Universiteit Leiden

Faculteit der Sociale Wetenschappen

# Generalized IRTree Models of Children's Analogical Reasoning Processes

Chong Zhang

Master Thesis Psychology, Methodology and Statistics Unit, Institute of Psychology
Faculty of Social and Behavioral Sciences – Leiden University
Date: April 2016
Student number: 1599577
Supervisor: Dr. Claire Stevenson & Dr. Minjeong Jeon
Second readers: Dr. Tom F. Wilderjans & Prof. dr. Henk Kelderman

## Abstract

**Introduction** The traditional item response theory (IRT) models have been often applied to analyze psychological and behavioral data. In the present study, a class of more flexible models called “the generalized IRTree models” was used to gain insights into the analogical reasoning process of children. Two research questions were addressed. (1) Which model is the best fit for the children’s analogical reasoning strategy dataset? (2) Which model is the best fit for the dataset including age and working memory capacity?

**Method** The dataset included analogical reasoning strategy responses of 1002 children. The response variable was classified into four categories (correct, partial correct, duplicate and other). Age and working memory capacity were used as person predictor variables. Four IRTree models with different tree structures have been conducted for both the original ordered response variable and adjusted ordered response variable.

**Results** The IRTree model with binary tree structures was the most appropriate model for the children’s analogical reasoning strategy, regardless of orders between “Other” and “Duplicate”. When including the age and working memory capacity, the IRTree Model with binary tree structure and “Other” as the lowest ordered category was the best fit among the four IRTree models.

**Discussion** The results of the IRTree models illustrated the analogical reasoning process of children followed a binary structure with three stages. Age and working memory capacity had influence on different stages of children’s strategy use during the analogical reasoning process.

**Acknowledgements**

This master thesis is the final project of the master program "Methodology and Statistics in Psychology". After finishing the research master program "Clinical, Psychosocial, Epidemiology" in the University of Groningen, my enthusiasm and interest in statistics had grown. This master program, as my second master's degree in the Netherlands, was valuable for my employment opportunities in the near future.

During this master thesis project, I was lucky to have the helpful guidance from two supervisors. First of all, I would like to thank dr. Claire Stevenson for providing me with interesting and feasible research ideas, and the dataset for the present thesis project. In addition, the rich and clear comments on the research proposal and previous versions of thesis drafts were helpful.

Second, I would like to thank dr. Minjeong Jeon for providing the supervision for the application of "FLIRT" package. The informative suggestions helped me solve statistical questions. I appreciate your quick responses via emails regardless of time difference.

Finally, I would like to thank my best friend Yixia for inspiring and supporting me during the master thesis process. I am also thankful for my parents' understanding and financial support for me in completing this master's program.

**Table of Contents**

Abstract ........ 1
1 Introduction ........ 5
1.1 Analogical reasoning ........ 5
1.2 Analogical reasoning process models ........ 5
1.3 Analogical reasoning process of children ........ 7
1.4 Strategies for solving analogical reasoning tasks ........ 7
1.5 Traditional IRT models for analogical reasoning process ........ 8
1.6 IRTree models to understand cognitive processes ........ 9
1.7 Relations between analogical reasoning theories and IRTree models ........ 11
1.8 Research questions ........ 12
2 Method ........ 13
2.1 Sample ........ 13
2.2 Design and procedure ........ 13
2.3 Material ........ 13
2.4 Variables ........ 13
2.5 Properties of the dataset ........ 14
2.6 Explanatory IRT ........ 15
2.6.1 Fitting the IRTree models ........ 15
2.6.2 Traditional IRT models ........ 18
2.6.3 Software ........ 20
2.7 Model selection ........ 21
3 Results ........ 22
3.1 Descriptive statistics ........ 22
3.2 Classical test theory (CTT) results ........ 22
3.3 What is the better order among categories of response variable? ........ 23
3.3.1 Graded Response Model for the original ordered response variable ........ 23
3.3.2 Graded Response Model for the adjusted ordered response variable ........ 25
3.3.3 Generalized Partial Credit Model for the original ordered response variable ........ 26
3.3.4 Generalized Partial Credit Model for the adjusted ordered response variable ........ 27
3.3.5 Model selection ........ 29
3.4 Research question 1, "Which model is the best fit for the dataset of children's analogical reasoning strategy?" ........ 30
3.4.1 IRTree models ........ 30
3.4.2 Model selection ........ 31
3.4.3 Interpretation of the best fit model ........ 31

3.5 Research question 2, "Which model is the best fit for the dataset including age and working memory capacity?" ........ 32
3.5.1 IRTree Models ........ 32
3.5.2 Model selection ........ 34
3.5.3 Interpretation of the best fit model ........ 34
4 Discussion ........ 35
4.1 Effect of age and working memory capacity ........ 35
4.2 Advantages ........ 36
4.3 Limitations ........ 36
4.4 Methodological considerations ........ 37
4.5 Recommendations for future research ........ 38
5 Conclusions ........ 40
References ........ 41
Appendix ........ 44

# 1 Introduction

## 1.1 Analogical reasoning

Analogical reasoning refers to the human ability to learn about a new situation, by relating to a familiar one with similar structure (Goswami & Brown, 1991). One example of analogical reasoning is that children can recognize the relations between a red bear and a blue bear, after they have shown a red dog and a blue dog. Analogical reasoning has been widely considered as the hallmark of human intelligence (Gentner, 1983). It used to represents formal operational thinking in cognition development (Piaget, 1977). Nowadays, researchers reach a consensus that analogy is available by the pre-operational period (Goswami, Leevers., Pressley, & Wheelwright, 1998). Researchers have developed theories and models to explain the process of analogical reasoning since 1970s. According to different perspectives and materials for testing, several types of analogical reasoning tasks have been developed, such as geometric analogies (Tunteler, Pronk, & Resing, 2008) and verbal analogies (Goswami & Brown, 1990; Whitely & Barnes, 1979). Recent studies mainly focused on children’s cognitive process and performance of figural matrices analogy tasks (Siegler & Svetina, 2002; Stevenson, Alberto, van den Boom, & De Boeck, 2014).

## 1.2 Analogical reasoning process models

Researchers have constructed various models to explain the analogical reasoning process among children and adults. The most well-known analogical reasoning process models are Sternberg’s (1977) componential theory, and Mulholland’s (1980) two-stage figural analogical reasoning process model (Mulholland, Pellegrino, & Glaser, 1980; Sternberg, 1977; Sternberg & Rifkin, 1979).

Sternberg and colleagues presented a componential theory of the analogical reasoning process based on people’s reactions times when solving analogies which involved six components: encoding, inference, mapping, application, justification and response (Sternberg, 1977; Sternberg & Rifkin, 1979). (1) The encoding indicated the process of translating the analogy information into an internal representation. In the same example in the first paragraph, children encode features of colours and animals in the first two blocks, (2) then infer the relation between red dog and blue dog, (3) maps the relation between bear and dog, (4) apply the relation analogous of red bear and blue bear to the inferred one, (5) justify the choice, (6) and finally give the response as a blue bear. Among these components, mapping and justification were optional processes, and others were mandatory. Four procedural models were formulated based on the six components. These models were different from each other according to two operations, exhausting and self-terminating. Exhausting operation means people compared between all attributed values for stimuli in analogical reasoning process. Self-terminating operation means people compared among a limited subset of possible relations. For example, younger children were more likely to use self-terminating operation instead of exhausting operation in analogical reasoning process, compared with older children and adults.

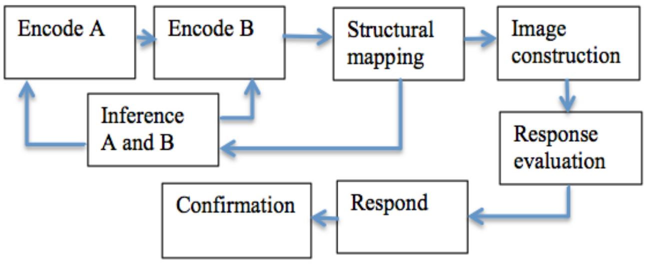



*Figure 1*. Processing model for analogical reasoning process by Embretson et al. (1989)

Embretson and colleague (1989) confirmed and extended Sternberg's componential theory. They examined the role of interactive processing on psychometrics analogies, especially on verbal analogies (Embretson & Schneider, 1989). It was found that mapping process could be replaced with structural mapping. Structural mapping was defined as an evaluation for common attributes relationships between base domain and target domain. In addition, inferences were contextualized. It was necessary to assess inference difficulty in analogical reasoning process. Furthermore, the application was separated as two components, which were image construction and response evaluation. Confirmation was added at the end of analogical reasoning process as a new component (Whitely & Barnes, 1979).

Mulholland et al. presented an analogical reasoning process model the for geometric analogies, referred as A:B::C:D (Mulholland et al., 1980). It assumed two stages of analogical reasoning process. The first stage was comparison and decomposition process; the second stage involved transformation analysis and rule generation. It focused on two components of processing, which were pattern comparison and transformation analysis. The features and transformations of pair A-B required to be recognized by subjects and stored in working memory, then applied the stored information to pair C-D. Thus, it could be possible to calculate item difficulty based on error rate, numbers of elements, as well as transformations. This method gave insights to processing stages and individual differences in cognitive abilities.

## 1.3 Analogical reasoning process of children

Previous studies have demonstrated that children have the ability to solve analogical reasoning tasks since early age (Brown & Kane, 1988; Goswami & Brown, 1991). For instance, 2-year-old children could be able to finish analogical reasoning tasks (Singer-Freeman, 2005), while they could not achieve adult-like performance until late adolescence. The analogical reasoning ability is with great variability during the childhood. The researchers concluded that the variability in strategy use on problem analogy tasks was common for both the children not in the training trials and the children in the training trials (Brown & Kane, 1988; Goswami & Brown, 1991; Siegler & Svetina, 2002; Tunteler et al., 2008; Tunteler & Resing, 2002, 2007a, 2007b).

One explanation for age-related change of children's analogical reasoning performance is that children have limited working memory capacity. They could be able to remember more rules and features as the working memory capacity increases (Primi & Paulo, 2002; Richland, Morrison, & Holyoak, 2006; Thibaut, French, & Vezneva, 2010). Working memory was shown to play a role as moderator in training and transfer of analogical reasoning (Stevenson, Resing, & Heiser, 2013). Limited capacity of working memory led children more likely to choose self-terminating operating, rather than exhausting each possibilities (Sternberg, 1977; Sternberg & Rifkin, 1979).

Another possible reason is that children have not gained enough knowledge to understand the rules of tasks in their early ages (Chen, Siegler, & Daehler, 2000; Goswami & Brown, 1991). The level of background knowledge differences among children might due to parenting style, the educational level of parents, the peer effect, and the neighbourhood environment, etc. The individual differences of background knowledge levels were not focused in the current study, since the possible causes for the individual differences were various.

## 1.4 Strategies for solving analogical reasoning tasks

Previous studies found that children used various strategies to solve the analogical reasoning tasks (Matzen, van der Molen, & Dudink, 1994; Siegler & Svetina, 2002; Tunteler et al., 2008). Different strategies resulted in several analogical reasoning errors.

Inhelder and Piaget (1964) found that children chose duplicates of the objects near the blank square of the matrix, before they responded correctly (Inhelder & Piaget, 1958). This finding influenced the matrix complete research in the analogical reasoning field. Siegler and

Svetina confirmed the previous finding. They conducted matrix completion experimental sessions among 6-8 year-old children. The results of their experiments showed that most errors in each session were duplicate errors (Siegler & Svetina, 2002).

More recently, Tunteler and Resing (2007) studied the performances on the problem analogy tasks among 5-7 year-old children (Tunteler & Resing, 2007b). They distinguished three groups of reasoners, (1) children who showed consistent analogical reasoning over trials; (2) children who showed consistent inadequate, non-analogical reasoning; and (3) children who showed variable, adequate and inadequate reasoning.

Based on their previous findings, Tunteler, Pronk and Resing (2008) studied the changes of analogical reasoning ability on the geometric analogical reasoning problems among 6-8 year-old children (Tunteler et al., 2008). The effect of a short training procedure was included to check inter-individual variability. They distinguished four kinds of analogical reasoning solutions, (1) explicit analogical solutions; (2) implicit analogical solutions; (3) incomplete analogical solutions; and (4) non-analogical solutions.

In general, children's analogical reasoning strategy was considered to be a polytomous variable, which contained four categories (correct, partial correct, duplicate and other). The item response theory (IRT) models were applied for analysing the polytomous response variable.

### 1.5 Traditional IRT models for analogical reasoning process

Item response theory (IRT) models have been widely applied to analyse test scores in analogical reasoning studies. IRT models include a family of measurement models, in which item responses are related to a latent variable. These models have been proven to be efficient in psychological and behavioural studies, because they indicated characteristics of items and characteristics of the respondent (van der Maas, Molenaar, Maris, & Kievit, 2011). IRT models have various advantages compared to classical test theory, because these models focus on the mathematical relations between the item responses, and a set of person and item parameters (De Boeck et al., 2011). The most well-known IRT models for the polytomous variable are the Partial Credit Model (PCM), the Graded Response Model (GRM), and the Generalized Partial Credit Model (GPCM) (Hoskens & De Boeck, 1995; Masters, 1982).

Cnossen (2015) analysed the children's analogical reasoning process by using three traditional IRT models. These three models were the Partial Credit Model (PCM), the graded-response model (GRM), and the Cumulative Response Model (CRM). The results showed that the GRM was the most appropriate model among the three traditional IRT models

(Cnossen, 2015). However, the stages of children's analogical reasoning process were not considered.

The traditional IRT models have several disadvantages. First, these models have limited flexibility for including different types of variables within one model, because each IRT model has its specification. The second disadvantage is that the traditional IRT models are difficult to be interpreted by the theories, especially cognitive process theories. For instance, the Sternberg's component theory demonstrated six important components in the children's analogical reasoning process (Sternberg, 1977; Sternberg & Rifkin, 1979). The traditional IRT models cannot relate the item parameters to certain components and stages during the analogical reasoning process.

### 1.6 IRTree models to understand cognitive processes

In order to increase the model flexibility, and to investigate features and reasoning process of the response categories, the IRTree models with a tree structure have been provided (De Boeck & Partchev, 2012). The IRTree models belong to the generalized linear mixed model (GLMM) family. Within a tree structure, squares represent nodes, arrows are branches, and leaves are the ends of nodes, which indicate the outcomes of item response processes. For instance, *Figure 2* displayed a linear tree model with three response categories. This IRTree model had two nodes, and each node had two branches. The end of the branches reached three response categories.

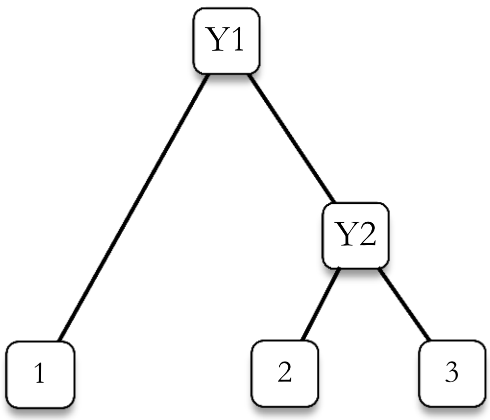



*Figure 2.* An example of IRTree model with three response categories

The response categories of the IRTree models can be either dichotomous (e.g. yes or no) or polytomous (e.g. agree, neural, or disagree). A binary tree with two branches represented a sequential process of item responses from the top of tree to the end nodes.

Based on the IRTree models, researchers attempted to represent cognitive processing mechanisms from statistical perspective, and to build connections between the IRTree models and theoretical models. Recently, a new IRTree model called generalized IRTree model has been developed (Jeon & De Boeck, 2015). The generalized IRTree model has three main advantages comparing to the traditional IRT models. First of all, it allows more flexibility of latent variables for analysing an item response process by utilizing a tree structure, instead of only focusing on the item responses. The second advantage is that the parameters of items can be node-specific or shared among nodes. Thirdly, the node-specific structure allows different IRT models specified in each node. For instance, if the first node of an IRTree model had two branches, and the second node had three responses. In this case, a binary IRT model can be conducted for the first node, and a multivariate IRT model can be applied for the second node. The IRTree model can combine the two models for specific nodes. Given these advantages, the generalized IRTree model was applied in the current study.

The mapping matrix $T$ is of size $M * K$, the element $T_{mk}$ ($m = 1, \ldots, M$, $k = 1, \ldots, K$) represents the outcome at the internal Node $k$. That is, the element $T_{mk}$ take values 0, 1, 2, …, (L - 1) when the Node $k$ includes $L$ branches, and it shows *NA* when node $k$ does not appear in the path to the observed outcome $m$. The conditional probability of internal outcome $T_{mk}$ at the Node $k$ can be calculated as follows,

$$\Pr\,(\mathrm{Y}_{pik} = \mathrm{T}_{mk} \mid \theta_{pk}) = g^{-1}\,(\alpha_{ik}\theta_{pk} + \beta_{ik}), \tag{1}$$

where $p$ refers to the subject ($p = 1, \ldots, N$), $i$ refers to the specific item ($i = 1, \ldots I$), and $k$ is node ($k = 1, \ldots, K$). $\theta_{pk}$ refers to the latent variable for person $p$ at Node $k$. For item $i$ at node $k$, $\alpha_{ik}$ indicate the parameter of item slope, and $\beta_{ik}$ is the item intercept parameter. The link function $g$ could adjust to different numbers of branches. For instance, when node $k$ includes two branches, the link function $g$ could be a logit or probit function for binary responses (e.g., $T_{mk} = 0$ or 1). When Node $k$ includes more than two branches, the link function $g$ could be adjacent logit or cumulative function (Jeon & De Boeck, 2015).

By using the conditional probabilities of internal outcomes $Y_{pik} = T_{mk}$ (1), the model for observed terminal outcome $Y_{pi} = m$ ($m = 1, \ldots, M$) is formulated as follows,

$$\Pr\,(Y_{pi} = m \mid \theta_{p1}, \ldots, \theta_{pK})$$

$$= \Pr\,(Y^*_{pi1} = T_{m1}, \ldots, Y^*_{pik} = T_{mK} \mid \theta_{p1}, \ldots, \theta_{pK})$$

$$= \prod_{k=1}^{K} \Pr\big(Y^*_{pik} = T_{mk} \mid \theta_{p1}, \ldots, \theta_{pk}\big)^{t_{mk}}, \tag{2}$$

where $t_{mk} = T_{mk}$ if $T_{mk} = 0$ or 1, and $t_{mk} = 0$ if $T_{mk}$ = NA ($k = 1, \ldots, K$, $m = 1, \ldots, M$). The K latent variables $\theta_p = (\theta_{p1, \ldots,} \theta_{pK})'$ are assumed to follow a multivariate normal distribution with $\theta_p \sim N(0, \sum)$, where $\sum$ is a $K * K$ covariance matrix. Thus, the $K$ node-specific latent traits are allowed to be correlated with each other (Jeon & De Boeck, 2015).

### 1.7 Relations between analogical reasoning theories and IRTree models

According to Sternberg’s three-node components theory for children’s analogical reasoning process (Sternberg, 1977) and Mulholland’s two-node theory (Mulholland et al., 1980), two tree structures of IRTree models have been formulated (See *Figure 3*). Each tree structure was argued in the following section based on the analogical reasoning theories.

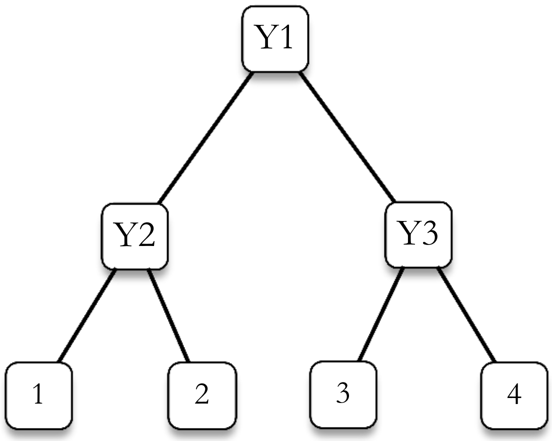



*Figure 3(a).* Binary tree structure for the four categories polytomous variable

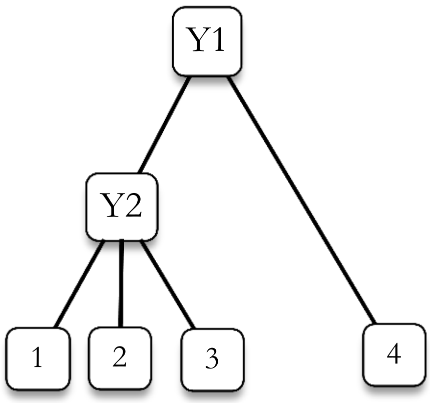



*Figure 3(b).* Tree structure for the four categories polytomous variable

Tree *3(a)* denoted a binary tree structure with three nodes, which is formulated based on the Sternberg’s component theory (Sternberg, 1977). It assumed that children’s analogical reasoning is a three-stage process. Y1 referred to encoding and inference stage. Children who used analogical reasoning strategy were in the stage Y2, while others who used non-analogical reasoning strategies went to the stage Y3. The stage Y2 indicated as the mapping stage. In the stage Y2, children processed all transformations correctly recorded "Correct" responses; the others who made mistakes in the mapping process recorded "Partial Correct" responses. The stage Y3 referred as the application. In the stage Y3, children mapped correctly but applied wrongly tended to choose "Duplicate", and others were classified as "Other".

Tree *3(b)* had one response category qualitatively different from the other three categories, which was based on the Mulholland's two-sage model (Mulholland et al., 1980). The stage Y1 represented as pattern comparison and decomposition. In this stage, each feature and pattern of the analogical tasks were isolated and compared. The stage Y2 represented as transformation analysis and rule generation. During this stage, children specified the rules for transforming the A stimulus into the B stimulus. This tree structure assumed that children who made mistakes in the stage Y1 of pattern comparison and decomposition are qualitatively different from others, probably due to age-related difference (Brown & Kane, 1988; Chen et al., 2000). Children who correctly compare the patterns in the stage Y1 need to make a second decision in the stage Y2 of transformation analysis and rule generation. This stage may relate to children's working memory capacity (Stevenson, Resing, et al., 2013; Swanson, 2008). Children who made mistakes during the transformation were most likely to choose duplicates Those children who missed some parts of features in the transforming and rule generating were recorded as "Partial Correct". Children who answered correctly in both stages were in the category of "Correct".

### 1.8 Research questions

The aim of this study is to gain insight into children's analogical reasoning processing while solving figural analogical reasoning tasks. To achieve this, the generalized IRTree models with four different tree structures have been applied to the current dataset. Two research questions have been addressed. (1) Which model is the best fit for the current dataset of children's analogical reasoning strategy? (2) Which model is the best fit for the dataset including person variables of age and working memory capacity?

## 2 Method

### 2.1 Sample

There were 1002 participants in the current dataset. The children were recruited from 28 public elementary schools of similar middle class social economic states (SES) in the southwest of the Netherlands. The sample consisted of 490 boys and 512 girls, with a mean age of 7 years, 3 months (range 4.9-11.3 years).

### 2.2 Design and procedure

The present cross-sectional study used the pretest data from a large project of children's analogical reasoning strategy, which combined six analogical reasoning experiments, and each experiment utilized a pretest-intervention-posttest-control group design (Stevenson, Hickendorff, Resing, Heiser, & De Boeck, 2013). The data was already collected before the present study.

### 2.3 Material

A computerized figural analogy task called AnimaLogica (Stevenson, Hickendorff, et al., 2013) has been used to test children's analogical reasoning process. As it showed in *Figure 3*, the figural analogies task consisted of 2 x 2 matrices with coloured animals pictures. These animals had six transformation features, animals (camel, bear, dog, horse, lion or elephant), colour (yellow, blue or red), orientation (left or right), position (top or bottom), quantity (one or two) and size (small or large). Children were asked to fill in the empty box by choosing an animal card, so that the bottom two figures shared the same relation as the top two figures (A:B::C:?).

### 2.4 Variables

The response variable in the present study is the strategy, which used by children when solving figural analogical tasks. The strategy was classified into four categories (correct, partial correct, duplicate, or other). It was an ordinal variable. The "Correct" analogical strategy was the highest level of reasoning performance, and then followed by "Partial Correct", which both were analogical reasoning strategies. The other two categories, "Other" and "Duplicate", were considered as non-analogical reasoning strategies. The orders between other and duplicate can be reversed, based on different interpretation of analogical reasoning theories (See *Section 1.4*). An example of four strategies has been presented as *Figure 4*. The "Correct" analogical strategy was recorded when the answer of item was correct. "Partial Correct" was recorded when one or two transformations were missing in the answer. "Other" was recorded when three or more transformations were missing. "Duplicate" was recorded when the answer was copied from one of already existed matrix. (Stevenson, Hickendorff, et al., 2013).

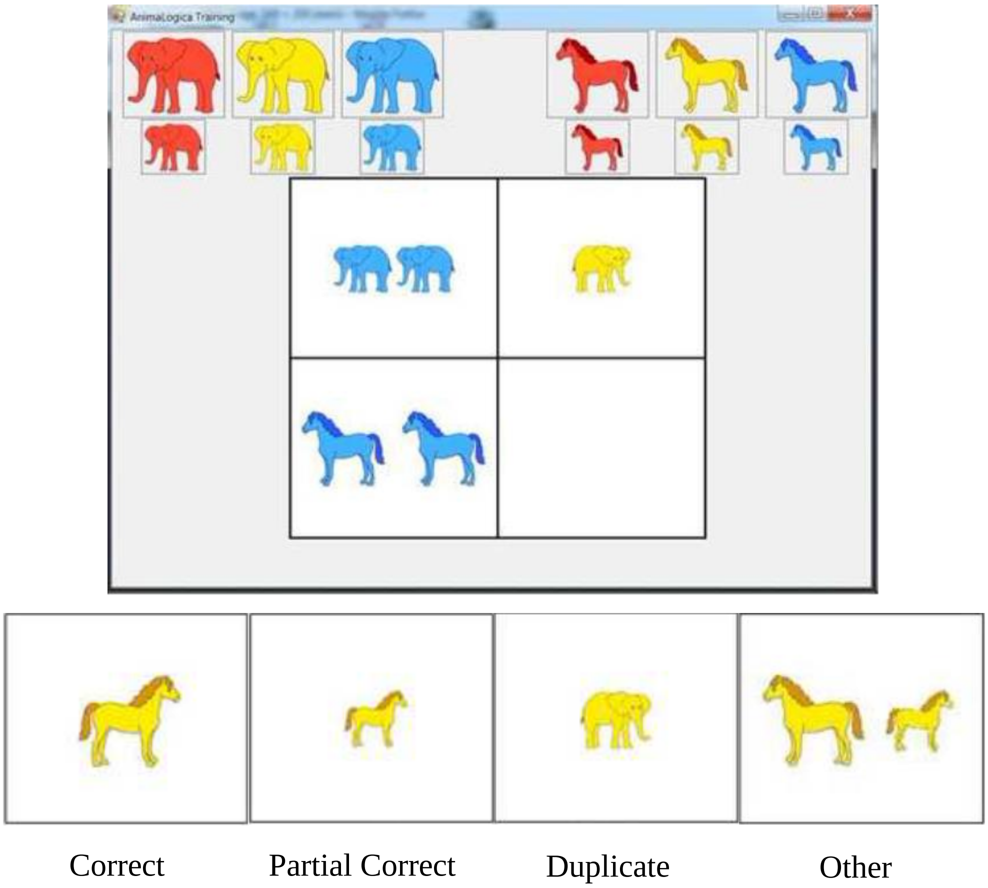



*Figure 4.* An example of task screen and four categories of strategy

In addition to the response variable, two person variables were collected. First, age of each child was recorded. Second, working memory capacity was measured for each children by an age appropriate verbal memory test, which included AWMA listening recall (Alloway, 2007), WISC-IV digit span (Wechsler, 2003), and RAKIT memory span (Bleichrodt, Drenth, Zaal, & Resing, 1984).

## 2.5 Properties of the dataset

Three specific properties of the dataset have been considered during the exploration of current dataset.

First of all, different orders between the two categories "Other" and "Duplicate" are explored. In the original dataset for the present study, the category "Other" was coded as the lowest order category (Stevenson, Hickendorff, et al., 2013). The category "Duplicate" was defined as the subjects copied one of the already showed figures, which indicated the subject might recognize certain features of already visible figures while could not understand the relations among the features. The "Other" category was recorded when three or more features were missing, which indicated the subject made mistakes of recognizing the features of already visible figures in the first place. However, the category "Duplicate" was considered as a qualitatively different response comparing with other responses in previous studies, because it was the most common non-analogical response from children (Siegler, 1999; Siegler & Svetina, 2002). Thus, the "Other" and "Duplicate" both could be the lowest ordered category among the four categories of strategy.

Secondly, all the sample children gave responses to 7 out of 21 items from different schools. The seven items were common items, which were used in the following IRTree modelling analysis. The reliability of the seven common items was checked in the following section of results. Previous study showed that the seven common items fitted well by the traditional IRT models (Cnossen, 2015).

Thirdly, the person variable working memory capacity contained 256 missing data. This affects the IRTree model analysis. Since the working memory scores were normally distributed, the missing data were replaced by the means before conducting the IRTree models.

## 2.6 Explanatory IRT

### 2.6.1 Fitting the IRTree models

Two tree structures of IRTree models were applied for both the original ordered response variable and the adjusted ordered response variable. Thus, four IRTree models were conducted in the present study.

Model 1 is a nested tree structure with three nodes. The lowest order category is "Other", followed by "Duplicate", "Partial correct" and "Correct".

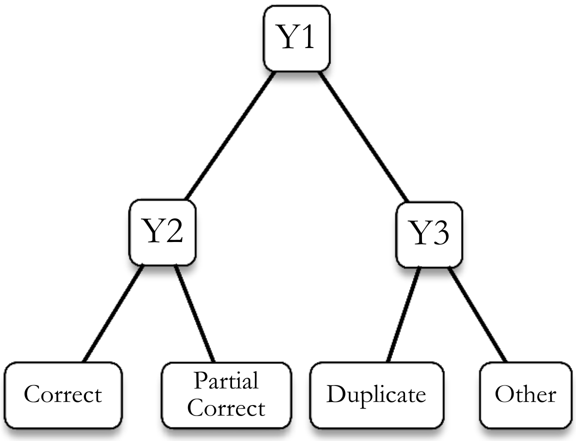



*Figure 5.* Model 1 tree structure

Table 1.

*Model 1 Mapping matrixes of four categories of response*

| | $Y_{pi1}$ | $Y_{pi2}$ | $Y_{pi3}$ |
|---|---|---|---|
| $Y_{pi}$ = 1 (Other) | 0 | NA | 0 |
| $Y_{pi}$ = 2 (Duplicate) | 0 | NA | 1 |
| $Y_{pi}$ = 3 (Partial) | 1 | 0 | NA |
| $Y_{pi}$ = 4 (Correct) | 1 | 1 | NA |

Model 2 is a nested tree structure with three nodes. Comparing with Model 1, the order between two categories "duplicate" and "other" have been reversed in Model 2. The lowest order category is "Duplicate", followed by "Other", "Partial Correct", and "Correct".

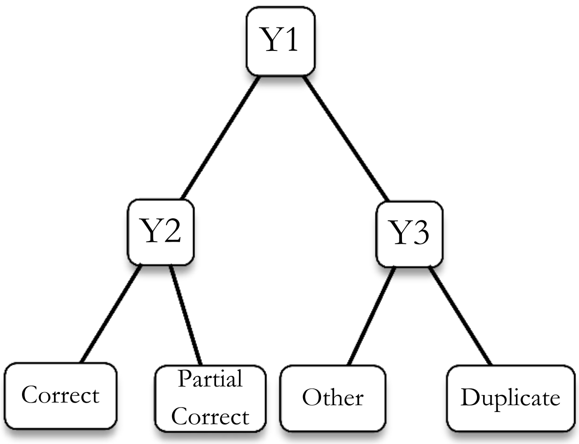



*Figure 6.* Model 2 tree structure

Table 2.

*Model 2 Mapping matrixes of four categories of response*

| | $Y_{pi1}$ | $Y_{pi2}$ | $Y_{pi3}$ |
|---|---|---|---|
| $Y_{pi}$ = 1 (Duplicate) | 0 | NA | 0 |
| $Y_{pi}$ = 2 (Other) | 0 | NA | 1 |
| $Y_{pi}$ = 3 (Partial) | 1 | 0 | NA |
| $Y_{pi}$ = 4 (Correct) | 1 | 1 | NA |

Model 3 is a two-node IRTree model. One category is qualitatively different from the other three. The lowest order category is "other", followed by "duplicate", "partial correct", and "correct".

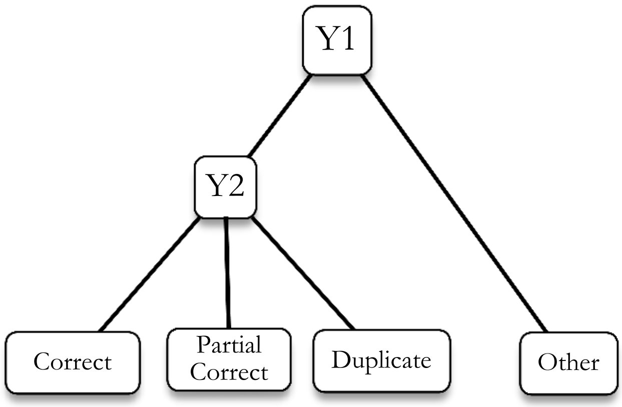



*Figure 7.* Model 3 tree structure

Table 3.

*Model 3 Mapping matrixes of four categories of response*

| | $Y_{pi1}$ | $Y_{pi2}$ |
|---|---|---|
| $Y_{pi}$ = 1 (Other) | 0 | NA |
| $Y_{pi}$ = 2 (Duplicate) | 1 | 0 |
| $Y_{pi}$ = 3 (Partial) | 1 | 1 |
| $Y_{pi}$ = 4 (Correct) | 1 | 2 |

Model 4 also has a two-node tree structure, with one category deviated from the other three. Comparing with Model 3, the order between two categories "duplicate" and "other" have been reversed in Model 4.The lowest order category is "duplicate", followed by "other", "partial correct", and "correct".

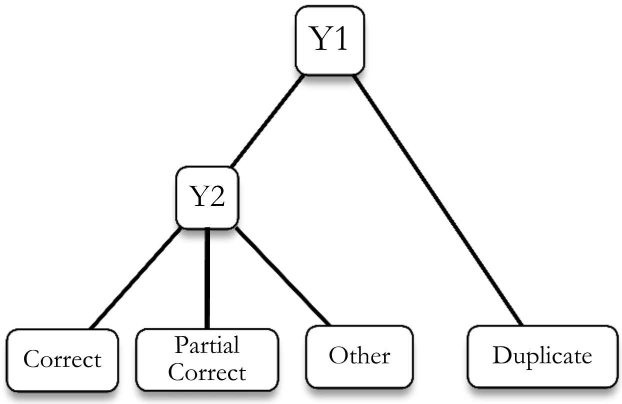



*Figure 8.* Model 4 tree structure

Table 4.

*Model 4 Mapping matrixes of four categories of response*

| | $Y_{pi1}$ | $Y_{pi2}$ |
|---|---|---|
| $Y_{pi}$ = 1 (Duplicate) | 0 | NA |
| $Y_{pi}$ = 2 (Other) | 1 | 0 |
| $Y_{pi}$ = 3 (Partial) | 1 | 1 |
| $Y_{pi}$ = 4 (Correct) | 1 | 2 |

### 2.6.2 Traditional IRT models

The response variable "strategy" is an ordered polytomous variable with four categories. Previous study claimed that the Graded Response Model (GRM) was the most appropriate model for the children's analogical reasoning strategy, comparing with the Partial Credit Model (PCM) and the Continuation Ratio Model (CRM) (Cnossen, 2015). Therefore, the GRM was also chosen for the analysis in the present study. In addition, the generalized Partial Credit Model (GPCM) was applied for the response variable (Muraki, 1992), which has not been tested by previous study (Cnossen, 2015).

#### *2.6.2.1 Graded Response Model*

The GRM is an extension of the two-parameter logistic (2PL) model, which belongs to the class of cumulative probability models (Hemker, van der Ark, & Sijtsma, 2001; Samejima, 1969). Each item is described by the slope parameter ($\alpha_i$) and $j$ ($j$ = 1, 2, …, $m_i$), in addition to the item difficulty parameter ($\beta_i$). (Embretson & Reise, 2000). In the GRM, the probability of a person $p$'s item response ($x$) to be equal or greater than a given category threshold ($j$) on the item $i$ can be calculated as follows:

$$P^{*}_{ix}(\theta) = \frac{\exp[\alpha_i(\theta_p - \beta_{ij})]}{1+\exp[\alpha_i(\theta_p - \beta_{ij})]} \qquad (3)$$

where $P^{*}_{i0}(\theta) = 1$, $P^{*}_{im}(\theta) = 0$ and $x = j$. In this study, the subjects' latent traits ($\theta_p$) are normally distributed and means equal to zero ($\theta \sim N(0, \sigma_\theta^{2})$). The GRM is suitable for the polytomous response variables. In the GRM, the $\alpha_i$ parameters are not item discrimination parameters as in other 2PL models, but instead they are slope parameters. This is due to the discrimination of categorical items also depends on the category thresholds $j$ spread. For the response variable in present study, the probabilities of responses x = 0 versus 1, 2 and 3, x = 0, 1 versus 2, 3 and x = 0, 1, 2 versus 3 are calculated with constraint that the item slopes are equal (see *Figure 9*).

<table>
<tr><th rowspan="2">Four ordered categories</th><th colspan="3">Cumulative Probabilities</th></tr>
<tr><th>Categories 1, 2 and 3 vs. 0</th><th>Categories 2 and 3 vs. 0 and 1</th><th>Categories 3 vs. 0, 1 and 2</th></tr>
<tr><td>0</td><td>0</td><td rowspan="2">0 & 1</td><td rowspan="3">0 & 1 & 2</td></tr>
<tr><td>1</td><td rowspan="3">1 & 2 & 3</td></tr>
<tr><td>2</td><td rowspan="2">2 & 3</td></tr>
<tr><td>3</td><td>3</td></tr>
</table>

*Figure 9.* Cumulative probability model

The probability of a subject responding in the category x to item I is calculated by subtracting the cumulative probabilities (Samejima, 1969). For the same example in Figure 9, the probabilities of responding in each category are given by equations (4.1) to (4.4). These four equations can be generated into one equation (5) with the total probability equals 1.

$$P_{i0}(\theta) = 1 - P_{i1}(\theta) \qquad (4.1)$$

$$P_{i1}(\theta) = P_{i1}(\theta) - P_{i2}(\theta) \qquad (4.2)$$

$$P_{i2}(\theta) = P_{i2}(\theta) - P_{i3}(\theta) \qquad (4.3)$$

$$P_{i3}(\theta) = P_{i3}(\theta) - 0 \qquad (4.4)$$

$$P_{i3}(\theta) = P_{ix}(\theta) - P_{i(x+1)}(\theta) \qquad (5)$$

#### *2.6.2.2 Generalized Partial Credit Model*

The GPCM is formulated according to the assumption that the probability of choosing the *k*th category over the *k* minus the first (*k* - 1) category is controlled by the dichotomous response model (Muraki, 1992). The GPCM extended the 1PL Partial Credit Model (PCM) (Masters, 1982), and retained the item discriminating power in the model. Therefore, the GPCM is suitable for the polytomous response variable. Let $P_{jk}(\theta)$ denote the specific probability of selecting the *k*th category from $m_j$ categories of item *j*. The probability of a specific categorical response *k* over *k* – 1 is given by the conditional probability:

$$C_{jk} = P_{jk/k\text{-}1,k}(\theta) = \frac{P_{jk}(\theta)}{P_{jk-1}(\theta) + P_{jk}(\theta)} = \frac{\exp[\alpha_j(\theta - b_{jk})]}{1+\exp[\alpha_j(\theta - b_{jk})]} \quad (6)$$

Where the $k = 1, 2, \ldots, m_j$. After normalizing each $P_{jk}(\theta)$, the total sum of $P_{jk}(\theta)$ equals 1. The GPCM is an adjacent category model, the adjacent ratios can be calculated for probabilities of responses $x$ = 1 versus 0, $x$ = 2 versus 1, and $x$ = 3 versus 2. (see *Figure 10*).

| Four ordered | Adjacent Categories | | |
|---|---|---|---|
| categories | Categories 1 vs. 0 | Categories 2 vs. 1 | Categories 3 vs. 2 |
| 0 | 0 | | |
| 1 | 1 | 1 | |
| 2 | | 2 | 2 |
| 3 | | | 3 |

*Figure 10.* Adjacent category model

### 2.6.3 Software

The maximum likelihood estimation proposed generalized IRTree models have been estimated with the freely available R package "FLIRT" (Jeon, Rijmen, & Rabe-Hesketh, 2014). A major advantage of "FLIRT" is that a variety of one and two parameter logistic and bi-factor IRT models could be built and explored by a rich number of modeling options, except three parameter logistic IRT models for now. The "FLIRT" package provides an IRT-friendly approach of modeling different hypotheses on item and person parameters. Therefore, it is suitable for exploring different tree models and analogical processes.

The "ltm" package was applied for analyzing the Graded Response Model (GRM) and the generalized Partial Credit Model (GPCM), for the original ordered response variable and the adjusted ordered response variable (Rizopoulos, 2006).

### 2.7 Model selection

The fit indices AIC and BIC values were used to compare among different models in the present study (Akaike, 1974; Schwarz, 1969). Both values could be calculated for each model in the R packages "FLIRT" and "ltm" (Jeon et al., 2014; Rizopoulos, 2006). The final model was assumed to have the lowest AIC and BIC values, and included the most number of parameters of the dataset. In addition, it is expected that the final model can be easily interpreted by the analogical reasoning theories.

## 3 Results

### 3.1 Descriptive statistics

Descriptive statistics of the seven items with original orders are shown in *Table 5*. Age and working memory were not correlated ($r$ = .004, $p$ = .91). 737 out of 1002 respondents have reported working memory scores. Missing data has been taken into consideration in the following analysis.

Table 5.

*Descriptive statistics for the seven items in original orders*

| Item | *N* | Minimum | Maximum | Mean | Median | *SE* | *Sd* | Variance |
|---|---|---|---|---|---|---|---|---|
| 201 | 1002 | 1 | 4 | 3.00 | 3.00 | .029 | .919 | .844 |
| 204 | 1002 | 1 | 4 | 2.99 | 3.00 | .030 | .951 | .905 |
| 301 | 1002 | 1 | 4 | 2.98 | 3.00 | .029 | .908 | .824 |
| 404 | 1002 | 1 | 4 | 2.57 | 3.00 | .032 | 1.005 | 1.010 |
| 502 | 1002 | 1 | 4 | 2.18 | 2.00 | .033 | 1.036 | 1.073 |
| 505 | 1002 | 1 | 4 | 2.18 | 2.00 | .033 | 1.029 | 1.059 |
| 604 | 1002 | 1 | 4 | 2.09 | 2.00 | .032 | 1.026 | 1.052 |

### 3.2 Classical test theory (CTT) results

The Cronbach's alpha of the seven items with original orders equalled 0.843 (95% CI: 0.828-0.856), which indicated good reliability of the test. When the orders between "Duplicate" and "Other" reversed, the Cronbach's alpha of the seven items was slightly increased as 0.853 (*95% CI*: 0.837-0.867).

Table 6.

*The proportion of strategy used per item*

| Item | Non-analogical | | Analogical | |
|---|---|---|---|---|
| | Duplicate | Other | Partial Correct | Correct |
| 201 | 0.26 | 0.05 | 0.32 | 0.37 |
| 204 | 0.35 | 0.04 | 0.20 | 0.41 |
| 301 | 0.25 | 0.06 | 0.35 | 0.34 |
| 404 | 0.23 | 0.19 | 0.39 | 0.19 |
| 502 | 0.23 | 0.35 | 0.31 | 0.11 |
| 505 | 0.25 | 0.34 | 0.30 | 0.11 |
| 604 | 0.21 | 0.39 | 0.31 | 0.09 |

The proportion of strategy used per item showed that Item 204 was the easiest item with the highest proportion of "Correct" and the lowest proportion of "Other". Item 604 was the most difficult one with the highest proportion of "Other" and lowest proportion of "Correct". The proportion of response category "Duplicate" did not vary much among the seven items.

Since the proportion of the category "Duplicate" did not vary much among the seven items, it might belong to another distinct category, which was different from the other three categories. The traditional IRT models and the IRTree models were used to analyse two categorical orders of response variable.

### 3.3 What is the better order among categories of response variable?

Two traditional IRT models, the Graded Response Model (GRM) and the Generalized Partial Credit Model (GPCM), were conducted for analysing both the original ordered response variable and the adjusted response variable with reversed orders between "Duplicate" and "Other".

#### 3.3.1 Graded Response Model for the original ordered response variable

Table 7.

*Coefficients parameters for each category per item of original response variable*

| Item | Category 1 | Category 2 | Category 3 | Slope |
|---|---|---|---|---|
| 201 | -2.224 | -0.738 | 0.369 | 1.784 |
| 204 | -2.762 | -0.483 | 0.283 | 1.511 |
| 301 | -3.134 | -1.015 | 0.711 | 1.018 |
| 404 | -1.153 | -0.327 | 1.175 | 2.036 |
| 502 | -0.551 | 0.106 | 1.475 | 3.020 |
| 505 | -0.603 | 0.180 | 1.591 | 2.204 |
| 604 | -0.493 | 0.230 | 1.864 | 1.892 |

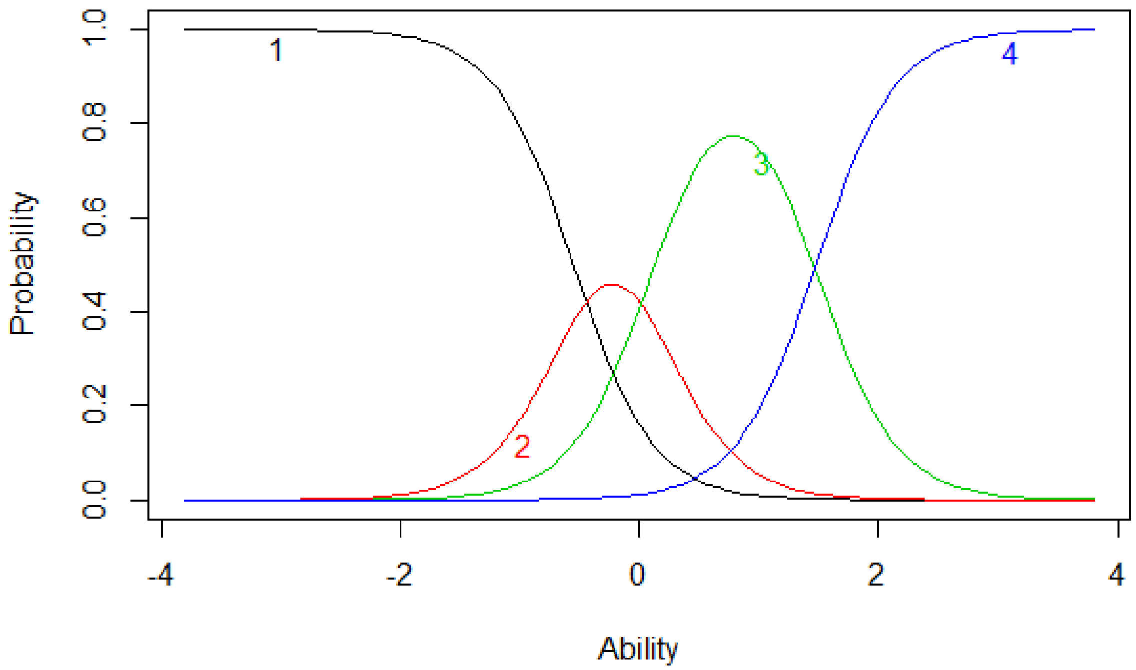



*Figure 11*. Category Response Curves of item 502 under the GRM

The results of coefficients parameters of GRM for each category per item have displayed in the *Table 10*. The coefficients represented the point on the latent scale where a subject had a.50 probability of responding within or above the category $j = x$. For instance, for the Item 502, a subject with a trait level of -0.551 had a probability of responding in or above the category 1; and the subject with a trait level of 0.106 had .50 probability of responding in or above the category 2. In the *Figure 11*, the category response curves of the item 502 are presented.

The slope parameters (α) were included in the GRM, since it is a 2PL model. The value of the item slope parameter represented the amount of information that was provided by the item. For instance, the Item 502 had the largest slope parameter among the seven common items. This indicated that the item functions well for distinguishing between subjects with different trait levels.

### 3.3.2 Graded Response Model for the adjusted ordered response variable

Table 8.

*Coefficients parameters for each category per item of reversed response variable*

| Item | Category 1 | Category 2 | Category 3 | Slope |
|---|---|---|---|---|
| 201 | -0.892 | -0.695 | 0.412 | 1.878 |
| 204 | -0.561 | -0.431 | 0.292 | 1.769 |
| 301 | -1.256 | -0.937 | 0.717 | 1.097 |
| 404 | -0.915 | -0.246 | 1.113 | 2.483 |
| 502 | -0.894 | 0.186 | 1.480 | 2.874 |
| 505 | -0.869 | 0.237 | 1.548 | 2.368 |
| 604 | -0.966 | 0.306 | 1.636 | 2.674 |

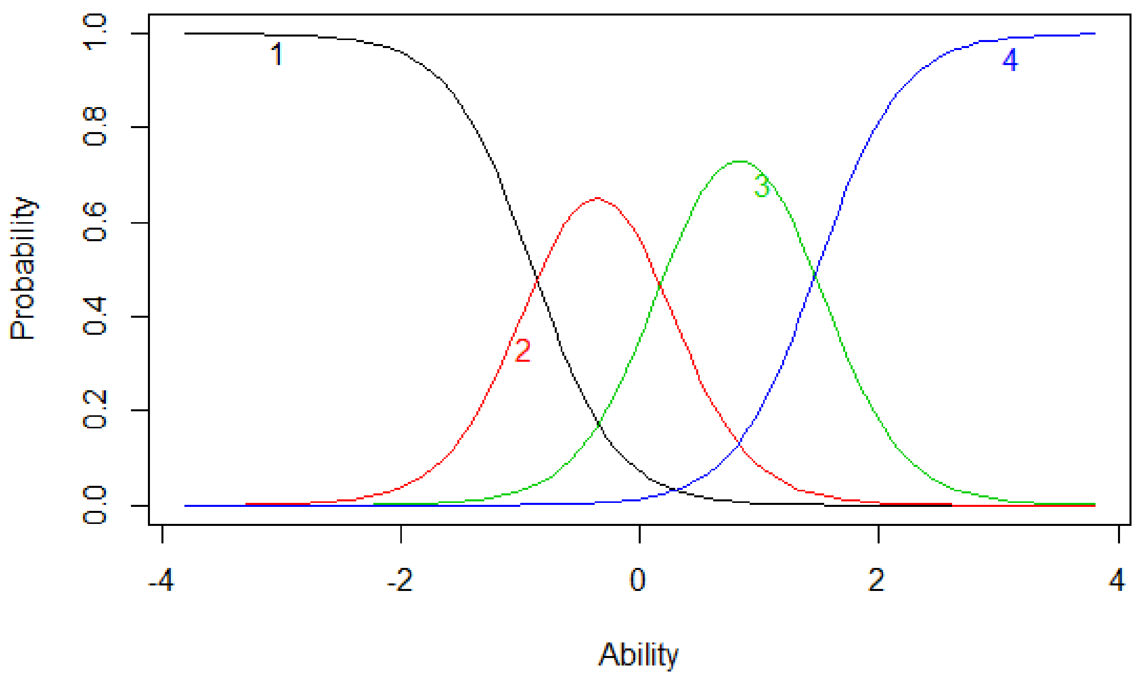



*Figure 12*. Category Response Curves of item 502 under the GRM

The orders between categories "Duplicate" and "Other" have been reversed in this GRM. The results of coefficients parameters of GRM for each category per item have displayed in the *Table 11*. For the Item 502 in the adjusted ordered response variable dataset, a subject with a trait level of -0.894 had a probability of responding in or above the category 1; and the subject with a trait level of 0.237 had a .50 probability of responding in or above the category 2. In the *Figure 12*, the category response curves of the item 502 have been presented.

The slope parameters ($\alpha$) were also included in this GRM. The value of the item slope parameter represented the amount of information that was provided by the item. For instance, the Item 502 had the largest slope parameter among the seven common items, which indicated that the item functions well for distinguishing between subjects with different trait levels.

### 3.3.3 Generalized Partial Credit Model for the original ordered response variable

Table 9.

*Coefficients parameters for each category per item of original response variable*

| Item | Category 1 | Category 2 | Category 3 | Discrimination |
|---|---|---|---|---|
| 201 | -2.196 | -0.575 | 0.149 | 1.302 |
| 204 | -3.226 | 0.204 | -0.471 | 0.965 |
| 301 | -2.849 | -0.802 | 0.259 | 0.687 |
| 404 | -0.846 | 0.520 | 1.134 | 1.470 |
| 502 | -0.342 | 0.009 | 1.519 | 2.168 |
| 505 | -0.287 | 0.012 | 1.609 | 1.473 |
| 604 | 0.084 | -0.117 | 1.972 | 1.189 |

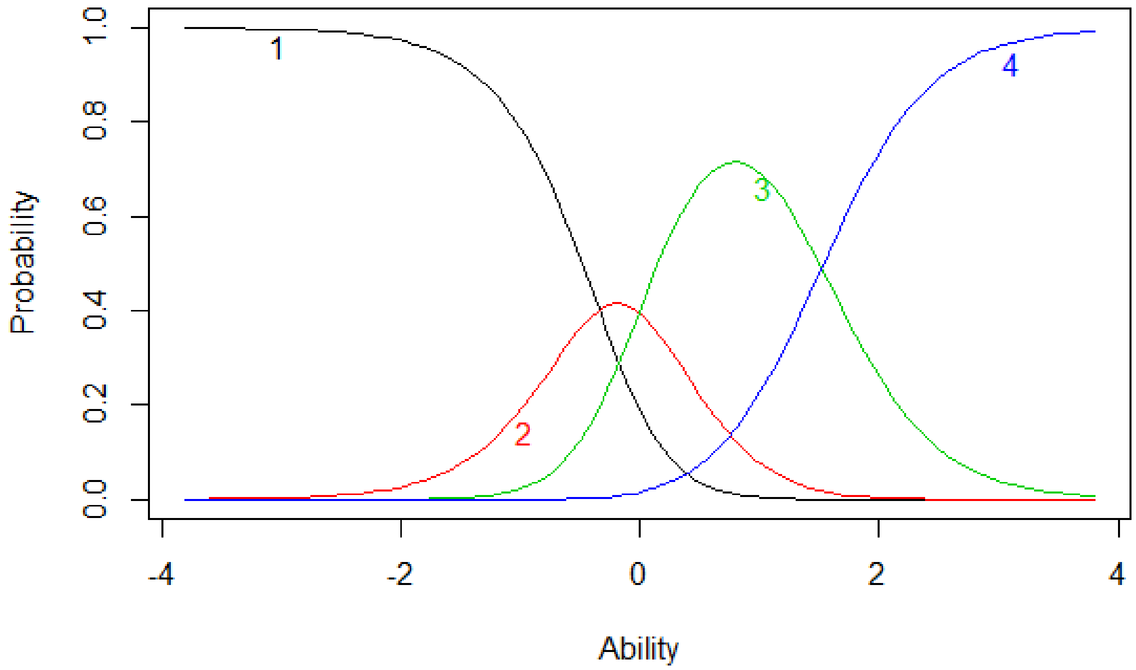



*Figure 13.* Category Response Curves of item 502 under the GPCM.

The results of coefficients parameters of GPCM for each category per item have displayed in the *Table 12*. For the Item 502 in the adjusted ordered response variable dataset, a subject with a trait level of -0.342 had a probability of responding in or above the category 1; the subject with a trait level of 0.009 had .50 probability of responding in or above the category 2; and the subject with a trait level of 1.519 had .50 probability of responding in or above the category 3. In the *Figure 13*, the category response curves of the item 502 have been presented.

The GPCM is a 2PL model, which presented the item discrimination parameter for each item. The item 502 had the largest value of item discrimination parameter. It indicated that the item is very capable of distinguishing subjects with different trait levels. This can also be seen in the *Figure 13* that the item 502 had peaked category response curves.

### 3.3.4 Generalized Partial Credit Model for the adjusted ordered response variable

Table 10.

*Coefficients parameters for each category per item in GPCM*

| Item | Category 1 | Category 2 | Category 3 | Discrimination |
|---|---|---|---|---|
| 201 | 0.980 | -2.084 | 0.176 | 0.953 |
| 204 | 2.272 | -2.282 | -0.503 | 0.804 |
| 301 | 2.252 | -3.590 | 0.311 | 0.522 |
| 404 | -0.632 | -0.465 | 1.096 | 1.848 |
| 502 | -0.837 | 0.185 | 1.474 | 2.322 |
| 505 | -0.771 | 0.229 | 1.514 | 1.804 |
| 604 | -0.927 | 0.277 | 1.632 | 2.268 |

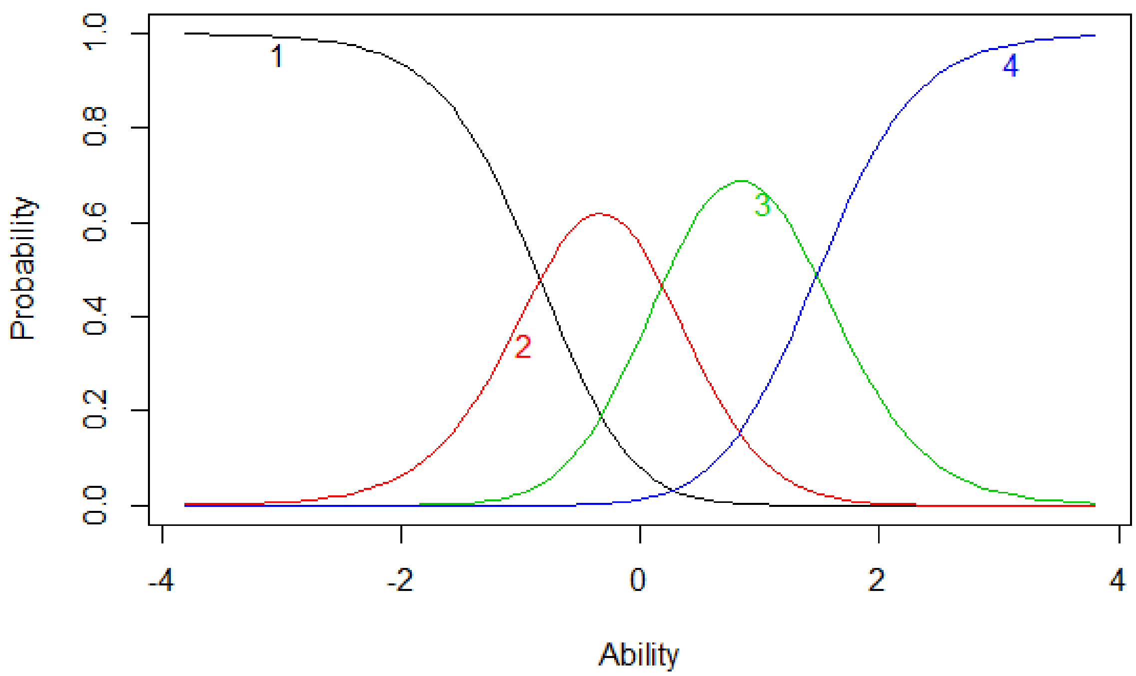



*Figure 14.* Category Response Curves of item 502 under the GPCM

The orders between categories "Duplicate" and "Other" have been reversed in this GPCM. The results of coefficients parameters of GPCM for each category per item have displayed in the *Table 13*. For the Item 502 in the adjusted ordered response variable dataset, a subject with a trait level of -0.837 had a probability of responding in or above the category 1; the subject with a trait level of 0.185 had .50 probability of responding in or above the category 2; and the subject with a trait level of 1.474 had .50 probability of responding in or above the category 3. In the *Figure 14*, the category response curves of the item 502 have been presented.

The Item 502 still had the largest value of item discrimination, and the Item 604 showed the second large value of item discrimination. This indicated that these two items are capable of distinguishing subjects with different trait levels. In addition, the most likely trait level for responding the adjusted ordered Item 502 and Item 604 correctly is higher than the trait level for responding these two items with original orders correctly. This can be proved by the *Figure 14,* which presented more peaked category response curves of Item 502 than the *Figure 13*.

### 3.3.5 Model selection

Table 11.

*Model fit indices of traditional IRT models for two orders of response variable*

| Category Orders | Models | *AIC* | *BIC* | *Log-Likelihood* |
|---|---|---|---|---|
| Original | GRM1 | 15592.42 | 15729.89 | -7768.21 |
| | GPCM1 | 15645.57 | 15783.04 | -7794.78 |
| Adjusted | GRM2 | 14975.19 | 15112.67 | -7459.59 |
| | GPCM2 | 15125.56 | 15263.03 | -7534.78 |

Generally, the values of fit indices were lower in the two IRT models for the adjusted ordered response variable, comparing with the values of fit indices in the two IRT models for the original ordered response variable. The finding indicated that the reversed orders between categories "Duplicate" and "Other" may influence the model fit. The "Duplicate" response category may be qualitatively different from the other three response categories, which can be assumed as the lowest-order category among the four categories of response variable. The IRTree models were conducted for both the original ordered response variable and the adjusted ordered response variable in following sessions, in order to compare with the findings of the two traditional IRT models.

In addition, GRMs fitted better than the GPCMs for both the original ordered response variable and the adjusted ordered response variable. This result extended the findings of previous study (Cnossen, 2015). For the ordered polytomous response variable, the GRM was the best-fit model among the PCM, CRM and GPCM.

## 3.4 Research question 1, "Which model is the best fit for the dataset of children's analogical reasoning strategy?"

Four IRTree models with two tree structures were conducted to answer this research question. The first tree structure was a binary tree structure, which assumed the category "Other" of children's analogical reasoning strategy belonged to a general category of "Non-analogical reasoning". While the second tree structure assumed that the category "Other" belonged to the general category of "Analogical reasoning" strategy. Both tree structures of IRTree models were tested for the original ordered response variable and the adjusted ordered response variable.

### 3.4.1 IRTree models

#### *3.4.1.1 Model 1*

Model 1 is a binary tree structure IRTree model for the original ordered response variable (see *Figure 5*). The covariance between the first and second node is 0.858 in Model 1. The covariance between the first and third node is approximately -0.431. The covariance between the second and third node is approximately -0.223. The relationships indicated that when "Other" is the lowest ordered category of strategy, the stage Y3 of application was in the opposite direction of stage Y1 of encoding and inference and Y2 of mapping during the process of children's analogical reasoning.

#### *3.4.1.2 Model 2*

Model 2 is a binary tree structure IRTree model for the adjusted ordered response variable (see *Figure 6*). The covariance between the first and second node is 0.858 in Model 2, which is the same as the covariance in Model 1. The covariance between the first and third node is approximately 0.431. The covariance between the second and third node is approximately 0.223. The relationships of each two nodes were positive. This indicated that when "Duplicate" is the lowest ordered category of strategy, the three stages were in the same direction during the process of children's analogical reasoning.

#### *3.4.1.3 Model 3*

Model 3 assumed the category "Other" is the lowest-order category of children's analogical reasoning strategy, which is qualitatively different than the other three categories (see *Figure 7*). The covariance between the first and second node is approximately 0.462, which indicated the relationship between the first and second node is positive. 46.2% of the sample children who responded in the stage Y1 of pattern comparison and decomposition went to the stage Y2 of transformation analysis and rule generation.

#### *3.4.1.4 Model 4*

Model 4 assumed the category "Duplicate" is the lowest-order category of children's analogical reasoning strategy, which is qualitatively different than the other three categories (see *Figure 8*). The covariance between the first and second node is approximately 0.621, which indicated the relationship between the first and second node is positive. Approximately 62% of the sample children who responded in the stage Y1 of pattern comparison and decomposition went to the stage Y2 of transformation analysis and rule generation.

### 3.4.2 Model selection

Table 12.

*Fit indices of the estimated IRTree models.*

| | AIC | BIC | Number of parameters | Log-likelihood |
|---|---|---|---|---|
| Model 1 | 14639 | 14860 | 45 | -7275 |
| Model 2 | 14639 | 14860 | 45 | -7275 |
| Model 3 | 14913 | 15090 | 36 | -7420 |
| Model 4 | 14692 | 14869 | 36 | -7310 |

The *Table 12* presented the model fit indices and the number of parameters of the four IRT tree models. Based on the values of AIC and BIC, Model 1 and Model 2 were the most appropriate model for the current dataset with same model fit indices values.

### 3.4.3 Interpretation of the best fit model

For the first research question, the Model 1 and Model 2 with binary tree structure fitted better than the other two IRTree models. This indicated that the children's analogical reasoning process followed a binary structure with three stages. In the stage Y1 of encoding and inference, children chose between two general categories of strategy, which were analogical strategy and non-analogical strategy. In the stage Y2 of mapping, children with analogical reasoning skills chose between "Correct" and "Partial Correct" strategies. In the stage Y3 of application, children with non-analogical reasoning skills chose between "Duplicate" and "Other" strategies.

The Model 1 and Model 2 presented the same model fit indices values. It is interesting to find out that the orders between categories "Duplicate" and "Other" did not matter for the IRTree models, as long as they both belonged to the general category of "Non-analogical". This finding is contrast with the result of the traditional IRT models in previous session. This might due to the IRTree models gave more in-depth information of children's analogical reasoning process than the traditional IRT models.

## 3.5 Research question 2, "Which model is the best fit for the dataset including age and working memory capacity?"

To answer this research question, the person covariates age and working memory scores for each subject were included in the dataset. The same structured IRTree models as in previous session were conducted for both the original ordered response variable and the adjusted ordered response variable together with the age and working memory capacity scores.

### 3.5.1 IRTree Models

The person variables age and working memory capacity scores have been normalized before conducting the IRTree modelling analysis. Therefore, we used the standard scores of age and working memory capacity instead of original scores when interpreting the results of each IRTree model.

#### *3.5.1.1 Model 5*

Model 5 is a binary tree structure IRTree model for the original ordered response variable (see *Figure 5*). The estimated parameter of age for the stage Y2 of Model 5 is .888, which indicated that with 1 standard deviation of increasing in standard age, the likelihood of choosing "Correct" instead of "Partial Correct" at the stage Y2 increased by .888 logits. The estimated parameter of working memory for the stage Y3 is .356. With 1 standard deviation increased in the standard scores of working memory, the likelihood of choosing "Duplicate" instead of "Other" at the stage Y3 increased by .356 logits. The stage Y1 was not related to any person covariate variable in this model, because it might associate with children's IQ levels or background information levels (Primi & Paulo, 2002; Siegler, 1999; Siegler & Svetina, 2002), which were not concerned by this research question.

#### *3.5.1.2 Model 6*

Model 6 is a binary tree structure IRTree model for the adjusted ordered response variable (see *Figure 6*). The estimated parameter of age for the stage Y2 of Model 6 is .919, which indicated that with 1 standard deviation of increasing in standard age, the likelihood of choosing "Correct" instead of "Partial Correct" at the stage Y2 increased by .919 logits. The estimated parameter of working memory for the stage Y3 is .293. With 1 standard deviation increased in the standard score of working memory, the likelihood of choosing "Other" instead of "Duplicate" at the stage Y3 increased by .293 logits. The stage Y1 was not related to any person covariate variable in this model, because it might associate with children's IQ levels or background information levels (Primi & Paulo, 2002; Siegler, 1999; Siegler & Svetina, 2002), which were not concerned by our current research question.

#### *3.5.1.3 Model 7*

Model 7 assumed that the category "Other" was the lowest-order category of children's analogical reasoning strategy, which was qualitatively different than the other three categories (see *Figure 7*). The estimated parameter of age for the stage Y1 is .599, which indicated that with 1 standard deviation of increasing in standard age, the likelihood of processing towards the stage Y2 instead of "Other" at the stage Y1 was increased by .599 logits. The estimated parameter of working memory for the stage Y2 is .352. This indicated that the with 1 standard deviation increased in the standard scores of working memory, the likelihood of choosing "Correct" instead of "Partial Correct" and "Duplicate" at the stage Y2 increased by .352 logits.

#### *3.5.1.4 Model 8*

Model 8 assumed that the category "Duplicate" was the lowest-order category of children's analogical reasoning strategy, which was qualitatively different than the other three categories (see *Figure 8*). The estimated parameter of age for the stage Y1 is .811, which indicated that with 1 standard deviation of increasing in standard age, the likelihood of processing towards the stage Y2 instead of "Duplicate" at the stage Y1 was increased by .811 logits. The estimated parameter of working memory for the stage Y2 is .429, which indicated with 1 standard deviation increased in the standard score of working memory, the likelihood of choosing "Correct" instead of "Partial Correct" and "Other" at the stage Y2 increased by .429 logits.

### 3.5.2 Model selection

Table 13.

*Model selection of four IRTree models including person covariates*

| Models | *AIC* | *BIC* | *Number of parameters* | *Log-likelihood* |
|---|---|---|---|---|
| Model 5 | 14606 | 14837 | 47 | -7256 |
| Model 6 | 14701 | 14931 | 47 | -7303 |
| Model 7 | 14875 | 15062 | 38 | -7400 |
| Model 8 | 14733 | 14919 | 38 | -7328 |

According to the values of model fit indices in *Table 13*, the Model 5 contained more parameters than the Model 7 and Model 8. The AIC and BIC values of Model 5 were lower than the Model 6. Therefore, Model 5 was the best fitting model for the dataset including person predictors age and working memory capacity.

### 3.5.3 Interpretation of the best fit model

For the second research question, the IRTree Model 5 with binary tree structure was the most appropriate model than the other three models. Age had influence on the stage Y2, which demonstrated the age-relate differences in the mapping stage among children who chose “Correct” and “Partial Correct” strategies. Children who used the “Correct” strategy were probably older than those children who used the “Partial Correct” strategy. The working memory capacity was related to the stage Y3 of application. The working memory capacity was distinguished between children who chose “Other” and “Duplicate”. Children who used “Duplicate” strategy might have larger working memory capacity to remember the analogical tasks and features, than children who used “Other” strategy.

## 4 Discussion

In the present study, the strategy children applied for solving the analogical reasoning tasks was classified as four categories. Two research questions were targeted. Firstly, which model was the best fit for the dataset of children's analogical reasoning strategy? Secondly, which model was the most appropriate one considering two common predictors of analogical reasoning ability, age and working memory capacity?

### 4.1 Effect of age and working memory capacity

The present study included the age and working memory capacity as person predictors as these have often been found to be related to analogical reasoning ability (Stevenson, Hickendorff, et al., 2013; Stevenson, Resing, et al., 2013). The results found that age was an important factor in the prediction of children's analogical reasoning skills. This finding was consistent with previous studies results (Cnossen, 2015; Tunteler & Resing, 2007a, 2007b). More specifically, the present study concluded that age was correlated with the stage of mapping among the children who used analogical reasoning strategy. According to Sternberg's component theory, the stage mapping indicated that the subject linked the already showed figures by discovering the relation between the features of the figures (Sternberg, 1977; Sternberg & Rifkin, 1979). Older children tended to use "Correct" strategy, while younger children were more likely to make mistakes during the mapping stage and use "Partial Correct" strategy.

In addition, working memory capacity was proved to be important in the prediction of children's analogical reasoning skills (Stevenson, Resing, et al., 2013; Swanson, 2008). The present study further explained the working memory capacity was specifically related to the application stage among the children who used non-analogical reasoning strategy, according to Sternberg's component theory (Sternberg, 1977; Sternberg & Rifkin, 1979). Children with less working memory capacity tended to use "Other" strategy since they might forgot the task content or the features of the already existed figures. In contrast. children with more working memory capacity were more likely to use "Duplicate" strategy since they remembered the features of the already existed figures.

### 4.2 Advantages

In general, the present study has multiple advantages for the research of children's analogical reasoning process.

The first advantage of current study is that the generalized IRTree models have been applied by using the package "FLIRT". The IRTree models gave insights of children's analogical reasoning process by comparing IRTree models with different tree structures. In addition, the IRTree models can be better interpreted by analogical reasoning theories than the traditional IRT models. Since the analogical reasoning theories included stages and components, the IRTree models can be explained by each node representing specific stage or component (Mulholland et al., 1980; Sternberg & Rifkin, 1979).

Secondly, the present study included both the original ordered response variable, and the adjusted response variable with reversed orders between categories "Duplicate" and "Other". The traditional IRT models and the IRTree models were conducted for the response variable in both orders. The traditional IRT models gave better performance for the adjusted ordered response variable, while the IRTree models performed no difference between the original ordered response variable and the adjusted ordered response variable. This indicated that the IRTree models were more sensitive and accurate than the traditional IRT models.

Another advantage is that, the model fit results of the IRTree models were improved, comparing with the results of the traditional IRT models in Cnossen's study for the same dataset (Cnossen, 2015). This indicated that the IRTree models are more suitable for the analysing polytomous response variable, comparing with the traditional IRT models.

### 4.3 Limitations

The present study comprised additional analyses on the children's analogical reasoning strategy dataset that was collected from 2009 to 2012. The present two research questions were formed after the data collection. Thus, several methodological limitations existed during the analysis of the IRTree models and the traditional IRT models.

First of all, the person covariate variable working memory capacity included missing data. Missing-data imputation was conducted by replacing the missing values into the mean of working memory variable. The method of imputation increased the risk of bias, although the working memory scores were normally distributed and the sample mean did not change after imputation.

The second limitation is that the item difficulty was not considered as a predictor in the present study. There were six transformation features for each figural analogical reasoning task, with combination of animal, size, quantity, colour, orientation and position. The item difficulty level increased when more features included in the task. The strategy was assumed to change regarding to different item difficulty level. In the present study, the item difficulty levels of the seven common items were fixed effects for all the sample children.

Third, the comparison between the IRTree models and the traditional IRT models seem not appropriate. Since the IRTree models were process models, which gave different results with the same dataset and different tree structures. While the traditional IRT models led to the same model fit results as long as the dataset was the same.

The final limitation is that there were only seven common items, which were answered by all the sample children from different schools. Therefore, linking the results was based on the seven common items. However, these seven items were appropriate as link items, because they represented figural analogical reasoning tasks with good reliability.

### 4.4 Methodological considerations

Methodologically, we started with a complex dataset, which contained missing data in multiple items. All the subjects responded seven common items. Therefore, these seven items were considered as anchored items, which used in the following analysis of IRT models. We were interested in finding a model structure to fit the dataset appropriately, and to be interpreted easily by the analogical reasoning theories. We argued that the IRTree models are appropriate for analysing the polytomous response variable. Specifically, the IRTree models can gain insights into the stages of children's analogical reasoning process.

When looking into the proportions of dataset, we realized that the probability of category "Duplicate" seemed to be stable among the seven common items, regardless of the changing of item difficulty and item discrimination levels. This category of strategy might be qualitatively different from the other three categories. The original ordered response variable and the adjusted ordered response variable with reversed orders between "Duplicate" and "Other" have been tested by the traditional IRT models and the IRTree models. Two traditional IRT models, Graded Response Model and Generalized Partial Credit Model, were included in present study. The findings were mixed according to different models. The traditional IRT models better fitted the adjusted ordered response variable, rather than the original ordered response variable. However, the IRTree models showed no difference of different ordered response variable, as long as “Duplicate” and “Other” belonged to a general category of non-analogical reasoning strategy.

Four IRTree models of two tree structures were applied for both the original ordered response variable, and the adjusted ordered response variable. All the approaches resulted in the IRTree model with binary tree structure as the best fitting model. The interpretation of the parameter estimates of the IRTree model was clear and reasonable according to the analogical reasoning theories. Therefore, we presented the results of the most appropriate model for the dataset.

### 4.5 Recommendations for future research

Firstly, the present study analysed two tree structures of IRTree models based on the previous analogical reasoning theories (Mulholland et al., 1980; Sternberg, 1977; Sternberg & Rifkin, 1979). Further research can conduct the IRTree models with more complex tree structures, according to other analogical reasoning theories. For instance, considering the Embretson’s cognitive component model, an interactive structural mapping component was added based on the Sternberg’s six component theory (Embretson & Schneider, 1989). An IRTree model with interactive processing structure can be formulated in the future.

Secondly, the individual difference and variability of analogical reasoning ability among children was not considered in the present study.  There were various reasons led to individual difference of analogical reasoning ability. For instance, children’s IQ levels, the reaction time for responding the analogical reasoning tasks, and the background knowledge of the analogical reasoning (Primi & Paulo, 2002; Swanson, 2008; Tunteler & Resing, 2007b). It is important to take these factors into account in the future analogical reasoning study.

Thirdly, model selection in the present study was based on the fit indices and the interpretation of parameters. However, the maximum likelihood could be used in the model selection in the future studies. The maximum likelihood estimated by using a modified expectation-maximization (EM) algorithm based on graphical model theory (Lauritzen, 1995; Rijmen, Vansteelandt, & De Boeck, 2008). The modified EM algorithm applies the expectation (E) step efficiently, so that computations can be conducted in lower dimensional latent spaces with higher speed than regular ML methods (Jeon et al., 2014).

Last but not the least, future studies can try different methods for missing data imputation. In the present study, we replaced the missing data in the working memory variable into means. The multiple imputations can be applied when data are missing at random. The Type I error and power of the imputed new data are comparable to the complete data when the random missing data is less than 40% of the whole dataset (Graham, 2009).

## 5 Conclusions

In the present study, two research questions have been answered. Two traditional IRT models have conducted for the different orders of the response variable categories. The result of the traditional IRT models showed that the Grade Response Model was better fit than the Generalized Partial Credit Model. The fit indices values of both models were improved when the "Duplicate" was the lowest-order category of the response variable, followed by "Other", "Partial Correct", and "Correct".

For the first research question, the IRTree models with binary tree structures were better fit than other IRTree models. It indicated that children's analogical reasoning process was binary structured with three stages, regardless of orders between categories "Other" and "Duplicate". According to the Sternberg's component theory, the first stage represented the encoding and inferences, the second stage represented mapping, and the last stage was application (Sternberg, 1977; Sternberg & Rifkin, 1979).

For the second research question, the binary structured IRTree model with "Other" as the lowest order category was the most appropriate model among the four IRTree models. It indicated that age was highly correlated to the mapping stage for children who chose analogical reasoning strategy, and working memory capacity was slightly related to the application stage for children who chose non-analogical reasoning strategy.

**References**

Akaike, M. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*, 716-723.

Alloway, T. P. (2007). *Automated working memory assessment manual*. London: Pearson Assessment.

Bleichrodt, N., Drenth, P. J. D., Zaal, J. N., & Resing, W. C. M. (1984). Revisie amsterdamse kinder intelligentie Test, RAKIT: Lisse: Swets & Zeitlinger.

Brown, A. L., & Kane, M. J. (1988). Preschool children can learn to transfer: Learning to learn and learning from example. *Cognitive Psychology, 20*(4), 493-523.

Chen, Z., Siegler, R. S., & Daehler, M. W. (2000). Across the great divide: Bridging the gap between understanding of toddlers' and older children's thinking. *Monographs of the Society for Research in Child Development, 65*(2), 1-96.

Cnossen, S. (2015). *Statistical models of children's strategy change in analogical reasoning.* (Master of Science Master Thesis), Leiden University, The Netherlands.

De Boeck, P., Bakker, M., Zwitser, R., Nivard, M., Hofman, A., Tuerlinckx, F., & Partchev, I. (2011). The estimation of item response models with the lmer function from the lme4 package in R. *Journal of Statistical Software, 39*(12), 1-28.

De Boeck, P., & Partchev, I. (2012). IRTrees: Tree-based item response models of the GLMM family. *Journal of Statistical Software, 48*, 1-28.

Embretson, S., & Reise, S. P. (2000). *Item Response Theory for Psychologists*. New York: Psychology Press.

Embretson, S., & Schneider, L. M. (1989). Cognitive component models for psychometric analogies: Conceptually driven versus interactive process models. *Learning and Individual Differences, 1*(2), 155-178.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science, 7*(2), 155-170.

Goswami, U., & Brown, A. L. (1990). Higher-order structure and relational reasoning: contrasting analogical and thematic relations. *Cognition, 36*(3), 207-226.

Goswami, U., & Brown, A. L. (1991). Analogical reasoning: What develops? A review of research and theory. *Child Development: Abstracts & Bibliography, 62*(1), 1-22.

Goswami, U., Leevers., H., Pressley, S., & Wheelwright, S. (1998). Causal reasoning about pairs of relations and analogical reasoning in young children. *British Journal of Developmental Psychology, 16*(4), 553-569.

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual review of psychology, 60*, 549-576.

Hemker, B. T., van der Ark, L. A., & Sijtsma, K. (2001). On Measurement Properties of Continuation Ratio Models. *Psychometrika, 66*(4), 487-506.

Hoskens, M., & De Boeck, P. (1995). Componential IRT models for polytomous items. *Journal of Educational Measurement, 32*(4), 364-384.

Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence*. New York: Basic Books.

Jeon, M., & De Boeck, P. (2015). A generalized item response tree model for psychological assessments. *Behavioral Research Methods,* 1-16.

Jeon, M., Rijmen, F., & Rabe-Hesketh, S. (2014). Flexible item response theory modeling with FLIRT. *Applied Psychological Measurement, 38*, 404-405.

Lauritzen, S. (1995). The EM algorithm for graphical association models with missing data. *Computational Statistics & Data Analysis, 19*, 191-201.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*(2), 149-174.

Matzen, L. B. L. V., van der Molen, M. W., & Dudink, A. C. M. (1994). Error analysis of raven test performance. *Personality and individual differenfces*, *16*(3), 433-445.

Mulholland, T. M., Pellegrino, J. W., & Glaser, R. (1980). Components of geometric analogy solution. *Cognitive Psychology, 12*(2), 252-284.

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement, 16*(2), 159-176.

Piaget, J. (1977). *The development of thought: Equilibration of cognitive structures. (Trans A. Rosin).* Oxford, England: Viking The development of thought: Equilibration of cognitive structures.

Primi, R., & Paulo, B. (2002). Complexity of geometric inductive reasoning tasks: Contribution to the understanding of fluid intelligence. *Intelligence, 30*(1), 41-70.

Richland, L. E., Morrison, R. G., & Holyoak, K. J. (2006). Children's development of analogical reasoning: insights from scene analogy problems. *Journal Experimemtal Children Psychology, 94*(3), 249-272.

Rijmen, F., Vansteelandt, K., & De Boeck, P. (2008). Latent class models for diary method data: Parameter estimation by local computations. *Psychometrika, 73,* 167-182.

Rizopoulos, D. (2006). An R package for Latent Variable Modelling and Item Response Theory Analyses. *Journal of Statistical Software, 17*(5), 1-25.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychmetrika Monograph, No. 17*.

Schwarz, G. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychmetrika Monograph, No. 17*.

Siegler, R. S. (1999). Strategic development. *Trends in Cognitive Sciences, 3*(11), 430-435.

Siegler, R. S., & Svetina, M. (2002). A microgenetic/cross-sectional study of matrix completion: comparing short-term and long-term change. *Child Development, 73*(3).

Singer-Freeman, K. E. (2005). Analogical reasoning in 2-year-olds: The development of access and relational inference. *Cognitive Development*, *20*(2).

Sternberg, R. J. (1977). Component processes in analogical reasoning. *Psychological Review, 84*(4).

Sternberg, R. J., & Rifkin, B. (1979). The Development of Analogical Reasoning Processes. *Journal of Experimental Child Psychology*, *27*(2).

Stevenson, C. E., Alberto, R. A., van den Boom, M. A., & De Boeck, P. A. (2014). Visual relations children find easy and difficult to process in figural analogies. *Frontiers in Psychology, 5*, 827.

Stevenson, C. E., Hickendorff, M., Resing, W. C. M., Heiser, W. J., & De Boeck, P. A. L. (2013). Explanatory item response modeling of children's change on a dynamic test of analogical reasoning. *Intelligence*, *41*(3).

Stevenson, C. E., Resing, W. C. M., & Heiser, W. J. (2013). Working memory as a moderator of training and transfer of analogical reasoning in children. *Contemporary Educational Psychology, 38*(3).

Swanson, H. L. (2008). Working Memory and Intelligence in Children: What Develops? *Journal of Educational Psychology, 100*(3).

Thibaut, J., French, R., & Vezneva, M. (2010). The development of analogy making in children: Cognitive load and executive functions. *Journal of Experimental Child Psychology, 106*.

Tunteler, E., Pronk, C. M. E., & Resing, W. C. M. (2008). Inter- and intra-individual variability in the process of change in the use of analogical strategies to solve geometric tasks in children: A microgenetic analysis. *Learning and Individual Differences, 18*(1).

Tunteler, E., & Resing, W. C. M. (2002). Spontaneous analogical transfer in 4-year-olds: A microgenetic study. *Journal Experimemtal Child Psychology, 83*.

Tunteler, E., & Resing, W. C. M. (2007a). Change in spontaneous analogical transfer in young children: A microgenetic study. *Infant and Child Development, 16*(1).

Tunteler, E., & Resing, W. C. M. (2007b). Effects of prior assistance in using analogies on young children's unprompted analogical problem solving over time: A microgenetic study. *British Journal of Educational Psychology, 77*.

van der Maas, H. L. J., Molenaar, D., Maris, G., & Kievit, R. A. (2011). Cognitive psychology meets psychometric theory: On the relation between process models for decision making and latent variable models for individual differences. *Psychological Review, 118*(2).

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children-WISC-IV*: Psychological Corporation.

Whitely, S. E., & Barnes, G. M. (1979). The implications of processing event sequences for theories of analogical reasoning. *Memory & Cognition, 7*(4).

**Appendix**

1.R-codes of GRM and g-PCM fitted on the common pretest items

1.1 Original ordered response variable

```
> summary(MD1)
  i1201_strat      i1204_strat      i1301_strat      i1404_strat
 Min.   :0.000   Min.   :0.000   Min.   :0.000   Min.   :0.000
 1st Qu.:1.000   1st Qu.:1.000   1st Qu.:1.000   1st Qu.:1.000
 Median :2.000   Median :2.000   Median :2.000   Median :2.000
 Mean   :2.005   Mean   :1.986   Mean   :1.981   Mean   :1.571
 3rd Qu.:3.000   3rd Qu.:3.000   3rd Qu.:3.000   3rd Qu.:2.000
 Max.   :3.000   Max.   :3.000   Max.   :3.000   Max.   :3.000
  i1502_strat      i1505_strat      i1604_strat
 Min.   :0.000   Min.   :0.000   Min.   :0.000
 1st Qu.:0.000   1st Qu.:0.000   1st Qu.:0.000
 Median :1.000   Median :1.000   Median :1.000
 Mean   :1.178   Mean   :1.178   Mean   :1.094
 3rd Qu.:2.000   3rd Qu.:2.000   3rd Qu.:2.000
 Max.   :3.000   Max.   :3.000   Max.   :3.000
```

1.2 Adjusted ordered response variable

```
> summary(MD2)
    re_1201          re_1204          re_1301          re_1404
 Min.   :0.000   Min.   :0.000   Min.   :0.000   Min.   :0.000
 1st Qu.:0.000   1st Qu.:0.000   1st Qu.:1.000   1st Qu.:1.000
 Median :2.000   Median :2.000   Median :2.000   Median :2.000
 Mean   :1.797   Mean   :1.671   Mean   :1.791   Mean   :1.534
 3rd Qu.:3.000   3rd Qu.:3.000   3rd Qu.:3.000   3rd Qu.:2.000
 Max.   :3.000   Max.   :3.000   Max.   :3.000   Max.   :3.000
    re_1502          re_1505          re_1604
 Min.   :0.000   Min.   :0.000   Min.   :0.000
 1st Qu.:1.000   1st Qu.:1.000   1st Qu.:1.000
 Median :1.000   Median :1.000   Median :1.000
 Mean   :1.301   Mean   :1.272   Mean   :1.273
 3rd Qu.:2.000   3rd Qu.:2.000   3rd Qu.:2.000
 Max.   :3.000   Max.   :3.000   Max.   :3.000
```

1.3 GRM1 on toe dataset original ordered response variable

```
> fit_grm1 <- grm(data=MD1,Hessian = TRUE)
> summary(fit_grm1)

Call:
grm(data = MD1, Hessian = TRUE)

Model Summary:
  log.Lik      AIC      BIC
 -7768.21 15592.42 15729.89

Coefficients:
$i1201_strat
         value std.err  z.vals
Extrmt1 -2.224   0.132 -16.846
Extrmt2 -0.738   0.119  -6.201
Extrmt3  0.369   0.096   3.837
Dscrmn   1.784   0.120  14.807

$i1204_strat
         value std.err  z.vals
Extrmt1 -2.762   0.184 -15.047
Extrmt2 -0.483   0.131  -3.677
Extrmt3  0.283   0.124   2.272
Dscrmn   1.511   0.108  14.011

$i1301_strat
```

```
        value std.err  z.vals
Extrmt1 -3.134   0.249 -12.571
Extrmt2 -1.015   0.159  -6.386
Extrmt3  0.711   0.141   5.024
Dscrmn   1.018   0.082  12.451

$i1404_strat
        value std.err  z.vals
Extrmt1 -1.153   0.067 -17.187
Extrmt2 -0.327   0.069  -4.775
Extrmt3  1.175   0.274   4.292
Dscrmn   2.036   0.131  15.589

$i1502_strat
        value std.err  z.vals
Extrmt1 -0.551   0.046 -11.942
Extrmt2  0.106   0.044   2.410
Extrmt3  1.475   1.991   0.741
Dscrmn   3.020   0.211  14.339

$i1505_strat
        value std.err  z.vals
Extrmt1 -0.603   0.052 -11.495
Extrmt2  0.180   0.048   3.723
Extrmt3  1.591   0.925   1.721
Dscrmn   2.204   0.138  15.949

$i1604_strat
        value std.err z.vals
Extrmt1 -0.493   0.055 -8.992
Extrmt2  0.230   0.057  4.038
Extrmt3  1.864   1.047  1.780
Dscrmn   1.892   0.119 15.862


Integration:
method: Gauss-Hermite
quadrature points: 21

Optimization:
Convergence: 0
max(|grad|): 0.017
quasi-Newton: BFGS

> coef(fit_grm1, IRTpars=TRUE)
            Extrmt1 Extrmt2 Extrmt3 Dscrmn
i1201_strat  -2.224  -0.738   0.369  1.784
i1204_strat  -2.762  -0.483   0.283  1.511
i1301_strat  -3.134  -1.015   0.711  1.018
i1404_strat  -1.153  -0.327   1.175  2.036
i1502_strat  -0.551   0.106   1.475  3.020
i1505_strat  -0.603   0.180   1.591  2.204
i1604_strat  -0.493   0.230   1.864  1.892
```

### 1.4 GRM2 on the dataset adjusted ordered response variable

```
> fit_grm2 <- grm(data=MD2,Hessian = TRUE)
> summary(fit_grm2)

Call:
grm(data = MD2, Hessian = TRUE)

Model Summary:
   log.Lik      AIC      BIC
 -7459.596 14975.19 15112.67

Coefficients:
$re_1201
```

```
         value std.err  z.vals
Extrmt1 -0.892   0.066 -13.466
Extrmt2 -0.695   0.081  -8.554
Extrmt3  0.412   0.057   7.177
Dscrmn   1.878   0.120  15.666

$re_1204
         value std.err z.vals
Extrmt1 -0.561   0.061 -9.161
Extrmt2 -0.431   0.076 -5.687
Extrmt3  0.292   0.068  4.273
Dscrmn   1.769   0.118 14.939

$re_1301
         value std.err  z.vals
Extrmt1 -1.256   0.108 -11.648
Extrmt2 -0.937   0.114  -8.183
Extrmt3  0.717   0.093   7.720
Dscrmn   1.097   0.083  13.260

$re_1404
         value std.err  z.vals
Extrmt1 -0.915   0.059 -15.461
Extrmt2 -0.246   0.060  -4.095
Extrmt3  1.113   0.329   3.385
Dscrmn   2.483   0.148  16.792

$re_1502
         value std.err  z.vals
Extrmt1 -0.894   0.056 -15.988
Extrmt2  0.186   0.044   4.256
Extrmt3  1.480   1.521   0.973
Dscrmn   2.874   0.178  16.142

$re_1505
         value std.err  z.vals
Extrmt1 -0.869   0.060 -14.579
Extrmt2  0.237   0.045   5.250
Extrmt3  1.548   0.973   1.591
Dscrmn   2.368   0.139  17.076

$re_1604
         value std.err  z.vals
Extrmt1 -0.966   0.058 -16.543
Extrmt2  0.306   0.043   7.065
Extrmt3  1.636   1.780   0.919
Dscrmn   2.674   0.161  16.567


Integration:
method: Gauss-Hermite
quadrature points: 21

Optimization:
Convergence: 0
max(|grad|): 0.044
quasi-Newton: BFGS

> coef(fit_grm2, IRTpars=TRUE)
        Extrmt1 Extrmt2 Extrmt3 Dscrmn
re_1201  -0.892  -0.695   0.412  1.878
re_1204  -0.561  -0.431   0.292  1.769
re_1301  -1.256  -0.937   0.717  1.097
re_1404  -0.915  -0.246   1.113  2.483
re_1502  -0.894   0.186   1.480  2.874
re_1505  -0.869   0.237   1.548  2.368
re_1604  -0.966   0.306   1.636  2.674
```

1.5 GPCM1 on the dataset original ordered response variable

```
> fit_gpcm1 <- gpcm(MD1, constraint = "gpcm")
> summary(fit_gpcm1)

Call:
gpcm(data = MD1, constraint = "gpcm")

Model Summary:
   log.Lik      AIC      BIC
 -7794.784 15645.57 15783.04

Coefficients:
$i1201_strat
         value std.err z.value
Catgr.1 -2.196   0.153 -14.378
Catgr.2 -0.575   0.077  -7.498
Catgr.3  0.149   0.075   1.975
Dscrmn   1.302   0.104  12.504

$i1204_strat
         value std.err z.value
Catgr.1 -3.226   0.255 -12.657
Catgr.2  0.204   0.117   1.751
Catgr.3 -0.471   0.117  -4.016
Dscrmn   0.965   0.079  12.266

$i1301_strat
         value std.err z.value
Catgr.1 -2.849   0.271 -10.524
Catgr.2 -0.802   0.132  -6.101
Catgr.3  0.259   0.120   2.151
Dscrmn   0.687   0.061  11.205

$i1404_strat
         value std.err z.value
Catgr.1 -0.846   0.080 -10.625
Catgr.2 -0.520   0.073  -7.104
Catgr.3  1.134   0.085  13.329
Dscrmn   1.470   0.115  12.756

$i1502_strat
         value std.err z.value
Catgr.1 -0.342   0.061  -5.622
Catgr.2 -0.009   0.061  -0.142
Catgr.3  1.519   0.082  18.425
Dscrmn   2.168   0.186  11.651

$i1505_strat
         value std.err z.value
Catgr.1 -0.287   0.074  -3.854
Catgr.2  0.012   0.074   0.169
Catgr.3  1.609   0.102  15.725
Dscrmn   1.473   0.114  12.882

$i1604_strat
         value std.err z.value
Catgr.1  0.084   0.096   0.874
Catgr.2 -0.117   0.091  -1.289
Catgr.3  1.972   0.132  14.976
Dscrmn   1.189   0.091  13.122


Integration:
method: Gauss-Hermite
quadrature points: 21

Optimization:
Convergence: 0
max(|grad|): 0.041
optimizer: nlminb
```

```
> coef(fit_gpcm1, IRTpars=TRUE)
            Catgr.1 Catgr.2 Catgr.3 Dscrmn
i1201_strat  -2.196  -0.575   0.149  1.302
i1204_strat  -3.226   0.204  -0.471  0.965
i1301_strat  -2.849  -0.802   0.259  0.687
i1404_strat  -0.846  -0.520   1.134  1.470
i1502_strat  -0.342  -0.009   1.519  2.168
i1505_strat  -0.287   0.012   1.609  1.473
i1604_strat   0.084  -0.117   1.972  1.189
```

1.6 GPCM2 on the dataset adjusted ordered response variable

```
> fit_gpcm2 <- gpcm(MD2, constraint = "gpcm")
> summary(fit_gpcm2)

Call:
gpcm(data = MD2, constraint = "gpcm")

Model Summary:
   log.Lik      AIC      BIC
 -7534.779 15125.56 15263.03

Coefficients:
$re_1201
         value std.err z.value
Catgr.1  0.980   0.220   4.447
Catgr.2 -2.084   0.214  -9.738
Catgr.3  0.176   0.095   1.862
Dscrmn   0.953   0.075  12.638

$re_1204
         value std.err z.value
Catgr.1  2.272   0.324   7.020
Catgr.2 -2.282   0.282  -8.091
Catgr.3 -0.503   0.138  -3.650
Dscrmn   0.804   0.063  12.699

$re_1301
         value std.err z.value
Catgr.1  2.252   0.389   5.783
Catgr.2 -3.590   0.403  -8.911
Catgr.3  0.311   0.153   2.029
Dscrmn   0.522   0.046  11.434

$re_1404
         value std.err z.value
Catgr.1 -0.632   0.072  -8.722
Catgr.2 -0.465   0.071  -6.528
Catgr.3  1.096   0.073  15.094
Dscrmn   1.848   0.136  13.555

$re_1502
         value std.err z.value
Catgr.1 -0.837   0.061 -13.714
Catgr.2  0.185   0.056   3.328
Catgr.3  1.474   0.077  19.265
Dscrmn   2.322   0.174  13.339

$re_1505
         value std.err z.value
Catgr.1 -0.771   0.067 -11.469
Catgr.2  0.229   0.063   3.631
Catgr.3  1.514   0.087  17.401
Dscrmn   1.804   0.130  13.827

$re_1604
         value std.err z.value
Catgr.1 -0.927   0.063 -14.730
Catgr.2  0.277   0.056   4.933
```

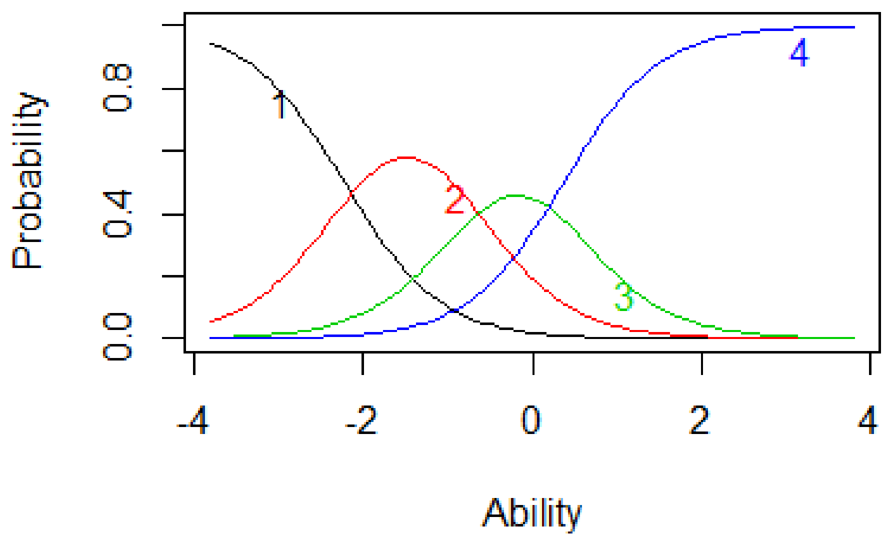
```
Catgr.3  1.632    0.083  19.575
Dscrmn   2.268    0.168  13.496


Integration:
method: Gauss-Hermite
quadrature points: 21

Optimization:
Convergence: 0
max(|grad|): 0.018
optimizer: nlminb
```

```
> coef(fit_gpcm2, IRTpars=TRUE)
        Catgr.1 Catgr.2 Catgr.3 Dscrmn
re_1201   0.980  -2.084   0.176  0.953
re_1204   2.272  -2.282  -0.503  0.804
re_1301   2.252  -3.590   0.311  0.522
re_1404  -0.632  -0.465   1.096  1.848
re_1502  -0.837   0.185   1.474  2.322
re_1505  -0.771   0.229   1.514  1.804
re_1604  -0.927   0.277   1.632  2.268
```

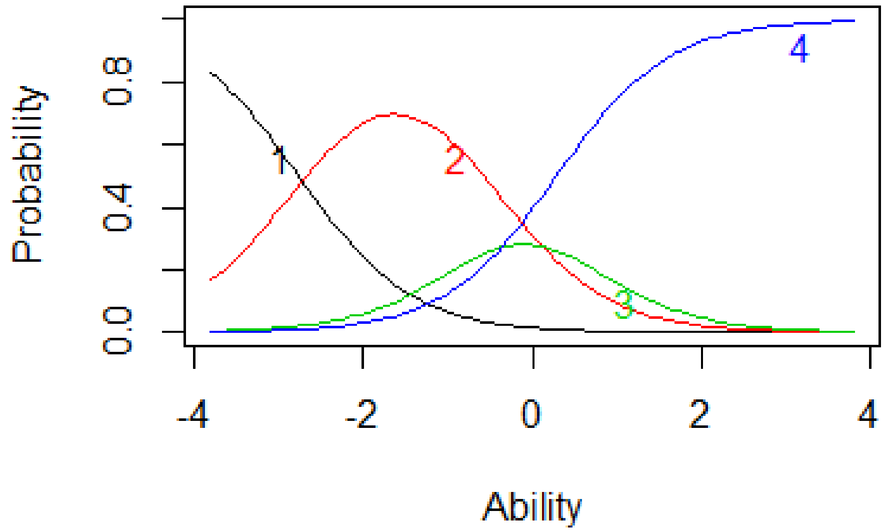
2. Item plots

2.1 Item plots of GRM1

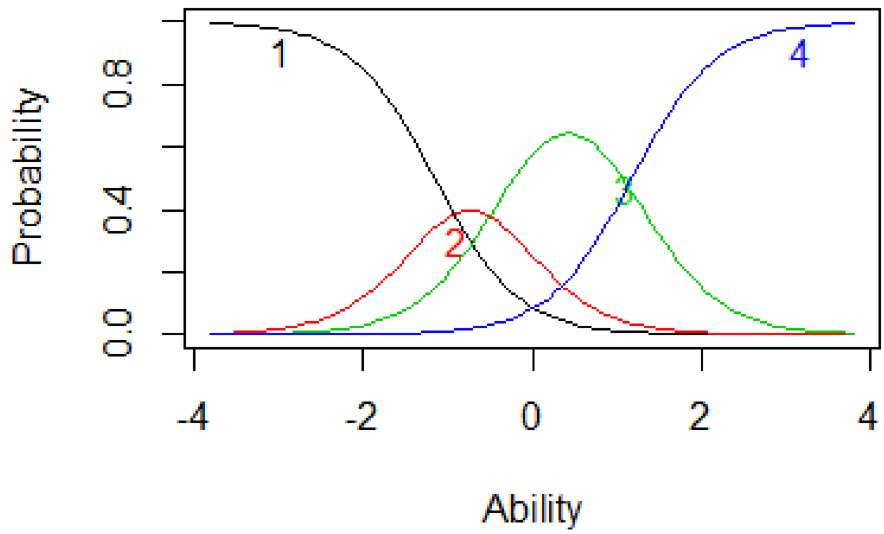
```
> plot(fit_grm1) # ICCS

# item 201
```



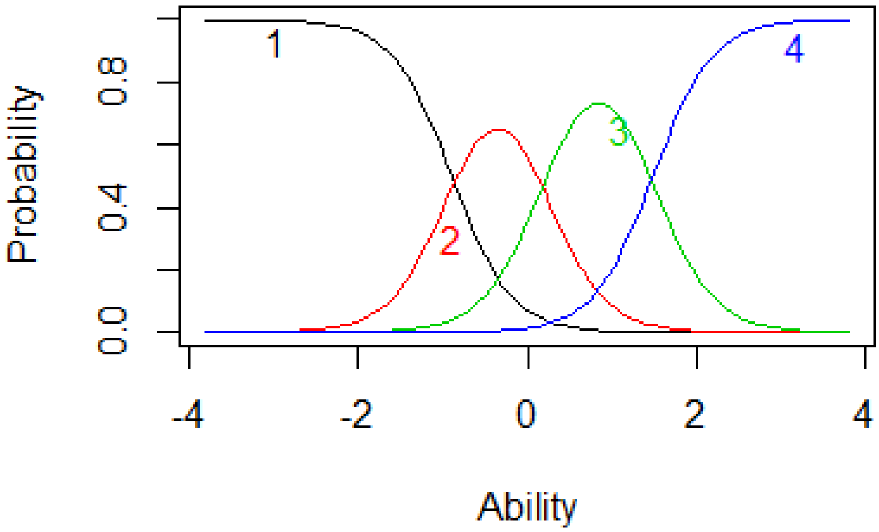
```
# item 204
```



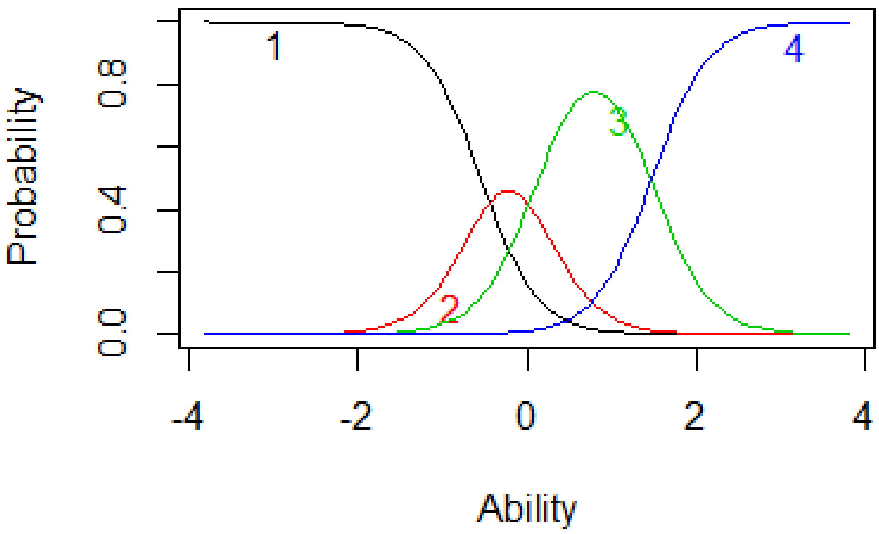

```
# item 301
```



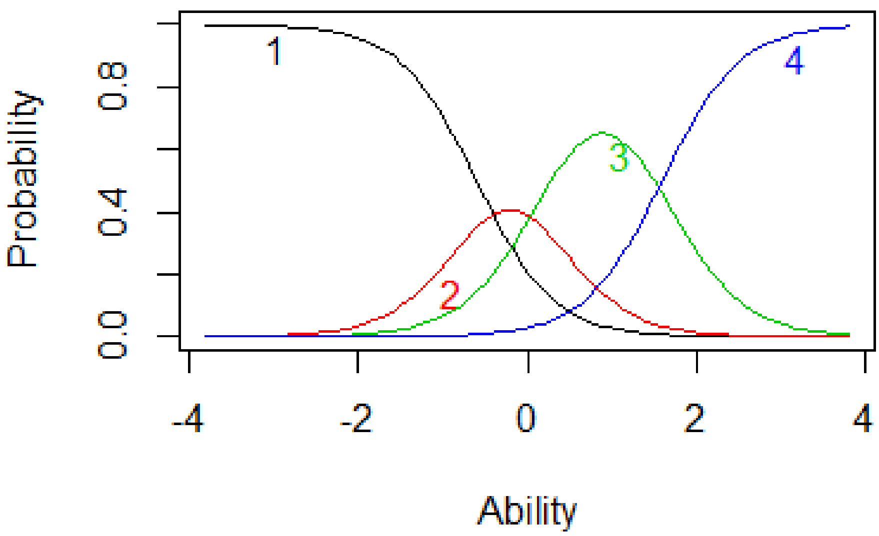
```
# item 404
```



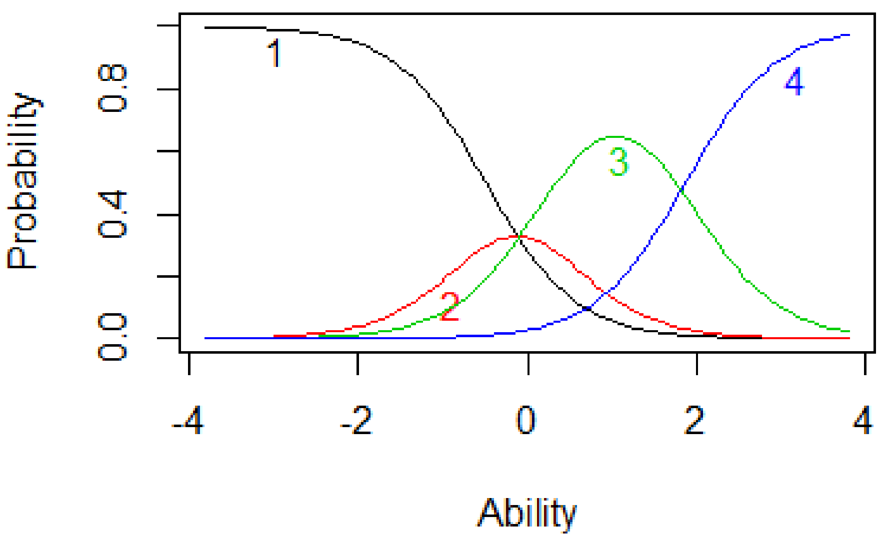

```
# item 502
```



```
# item 505
```



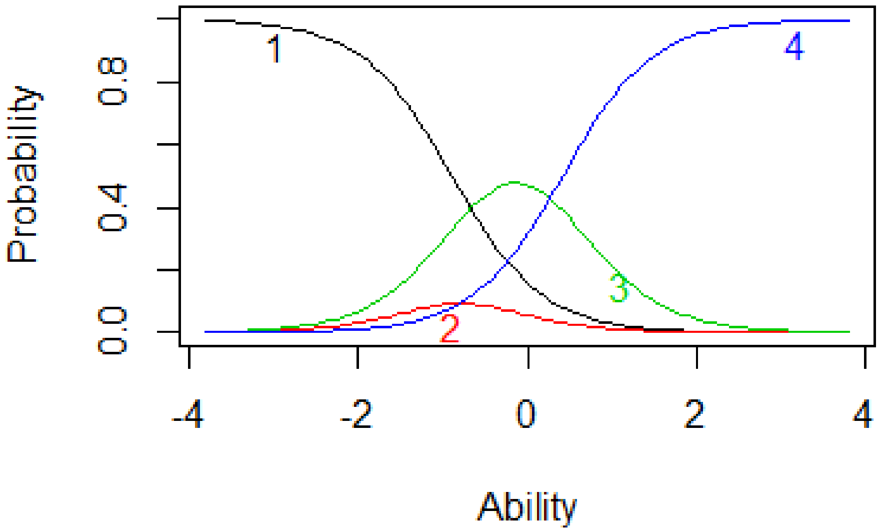

```
# item 604
```



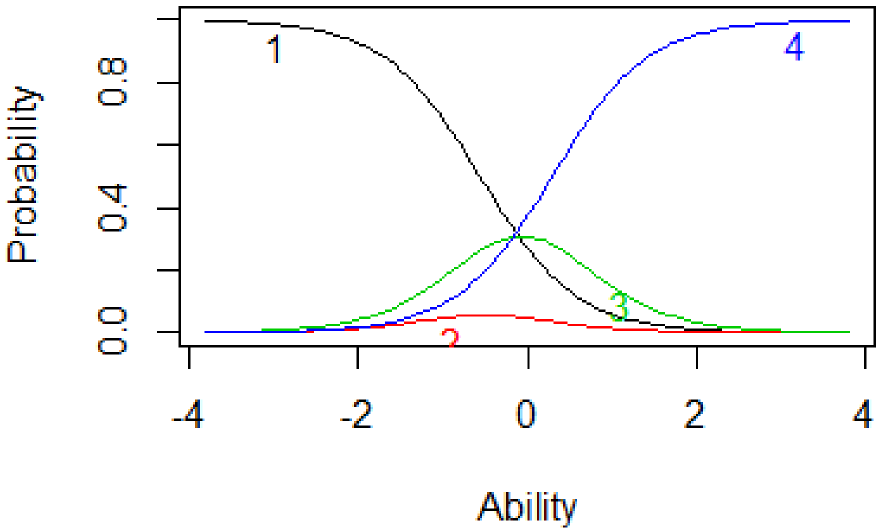

2.2 Item plots of GRM2

```
> plot(fit_grm2) # ICCS
```

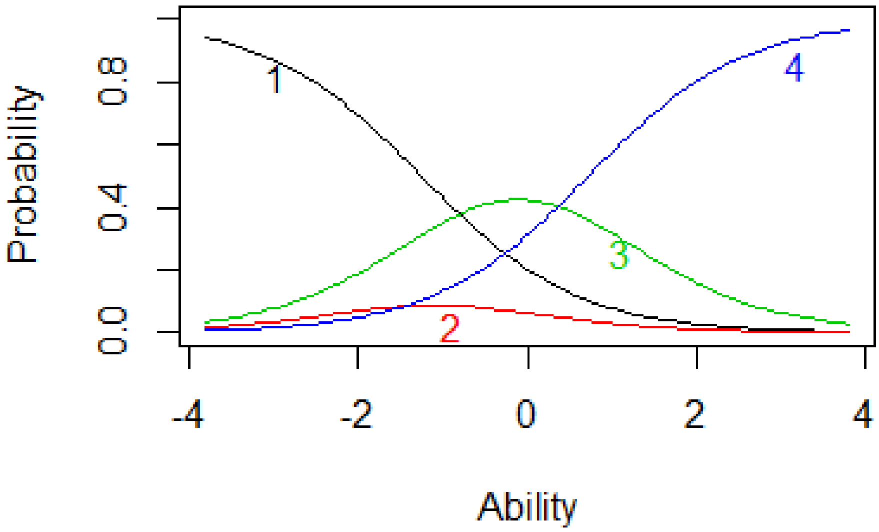
```
# item 201
```



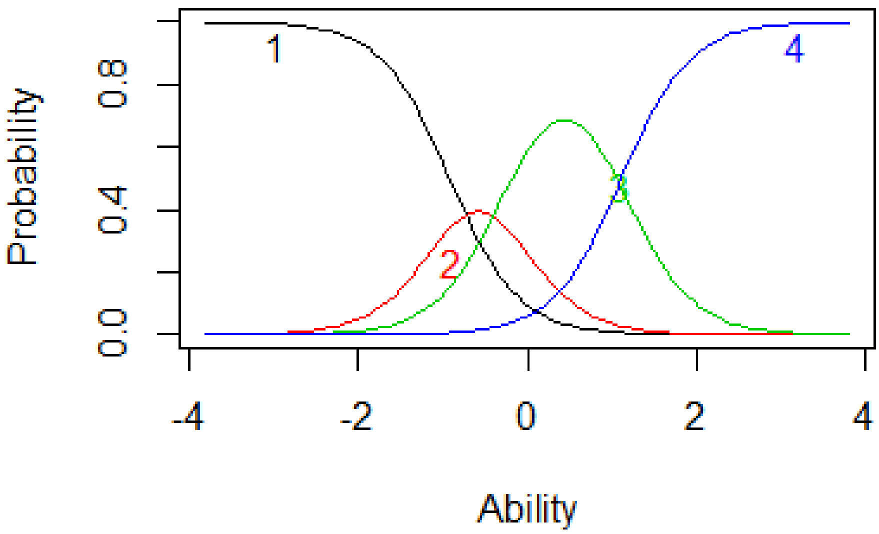

```
# item 204
```



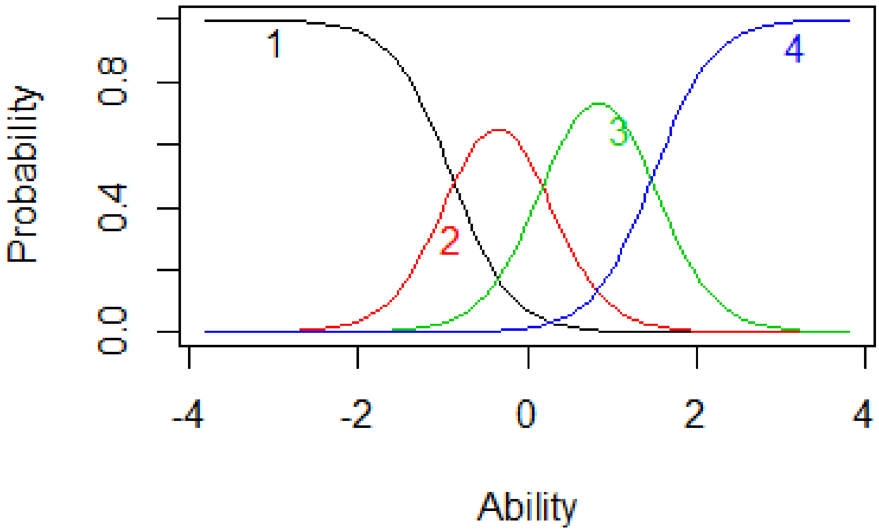
```
# item 301
```



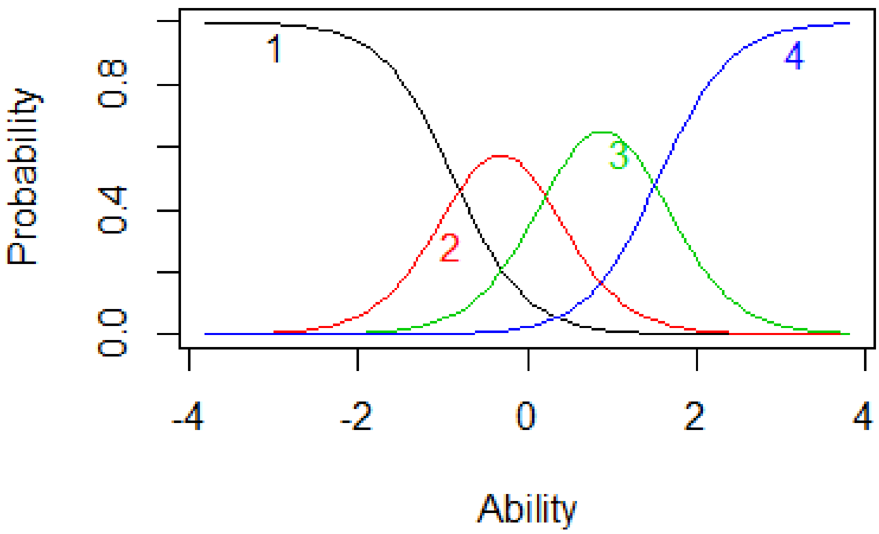

```
# item 404
```



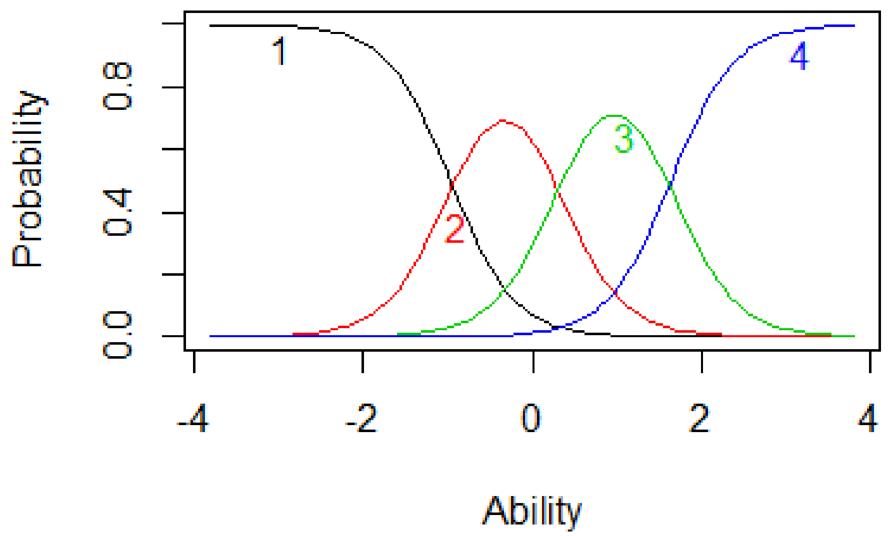
```
# item 502
```




```
# item 505
```

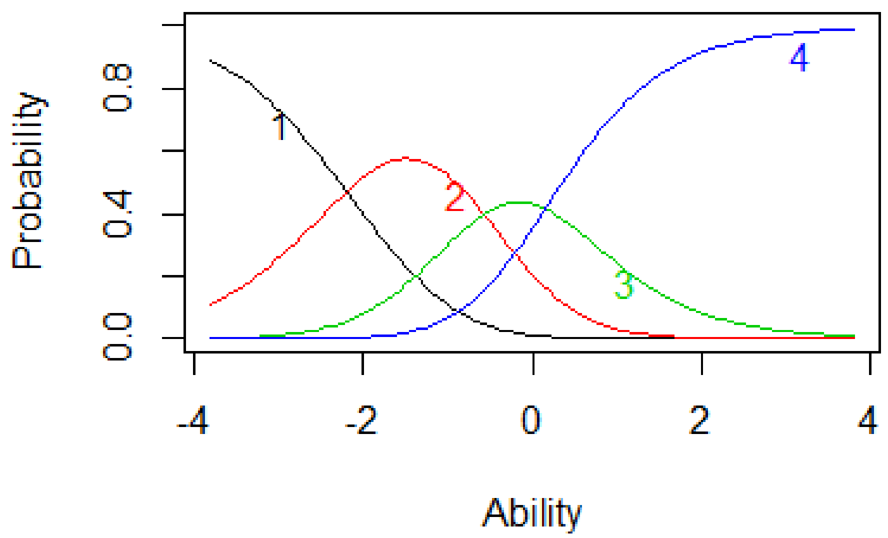


```
# item 604
```

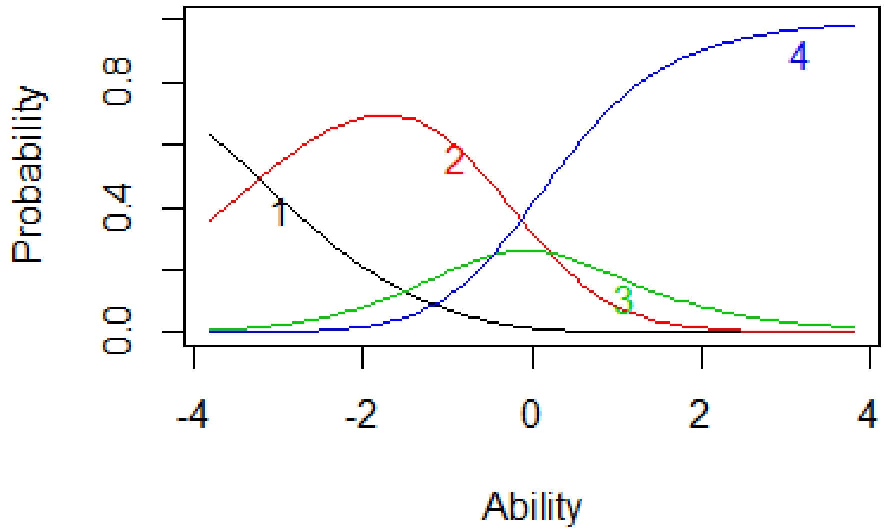



2.3 Item plots of GPCM1

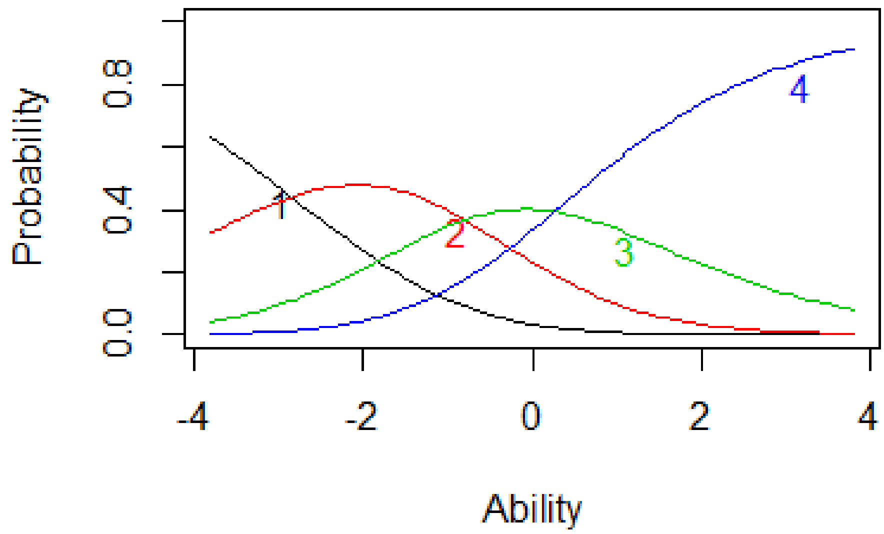
```
> plot(fit_gpcm1) #ICCs
# item 201
```



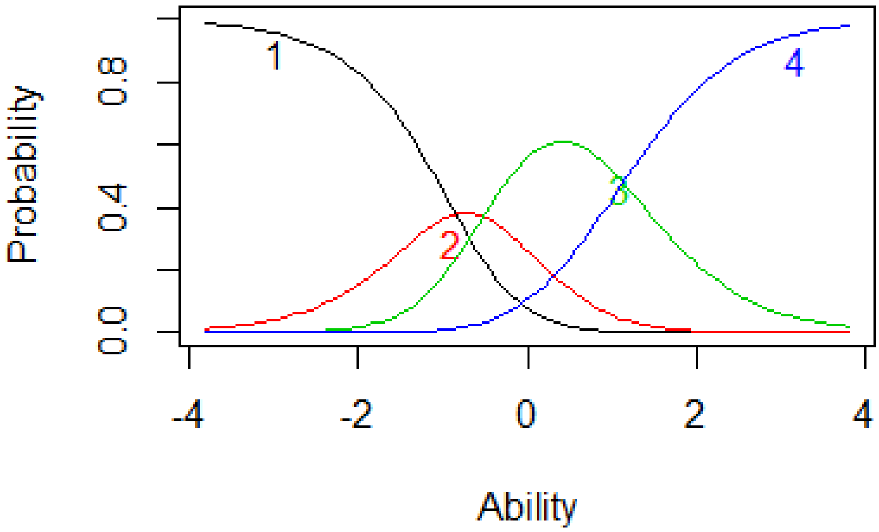
```
# item 204
```



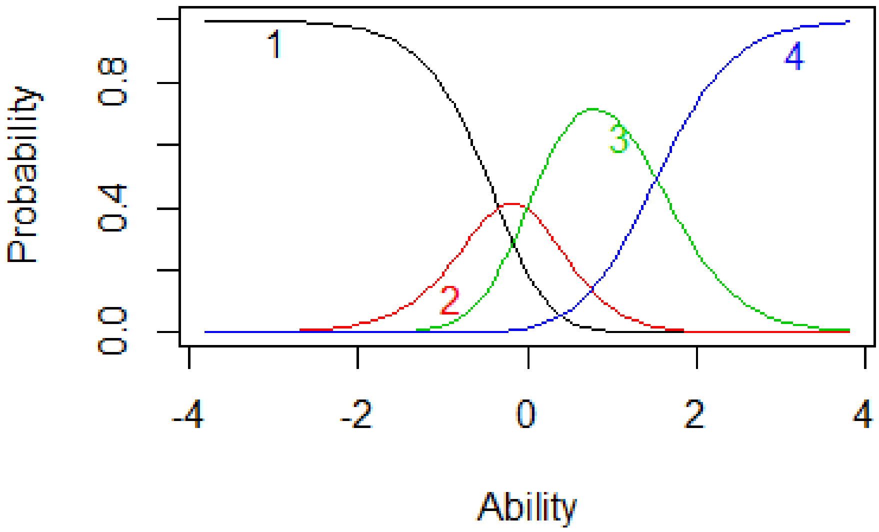

```
# item 301
```



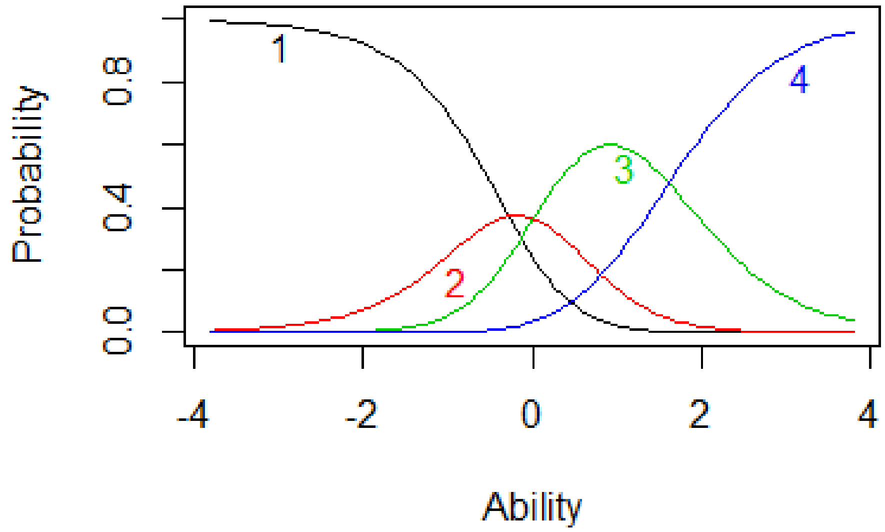
```
# item 404
```



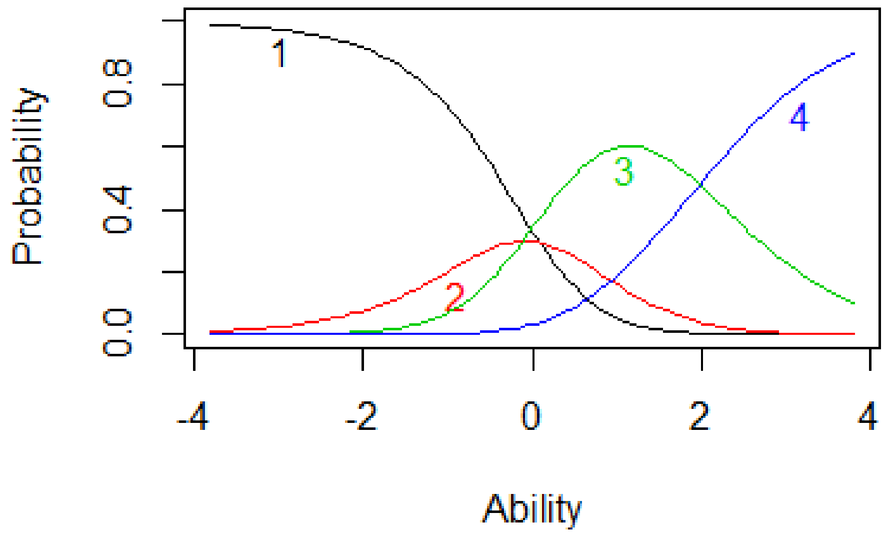

```
# item 502
```



```
# item 505
```



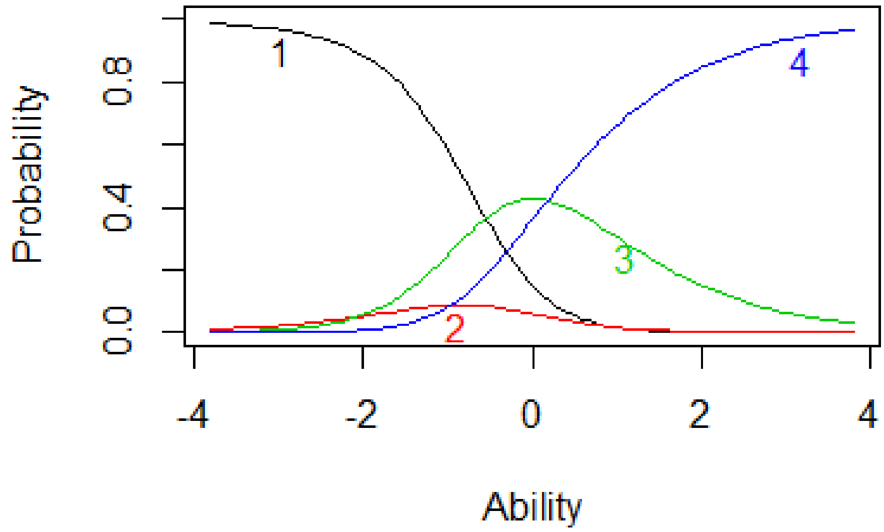

```
# item 604
```



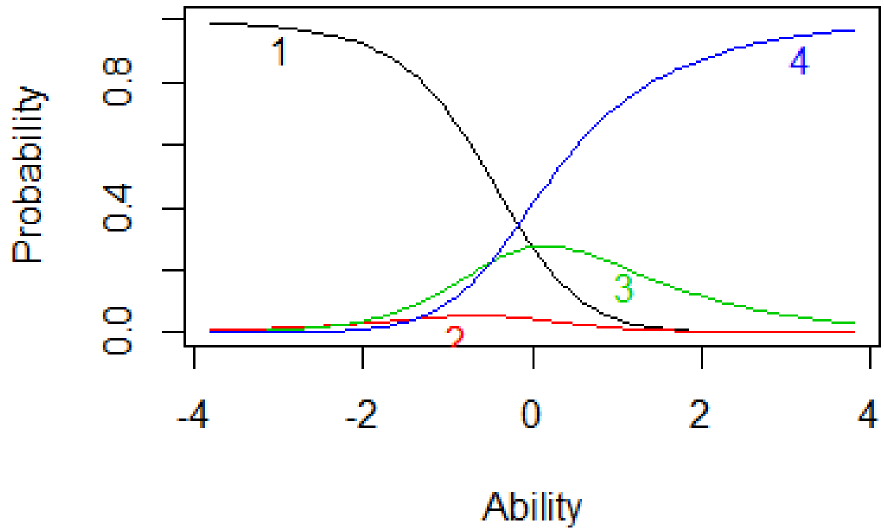

### 2.4 Item plots of GPCM2

```
> plot(fit_gpcm2) #ICCs
```

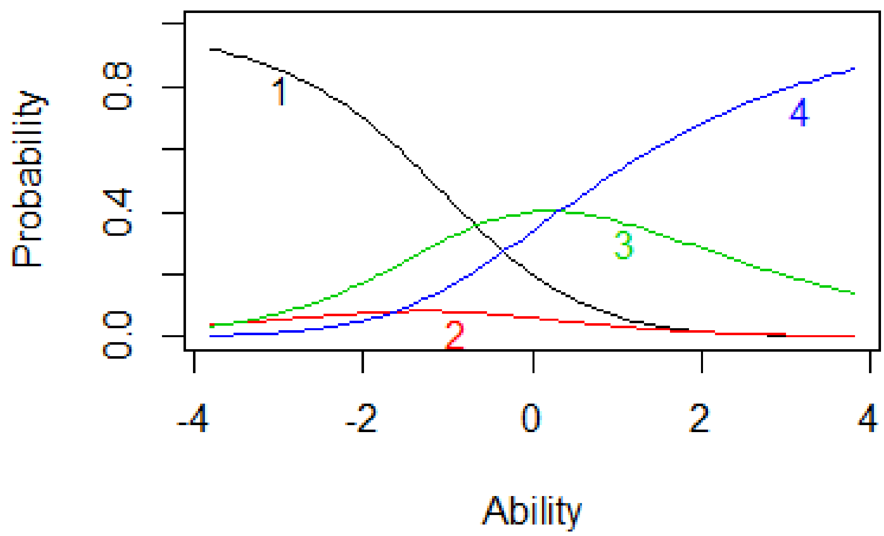
```
# item 201
```



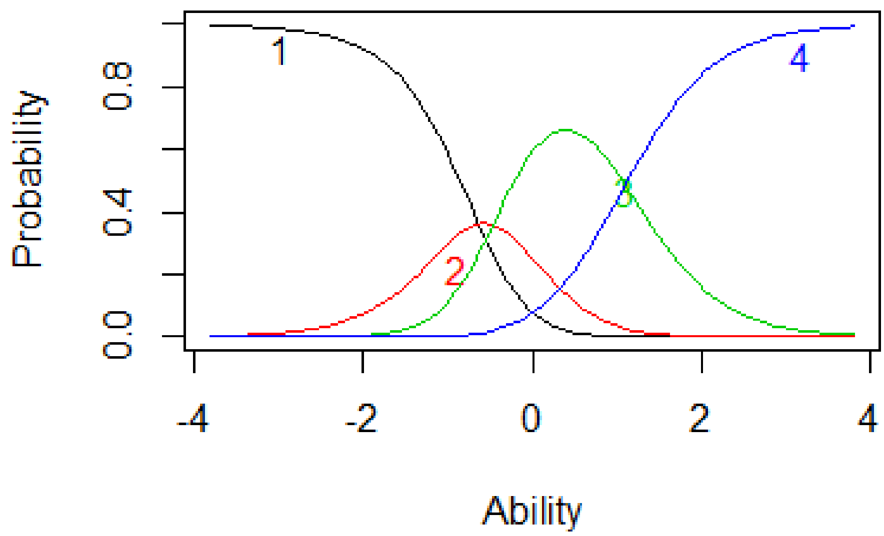

```
# item 204
```



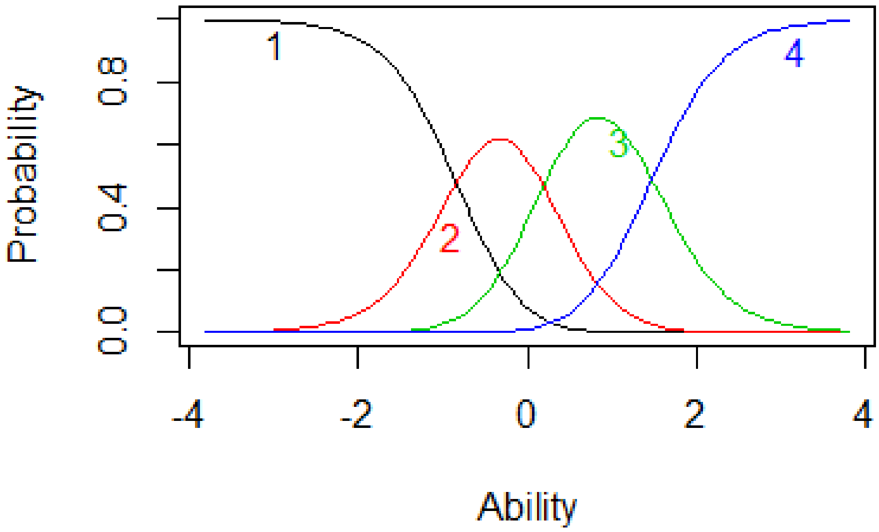
```
# item 301
```



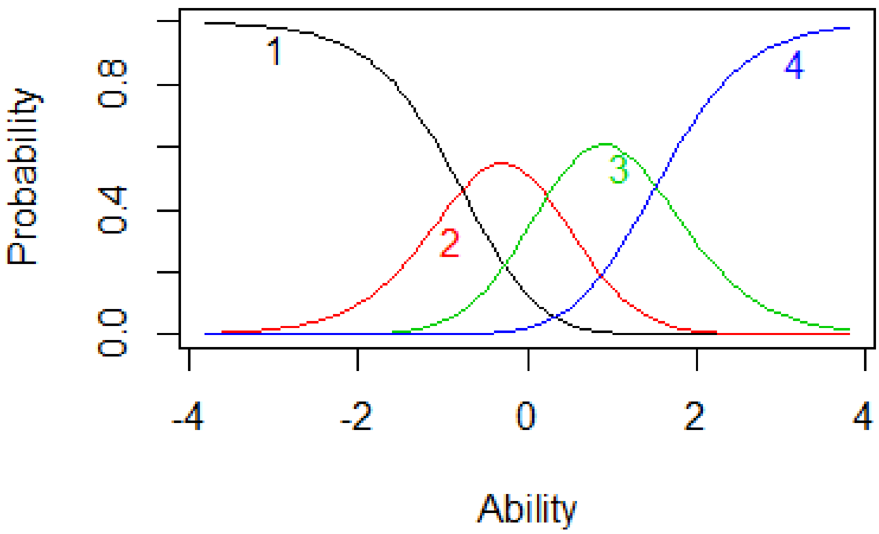

```
# item 404
```



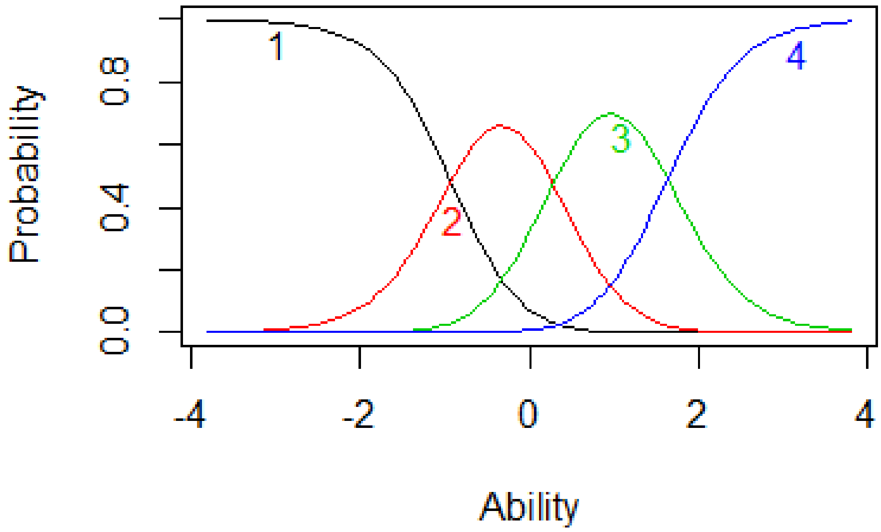
```
# item 502
```




```
# item 505
```



```
# item 604
```




3. R codes of the IRTree models

3.1 Model 1 with original ordered response variable

```
> mapping1 <- cbind(c(0, 0, 1, 1), c(NA, NA, 0, 1), c(0, 1, NA, NA))
> wide1 <- dendrify2(MD1, mapping1, wide=T)
> model1 <- flirt(data=wide1[,-1], loading=list(on=T, inside=F),
+                   mul=list(on=T, dim_info=list(dim1=1:7, dim2=8:14, dim3=15
:21)),
+                   control=list(nq=5, link = "adjacent", show=T) )
> summary(model1)

Estimation of Multidimensional 2PL Model Family

Data:
  nobs  nitem maxcat ngroup
  1002     21      2      1

Model fit:
  npar    AIC    BIC loglik
    45  14639  14860  -7275

Parameterization:

"a*th+b"

Type:

"between-item"

Dimension:
ndim dim1 dim2 dim3
   3    7    7    7

Parameter estimates:
               Est     SE
alp1       1.34561 0.1138
alp2       1.59444 0.1364
alp3       1.00802 0.0967
```

```
alp4       3.30945 0.5860
alp5       3.97216 0.6467
alp6       2.67812 0.2344
alp7       2.53754 0.1945
alp8       1.38409 0.2307
alp9       0.82206 0.1309
alp10      0.55469 0.0916
alp11      0.82489 0.1249
alp12      1.37142 0.2700
alp13      1.04023 0.1962
alp14      0.60411 0.1269
alp15      2.20022 0.4769
alp16      1.68380 0.3574
alp17      0.78211 0.1855
alp18      2.07182 0.3307
alp19      2.06965 0.2805
alp20      2.30491 0.3368
alp21      2.82375 0.5749
bet1       0.93787 0.0900
bet2       0.51629 0.0885
bet3       0.90519 0.0818
bet4       0.33010 0.1084
bet5      -1.01650 0.1206
bet6      -0.99155 0.1135
bet7      -1.07556 0.1146
bet8      -0.70735 0.1627
bet9       0.21870 0.1230
bet10     -0.38272 0.1009
bet11     -1.97293 0.2117
bet12     -4.08284 0.7194
bet13     -3.15885 0.4457
bet14     -2.44333 0.3199
bet15      2.14363 0.3478
bet16      2.90927 0.3661
bet17      1.41547 0.1661
bet18     -0.41545 0.1916
bet19     -1.20829 0.1930
bet20     -1.13101 0.2079
bet21     -1.84821 0.3292
th_mean11  0.00000     NA
th_mean21  0.00000     NA
th_mean31  0.00000     NA
th_sd11    1.00000     NA
th_sd21    1.94596     NA
th_sd31    1.16907     NA
th_cov11   1.66936     NA
th_cov21  -0.50422     NA
th_cov31  -0.50635     NA
```

```
> est_alp1 <- model1@pars[1:21,1]  # vector of alpha estimates
> est_cov1 <- model1@pars[46:51,1]  # vector of covariance matrix estimates
> cov_matrix1 <- matrix(c( est_cov1[1]^2, est_cov1[4], est_cov1[5],
est_cov1[4], est_cov1[2]^2, est_cov1[6], est_cov1[5], est_cov1[6],
est_cov1[3]^2), 3, 3, byrow=F)
> dim_info1 <- list(dim1=1:7, dim2=8:14, dim3=15:21) # list
>
> test1 <- std_coef(est = est_alp1, dim_info = dim_info1, cov_matrix =
cov_matrix1)
> # correlation
> test1$cor_mat
```

```
           [,1]       [,2]       [,3]
[1,]  1.0000000  0.8578594 -0.4313036
[2,]  0.8578594  1.0000000 -0.2225747
[3,] -0.4313036 -0.2225747  1.0000000
```

3.2 Model 2 with adjusted ordered response variable

```
> mapping2 <- cbind(c(0, 0, 1, 1), c(NA, NA, 0, 1), c(0, 1, NA, NA))
> wide2 <- dendrify2(MD2, mapping2, wide=T)
> model2 <- flirt(data=wide2[,-1], loading=list(on=T, inside=F),
+                 mul=list(on=T, dim_info=list(dim1=1:7, dim2=8:14, dim3=15
:21)),
+                 control=list(nq=5, link = "adjacent", show=T) )
> summary(model2)

Estimation of Multidimensional 2PL Model Family

Data:
  nobs  nitem maxcat ngroup
  1002     21      2      1

Model fit:
  npar    AIC    BIC loglik
    45  14639  14860  -7275

Parameterization:

"a*th+b"

Type:

"between-item"

Dimension:
ndim dim1 dim2 dim3
   3    7    7    7

Parameter estimates:
               Est     SE
alp1       1.34561 0.1138
alp2       1.59444 0.1364
alp3       1.00802 0.0967
alp4       3.30945 0.5860
alp5       3.97215 0.6466
alp6       2.67812 0.2344
alp7       2.53754 0.1945
alp8       1.38407 0.2307
alp9       0.82205 0.1309
alp10      0.55468 0.0916
alp11      0.82489 0.1249
alp12      1.37142 0.2700
alp13      1.04022 0.1962
alp14      0.60410 0.1269
alp15      2.20021 0.4769
alp16      1.68380 0.3574
alp17      0.78210 0.1855
alp18      2.07181 0.3307
alp19      2.06964 0.2805
alp20      2.30491 0.3368
alp21      2.82371 0.5749
bet1       0.93787 0.0900
bet2       0.51629 0.0885
bet3       0.90519 0.0818
bet4       0.33010 0.1084
bet5      -1.01650 0.1206
bet6      -0.99155 0.1135
bet7      -1.07556 0.1146
bet8      -0.70735 0.1627
bet9       0.21870 0.1230
bet10     -0.38272 0.1009
bet11     -1.97294 0.2117
bet12     -4.08287 0.7194
bet13     -3.15886 0.4457
bet14     -2.44334 0.3199
bet15     -2.14362 0.3478
bet16     -2.90927 0.3661
```

```
bet17     -1.41547 0.1661
bet18      0.41545 0.1916
bet19      1.20830 0.1930
bet20      1.13102 0.2079
bet21      1.84820 0.3292
th_mean11  0.00000     NA
th_mean21  0.00000     NA
th_mean31  0.00000     NA
th_sd11    1.00000     NA
th_sd21    1.94598     NA
th_sd31    1.16907     NA
th_cov11   1.66939     NA
th_cov21   0.50423     NA
th_cov31   0.50635     NA

> est_alp2 <- model2@pars[1:21,1]  # vector of alpha estimates
> est_cov2 <- model2@pars[46:51,1]  # vector of covariance matrix estimates
> cov_matrix1 <- matrix(c( est_cov2[1]^2, est_cov2[4], est_cov2[5],
est_cov2[4], est_cov2[2]^2, est_cov2[6], est_cov2[5], est_cov2[6],
est_cov2[3]^2), 3, 3, byrow=F)
> dim_info2 <- list(dim1=1:7, dim2=8:14, dim3=15:21) # list
>
> test2 <- std_coef(est = est_alp2, dim_info = dim_info2, cov_matrix =
cov_matrix2)
> # correlation
> test2$cor_mat

          [,1]      [,2]      [,3]
[1,] 1.0000000 0.8578625 0.4313039
[2,] 0.8578625 1.0000000 0.2225734
[3,] 0.4313039 0.2225734 1.0000000
```

3.3 Model 3 with original ordered response variable

```
> mapping3 <- as.matrix(cbind(c(0, 1, 1, 1), c(NA, 1, 2, 3)))
> wide3 <- dendrify2(MD1, mapping3, wide=T)
> # multidimensional model
> #(dimension 1: binary data, dimension 2: ordinal data with graded
response model)
> model3 <- flirt(data=wide3[,-1], loading=list(on=T, inside=F),
+                 mul=list(on=T, dim_info=list(dim1=1:7, dim2=8:14 )),
+                 control=list(nq=5, link="adjacent"))
> summary(model3)

Estimation of Multidimensional 2PL Model Family

Data:
  nobs  nitem maxcat ngroup
  1002     14      4      1

Model fit:
  npar    AIC    BIC loglik
    36  14913  15090  -7420

Parameterization:

"a*th+b"

Type:

"between-item"

Dimension:
ndim dim1 dim2
   2    7    7

Parameter estimates:
                 Est     SE
```

```
alp1         2.3361389 0.3690
alp2         2.3821070 0.4873
alp3         1.2798415 0.1918
alp4         1.6952656 0.1655
alp5         2.3870479 0.2069
alp6         1.9329429 0.1663
alp7         1.8519254 0.1756
alp8         1.2910832 0.1045
alp9         1.0649074 0.0905
alp10        0.6481923 0.0606
alp11        2.0046050 0.1687
alp12        2.7777131 0.2789
alp13        2.0123157 0.1815
alp14        2.3306146 0.2100
bet1         4.9832220 0.5148
bet2         5.5998159 0.7071
bet3         3.5850411 0.2474
bet4         2.2874276 0.1762
bet5         1.3016353 0.1442
bet6         1.2150679 0.1263
bet7         0.8404831 0.1096
bet8         0.9818735 0.1340
bet8        -0.2649985 0.1090
bet9        -0.0061781 0.1173
bet9         0.3520943 0.1087
bet10        0.6255855 0.0985
bet10       -0.1854649 0.0867
bet11        1.4360911 0.1639
bet11       -2.2774912 0.2046
bet12        0.7774223 0.1740
bet12       -4.2295208 0.4204
bet13        0.5212557 0.1442
bet13       -3.1350825 0.2808
bet14        0.9792567 0.1763
bet14       -3.9104238 0.3364
th_mean11    0.0000000     NA
th_mean21    0.0000000     NA
th_sd11      1.0000000     NA
th_sd21      1.1275192     NA
th_cov11     0.5208642     NA
```

```
> est_cov3 <- model3@pars[,1] # vector of covariance matrix estimates
> cov_matrix3 <- matrix(c( est_cov3[38]^2, est_cov3[40], est_cov3[40],
est_cov3[39]^2), 2, 2, byrow=F)
> dim_info3 <- list(dim1=1:7, dim2=8:14 )
> # correlation
> std_cov(cov_matrix3, dim_info = dim_info3)
```

```
          [,1]      [,2]
[1,] 1.0000000 0.4619559
[2,] 0.4619559 1.0000000
```

3.4 Model 4 with adjusted ordered response variable

```
> mapping4 <- as.matrix(cbind(c(0, 1, 1, 1), c(NA, 1, 2, 3)))
> wide4 <- dendrify2(MD2, mapping4, wide=T)
> model4 <- flirt(data=wide4[,-1], loading=list(on=T, inside=F),
+                 mul=list(on=T, dim_info=list(dim1=1:7, dim2=8:14 )),
+                 control=list(nq=5, link="adjacent"))
> summary(model4)
```

```
Estimation of Multidimensional 2PL Model Family

Data:
  nobs  nitem maxcat ngroup
  1002     14      4      1

Model fit:
```

```
  npar    AIC    BIC loglik
    36  14692  14869  -7310

Parameterization:

"a*th+b"

Type:

"between-item"

Dimension:
ndim dim1 dim2
   2    7    7

Parameter estimates:
               Est     SE
alp1       1.49109 0.1261
alp2       1.66574 0.1521
alp3       1.11699 0.1131
alp4       2.50793 0.1929
alp5       2.00838 0.1616
alp6       2.03763 0.1638
alp7       2.66757 0.2036
alp8       1.70844 0.1729
alp9       1.00286 0.1134
alp10      0.86810 0.0957
alp11      1.52412 0.1388
alp12      2.52196 0.2773
alp13      1.70370 0.1613
alp14      1.56332 0.1428
bet1       1.32270 0.1058
bet2       0.75072 0.0945
bet3       1.28369 0.0940
bet4       2.01393 0.1634
bet5       1.79706 0.1329
bet6       1.62626 0.1280
bet7       2.31777 0.1904
bet8       3.46342 0.2816
bet8      -0.47310 0.1376
bet9       2.51934 0.2454
bet9       0.35934 0.1119
bet10      2.38040 0.1854
bet10     -0.32961 0.0986
bet11      1.14381 0.1462
bet11     -2.19734 0.1975
bet12     -0.35309 0.1607
bet12     -4.63627 0.4350
bet13     -0.26799 0.1307
bet13     -3.24840 0.2835
bet14     -0.58117 0.1240
bet14     -3.43368 0.2785
th_mean11  0.00000     NA
th_mean21  0.00000     NA
th_sd11    1.00000     NA
th_sd21    1.27645     NA
th_cov11   0.79330     NA

> est_cov4 <- model4@pars[,1] # vector of covariance matrix estimates
> cov_matrix4 <- matrix(c( est_cov4[38]^2, est_cov4[40], est_cov4[40],
est_cov4[39]^2), 2, 2, byrow=F)
> dim_info4 <- list(dim1=1:7, dim2=8:14 )
> # correlation
> std_cov(cov_matrix4, dim_info = dim_info4)

          [,1]      [,2]
[1,] 1.0000000 0.6214913
[2,] 0.6214913 1.0000000
```

4. IRTree models for dataset including age and working memory capacity

4.1 Recode the person variables age and working memory capacity as matrix

```
> age <- as.numeric(MD[,36])  #standardized age, with 2 NAs
> memo <- as.numeric(MD[,37])  #standardized working memory scores, with
265 NAs
> #recode mising values of person covariates
> age[is.na(age)] <- mean(age,na.rm=TRUE)
> memo[is.na(memo)] <- mean(memo,na.rm=TRUE)

> #person variables matrix
> person_mat <- cbind(age, memo) #first column of person_mat is age, second
column of person_mat is working memory scores.

> person_mat <- as.data.frame(person_mat)
> summary(person_mat)
      age                memo
 Min.   :-1.5141   Min.   :-2.61358
 1st Qu.:-0.8646   1st Qu.:-0.58574
 Median :-0.1610   Median :-0.30197
 Mean   : 0.0000   Mean   :-0.30197
 3rd Qu.: 0.6509   3rd Qu.:-0.06828
 Max.   : 2.6536   Max.   : 2.63441
```

4.2 Model 5 with original ordered response variable

```
> mapping5 <- cbind(c(0, 0, 1, 1), c(NA, NA, 0, 1), c(0, 1, NA, NA))
> wide5 <- dendrify2(MD1, mapping5, wide=T)
> model5 <- flirt(data=wide5[,-1], loading=list(on=T, inside=F),
+                 person_cov=list(on=T, person_matrix=person_mat),
+                 mul=list(on=T, dim_info=list(dim1=1:7, dim2=8:14,
dim3=15:21),
+                          cov_info=list(dim1=0,dim2=1,dim3=2)),
#first column of person_mat for node 2, second column of person_mat for
node 3.
+                 post = TRUE, # the EAP estimates
+                 control=list(nq=2, link = "adjacent", show=T,se_num=F,
se_emp=F) )
> summary(model5)

Estimation of Multidimensional 2PL Model Family

Data:
  nobs  nitem maxcat ngroup
  1002     21      2      1

Model fit:
  npar    AIC    BIC loglik
    47  14606  14837  -7256

Parameterization:

"a*th+b"

Type:

"between-item"

Dimension:
ndim dim1 dim2 dim3
   3    7    7    7

Parameter estimates:
              Est
alp1       1.2306
alp2       1.4162
alp3       0.7242
alp4       1.5503
```

```
alp5        1.3054
alp6        1.3334
alp7        1.2727
alp8        0.8091
alp9        0.3282
alp10       0.6095
alp11       1.0483
alp12       1.1154
alp13       1.0771
alp14       0.9337
alp15       2.1210
alp16       1.6026
alp17       0.8337
alp18       1.9065
alp19       1.9001
alp20       1.8677
alp21       2.3124
bet1        1.3640
bet2        0.9352
bet3        1.2264
bet4        0.7101
bet5       -0.3921
bet6       -0.4748
bet7       -0.5465
bet8        0.0150
bet9        0.6970
bet10      -0.0893
bet11      -1.7590
bet12      -2.6513
bet13      -2.5208
bet14      -2.6905
bet15       3.1010
bet16       3.6113
bet17       1.9963
bet18       0.4422
bet19      -0.3994
bet20      -0.2705
bet21      -0.9003
gam_age1    0.8881
gam_memo2   0.3562
th_mean11   0.0000
th_mean21   0.0000
th_mean31   0.0000
th_sd11     1.0000
th_sd21     1.3418
th_sd31     1.2207
th_cov11    0.8947
th_cov21   -0.1925
th_cov31    0.5007
```

4.2 Model 6 with adjusted ordered response variable

```
> mapping6 <- cbind(c(0, 0, 1, 1), c(NA, NA, 0, 1), c(0, 1, NA, NA))
> wide6 <- dendrify2(MD2, mapping6, wide=T)
> model6 <- flirt(data=wide6[,-1], loading=list(on=T, inside=F),
+                 person_cov=list(on=T, person_matrix=person_mat),
+                 mul=list(on=T, dim_info=list(dim1=1:7, dim2=8:14,
dim3=15:21),
+                          cov_info=list(dim1=0,dim2=1,dim3=2)),
#first column of person_mat for node2, second column of person_mat for
node3.
+                 post = TRUE, # the EAP estimates
+                 control=list(nq=2, link = "adjacent", show=T,se_num=F,
se_emp=F) )
> summary(model6)

Estimation of Multidimensional 2PL Model Family

Data:
```

```
  nobs  nitem maxcat ngroup
  1002     21      2      1

Model fit:
  npar    AIC    BIC loglik
    47  14701  14931  -7303

Parameterization:

"a*th+b"

Type:

"between-item"

Dimension:
ndim dim1 dim2 dim3
   3    7    7    7

Parameter estimates:
                Est
alp1         1.2867
alp2         1.4304
alp3         0.8594
alp4         1.4599
alp5         1.1910
alp6         1.3066
alp7         1.1706
alp8         0.6740
alp9         0.2707
alp10        0.4172
alp11        0.9497
alp12        1.0309
alp13        1.0023
alp14        0.9474
alp15        1.5224
alp16        1.3100
alp17        0.9521
alp18        1.3111
alp19        1.1966
alp20        1.6894
alp21        1.2979
bet1         1.3557
bet2         0.9117
bet3         1.2423
bet4         0.6654
bet5        -0.3639
bet6        -0.4607
bet7        -0.5130
bet8        -0.0989
bet9         0.6526
bet10       -0.1322
bet11       -1.9635
bet12       -2.8962
bet13       -2.7623
bet14       -3.0179
bet15       -1.9285
bet16       -2.7600
bet17       -1.2494
bet18        0.7590
bet19        1.5081
bet20        1.6543
bet21        1.8559
gam_age1     0.9196
gam_memo2    0.2934
th_mean11    0.0000
th_mean21    0.0000
th_mean31    0.0000
th_sd11      1.0000
th_sd21      1.5440
```

```
th_sd31    1.8894
th_cov11   1.1764
th_cov21   0.6403
th_cov31  -0.7164
```

4.3 Model 7 with original ordered response variable

```
> mapping7 <- as.matrix(cbind(c(0, 1, 1, 1), c(NA, 1, 2, 3)))
> wide7 <- dendrify2(MD1, mapping7, wide=T)
> #(dimension 1: binary data, dimension 2: ordinal data with graded
response model)
> model7 <- flirt(data=wide7[,-1], loading=list(on=T, inside=F),
+                 person_cov=list(on=T, person_matrix=person_mat),
+                 mul=list(on=T, dim_info=list(dim1=1:7, dim2=8:14 ),
+                          cov_info=list(dim1=1,dim2=2) ), #first column of
person_mat for node 1, second column of person_mat for node 2.
+                 post = TRUE, # the EAP estimates
+                 control=list(nq=2, link="adjacent",se_num=F, se_emp=F))
> summary(model7)

Estimation of Multidimensional 2PL Model Family

Data:
  nobs  nitem maxcat ngroup
  1002     14      4      1

Model fit:
  npar    AIC    BIC loglik
    38  14875  15062  -7400

Parameterization:

"a*th+b"

Type:

"between-item"

Dimension:
ndim dim1 dim2
   2    7    7

Parameter estimates:
              Est
alp1       1.5179
alp2       1.6127
alp3       0.5302
alp4       1.3474
alp5       1.6249
alp6       1.4669
alp7       1.2067
alp8       0.9523
alp9       0.8896
alp10      0.4931
alp11      1.5728
alp12      1.6285
alp13      1.6722
alp14      1.9307
bet1       4.0410
bet2       4.5594
bet3       3.2191
bet4       2.1930
bet5       1.1604
bet6       1.1484
bet7       0.7278
bet8       0.7509
bet8      -0.3836
bet9      -0.0765
bet9       0.1806
```

```
bet10      0.7117
bet10     -0.2882
bet11      0.9602
bet11     -2.3467
bet12      0.2390
bet12     -3.1514
bet13      0.1262
bet13     -3.1415
bet14      0.5207
bet14     -3.7426
gam_age1   0.5989
gam_memo2  0.3522
th_mean11  0.0000
th_mean21  0.0000
th_sd11    1.0000
th_sd21    1.0688
th_cov11   0.3774
```

4.4 Model 8 with adjusted ordered response variable

```
> mapping8 <- as.matrix(cbind(c(0, 1, 1, 1), c(NA, 1, 2, 3)))
> wide8 <- dendrify2(MD2, mapping8, wide=T)
> #(dimension 1: binary data, dimension 2: ordinal data with graded
response model)
> model8 <- flirt(data=wide8[,-1], loading=list(on=T, inside=F),
+                  person_cov=list(on=T, person_matrix=person_mat),
+                  mul=list(on=T, dim_info=list(dim1=1:7, dim2=8:14 ),
+                           cov_info=list(dim1=1,dim2=2)), #first column of
person_mat for node 1, second column of person_mat for node 2.
+                  post = TRUE, # the EAP estimates
+                  control=list(nq=2, link="adjacent",se_num=F, se_emp=F))
> summary(model8)

Estimation of Multidimensional 2PL Model Family

Data:
  nobs  nitem maxcat ngroup
  1002     14      4      1

Model fit:
  npar    AIC    BIC loglik
    38  14733  14919  -7328

Parameterization:

"a*th+b"

Type:

"between-item"

Dimension:
ndim dim1 dim2
   2    7    7

Parameter estimates:
              Est
alp1       1.2826
alp2       1.2057
alp3       1.0017
alp4       1.6047
alp5       1.4145
alp6       1.7834
alp7       1.8036
alp8       0.9368
alp9       0.6129
alp10      0.6410
alp11      1.3681
alp12      1.2474
```

```
alp13      1.2139
alp14      1.0870
bet1       1.4414
bet2       0.8168
bet3       1.4557
bet4       1.8136
bet5       1.7355
bet6       1.7587
bet7       2.0925
bet8       2.8402
bet8      -0.1159
bet9       2.6081
bet9       0.5892
bet10      2.6230
bet10     -0.1854
bet11      1.3726
bet11     -2.1023
bet12     -0.0945
bet12     -2.7541
bet13     -0.0713
bet13     -2.6571
bet14     -0.3472
bet14     -2.8319
gam_age1   0.8110
gam_memo2  0.4294
th_mean11  0.0000
th_mean21  0.0000
th_sd11    1.0000
th_sd21    1.1187
th_cov11   0.5016
```