

VALIDATING LISTENING STRATEGIES USING ORDINAL RESPONSE MODELS

YOUNG-MIN (BRIAN) KWON

SCHOOL OF MATHEMATICS AND STATISTICS

SUPERVISOR: DR IVY LIU

A THESIS SUBMITTED IN FULFILLMENT OF THE REQUIREMENT FOR THE DEGREE OF
MASTERS OF SCIENCE IN STATISTICS AND OPERATIONS RESEARCH,
VICTORIA UNIVERSITY OF WELLINGTON

2016

Acknowledgements

I would like to express my sincere gratitude towards Dr Ivy Liu for her invaluable support, remarkable patience and guidance throughout this work. This final summation would not have been possible if it was not for her knowledge and experience in the area.

I would also like to express my gratitude towards my parents Daniel and Rachel Kwon who have always believed in me. Thanks to my beloved brother Alex Kwon, for being supportive and caring sibling.

Special thanks go to my colleagues as well, especially my Manager Andrew Rae for his support and wisdom throughout this work.

Finally, I owe my gratitude to Jane Choi for the support she has shown and given me during the course for this work.

To them I dedicate this thesis.

Abstract

This thesis illustrates statistical methodology for identifying the effects of explanatory variables, for the response variables with an ordinal nature. The dataset applied to this methodology is a Listening Strategy dataset collected by The Language Learner Strategy Team at the National Institute of Education from Singapore. In this dataset, eight strategies were formed from 38 questions based on Linguistic theory. The core objective of this thesis is to validate whether 38 questions were aggregated appropriately. We use the proportional odds model, which is the most popular for ordinal responses, and the generalised estimating equations (GEE) method to analyse repeated measurements. Although there are several ways to analyse repeated categorical responses, this thesis only demonstrates the marginal approach using the GEE method. By fitting proportional odds models, we evaluate whether student's English Language test result associated with the questions are at the same level within each strategy. Results show that the English Language test result effects for the questions associated with Self-initiation, Planning, Monitoring and Evaluating, Prediction and Utilisation strategies are similar. On the other hand, the effects for the questions associated with Perceptual processing, Inferencing and Socio-affective strategies are significantly different. We also use a simulation study to show that when the ordinal response is treated as continuous, ordinary least square regression might have misleading results.

Contents

1	Introduction	11
1.1	Background	11
1.2	Review of related literature	16
1.3	Research objectives	17
1.4	Thesis structure	18
2	Preliminary Data Analysis	19
2.1	General overview	20
2.2	Data collection	21
2.3	Data cleaning	22

2.4	Listening strategies	24
2.5	Data description	26
2.5.1	Data transformation	26
2.5.2	Demographic characteristics	27
2.5.3	Other variables	28
2.5.4	English language result	30
3	General Methods	41
3.1	Proportional Odds Model	41
3.1.1	Proportional Odds Property	42
3.1.2	Latent Variable Motivation	43
3.1.3	Interpretation - Meaning of β	45
3.2	Repeated Measurements model	46
3.2.1	Marginal Approach	46
3.2.2	Generalised Estimating Equations	47
3.3	Wald Statistics	49

4	Application	51
4.1	Data preparation	52
4.2	Analysis	53
4.2.1	Self-initiation strategy	57
4.2.2	Planning strategy	58
4.2.3	Monitoring and Evaluating strategy	59
4.2.4	Perceptual processing strategy	60
4.2.5	Inferencing strategy	64
4.2.6	Prediction strategy	68
4.2.7	Utilisation strategy	69
4.2.8	Socio-affective strategy	70
5	Simulation	77
5.1	Methodology	78
5.2	First scenario	79
5.3	Second scenario	80

5.4	Third scenario	80
5.5	Fourth scenario	81
6	Discussion	83
A	Listening Strategy Questionnaire	87
B	Model Selection - Backward Elimination	91
C	Standardised EL test result effect	99
D	Simulation	103
E	R code	109
E.1	Proportional odds model using GEE approach	109
E.2	Simulation - R code	128
	References	131

List of Tables

2.1	Students' Grade Level Distribution	22
2.2	Number of grade 6 students by ethnicity	27
2.3	Number of grade 6 students by sex	28
2.4	Number of grade 6 students by school	29
2.5	Number of grade 6 students by order	29
4.1	Coefficient estimates - Perceptual processing strategy	62
4.2	Pairwise standardised EL test result effect comparisons ($\hat{\beta}_{Q_k}^{EL} - \hat{\beta}_{Q_{k'}}^{EL*Q}$) for questions within the Perceptual processing strategy	63
4.3	Coefficient estimates - Inferencing strategy	66

4.4	Pairwise standardised EL test result effect comparisons ($\hat{\beta}_{Q_k}^{EL} - \hat{\beta}_{Q_{k'}}^{EL*Q}$) for questions within the Inferencing strategy	67
4.5	Coefficient estimates - Socio-affective strategy	72
4.6	Pairwise standardised EL test result effect comparisons ($\hat{\beta}_{Q_k}^{EL} - \hat{\beta}_{Q_{k'}}^{EL*Q}$) for questions within the Socio-Affective strategy	74
B.1	Backward elimination for Self-initiation strategy using GEE approach	91
B.2	Backward elimination for Planning strategy using GEE approach	92
B.3	Backward elimination for Monitoring and Evaluating strategy using GEE approach	93
B.4	Backward elimination for Perceptual processing strategy using GEE approach	94
B.5	Backward elimination for Inferencing strategy using GEE approach	95
B.6	Backward elimination for Prediction strategy using GEE approach	96
B.7	Backward elimination for Utilisation strategy using GEE approach	97

B.8	Backward elimination for Socio-affective strategy using GEE approach	98
C.1	Pairwise standardised EL test result effect comparisons ($\hat{\beta}_{Q_k}^{EL} - \hat{\beta}_{Q_{k'}}^{EL}$) for questions within the Perceptual Processing Strategy .	99
C.2	Pairwise standardised EL test result effect comparisons ($\hat{\beta}_{Q_k}^{EL} - \hat{\beta}_{Q_{k'}}^{EL}$) for questions within the Inferencing Strategy	100
C.3	Pairwise standardised EL test result effect comparisons ($\hat{\beta}_{Q_k}^{EL} - \hat{\beta}_{Q_{k'}}^{EL}$) for questions within the Socio-affective Strategy	101
D.1	The proportion of times that the null hypothesis $H_0 : \beta_1 = 0$ is rejected at a 5% level	104
D.2	The proportion of times that the null hypothesis $H_0 : \beta_2 = 0$ is rejected at a 5% level	105
D.3	The proportion of times that the null hypothesis $H_0 : \beta_{12} = 0$ is rejected at a 5% level	106
D.4	The proportion of times that the null hypothesis $H_0 : \beta_{13} = 0$ is rejected at a 5% level	107

List of Figures

2.1	Median Standardised EL Result per Score for Questions Associated with Self-Initiation Strategy	32
2.2	Median Standardised EL Result per Score for Questions Associated with Planning Strategy	33
2.3	Median Standardised EL Result per Score for Questions Associated with Monitoring & Evaluating Strategy	34
2.4	Median Standardised EL Result per Score for Questions Associated with Perceptual Processing Strategy	35
2.5	Median Standardised EL Result per Score for Questions Associated with Inferencing Strategy	36
2.6	Median Standardised EL Result per Score for Questions Associated with Prediction Strategy	37

2.7	Median Standardised EL Result per Score for Questions Associated with Utilisation Strategy	38
2.8	Median Standardised EL Result per Score for Questions Associated with Socio-Affective Strategy	39

1 Introduction

1.1 Background

In 2005, a watershed research has been conducted by the Language Learner Strategy Research Team at the National Institute of Education (NIE) in the field of language learning strategies. The research team designed a Listening Strategy Survey and administered the questionnaire across 6 primary schools in Singapore with aims to empower learners, help teachers, and inform policy makers in regards to language education strategies. The Listening Strategy Questionnaire uses initial dataset provided by Dr. Peter Gu, Associate Professor at the School of Linguistics and Applied Language Studies, Victoria University of Wellington (for the original dataset, refer to: <http://www.victoria.ac.nz/lals/about/staff/peter-gu>). This original questionnaire consists of two main sections: first section covers students' demographical background and the second section allows the students to answer 38 listening strategy questions. These 38 listening strategy questions were then further categorised into eight different strategies drawn out from students responses which varied in a scale from 1 to 5 (1 = Never, 2 = rarely,

3 = Sometimes, 4 = Often and 5 = Always). A scale as such, consisting of a neutral middle category is known as the Likert scale, which is an “ordered categorical scale” type often referred to as an ordinal variable.

An ordinal variable is similar to a categorical variable. The difference between the two is that there is a clear ordering of the variables. Examples of ordinal variables are frequently found in measuring job satisfaction (very dissatisfied, dissatisfied, moderate, satisfied, very satisfied), or quality of life in relation to the frequency of going out to have fun (never, rarely, occasionally, often). Also, ordinal scales are extensively used in social sciences for measuring attitudes and opinions. For example, each subject’s opinion towards same-sex marriage can be effectively measured by using categories such as (strongly disagree, disagree, undecided, agree, strongly agree). Furthermore, ordinal scales are common features in medical and public health disciplines: for example, variables for describing pain (none, mild, discomforting, distressing, intense, excruciating), and injury severity (uninjured, mild injury, moderate injury, serious injury, fatal). Sometimes ordinal scales are used to describe the distribution of a continuous variable at an aggregated level, for instance age measured in years (0-20, 21-40, 41-60, 61-80, 80+).

As mentioned above, for ordinal categorical scales, there is a clear ordering of the level but a major shortcoming is that the absolute distances among them are unknown; consider the job satisfaction level measured in categories (very dissatisfied, dissatisfied, moderate, satisfied, very satisfied). There is no absolute numerical measure that effectively describes how different ‘very dissatisfied’ is from ‘dissatisfied’ and ‘satisfied’ is from ‘very satisfied’.

It is quite often that researchers treat ordinal variables as a continuous variable and use a standard linear model for analysis of an ordinal variable. This approach involves assigning scores to the ordered categories and then uses ordinary least squares (OLS) methods such as linear regression and analysis of variance (ANOVA) (Agresti, 2010). Using OLS method will still allow us to identify variables that significantly affect a response variable. However it is recommended that the simplistic approach of posing linear regression models for ordinal response scores and fitting them using OLS methods is to be avoided. Agresti (2010) has listed 5 reasons why ordinal response scores should not be treated as a continuous variable:

1. Due to the unknown distances between levels in the scale, clear-cut choice for the scores is usually not crystal clear.
2. A particular response outcome is likely to be consistent with a range of values for some underlying latent variable, and an ordinary regression analysis does not capture this effect by simply replacing a range by a single numerical value.
3. OLS does not generate estimated probabilities for the response categories at fixed settings of the explanatory variable, whereas the methods presented in this thesis is capable.
4. That approach is capable of yielding predicted values above the highest category score or below the lowest.
5. That approach ignores the fact that the variability of the responses is naturally non-constant for categorical data. For an ordinal response

variable, there is little variability at predictor values for which observations fall mainly in the highest category (or mainly in the lowest category), but there is considerable variability at predictor values for which observations tend to spread among the categories.

From the five reasons presented above, second, fourth and fifth reasons in particular may contribute to misleading results due to the effect known as “ceiling effects” and “floor effects”.

- Ceiling effect – the effect of an explanatory variable being misled, due to the ordinal response not being able to distinguish between participants who have somewhat high and those who have very high levels of response measure. The response variable puts an artificially low ceiling on how high a participant may score.
- Floor effect – the effect of an explanatory variable being misled, due to the dependent ordinal response not being able to distinguish between participants who have somewhat low and those who have very low levels of response measure. The response variable puts an artificially high floor on how low a participant may score.

In addition, it is worth noting that some researchers routinely treat ordinal variables as nominal variables. Nominal variable is a variable that has two or more categories, but no ordering nature is present. Hence, the statistical interpretation of the above, disregards the ‘ordering feature of ordinal variables. An example includes religious affiliation (Protestant, Catholic,

Jewish, Muslim, other), marital status (married, divorced, widowed, never married), favourite type of music (classical, folk, jazz, rock, other). However, it is suggested by Agresti (2010) that for categorical variable where the order matters, ordinal rather than nominal may be used for the following advantages:

- 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, slopes and means.
- Ordinal analysis is capable of using a greater range of models, and those models are more parsimonious and easier to interpret than those standard models for nominal variables.
- Ordinal methods have greater power of detecting relevant trend to the null hypothesis of no effect of an explanatory variable on the response variable.
- Interesting ordinal models apply in setting for which standard nominal models are trivial or else have too many parameters to be tested for goodness of fit.
- An ordinal analysis is capable of providing much more powerful results than an analysis that ignores ordinal nature.

1.2 Review of related literature

Some case studies have shown how treating an ordinal response, when treated as continuous, can provide misleading results. Hastie et al. (1989) illustrated a study of women in South Africa on modelling an ordinal measurement of osteoporosis in terms of age and an indicator variable for whether the woman had osteoarthritis. When using OLS, the age effect on osteoporosis score was different for those who had osteoarthritis when compared to those who had not osteoarthritis. However, when ordinal model was used, the authors found that there was no difference in the age effect towards osteoporosis score between those who had osteoarthritis and those who had not osteoarthritis. Another study that was done by McKelvey and Zavoina (1975) used an educational mobility data derived from the 1932 Occupational Changes in a Generation Survey (OCG). This particular example was to measure the effectiveness of the father's schooling and race colour on the son's schooling. An ordinal measurement was taken on both father's and son's schooling and a binary for race colour; white and non-white. The result showed that the father's schooling effect on son's schooling was 25% higher for white sons than for non-white sons when ordinal model was used. However, OLS suggested that father's schooling effect on son's schooling was no different between white sons and non-white sons.

For the listening strategy data used in this thesis, Dobbie (2015) presented a comparison of statistical methodologies for cluster analyses on ordinal response variable. Dobbie (2015) demonstrated how well student's responses

to all questions were categorised into eight listening strategies. This thesis focuses on validating eight listening strategies by taking student's EL test results into account.

1.3 Research objectives

There are two main objectives for this thesis and they are as follows:

1. Application

Validating eight listening strategies formed from 38 questions from applying Linguistic theory using each student's English language test result. If the questions are properly aggregated, the English Language test result effect for each questions within strategy should be similar.

2. Simulation

Using simulations to show the impact of ignoring ordinal nature by treating the ordinal variable as if they were continuous variable.

1.4 Thesis structure

This thesis uses ordinal response models to validate listening strategies. Chapter 2 shows preliminary analysis on listening strategy questionnaire data, designed and administered by The Language Learner Strategy Research Team at the National Institute of Education. Chapter 3 introduces the ordinal model used in this thesis. It is called the proportional odds model (McCullagh, 1980) that uses the cumulative probabilities of an ordinal response. Also a method for repeated measurements will be introduced, specifically for generalised estimating equations (GEE) (Liang and Zeger, 1986). Chapter 3 also describes Wald statistics which are test statistics used for GEE method for model selections. In Chapter 4, methods mentioned in Chapter 3 are demonstrated by applying listening strategy questionnaire data as an example. Chapter 4 presents the first objective of this thesis. The second objective is discussed in Chapter 5 by conducting simulations and lastly, Chapter 6 gives conclusions to wrap up this thesis.

2 Preliminary Data Analysis

This chapter provides preliminary description and summarisation of listening strategy questionnaire data presented by The Language Learner Strategy Research (LLSR) Team from National Institute of Education (NIE). The more advanced statistical analysis on the data, model building process on ordinal response variable, will be demonstrated later on in Chapter 4, which will focus on answering the first objective of this thesis. Chapter 2 also explains the data collection progress as well as eight listening strategies formed by using 38 questions from the listening questionnaire along with each strategy's characteristics. A crucial piece of information on each student's English Language (EL) test result is introduced. This variable is very important in this thesis as its effectiveness is measured to validate whether questions with each strategies are similar.

2.1 General overview

Gu et al. (2005) states that, to this date, only a handful of studies can be found on primary school student's use of language learner strategies. A few existing studies only apply to adults that have very little knowledge about children in primary schools. However, researchers know very little about whether the same technique can be applied to lower primary school in regards to identifying the best English learning strategy (Gu et al., 2005). Prior to identify the best English learning strategy, we need to check whether the questions formed within each strategies are suitable. This particular thesis focuses on evaluating whether these questions have similar contributions to student's EL test results. The LLSR Team from NIE has stepped in the game, seeking for the best English learning strategy by designing and distributing listening strategy questionnaire to six lower primary schools in Singapore in 2005. The purpose was to empower learners, help teachers, and inform policy makers with detailed information about language learning strategies. The questionnaire was designed based on an exploratory phase of their study.

2.2 Data collection

In selecting subjects suitable for the research purpose, LLSR Team approached the principal and the head of the English Department of each primary school to share their vision and initiatives. Cooperation from the schools was provided based on the fact that the LLSR Team was seen as a group of experts from the only teacher training institution in Singapore who were carrying out a Government-sponsored research project to help children learn better in school.

Approval were sought from schools and selected student's parents. Six Singapore primary schools participated in the research and materials were distributed to these schools. The materials included four strategy questionnaires, corresponding answer sheets, and a consent form for the teacher and student. Teachers were given a checklist and instructions for the conduction of the questionnaire, which standardised the method of data collection and facilitated the process.

The potential questions and answers part had examples of questions, where further explanation may be required for students to comprehend. One NIE researcher was present in the classroom while the questionnaire was administered in the classroom. This was to ensure that student's questions, which teachers may not know the answer to could be provided.

2.3 Data cleaning

This section will explain the listening strategy questionnaire dataset cleaning process, specifically subsetting the questionnaire dataset to older students so that we get more reliable results and in-depth details about missing observations.

Grade level

The listening strategy questionnaire was administered to grade 4, 5 and 6 students. It is suggested by the data owner, Associate Professor Peter Gu, that we limit the dataset by carrying out the analysis on grade 6 students only. The theory behind this is that grade 6 students are older than the others and would have a better sense and understanding of questions being asked in the questionnaire. From this, number of participated students has decreased from 3618 students to 1226 students. Table 2.1 illustrates the number of students by their grade level.

Table 2.1: Students' Grade Level Distribution

Grade Level	Number of students	Percent
4	1182	32.7%
5	1204	33.3%
6	1226	33.9%
Missing	6	0.2%
Total	3618	100.0%

Missing data

The problem of missing data arises quite frequently when one wishes to undertake a data analysis. The main method utilised for this thesis is the generalised estimating equations (GEE) approach (Liang and Zeger, 1986). It is stated that, in order to utilise this approach, it is required that the missing data to be missing completely at random (MCAR) (Little, 1988). The data are MCAR if the missing data represents a random sample of all observations. This means that the probability of any observation being missing does not depend on both observed and unobserved data (Little and Rubin, 2002, p. 12). MCAR is a special case of data missing at random (MAR). The data is missing at random if the missingness of the data does not depend on the components that are missing (Little and Rubin, 2002, p. 12). This means that the probability of an observation being missing depends on observed values.

For the simplicity sake, the missing observations in the listening questionnaire data were deleted before the model building process began. As a result, 187 grade 6 students' records of 1226 (15.3%) were deleted from the listening questionnaire dataset.

2.4 Listening strategies

The 38 questions which are listed in Appendix A were categorised in to eight listening strategies based on Linguistic theory. The following shows the eight strategies and their corresponding questions:

- Self-initiation strategy – Q11, Q12, Q14, Q18
- Planning strategy – Q3, Q15, Q23
- Monitoring and evaluating strategy – Q5, Q9, Q21, Q32, Q36
- Perceptual processing strategy – Q8, Q24, Q34, Q37
- Inferencing strategy – Q1, Q4, Q17, Q26, Q27, Q31, Q33
- Prediction strategy – Q6, Q7, Q13, Q29, Q35
- Utilisation strategy – Q16, Q19, Q20, Q22, Q30
- Socio-affective strategy – Q2, Q10, Q25, Q28, Q38

There are three language learning strategy types which these eight listening strategies can further be categorised into. They are metacognitive, cognitive and social-affective strategies.

The metacognitive strategies consist essentially of reflecting on the personal learning process, on the understanding of the conditions which favour it, on the organizing and planning of the activities with the view to learning, on

the self-evaluation and self-correct (Ungureanu and Georgescu, 2012). The following listening strategies fall under the metacognitive strategy:

- Self-initiation
- Planning
- Monitoring and Evaluating

Cognitive strategies are directly related to learning tasks and are used by learners when they mentally and/or physically manipulate material to be learned, or when they apply a specific technique to a learning task (O'Malley and Chamot, 1990). The following listening strategies are associated with the cognitive strategy:

- Perceptual processing
- Inferencing
- Predicting
- Utilisation and Elaboration

The socio-affective strategies involve the interaction with another person with the view to helping the language learning and the control of the affective dimension accompanying the learning process (Ungureanu and Georgescu, 2012).

2.5 Data description

This section discusses the transformation process of listening questionnaire dataset and then the potential variables of the population from the listening questionnaire dataset are introduced. The demographic characteristics are illustrated to explore the characteristics of students who participated in the study along with other potential variables. Lastly, median English Language (EL) results for each questions within the listening strategy were calculated and plotted against for the response score. This was to visually explore the relationship between the EL results and scores for each questions within each strategy.

2.5.1 Data transformation

In order to get the dataset ready for more advanced statistical analysis which will be demonstrated in Chapter 4, the listening strategy dataset is required to be transformed. The original dataset was reshaped from wide format to long format. The original dataset is in a wide format since 38 listening strategy questions of data are wide and in order to take GEE approach (Liang and Zeger, 1986), the data is required to be in a long format. This process was done by using the melt function under the R packaged known as reshape (Wickham and Hadley, 2007).

2.5.2 Demographic characteristics

There are two demographic variables to be explored for the listening questionnaire dataset; ethnicity and sex. Each demographic variable is presented in a table by counting the number of students for the corresponding variable.

Ethnicity

Table 2.2 shows the number of grade 6 students who participated in filling out the questionnaire by their ethnicity. Nearly 70% of students who participated in the study were Chinese.

Table 2.2: Number of grade 6 students by ethnicity

Ethnicity	Number of students	Percent
Chinese	711	68.4%
Malaysian	260	25.0%
Indian	47	4.5%
Eurasian	6	0.6%
Others	15	1.4%
Total	1039	100.0%

For the sake of simplicity and to validate the model building process, students that fall under the categories of non-Chinese were aggregated.

Sex

According to Table 2.3, 56.5% of students were boys and 43.5% were girls.

Table 2.3: Number of grade 6 students by sex

Sex	Number of students	Percent
Male	587	56.5%
Female	452	43.5%
Total	1039	100.0%

2.5.3 Other variables

Other variables to consider include the school students attend and the order they have received the listening questionnaire dataset.

School

Table 2.4 shows the number of grade 6 students by the school they attend. Fuchun had the most grade 6 students who participated in the study at 251, following closely was Junyuan with 223 grade 6 students. The least grade 6 students who participated in the study was Clementi at 99.

Table 2.4: Number of grade 6 students by school

School	Number of students	Percent
Fuchun	251	24.2%
Junyuan	223	21.5%
Lakeside	176	16.9%
CathHigh	161	15.5%
Jurwest	129	12.4%
Clementi	99	9.5%
Total	1039	100.0%

Order

Order variable shows whether the students received listening strategy questionnaire first or second. There were other types of strategy questionnaires that the students were asked to complete.

Table 2.5: Number of grade 6 students by order

Order	Number of students	Percent
1	511	49.2%
2	528	50.8%
Total	1039	100.0%

2.5.4 English language result

Each student had undertaken an English Language (EL) test which were marked out of 100. It was indicated that each school had their own version of EL test. As a result, each student's EL test result needs to be standardised. The following illustrates how this was taken into account:

$$x_{StdELResult} = \frac{x_{ij} - \bar{x}_j}{s_j}$$

The notation x_{ij} is the raw EL test result for the i^{th} student at j^{th} school.

The mean \bar{x}_j is

$$\bar{x}_j = \sum_i^{n_j} \frac{x_{ij}}{n_j}$$

is the sample mean EL test result for j^{th} school, where $i = 1, \dots, n_j$. The standard deviation, s_j is

$$s_j = \sqrt{\frac{\sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2}{n_j - 1}}$$

Each student's standardised EL test result is a crucial variable for this thesis in a way that its effectiveness will be measured against the scores each student has provided for 38 listening strategy questions while controlling the other available variables from the listening strategy questionnaire dataset.

For every response on each question, the median standardised EL test result was calculated, and then plotted against the score responses. The following plots provide a visual understanding around the relationship between the standardised EL test result and score responses for each question per strategy. Notice that the "score response" for each question refers to the five-likert scale

from 1 (never) to 5 (always). We calculated the median standardised EL test result for the students who responded on each categories of each questions. For example, figure 2.1, the median standardised EL test result is -0.319 for the students who responded with score 1 (never) on Q11 and the median standardised EL test result is 0.114 for the students who responded with score 5 (always). For the sake of consistency, y-axis (standardised EL test result) for each plot was fixed with the minimum of -1 and the maximum of 1. The positive relationship between the median standardised EL test result and score responses for each question is expected, which was not tested statistically as the purpose of these plots is to provide a visual understanding of the relationship between EL test result and score responses.

Figure 2.1: Median Standardised EL Result per Score for Questions Associated with Self-Initiation Strategy

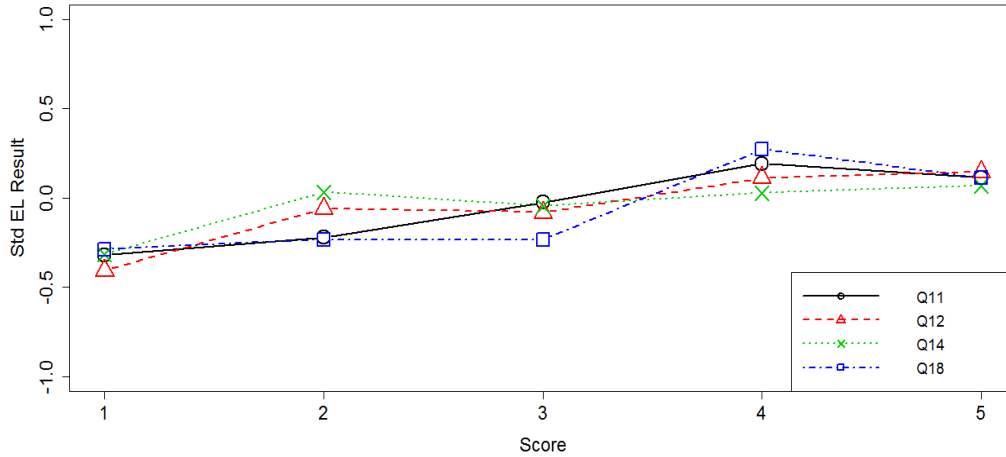


Figure 2.1 shows the following:

- The median standardised EL test result for all questions associated with self-initiation strategy tend to increase as score level increases from 1 (never) to 5 (always).
- Question 11 and question 18 have the highest median standardised EL result at score 4 (often), which then decreases at score 5 (always).
- The greatest change within the range of score level with the median standardised EL test result is the jump from score 3 (sometimes) to score 4 (often) for question 18.

- It seems that there seems to be a positive relationship between EL test result and the score level.

Figure 2.2: Median Standardised EL Result per Score for Questions Associated with Planning Strategy

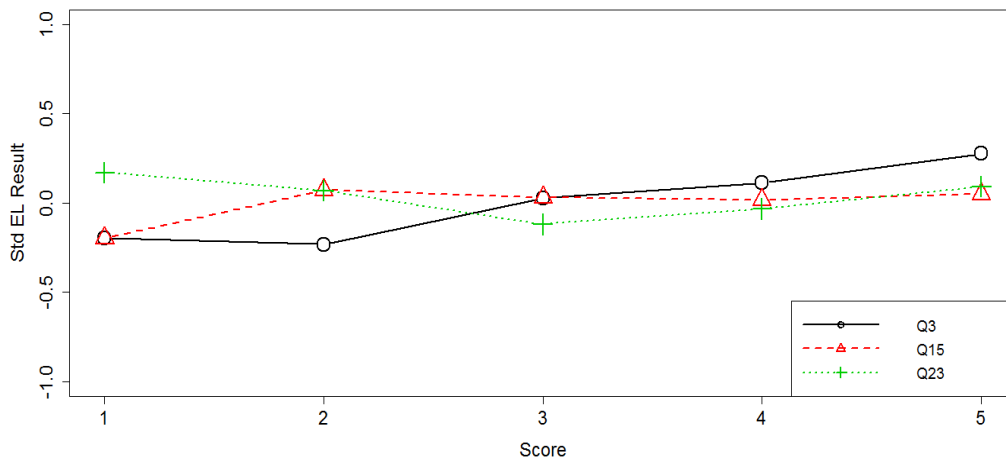


Figure 2.2 shows the following:

- The changes with the median standardised EL test result for question 15 and 23 is very small at a various level of scores.
- Question 3 illustrates a slight drop in the median standardised EL test result at score 2 (rarely) from score 1 (never), which then increases all the way up to score 5 (always).
- There appears to be positive relationship between the median standardised EL test result and the score for the questions associated with

planning strategy.

Figure 2.3: Median Standardised EL Result per Score for Questions Associated with Monitoring & Evaluating Strategy

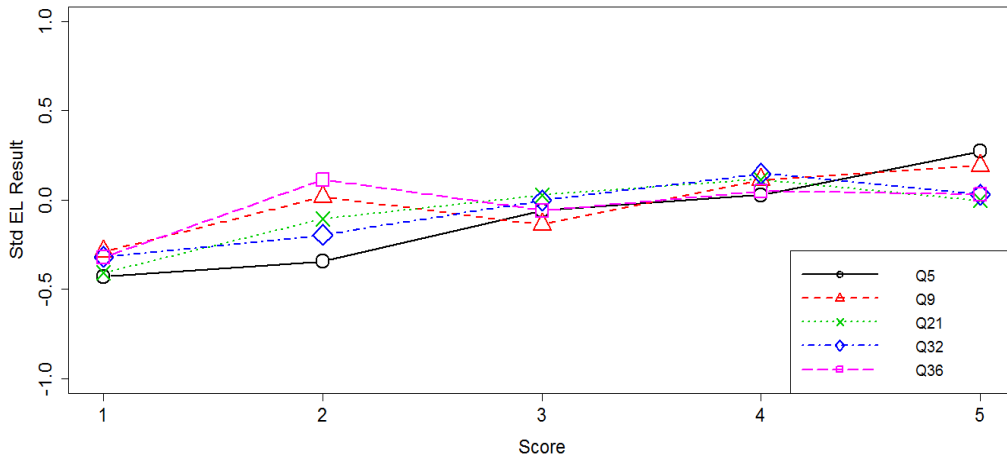


Figure 2.3 indicates the following:

- The median standardised EL test result for question 5 and 9 are the only questions that increase from score 4 (often) to 5 (always) and the rest of the questions decrease.
- Question 5 tends to show a crystal clear positive relationship between the median standardised EL test result and the score level, where it manages to score the lowest median standardised EL test result at score 1 (never) and the highest medians standardised EL test result at score 5 (always).

- The median standardised EL test result for the questions associated with monitoring & evaluating strategy appear to behave positively with the score level.

Figure 2.4: Median Standardised EL Result per Score for Questions Associated with Perceptual Processing Strategy

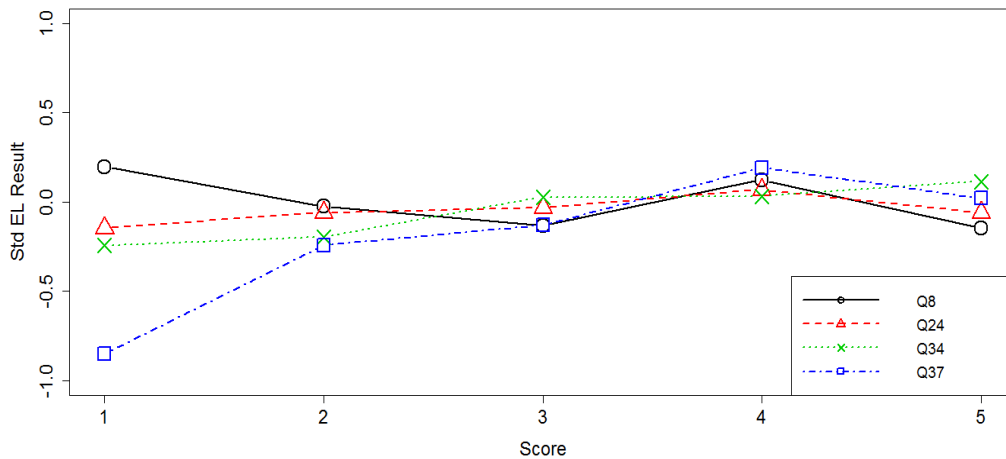


Figure 2.4 demonstrates the following:

- The median standardised EL test result for question 24 shows very little change across the level of score range.
- Question 8 has the highest median standardised EL test result at score 1 (never).
- All the median standardised EL test result appears to decrease from score 4 (often) to score 5 (always), except for question 34.

- The lowest median standardised EL test result is recorded at score 1 (never) for question 37.
- The difference in the median standardised EL test result is at its greatest at score 1 (never).

Figure 2.5: Median Standardised EL Result per Score for Questions Associated with Inferencing Strategy

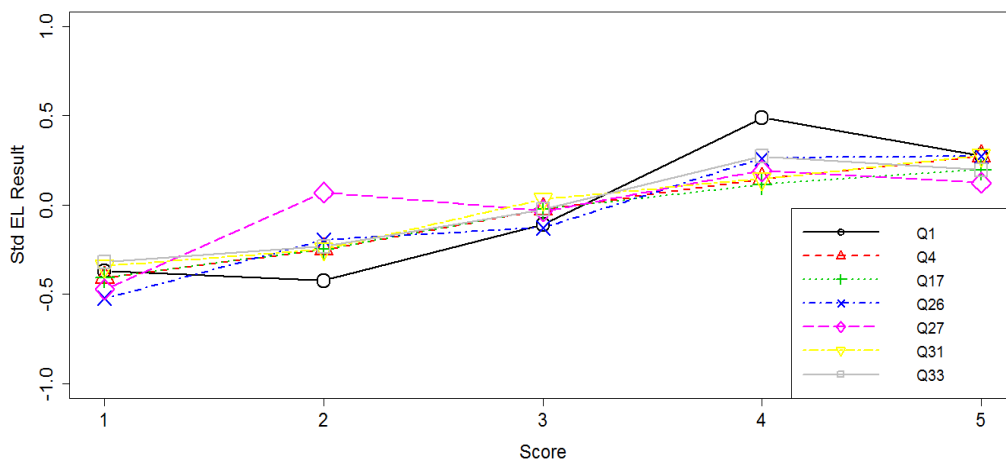


Figure 2.5 indicates the following:

- Question 1 has the highest median standardised EL test result at score 4 (often).
- The median standardised EL test result for question 27 is higher than the other questions at score 2 (rarely).

- Despite the decrease in the median standardised EL test result for question 1, 27 and 33, the median standardised EL test result for the questions associated with inferencing strategy appears to be positively related with the score level.

Figure 2.6: Median Standardised EL Result per Score for Questions Associated with Prediction Strategy

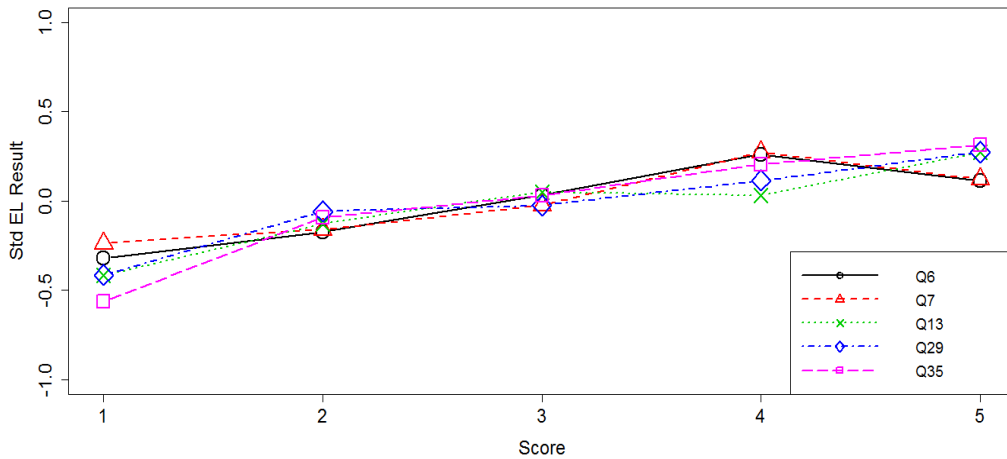


Figure 2.6 shows the following:

- Question 35 has the lowest median standardised EL test result at score 1 and the highest at score 5 (always).
- Question 6 and 7 maintains a positive relationship between the median standardised EL test result and score level until score 4 (often), which then drops at score 5 (always).

- The median standardised EL test result for question 13, 29 and 35 maintains a positive relationship to the score level.

Figure 2.7: Median Standardised EL Result per Score for Questions Associated with Utilisation Strategy

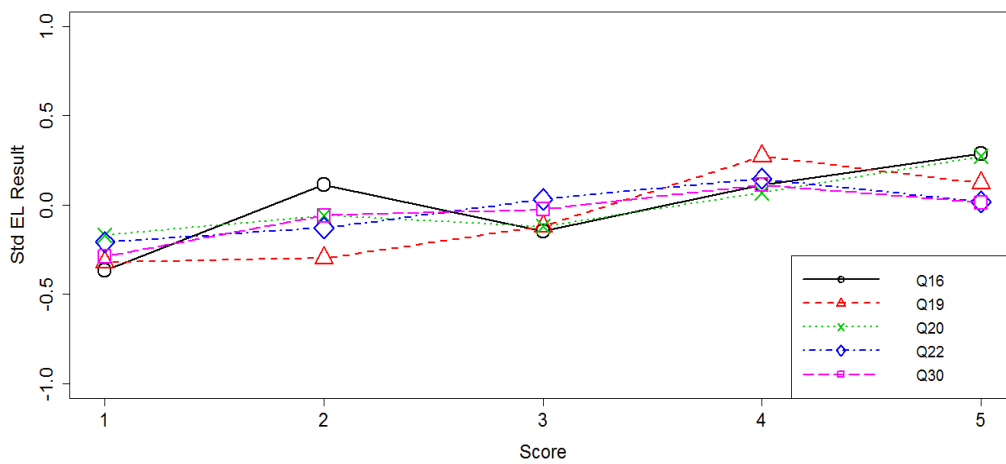


Figure 2.7 illustrates the following:

- The median standardised EL test result appears to behave similar at score 3 (sometimes), 4 (often) and 5 (always) level for question 16 and 20.
- Question 19 has the lowest median standardised EL test result at score 2 (rarely) and the highest at score 4 (often).
- There appears to be a positive relationship between median standardised EL test result and the score level.

Figure 2.8: Median Standardised EL Result per Score for Questions Associated with Socio-Affective Strategy

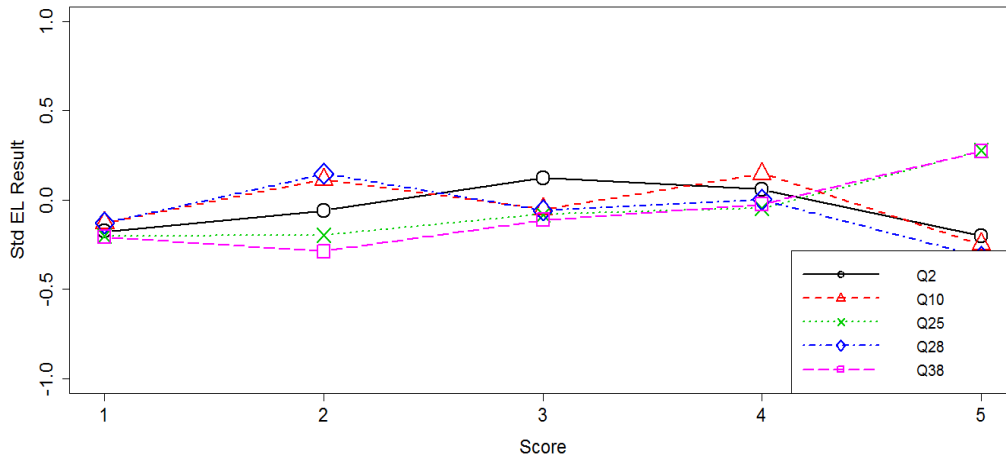


Figure 2.8 demonstrates the following:

- The median standardised EL test result for question 25 and 38 tend to behave very similar along the whole range of score level.
- Question 25 and 38 scored the high median standardised EL test result at score 5 (always), whereas question 2, 10 and 28 scored low.
- There appears to be positive relationship between the median standardised EL test result and score for question 25 and 38, but the same may not apply to the rest of the questions.

3 General Methods

This chapter discusses various statistical methods that were applied to analyse the listening strategy dataset. Section 3.1 introduces the most popular model of analysing ordinal responses, called the proportional odds model. Its property, motivation and the meaning of parameters is discussed. Section 3.2 describes the method for repeated measurements. Specifically the marginal approach (e.g. Generalised Estimating Equations) is discussed. Lastly, Section 3.3 describes the Wald test and its usage towards model selection.

3.1 Proportional Odds Model

The logistic regression is widely used for identifying the effects of explanatory variables, given the response variable is binary. So how do we analyse the effects of explanatory variables when the response variable is ordinal. Currently the most popular model for identifying the effects of explanatory variables for the ordinal response variable is the proportional odds model (McCullagh, 1980).

3.1.1 Proportional Odds Property

The proportional odds model is a class of multivariate generalised linear models used for modelling ordinal responses. This model uses cumulative probabilities up to a threshold, making the whole range of ordinal categories binary at that threshold. Let Y denote the response categorical variable that has c ordered categories. Then the general form of proportional odds model can be written as the following:

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

The column vector β of parameters describes the effects of the explanatory variables. Model (3.1) illustrates that the logit for cumulative probability j has its own intercept. The α parameters are known as cutpoints and are increasing in j , $\alpha_1 < \alpha_2 < \dots < \alpha_{c-1}$. This is the case as $P(Y \leq j)$ increases in j for each fixed value of x , and the logit is an increasing function of this probability. The ‘ $-$ ’ sign makes the interpretation of β consistent with an ordinal linear model. That is, when $\beta > 0$, the response Y is more likely to fall in a higher category as x increases. The equivalent model version of cumulative probabilities can be written as the following:

$$P(Y \leq j) = \frac{\exp(\alpha_j - \beta' \mathbf{x})}{1 + \exp(\alpha_j - \beta' \mathbf{x})} \quad j = 1, \dots, c-1 \quad (3.2)$$

and the probability at j can be written as the following:

$$P(Y = j) = \frac{\exp(\alpha_j - \beta' \mathbf{x})}{1 + \exp(\alpha_j - \beta' \mathbf{x})} - \frac{\exp(\alpha_{j-1} - \beta' \mathbf{x})}{1 + \exp(\alpha_{j-1} - \beta' \mathbf{x})} \quad (3.3)$$

with $\alpha_0 = -\infty$ and $\alpha_c = \infty$. This formula has the form of a linear combination of inverse link function, namely, inverse logit links with coefficients 1

and -1 (Agresti, 2010). The link function for the j probability is known as *composite link function* (Thompson and Baker, 1981). Model (3.1) has the same parameter β across j . This means that the explanatory effects are the same for each cumulative logits.

3.1.2 Latent Variable Motivation

A proportional odds model can be motivated by a latent variable. A latent variable, as opposed to an observable variable, is a variable that is not directly observed but is rather inferred from other variables that are observed. In order to justify the effect of β for different logits in the proportional odds model, a regression model is used for an unobserved continuous variable assumed to underlie Y (Anderson and Philips, 1981).

Let Y^* denote underlying latent variable and suppose that Y^* varies around a location parameter η , such as a mean for fixed values of explanatory variables \mathbf{x} . Given that η depends on X through $\eta(\mathbf{x}) = \beta'\mathbf{x}$, the conditional cumulative distribution function (cdf) of Y^* can be written as the following:

$$P(Y^* \leq y^* \mid \mathbf{x}) = G(y^* - \eta) = G(y^* - \beta'\mathbf{x}) \quad (3.4)$$

Suppose that $-\infty = \alpha_0 < \alpha_1 < \dots < \alpha_c = \infty$ are *cutpoints* of the continuous scale such that the observed response Y satisfies $Y = j$, given that $\alpha_{j-1} < Y^* \leq \alpha_j$. The response Y falls in category j when the latent variable falls in the j th interval of values (Agresti, 2010). Under this latent variable structure,

the following can be achieved,

$$P(Y \leq j \mid \mathbf{x}) = P(Y^* \leq \alpha_j \mid \mathbf{x}) = G(\alpha_j - \beta' \mathbf{x}). \quad (3.5)$$

The link function to apply to $P(Y \leq j \mid \mathbf{x})$ to obtain a linear predictor is G^{-1} , the inverse of the cdf for Y^* .

$$G^{-1}[P(Y \leq j \mid \mathbf{x})] = \alpha_j - \beta' \mathbf{x} \quad (3.6)$$

When G is the cdf of the standard logistic distribution, which is $G(\epsilon) = e^\epsilon / (1 + e^\epsilon)$, then G^{-1} is the logit link function (Agresti, 2010). That is, the logistic latent variable model implies the model for the observed response,

$$\text{logit}[P(Y \leq j \mid \mathbf{x})] = \alpha_j - \beta' \mathbf{x}. \quad (3.7)$$

This is the proportional odds model that has the same effects for each cumulative probability. The above derivation shows that using a cdf of the form $G(y^* - \beta' \mathbf{x})$ for the latent variable, results in $\alpha_j - \beta' \mathbf{x}$ rather than $\alpha_j + \beta' \mathbf{x}$. In practice, the negative parameterisation for the explanatory variables does not matter as long as we interpret effects appropriately (Agresti, 2010).

Agresti (2010) states that the latent variable motivation for the model explains why distribution of Y at different settings of explanatory variables are stochastically ordered. The model is sensitive to location effects, not effects whereby the variability of Y changes as the explanatory variables changes. The model usually fits quite poorly if the variability of an underlying latent continuous variable changes dramatically over the range of observed values.

3.1.3 Interpretation - Meaning of β

To help explain the proportional odds model and its interpretation of the β coefficient, let's consider the case where an explanatory variable is a single continuous variable x . For a continuous explanatory variable x and a fixed j , the cumulative logit curve is a logistic regression curve for a binary response with outcomes $Y \leq j$ and $Y > j$ (Agresti, 2013).

Let $\text{logit}[P(Y \leq j)] = \alpha_j - \beta \mathbf{x}$. The odds of having $Y \leq j$ are multiplied by $e^{-\beta}$ for every additional unit increase in x , or we say that the odds of having $Y \geq j$ are multiplied by e^{β} for every additional unit increase in x . For example, if $\beta = 0.5$, the odds of having $Y \geq j$ are multiplied by $e^{0.5} = 1.65$ for every additional unit increase in x . An additional unit increase in x has a 65% increase in the odds of $Y \geq j$.

3.2 Repeated Measurements model

Many studies often observe multiple response variables per subject. For example, a physician might want to evaluate patients who are using a placebo or a new drug treatment for a curable condition at weekly intervals using the scale (cured, improved, no change, worse). There would be a set of responses for a particular patient in this study. For the questionnaire dataset used in this thesis, each student was asked to provide a score from 1 to 5 for 38 questions. It is required that the dependency amongst each student to be taken into account for analysis.

A marginal approach is one of the ways to analyse repeated categorical response variables (Agresti, 2010). In this thesis, we will focus on the Generalised Estimating Equations (GEE) method (Liang and Zeger, 1986).

3.2.1 Marginal Approach

Let Y_t be the response at time t , where $t = 1, \dots, T$. Given observation Y_1, Y_2, \dots, Y_T on a c -category scale, by using $c - 1$ cumulative logits, we can model the marginal response distribution as follows

$$\text{logit}[P(Y_t \leq j)] = \alpha_j - \beta'_j \mathbf{x}_t, \quad j = 1, \dots, c - 1, \quad t = 1, \dots, T \quad (3.8)$$

If we replace β_j with β , the model takes the proportional odds form with the same effects for each logit (Agresti, 2010).

$$\text{logit}[P(Y_t \leq j)] = \alpha_j - \beta' \mathbf{x}_t = \alpha_j - \beta_1 x_{1t} - \beta_2 x_{2t} - \dots - \beta_k x_{kt} \quad (3.9)$$

It is said that fitting marginal models for categorical responses by Maximum Likelihood (ML) is not simple, as the models refer to marginal probabilities, whereas the likelihood function specifies the joint distribution of the multiple response variables per subject. ML fitting becomes even more difficult when explanatory variables are added (Agresti, 2010).

An alternative to ML fitting uses a multivariate generalisation of *quasi-likelihood*. The Quasi-Likelihood function was introduced by (Wedderburn, 1974), where only an assumption about the mean-variance is established without having to specify the underlying distribution. For a univariate response, the quasi-likelihood method specifies only a linear model for a link function applied to $\mu = E(Y)$ and a formula $v(\mu)$ for how the variance of Y depends on the mean. The multivariate generalisation does this for each Y_t and uses a guess for the correlation structure among Y_1, Y_2, \dots, Y_T without having to assume a particular joining probability distribution.

3.2.2 Generalised Estimating Equations

Originally, GEE methodology was introduced for modeling binary responses, and has since been extended to marginal modeling of multinomial responses (Agresti, 2010). It was then Lipsitz et al. (1994) who proposed a GEE approach for cumulative logit models with repeated ordinal responses.

The method was described by (Agresti, 2010, chp. 9) as follows. We assume a multinomial distribution for each $Y_t, t = 1, \dots, T$. For every i th subject, Y_{it} is

determined by a set of $c-1$ indicator variables, the j th one indicating whether Y_{it} falls in category j , where $j = 1, \dots, c-1$. Hence let $y_{ijt} = 1$ if observation t in subject i has outcome j ($j = 1, \dots, c-1$), and $y_{ijt} = 0$ otherwise. Let \mathbf{y}_i be the $T(c-1)$ binary indicators for i th subject for the T observations then the covariance matrix \mathbf{V}_i for \mathbf{y}_i is a $T(c-1) \times T(c-1)$ matrix. The covariance matrix \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 . Therefore, the covariance matrix \mathbf{V}_{it} for $y_{i1t}, \dots, y_{i,c-1,t}$ has entry $v_{ijtt} = P(Y_{ijt} = 1)[1 - P(Y_{ijt} = 1)]$ for the cell on the main diagonal in row i and column j and entry $-P(Y_{iht} = 1)P(Y_{ijt} = 1)$ for the cell in row h and column j with $h \neq j$.

The remaining parts of \mathbf{V}_i are $\text{Cov}(Y_{iht}, Y_{ijs})$ for $s \neq t$ which are not specified by the marginal multinomial covariances. For each pair (h, j) of outcome categories, the GEE applies a working correlation matrix for the pairs (Y_{is}, Y_{it}) of every i th subject. The working covariance matrix \mathbf{V}_i for \mathbf{y}_i specifies a pattern for $\text{Corr}(Y_{ijt}, Y_{iht})$ for each pair of outcome categories (h, j) and each pair (s, t) for every observations within an i th subject.

Let $\mu_i = E(\mathbf{y}_i)$. This is the function of the model parameters β that depends on the choice of model.

$$\mathbf{u}(\beta) = \sum_{i=1}^n \mathbf{D}_i' \mathbf{V}_i^{-1} (\mathbf{y}_i - \mu_i) = 0, \quad \text{where } \mathbf{D}_i' = \frac{\partial \mu_i'}{\partial \beta} \quad (3.10)$$

By substituting $\hat{\mu}_i$ from the model fit in \mathbf{D}_i and \mathbf{V}_i and replacing $\text{Cov}(\mathbf{Y}_i)$ by the empirical covariance matrix of \mathbf{y}_i gives an empirically adjusted sandwich

covariance matrix for the GEE estimate $\hat{\beta}$.

$$\left[\sum_{i=1}^n \mathbf{D}_i' \mathbf{V}_i^{-1} \mathbf{D}_i \right]^{-1} \left[\sum_{i=1}^n \mathbf{D}_i' \mathbf{V}_i^{-1} \text{Cov}(\mathbf{Y}_i \mathbf{V}_i^{-1} \mathbf{D}_i) \right] \left[\sum_{i=1}^n \mathbf{D}_i' \mathbf{V}_i^{-1} \mathbf{D}_i \right]^{-1} \quad (3.11)$$

The GEE method is appealing for categorical variable given its computational simplicity compared to ML approach. However as GEE method does not have a likelihood function, likelihood-based methods such as likelihood-ratio tests are not available for checking model fit and conducting inference about parameters. Therefore, we use a Wald type of test statistics to compare models and make inferences.

3.3 Wald Statistics

Wald test (Wald, 1943) was introduced for a significance test of a null hypothesis $H_0 : \beta = 0$. If the Wald test is significant for the coefficient that is being tested, this means that the β coefficient is not equal to zero and should be included in the model. On the other hand, if the Wald test is not significant, the corresponding explanatory variable is to be neglected from the model.

Consider the following, where SE is the standard error for the arbitrary parameter, β

$$z = (\hat{\beta} - \beta_0) / SE \quad (3.12)$$

Equation (3.12) has an approximate standard normal distribution when $\beta =$

β_0 and z^2 has approximately a chi-squared distribution with $df = 1$ (Agresti, 2013).

The multivariate extension for the Wald test of $H_0 : \beta = \beta_0$ has test statistic

$$W = (\hat{\beta} - \beta_0)^T [\text{cov}(\hat{\beta})]^{-1} (\hat{\beta} - \beta_0) \quad (3.13)$$

The asymptotic multivariate normal distribution for $\hat{\beta}$ implies an asymptotic chi-squared distribution for W . The df equals the number of parameters, β 's being tested (Agresti, 2013).

Model selection - backward elimination

Backward elimination (Goodman, 1971) begins with a full complex model and sequentially removes explanatory variables one by one. At each step, it removes the explanatory variable that has the least damaging effect on the model. The process stops when any further deletion leads to a significantly poorer fit. For qualitative explanatory variable with more than two categories, the process considers the entire explanatory variable rather than just individual indicator variables (Agresti, 2013, chp. 6).

For models that include interaction effect, it would not be appropriate to exclude main effect terms if interaction effect terms were to be significant (Agresti, 2013).

4 Application

In this chapter, 38 questions categorised into eight listening strategy are validated by applying techniques and methods discussed in Chapter 3. The statistical model is established for each listening strategy. The model will focuses on examining the relationship between student's standardised EL test result and the score responses that each student has provided for 38 questions while controlling the other demographic variables.

Analysis presented in this chapter was performed by statistical software, R. All the relevant code is given in Appendix E.1. The function used is `ordLORgee` from the package `multgee` (Touloumis, 2014).

4.1 Data preparation

This section describes the data manipulation process so that the structure of listening strategy dataset is appropriately prepared for the model building process.

- Certain proportion of observations was removed; only selecting grade 6 students meant number of students has decreased from 3,618 to 1,226 and the missing observations meant further 15.3% reduction in number observations, as mentioned in Section 2.3. As a result the final subset gives 1,039 students.
- Due to the lack of observations for certain ethnicities, students with the following ethnicity were categorised into non-Chinese group.
 - Malaysian
 - Indian
 - Eurasian
 - Others

Table 2.2 shows the number of students by their ethnicity.

- Student's EL test result variable was standardised by school, as each school had their own version of EL test. Therefore all the interpretation will be based on standardised EL test result rather than the raw EL test result. Equation (2.5.4) shows how EL test results were standardised.

- Score response variable for 38 questions and question variables were transposed by each student along with the demographic variables. As a result, 39,482 observations are derived (1,039 students*38 questions).

4.2 Analysis

This section describes the proportional odds model by applying GEE method. To validate the questions aggregated into eight listening strategies, by applying linguistic theory, we check whether the standardised EL test result effect is the same for the questions within each listening strategies, while controlling for the other possible confounders. In the model, the 5 Likert-scale response is the response variable, the explanatory variables include:

- Standardised EL test result (EL) - Numerical variable
- Sex (S) - Categorical variable with 2 categories (Male, Female)
- Ethnicity (E) - Categorical variable with 2 categories (Chinese, Non-Chinese)
- Order (O) - Categorical variable with 2 categories (1, 2)
- School (Sc) - Categorical variable with 6 categories (CathHigh, Clementi, Fuchun, Junyuan, JurWest, Lakeside)
- Question (Q) - Categorical variable with 38 questions (Q1,...,Q38)

We also consider many interaction terms for the model. Interactions between Sex and School is excluded as one of the schools was a boys only school. Given that, the following interaction terms are included in the models.

- Standardised EL test result \times Sex (EL*S)
- Standardised EL test result \times Ethnicity (EL*E)
- Standardised EL test result \times Order (EL*O)
- Standardised EL test result \times School (EL*Sc)
- Standardised EL test result \times Question (EL*Q)
- Sex \times Ethnicity (S*E)
- Sex \times Order (S*O)
- Sex \times Question (S*Q)
- Ethnicity \times Order (E*O)
- Ethnicity \times School (E*Sc)
- Ethnicity \times Question (E*Q)
- Order \times School (O*Sc)
- Order \times Question (O*Q)
- School \times Question (Sc*Q)

The following shows the proportional odds model including two-factor interactions.

$$\begin{aligned}
\text{logit}[P(Y_{it} \leq j)] = & \alpha_j - \beta^{EL}x_i - \beta_{S_i}^S - \beta_{E_i}^E - \beta_{O_i}^O - \beta_{Sc_i}^{Sc} \\
& - \beta_{Q_t}^Q - \beta_{S_i}^{EL*S}x_i - \beta_{E_i}^{EL*E}x_i - \beta_{O_i}^{EL*O}x_i \\
& - \beta_{Sc_i}^{EL*Sc}x_i - \beta_{Q_t}^{EL*Q}x_i - \beta_{(S_i,E_i)}^{S*E} - \beta_{(S_i,O_i)}^{S*O} \\
& - \beta_{(S_i,Q_t)}^{S*Q} - \beta_{(E_i,O_i)}^{E*O} - \beta_{(E_i,Sc_i)}^{E*Sc} - \beta_{(E_i,Q_t)}^{E*Q} \\
& - \beta_{(O_i,Sc_i)}^{O*Sc} - \beta_{(O_i,Q_t)}^{O*Q} - \beta_{(Sc_i,Q_t)}^{Sc*Q},
\end{aligned} \tag{4.1}$$

where α_j are the cutpoints and $j = 1, 2, 3, 4$ and Y_{it} indicates a likert-scale response for i^{th} student at t^{th} question with $i = 1, \dots, 1,039$ and $t = 1, \dots, 38$. As (Y_{i1}, \dots, Y_{i38}) are not independent, we use the GEE method to take the dependency into account to estimate model parameters.

In model (4.1), for i^{th} student,

- x_i = Standardised EL test result

- $S_i = \begin{cases} 1 & \text{Female} \\ 2 & \text{Male} \end{cases}$

- $E_i = \begin{cases} 1 & \text{Chinese} \\ 2 & \text{Non-Chinese} \end{cases}$

- $O_i = \begin{cases} 1 & 1 \\ 2 & 2 \end{cases}$

$$\begin{aligned}
\bullet \quad Sc_i &= \begin{cases} 1 & \text{CathHigh} \\ 2 & \text{Clementi} \\ 3 & \text{Fuchun} \\ 4 & \text{Junyuan} \\ 5 & \text{JurWest} \\ 6 & \text{Lakeside} \end{cases} \\
\bullet \quad Q_t &= \begin{cases} 1 \\ 2 \\ \vdots \\ 38 \end{cases}
\end{aligned}$$

Note that t varies for Q_t depending on which questions are categorised into the particular strategy. For example, for Planning strategy, β_3^Q becomes 1. R uses level 1 as the reference level when dealing with a categorical variable. Therefore, $\beta_1^S = \beta_1^E = \beta_1^O = \beta_1^{Sc} = \beta_1^Q = 0$.

For the model selection process, the Wald test is used in the backward elimination. Each removal of a term is tested until further deletion of a term leads to a significantly poorer fit. More specifically a term removing process is repeated until the left terms are significant at the 5% level. All the removed terms for eight listening strategies can be found in Appendix B. If the standardised EL test result effect remains the same for all questions, the interaction coefficient $\beta_{Q_t}^{EL*Q}$ should equal to zero.

4.2.1 Self-initiation strategy

Self-initiation strategy was formed by aggregating the following 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).

Table B.1 indicates that all the main terms are significant with the likert-scale response. The first interaction term to be removed was the interaction between standardised EL test result and School (EL*S), with the highest level of no evidence against the null hypothesis of no association with the response. At the end, the best fitted model included all the main terms along with interaction terms between Sex and Questions, Ethnicity and School and Order and Questions. The following shows the final proportional odds model using GEE method for Self-initiation strategy.

$$\begin{aligned} \text{logit}[P(Y_{it} \leq j)] = & \alpha_j - \beta_{x_i}^{EL} - \beta_{S_i}^S - \beta_{E_i}^E - \beta_{O_i}^O - \beta_{Sc_i}^{Sc} - \beta_{Q_t}^Q \\ & - \beta_{(S_i, Q_t)}^{S*Q} - \beta_{(E_i, Sc_i)}^{E*Sc} - \beta_{(O_i, Q_t)}^{O*Q} - \beta_{(Sc_i, Q_t)}^{Sc*Q} \end{aligned} \quad (4.2)$$

The best fitted model for Self-initiation listening strategy indicates that, while controlling for the other variables in the model, there are no significant

differences in the standardised EL test result effects between Q11, Q12, Q14 and Q18, because the interaction terms β_{Qt}^{EL*Q} are not significantly different from zero.

4.2.2 Planning strategy

Planning strategy was formed by aggregating the following questions:

- Q3 - Before I listen to something important, I ask myself what I already know about the topic.
- 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.

According to table B.2, all the main terms for Planning strategy are significant at 5% level, except for the school variable. The only significant interaction term at 5% significant level is between order and question (O*Q). The following is the best fitted proportional odds model using the GEE method for Planning listening strategy.

$$\text{logit}[P(Y_{it} \leq j)] = \alpha_j - \beta^{EL}x_i - \beta_{Si}^S - \beta_{Ei}^E - \beta_{Oi}^O - \beta_{Qt}^Q - \beta_{(O_i, Q_t)}^{O*Q} \quad (4.3)$$

The above fitted model shows that there are no significant differences in

the standardised EL test result effects between Q3, Q15 and Q23, while controlling for other variables, for Planning listening strategy.

4.2.3 Monitoring and Evaluating strategy

Monitoring and evaluating strategy was formed by aggregating the following 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 B.3 shows that all the main terms are removed except for the Sex variable for Monitoring and Evaluating strategy. Interaction term between Sex and Questions was the first term to be removed from the full model with the highest level of no evidence against the null hypothesis. Interaction terms that provided the highest level of evidence against the null hypothesis were

between Ethnicity and standardised EL test result, School and Questions. Interaction term between Order and Questions was also removed. The following shows the best fitted proportional odds model using GEE method for Monitoring and Evaluating listening strategy.

$$\begin{aligned} \text{logit}[P(Y_{it} \leq j)] = & \alpha_j - \beta_{x_i}^{EL} - \beta_{E_i}^E - \beta_{O_i}^O - \beta_{Sc_i}^{Sc} - \beta_{Q_t}^Q \\ & - \beta_{E_i}^{EL*E} x_i - \beta_{(E_i, Sc_i)}^{E*Sc} - \beta_{(E_i, Q_t)}^{E*Q} - \beta_{(O_i, Q_t)}^{O*Q} \end{aligned} \quad (4.4)$$

The interaction term between standardised EL test result and Questions was removed from the final model. This indicates that there are no significant differences in the standardised EL test result effects for the questions associated with Monitoring and Evaluating strategy, while controlling for other variables.

4.2.4 Perceptual processing strategy

Perceptual processing strategy was formed by aggregating the following questions:

- 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 B.4 shows that the first explanatory variable to be removed was the interaction between sex and questions with the highest level of no evidence against the null hypothesis of no association with the score response variable. The interaction between order and school was very marginal, which was removed at the last step of backward elimination process. All the main terms are included in the model as they are all significantly related to the likert-scale response variable at a 5% significant level. The following is the final best fitted proportional odds model using GEE method for Perceptual processing strategy.

$$\begin{aligned} \text{logit}[P(Y_{it} \leq j)] = & \alpha_j - \beta^{EL}x_i - \beta_{S_i}^S - \beta_{E_i}^E - \beta_{O_i}^O - \beta_{S_{ci}}^{Sc} - \beta_{Q_t}^Q \\ & - \beta_{Q_t}^{EL*Q}x_i - \beta_{(S_i,O_i)}^{S*O} - \beta_{(E_i,SC_i)}^{E*Sc} - \beta_{(E_i,Q_t)}^{E*Q} - \beta_{(O_i,Q_t)}^{O*Q} \end{aligned} \quad (4.5)$$

The interactions between questions and standardised EL test result are significant at a 5% significant level for Perceptual processing strategy. This suggests that there are significant differences in standardised EL test result effects among the questions associated with Perceptual processing strategy, while controlling for other variables. Figure 2.4 supports the idea of the significant interaction between questions and standardised EL test result for Perceptual processing strategy as Q8 has a different pattern to Q24, Q34 and Q37 across the response levels. Table 4.1 shows the estimates for the standardised EL test result effects for proportional odds model using GEE method for Perceptual processing strategy.

Table 4.1: Coefficient estimates - Perceptual processing strategy

Coefficient	Estimate	Standard Error	P-value
β^{EL}	-0.075	0.054	0.169
$\beta_{Q_8}^{EL*Q}$	0	—	—
$\beta_{Q_{24}}^{EL*Q}$	0.244	0.066	0.000
$\beta_{Q_{34}}^{EL*Q}$	0.246	0.065	0.000
$\beta_{Q_{37}}^{EL*Q}$	0.244	0.070	0.000

As briefly described, R treats the lowest category level as the reference level. Therefore $\beta_8^{EL*Q} = 0$. In other words, the estimated standardised EL test result effect on the response for Q8 is $\hat{\beta}^{EL} + \hat{\beta}_{Q_8}^{EL*Q} = -0.075$. As for Q24, Q34 and Q37, standardised EL test result effect can be derived by adding β^{EL} with $\beta_{Q_k}^{EL*Q}$, where $k = 24, 34, 37$. For example, the standardised EL test result effect for Q24 is estimated by $(-0.075 + 0.244) = 0.170$.

Table 4.1 shows that the standardised EL test result has no effect on the response for Q8. However, Q24, Q34 and Q37 have strong evidence against the null hypothesis of no differences in standardised EL test result effects when compared to Q8. The pairwise standardised EL test result effect comparisons amongst questions associated with Perceptual processing strategy is shown in Table 4.2.

For Perceptual processing strategy, at every additional unit increase in standardised EL test result, while controlling for the other variables, the estimated odds of students responding at a high level on

- question 8 are multiplied by 0.928 ($e^{-0.075}$)
- question 24 are multiplied by 1.185 ($e^{-0.075+0.244}$)
- question 34 are multiplied by 1.187 ($e^{-0.075+0.246}$)
- question 37 are multiplied by 1.185 ($e^{-0.075+0.244}$)

Table 4.2: Pairwise standardised EL test result effect comparisons ($\hat{\beta}_{Q_k}^{EL} - \hat{\beta}_{Q_{k'}}^{EL*Q}$) for questions within the Perceptual processing strategy

$Q_{k'}$	Q_k			
	Q8	Q24	Q34	Q37
Q8	0	0.244**	0.246**	0.244**
Q24		0	0.000	0.000
Q34			0	0.000
Q37				0

* Significant level at between 0.01 and 0.05

** Significant level at < 0.01

Table 4.2 indicates that, while controlling for the other variables in the model, the standardised EL test result effect for Q8 is significantly different from Q24, Q34 and Q37 and there are no differences in the standardised EL test

result effect amongst Q24, Q34 and Q37. The p-values of testing the null hypothesis of $\beta_{Q_k}^{EL*Q} - \beta_{Q_{k'}}^{EL*Q} = 0$ for $k \neq k' = 8, 24, 34, 37$ can be found in Appendix C.

4.2.5 Inferencing strategy

Inferencing strategy was formed by aggregating the following 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.
- 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 what I will hear next.
- Q31 - When I don't understand something, I use my knowledge about text type to help me understand.
- Q33 - When I don't understand something, I use my knowledge about the topic to guess.

Table B.5 indicates that all the interaction terms are removed except for the interaction between order and questions, and standardised EL test result and questions. This indicates that the standardised EL test result effects for those questions associated with Inferencing strategy are significantly different. All the main terms have evidence against the null hypothesis of no association with the score response variable. The following shows the best fitted proportional odds model for using GEE method for Inferencing strategy.

$$\begin{aligned} \text{logit}[P(Y_{it} \leq j)] = & \alpha_j - \beta^{EL}x_i - \beta_{S_i}^S - \beta_{E_i}^E - \beta_{O_i}^O \\ & - \beta_{Sc_i}^{Sc} - \beta_{Q_t}^Q - \beta_{Q_t}^{EL*Q}x_i - \beta_{(O_i, Q_t)}^{O*Q} \end{aligned} \quad (4.6)$$

The interaction between standardised EL test result and questions have evidence against the null hypothesis. This suggests that the standardised EL test result effects are the same for the questions associated with Inferencing strategy. It is required then that standardised EL test result effect for the Q1, Q4, Q17, Q26, Q27, Q31 and Q33 to be quantified separately. Figure 2.5 supports this logic as questions 1 and 31 have a different pattern compared to other questions under the Inferencing strategy. Table 4.3 shows the estimates for the standardised EL test result effects for proportional odds model using the GEE method for Inferencing strategy.

Table 4.3: Coefficient estimates - Inferencing strategy

Coefficient	Estimate	Standard Error	P-value
β^{EL}	0.435	0.049	$< 2e-16$
$\beta_{Q_1}^{EL*Q}$	0	—	—
$\beta_{Q_4}^{EL*Q}$	-0.123	0.061	0.044
$\beta_{Q_{17}}^{EL*Q}$	-0.122	0.057	0.032
$\beta_{Q_{26}}^{EL*Q}$	-0.112	0.058	0.054
$\beta_{Q_{27}}^{EL*Q}$	-0.198	0.062	0.001
$\beta_{Q_{31}}^{EL*Q}$	-0.142	0.063	0.024
$\beta_{Q_{33}}^{EL*Q}$	-0.151	0.060	0.012

The $\beta_{Q_1}^{EL*Q}$ was used as a reference level for the Inference strategy, therefore the estimated standardised EL test result effect on the response for Q1 is $\hat{\beta}^{EL} + \hat{\beta}_{Q_1}^{EL*Q} = 0.435$. Table 4.3 shows that Q1 has strong evidence against the null hypothesis of no association between the standardised EL test result and the response. The rest of the questions associated with the Inference strategy provides evidence against the null hypothesis of no difference in the standardised EL test result effect compared to Q1. The pairwise standardised EL test result effect comparisons amongst questions under Inferencing strategy is shown in Table 4.4.

For Inference strategy, at every additional unit increase in standardised EL test result, while controlling for the other variables, the estimated odds of students responding at a high level on

- question 1 are multiplied by 1.545 ($e^{0.435}$)
- question 4 are multiplied by 1.366 ($e^{0.435-0.123}$)
- question 17 are multiplied by 1.368 ($e^{0.435-0.122}$)
- question 26 are multiplied by 1.381 ($e^{0.435-0.112}$)
- question 27 are multiplied by 1.267 ($e^{0.435-0.198}$)
- question 31 are multiplied by 1.341 ($e^{0.435-0.142}$)
- question 33 are multiplied by 1.328 ($e^{0.435-0.151}$)

Table 4.4: Pairwise standardised EL test result effect comparisons ($\hat{\beta}_{Q_k}^{EL} - \hat{\beta}_{Q_{k'}}^{EL*Q}$) for questions within the Inferencing strategy

$Q_{k'}$	Q_k						
	Q1	Q4	Q17	Q26	Q27	Q31	Q33
Q1	0	-0.123*	-0.122*	-0.112*	-0.198**	-0.142*	-0.151*
Q4		0	0.001	0.011	-0.076	-0.019	-0.028
Q17			0	0.010	-0.076	-0.020	-0.029
Q26				0	-0.086	-0.030	-0.039
Q27					0	0.057	0.047
Q31						0	-0.009
Q33							0

* Significant level at between 0.01 and 0.05

** Significant level at < 0.01

Table 4.4 shows that, while controlling for the other variables, the standardised EL test result effect for Q1 is significantly different from the rest of the questions associated with the Inferencing strategy and there is no evidence to state that the differences between Q4, Q17, Q26, Q27, Q31 and Q33 are significant. The p-value of testing the null hypothesis of $\beta_{Q_k}^{EL*Q} - \beta_{Q_{k'}}^{EL*Q}$ for $k \neq k' = 1, 4, 17, 26, 27, 31, 33$ can be found in Appendix C.

4.2.6 Prediction strategy

Prediction strategy was formed by aggregating the following 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 B.6 shows that the first explanatory variable to be removed is interaction between ethnicity and school (E*Sc), providing no evidence against the null hypothesis of no association with the score response variable. All the main terms are being kept in the model, except for sex variable. The following model shows the best fitted proportional odds model using GEE method for Prediction strategy.

$$\begin{aligned} \text{logit}[P(Y_{it} \leq j)] = & \alpha_j - \beta^{EL}x_i - \beta_{E_i}^E - \beta_{O_i}^O - \beta_{Sc_i}^{Sc} - \beta_{Q_t}^Q \\ & - \beta_{Sc_i}^{EL*Sc}x_i - \beta_{(E_i,O_i)}^{E*O} - \beta_{(O_i,Sc_i)}^{O*Sc} - \beta_{(O_i,Q_t)}^{O*Q} \end{aligned} \quad (4.7)$$

The final model suggests that, while controlling for other variables, there is no significant differences in standardised EL test result effect for the questions associated with Prediction strategy.

4.2.7 Utilisation strategy

Utilisation strategy was formed by aggregating the following questions:

- Q16 - I try to find problems with what the speaker has said so that I can avoid the same mistakes.
- 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 summarise what I heard in my mind.

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

Table B.7 shows that all the interaction terms, except for the interaction between order and questions (O*Q), are being removed as their relationship with score response variable is not statistically significant at 5% level. The least significant interaction term is the interaction between standardised EL test result and order (EL*O). Sex is the only main term being removed from the final model. The following shows the best fitted proportional odds model using GEE method for Utilisation strategy.

$$\text{logit}[P(Y_{it} \leq j)] = \alpha_j - \beta^{EL}x_i - \beta_{E_i}^E - \beta_{O_i}^O - \beta_{Sc_i}^{Sc} - \beta_{Q_t}^Q - \beta_{(O_i, Q_t)}^{O*Q} \quad (4.8)$$

The final model suggests that the differences in the standardised EL test result effects for questions associated with the Utilisation strategy, while controlling for the other variables, are not significant.

4.2.8 Socio-affective strategy

Socio-affective strategy was formed by aggregating the following 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 in 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.

Table B.8 shows that all the interaction terms are being removed except for the interaction between standardised EL test result and questions, sex and questions, order and questions and school and questions. This indicates that standardised EL test result effects for the questions associated with Socio-affective strategy are different. All the main terms have evidence against the null hypothesis of no association with the likert-scale response. The following shows the best fitted proportional odds model for using the GEE approach for the Socio-affective strategy.

$$\begin{aligned} \text{logit}[P(Y_{it} \leq j)] = & \alpha_j - \beta^{EL}x_i - \beta_{S_i}^S - \beta_{E_i}^E - \beta_{O_i}^O - \beta_{Sc_i}^{Sc} \\ & - \beta_{Q_t}^Q - \beta_{Q_t}^{EL*Q}x_i - \beta_{(S_i,Q_t)}^{S*Q} - \beta_{(O_i,Q_t)}^{O*Q} - \beta_{(Sc_i,Q_t)}^{Sc*Q} \end{aligned} \quad (4.9)$$

As model (4.9) indicates, the interaction terms between standardised EL test result and questions under Socio-affective strategy, $\beta_{Q_t}^{EL*Q}$, are significant. As a result, it is required that standardised EL test result effects for Q2, Q10, Q25, Q28 and Q38 to be quantified individually. Figure 2.8 supports this logic as standardised EL test result for Q25 and Q38 is higher than the rest of the questions at the response level 5. Table 4.5 shows the summary statistics for the standardised EL test result effects for proportional odds model using GEE method for the Socio-affective strategy.

Table 4.5: Coefficient estimates - Socio-affective strategy

Coefficient	Estimate	Standard Error	P-value
β^{EL}	0.068	0.051	0.186
$\beta_{Q_2}^{EL*Q}$	0	—	—
$\beta_{Q_{10}}^{EL*Q}$	-0.094	0.065	0.149
$\beta_{Q_{25}}^{EL*Q}$	0.159	0.069	0.022
$\beta_{Q_{28}}^{EL*Q}$	-0.092	0.063	0.145
$\beta_{Q_{38}}^{EL*Q}$	0.223	0.072	0.002

$\beta_{Q_2}^{EL*Q}$ was used as a reference level for the Socio-affective strategy, hence the estimated standardised EL test result effect on the response for Q2 is $\hat{\beta}^{EL} + \hat{\beta}_{Q_2}^{EL*Q} = 0.068$. Table 4.5 shows that Q2 has no evidence against the null hypothesis of no association with the likert-scale response. The standardised EL test result effects for Q10 and Q28 show no evidence against the null hypothesis of no difference in standardised EL test result effect compared to Q2. Whereas, Q25 and Q38 show evidence against the same null hypothesis. The pairwise standardised EL test result effect comparisons amongst questions under Socio-affective strategy are shown in Table 4.6.

For Socio-affective strategy, at every additional unit increase in standardised EL test result, while controlling for the other variables, the estimated odds of students responding at a high level on

- question 2 are multiplied by 1.070 ($e^{0.068}$)
- question 10 are multiplied by 0.974 ($e^{0.068-0.094}$)
- question 25 are multiplied by 1.254 ($e^{0.068+0.159}$)
- question 28 are multiplied by 0.976 ($e^{0.068-0.092}$)
- question 38 are multiplied by 1.338 ($e^{0.068+0.223}$)

Table 4.6 shows the pairwise standardised EL test result effects between questions under the Socio-affective strategy. The table shows $\beta^{EL*Q_k} - \beta^{EL*Q_{k'}}$ for $k \neq k' = 2, 10, 25, 28, 38$.

Table 4.6: Pairwise standardised EL test result effect comparisons ($\hat{\beta}_{Q_k}^{EL} - \hat{\beta}_{Q_{k'}}^{EL*Q}$) for questions within the Socio-Affective strategy

$Q_{k'}$	Q_k				
	Q2	Q10	Q25	Q28	Q38
Q2	0	-0.094	0.159*	-0.092	0.223**
Q10		0	0.253**	0.002	0.317**
Q25			0	-0.253**	0.064
Q28				0	0.315**
Q38					0

* Significant level at between 0.01 and 0.05

** Significant level at < 0.01

According to the Table 4.6, the following pairwise questions shows significant differences in the standardised EL test result effects, while controlling for the other variables:

- Q2 and Q25
- Q2 and Q38
- Q10 and Q25
- Q10 and Q38
- Q25 and Q28
- Q28 and Q38

On the other hand, the following shows no significant differences in the standardised EL test result effects, while controlling for the other variables:

- Q2 and Q10
- Q2 and Q28
- Q10 and Q28
- Q25 and Q38

The p-value of testing the null hypothesis of $\beta_{Q_k}^{EL*Q} - \beta_{Q_{k'}}^{EL*Q}$ for $k \neq k' = 2, 10, 25, 28, 38$ can be found in Appendix C.

5 Simulation

In this chapter, simulation is illustrated to show the impact of ignoring ordinal nature by treating the ordinal response as if they were continuous. The data were simulated from a proportional odds model for 4 different scenarios. Each simulated dataset was then fitted with the proportional odds model, the true model, and an OLS regression model. The nominal level (type I error) was evaluated for both models. Type I error refers to the error of rejecting the null hypothesis given that it is actually true. For example, when x does not have an effect on the response, the probability of making a wrong conclusion to claim that there is a significant effect is the type I error. The simulation was done in R and the code can be found in Appendix E.2. The function used is `vglm` from the package VGAM (Yee et al., 2010).

5.1 Methodology

One hundred observations are generated (sample size $n = 100$). These were generated as follows; $\mathbf{x}_1 \sim N(\mu = 1, \sigma^2 = 1)$, $\mathbf{x}_2 \sim Bin(n = 1, p = 0.5)$ and $\mathbf{x}_3 \sim N(\mu = 1, \sigma^2 = 1)$. These explanatory variables are then applied along with other associated parameters to the model (3.2) to generate a response variable, Y . The associated parameters are as follows:

- Tables D.1 - D.4 show the value of β coefficients.
- We only consider three explanatory variables \mathbf{x}_1 , \mathbf{x}_2 and \mathbf{x}_3 .
- Cutpoints, α_j where $j = 1, \dots, c - 1$, which are obtained such that the response categories have equal probabilities when $x_1 = x_2 = x_3 = 0$. For example, when $c = 5$; $\alpha_1 = \log(\frac{1}{4})$, $\alpha_2 = \log(\frac{2}{3})$, $\alpha_3 = \log(\frac{3}{2})$ and $\alpha_4 = \log(4)$.
- Let the number of categories, $c = 3, 5, 7$.

The generated dataset is then fitted with the proportional odds model and OLS regression model. For each scenario, 5,000 datasets were simulated. Table D.1 - D.4 show the proportion of times that the null hypothesis are rejected over 5,000 simulated dataset, where the null hypothesis varies depending on each scenario. The significance level is chosen at 5%. If the model is appropriate, the proportion of times that the null hypothesis is rejected should be similar to the 5% level.

5.2 First scenario

The true model in the first scenario is as follows:

$$\text{logit}[P(Y_i \leq j)] = \alpha_j - \beta_1 \mathbf{x}_1 - \beta_2 \mathbf{x}_2 \quad \text{where } j = 1, \dots, c-1 \quad (5.1)$$

where $\beta_1 = 0$ and β_2 is in the range of $(0.25, 2.50)$. This model implies that \mathbf{x}_1 has no effect on Y . Two models were fitted – the proportional odds model (5.1) and the regression model (5.2) with the form:

$$E(Y) = \alpha_j + \beta_1 \mathbf{x}_1 + \beta_2 \mathbf{x}_2 \quad (5.2)$$

We tested $H_0 : \beta_1 = 0$ against $H_1 : \beta_1 \neq 0$.

Table D.1 shows that the results can be misleading when β_2 is high. When $\beta_2 = 2.25$ and $c = 3$, type I error for the null hypothesis under the OLS method is conservative. It does not reject the null hypothesis as often as it should. In contrast, the same β_2 coefficient and $c = 5, 7$, type I error for the null hypothesis under the OLS is liberal. It rejects the null hypothesis too often. However, overall speaking, the performance of the OLS method is quite similar to the proportional odds model.

5.3 Second scenario

The true model (5.1) in the second scenario sets $\beta_2 = 0$ and β_1 in the range of (0.25, 2.50). This model implies that \mathbf{x}_2 has no effect on Y . Two models were fitted – the proportional odds model (5.1) and the regression model (5.2) and tested $H_0 : \beta_2 = 0$ against $H_1 : \beta_2 \neq 0$.

Table D.2 shows that the performance of the OLS method is as good as the proportional odds model. For this case, treating an ordinal response as continuous does not have disadvantages in terms of keeping the nominal level as 5%.

5.4 Third scenario

The true model for the third scenario is as follows:

$$\text{logit}[P(Y \leq j)] = \alpha_j - \beta_1 \mathbf{x}_1 - \beta_2 \mathbf{x}_2 - \beta_{12} \mathbf{x}_1 \mathbf{x}_2 \quad \text{where } j = 1, \dots, c-1 \quad (5.3)$$

where $\beta_{12} = 0$, β_1 is in the range of (0.25, 1.00), and β_2 is in the range of (1.00, 2.50). This model implies that there is no $\mathbf{x}_1 * \mathbf{x}_2$ interaction effect on Y . Two models were fitted – the proportional odds model (5.3) and the regression model (5.4) with the form of:

$$E(Y) = \alpha_j + \beta_1 \mathbf{x}_1 + \beta_2 \mathbf{x}_2 + \beta_{12} \mathbf{x}_1 \mathbf{x}_2 \quad (5.4)$$

We tested $H_0 : \beta_{12} = 0$ against $H_1 : \beta_{12} \neq 0$.

Table D.3 indicates that type I error for $H_0 : \beta_{12} = 0$ under OLS method appears to be valid until β_1 and β_2 starts increasing. Given that β_1 and β_2 have high coefficients, the OLS method becomes too liberal. It rejects the null hypothesis too often. For example, at $\beta_1 = 1.00$ and $\beta_2 = 2.50$, the proportions of times that the null hypothesis is rejected under OLS method are 37.9%, 34.7% and 30.1% when the numbers of response categories are 3, 5 and 7 respectively.

5.5 Fourth scenario

The true model in the fourth scenario is as follows:

$$\text{logit}[P(Y \leq j)] = \alpha_j - \beta_1 \mathbf{x}_1 - \beta_3 \mathbf{x}_3 - \beta_{13} \mathbf{x}_1 \mathbf{x}_3 \quad \text{where } j = 1, \dots, c-1 \quad (5.5)$$

where $\beta_{13} = 0$, β_3 is in the range of (0.25, 1.00), and β_1 is in the range of (0.25, 2.50). This model implies that there is no $\mathbf{x}_1 * \mathbf{x}_3$ intersection effect on Y . Two models were fitted – the proportional odds model (5.5) and the regression model (5.6) with the form of:

$$E(Y) = \alpha_j + \beta_1 \mathbf{x}_1 + \beta_3 \mathbf{x}_3 + \beta_{13} \mathbf{x}_1 \mathbf{x}_3 \quad (5.6)$$

We tested $H_0 : \beta_{13} = 0$ against $H_1 : \beta_{13} \neq 0$.

Table D.4 shows that when β_3 is low, the type I error for $H_0 : \beta_{13} = 0$ under OLS method is too conservative. This means that it does not reject the null hypothesis as often as it should and for high β_3 , the null hypothesis is too liberal under OLS method, meaning that it rejects the null hypothesis

too often. For example, at $\beta_1 = \beta_3 = 0.50$, the proportions of times that the null hypothesis is rejected are 3.5%, 3.9% and 4.0% when number of response categories are 3, 5 and 7 respectively. At $\beta_1 = 0.75$, $\beta_3 = 2.50$, the proportions of times that the null hypothesis is rejected are 11.7%, 13.3% and 13.5% when number of response categories are 3, 5 and 7 respectively.

From Tables D.1 - D.4, we can conclude that treating the ordinal response as continuous could result in misleading results on the interaction effects, but it does not affect the main effects too severely.

6 Discussion

In this thesis, listening strategy questionnaire data presented by The LLSR Team from NIE, along with student's EL test result, were utilised to illustrate statistical models when dealing with ordinal response variable. The proportional odds model using GEE method was applied to validate how well questions were aggregated to form eight strategies, which was based on Linguistic theory. This was done by measuring the effectiveness of student's standardised EL test result for questions within each strategy. If the standardised EL test result effect was similar for the questions associated with the strategy, we concluded that the strategy was formed appropriately.

The proportional odds model using the GEE approach showed that there were no significant differences in the standardised EL test result effects for the questions associated with the Self-initiation, Planning, Monitoring and Evaluating, Prediction and Utilisation strategies. It implies that the questions were formed well for these strategies. On the other hand, the standardised EL test result effects for the questions associated with the Perceptual processing, Inferencing and Socio-affective strategies were significantly different. We further identified the questions which are different from the others.

For the Perceptual processing strategy, only one question showed significant difference in the standardised EL test result effect compared to the rest of the questions associated with the strategy. Question 8 had a significantly lower standardised EL test result effect towards the likert-scale response compared to questions 24, 34 and 37. The pairwise standardised EL test result effect comparisons amongst questions 24, 34 and 37 showed no significant difference.

Questions associated with the Inferencing strategy had a similar characteristic as the Perceptual processing strategy where only one question showed significant difference in the standardised EL test result effect. Question 1 had a significantly higher standardised EL test result effect towards the likert-scale response compared to questions 4, 17, 26, 27, 31 and 33. The pairwise standardised EL test result effect comparisons amongst questions 4, 17, 26, 27, 31 and 33 showed no significant differences.

For Socio-affective strategy, the pairwise standardised EL test result effect comparisons amongst questions 2 and 25, questions 2 and 38, questions 10 and 25, questions 10 and 38, questions 25 and 38 and questions 28 and 38 were significantly different. On the other hand, no significant differences in the standardised EL test result effects were detected between questions 2 and 10, questions 2 and 28, questions 10 and 28 and questions 25 and 38. The difference in the standardised EL test result effects was greatest between questions 28 and 38.

The second main objective of this thesis was to show the impact of ignoring ordinal nature by treating the ordinal variable as if they were a continuous variable. This was illustrated in Chapter 5 by running a simulation. The result showed that if we were to treat the ordinal response as continuous, the result can be misleading on the interaction effects. However it is not severe for the main effects.

The opportunity for future work in this research could identify the best listening strategy related to the English Language test result by fitting the proportional odds model using the GEE method. However we need to aggregate the questions for each strategies so that we can compare either strategies, instead of comparing 38 questions.

A Listening Strategy Questionnaire

This part will list thirty eight questions that were listed in the Listening Strategy Questionnaire designed by The Language Learner Strategy Research Team at National Institute Education.

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.
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 what the speaker has said so that I can avoid the same mistakes.
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 summarise what I heard in my mind.
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 in 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 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 hear 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 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.

B Model Selection - Backward Elimination

This part shows the backward elimination using GEE approach for eight listening strategies at the 5% significant level.

Table B.1: Backward elimination for Self-initiation strategy using GEE approach

Order	Explanatory variable	P-value
1	$EL \times S$	0.944
2	$S \times R$	0.935
3	$EL \times R$	0.866
4	$EL \times O$	0.289
5	$S \times O$	0.182
6	$EL \times S$	0.153
7	$EL \times S$	0.129
8	$EL \times S$	0.240
9	$EL \times S$	0.380
10	$EL \times S$	0.088

Table B.2: Backward elimination for Planning strategy using GEE approach

Order	Explanatory variable	P-value
1	$S \times O$	0.865
2	$S \times Q$	0.884
3	$EL \times Sc$	0.749
4	$EL \times O$	0.992
5	$EL \times R$	0.906
6	$R \times O$	0.731
7	$S \times R$	0.684
8	$Sc \times Q$	0.640
9	$R \times Q$	0.607
10	$O \times Sc$	0.515
11	$EL \times S$	0.366
12	$EL \times Q$	0.090
13	$R \times Sc$	0.090
14	Sc	0.080

Table B.3: Backward elimination for Monitoring and Evaluating strategy using GEE approach

Order	Explanatory variable	P-value
1	$S \times Q$	0.899
2	$Sc \times Q$	0.517
3	$EL \times O$	0.431
4	$S \times O$	0.377
5	$EL \times S$	0.306
6	$R \times O$	0.237
7	$S \times R$	0.201
8	$EL \times Q$	0.160
9	$O \times Sc$	0.111
10	$EL \times Sc$	0.442
11	S	0.675

Table B.4: Backward elimination for Perceptual processing strategy using GEE approach

Order	Explanatory variable	P-value
1	$S \times Q$	0.824
2	$EL \times O$	0.795
3	$S \times R$	0.779
4	$R \times O$	0.517
5	$EL \times S$	0.386
6	$EL \times R$	0.288
7	$Sc \times Q$	0.161
8	$EL \times Sc$	0.114
9	$O \times Sc$	0.051

Table B.5: Backward elimination for Inferencing strategy using GEE approach

Order	Explanatory variable	P-value
1	$EL \times R$	0.879
2	$R \times Q$	0.811
3	$EL \times S$	0.786
4	$S \times Q$	0.396
5	$S \times R$	0.359
6	$Sc \times Q$	0.307
7	$R \times Sc$	0.313
8	$R \times O$	0.193
9	$EL \times Sc$	0.201
10	$O \times Sc$	0.251
11	$S \times O$	0.330
12	$EL \times O$	0.062

Table B.6: Backward elimination for Prediction strategy using GEE approach

Order	Explanatory variable	P-value
1	$R \times Sc$	0.924
2	$EL \times S$	0.837
3	$Sc \times Q$	0.800
4	$S \times R$	0.711
5	$S \times Q$	0.585
6	$EL \times Q$	0.580
7	$EL \times O$	0.527
8	$EL \times R$	0.369
9	$S \times O$	0.294
10	$R \times Q$	0.173
11	S	0.144

Table B.7: Backward elimination for Utilisation strategy using GEE approach

Order	Explanatory variable	P-value
1	$EL \times O$	0.894
2	$EL \times S$	0.866
3	$EL \times R$	0.595
4	$S \times R$	0.562
5	$Sc \times Q$	0.519
6	$O \times Sc$	0.284
7	$R \times O$	0.534
8	$EL \times Sc$	0.439
9	$R \times Q$	0.217
10	$R \times Sc$	0.191
11	$EL \times Q$	0.140
12	$S \times O$	0.075
13	$S \times Q$	0.062
14	S	0.848

Table B.8: Backward elimination for Socio-affective strategy using GEE approach

Order	Explanatory variable	P-value
1	$S \times R$	0.473
2	$EL \times Sc$	0.389
3	$O \times Sc$	0.499
4	$S \times O$	0.235
5	$EL \times O$	0.288
6	$EL \times R$	0.207
7	$R \times Sc$	0.197
8	$R \times O$	0.134
9	$R \times Q$	0.100
10	$EL \times S$	0.062

C Standardised EL test result effect

This part illustrates tables that summarises the pairwise standardised EL test result effect comparisons amongst questions associated with those strategies that has significant differences in standardised EL test result effects.

Table C.1: Pairwise standardised EL test result effect comparisons ($\hat{\beta}_{Q_k}^{EL} - \hat{\beta}_{Q_{k'}}^{EL}$) for questions within the Perceptual Processing Strategy

$Q_{k'}$	Q_k			
	Q8	Q24	Q34	Q37
Q8	0	0.24** (0.00)	0.25** (0.00)	0.24** (0.00)
Q24		0	0.00 (0.98)	0.00 (1.00)
Q34			0	0.00 (0.98)
Q37				0

* Significant level at between 0.01 and 0.05

** Significant level at < 0.01

Table C.2: Pairwise standardised EL test result effect comparisons ($\hat{\beta}_{Q_k}^{EL} - \hat{\beta}_{Q_{k'}}^{EL}$) for questions within the Inferencing Strategy

$Q_{k'}$	Q_k						
	Q1	Q4	Q17	Q26	Q27	Q31	Q33
Q4		0	0.00 (0.99)	0.01 (0.86)	-0.08 (0.26)	-0.02 (0.78)	-0.03 (0.65)
Q17			0	0.01 (0.87)	-0.08 (0.28)	-0.02 (0.78)	-0.03 (0.64)
Q26				0	-0.09 (0.19)	-0.03 (0.66)	-0.04 (0.52)
Q27					0	0.06 (0.37)	0.05 (0.45)
Q31						0	-0.01 (0.88)
Q33							0

* Significant level at between 0.01 and 0.05

** Significant level at < 0.01

Table C.3: Pairwise standardised EL test result effect comparisons ($\hat{\beta}_{Q_k}^{EL} - \hat{\beta}_{Q_{k'}}^{EL}$) for questions within the Socio-affective Strategy

$Q_{k'}$	Q_k			
	Q2	Q10	Q25	Q38
Q2	0	-0.09 (0.15)	0.16* (0.01)	-0.09 (0.14) 0.22** (0.00)
Q10		0	0.25** (0.00)	0.00 (0.97) 0.32** (0.00)
Q25			0	-0.25** (0.00) 0.06 (0.34)
Q28				0 0.31** (0.00)
Q38				0

* Significant level at between 0.01 and 0.05

** Significant level at < 0.01

D Simulation

This part shows the table of the proportion of times the following null hypothesis are being rejected at a 5% level:

1. $H_0 : \beta_1 = 0$ against $H_1 : \beta_1 \neq 0$
2. $H_0 : \beta_2 = 0$ against $H_1 : \beta_2 \neq 0$
3. $H_0 : \beta_{12} = 0$ against $H_1 : \beta_{12} \neq 0$
4. $H_0 : \beta_{13} = 0$ against $H_1 : \beta_{13} \neq 0$

Table D.1: The proportion of times that the null hypothesis $H_0 : \beta_1 = 0$ is rejected at a 5% level

β_1, β_2	Proportional Odds Model			OLS Regression		
	c=3	c=5	c=7	c=3	c=5	c=7
(0.00, 0.25)	5.0%	5.5%	6.1%	5.0%	5.5%	5.6%
(0.00, 0.50)	5.0%	4.8%	5.2%	5.1%	4.7%	5.1%
(0.00, 0.75)	5.1%	5.3%	5.5%	5.2%	5.1%	5.4%
(0.00, 1.00)	5.3%	6.0%	5.4%	5.3%	5.9%	5.0%
(0.00, 1.25)	5.1%	5.1%	5.0%	5.5%	5.4%	4.8%
(0.00, 1.50)	5.6%	5.5%	6.0%	6.4%	5.5%	5.8%
(0.00, 1.75)	5.6%	5.3%	5.6%	6.4%	5.7%	5.7%
(0.00, 2.00)	4.9%	4.9%	5.6%	6.0%	5.3%	6.5%
(0.00, 2.25)	5.5%	5.4%	5.4%	3.3%	6.3%	6.3%
(0.00, 2.50)	5.4%	4.7%	5.0%	7.0%	5.7%	5.7%

Table D.2: The proportion of times that the null hypothesis $H_0 : \beta_2 = 0$ is rejected at a 5% level

β_1, β_2	Proportional Odds Model			OLS Regression		
	c=3	c=5	c=7	c=3	c=5	c=7
(0.25, 0.00)	4.8%	5.2%	6.0%	4.7%	4.9%	5.6%
(0.50, 0.00)	4.7%	5.6%	5.6%	4.7%	5.3%	5.1%
(0.75, 0.00)	5.1%	5.6%	5.7%	5.3%	5.2%	5.4%
(1.00, 0.00)	4.9%	5.3%	4.6%	4.7%	5.3%	4.4%
(1.25, 0.00)	5.0%	5.3%	5.0%	4.9%	5.3%	4.8%
(1.50, 0.00)	4.7%	5.3%	5.8%	4.9%	4.8%	5.4%
(1.75, 0.00)	4.9%	5.3%	5.6%	4.8%	5.4%	5.0%
(2.00, 0.00)	4.6%	5.4%	5.7%	4.6%	4.7%	4.9%
(2.25, 0.00)	4.9%	5.2%	5.1%	4.9%	5.1%	5.0%
(2.50, 0.00)	4.5%	4.7%	4.9%	4.5%	4.5%	4.7%

Table D.3: The proportion of times that the null hypothesis $H_0 : \beta_{12} = 0$ is rejected at a 5% level

β_1, β_2	Proportional Odds Model			OLS Regression		
	c=3	c=5	c=7	c=3	c=5	c=7
(0.25, 1.00)	5.5%	5.3%	5.9%	5.0%	4.5%	5.2%
(0.50, 1.00)	5.7%	4.9%	5.3%	5.1%	4.2%	4.3%
(0.75, 1.00)	5.4%	5.3%	5.3%	4.1%	3.7%	4.0%
(1.00, 1.00)	4.1%	5.3%	5.1%	2.8%	2.8%	2.9%
(0.25, 1.50)	5.4%	5.7%	5.9%	5.3%	5.4%	5.4%
(0.50, 1.50)	5.2%	5.5%	5.8%	6.7%	5.9%	5.8%
(0.75, 1.50)	5.0%	5.6%	5.8%	7.0%	6.8%	6.4%
(1.00, 1.50)	4.8%	4.9%	5.7%	6.7%	5.7%	5.5%
(0.25, 2.00)	5.1%	5.2%	5.1%	6.5%	6.1%	5.9%
(0.50, 2.00)	4.9%	5.7%	5.7%	10.8%	9.8%	9.5%
(0.75, 2.00)	4.8%	5.0%	5.2%	16.7%	12.8%	12.3%
(1.00, 2.00)	5.1%	5.1%	5.1%	18.1%	15.1%	14.2%
(0.25, 2.50)	5.0%	5.6%	5.0%	8.7%	7.9%	7.3%
(0.50, 2.50)	4.5%	5.2%	4.8%	18.8%	15.8%	14.8%
(0.75, 2.50)	4.0%	4.9%	5.2%	30.1%	26.6%	23.1%
(1.00, 2.50)	4.7%	4.8%	5.6%	37.9%	34.7%	30.1%

Table D.4: The proportion of times that the null hypothesis $H_0 : \beta_{13} = 0$ is rejected at a 5% level

β_1, β_3	Proportional Odds Model			OLS Regression		
	c=3	c=5	c=7	c=3	c=5	c=7
(0.25, 0.25)	5.5%	5.1%	5.5%	4.7%	4.4%	4.6%
(0.25, 0.50)	5.5%	5.5%	5.6%	4.2%	4.5%	4.6%
(0.25, 0.75)	4.6%	5.2%	6.0%	3.7%	4.0%	4.0%
(0.25, 1.00)	5.7%	5.1%	5.1%	4.0%	3.8%	3.2%
(0.25, 1.50)	4.6%	5.7%	4.9%	4.6%	4.4%	4.1%
(0.25, 2.00)	4.6%	5.2%	5.6%	6.9%	7.1%	6.7%
(0.25, 2.50)	5.0%	4.7%	5.1%	9.4%	10.5%	12.1%
(0.50, 0.25)	5.3%	5.5%	5.4%	4.1%	3.8%	4.1%
(0.50, 0.50)	5.0%	5.4%	5.6%	3.5%	3.9%	4.0%
(0.50, 0.75)	4.5%	4.8%	5.5%	3.1%	3.4%	3.8%
(0.50, 1.00)	4.7%	5.7%	4.8%	3.3%	3.5%	3.2%
(0.50, 1.50)	4.6%	5.2%	5.5%	4.3%	4.7%	3.9%
(0.50, 2.00)	4.9%	5.1%	4.8%	7.2%	7.7%	7.3%
(0.50, 2.50)	4.9%	5.0%	5.3%	11.8%	12.6%	13.9%
(0.75, 0.25)	5.8%	5.2%	5.6%	3.7%	3.5%	3.5%
(0.75, 0.50)	5.3%	5.0%	5.2%	3.3%	3.0%	2.9%
(0.75, 0.75)	5.1%	5.1%	5.7%	2.9%	2.9%	3.1%
(0.75, 1.00)	5.5%	5.4%	5.9%	2.8%	3.0%	3.5%
(0.75, 1.50)	4.7%	5.2%	5.5%	3.7%	4.2%	3.9%
(0.75, 2.00)	4.8%	5.4%	5.3%	7.7%	7.3%	7.8%
(0.75, 2.50)	4.8%	4.8%	5.1%	11.7%	13.3%	13.5%
(1.00, 0.25)	5.1%	4.9%	5.2%	2.6%	2.6%	2.7%
(1.00, 0.50)	5.1%	5.0%	5.3%	2.6%	2.4%	2.7%
(1.00, 0.75)	5.4%	5.2%	5.7%	2.4%	2.3%	4.0%
(1.00, 1.00)	4.9%	5.1%	5.8%	2.3%	2.3%	2.7%
(1.00, 1.50)	4.5%	5.5%	5.8%	3.4%	3.3%	3.4%
(1.00, 2.00)	4.9%	4.8%	5.7%	6.9%	6.3%	6.2%
(1.00, 2.50)	5.1%	5.5%	5.4%	11.6%	12.1%	12.0%

E R code

This part shows R code associated with this thesis. A line that starts with `#` refers to comments, which R ignores.

E.1 Proportional odds model using GEE approach

This part shows how data were coded for fitting the proportional odds model by using GEE approach for the Perceptual processing strategy. It also shows how line chart for the Inferencing strategy was produced in Chapter 2.

```
# Required packages
library(MASS)
library(VGAM)
library(repolr)
library(multgee)
library(reshape)
library(geepack)
```



```

library(plyr)
library(ggplot2)

## ----- Data preparation ----- ##
## The following shows the data preparation process so that
## the data is appropriate to fit the proportional odds model
## by using the GEE method.

## Converts score response variable from a numeric
## to categorical variable
for (i in 7:44) {
temp_data1[,i] = factor(temp_data1[,i], ordered = T)}

## Select grade 6 students only
temp_data2 = subset(temp_data1, GradeNAlevel == 6)

## Remove the observations with missing values
temp_data2 = na.omit(temp_data2)

## Standardise score response variable by school variable
Std.EL = ddply(temp_data2, .(School), summarize,
IC=IC, Std.Y = scale(Results))
temp_data3 = merge(temp_data2, Std.EL, by = "IC")

## Transpose score response variable by

```

```

temp_data4 = melt(temp_data3, id.vars = c("IC", "Sex",
"Race", "Order", "School.x", "Std.Y"),
measure.vars = c("L1", "L2", "L3", "L4", "L5", "L6", "L7",
"L8", "L9", "L10", "L11", "L12", "L13", "L14", "L15", "L16",
"L17", "L18", "L19", "L20", "L21", "L22", "L23", "L24",
"L25", "L26", "L27", "L28", "L29", "L30", "L31", "L32",
"L33", "L34", "L35", "L36", "L37", "L38"))

## Sort temp_data4 by IC (unique obs id)
data = with(temp_data4, temp_data4[order(IC),])

## Rename the variables
names(data) = c("ID", "Sex", "Eth", "Order", "School",
"EL_Test_Result", "Q", "Response")

## Convert variable Q to a categorical variable
data$Q = factor(data$Q)

## Aggregate non-Chinese students to "Non-Chinese"
data$Eth = ifelse(data$Eth != "Chinese",
"Non-Chinese", "Chiense")

## ----- Line graph ----- ##
## The following shows the code for creating a line plot.
## For this part, Inference strategy will be used as

```

```

## an example

## Select obs where questions are associated with
## Inferencing strategy
inference = subset(data, Q == "L1" | Q == "L4" |
Q == "L17" | Q == "L26" | Q == "L27" |
Q == "L31" | Q == "L33")
inference$Question = ifelse((inference$Q == "L1"), 1,
ifelse((inference$Q == "L4"), 2,
ifelse((inference$Q == "L17"), 3,
ifelse((inference$Q == "L26"), 4,
ifelse((inference$Q == "L27"), 5,
ifelse((inference$Q == "L31"), 6, 7))))))
inference$Q = factor(inference$Q)

## Calculate median standardised EL test result
inference_med = aggregate(EL_Test_Result ~ Q + Response, inference, median)
inference_med = data.frame(reshape(inference_med, timevar="Q",
idvar="Response", direction="wide"))

## Plot median standardised EL test result for each questions at:
## 1 - Never
## 2 - Rarely
## 3 - Sometimes
## 4 - Often

```

```

## 5 - Always
plot(inference_med[,2], type="b", ylim=c(-1,1), col=1, lty=1,
ylab="Std EL Result", lwd=2, xlab="Score", pch=1, xaxt="n",
cex.axis=1.5, cex.lab=1.5, cex=2.5)
lines(inference_med[,3], type="b", col=2, pch=2, lty=2, lwd=2, cex=2.5)
lines(inference_med[,4], type="b", col=3, pch=3, lty=3, lwd=2, cex=2.5)
lines(inference_med[,5], type="b", col=4, pch=4, lty=4, lwd=2, cex=2.5)
lines(inference_med[,6], type="b", col=6, pch=5, lty=5, lwd=2, cex=2.5)
lines(inference_med[,7], type="b", col=7, pch=6, lty=6, lwd=2, cex=2.5)
lines(inference_med[,8], type="b", col=8, pch=22, lty=7, lwd=2, cex=2.5)
legend("bottomright", legend=c("Q1", "Q4", "Q17", "Q26", "Q27", "Q31", "Q33"),
col=c(1,2,3,4,6,7,8), lty=c(1,2,3,4,5,6,7), pch=c(1,2,3,4,5,6,22),
bg="white",lwd=2,cex=1.2)
axis(1, at=c(1:nrow(inference_med)), cex.axis=1.5)

## ----- Functions ----- ##
## Wald test statistics for an explanatory variable with more than 2 categories
wald = function(Y, beta0, beta1){
H = diag(length(Y$coefficients))[c(beta0:beta1),]
Beta = as.matrix(Y$coefficients[1:length(Y$coefficients)])
HBeta = H%*%Beta
cov = as.matrix(Y$robust.variance)
covHBeta = H%*%cov%*%t(H)
W = t(H%*%Beta)%*%solve(covHBeta)%*%HBeta
Pvalue = 1 - pchisq(W, beta1-beta0+1)

```

```

        cat("P-value")
print(Pvalue)
}

## Wald test statistics for pairwise
wald_pair= function(Y, beta0, beta1){
H = matrix(nrow=1, ncol=length(Y$coefficients))
H[is.na(H)] = 0
H[,beta0] = 1
H[,beta1] = -1
Beta = as.matrix(Y$coefficients[1:length(Y$coefficients)])
HBeta = H%%Beta
cov = as.matrix(Y$robust.variance)
covHBeta = H%%cov%%t(H)
W = t(H%%Beta)%%solve(covHBeta)%%HBeta
Pvalue = 1 - pchisq(W,1)
        cat("P-value")
print(Pvalue)
}

## Proportional odds model with the GEE method
## for Perceptual processing strategy

## Selecting obs where questions are associated with
## Perceptual processing strategy

```

```

perceptual = subset(data, Q == "L8" | Q == "L24" |
Q == "L34" | Q== "L37")
perceptual$Q = ifelse((perceptual$Q == "L8"), 1,
ifelse((perceptual$Q == "L24"), 2,
ifelse((perceptual$Q == "L34"), 3, 4)))
perceptual$Q = factor(perceptual$Q)

## ----- perceptual.fit ----- ###
perceptual.fit = ordLORgee(Response ~ EL_Test_Result + Sex + Eth +
Order + School + Q + EL_Test_Result*Sex + EL_Test_Result*Eth +
EL_Test_Result*Order + EL_Test_Result*School + EL_Test_Result*Q +
Sex*Eth + Sex*Order + Sex*Q + Eth*Order + Eth*School + Eth*Q +
Order*School + Order*Q + School*Q, id = perceptual$ID,
repeated = perceptual$Q, data = perceptual)

summary(perceptual.fit)

## Model selection by using backward elimination
## The following shows the list of explanatory variables from
## perceptual.fit along with their corresponding p-values.

# EL_Test_Result*Sex (p-value = 0.373)
# EL_Test_Result*Eth (p-value = 0.316)
# EL_Test_Result*Order (p-value = 0.795)
# EL_Test_Result*School (p-value = 0.125)

```

```

wald(perceptual.fit, 20, 24)
# EL_Test_Result*Q (p-value = 0.000)
wald(perceptual.fit, 25, 27)
# Sex*Eth (p-value = 0.796 )
# Sex*Order (p-value = 0.005)
# Sex*Q (p-value = 0.824)
wald(perceptual.fit, 30, 32)
# Eth*Order (p-value = 0.539)
# Eth*Q (p-value = 0.008)
wald(perceptual.fit, 39, 41)
# Order*School (p-value = 0.008)
wald(perceptual.fit, 42, 46)
# Order*Q (p-value = 0.022)
wald(perceptual.fit, 47, 49)
# School*Q (p-value = 0.167)
wald(perceptual.fit, 50, 64)

## Sex*Q was removed with the highest p-value of 0.824
## from perceptual.fit

## ----- perceptual.fit2 ----- ###
perceptual.fit2 = ordLORgee(Response ~ EL_Test_Result + Sex +
Eth + Order + School + Q + EL_Test_Result*Sex +
EL_Test_Result*Eth + EL_Test_Result*Order +
EL_Test_Result*School + EL_Test_Result*Q +

```

```

Sex*Eth + Sex*Order + Eth*Order + Eth*School + Eth*Q +
Order*School + Order*Q + School*Q, id = perceptual$ID,
repeated = perceptual$Q, data = perceptual)

summary(perceptual.fit2)

## Model selection by backward elimination for perceptual.fit2
## EL_Test_Result*Sex (p-value = 0.373)
## EL_Test_Result*Eth (p-value = 0.318)
## EL_Test_Result*Order (p-value = 0.795)
## EL_Test_Result*School (p-value = 0.126)
wald(perceptual.fit2, 20, 24)
## EL_Test_Result*Q (p-value = 0.000)
wald(perceptual.fit2, 25, 27)
## Sex*Eth (p-value = 0.784)
## Sex*Order (p-value = 0.005)
## Eth*Order (p-value = 0.534)
## Eth*School (p-value = 0.012)
wald(perceptual.fit2, 31, 35)
## Eth*Q (p-value = 0.008)
wald(perceptual.fit2, 36, 38)
## Order*School (p-value = 0.008)
wald(perceptual.fit2, 39, 43)
## Order*Q (p-value = 0.022)
wald(perceptual.fit2, 44, 46)

```



```

## School*Q (p-value = 0.159)
wald(perceptual.fit2, 47, 61)

## EL_Test_Result*Order was removed with the highest
## p-value of 0.795 from perceptual.fit2.

## ----- perceptual.fit3 ----- ###
perceptual.fit3 = ordLORgee(Response ~ EL_Test_Result + Sex +
Eth + Order + School + Q + EL_Test_Result*Sex +
EL_Test_Result*Eth + EL_Test_Result*School +
EL_Test_Result*Q + Sex*Eth + Sex*Order + Eth*Order +
Eth*School + Eth*Q + Order*School + Order*Q + School*Q,
id = perceptual$ID, repeated = perceptual$Q, data = perceptual)

summary(perceptual.fit3)

## Model selection by backward elimination for perceptual.fit3
## EL_Test_Result*Sex (p-value = 0.371)
## EL_Test_Result*Eth (p-value = 0.317)
## EL_Test_Result*School (p-value = 0.106)
wald(perceptual.fit3, 19, 23)
## EL_Test_Result*Q (p-value = 0.000)
wald(perceptual.fit3, 24, 26)
## Sex*Eth (p-value = 0.779)
## Sex*Order (p-value = 0.005)

```

```

## Eth*Order (p-value = 0.524)
## Eth*School (p-value = 0.013)
wald(perceptual.fit3, 30, 34)
## Eth*Q (p-value = 0.008)
wald(perceptual.fit3, 35, 37)
## Order*School (p-value = 0.009)
wald(perceptual.fit3, 38, 42)
## Order*Q (p-value = 0.022)
wald(perceptual.fit3, 43, 45)
## School*Q (p-value = 0.159)
wald(perceptual.fit3, 46, 60)

## Sex*Eth was removed with the highest
## p-value of 0.779 from perceptual.fit3.

## ----- perceptual.fit4 ----- ###
perceptual.fit4 = ordLORgee(Response ~ EL_Test_Result + Sex +
Eth + Order + School + Q + EL_Test_Result*Sex +
EL_Test_Result*Eth + EL_Test_Result*School +
EL_Test_Result*Q + Sex*Order + Eth*Order +
Eth*School + Eth*Q + Order*School + Order*Q + School*Q,
id = perceptual$ID, repeated = perceptual$Q, data = perceptual)

summary(perceptual.fit4)

```

```

## Model selection by backward elimination for perceptual.fit4
## EL_Test_Result*Sex (p-value = 0.382)
## EL_Test_Result*Eth (p-value = 0.300)
## EL_Test_Result*School (p-value = 0.106)
wald(perceptual.fit4, 19, 23)
## EL_Test_Result*Q (p-value = 0.000)
wald(perceptual.fit4, 24, 26)
## Sex*Order (p-value = 0.005)
## Eth*Order (p-value = 0.517)
## Eth*School (p-value = 0.013)
wald(perceptual.fit4, 29, 33)
## Eth*Q (p-value = 0.008)
wald(perceptual.fit4, 34, 36)
## Order*School (p-value = 0.008)
wald(perceptual.fit4, 37, 41)
## Order*Q (p-value = 0.022)
wald(perceptual.fit4, 42, 44)
## School*Q (p-value = 0.159)
wald(perceptual.fit4, 45, 59)

## Eth*Order was removed with the highest
## p-value of 0.517 from perceptual.fit4.

## ----- perceptual.fit5 ----- ###
perceptual.fit5 = ordLORgee(Response ~ EL_Test_Result + Sex +

```

```

Eth + Order + School + Q + EL_Test_Result*Sex +
EL_Test_Result*Eth + EL_Test_Result*School +
EL_Test_Result*Q + Sex*Order + Eth*School + Eth*Q +
Order*School + Order*Q + School*Q, id = perceptual$ID,
repeated = perceptual$Q, data = perceptual)

summary(perceptual.fit5)

## Model selection by backward elimination for perceptual.fit5
## EL_Test_Result*Sex (p-value = 0.386)
## EL_Test_Result*Eth (p-value = 0.272)
## EL_Test_Result*School (p-value = 0.117)
wald(perceptual.fit5, 19, 23)
## EL_Test_Result*Q (p-value = 0.000)
wald(perceptual.fit5, 24, 26)
## Sex*Order (p-value = 0.005)
## Eth*School (p-value = 0.009)
wald(perceptual.fit5, 28, 32)
## Eth*Q (p-value = 0.008)
wald(perceptual.fit5, 33, 35)
## Order*School (p-value = 0.010)
wald(perceptual.fit5, 36, 40)
## Order*Q (p-value = 0.022)
wald(perceptual.fit5, 41, 43)
## School*Q (p-value = 0.159)

```

```

wald(perceptual.fit5, 44, 58)

## EL_Test_Result*Sex was removed with the highest
## p-value of 0.386 from perceptual.fit5.

## ----- perceptual.fit6 ----- ###
perceptual.fit6 = ordLORgee(Response ~ EL_Test_Result + Sex +
Eth + Order + School + Q + EL_Test_Result*Eth +
EL_Test_Result*School + EL_Test_Result*Q + Sex*Order +
Eth*School + Eth*Q + Order*School + Order*Q + School*Q,
id = perceptual$ID, repeated = perceptual$Q, data = perceptual)

summary(perceptual.fit6)

## Model selection by backward elimination for perceptual.fit6
## EL_Test_Result*Eth (p-value = 0.288)
## EL_Test_Result*School (p-value = 0.106)
wald(perceptual.fit6, 18, 22)
## EL_Test_Result*Q (p-value = 0.000)
wald(perceptual.fit6, 23, 25)
## Sex*Order (p-value = 0.005)
## Eth*School (p-value = 0.009)
wald(perceptual.fit6, 27, 31)
## Eth*Q (p-value = 0.008)
wald(perceptual.fit6, 32, 34)

```

```

## Order*School (p-value = 0.011)
wald(perceptual.fit6, 35, 39)
## Order*Q (p-value = 0.022)
wald(perceptual.fit6, 40, 42)
## School*Q (p-value = 0.160)
wald(perceptual.fit6, 43, 57)

## EL_Test_Result*Eth was removed with the highest
## p-value of 0.288 from perceptual.fit6.

## ----- perceptual.fit7 ----- ###
perceptual.fit7 = ordLORgee(Response ~ EL_Test_Result + Sex +
Eth + Order + School + Q + EL_Test_Result*School +
EL_Test_Result*Q + Sex*Order + Eth*School + Eth*Q +
Order*School + Order*Q + School*Q, id = perceptual$ID,
repeated = perceptual$Q, data = perceptual)

summary(perceptual.fit7)

## Model selection by backward elimination for perceptual.fit7
## EL_Test_Result*School (p-value = 0.117)
wald(perceptual.fit7, 17, 21)
## EL_Test_Result*Q (p-value = 0.000)
wald(perceptual.fit7, 22, 24)
## Sex*Order (p-value = 0.005)

```

```

## Eth*School (p-value = 0.012)
wald(perceptual.fit7, 26, 30)
## Eth*Q (p-value = 0.008)
wald(perceptual.fit7, 31, 33)
## Order*School (p-value = 0.013)
wald(perceptual.fit7, 34, 38)
## Order*Q (p-value = 0.023)
wald(perceptual.fit7, 39, 41)
## School*Q (p-value = 0.161)
wald(perceptual.fit7, 42, 56)

## School*Q was removed with the highest
## p-value of 0.161 from perceptual.fit7.

## ----- perceptual.fit8 ----- ###
perceptual.fit8 = ordLORgee(Response ~ EL_Test_Result + Sex +
Eth + Order + School + Q + EL_Test_Result*School +
EL_Test_Result*Q + Sex*Order + Eth*School + Eth*Q +
Order*School + Order*Q, id = perceptual$ID,
repeated = perceptual$Q, data = perceptual)

summary(perceptual.fit8)

## Model selection by backward elimination for perceptual.fit8
## EL_Test_Result*School (p-value = 0.114)

```

```

wald(perceptual.fit8, 17, 21)
## EL_Test_Result*Q (p-value = 0.000)
wald(perceptual.fit8, 22, 24)
## Sex*Order (p-value = 0.005)
## Eth*School (p-value = 0.010)
wald(perceptual.fit8, 26, 30)
## Eth*Q (p-value = 0.021)
wald(perceptual.fit8, 31, 33)
## Order*School (p-value = 0.013)
wald(perceptual.fit8, 34, 38)
## Order*Q (p-value = 0.015)
wald(perceptual.fit8, 39, 41)

## EL_Test_Result*School was removed with the
## highest p-value of 0.114 from perceptual.fit8.

## ----- perceptual.fit9 ----- ###
perceptual.fit9 = ordLORgee(Response ~ EL_Test_Result + Sex +
Eth + Order + School + Q + EL_Test_Result*Q + Sex*Order +
Eth*School + Eth*Q + Order*School + Order*Q,
id = perceptual$ID, repeated = perceptual$Q, data = perceptual)

summary(perceptual.fit9)

## Model selection by backward elimination for perceptual.fit9

```



```

## EL_Test_Result*Q (p-value = 0.000)
wald(perceptual.fit9, 17, 19)
## Sex*Order (p-value = 0.005)
## Eth*School (p-value = 0.009)
wald(perceptual.fit9, 21, 25)
## Eth*Q (p-value = 0.022)
wald(perceptual.fit9, 26, 28)
## Order*School (p-value = 0.051)
wald(perceptual.fit9, 29, 33)
## Order*Q (p-value = 0.015)
wald(perceptual.fit9, 34, 36)

## Order*School was removed with the highest
## p-value of 0.051 from perceptual.fit8.

## ----- perceptual.fit10 ----- ###
perceptual.fit10 = ordLORgee(Response ~ EL_Test_Result + Sex +
Eth + Order + School + Q + EL_Test_Result*Q + Sex*Order +
Eth*School + Eth*Q + Order*Q, id = perceptual$ID,
repeated = perceptual$Q, data = perceptual)

summary(perceptual.fit10)

## Model selection by backward elimination for perceptual.fit10
## EL_Test_Result*Q (p-value = 0.000)

```

```

wald(perceptual.fit10, 17, 19)
## Sex*Order (p-value = 0.010)
## Eth*School (p-value = 0.007)
wald(perceptual.fit10, 21, 25)
## Eth*Q (p-value = 0.023)
wald(perceptual.fit10, 26, 28)
## Order*Q (p-value = 0.016)
wald(perceptual.fit10, 29, 31)

## perceptual.fit10 is the final proportional odds model
## using GEE method for Perceptual processing strategy.

## Pairwise in the effect of standardised EL test result
## for the questions associated with the Perceptual
## processing strategy

## Q24 vs Q34
wald_pair(perceptual.fit10, 17, 18)
## P-value = 0.983
## Q24 vs Q37
wald_pair(perceptual.fit10, 17, 19)
## P-value = 0.998
## Q34 vs Q37
wald_pair(perceptual.fit10, 17, 19)
## P-value = 0.981

```

E.2 Simulation - R code

This section reflects R code for the proportion of times the null hypothesis, $H_0 : \beta_{12} = 0$ is rejected at a 5% level, Section 5.4, the third scenario. This particular example has the following parameters: $c = 7$, $\beta_1 = 1.00$ and $\beta_2 = 2.50$.

```
library(VGAM)

x1 = rnorm(n=100, m=0, sd=1)
x2 = runif(100)<=0.5
x2 = ifelse(x2 == "TRUE", 1, 0)
data = data.frame(x1,x2)

c = 7
alpha = c(log(1/6), log(2/5), log(3/4), log(4/3), log(5/2), log(6))
# when c = 3, then alpha = c(log(1/2), log(2))
# when c = 5, then alpha = c(log(1/4), log(2/3), log(3/2), log(4))

beta1 = 1.00
beta2 = 2.50
rpt = 5000
```

```

ptm = proc.time()
sim = (replicate(rpt, {

y = function(x1, x2){
p = array(NA, dim=c)
p[1] = exp(alpha[1] - beta1*x1 - beta2*x2) /
(1 + exp(alpha[1] - beta1*x1 - beta2*x2))
for(i in 2:(c-1)){
p[i] = (exp(alpha[i] - beta1*x1 - beta2*x2) /
(1 + exp(alpha[i] - beta1*x1 - beta2*x2))) -
(exp(alpha[i-1] - beta1*x1 - beta2*x2) /
(1 + exp(alpha[i-1] - beta1*x1 - beta2*x2)))
}
p[c] = 1 - (exp(alpha[c-1] - beta1*x1 - beta2*x2) /
(1 + exp(alpha[c-1] - beta1*x1 - beta2*x2)))

y_temp1 = c(1:c) * rmultinom(n=1, size=1, prob=p)
y = y_temp1[y_temp1!=0]
y
}

data$y = mapply(y, data$x1, data$x2)

```

```

ols = lm(y ~ x1 + x2 + x1*x2, data=data)
summary(ols)
p_ols = coef(summary(ols))[16]
p_ols

data$y_pom = factor(data$y, ordered = T)
pom = vglm(y_pom ~ x1 + x2 + x1*x2, family = cumulative(
parallel = TRUE, reverse = TRUE), data=data)
ctable_pom = coef(summary(pom))
p = pnorm(abs(ctable_pom[, "z value"]), lower.tail = FALSE) * 2
p_pom = p[c+2]
p_pom

rej_ols = ifelse(p_ols <= 0.05, 1, 0)
rej_pom = ifelse(p_pom <= 0.05, 1, 0)

return(c(rej_ols, rej_pom))

}))
proc.time() - ptm

ols = round(sum(na.omit(sim[1,]))/rpt, digits=3)
ols
pom = round(sum(na.omit(sim[2,]))/rpt, digits=3)
pom

```

References

- Agresti, A. (2010). *Analysis of ordinal categorical data*. John Wiley & Sons, 2nd edition.
- Agresti, A. (2013). *Categorical Data Analysis*. John Wiley & Sons, 3rd edition.
- Anderson, J. and Philips, P. (1981). Regression, discrimination and measurement models for ordered categorical variables. *Applied statistics*, pages 22–31.
- Dobbie, S. (2015). Cluster verification for listening learner strategies. Victoria University of Wellington, <http://hdl.handle.net/10063/4725>.
- Goodman, L. A. (1971). The analysis of multidimensional contingency tables: Stepwise procedures and direct estimation methods for building models for multiple classifications. *Technometrics*, 13(1):33–61.
- Gu, P. Y., Hu, G., and Zhang, L. J. (2005). Investigating language learner

- strategies among lower primary school pupils in singapore. *Language and Education*, 19(4):281–303.
- Hastie, T., Botha, J., and Schnitzler, C. (1989). Regression with an ordered categorical response. *Statistics in medicine*, 8(7):785–794.
- Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13–22.
- Lipsitz, S. R., Kim, K., and Zhao, L. (1994). Analysis of repeated categorical data using generalized estimating equations. *Statistics in medicine*, 13(11):1149–1163.
- Little, R. J. (1988). Missing-data adjustments in large surveys. *Journal of Business & Economic Statistics*, 6(3):287–296.
- Little, R. J. and Rubin, D. B. (2002). *Statistical analysis with missing data*. John Wiley & Sons.
- McCullagh, P. (1980). Regression models for ordinal data. *Journal of the royal statistical society. Series B (Methodological)*, pages 109–142.
- McKelvey, R. D. and Zavoina, W. (1975). A statistical model for the analysis of ordinal level dependent variables. *Journal of mathematical sociology*, 4(1):103–120.
- O’Malley, J. M. and Chamot, A. U. (1990). *Learning strategies in second language acquisition*. Cambridge University Press.
- Thompson, R. and Baker, R. (1981). Composite link functions in generalized linear models. *Applied Statistics*, pages 125–131.

- Touloumis, A. (2014). R package multgee: A generalized estimating equations solver for multinomial responses. *arXiv preprint arXiv:1410.5232*.
- Ungureanu, C. and Georgescu, C. A. (2012). Learners strategies in language learning. *Procedia-Social and Behavioral Sciences*, 46:5000–5004.
- Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical society*, 54(3):426–482.
- Wedderburn, R. W. (1974). Quasi-likelihood functions, generalized linear models, and the gaussnewton method. *Biometrika*, 61(3):439–447.
- Wickham and Hadley (2007). Reshaping data with the reshape package. *Journal of Statistical Software*, 21(12).
- Yee, T. W. et al. (2010). The vgam package for categorical data analysis. *Journal of Statistical Software*, 32(10):1–34.