n,ncol=10)
utilisation = matrix(NA,nrow=n,ncol=10)
socioaffective = matrix(NA,nrow=n,ncol=10)

## A loop that runs n times. Creates a sample of the 1,039 students, with replacement.
Runs each pairwise question through the 'ordinal' function created previously and
adding each beta output into the appropriate strategy matrix created above.
At the end of the loop, there would be, for example, 6 columns and 1,000 rows
of beta values for the Self-initiation strategy.

for (i in 1:n)
{
  sample_year = data_Year6[sample(nrow(data_Year6),dim(data_Year6)[1],replace=TRUE),]

  #Self initiation
  self_initiation[i,1] = ordinal(sample_year$V17,sample_year$V18)[,1]
  self_initiation[i,2] = ordinal(sample_year$V17,sample_year$V20)[,1]
  self_initiation[i,3] = ordinal(sample_year$V17,sample_year$V24)[,1]
  self_initiation[i,4] = ordinal(sample_year$V18,sample_year$V20)[,1]
  self_initiation[i,5] = ordinal(sample_year$V18,sample_year$V24)[,1]
  self_initiation[i,6] = ordinal(sample_year$V20,sample_year$V24)[,1]

  #Planning
  planning[i,1] = ordinal(sample_year$V9,sample_year$V21)[,1]
  planning[i,2] = ordinal(sample_year$V9,sample_year$V29)[,1]
  planning[i,3] = ordinal(sample_year$V21,sample_year$V29)[,1]

  #Monitoring and Evaluating
  MandE[i,1] = ordinal(sample_year$V11,sample_year$V15)[,1]

```

```
MandE[i,2] = ordinal(sample_year$V11,sample_year$V27) [,1]
MandE[i,3] = ordinal(sample_year$V11,sample_year$V38) [,1]
MandE[i,4] = ordinal(sample_year$V11,sample_year$V42) [,1]
MandE[i,5] = ordinal(sample_year$V15,sample_year$V27) [,1]
MandE[i,6] = ordinal(sample_year$V15,sample_year$V38) [,1]
MandE[i,7] = ordinal(sample_year$V15,sample_year$V42) [,1]
MandE[i,8] = ordinal(sample_year$V27,sample_year$V38) [,1]
MandE[i,9] = ordinal(sample_year$V27,sample_year$V42) [,1]
MandE[i,10] = ordinal(sample_year$V38,sample_year$V42) [,1]

#Perceptual Processing
pp[i,1] = ordinal(sample_year$V14,sample_year$V28) [,1]
pp[i,2] = ordinal(sample_year$V14,sample_year$V38) [,1]
pp[i,3] = ordinal(sample_year$V14,sample_year$V43) [,1]
pp[i,4] = ordinal(sample_year$V28,sample_year$V38) [,1]
pp[i,5] = ordinal(sample_year$V28,sample_year$V43) [,1]
pp[i,6] = ordinal(sample_year$V38,sample_year$V43) [,1]

#Inferencing
inferencing[i,1] = ordinal(sample_year$V7,sample_year$V10) [,1]
inferencing[i,2] = ordinal(sample_year$V7,sample_year$V23) [,1]
inferencing[i,3] = ordinal(sample_year$V7,sample_year$V32) [,1]
inferencing[i,4] = ordinal(sample_year$V7,sample_year$V33) [,1]
inferencing[i,5] = ordinal(sample_year$V7,sample_year$V37) [,1]
inferencing[i,6] = ordinal(sample_year$V7,sample_year$V39) [,1]
inferencing[i,7] = ordinal(sample_year$V10,sample_year$V23) [,1]
inferencing[i,8] = ordinal(sample_year$V10,sample_year$V32) [,1]
inferencing[i,9] = ordinal(sample_year$V10,sample_year$V33) [,1]
inferencing[i,10] = ordinal(sample_year$V10,sample_year$V37) [,1]
inferencing[i,11] = ordinal(sample_year$V10,sample_year$V39) [,1]
inferencing[i,12] = ordinal(sample_year$V23,sample_year$V32) [,1]
```

```
inferencing[i,13] = ordinal(sample_year$V23,sample_year$V33)[,1]
inferencing[i,14] = ordinal(sample_year$V23,sample_year$V37)[,1]
inferencing[i,15] = ordinal(sample_year$V23,sample_year$V39)[,1]
inferencing[i,16] = ordinal(sample_year$V32,sample_year$V33)[,1]
inferencing[i,17] = ordinal(sample_year$V32,sample_year$V37)[,1]
inferencing[i,18] = ordinal(sample_year$V32,sample_year$V39)[,1]
inferencing[i,19] = ordinal(sample_year$V33,sample_year$V37)[,1]
inferencing[i,20] = ordinal(sample_year$V33,sample_year$V39)[,1]
inferencing[i,21] = ordinal(sample_year$V37,sample_year$V39)[,1]
```

#### #Prediction

```
prediction[i,1] = ordinal(sample_year$V12,sample_year$V13)[,1]
prediction[i,2] = ordinal(sample_year$V12,sample_year$V19)[,1]
prediction[i,3] = ordinal(sample_year$V12,sample_year$V35)[,1]
prediction[i,4] = ordinal(sample_year$V12,sample_year$V41)[,1]
prediction[i,5] = ordinal(sample_year$V13,sample_year$V19)[,1]
prediction[i,6] = ordinal(sample_year$V13,sample_year$V35)[,1]
prediction[i,7] = ordinal(sample_year$V13,sample_year$V41)[,1]
prediction[i,8] = ordinal(sample_year$V19,sample_year$V35)[,1]
prediction[i,9] = ordinal(sample_year$V19,sample_year$V41)[,1]
prediction[i,10] = ordinal(sample_year$V35,sample_year$V41)[,1]
```

#### #Utilisation

```
utilisation[i,1] = ordinal(sample_year$V22,sample_year$V25)[,1]
utilisation[i,2] = ordinal(sample_year$V22,sample_year$V26)[,1]
utilisation[i,3] = ordinal(sample_year$V22,sample_year$V28)[,1]
utilisation[i,4] = ordinal(sample_year$V22,sample_year$V36)[,1]
utilisation[i,5] = ordinal(sample_year$V25,sample_year$V26)[,1]
utilisation[i,6] = ordinal(sample_year$V25,sample_year$V28)[,1]
utilisation[i,7] = ordinal(sample_year$V25,sample_year$V36)[,1]
utilisation[i,8] = ordinal(sample_year$V26,sample_year$V28)[,1]
```



```
utilisation[i,9] = ordinal(sample_year$V26,sample_year$V36)[,1]
utilisation[i,10] = ordinal(sample_year$V28,sample_year$V36)[,1]

#Socio-affective
socioaffective[i,1] = ordinal(sample_year$V8,sample_year$V16)[,1]
socioaffective[i,2] = ordinal(sample_year$V8,sample_year$V31)[,1]
socioaffective[i,3] = ordinal(sample_year$V8,sample_year$V34)[,1]
socioaffective[i,4] = ordinal(sample_year$V8,sample_year$V44)[,1]
socioaffective[i,5] = ordinal(sample_year$V16,sample_year$V31)[,1]
socioaffective[i,6] = ordinal(sample_year$V16,sample_year$V34)[,1]
socioaffective[i,7] = ordinal(sample_year$V16,sample_year$V44)[,1]
socioaffective[i,8] = ordinal(sample_year$V31,sample_year$V34)[,1]
socioaffective[i,9] = ordinal(sample_year$V31,sample_year$V44)[,1]
socioaffective[i,10] = ordinal(sample_year$V34,sample_year$V44)[,1]
}
```

### D.3 Summary of overall agreement $\hat{\beta}$

This section shows how to calculate a vector of  $\hat{\beta}$  used in the estimation of pairwise agreement.

```
## Creates eight variables that has the value 0.
```

```
mean_self_initiation = numeric(0)
mean_planning = numeric(0)
mean_MandE = numeric(0)
mean_pp = numeric(0)
mean_inferencing = numeric(0)
mean_prediction = numeric(0)
mean_utilisation = numeric(0)
mean_socioaffective = numeric(0)
```

```
## A loop that runs through each row of the eight Linguistic strategy dataset
of beta's. For each strategy, the mean of each row is taken. So, for example,
the Self-initiation strategy has a matrix of 6 columns and 1,000 rows of beta
values. The mean of each row is taken, so that the new vector has 1 column
and 1,000 rows.
```

```
for (j in 1:n)
{ mean_self_initiation[j] = mean(self_initiation[j,])
  mean_planning[j] = mean(planning[j,])
  mean_MandE[j] = mean(MandE[j,])
  mean_pp[j] = mean(pp[j,])
  mean_inferencing[j] = mean(inferencing[j,])
  mean_prediction[j] = mean(prediction[j,])
  mean_utilisation[j] = mean(utilisation[j,])
  mean_socioaffective[j] = mean(socioaffective[j,]) }
```

## D.4 Estimation of pairwise agreement

This section shows how the  $\hat{\beta}$  and  $\hat{\beta}_{j,j'}$  of each pairwise combination for each strategy is calculated.

```
## A function called Conf.Int is created here. This function calculates several values including the mean value of a column, lower and upper 95% confidence interval for the mean value.
```

```
Conf.Int = function(data)
{
  x = numeric(0)
  for (i in 1:length(data))
  {
    x[i]= (data[i]-mean(data))^2
  }

  se = sqrt(sum(x)/(length(x)-1))

  ci=numeric(0)
  ci[1]=mean(data)
  ci[2]=mean(data)-qnorm(0.975)*se
  ci[3]=mean(data)+qnorm(0.975)*se
  ci
}
```

```
## Calls the Conf.Int function for each of the strategy vectors created previously.  
Here the strategy mean beta is calculated.
```

```
Conf.Int(mean_self_initiation)  
Conf.Int(mean_planning)  
Conf.Int(mean_MandE)  
Conf.Int(mean_pp)  
Conf.Int(mean_inferencing)  
Conf.Int(mean_prediction)  
Conf.Int(mean_utilisation)  
Conf.Int(mean_socioaffective)
```

```
## Calls the Conf.Int function for the Self-initiation strategy for each column  
in the Self-initiation matrix that was created first, to give the pairwise mean  
beta's. This function is also called for all the other strategies and the  
columns in those matrices.
```

```
Conf.Int(self_initiation[,1])  
Conf.Int(self_initiation[,2])  
Conf.Int(self_initiation[,3])  
Conf.Int(self_initiation[,4])  
Conf.Int(self_initiation[,5])  
Conf.Int(self_initiation[,6])
```