DIPLOMA THESIS

Title of the diploma thesis

“About the psychometric quality of various multiple choice response formats in the context of cultural differences between Austria and the United States of America.”

Author

Brigitte Christine Hansmann

Degree in progress

“Magistra der Naturwissenschaften” (Mag[a]. rer. nat.)

Vienna, February 2010

Registration number of study field/ Studienkennzahl lt. Studienblatt:	A 298
student identification number/Matrikelnummer:	0200535
Field of Study/ Studienrichtung lt. Studienblatt:	Psychology
Supervisor:	Univ. -Prof. Mag. Dr. Klaus D. Kubinger

„Diploma theses are not made to make students happy, but to graduate from University."

Klaus D. Kubinger (2009)

"This is my attempt to write an acceptable thesis out of difficult circumstances – not just to graduate but also to look back in proudness and good memories."

Brigitte C. Hansmann (2010)

Acknowledgments

First of all I want to thank Univ.-Prof. Dr. Mag. Klaus D. Kubinger who introduced me to this topic and who awakened my interest not only for this topic but also for research in the field of psychological assessment.

I also want to thank Mag. Herbert Poinstingl and Mr. Gottfried Berndl for their work on this project. Mag. Herbert Poinstingl who programmed the computer test battery and Mr. Berndl who helped me to upload the data on the computer were a great help. Thank you.

A special thank you, concerning the motivational and emotional aspects of my thesis are dedicated to my parents DI Robert M. Hansmann and Christine M. Hansmann and also to my boyfriend Michael G. Maschler who supported me trough rough times. Thank you.

A big thank you shall be given to the American test assessors Dr. Eduard Dunbar, Mr. Ignacio Allegre MSc and Mrs. Nicole Magdaleno who made a data collection in the United States of America possible. Thank you for your generous help.

Next, I want to thank all the instructors who allowed me to advertise my thesis in their seminars. I know that your teaching time is short and precious. So, thank you very much.

Concerning the "goodies" which the students got in this study, I want to thank Mr. Andree Henneberg, manager of the discotheque "Praterdome" in Vienna, for supporting me with free V.I.P. entry tickets and free drinking vouchers for participants of this study.

Last but not least I want to thank my native speaking friends all over the world who read over my thesis with great patience und who taught me very friendly that "high school English" is far away of "scientific English". In this cohesion a special thank you has to be dedicated to Mag[a]. Carrie Kovacs who read over the English version of the verbal test battery.

Brigitte Hansmann

Abstract (English)

One problem within multiple choice tests is that even if a test taker does not know the answer of an item he or she can solve it, simply through lucky guessing. The aim of this study was not only to investigate if there are different guessing effects between the multiple choice format “1 out of 5” and the sequential response format “1 out of 5” but also if there are cultural differences in this cohesion between Austria and the USA. To investigate this topic 225 Psychology students of the University of Vienna, Austria and 104 Psychology students at the Cypress College, California executed a verbal computer test battery, which consisted out of the FRRT (Family Relations Reasoning Test) with 60 items and the Syllogisms 2009 with 10 items. Unfortunately it turned out that about 50 % of the American data had to be excluded of further analyses due to an extreme answering behaviour (editing times under one minute combined with low raw scores) of the test takers. Because of this fact, no Rasch model analyses of the English items had been feasible and therefore no test of equivalency *via* Rasch model between the two languages was possible. So, the focus was only put on the Austrian sample which showed significant results: It turned out that the items are significantly easier if they are in the multiple choice format “1 out of 5” – which means that they show a significant higher guessing probability – than the same items but in the sequential response format “1 out of 5”.

Keywords: multiple choice “1 out of 5”, sequential response format, guessing effect, Rasch model

Abstract (German)

Problematisch bei Multiple Choice Tests ist, dass selbst eine Testperson, die keine Fähigkeiten besitzt ein bestimmtes Item zu lösen, dieses durch simples Raten doch lösen kann. Ziel dieser Studie war es nun, unterschiedliche Rateeffekte in den beiden Formaten Multiple-Choice „1 aus 5“ und dem sequentiellen Antwortformat „1 aus 5“ zu eruieren, wobei ein zusätzlicher Fokus auch auf Kulturunterschiede zwischen Österreich und den USA gerichtet wurde. Dazu wurde eine verbale Computer-Testbatterie, die aus den beiden Untertests FRRT (Family Relations Reasoning Test), bestehend aus 60 Items und den Syllogismen 2009, bestehend aus 10 Items, 225 deutschsprachigen Psychologie Studierenden an der Universität Wien und 104 englischsprachigen Psychologie Studierenden am Cypress College, California vorgegeben. Das Itemformat jedes Untertests variierte je Satz zufällig zwischen den beiden Antwortformaten. Es stellte sich heraus, dass ca. die Hälfte der Daten aus der amerikanischen Stichprobe für weitere Analysen ausgeschlossen werden musste, da die Studierenden ein auffälliges Antwortverhalten (Testbearbeitungszeiten unter einer Minute mit gleichzeitig niedrigen Rohscores) zeigten. Aus der daraus folgenden geringen Teilnehmerzahl in der amerikanischen Stichprobe war eine Analyse der englischen Items mit dem Rasch Modell nicht mehr möglich. Aus diesem Grund konnte keine Äquivalenzprüfung mittels Rasch Modell durchgeführt werden. Für die deutschsprachige Stichprobe konnte gezeigt werden, dass jene mit dem Rasch Modell konformen Items im Multiple Choice Antwortformat signifikant leichter waren, als dieselben Items im sequentiellen Antwortformat. Somit wiesen die Items im Multiple Choice Format einen signifikant höheren Rateeffekt auf, als im sequentiellen Format.

Schlüsselwörter: Multiple-Choice „1 aus 5“, Sequentielles Antwortformat, Rateeffekt, Rasch Modell

Bibliography

1 Introduction 11

2 Scientific aim of this study 14

3 Method 16

3.1 Method of Measurement 16

3.2 The Family Relations Reasoning Test 19

3.2.1 Item construction rules of the FRT 21

3.2.2 Item construction rules of the FRRT 24

3.3 Syllogisms 2009 28

3.4 Multiple choice formats 29

3.5 Parallel groups 31

3.6 Sample 33

3.7 Research questions and study hypotheses 36

3.8 Psychometric analyses 38

3.8.1 Rasch model analyses 38

3.8.2 Welch-Test 40

3.8.3 General important facts about the results 42

4 Results 44

4.1 Important remarks about the data 44

4.2 Results of the Rasch model analyses within the Austrian sample 47

4.2.1 Results of the Family Relations Reasoning Test 47

4.2.2 Results of the Syllogisms 2009 50

4.2.3 Recapitulatory examination of the Rasch model results within the Austrian sample 55

4.2.4 Estimated item parameters of the Austrian Syllogisms 2009 and FRRT items 55

4.3 Intercultural results of the Rasch model analyses 56

4.3.1 Results of the Family Relations Reasoning Test 57

4.3.2 Results of the Syllogisms 2009 59

4.4 Results of the hypotheses testings 62

4.4.1 Results of the first study hypothesis 62

4.4.2 Results of the second study hypothesis 68

4.4.3 Results of the third study hypothesis 71

4.4.4 Results of the fourth study hypothesis 78

4.4.5 Additional results of the study and a critical retrospective 82

5 Interpretation and summary of the study results 84

6 Discussion and focus on further research 87

7 Summary 90
8 Indices 92
8.1 Bibliography 92
8.2 List of figures 94
8.3 List of tables 95
9 Appendix 96
Appendix 9.1: The verbal test battery: Access to the verbal test battery 96
Appendix 9.2: The Verbal Test battery: Handout and soziodemographic questionnaire 98
Appendix 9.3: Family Relations Reasoning Test instruction, items, group distribution and unpublished logo draft. 100
Appendix 9.4: Syllogisms 2009 – instruction, items and group distribution 107
Appendix 9.5: Continuative sample descriptions 110
Appendix 9.6: Feedback about the results (to the students of the University of Vienna and the Cypress College in Anaheim, California) 112

1 Introduction

Tests without a free response format are called multiple choice tests and within these, there is a broad variety of answering formats. If doing a multiple choice test (short: MC test), test takers have to choose between two or more given answering options and – depending on the denoted answering format - they have to decide which one is or which ones are the correct answer/ the correct answers. Commonly used answering designs are “1 out of 6” (one correct answer among 5 distractors[1]), “1 out of 5” (one correct answer among 4 distractors”), “x out of 5” (5 response options and any amount of solution between 0 and 5 can be possible) and “1 out of 4” (one correct answer among 3 distractors). However, multiple choice formats can vary within a very wide range and any constellation of right answers and distractors can be possible if expedient.

An important fact about the different multiple choice response formats is that they do not only vary in their designs but – more importantly – in their probability to be solved. Kubinger and Gottschall (2007) could indicate in their study that item difficulty varies significantly according to the conceptualization of different multiple choice response formats. They were able to show that with the multiple choice format “1 out of 6” it is significantly easier to solve an item than the same item but with the format “x out of 5” (which did not significantly differ from the free response format). So, “1 out of 6” has a lower degree of difficulty than the multiple choice format “x out of 5”, which indicates a relevant guessing effect within the first format.

However, not only the response format has an impact of the solution probability of an item but also the construction of the distractors. A research study of Mittring & Rost (2008) could show that if distractors are not designed very carefully, a systematic exclusion of those unfitting – obviously wrong – distractors lead to a higher solution frequency of the items.

[1] Distractors are wrong answers with the aim to distract the test taker from the right solution, if he/she does not know the right answer. For more information see Kubinger (2009).

A very important and in this research study also examined fact about multiple choice tests is the *a priori guessing probability*. Doing a multiple choice test means that a test taker only has to decide which of the given answers is correct and which ones are not. He/She does not have to actively reproduce knowledge. A test taker who does not know which of the answers are true or false can still guess. So for example for the multiple choice format "1 out of 5" – as it was used in this study - the *a priori guessing probability* is 1/5 or 20 %. This indicates that a person who executes a multiple choice test with the format "1 out of 5" could still solve 1/5 of the items by simply guessing.

In addition, another type of multiple choice format is going to be examined in this research study: The sequential response format, which differs from other multiple choice formats in that way that it does not offer the test takers all the possible answers at one time but provides one answer after the other step by step. Once the test taker has chosen an answer to be true, he/she continues to the next item. He/She cannot go back to change the decision. This indicates that it can only be used *via* computer testing (if it is assumed that no assessor wants to replace the work of the computer). To focus on the *a priori guessing probability* within the sequential response format, it is shown, that if the correct answer is offered in the first position, the *a priori guessing probability* to solve the item is ½= 50 % (A person without any solving skills has a chance of 50 % to guess right.). The later the computer offers the solution, the more the *a priori guessing probability* to solve the item decreases (Kubinger, 2003). If a solution is offered in the fifth position, the *a priori guessing probability* to solve the item is only 3.125 % ($^{1}/_{32}$).

It is important to mention that this type of response format takes longer to be executed (and therefore the motivational aspects of the test takers to edit the test may be lower than usual) in comparison to *regular* multiple choice formats.

In this research study possible cultural effects concerning the response styles of the two formats: "sequential 1 out of 5" and "multiple choice 1 out of 5" are investigated. For this reason psychology students at the University of Vienna, Austria and the Cypress College, USA were tested with a verbal test battery which consisted out of two sub-tests that measured verbal reasoning [2] and which response options were changed randomly between the two described formats.[3]

The actual testing took place in Vienna from June to July 2009 and in Cypress, California from September to October 2009.

[2] For more information see Arthur S. Reber & Emily Reber (2001).

[3] At the beginning of the data collection in the United States, a short visit to the supervisors (Dr. Eduard Dunbar and Mr. Ignacio Allegre MSc) was arranged in August 2009 to explain the aims of the study and to clear up any ambiguities about the research scheme. Furthermore, the student tutor who would later be the test assessor was instructed on how to use the test program.

2 Scientific aim of this study

For many years now one of the main interests of the Division of Psychological Assessment at the University of Vienna are the construction and evaluation of multiple choice formats. Research has been done about many different aspects of multiple choice formats like the difficulty of different multiple choice formats (Gottschall, 2007) or the acceptance of various multiple choice formats among students (Liedlbauer, 2009). Litzenberger, Gnambs and Punter, (2005) even composed guidelines about the construction of multiple choice formats.

In some respects, this diploma thesis can be considered as a further (small) investigation into this wide study field of the Division of Psychological Assessment at the University of Vienna. As the title of this thesis indicates, the main focus of this thesis is put on cultural differences between Austria and the United States of America concerning the psychometric qualities of different multiple choice response formats. On this account, a verbal test battery (consisting of the Family Relations Reasoning Test and a short version of the: *Syllogisms*) has been developed and presented to the students *via* remote desktop. Students at the University of Vienna, Austria and students at the Cypress College in California, USA were tested with the verbal test battery (which was designed with the software program "Visual Studio.NET"). The response formats did not vary within one sub-test but could vary within the test battery. So, each sub-test could be executed with a different response format. This leads to the conclusion, that the students executed the test in either one of four response combinations, as it can be seen in table 2-1:

Table 2-1
Possible response options of the sub-tests of the verbal test battery. Students were randomly allocated to one of the four response combinations.

FRRT	**Syllogisms 2009**
Multiple Choice 1 out of 5	Multiple Choice 1 out of 5
Multiple Choice 1 out of 5	Sequential Response Format 1 out of 5
Sequential Response Format 1 out of 5	Sequential Response Format 1 out of 5
Sequential Response Format 1 out of 5	Multiple Choice 1 out of 5

It is important to mention that also the sequential response format belongs to the *family* of the multiple choice formats – as it also offers various solution opportunities and hence guessing effects. However, to make it easier to verbally differ between the two formats, in this thesis, the sequential response format "1 out of 5" will be called "sequential response format" and the multiple choice format "1 out of 5" will be called multiple choice format. Hence, this formulation can be deceptive, the reader is to be asked to remember this statement.

So the primary focus of this thesis was to find out if there are any differences in the answering styles and scores between those two multiple choice formats with regards to cultural differences.

Moreover, it was of interest to find out, if items with middle correct answers (with correct answers in the center position) are easier to solve than those with extreme correct answers (correct answers in extreme – in this case: extreme top or bottom positions). Especially for the sequential response format, this was of interest because this format does not show all the possible solutions at one time. This interest was based on a scientific study of Attali and Bar-Hillel (2003) where they could show that test takers have a strong tendency for seeking correct answers in the center positions. On this account, they postulate that questions with middle correct answers are easier and less discriminating than questions with extreme correct answers.

3 Method

This chapter provides information about the study design of this thesis. Step by step, the single components of this study will be presented to give the reader a detailed overview of the configuration of this study. First, the method of measurement - the verbal test battery will be explained. Then, more information about the various multiple choice formats, the study sample, the research questions and the study hypotheses will be presented to the reader. Finally, the last sub-chapter will provide a description of the psychometric analyses of the items.

To implement the aims of this study (see chapter 2), a verbal test battery was established. This verbal test battery consisted out of two subtests: the Family Relations Reasoning Test ("FRRT") and a short version of the Syllogisms. Psychology students at the University of Vienna, Austria and the Cypress College in California, USA could participate voluntarily to execute the test battery. The students were assigned to one of sixteen parallel groups (the group distribution can be found in table 3-8) and had to execute either items with the format multiple choice "1 out of 5" or the sequential response format "1 out of 5" (the allocation of the students to the various format groups was at random). So, both formats offered one solution (or also called "activator") among four wrong answers (or also called distractors). The answering format (multiple choice or sequential) could vary between the sub-tests of the test battery but did not vary within the subtests. After the data collection, psychometric analyses were calculated to see if the items fit the Rasch model assumptions.

3.1 Method of Measurement

According to the aims of this research, a computer based test program – called Verbal Test battery 2009 – was developed with the program *Visual Studio.NET*. The test program could be entered through a link on the homepage of the University of Vienna. Screenshots of the entering mode can be seen in appendix 9.1. Depending on the language selected, the test takers could either execute the test in English or in German. The English version of the test battery was translated and then double checked by a native English speaking psychologist at the University of Vienna. It was considered to collect the data under controlled conditions to avoid giving the test takers any chance to use helping facilities like mobile phones or the internet to solve the items. While in Austria one main test assessor was responsible of the data collecting process, two American test assessors shared this function at the Cypress College in Anaheim. Mrs. Nicole Magdaleno, psychology student and student tutor and Mr. Ignacio Allegre MSc, teaching psychologist at the Cypress College in Anaheim offered their help to control the data collection in the United States of America.

After having opened the link of the German or English verbal test battery version, the test takers were connected to the remote desktop and led to three welcoming pages, which extensively explained the further steps on how to enter the test battery. It was very important to clear up any ambiguity about the test program before the data collection started to create a well organized test situation.
On the last welcoming page of the verbal test battery, a link through which the connection to the test program was established could be found. The test program could be entered with the help of the test assessors who primarily were instructed about the username and password to log in (A visualized version of the entering process can be seen in appendix 9.1.).

After having successfully completed the login, the entry mask of the test battery appeared. The test takers had to fill in the following questions:

- Name/Identification: The students were instructed to use a special code system instead of their names. This guaranteed that the test takers did not have to worry about their names being linked to their test results. While the code system in Austria consisted out of a combination of the first two letters of the first name, the first two letters of the family name and the last two numbers of the student identification number, the American test assessors arranged code-numbers for their students. (The advantage of the German code system is that the students do not have to remember what number or code name they typed in. With the help of the instruction of the code composition it is possible to recreate the entry name even after a longer period of time).
- Age: Only the number of years was of interest.

- Nationality: The focus was on the cultural background in reference to the country in which the test was executed.
- Native tongue: This question asked about the language the test taker grew up with.
- Profession (Study field): Only the major subject at the University was asked (and of interest).
- Degree in progress: This question was designed to ask about the next formal degree which the students want to gain in their Psychology studies (For the USA, the students were supposed to write in “bachelor”, “master of science” or “PhD”, while students in Austria could either fill in “1. Abschnitt”, “2. Abschnitt“ or „Doktorat“.).
- Sex: This question simply asked if the test taker was male or female.

After having completed to fill in the entry mask (which can be seen in appendix 9.1), the actual test program started. All students had to execute the Family Relations Reasoning Test first and then work on the short version of the Syllogisms. For the FRRT which actually consisted out of 60 items, four parallel test forms were created. This was necessary to keep the required time to complete the test less than one hour (There were concerns that the motivational aspects of doing this test without any financial compensation would be very low, if the testing time would take too long.) So called “linking items” (items which occurred in two or more of the four parallel test forms) were created to make it possible to compare the four groups. The various groups were considered not to differ in difficulty. To inspect the equivalency of difficulty between the groups, the item easiness parameters of the items were identified, based on earlier research, done by Schechtner in 2009. Then the items were divided into four groups so that each group was approximately as difficult as the others (for more information about the groups, see appendix 9.5).
The allocation of the students to the groups was at random. The order of the combination of the response formats was perseverative. So each student could get one of four response style combinations. Table 3-1 shows the distribution of the students to the four format combinations.

Table 3-1
Overview of the distribution of the four different response format versions of the verbal test battery. Instead of “multiple choice”, the short form “MC” is used and instead of “sequential response format”: “SEQ” is used. 4 students of the Austrian sample and 28 students of the U.S. American sample executed only the FRRT. This is why their results are defined as “missing”.

Distribution of the various multiple choice formats

	Austrian sample		U.S. sample	
Response format combinations	Frequency	Percent	Frequency	Percent
MC (FFRT) – MC (Syllogisms 2009)	55	24.4	13	21.3
SEQ (FRRT) – SEQ (Syllogisms 2009)	54	24.0	7	11.5
MC (FRRT) – SEQ (Syllogisms 2009)	57	25.3	6	9.8
SEQ (FRRT) – MC (Syllogisms 2009)	54	24.0	7	11.5
Missing	4	1.8	28	45.9
Total	225	100.0	33	54.1

The students had to execute both sub tests in order to successfully finish the verbal test battery.

After having completed both subtests, the program closed automatically and the test takers were led back to the homepage of the University of Vienna.

3.2 The Family Relations Reasoning Test

The Family Relations Reasoning Test or shortly called FRRT was designed to measure verbal reasoning. It is a further development of the *Verwandtschaften-Resoning-Test* (Skoda, 2005) and the Family Reasoning Test (short FRT; Schechtner, 2009). Psychometric analyses about the FRT can be found by Poinstingl, 2009. Concerning the FRT, Schechtner developed fifty items with specific item construction rules in a paper pencil format. Those items were tested on 506 students between 13 and 20 years at different high schools in Styria. The response format of the FRT was “1 out of 8” (One solution among seven distractors.). 39 of 50 items fit the assumptions of the Rasch model but all items were rather easy to solve for the test takers (no item easiness parameters higher than -1.8; many parameters between 0.1 and 2.5). These facts lead to the aim to design new items (and item construction rules) for the FRRT to create more difficult items. So, the refined item pool counted a total number of sixty FRRT items. Schechtner’s fifty original items were retained plus ten new items (partly with new constructions rules; see chapter 3.2.2) were developed.

Schechtner's 50 original items were adapted and imprecise articulations, like unisex, very rare or too long names (more than 5 letters) were replaced through gender-specific, commonly used and shorter names. Further attention was paid to the equivalency between the German and the English names in term of their popularity, length and number of syllables. An official website, the "UK national statistic" website for popular English names worldwide ("http://www.statistics.gov.uk") was used to find suitable English equivalents to the German names. It was also an effort to find English names with the same number of syllables as their German counterparts. Another important consideration was that the German names were always replaced with the same English names, if a name appeared more often than in one item. (So for example, the German name *Ute* was always replaced with the English name *Uma* in every item in which it appeared.) This accuracy was important to reduce the number of interfering variables caused by imprecise articulation or bad translation.

Among the ten new items, six are based on Schechtner's original item construction rules (see chapter 3.2.1), while four items were constructed with an extended rule (see chapter 3.2.2).
However, first a general buildup of a simple FRRT item will be presented to give the reader an impression on how the items look like. Then, beginning with chapter 3.2.1, a closer look at the construction rules will be taken.
The general content of an FRRT item can be explained very simply: Each item consists of a story text which describes a family with its different relations among the individual family members. The test taker has to find the right relatedness between two family members.

<u>Example of a story text:</u>
Claire is the daughter of Sonya. Sonya is the daughter of Maria. Maria is the _________ of Claire.

Because of the focus on the multiple choice formats "1 out of 5", five possible familial relationships were offered and the test taker had to find the only correct relationship.

Example for possible family relationships that were offered:

a) *Maria is the grandmother of Claire.*
b) *Maria is the aunt of Claire.*
c) *Maria is the niece of Claire.*
d) *Maria is the daughter of Claire.*
e) *Maria is the granddaughter of Claire.*

In this case, the test taker should have clicked at "*a) Maria is the* ***grandmother*** *of Claire*" to solve this item.

To avoid too much complication, the test did not deal with half siblings, step siblings and step parents. In other words, the FRRT only handles married couples who exclusively have children with each other. This information was given to the test takers at the very beginning of the test. *Via* a help button on the bottom right of the screen, this information was available at any time during the whole sub-test.

3.2.1 Item construction rules of the FRT

As explained in chapter 3.2 the forerunner of the Family Relations Reasoning Test was a test called Family Reasoning Test (or sortly called FRT) by Schechtner (2009). In her thesis, Schechtner postulates that the difficulty of the items increases with the rise of: generations combined with the rise of relations and redundancies. All the three factors will now be described in this chapter and examples will be presented to get a better illustration of the quoted points.

I. **Generations**. The maximum number of generations that is used in the Family Reasoning Test (and also in the Family Relations Reasoning Test) is three: *child – parents – grandparents*. Table 3-2 gives an overview of the three different generations distinguished by Schechtner in the Family Reasoning Test and which are also used in the Family Relations Reasoning Test.

Table 3-2
Overview of the generations that are used in the FRT and FRRT. The very left column contains the generic terms of the different generations. The second column contains the explanations of those generic terms. The third column gives item examples and the fourth shows the generation diagrams.

Generation Term Schechtner (2009)	**Explanation**	**Item example**	**Generation diagram**
2 G	2 generations are combined	Laura is the **mother** of Nina.	Mother (Laura) ↓ daughter (Nina)
3 G	3 generations are combined	Joe is the **grandfather** of Sonya	Grandfather (Joe) ↓ unknown person ↓ granddaughter (Sonya)
3 G+	3 generations are combined but the right answer is either: cousin/cousin niece/nephew aunt/ uncle	Joe is the grandfather of Sonya. Sonya is the cousin of Lara.	Grandfather (Joe) ↓ Nina → unknown person ↓ ↓ granddaughter (Sonya) niece (Lara)

II. **Number of relations**. Relations are important information which are needed to solve the tasks. An example for relations are:

Relation 1

"Tina is the mother of Sonya. Joe is the father of Tina. Joe is the ________ of Sonya."

Relation 2

Important information to be able to solve this example item is to know that Joe is the father of Tina and that Tina is the mother of Sonya. So this example contains two relations:

- Relation 1: *Tina is the mother of Sonya.*
- Relation 2: *Joe is the father of Tina.*

III. **Number of redundancies**. Redundancies are unimportant information which are not necessarily needed to solve the tasks. The item above is now written with two relations plus one redundancy:

Redundancy

"Tina is the mother of Sonya. Joe is the father of Tina. ***Tim is the son of Tina.*** *Joe is the ________ of Sonya."*

To solve this example item it is not necessary to know that Tim is the son of Tina. So, this sentence is a redundancy.

Furthermore, Schechtner designated two different ways of linkages between two family members:

- Paul **is the** brother of Simon. → "is- the- linkage"
- Paul **has a** brother called Simon. → „has- the- linkage"

These linkages only allow the connection of two family members in one sentence or phrase.

Schechtner (2009) combines various generations, relations and redundancies in her items through the linkages named above. Table 3-3 gives an overview of the combinations that are used in the Family Reasoning Test. The numbers in the field provide information about how many times this combination was used. The grey coloured fields are new combinations that are only used in the Family Relations Reasoning Test.

Table 3-3
Overview of the combinations of generations, relations and redundancies in the FRRT. The very left column shows the different generations (G) that are used in the FRRT. The first row gives an overview of the various combinations of relations ("Rel") and redundancies ("Red"). The numbers in the crossed fields show the quantity of items of that specific combination. The grey coloured fields with the numbers printed in bold letters are new items that were only used for the FRRT.

	2 Rel 0 Red	2 Rel 1 Red	2 Rel 2 Red	3 Rel 0 Red	3 Rel 1 Red	3 Rel 2 Red	4 Rel 0 Red	4 Rel 1 Red	4 Rel 2 Red	5 Rel 0 Red	5 Rel 1 Red	5 Rel 2 Red
2 G	2	2		2	2	2						
3 G	2	2		2	2	2	2	2	2	2	2	**3**
3+G	2	2		2	2	2	2	2	2	2	2	**3**

3.2.2 Item construction rules of the FRRT

To create more difficult items the hitherto existing generation rules were extended. It was assumed that not only the numbers of generations, relations and redundancies can increase the difficulty of the items, but also the type of linkage. A first approach can be already seen by Schechtner (2009) in her “3+G” generation step, where she assumes that an item with three generations is even more difficult if the answer is either cousin/cousin, niece/nephew or aunt/uncle.

The first step to create new item construction rules was to draw tree diagrams about the various familial connections that could exist. This lead to the first new item construction rule:

I. <u>The first way of linkage: The kind of affinity</u>

It is assumed that there is a difference in the level of difficulty if two family members are “directly” connected (like mother and daughter) or if there is/are one person/ more persons that are needed to connect two family members.

An example of a so called *“direct connection”* would be: mother and daughter.
An example of an *“indirect connection”* (through other people) would be: grandmother (– *via* the parent) – granddaughter. The connection between grandmother and granddaughter leads over the mother or the father of the granddaughter.

So, the difficulty of the items increases not only with the number of generations, relations and redundancies but also with the kind of affinity between two family members. Table 3-4 gives an overview about the different ways of linkage related to the various kinds of affinity.

Table 3-4
Overview of the different kinds of relations and their ways of linkage. The five columns describe: the kind of linkage (column 1), the explanation of this linkage (column 2), an example of this linkage (column 3), a graphical drawing of the linkage (column 4) and a list of relatives that belong to this linkage (column 5).

Way of Linkage	Explanation	Example	Graphical note of the example	Relatives that belong to this category
simple linkage	direct connection	A is **the mother** of B.	A ↓ B	Daughter/ Son Sister/ Brother Mother/ Father Wife/ Husband
dual linkage	indirect connection *via* one more person	A is **the granddaughter** of B. (The connection is *via* the parent of A)	B ↓ (parent of A) ↓ A	Granddaughter/ Grandson, Grandmother/ Grandfather, Aunt/ Uncle, Niece/ Nephew
ternary linkage	indirect connection *via* two more persons	A is **the cousin** of B. (the connection is *via* the parents of A and B)	parent ↔ parent of A of B ↑ ↓ A B	Cousin (female)/ Cousin (male)

II. <u>The second way of linkage: Number of connections</u>

The second new item construction rule focuses on the number of connections that appear in one sentence or phrase. This is a big difference to the original item construction rules, where only two family members were connected with each other in one sentence or phrase. The new rule allows the connections of three to four people in one sentence or phrase through a simple method: the normal “is- the- connections” and the “has- the- connections” are extended with additional connections. It is assumed that an increased number of relations in one sentence or phrase lead to more difficult items. Table 3-5 gives an overview of the various additional connections in the original “is- the- connections” and the “has- the- connections”.

Table 3-5
Overview of the old and the new item construction rules. The left three columns explain the original item construction rules of Schechtner (2009), while the three columns on the right side explain the new item construction rules with the additional connections.

Original item construction rule	**Example of the original item construction rule**	**Tree diagram** (original item construction rule)	**Extended item construction rule**	**Explanation of the extended item construction rule**	**Tree diagram** (extended item construction rule)
is- the- linkage	K is the sister of B.	K ↔ B	**A's mother** K is the sister of B.	**One** additional connection interferes with the original "is- the- linkage".	K ↔ B ↑ A
is- the- linkage	K is the sister of B.	K ↔ B	A**'s mother** K is the sister of M's wife B.	**Two** additional connections interfere with the original "is- the- linkage".	K ↔ B ∞ M ↑ A
is- the- linkage	K is the sister of B.	K ↔ B	A**'s mother** K is the sister of **the mother of** L**'s wife** O. *or* B, the **mother of L's wife O** is the sister of **A's mother K**.	**Three** additional connections interfere with the original "is- the- linkage".	K ↔ (B) ↑ ↑ A O ∞ L *or* K ↔ B ↑ ↑ A O ∞ L

The two new item construction rules can be combined with Schechtner's (2009) rules but also with the other extended item construction rules that handle the kind of affinity. In this case, a great variety of familial connections can be created. Table 3-6 shows all the possible connections of the new item constructions rules.

Table 3-6
Overview of all the possible connections of the new item constructions rules

Way of linkage	Item example	tree diagram
simple linkage with one additional connection	A’s mother K is the sister of B	K ↔ B ↑ A
dual linkage with one additional connection	A’s mother K is the aunt of B.	K ↔ (?) ↑ ↑ A B
ternary linkage with one additional connection	A’s mother K is the cousin of B.	(?)↔ (?) ↑ ↑ K B ↑ A
simple linkage with two additional connection	A’s mother K is the sister of M’s wife B.	K ↔ B ∞ M ↑ A
dual linkage with two additional connection	A’s mother K is the aunt of M’s wife B.	K ↔ (?) ↑ ↑ A B ∞ M
ternary linkage with two additional connection	A’s mother K is the cousin of M’s wife B.	(?)↔ (?) ↑ ↑ K B ∞ M ↑ A
simple linkage with three additional connection	A’s mother K is the sister of the mother of M’s wife B. *or* X, the mother of M’s wife B is the sister of A’s mother K.	K↔(mother of B) ↑ ↑ A B ∞ M K ↔ X ↑ ↑ A B ∞ M
dual linkage with three additional connection	A’s mother K is the aunt of M’s wife B.	(?)↔ (?) ↑ ↑ K B ∞ M ↑ A
ternary linkage with three additional connection	A’s mother K is the cousin of the mother of M’s wife B. *or* X, the mother of M’s wife B is the cousin of A’s mother K.	(?)↔ (?) ↑ ↑ K (mother of B) ↑ ↑ A B ∞ M *or* (?)↔ (?) ↑ ↑ K X ↑ ↑ A B ∞ M

Following these rules, four more FRRT items were created. The other six additional FRRT items were designed with Schechtner’s (2009) rules (with an increased number of relations and redundancies). All the items of the FRRT (old and new) can be found in appendix 9.3.

3.3 Syllogisms 2009

For this study, only 10 items of the test *Syllogisms (Srp, 1994)* which was developed to detect verbal reasoning and fluid intelligence (Poinstingl, 2001) were used. To make it easy to differentiate the “short version” with the original test, these ten items were merged to the sub-test of the verbal testbattery called “Syllogisms 2009”.
The original test consists out of 75 items - each one in the multiple choice format “1 out of 4” which measure verbal reasoning. For this study, a short version with only 10 items was used, because the whole testing period for each student was considered to last no longer than 30 to 60 minutes. Like the number of items, also the response format was changed. The original format “one solution among four distractors” was extended – under the respect of the Syllogisms item construction rules – with one more distractor to create a “1 out of 5” response format. All of the Syllogisms 2009 items can be found in appendix 9.4. More information about the construction of Syllogisms can be found by Nortmann (1996).

Now, the items of the Syllogisms 2009 will be explained: Each item offers two statements (or also called premises) and with the help of these two statements, logical conclusions have to be drawn. Thereby a series of possible conclusions is offered and the test taker has to find the right conclusion among the distractors.

<u>An example of two premises and five conclusions:</u>

Premise 1: All A are B
Premise 2: All B are C

---

a) *Conclusion 1: All A are C*
b) *Conclusion 2: Some A are no C*
c) *Conclusion 3: All A are no C*
d) *Conclusion 4: Some A are C*
e) *Conclusion 5: No A are C*

In this example, the right answer is *a) All A are C.*

Via a help button on the bottom right of the screen, information about the four main Syllogisms rules was available at any time during the whole (sub-) test situation. Figure 3-1 shows a screen shot of the four rules.

Figure 3-1: Screenshot of the guidelines to solve the Syllogisms 2009 items. This help page could be opened at any time during the testing period of the Syllogisms 2009.

3.4 Multiple choice formats

As already mentioned in chapter 2, one of the aims of this study was to examine the differences between the two multiple choice formats: multiple choice "1 out of 5" and sequential response format "1 out of 5". In this chapter, these two answering formats will be described to point out the difference between them:

- **Multiple choice format "1 out 5" (short MC "1 out of 5"):** If using this format, each item consists out of a story text or question and five possible solutions. The answering suggestions are all given at once and the test taker has to find the one correct answer among the 4 distractors. In this study, the story text was given on the top half of the screen, while all the possible answers were lined up vertically underneath. An example of a multiple choice format "1 out of 5" would be:

What is the capital of Austria?

- Graz
- Vienna
- Linz
- Salzburg
- Innsbruck

- **Sequential response format „1 out of 5" (short: SEQ "1 out of 5):** Using the sequential response format *via* computer, each task only offers the story text or the problem and the first response option. The test taker has to decide whether this first given answer is true or false. If he/she decides that the first presented answer is correct, then the computer switches to the next task. If the test taker decides that the first presented respond option is false, this option disappears and a second possible solution is presented. Then the appraisal process starts all over again and the test taker has to decide again, if this second suggested solution is correct or not. If it is decided to be correct, then the computer switches to the next task. If the test taker also refused the second solution option, the computer presents a third possible answer and so on. If all five answers are rejected, the computer switches to the next task. However, the person does not know how many solution options will follow. So, even if the format is sequential "1 out of 5", the test taker does not know that he/she gets five solution possibilities offered if he/she neglects the first four answers.

  Figure 3-2 shows a block chart about the possible answering possibilities during the sequential testing.

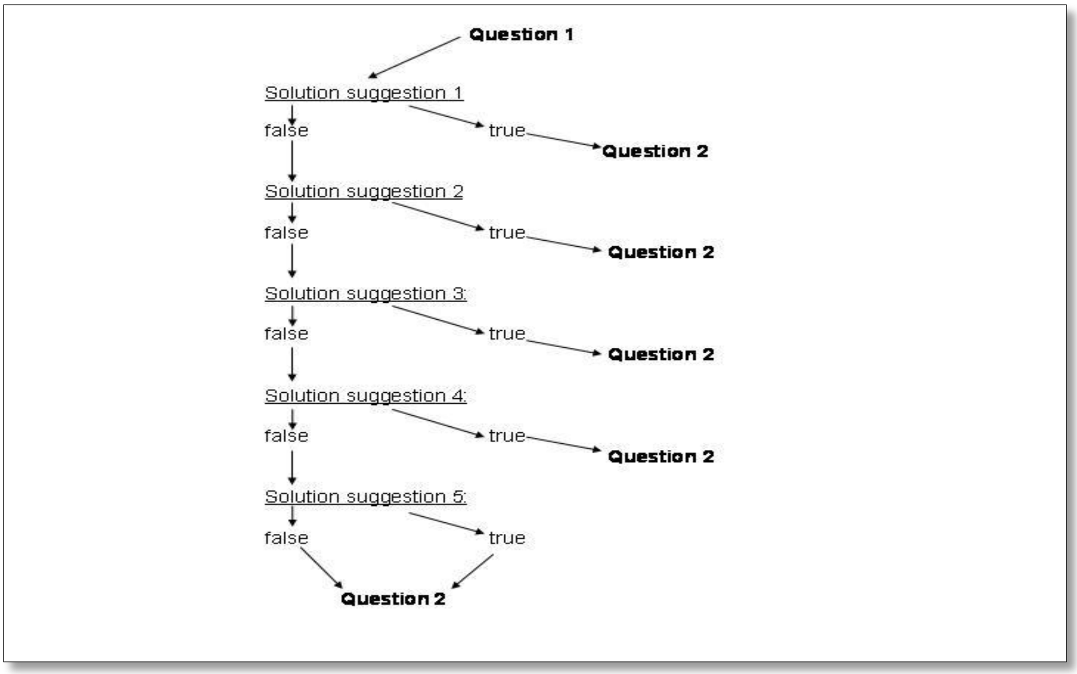


Figure 3-2: Tree diagram of the possible ways to edit an item in the sequential response format. If a given solution possibility is rejected, another solution suggestion is given. If an answer is found to be correct, then the computer switches to the next question.

It is important to mention that the sequential response format does not allow the test taker to switch back- and forwards between the given answering opportunities. Once an answer is rejected and the next possible solution is already presented, the person cannot revise his/her decision and go back to the previous response option.

3.5 Parallel groups

Each item of the FRRT and the Syllogisms 2009 was designed as a pair (with the exact same content but with the two different answering formats: multiple choice "1 out of 5" and sequential response format "1 out of 5". Because of the high number of FRRT items (k= 60), the items were allocated to four parallel groups. Those groups were connected *via* linking items (items which appear more often than in one group) to allow a direct comparison. Table 3-7 gives an overview of the distribution of the FRRT items to the four parallel groups. Each item is listed up with its name and its estimated item easiness parameter. (Though it might sound strange to describe an item by its "easiness", in this study, all the estimated item parameters are easiness parameters because of the extended Rasch modelling program *eRm* which only calculates easiness parameters. So, it has to be mentioned, that the higher the easiness parameter, the easier is the item.)

Table 3-7
Overview of the distribution of the items to the four parallel groups of the FRRT. The items are presented with their number and their item easiness parameter. Items written in thick letters are linking items and items which are marked in grey are deleted items. The items are ordered in the same order as they were offered to the students.

FRRT group 1		FRRT group 2		FRRT group 3		FRRT group 4	
Item number	easiness parameter	Item number	easiness parameter	Item number	easiness parameter	Item number	easiness parameter
3	0.339	2	1.671	1	deleted	5	-0.627
6	0.929	4	0.225	9	deleted	12	0.740
8	1.374	7	-0.377	**10**	**2.384**	**14**	**0.244**
14	**0.244**	**10**	**2.384**	11	1.203	15	-0.087
17	-0.807	**14**	**0.244**	13	deleted	**18**	**-0.096**
22	**0.244**	**18**	**-0.096**	**14**	**0.244**	21	0.244
25	**0.903**	19	0.637	16	0.625	**22**	**0.244**
27	-0.075	23	0.899	20	-0.251	26	0.140
30	0.382	24	1.67	**25**	**0.903**	29	0.408
31	0.602	28	0.225	32	0.882	35	-0.087
37	-1.251	**30**	**0.382**	33	0.048	36	1.927
40	1.374	34	0.225	38	-0.633	**42**	**-0.562**
41	0.602	45	-0.245	39	0.218	43	-1.626
44	-0.682	**48**	**-0.608**	**42**	**-0.562**	47	1.190
48	**-0.608**	49	0.054	46	-1.772	**48**	**-0.608**
56	-2.322	54	-1.782	**48**	**-0.608**	50	0.740
60	-0.682	58	0.225	53	0.408	52	-2.227
62	-0.926	61	-2.322	64	-1.674	63	-1.674

The students who took part in this study were randomised allocated (*via* computer) to one of the sixteen test forms. An overview of the distribution to the sixteen test forms can be seen in table 3-8.

Table 3-8
Overview of the sixteen parallel groups within the verbal test battery. The category *Missing* shows the amount of people who only executed one sub-test and therefore could not be allocated to one of the 16 (combined) test forms.

	Austrian sample		U.S. sample	
Sub-test combinations	**Frequency**	**Percent**	**Frequency**	**Percent**
FRRT(1) - Syllogisms: MC – MC	14	6.4	4	12.1
FRRT(1) - Syllogisms: MC- SEQ	14	6.4	1	3.0
FRRT(1) - Syllogisms: SEQ - SEQ	13	5.9	1	3.0
FRRT(1) - Syllogisms: SEQ – MC	12	5.5	2	12.1
FRRT(2) - Syllogisms: MC – MC	14	6.4	4	6.1
FRRT(2) - Syllogisms: MC – SEQ	14	6.4	1	3.0
FRRT(2) - Syllogisms: SEQ – SEQ	14	6.4	3	9.1
FRRT(2) - Syllogisms: SEQ – MC	15	6.8	1	3.0
FRRT(3) - Syllogisms: MC – MC	14	6.4	0	0.0
FRRT(3) - Syllogisms: MC – SEQ	14	6.4	3	9.1
FRRT(3) - Syllogisms: SEQ – SEQ	14	6.4	2	6.1
FRRT(3) - Syllogisms: SEQ – MC	15	6.8	3	9.1
FRRT(4) - Syllogisms: MC – MC	11	5.0	5	15.2
FRRT(4) - Syllogisms: MC – SEQ	15	6.8	1	3.0
FRRT(4) - Syllogisms: SEQ – SEQ	13	5.9	1	3.0
FRRT(4) - Syllogisms: SEQ – MC	14	6.4	1	3.0
Missing	5	2.2	28	45.9
Total	220	97.8	33	54.1

As it can be seen in table 3-9, the students were allocated to one of four multiple choice format combinations. To have a closer look at the distribution of the students to only the response format combinations (without consideration of the FRRT groups), table 3-9 shall give an overview.

Table 3-9
Distribution of the students to the response formats.

Country	Sub-test	Response format	Frequency	Percent
Austria	FRRT	MC	112	49.8
		SEQ	113	50.2
Austria	Syllogisms 2009	MC	109	49.3
		SEQ	112	50.7
USA	FRRT	MC	52	50.0
		SEQ	52	50.0
USA	Syllogisms 2009	MC	53	52.0
		SEQ	49	48.0

3.6 Sample

All together 329 psychology students participated in this study. Among these 329 students, 225 were students at the University of Vienna, Austria. The other 104 students were psychology students at the Cypress College in Anaheim, California. The aim of this chapter is, to describe both sub-samples and to give the reader a good overview of the sample.

The data collection in Vienna took part from June 20th to July 10th. One main test assessor coordinated the testing times of all 225 Austrian students. The testing took place in the computer laboratory of the Psychological Division in Vienna. The data collection in the United States of America was split into two testing times. The first period of data collection took place from September 11th to September 18th at the Cypress College in Anaheim under the guidance of Dr. Eduard Dunbar, UCLA professor in clinical psychology and Mrs. Nicole Magdaleno, psychology student at UCLA. The second period of data collection took also place at the Cypress College in Anaheim, from October 28th to October 31st 2009 under the guidance of Mr. Ignacio Allege MSc., statistic teacher at the Cypress College. A short visit to Los Angeles and Anaheim in August 2009 was done to introduce the test program to the assessors, to clear up any ambiguity about the test program. Also, it was double-checked if the program works on the computers at the Cypress College.

During the whole data collection, students could participate voluntarily to execute the testbattery. Unfortunately, no financial support was possible to be provided to the students, but a feedback (see appendix 9.6) about their results and sweets were offered. Austrian students could also decide for V.I.P. entry tickets to a big new dance club in Vienna. The Austrian test takers were recruited from several psychology seminars, where the study was presented to them and folders with the test schedules were handed out. Also, posters that were put on the walls around the computer laboratory of the Psychology Division invited to participate. Little handouts with the contact data of the Austrian main test assessor were provided to the Austrian students. These handouts also had an extra space where the students could write down their testing times ("*Post-testing* feedback" of the students indicated that these little handouts helped them to remember their testing time.). The handout can be seen in appendix 9.2.

All of the students were fully enrolled psychology students at their university or college. It was underlined to them that their participation was voluntary and that the data collection was anonymous and only used for scientific reasons.

Concerning the Austrian sample, 170 female and 55 male students executed the test. The unequal allocation of the gender among the students who participated can be lead back to the unequal allocation of the gender of psychology students at the University of Vienna. (More female students than male students study Psychology.) The range of the age was 40 years with a minimum of 19 years and a maximum of 59 years. Figure 3-3 shows a histogram of the age distribution among the Austrian students. It can be seen that most of the students were between 20 and 30 years old. Additional sample descriptions can be found in appendix 9.5.

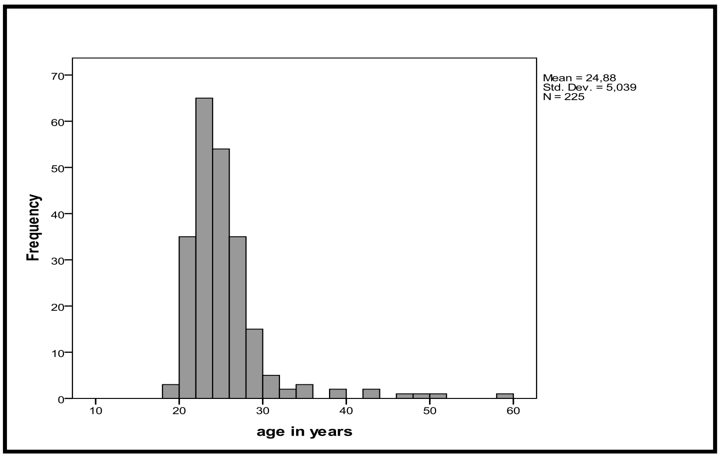


Figure 3-3: Histogram of the age distribution (in years) of the Austrian students (n= 225).

Now focusing on the U.S. American sample, it can be said that 64 % of the test takers were female (n= 67 students). The other 36 % were male test takers (n= 37 students). So also in the U.S. American sample an unequal distribution of the gender of the test takers to the favour of the women occurs.

Concerning the age of the U.S. American test takers, a range of 41 years can be calculated with a minimum of 18 and a maximum of 59 years. Most students reported an age between 20 and 25 years. The age distribution (in years) of the U.S. American sample can be seen in Figure 3-4.

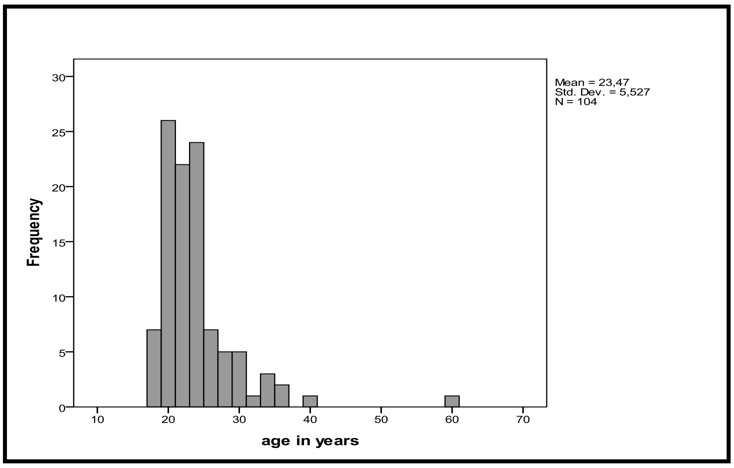


Figure 3-4: Histogram of the age (in years) distribution of the U.S. American students (n= 104).

Further, it has to be mentioned that the U.S. American sample can be split into two times of data collection. The first part took place from September 11th to September 18th 2009 and was supervised by Dr. Dunbar, psychology professor at the University of California, Los Angeles and Mrs. Nicole Magdaleno, psychology student at the

University of California, Los Angeles. 54 students (52 %) of the U.S. American sample were tested in a group testing at the Cypress College. The other 50 students (48 %) were tested serially under the guidance of Mr. Ignacio Allegre MSc, statistic teacher at the Cypress College in Anaheim. This second data collection in the United States took place from October 28th to October 31st 2009 and was also done at the Cypress College in Anaheim.

3.7 Research questions and study hypotheses

Scientific research about multiple choice tests is a very important and wide investigated study field within the Division of Psychological Assessment of the University of Vienna. What is special about this research study is that the focus is not only put on multiple choice responding behavior but also on cultural differences between Austria and the United States of America.
In this context the following research questions were developed:

a) Research questions concerning the tests and the multiple choice formats were:

- Do students who score high on the FRRT also score high on the Syllogisms?
- Does the answering format have an influence on how we score in a test: Is one multiple choice format more helpful for the participant to solve an item than the other?

b) Research questions concerning possible differences in the answering style and behaviour between Austria and the United States of America were based on considerations about the university principles, like the frequency on how often students have to use multiple choice formats during their educational career, or how often students at the Cypress College were in contact with psychological computer tests during their education. This lead to the following questions:

- Are there (cultural) differences in the answering style between Psychology students of the Cypress College and Psychology students of the University of Vienna, based on differences in education, experience in working with MC formats or university principles?

- Are there (cultural) differences in the scores between Psychology students of the Cypress College and Psychology students of the University of Vienna, based on differences in education, experience in working with MC formats or university principles?

The research questions above lead to the following study hypotheses:

H_0 (1): There is no mean difference of the solution frequency between paired items (same content but different response formats) with the multiple choice format “1 out of 5” and the sequential response format “1 out of 5”.

$$\mu 1 = \mu 2$$

H_1 (1): There is a mean difference of the solution frequency between paired items (same content but different response formats) with the multiple choice format “1 out of 5” and the sequential response format “1 out of 5”.

$$\mu 1 \neq \mu 2$$

H_0 (2): There is no coherence between the scores of the test takers on the FRRT and the scores of the test takers on the Syllogisms 2009 – if a test taker scores high on the FRRT, he/she must not necessarily score high on the Syllogisms 2009 (and the other way round).

H_1 (2): There is a coherence between the scores of the test takers on the FRRT and the scores of the test takers on the Syllogisms 2009 – test takers who score high on the FRRT, also score high on the Syllogisms 2009 (and the other way round).

H_0 (3): There is no difference in the answering style (declared through editing time and choice of solution) between psychology students at the Cypress College in California, Los Angeles and psychology students at the University of Vienna.

H_1 (3): There is a difference in the answering style (declared through editing time and choice of solution) between psychology students at the Cypress College in California and psychology students at the University of Vienna.

H_0 (4): There is no difference in the scores between psychology students at the Cypress College in California and psychology students at the University of Vienna.

H_1 (4): There is a difference in the scores between psychology students at the Cypress College in California and psychology students at the University of Vienna.

3.8 Psychometric analyses

This chapter is going to describe briefly the analysing test models that are used for this study. The aim is to give the reader an overview of the psychometric methods that are used to analyse the data. The descriptions of the models are cut short on purpose because they are not the preliminary topic of this thesis. Interested readers will find literature hints, where they can read more about the particular models.

3.8.1 Rasch model analyses

One condition before the hypotheses testings could have started was to prove, if the items fit the assumptions of the Rasch model. This is a necessary condition to be able to say that the raw score is a “fair” test value (Evidence can be found by Fisher, 1974 and 1995). If the items of a test fit the Rasch Model, then it can be assumed that they all measure in one dimension and therefore each person can be represented by only one number or also called parameter (Kubinger, 2006; Bühner, 2006).

If an item fit the Rasch model assumptions, then the probability that this particular item is solved by a test taker is depending on only two parameters:

- the person parameter (the ability level of a person) and
- the item parameter (the difficulty of the particular item).

Further it is very important to mention that the Rasch model is sample independent (Fischer, 1974). This is important to understand later calculations. Sample independency indicates that:

- A comparison between two individuals is independent of which particular item they were editing and also independent of which other individuals were also editing this item.
- The comparison of difficulty between two items is independent of which particular individuals were editing this item and which other items within the sample were or might also have been compared.

So, in this study the sample was separated in various parts of sub-samples. To fit the Rasch model, an item had to get similar easiness parameters (which describe the easiness of the item) within each of the two sub-samples of one criterion. If an item gets very different easiness parameters within the two sub-samples of a criterion, then it can be assumed that it does not measure in only one dimension (In this study, it was the aim to check, if the items measure verbal reasoning.) and it was discarded of further analyses. So, if an item fits the assumptions of the Rasch model, this can be proved *via* parameter tests but also *via* graphical model check. Both methods will be shown in this study. While the Likelihood ratio test of Andersen proves the overall model fit, the graphical model checks show the specific item fits: the item easiness parameters which are calculated in each sub-sample of the split criteria are positioned on the X- and the Y-axis, of an orthogonal coordinate system. If the items are conform to the Rasch model, they lay close (or on) a 45° line that goes straight through the origin of the coordinate system. If an item is too far away of diagonal, then it can be assumed that it has different item easiness parameters within the two sub- samples and therefore it does not fit the Rasch model assumptions.
To be able to prove if the items were conform to the Rasch model, the following three criteria were used to split the sample into the sub-samples:

<u>Raw score:</u> (low raw score *versus* high raw score); internal partition criterion. The sample was departed by the median of the distribution of the raw score.
<u>Sex:</u> (female *versus* male test takers); external partition criterion. The sample was departed by the gender of the test takers.

Response format: (multiple choice “1 out of 5” *versus* sequential response format “1 out of 5”), external partition criterion. The sample was departed by the response formats that were used by the test takers.

Table 3-10 gives an overview of the distribution within the sub-sample of each partition criterion.

Table 3-10
Overview of the distribution within the sub-samples of the partition criteria *sex* and *format*. The middle column represents the two sub-samples of each partition criterion which were compared to each other. *MC* is used as the short form of *multiple choice format “1 out of 5”* and *SEQ* is used as the short form of sequential *response format “1 out of 5”.* The column *Frequency* shows the absolute number of students in each sub-sample, while the column on the very right gives information about the percentage of the number of students in each sub-sample.

Subt-tests	External Partition Criteria	Sub-samples	Frequency	Percent
FRRT	sex	men	170	75.6 %
		women	55	24.4 %
	format	MC	112	49.8 %
		SEQ	113	50.2 %
Syllogisms 2009	sex	men	52	76.5 %
		women	169	23.5 %
	format	MC	109	49.3 %
		SEQ	112	50.7 %

It is also important to mention that if an item was solved by all test takers of one sub-sample (or even the whole sample) – as it happened in this study – the item parameter of this item could not be calculated by the Rasch model because in this particular case the item gives no information about its difficulty. However, important for this study is to know how this special case was preceded: If an item parameter could not be estimated within all partition criteria, then this particular item was deleted of further analyses. If an item parameter could not be estimated within “only” one or two sub-samples, then the item was kept for further analyses.

3.8.2 Welch-Test

After the Rasch model analyses had been completed – which were a necessary condition to do further analyses – the actual hypotheses testings could have been started. The Rasch model analyses were done with the program *R 2.10.0* – a freeware program to calculate Rasch models *inter alia*. The statistic program *PASW Statistics*

18 (formerly SPSS statistics) was used for further statistical analyses. All hypotheses were proved *via* Welch-Test, a modified Student's t-Test. The Welch-Test is predominantly intended for use with two samples having heterogeneous variances. However, Kubinger, Rasch & Moder (2009) could show that even in the case of homogeneous variances the power of the Welch-Test is equal to the power of the Student's t-Test. In case of heterogenious variances, the power of the Welch-Test is even higher than the power of the Student's t-Test. Due to these facts, the authors advice in cases of mean comparisons to use the Welch-Test even in cases without normal distributions or heterogenious variances because of the power of the Welch-Test. (Results show that the Mann-Whitney-U-Test is inferior to the Welch-Test in any cases and therefore can be neglected.) Further information can be found by Kubinger, Rasch & Moder (2009).

The Welch-Test calculations were done with the formula in figure 3-5 which expresses the *t*-distributed test statistic *t*. A comparison with the calculated *t*-value and the critical *t*-value provides information if there are significant differences in means between the two sample pairs. (If the empirical *t*-value is bigger than the critical *t*-value, then it can be assumed that there is a significant difference between the two compared groups.) The type-I-risk was set on an α-level of 0.05 for all Welch-Test calculations.

While the Welch-Test calculations were partly done with the statistic software program *PASW 18* (The means and the variances of the compared samples was calculated with *PASW 18*. Then those parameters were inserted to the formulas.), the critical *t*-value was calculated with the program *R 2.10.0.*

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{{s_1}^2}{n_1} + \frac{{s_1}^2}{n_2}}}$$

t:	t- distributed test statistic
$\bar{x}_1$:	mean of sample 1
$\bar{x}_2$:	mean of sample 2
s_1^2:	variance of sample 1
s_2^2:	variance of sample 2
n_1:	size of sample 1
n_2:	size of sample 2

Figure 3-5: Formula of the Welch-Test.

To compare the empirical *t*-value with the critical *t*-value, it is a necessary condition to calculate the degrees of freedom. The formula is presented in figure 3-6.

$$df = \frac{\left(\frac{s_1{}^2}{n_1} + \frac{s_2{}^2}{n_2}\right)^2}{\frac{\left(\frac{s_1{}^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2{}^2}{n_2}\right)^2}{n_2 - 1}}$$

df:	degrees of freedom
s_1^2:	variance of sample 1
s_2^2:	variance of sample 2
n_1:	size of sample 1
n_2:	size of sample 2

Figure 3-6: Formula to calculate the degrees of freedom to compare the critical *t-value* with the empirical *t-value* of the Welch-Test calculations.

All of the Welch-Test results, which are used for the hypotheses checks, will be presented in tables including the variable names of the compared sample pairs, the particular sample sizes (n), the variances and the empirical and the critical *t*-values.

3.8.3 General important facts about the results

The next chapter (chapter 4) is going to describe the study results. However, it is considered as useful to discuss the basic facts about the analysing methods first. So, in this coherence, two different ways to check if the items fit the Rasch model assumptions will be presented briefly:

- <u>Likelihood ratio test of Andersen (short: LRT)</u>: The LRT is a chi-square distributed model check that proves the overall model fit (Andersen, 1973).

  The partition criterion: country (USA *versus* Austria), was only used for the Rasch model analyses of the whole "censored" sample (see chapter 4.1). Due to the fact that no prove of equivalency between the two countries *via* Rasch model analyses was possible, the results within the whole sample cannot be used for serious scientific postulations. It was just the aim to explore and to see whether there are any differences between the two countries.

The commonly used partition criterion "age" was not considered to be a good partition criterion for this sample, because the students were almost the same age (A great number of students were in their twenties, only a few were older than thirty years which would lead to an unequal amount of students in the sub-samples.).

- Graphical model check: As already mentioned in chapter 3.4.1, the graphical model check shows the specific item fits in an orthogonal coordinate system.

It has to be mentioned that the Austrian sample which counts 225 (FRRT) / 221 (Syllogisms 2009) test takers is relatively small for Rasch model analyses. This implies that the power of the Likelihood ratio test was accordingly little. This is the reason why graphical model checks were of higher importance to decide about model fit or misfit than usual. (Results of the Wald-Test were additionally beholded in cases of undetermined item positions.)

4 Results

In this chapter the results of the data collection in Austria and the USA are going to be described. To give the reader a clear overview of the results, this chapter is divided into five sub-chapters. The first chapter (4.1) will point out the problems concerning the American data and the conclusions that could be deduced from those results. Chapter 4.2 will give some general important facts about the Rasch model analyses and right after, the outcomes of the Rasch Model analyses (*a-proiri* and *a-posteriori*) of the Austrian sample will be shown in chapter 4.3. Chapter 4.4 will handle the intercultural results of the Rasch model analyses. At last the examination of the study hypotheses will be shown in chapter 4.5.

4.1 Important remarks about the data

While the data collection in Austria took place from June 20th to July 10th 2009, the data collection in the USA was split into two separate testing times. The first part took place from September 11th to September 18th and the second part was from October 28th to October 31st 2009.

A closer look at the data concerning the editing times of the verbal test battery showed that about 50 % of the test takers within the U.S. American sample needed a very short time to edit the test.

Figure 4-1 shows the distribution of the FRRT editing times in comparison of the two countries. It can be seen that the peak of the distribution of the U.S. sample is further left than within the Austrian sample. The differences in the editing times between the two samples are even stronger in the second sub-test, the Syllogisms 2009. The distribution of the editing times of this sub-test (which had to be executed after the FRRT) differs strongly between the two countries. Figure 4-2 shows the distribution of the editing times of the Syllogisms 2009 items within both samples.

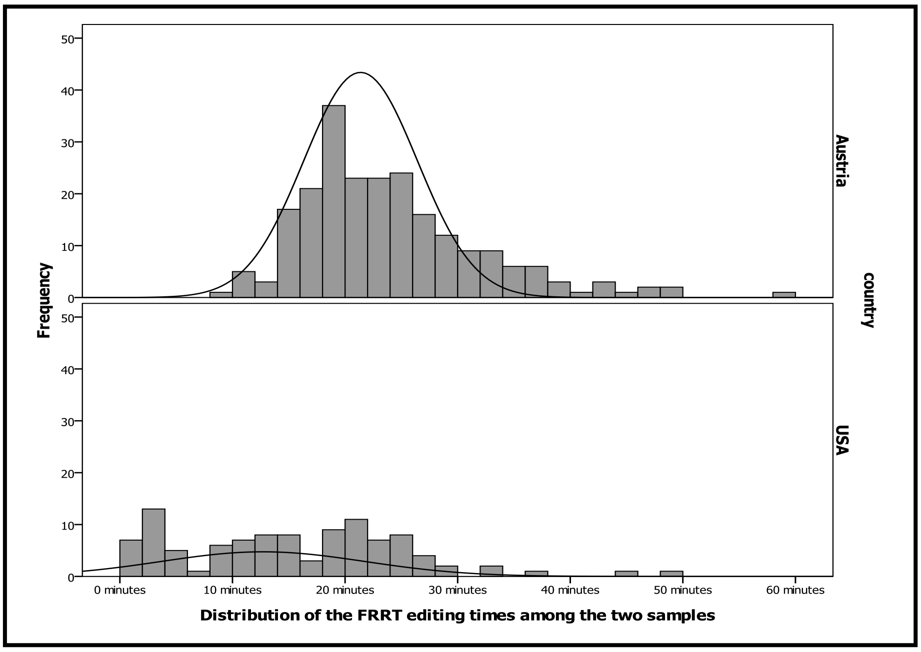


Figure 4-1: Distribution of the FRRT editing times. The upper distribution shows the editing times within the Austrian sample, while the distribution further low represents the U.S. editing times. It can be seen that the U.S. American data curve is more left than the Austrian data curve which indicates shorter editing times within the U.S. American sample.

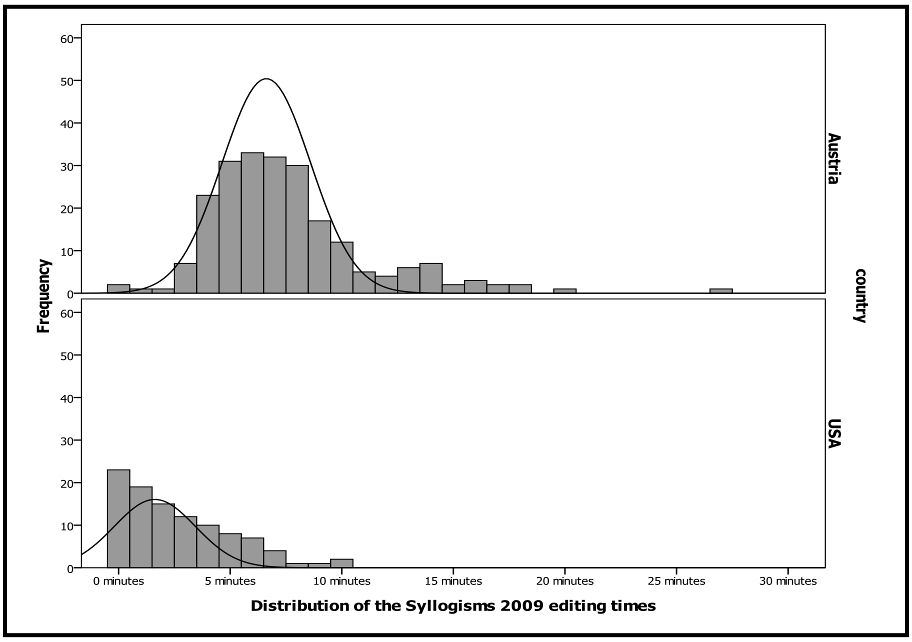


Figure 4-2: Distribution of the Syllogisms 2009 editing times. The upper distribution shows the editing times within the Austrian sample, while the distribution further low represents the U.S. editing times. It can be seen that the U.S. American data curve is far more left than the Austrian data curve which indicates shorter editing times within the U.S. American sample.

According to the results above, not all of the results of the American students were used for further analyses. Only the results of those test takers who needed longer to execute the test than the lower 5 % percentile of the Austrian editing times were kept for further investigations. A closer look at the differences in the editing times between the two countries will be done in chapter 4.4.3.1. Because of the fact that some data had to be deleted, the American sample shrank connotatively. Table 4-1 gives an overview of the remaining and the deleted amount of data within the U.S. American sample.

Table 4-1
The 5 % percentile cut off lead to a diminishment of the U.S. American sample. 54.8 % of the FRRT and 35.3 % of the Syllogisms 2009 data remained for further analyses. "

sub-test	percentile level	n	percentage of the U.S. sample
FRRT	> 5 %	57	54.8 %
	< 5 %	47	45.2 %
Syllogisms 2009	> 5 %	37	35.3 %
	< 5 %	65	64.7 %

The "censorship" of the U.S. American sample entailed a small number of remaining data, which made it impossible to calculate Rasch model analyses with the U.S. American sample (A calculation with such a small amount of data would lead to very imprecise and contorted results.). However, this would have been an important condition to check the equivalency of the two test battery versions. (Tests of equivalency contain the Rasch model analyses of each sample followed by a Rasch model analysis of the whole sample.) So for this thesis, no test of equivalency was possible. Only Rasch model analyses with the Austrian sample could have been done and so it was possible to see the German test theoretical qualities of the items of the FRRT and the Syllogisms 2009. Because of this fact, a statistical comparison of the two countries *via* complex analysing models was impossible because differences between the two countries could have been caused by different constructs that were measured with the two different verbal test battery versions. Without a test of equivalency it was impossible to see if both versions of the verbal test battery measure the same construct (verbal reasoning). This is the reason why complex statistical analyses were not possible to be calculated with the U.S. American sample. However, to get a slight impression about the English items, Rasch model analyses of the whole sample[4] were calculated just to get an approximate look at the item positions.

[4] excluding the results of the test takers that needed less time to edit the sub-tests of the verbal test battery than the lower 5% percentile of the Austrian sample.

4.2 Results of the Rasch model analyses within the Austrian sample

In this chapter, the results of the Rasch model analyses within the Austrian sample are presented. For each sub-test, the results of the Likelihood-Ratio-Tests (short "LRT") of Andersen as well as graphical model checks are given. The α-level for the LRT was always set on 0.05.

4.2.1 Results of the Family Relations Reasoning Test

In this chapter, the results of the Rasch Model analyses of the FRRT (k= 60 items) are presented. The results showed that some items of the FRRT had to be deleted due to complete correct response patterns. The items which had to be excluded of the Rasch model analyses were: *Item 1, Item 9 and Item 13*. For those items which had been solved by every test taker no item easiness parameters could have been estimated (They do not provide any information.). The following sub-chapters present the results of the Rasch model analyses sorted by the particular partition criteria. The analyses show that in each criterion, particular item parameters could not be estimated due to extreme response behaviour of the test takers. This is the reason why there is a different amount of remaining items in each partition criterion. Though some item easiness parameters could have not been calculated in every partition criterion – due to an extreme response pattern in one of the sub-groups – no item showed extreme response patterns in all of the partition criteria and therefore all of the remaining items were kept for the hypotheses testings.

The absolute number of test takers in each sub-sample can be found in appendix 9.1. All the items of the FRRT are listed up in appendix 9.3.

4.2.1.1 Partition criterion: raw score

The partition criterion *raw score* is an internal criterion. The examined sample (n= 225) was parted by the median (md= 16) of the score distribution into the two groups “low raw score” versus “high raw score”. Test takers who achieved the median score were allocated to the group “low raw score”.

The likelihood ratio test of Andersen showed no significant result on an α-level of 0.05. However, 28 of 57 items had to be excluded due to inappropriate response patterns within the sub-samples. All the remaining items fit the Rasch Model assumptions. Table 4-2 presents the results of the LRT of Andersen, while figure 4-3 shows the graphical model check.

Table 4-2
Partition criterion *raw score* (high raw score *versus* low raw score departed by the median) 28 of 57 items had to be excluded due to inappropriate response patterns within the sub-samples.

	χ^2(LRT)	df	χ^2(critical)	partition criterion
a-priori	31,877	28	41,33714	raw score

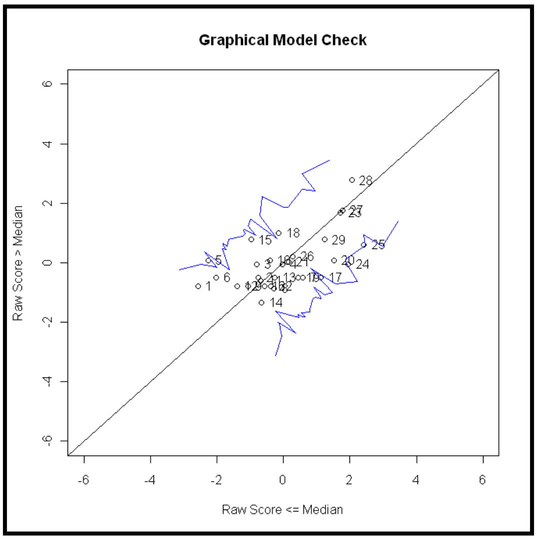


Figure 4-3: Graphical model check of 29 of all 60 FRRT items within the partition criteria *raw score*. 31 item easiness parameters could not be estimated because the items showed an extreme solution frequency in at least one sub-sample. The graphical model check shows the two sub-samples “high raw score” *versus* “low raw score” departed by the median and located in an orthogonal coordinate system.

4.2.1.2 Partition criterion: sex

The partition criterion *sex* is an external criterion. The examined sample (n= 225) was parted by the gender into the two groups “male test takers” *versus* “female test takers”.

Again, the Likelihood ratio test of Andersen did not show a significant result. However, 10 of 57 items had to be excluded due to extreme response behaviour of the test takers. . All of the remaining items fit the Rasch model assumptions.

Table 4-3
Partition criterion: *sex* (male *versus* female students). 10 of 57 items had to be deleted due to inappropriate response patterns within the sub-samples.

	χ^2(LRT)	df	χ^2(critical)	partition criterion
a-priori	33,95	46	62.82962	sex

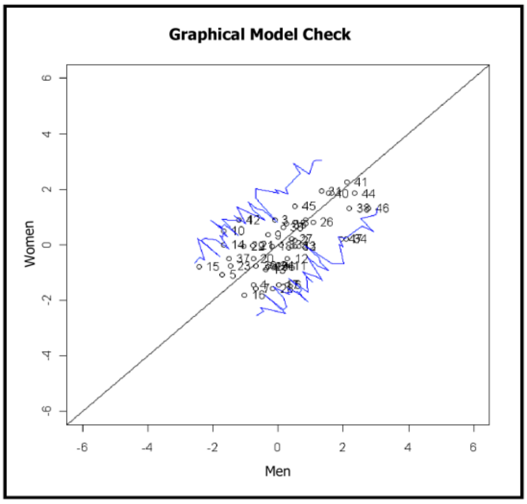


Figure 4-4: Graphical model check of 47 of all 60 items within the partition criteria *sex*. 13 item easiness parameters could not be estimated because the items showed an extreme solution frequency in at least one sub-sample. The graphical model check shows the two sub-samples "men" *versus* "women" departed by the gender and located in an orthogonal coordinate system.

4.2.1.3 Partition criterion: response format

The partition criterion *response format* is an external criterion. The examined sample (n= 225) was parted by the response format which the test taker used to edit the test. The two sub-groups are "sequential response format 1 out of 5" *versus* "multiple choice format 1 out of 5".

Also, for this partition criterion, the LRT of Andersen did not show a significant result. However, 9 of 57 items had to be deleted due to extreme response behaviour of the test takers. All of the remaining items fit the Rasch model assumptions.

Table 4-4
Partition criterion: format (multiple choice "1 out of 5" *versus* sequential response format "1 out of 5"). 9 of 57 items had to be deleted due to inappropriate response patterns within the sub-samples.

	χ^2(LRT)	df	χ^2(critical)	partition criterion
a-priori	49,767	47	64,00111	format

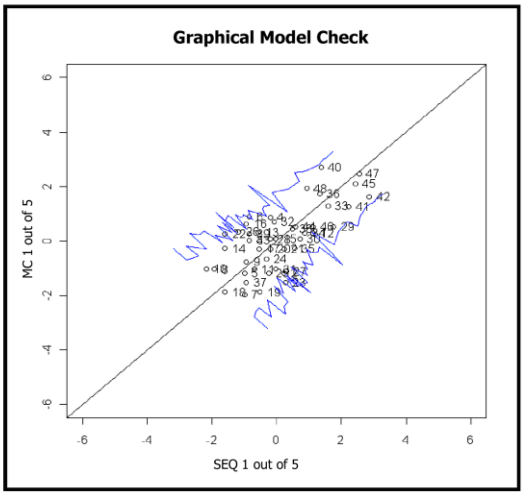


Figure 4-5: Graphical model check of 48 of all 60 items within the partition criteria *response format*. 12 item easiness parameters could not be estimated because the items showed an extreme solution frequency in at least one sub-sample. The graphical model check shows the two sub-samples: sequential response format "1 out of 5" (short: SEQ "1 out of 5") *versus* multiple choice format "1 out of 5" (short MC 1 out of 5) departed by the response format and located in an orthogonal coordinate system.

4.2.2 Results of the Syllogisms 2009

The following chapter presents the results of the Rasch model analyses of the Syllogisms (k= 10 items). Like for the FRRT, the results of the Likelihood Ratio Test of Andersen and the graphical model checks will be presented. Due to the fact, that 4 out of 10 items had to be excluded, the *a-priori* and the *a-posteriori* results of the Rasch model analyses of each partition criterion are presented. The item deletion process was step by step after looking at all of the partition criteria (!). The following items were deleted: *item 1* ("syl 1"), *item 9* ("syl 16"), *item 2* ("syl 4") and *item 6* ("syl 13"). The names in brackets are the original names of the Syllogisms 2009 items. Table 4-5 shows the stepwise item exclusion process of the 4 misfitting Syllogisms 2009 items.

Table 4-5
Overview of the item exclusion process during the Rasch model analyses for the sub-test Syllogisms 2009. Items which had been too far away of the 45° line were excluded step by step.

		partition criteria		
stepwise item exclusion	**item name**	**score**	**sex**	**response format**
1st step	syl 1	x		x
2nd step	syl 16			x
3rd step	syl 4	x		
4th step	syl 13			x

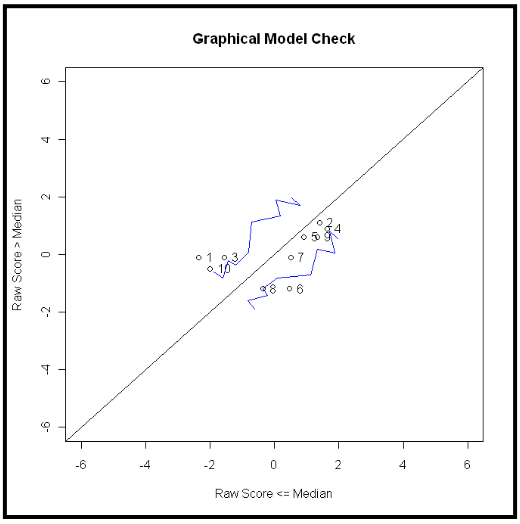
4.2.2.1 Partition criterion: raw score

The partition criterion *raw score* is an internal criterion. The examined sample (n= 221) was parted by the median (md= 8) of the score distribution into the two groups “low raw score” *versus* “high raw score”. Test takers who achieved the median score were allocated to the group “low raw score”.

The LRT of Andersen showed a significant result (α= 0.05). After the deletion of the four misfitting items, the Likelihood Ratio Test of Andersen does not show a significant results anymore and also the graphical model checks show only items within the confidence bands (close to the 45° diagonal). Table 4-6 presents the *a-priori* and the *a-posteriori* results of the Rasch model analyses. Figure 4-6 (*a-priori*) and figure 4-7 (*a-posteriori*) show the graphical model checks.

Table 4-6
Partition criterion *raw score* (high raw score *versus* low raw score departed by the median). The a-priori and the a-posteriori results can be seen.

	χ^2(LRT)	**df**	**χ^2(critical)**	**partition criterion**
a-priori	22,786	9	16,91898	raw score
a-posteriori	2,829	5	11,07050	raw score



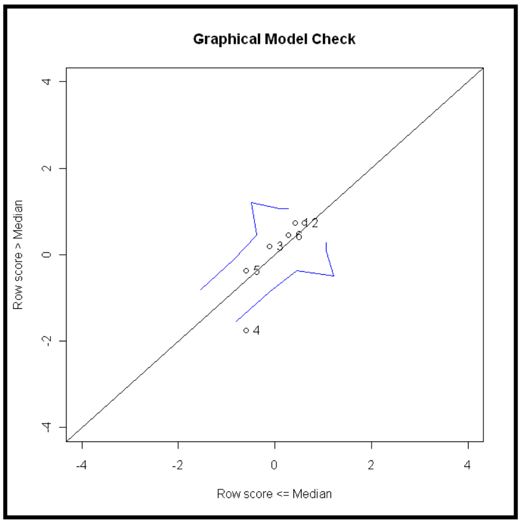


Figure 4-6: *A-priori* graphical model check of all 10 Syllogisms 2009 items within the partition criterion *score.* The graphical model check shows the two sub-samples “low row score” versus “high row score” departed by the median and located in an orthogonal coordinate system.

Figure 4-7: A-posteriori model check of the Syllogisms 2009 items within the partition criterion *raw* score and after the deletion of 4 items. The graphical model check shows the two sub-samples “low raw score” and “high raw score” departed by the median and located in an orthogonal coordinate system.

Because of the ambiguous position of item number 4 (*syl 10*) in the coordinate system, a second graphical model check with the partition criterion *raw score* - but this time with confidence ellipses – was drawn. Figure 4-8 shows the graphical model check with the confidence ellipses. Because of the fact the confidence ellipse of item 4 hits the 45° diagonal item 4 can be kept for further analyses.

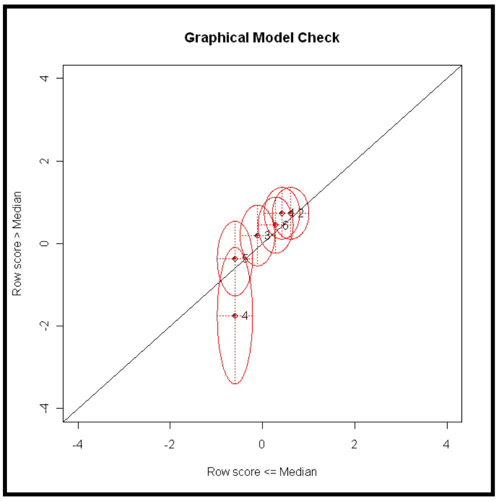


Figure 4-8: A-posteriori model check of the partition criterion *raw score* but this time with confidence ellipses to have a better few at item 4. The graphical model check shows the two sub-samples “low raw score” and “high raw score” departed by the median. The confidence ellipse of item 4 cut with the diagonal and therefore the item was kept for further analyses.

4.2.2.2 Partition criterion: sex

The partition criterion *sex* is an external criterion. The examined sample (n= 221) was parted by the gender into the two groups “male test takers” *versus* “female test takers”. While table 4-8 presents the results of the LRT of Andersen (*a- priori* and *a-posteriori*), figure 4-9 (*a-priori*) and figure 4-10 (*a-posteriori*) show the results of the graphical model checks.

Table 4-8
Results of the LRTs of Andersen within the partition criterion *sex* (male *versus* female students).

	χ^2(LRT)	**df**	**χ^2(critical)**	**partition criterion**
a-priori	5,21	9	16,91898	sex
a-posteriori	3,926	5	11,07050	sex

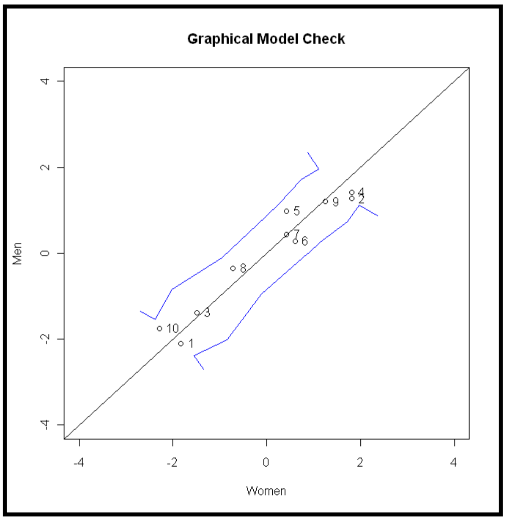


Figure 4-9: *A-priori* graphical model check of all 10 Syllogisms 2009 items within the partition criterion *sex.* The graphical model check shows the two sub-samples “women” and “men” departed by the gender and located in an orthogonal coordinate system.

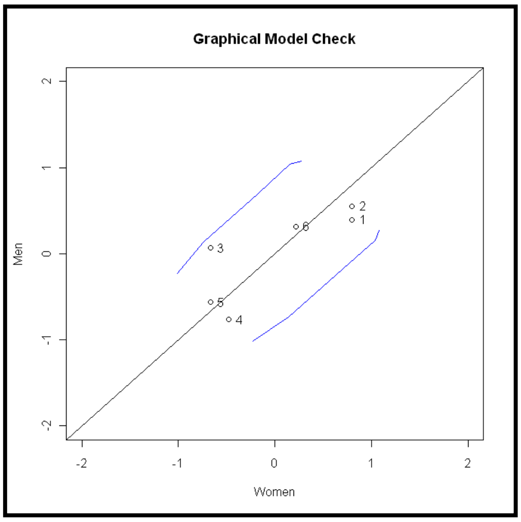


Figure 4-10: A-posteriori model check of the Syllogisms 2009 items within the partition criterion *sex* and after the deletion of 4 items. The graphical model check shows the two sub-samples “women” and “men” departed by the gender and located in an orthogonal coordinate system.

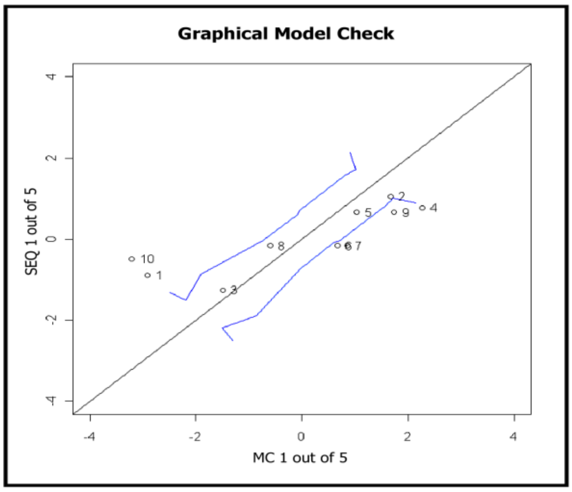
4.2.2.3 Partition criterion: response format

The partition criterion *response format* is an external criterion. The examined sample (n= 221) was parted by the response format which the test taker used to edit the sub-test. The two sub-groups are "sequential response format 1 out of 5" *versus* "multiple choice format 1 out of 5".

Table 4-9 presents the results (a-priori and a-posteriori) of the LRTs of Andersen. The graphical model checks can be seen in figure 4-11 (*a-priori*) and figure 4-12 (*a-posteriori*).

Table 4-9
Partition criterion *response format* (multiple choice format "1 out of 5" *versus* sequential response format "1 out of 5"). The a-priori and the a-posteriori results can be seen.

	χ^2(LRT)	df	χ^2(critical)	partition criterion
a-priori	56,909	9	16,91898	format
a-posteriori	6,279	5	11,07050	format



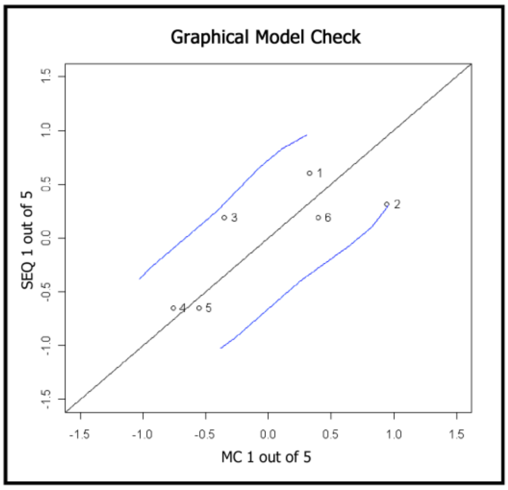


Figure 4-11: *A-priori* graphical model check of all 10 Syllogisms 2009 items within the partition criterion *response format.* The graphical model check shows the two sub-samples: sequential response format "1 out of 5" (short: SEQ 1 out of 5) and multiple choice format "1 out of 5" (short: MC 1 out of 5) departed by the response format and located in an orthogonal coordinate system.

Figure 4-12*:* *A-posteriori* model check of the Syllogisms 2009 items within the partition criterion *response format* and after the deletion of 4 items. The graphical model check shows the two sub-samples "multiple choice format 1 out of 5" (short: MC 1 out of 5) and "sequential response format 1 out of 5" (short: SEQ 1 out of 5) departed by the format and located in an orthogonal coordinate system.

4.2.3 Recapitulatory examination of the Rasch model results within the Austrian sample

Concerning the FRRT, three item parameters of all 60 items were not possible to be estimated due to an extreme response behavior of the test takers. The following items had to be excluded because they were solved by every test taker: *Item 1, Item 9 and Item 13.*

All of the remaining FRRT items fitted the assumption of the Rasch model and were kept for further analyses.

Concerning the sub-test Syllogisms 2009, it turned out that 4 out of 10 items had to be deleted due to a misfit of the Rasch model assumptions. While the partition criterion "sex" was inconspicuous, the other two partition criteria "raw score" and "response format" indicated deviations from the Rasch model assumptions. After the deletion of the four items (*syl 1, syl 16, syl 4, syl 13*), the Likelihood Ratio Test of Andersen and the graphical model checks showed no deviations of the Rasch model assumptions anymore.

4.2.4 Estimated item parameters of the Austrian Syllogisms 2009 and FRRT items

Subsequently to the results of the Likelihood ratio tests and the graphical model checks, this chapter will present the item easiness parameters with the standard errors and the solution frequencies in percent of all the items which fit the Rasch model. All the item parameters within one group are standardized on the sum zero. Three item easiness parameters could not be estimated due to extreme solution frequencies of these items (The items which item parameters could not be estimated are: *frrt1, frrt9, frrt13*). The results of the other items which item easiness parameters could be estimated are presented in table 4-10 (FRRT) and 4-11 (Syllogisms 2009). An overview of all the items of the FRRT and the Syllogisms 2009 can be found in appendix 9.3 and appendix 9.4.

Table 4-10
Schedule of all estimated item easiness parameters and standard errors of those items that are conform with the Rasch model.

Family Relations Reasoning Test

Item name	Item Easiness Parameter	Std. Error	Relative Solution Frequency MC/ SEQ in percent	Item name	Item Easiness Parameter	Std. Error	Relative Solution Frequency MC/SEQ in percent	Item name	Item Easiness Parameter	Std. Error	Relative Solution Frequency MC/SEQ in percent
frrt2	1.671	0.752	96.4/ 96.7	frrti23	0.899	0.568	100/ 86.7	frrti42	-0.562	0.298	87.2/ 81.5
frrti3	0.339	0.526	89.3/ 92.9	frrti24	1.671	0.752	96.4/ 96.7	frrti43	-1.626	0.383	71.4/ 67.9
frrti4	0.225	0.467	92.9/ 83.3	frrti25	0.903	0.419	91.1/ 96.4	frrti44	-0.682	0.414	82.1/ 78.6
frrti5	-0.627	0.439	92.9/ 75.0	frrti26	0.140	0.528	92.9/ 89.3	frrti45	-0.245	0.420	78.6/ 86.7
frrti6	0.929	0.638	96.4/ 92.9	frrti27	-0.075	0.470	89.3/ 85.7	frrti46	-1.772	0.385	60.7/ 51.9
frrti7	-0.377	0.409	92.9/ 70.0	frrti28	0.225	0.467	96.4/ 80.0	frrti47	1.190	0.754	96.4/ 96.4
frrti8	1.374	0.754	96.4/ 96.4	frrti29	0.408	0.572	89.3/ 96.4	frrti48	-0.608	0.189	82.1/ 77.9
frrti10	2.384	0.720	100/ 96.5	frrti30	0.382	0.352	91.1/ 89.7	frrti49	0.054	0.448	92.9/ 80.0
frrti11	1.203	0.636	96.4/ 92.6	frrti31	0.602	0.571	92.9/ 92.9	frrti50	0.740	0.638	100/ 89.3
frrti12	0.740	0.638	96.4/ 92.2	frrti32	0.882	0.570	100/ 85.2	frrti52	-2.227	0.376	75/ 42.9
frrti14	0.244	0.333	98.2/ 92.9	frrti33	0.048	0.455	92.9/ 77.8	frrti53	0.408	0.495	100/ 77.8
frrti15	-0.087	0.496	100/ 78.6	frrti34	0.225	0.467	82.1/ 93.3	frrti54	-1.782	0.364	50.0/ 63.3
frrti16	0.625	0.526	89.3/ 92.6	frrti35	-0.087	0.496	82.9/ 85.7	frrti56	-2.322	0.365	46.4/ 53.6
frrti17	-0.807	0.405	78.6/ 78.6	frrti36	1.927	1.025	100/ 96.4	frrti58	0.225	0.467	92.9/ 83.3
frrti18	-0.096	0.319	92.9/ 81.0	frrti37	-1.251	0.382	67.9/ 75.0	frrti60	-0.682	0.414	85.7/ 75.0
frrti19	0.637	0.523	96.4/ 86.7	frrti38	-0.633	0.406	71.4/ 81.5	frrti61	-2.322	0.374	46.4/ 46.7
frrti20	-0.251	0.429	82.1/ 81.5	frrti39	0.218	0.472	82.1/ 92.6	frrti62	-0.926	0.398	78.6/ 75.0
frrti21	0.244	0.365	96.4/ 78.6	frrti40	1.374	0.754	100/ 92.9	frrti63	-2.730	0.387	53.6/ 46.4
frrti22	0.244	0.365	96.4/ 78.5	frrti41	0.602	0.571	100/ 85.7	frrti64	-1.674	0.384	67.9/ 48.1

Table 4-11
Schedule of all estimated item easiness parameters and standard errors of those items that are conform with the Rasch model.

Syllogisms 2009

Item name	Item Easiness Parameter	Std. Error	Relative Solution Frequency MC/ SEQ in percent	Item name	Item Easiness Parameter	Std. Error	Relative Solution Frequency MC/ SEQ in percent	Item name	Item Easiness Parameter	Std. Error	Relative Solution Frequency MC/ SEQ in percent
syl 3	-0.478	0.164	71.6/ 43.8	syl 8	0.077	0.169	78.0/ 52.7	syl 11	0.587	0.180	88.1/ 55.4
syl 6	-0.602	0.164	76.1/ 35.7	syl 10	0.705	0.183	88.1/ 58.0	syl 14	-0.288	0.165	78.0/ 42.9

4.3 Intercultural results of the Rasch model analyses

As already described in chapter 4.2, no Rasch model analyses of the English version of the verbal test battery were possible because of the small size of the “censored” U.S. American sample. However, to receive a first impression of the psychometric qualities of the items within the whole sample, Rasch model analyses with the international sample (n_{FRRT}: 57 + 225= 282; $n_{Syllogisms\ 2009}$: 36 + 223= 259) were calculated. The partition criteria: *raw score, sex* and *response format* were chosen again and additionally the partition criterion *country* was used.

4.3.1 Results of the Family Relations Reasoning Test

Rasch model analyses within the whole sample showed that one item easiness parameter could not be estimated due to an extreme solution frequency. *Item 9* of the FRRT was solved by every test taker and therefore did not provide any information. All the other items fit the Rasch model assumptions. Some of the item parameters could not be estimated in all of the sub-samples due to an extreme response pattern within the sub-groups. This is why the number of FRRT items is different within each partition criterion. Concerning the Rasch model analyses of the mixed sample, only LRT results were used to describe the data. This was done out of economic reasons because the results could not be used for further analyses, but only to provide first information about the items in an international context.

4.3.1.1 Partition criterion: raw score

The partition criterion *raw score* is an internal criterion. The examined sample (n= 282) was parted by the median (md= 15) of the score distribution into the two groups "low raw score" versus "high raw score". Test takers who achieved the median score were allocated to the group "low raw score".
The likelihood ratio test of Andersen showed no significant result on an α-level of 0.05 (table 4-12). However, 26 of 59 items had to be excluded due to inappropriate response patterns within the sub-samples. All the remaining items fit the Rasch Model assumptions.

Table 4-12
Partition criterion: *raw score* (high raw score *versus* low raw score departed by the median) 26 out of 59 items had to be deleted due to inappropriate response patterns within the sub-samples.

	χ^2(LRT)	df	χ^2(critical)	partition criterion
a- priori	23.32	32	46.19426	raw score

4.3.1.2 Partition criterion: sex

The partition criterion *sex* is an external criterion. The examined sample (n= 282) was parted by the gender into the two groups "male test takers" *versus* "female test takers".

7 of 59 items had to be deleted due to inappropriate response patterns within the sub-samples. All of the remaining items fit the Rasch Model assumptions. Table 4-13 presents the results of the LRT of Andersen.

Table 4-13
Partition criterion: *sex* (male *versus* female students). 7 out of 59 items had to be deleted due to inappropriate response patterns within the subgroups.

	χ^2(LRT)	df	χ^2(critical)	partition criterion
a- priori	54.089	52	69.83216	sex

4.3.1.3 Partition criterion: response format

The partition criterion *response format* is an external criterion. The examined sample (n= 282) was parted by the response format which the test taker used to edit the test. The two sub-groups are "sequential response format 1 out of 5" *versus* "multiple choice format 1 out of 5".
4 of 59 items had to be excluded due to extreme response behaviour of the test takers. All of the remaining items fit the Rasch model assumptions. Table 4-14 presents the results of the LRT of Andersen.

Table 4-14
Partition criterion: response format (multiple choice "1 out of 5" *versus* sequential response format "1 out of 5"). 4 out of 59 items had to be deleted due to inappropriate response patterns within the subgroups.

	χ^2(LRT)	df	χ^2(critical)	partition criterion
a- priori	65.035	54	72.15322	format

4.3.1.4 Partition criterion: country

The partition criterion *country* is an external criterion. The examined sample (n= 282) was parted by the country in which the test taker executed the test. So, the two sub-groups are "Austria" *versus* "United States of America". 2 of 59 items had to be excluded due to an extreme response behaviour within one of the sub-groups. All of the remaining items fit the Rasch model assumptions. Table 4-15 presents the results of the LRT of Andersen.

Table 4-15
Partition criterion: “country” (Austria *versus* USA). 2 out of 59 items had to be excluded due to inappropriate response patterns within the subgroups.

	χ^2(LRT)	df	χ^2(critical)	partition criterion
a- priori	69.619	56	74.46832	country

4.3.2 Results of the Syllogisms 2009

Rasch model analyses of the Syllogisms 2009 items within the mixed (international) sample show, that unlike the Austrian sample, this time five out of ten items had to be excluded because they did not fit the Rasch model assumptions. Table 4-16 gives an overview of the results within the partition criterion *raw score*, while table 4-17, table 4-18 and table 4-19 will show the results of the criteria *sex*, *response format* and *country*.

4.3.2.1 Partition criterion: raw score

The partition criterion *raw score* is an internal criterion. The examined sample (n= 258) was parted by the median (md= 4) of the score distribution into the two groups “low raw score” *versus* “high raw score”. Test takers who achieved the median score were allocated to the group “low raw score”. The LRT of Andersen showed a significant result (α= 0.05). After the deletion of the five misfitting items, the LRT of Andersen does not show significant results anymore. Table 4-16 presents the results of the *a-priori* and the *a-posteriori* results of the LRT of Andersen.

Table 4-16
Partition criterion *raw score* (high raw score *versus* low raw score departed by the median). The a-priori and the a-posteriori results can be seen.

	χ^2(LRT)	df	χ^2(critical)	partition criterion
a- priori	17.823	9	16.91898	raw score
a- posteriori	5.791	4	9.487729	raw score

4.3.2.2 Partition criterion: sex

The partition criterion *sex* is an external criterion. The examined sample (n= 259) was parted by the gender into the two groups "women" *versus* "men". First, the LRT of Andersen showed a significant result (α= 0.05). After the deletion of the five misfitting items, the LRT does not show significant results anymore. Table 4-17 presents the results of the *a-priori* and the *a-posteriori* results of the LRT of Andersen.

Table 4-17
Partition criterion *sex* (male *versus* female students). The a-priori and the a-posteriori results can be seen.

	χ^2(LRT)	df	χ^2(critical)	partition criterion
a- priori	9.81	9	16.91898	sex
a- posteriori	0.385	4	9.487729	sex

4.3.2.3 Partition criterion: response format

The partition criterion *response format* is an external criterion. The examined sample (n= 259) was parted by the gender into the two groups "multiple choice 1 out of 5" *versus* "sequential response format 1 out of 5". After the deletion of the five misfitting items, the LRT does not show significant results anymore. Table 4-18 presents the results of the *a-priori* and the *a-posteriori* results of the LRT of Andersen.

Table 4-18
Partition criterion *response format* (multiple choice "1 out of 5" *versus* sequential response format "1 out of 5"). The a-priori and the a-posteriori results can be seen.

	χ^2(LRT)	df	χ^2(critical)	partition criterion
a- priori	62.808	9	16.91898	response format
a- posteriori	6,279	4	9.487729	response format

4.3.2.4 Partition criterion: country

The partition criterion *country* is an external criterion. The examined sample (n= 259) was parted by the country in which the test takers executed the verbal test battery. The two sub-samples are "Austria" *versus* "United States of America". After the deletion of the five misfitting items, the LRT does not show significant results anymore. Table 4-19 presents the results of the *a-priori* and the *a-posteriori* results of the LRT of Andersen.

Table 4-19
Partition criterion *country* (Austria *versus* USA). The a-priori and the a-posteriori results can be seen.

	χ^2(LRT)	df	χ^2(critical)	partition criterion
a- priori	3.152	9	16.91898	country
a- posteriori	6.601	4	9.487729	country

The deletion of 5 Syllogisms 2009 items indicates that 50 % of the items had to be excluded due to a misfit of the Rasch model assumptions. This partition criterion which indicated most often a misfit of the Rasch model assumptions was "response format". The stepwise item exclusion process can be seen in table 4-20.

Table 4-20
Overview of the item exclusion process during the Rasch model analyses for the sub-test Syllogisms 2009. Items which had been too far away of the 45° line were excluded step by step.

		partition criteria			
stepwise item exclusion	item name	score	sex	response format	country
1st step	syl 16			x	x
2nd step	syl 1			x	
3rd step	syl 13			x	
4th step	syl 4	x		x	x
5th step	syl 6			x	x

Though the results of the analyses within the whole sample cannot be used to postulate any conclusions, they show a direction where further research studies could tie in with. It seems that the items of the FRRT measure in one dimension, but many of them are too easy for the test takers (student sample). About the Syllogisms 2009 items it can be said that half of the items had to be deleted because they did not only measure in one dimension. Especially the partition criteria "response format" indicated a misfit of the items. Further Rasch model analyses with a bigger international sample would be of great interest – to get more accuarate information about the items of the verbal test battery.

4.4 Results of the hypotheses testings

In this chapter, the hypotheses testings will be illustrated. Each sub-chapter will handle one of the study hypotheses followed by the particular statistical analysing models to prove the hypothesis. At the end of each sub-chapter the results will be presented.

4.4.1 Results of the first study hypothesis

The first study hypothesis comprehends the focus on differences in the solution frequency between paired items:

H_0 (1): There is no mean difference of the solution frequency between paired items (same content but different response formats) with the multiple choice format “1 out of 5” and the sequential response format “1 out of 5”.

$$\mu 1 = \mu 2$$

H_1 (1): There is a mean difference of the solution frequency between paired items (same content but different response formats) with the multiple choice format “1 out of 5” and the sequential response format “1 out of 5”.

$$\mu 1 \neq \mu 2$$

To prove if there are differences in mean, only those items which fitted the Rasch model were used.

4.4.1.1 Austrian sample

Rasch model analyses of the Austrian sample showed that concerning the FRRT, three items had been solved by every test taker and therefore the particular item easiness parameters could not be estimated. Those items were excluded of further analyses – as explained in chapter 3.4.1. In particular, the following items had to be excluded: *frrt item 1, frrt item 9, frrt item 13.* An overview of the item exclusions can be found in appendix 9.5.

After having taken three items out of the FRRT item pool, the remaining items were used to explore if there are differences in their solution frequencies between the two response formats MC "1 out of 5" and SEQ "1 out of 5". Because of the fact that every item was designed with both response formats, it was possible to explore if there are differences in the solution frequencies depending on only the response formats. However, because of the fact that the FRRT items were allocated to four groups and that three items (items 1, 9 and 13) had to be excluded after the Rasch model analyses, one of the four FRRT parallel groups got a different amount of maximum raw score. Because of this fact, it was not possible to compare the absolute solution frequencies or raw scores of the students (Students who executed the FRRT group 3 could only achieve a maximum raw score of 15, while students who executed one of the other groups could still achieve a maximum amount of 18 points.). The amount of students who executed group 3 in the multiple choice format "1 out of 5" was bigger (n= 28) then the group of students who executed group 3 of the FRRT in the sequential response format "1 out of 5" (n= 27). So, a comparison of the two formats with the absolute solution frequencies achieved by the students would have lead to a bias. Table 4-21 gives an overview of the distribution of the four FRRT parallel groups and figure 4-13 presents the formula to calculate the relative solution frequencies:

Table 4-21
Overview of the distribution of the four FRRT parallel groups. It can be seen that in group three more students executed the test in the multiple choice format (short "mc") "1 out of 5" than in the sequential response format (short "seq") "1 out of 5". Because of this reason, the relative solution frequencies were used instead of the absolute raw scores of the students.

FRRT groups	Format	Frequency	Percent	Executed items *a-priori*	Executed item *a-posteriori*
	MC	28	50.0	18	18
FRRT group 1	SEQ	28	50.0	18	18
	MC	28	48.3	18	18
FRRT group 2	SEQ	30	51.7	18	18
	MC	28	50.9	18	15
FRRT group 3	SEQ	27	49.1	18	15
	MC	28	50.0	18	18
FRRT group 4	SEQ	28	50.0	18	18

Relative solution frequency = Absolute solution frequency / n
Absolute solution frequency = raw score achieved by the test taker n = absolute amount of maximum raw score that was possible to get

Figure 4-13: Formula to calculate the relative solution frequencies. This was necessary because the FRRT groups 1,2 and 4 consist out of 18 items, while group 3 only contains 15 items *a-posteriori*.

Though the Syllogisms 2009 items were not allocated to four groups – like the FRRT items - it would have been possible to calculate the further analyses with the raw scores (or absolute solution frequencies) instead of the relative solution frequencies to prove the differences between the two response formats. However, to have a well-arranged and clear overview of the results, it was considered to use the same units for both sub-tests.

To prove if there are mean differences (and accordingly differences in the distributions) of the two response formats, a Welch-Test was calculated. The results concerning the FRRT are presented in table 4-22, while the results of the Syllogisms 2009 are shown in table 4-23.

Table 4-22
Results of the Welch-Test concerning the FRRT. The Type-I-error was set on an α- level of 0.05. The results show that the empirical t-value is bigger than the critical t-value which indicates that there are significant differences between the two response formats within the FRRT.

Results of the Welch-Test (FRRT)

Variables	n	$\bar{x}$	s^2	$t_{empirical}$	$t_{critical}$	df
relative solution frequency (“MC 1 out of 5”)	112	87.5496	119.100			
relative solution frequency (“SEQ 1 out of 5”)	113	82.193	209.336	3.137	1.652	209

Table 4-23
Results of the Welch-Test concerning the Syllogisms 2009. The Type I error was set on an α- level of 0.05. The results show that the empirical t-value is bigger than the critical t-value which indicates that there are significant differences between the two response formats within the FRRT.

Results of the Welch-Test (Syllogisms 2009)

Variables	n	$\bar{x}$	s^2	$t_{empirical}$	$t_{critical}$	df
relative solution frequency (“MC 1 out of 5”)	109	79.969	451.542			
relative solution frequency (“SEQ 1 out of 5”)	112	48.066	1340.068	7.949	2.353	3

4.4.1.2 U.S. American sample

Within the American sample it was difficult to see if there are differences between the two response formats. Because of the fact that Rasch model analyses were not feasible to calculate (n= 57 test takers edited the FRRT items and n= 36 test takers edited the Syllogisms 2009 items), it was not possible to get any information if the English FRRT and Syllogisms 2009 items also fit the Rasch model assumptions. However, this would have been an important condition to prove if the items of the English version of the verbal test battery measure in one dimension and also to see, if they collect the same construct as in the German language. In this case, only a graphical exploration like bar charts were drawn to get an approximate impression about possible differences between the two response formats. All items of the Syllogisms 2009 and the FRRT were used in the diagrams because there were no assumptions that the same items as in the German sample would not fit the Rasch model in the American sample. Figure 4-14 and figure 4-15 show the bar charts of the two response formats of the FRRT.

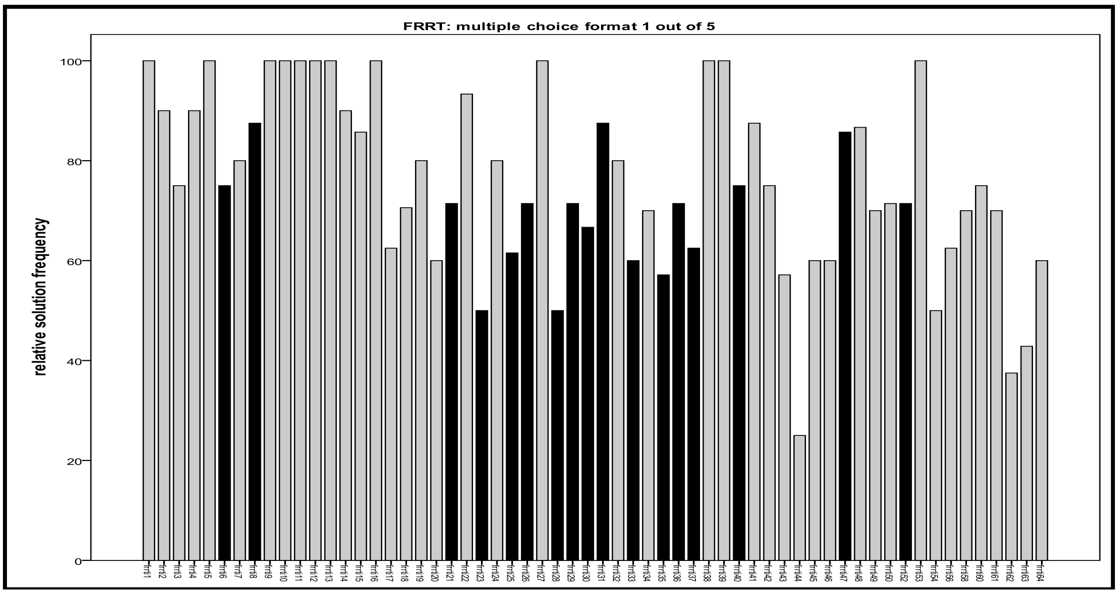


Figure 4-14: The items of the FRRT with the multiple choice format “1 out of 5”. The items are arranged according to their numbers, beginning with the number one on the very left side. If an item was solved more often with the sequential format “1 out nof 5”, then the bar is marked in black. If an item was solved more often with the multiple choice format “1 out of 5” the bar is coloured in grey.

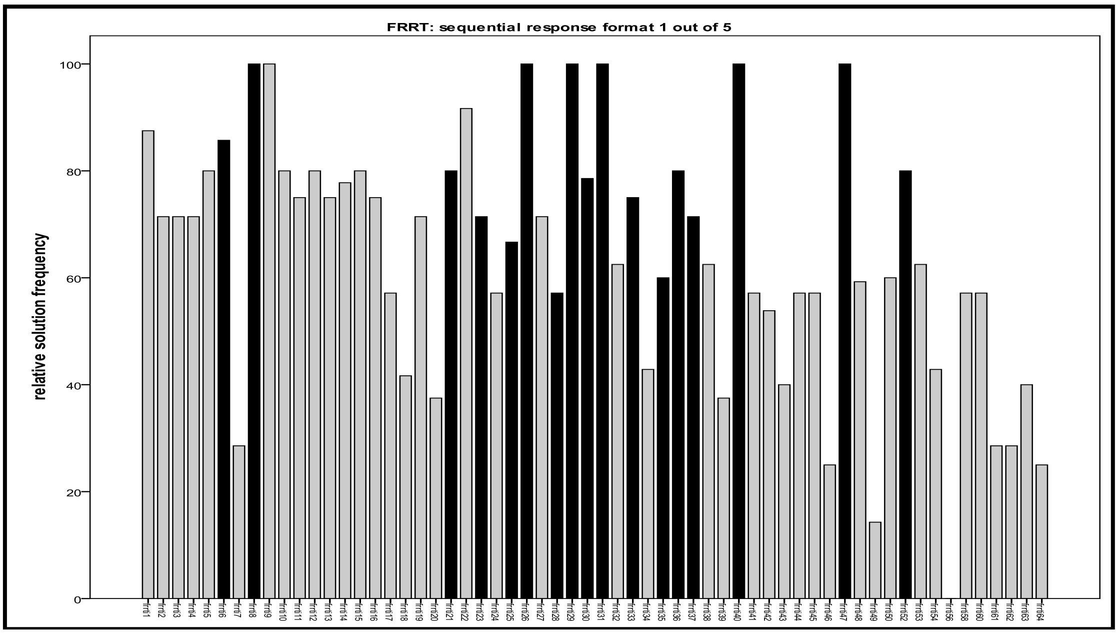


Figure 4-15: The items of the FRRT in the sequential response format “1 out of 5”. The items are arranged according to their numbers, beginning with the number one on the very left side. If an item was solved more often with the sequential format “1 out of 5”, then the bar is marked in black. If an item was solved more often with the multiple choice format “1 out of 5” the bar is coloured in grey.

The graphics show that 43 of 60 FRRT items have a higher solution frequency if they are in the multiple choice response format “1 out of 5”. Only 17 items have a higher solution frequency within the sequential response format “1 out of 5”. So, also for the U.S. sample, a trend into the direction of the alternative hypothesis can be seen among the items of the FRRT.

After the graphical analyses of the FRRT items, the focus is now put on the items of the Syllogisms 2009. It was of interest to see if a similar tendency can be seen for the Syllogisms 2009 items. A graphical check - also *via* bar charts - was designed. The results can be seen in figure 4-16.

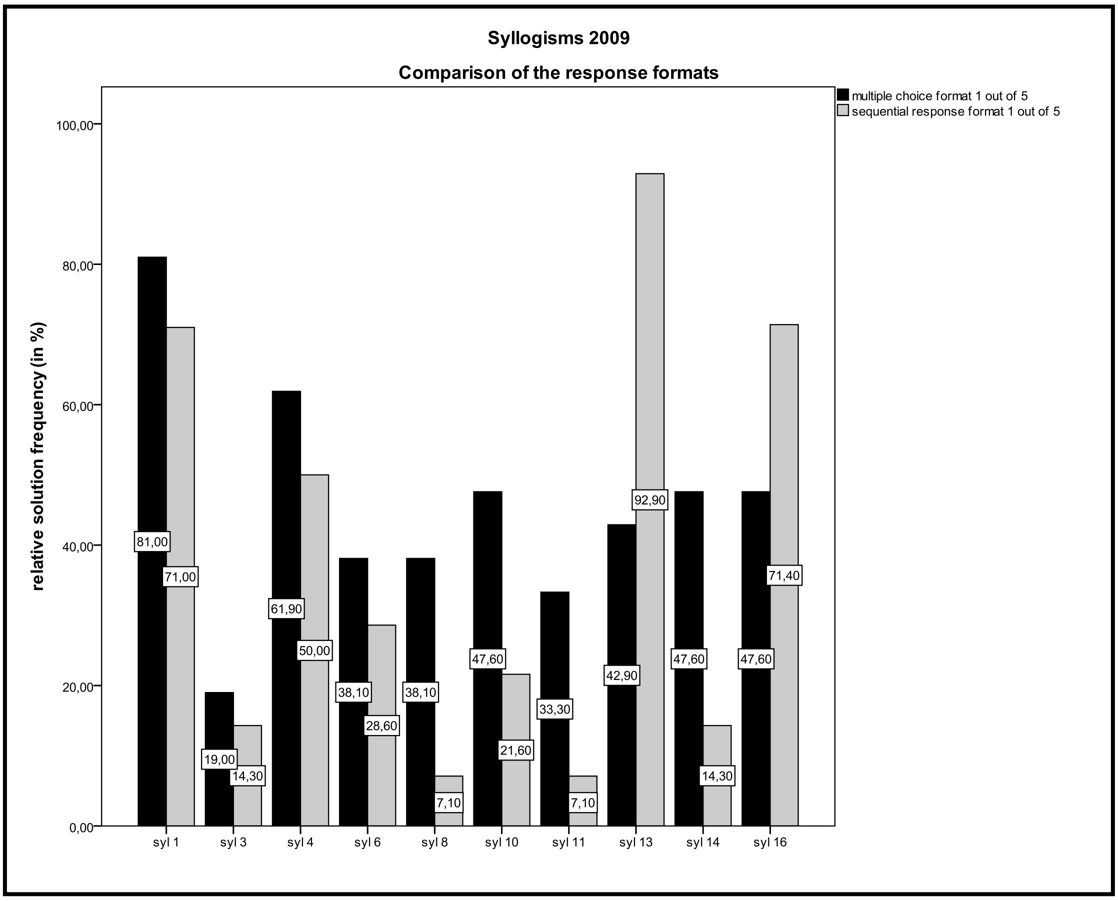


Figure 4-16: Graphical overview of the Syllogisms 2009 items. Bars coloured in grey represent the sequential response format “1 out of 5”, while black bars show the solution frequency of the same items but with the multiple choice format “1 out of 5”. The numbers within the bars indicate the percentage of the solution frequency of the particular Syllogisms 2009 items.

The graphic of the relative solution frequencies of the Syllogisms 2009 items show that within all ten items, only two are easier if they are in the sequential response format “1 out of 5” (Items *syl 13* and *syl 16*). All the other eight items have a higher solution frequency if they are in the multiple choice format “1 out of 5”. So, also the ten Syllogisms items that were used in this study seem to be more difficult if they are in the sequential response format “1 out of 5” than in the multiple choice format “1 out of 5”.

Although, no statistical analyses were calculated to prove the significance of the differences in the response formats, a trend into the direction of the alternative hypothesis can be seen for both sub-tests of the English version of the verbal test battery.

4.4.2 Results of the second study hypothesis

The second study hypothesis focuses on a possible coherence between the scores of the two sub-tests. It was of interest to inspect, if test takers who score high on the FRRT also score high on the Syllogisms 2009 and if test takers who score low on the FRRT also score low on the Syllogisms 2009. So, the second study hypothesis is:

H_0 (2): There is no coherence between the scores of the test takers on the FRRT and the scores of the test takers on the Syllogisms 2009.

H_1 (2): There is a coherence between the scores of the test takers on the FRRT and the scores of the test takers on the Syllogisms 2009. (Test takers who score high on the FRRT, also score high on the Syllogisms 2009 - and the other way round).

So, the focus was put on a possible coherence between the two sub-test results in respect to a first impression on the concordance validity of the FRRT - compared with the short form of the verbal reasoning test Syllogisms 2009.

4.4.2.1 Austrian sample

As it can be seen in chapter 4.4.1.1 the variables *relative solution frequency of the Syllogisms 2009* and *relative solution frequency of the FRRT* are not normal distributed in both response formats. Because of this fact, a Spearman- Correlation was calculated to see if there is a coherence between the scores of the two sub-tests of the German version of the verbal test battery. The results of the Spearman-Correlation (α= 0.05, two-sided) of the Austrian sample can be seen in Table 4-24.

Table 4-24
Results of the Spearman- Correlation of the Austrian sample.

Spearman- Correlation			
Variables	**n**	**Correlation Coefficient**	**sig. (α= 0.05)**
relative solution frequency (FRRT) – relative solution frequency (Syllogisms 2009)	221	0.227	0.001

As table 4-23 shows, there is a positive correlation between the two scores. Though, the correlation coefficient is relatively low (r= 0.227), the result is significant (p= 0.001; α= 0.05; two-sided) and meaningful. It can be said that there is a coherence between the scores of the two subtests of the German version of the verbal test battery: If people score high on the FRRT, they also score high on the Syllogisms 2009 and if people score low on the FRRT, they also score low on the Syllogisms 2009.

In a next step, the focus was put more on the two response formats of the sub-tests. It was the aim to see, if there are significant correlations, if the sub-tests are split into the various response format combinations. Because of the fact that now four correlations are calculated, an adjustment for multiple comparisons has to be made. When performing hypotheses tests with multiple contrasts, the overall significance level has to be adjusted from the significance levels for the included contrasts to reduce a Type I error. In this thesis, the Bonferroni correction was chosen. (Zöfel, 2003; Bortz, 2005). This method adjusts the observed significance level for the fact that multiple contrasts are being tested. The formula can be seen in figure 4-17.

Bonferroni correction:

$$\alpha\text{- level } 0.05/\, k^{1}$$

1: k: number of significance tests.

Figure 4-17: Formula of the Bonferroni correction.

In this case, the α- level of 0.05 sank to 0.0125. The results can be seen in table 4-25.

Table 4-25
The results of the Spearman-Correlation for the response format combinations of the two subtests of the German version of the verbal test battery. The FRRT in the sequential response format "1 out of 5" (short: SEQ) correlates significantly with the Syllogisms 2009 in both formats (multiple choice (short: MC) " 1 out of 5" and sequential "1 out of 5")

Spearman-Correlation

Format	n	Correlation Coefficient	sig. (α= 0.0125)
FRRT (MC) – Syllogisms 2009 (MC)	55	0.253	0.063
FRRT (SEQ) – Syllogisms 2009 (SEQ)	55	0.470	**0.000**
FRRT (MC) – Syllogisms 2009 (SEQ)	57	0.186	0.165
FRRT (SEQ) – Syllogisms 2009 (MC)	54	0.316	**0.020**

The results in table 4-23 show that two correlations are significant. If the FRRT is executed in the sequential response format “1 out of 5”, then there is a significant coherence to the results of the Syllogisms 2009 in both formats.
So, it can be said that test takers who executed the FRRT in the sequential response format “1 out of 5” and who scored high in this sub-test also scored high on the short version of the Syllogisms (in both response formats). Likewise it can be said, that test takers who executed the FRRT in the sequential response format “1 out of 5” and who scored low in this sub-test also scored low on the short version of the Syllogisms (in both formats).

However, though the results are significant, the effects are very small (22% and 10%). So, it can be said, that there is a coherence between the score on the FRRT in the sequential response format “1 out of 5” and the score on the short version of the Syllogisms (in both formats) but it is rather small and can be neglected. Validity studies with different – probably more connatural tests to the FRRT – might show more insightful results.

Now, if the focus is put back on the main (general) correlation, it can be said that there is a significant (but rather small) coherence between the relative solution frequencies of the two sub-tests. So, also the second null hypothesis has to be refused and the alternative hypothesis has to be accepted (within the Austrian sample).

4.4.2.2 U.S. American sample

Within the U.S. American sample – as already described in chapter 4.6.1.2 – no item check was possible. Because of this fact, no correlation could have been calculated. To still get information about this study hypothesis within the US-American sample, graphical methods were used. A scattergram was designed to see the “location of the test takers” in a orthogonal coordinate system. In this context, the X-axis represents the scores of the test takers on the FRRT (all 60 items were used), while the Y-axis represents the scores of the test takers on the sub-test Syllogisms 2009 (all 10 items were used).

Figure 4-18 shows the dispersion of the scores of the test takers in an orthogonal coordinate system. For the U.S. sample, it was possible to use the scores of the subtests instead of the relative solution frequencies because of the fact that all items were used and therefore all students could achieve the same maximum score.

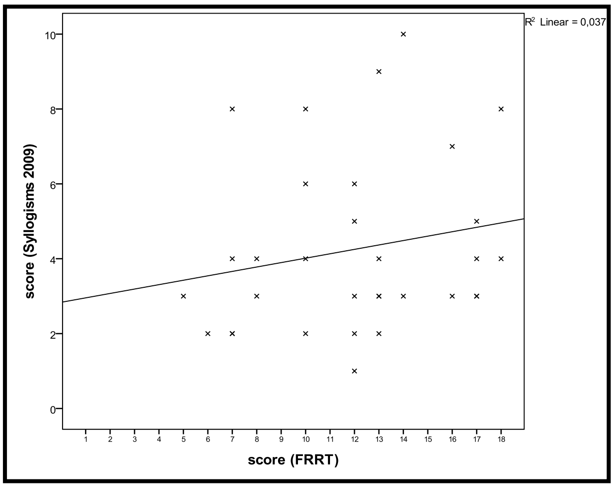


Figure 4-18: Scattergram of the scores of the two subtests FRRT and Syllogisms 2009. The X- axis shows the scores of the test takers on the FRRT, while the Y- axis shows the scores of the test takers on the subtest Syllogisms 2009.

As it can be seen in the scattergram in figure 4-18, there is a slight connection between the two scores of the sub-tests of the English version of the verbal test battery (R^2 = 0.037).

4.4.3Results of the third study hypothesis

The third study hypothesis focuses on the answering styles of the students:

H_0 (3): There is no difference in the answering style (declared through editing time and choice of solution) between psychology students at the Cypress College in California, Los Angeles and psychology students at the University of Vienna.

H_1 (3): There is a difference in the answering style (declared through editing time and choice of solution) between psychology students at the Cypress College in California and psychology students at the University of Vienna.

For the inspection of the third study hypothesis, the focus was put on the answering style of the test takers, which was defined through the following quantitative variables:

- editing time: How long did the test takers need to edit the test?
- choice of solution: Which answer was chosen by the test takers? The focus was on the student's choice of the solution position in each item.

First the results of the editing times will be presented. Then the analyses of the choices of solution will follow.

4.4.3.1 Editing time

A special interest was put on possible differences in the editing times of the two countries. As it is already explained in chapter 4.1, some of the U.S. American data was not used because the test takers executed the sub-tests in a very short time period. However, in this chapter, differences in the editing times between the two countries will first be analyzed with the whole U.S. American sample and then within the "censored" sample to give an overview of the analyzing process. All of the calculations were done *via* Welch-Tests. The type-I-risk was set on an α- level of 0.05. Table 4-26 shows the results of Welch-Test calculations concerning the "uncensored" FRRT editing times.

Table 4-26
Results of the Welch-Test concerning the FRRT editing times of the Austrian and the "uncensored" U.S. American data.

Results of the Welch-Test (FRRT)

Variables	n	$\bar{x}$	s^2	$t_{empirical}$	$t_{critical}$	df
FRRT editing time (Austria)	225	24.033	229669.317			
FRRT editing time (USA)	104	15.5	350848.173	0.129	1.654	168

The results in table 4-25 show, that there is a significant difference in the FRRT editing times concerning the two countries. In a next step, the focus was put on the Syllogisms 2009 editing times. Table 4-27 presents the results of the Welch-Test concerning the editing times within the two countries on the Syllogisms 2009.

Table 4-27
Results of the Welch-Test concerning the Syllogisms 2009 editing times of the Austrian and the "uncensored" U.S. American data.

Results of the Welch-Test (Syllogisms 2009)

Variables	n	$\bar{x}$	s^2	$t_{empirical}$	$t_{critical}$	df
Syllogisms 2009 editing time (Austria)	221	8.133	4429.177			
Syllogisms 2009 editing time (USA)	102	3.167	21514.909	0.244	1.650	274

The results show that within both sub-test of the verbal test battery, there is a significant difference between the editing times of the two countries. As it can be seen figures 4-1 and 4-2, the U.S. American students executed the test in a shorter time period than the Austrian test takers.

Because of these circumstances and the assumptions of formal response sets (Kubinger, 2009), the U.S. American sample was "censored": a 5 % percentile limit of the Austrian editing times was established and every U.S. American test taker whose editing time was within the lower 5 % percentile of the editing times, was cut out of the analyses. Table 4-28 shows the 5 % percentile limits of the Austrian sample.

Table 4-28
The 5 % percentile limits of the Austrian editing times.

Test	Format	5% percentile limit of the Austrian editing times
FRRT	MC	14min 05sec
	SEQ	14min 09sec
Syllogisms 2009	MC	3min 45sec
	SEQ	4min 06sec

The "censorship" of the U.S. American sample implicated that the U.S. American data shrank massively as it can be seen in table 4-29.

Table 4-29
The 5 % percentile limit of the Austrian editing times. ("n" represents the absolute number of people of each category.)

test	percentile level	n	percentage
FRRT	> 5 % percentile	57	54.8 %
	< 5 % percentile	47	45.2 %
Syllogisms 2009	> 5 % percentile	37	36.3 %
	< 5 % percentile	65	63.7 %

After the implementation of the 5 % percentile level, only 36 U.S American data of the Syllogisms 2009 and 57 U.S. American data of the FRRT remained. In a second step, possible differences between the editing times of the Austrian the “censored” U.S. American sample were analyzed. A Welch-Test was calculated to prove if there are differences in means (and accordingly in the distributions) of the editing times between the countries. The results of the Welch-Tests can be seen in table 4-30 (FRRT) and 4-31 (Syllogisms 2009).

Table 4-30
Results of the Welch-Test concerning the FRRT editing times of the Austrian and the “censored” U.S. American data.

Results of the Welch-Test (Family Relations Reasoning Test)

Variables	n	$\bar{x}$	s^2	$t_{empirical}$	$t_{critical}$	df
FRRT editing time (Austria)	225	24.033	229669.317			
FRRT editing time (USA)	57	22.783	161265.840	3.137	1.660	101

Table 4-31
Results of the Welch-Test concerning the editing times of the Syllogisms 2009.

Results of the Welch-Test (Syllogisms 2009)

Variables	n	$\bar{x}$	s^2	$t_{empirical}$	$t_{critical}$	df
Syllogisms 2009 editing time (Austria)	222	8.133	44929.177			
Syllogisms 2009 editing time (USA)	37	5.883	11224.369	0.0996	1.661	94

Concerning the editing times of the two sub-tests of the verbal test battery, it can be said that there are significant differences between the two countries if the U.S. American data is not “censored”. However, Welch-Test analyzes of the “censored” data showed that there is only a significant difference within the Syllogisms 2009. U.S American test takers executed the Syllogisms 2009 in a shorter time period than the Austrian test takers.

So, concerning the editing times, the null hypothesis H_0 (3) has to be rejected for the “uncensored” sample, but has to be partly accepted for the “censored” sample.

4.4.3.2 Choice of solution

This chapter will focus on the choices of solutions, done by the test takers. The analyses are done *via* crosstabulations and χ^2- Test. It is of interest to see, if the response format has an influence on the choice of solution. Do all the test takers choose the same distractors or does it depend on the response format what they choose? Test takers who used the multiple choice format “1 out of 5” could see all the possible solutions at one time, while test takers with the sequential response format “1 out of 5” had to decide step by step if one solution was correct or not. It is the aim of this chapter to briefly analyze this question to see if there is a tendency that if one response format leads to other choices of solutions than the other. First, the results of the Austrian sample will be presented, followed by the U.S. American results.

To see (and statistically prove) the distribution of the categorical variable *choice of solution*, crosstabulations were calculated. (The type-I-risk was set on an α-level of 0.05.) The focus was put on possible differences between the different choices of solution depending on the response formats. The results of the sub-test Syllogisms 2009 can be seen in table 4-32.

Table 4-32
Crosstabulations of the solution choices separated by the two different response formats: multiple choice (short: mc) “1 out of 5” and sequential (sort: seq) “1 out of 5” of the items of the Syllogisms 2009 (α= 0.05). Fields surrounded by a thick frame represent solution choices that were not used by any of the test takers. Fields marked in grey show the most preferred solution position of the particular item.

Crosstabulation (Syllogisms 2009)

		Choice of solution position									
item names	format	**0** absolute number/ percentage	**1** absolute number/ percentage	**2** absolute number/ percentage	**3** absolute number/ percentage	**4** absolute number/ percentage	**5** absolute number/ percentage	correct solution position	n	p α= 0.05	df
syl 3	mc	0/ 0%	0/ 0%	11/ 10.1%	10/ 9.2%	18/ 71.6%	10/ 9.2%	4	109		
	seq	8/ 7.1%	1/ 0.9%	34/ 30.4%	19/ 17%	49/ 43.8%	1/ 0.9%	4	112	**0.000**	5
syl 6	mc	0/ 0%	2/ 1.8%	17/ 15.6%	83/ 76.1%	5/ 4.6%	2/ 1.8%	3	109		
	seq	3/ 2.7%	3/ 2.7%	62/ 55.4%	40/ 35.7%	4/ 3.8%	0/ 0%	3	112	**0.000**	5
syl 8	mc	0/ 0%	0/ 0%	8/ 7.3%	5/ 4.6%	11/ 10.1%	85/ 78%	5	109		
	seq	5/ 4.5%	1/ 0.9%	13/ 11.7%	7/ 6.3%	27/ 24.1%	59/ 52.7%	5	112	**0.002**	5
syl 10	mc	0/ 0%	0/ 0%	2/ 1.8%	9/ 8.3%	96/ 88%	2/ 1.8%	4	109		
	seq	0/ 0%	4/ 3.8%	34/ 30.4%	9/ 8%	65/ 58%	0/ 0%	4	112	**0.000**	4
syl 11	mc	0/ 0%	1/ 0.9%	4/ 3.7%	8/ 7.3%	0/ 0%	96/ 88%	5	109		
	seq	7/ 6.3%	0/ 0%	31/ 27.7%	6/ 5.4%	6/ 5.4%	62/ 55.4%	5	112	**0.000**	5
syl 14	mc	0/ 0%	0/ 0%	14/ 12.8%	10/ 9.2%	85/ 78%	0/ 0%	4	109		
	seq	1/ 0.9%	2/ 1.8%	55/ 49.1%	6/ 5.4%)	48/ 42.9%	0/ 0%	4	112	**0.000**	4

The crosstabulations show that within all Syllogisms 2009 items, there are significant differences which “solution” was chosen subject to the response format ($0.00 \leq p \leq 0.002$).

This indicates that test takers who used the multiple choice format “1 out of 5” decided for different solutions (or solution positions) than test takers who executed the same items but with the sequential response format “1 out of 5”. A closer look on the crosstabulations shows that especially the response category “zero” (which means that all presented response options in the sequential version were neglected) was used in 5 of 6 Syllogisms 2009 items. If using the multiple choice format “1 out of 5”, it is not possible to neglect all presented answers (test takers had to click at one answer to switch to the next task) and this fact could be another hint why the multiple choice format “1 out of 5” seems easier than the sequential response format “1 out of 5” (which offers – in a strict sense – six response possibilities).

Concerning the two distribution curves of the choices of solutions of the two response formats, it can be seen that there is a difference in the distributions in two of the six Syllogisms 2009 items. Two times, there are different peaks of the distributions and in both cases, the most preferred solution position within the multiple choice format (marked in grey) is also the position where the solution was. In summary, it can be said, that concerning the multiple choice format “1 out of 5” the solution position which was selected most often (fields marked in grey) was always the correct solution position, while concerning the sequential response format “1 out of 5”, in 2 out of six items, the most selected solution position (also marked in grey) was not the correct solution position.

In a further step, it was the aim to see, if the same effects can be found among the items of the Family Relations Reasoning test. However, because of the big amount of FRRT items, table 4-33 only shows the significant results. Nine of all fifty-seven FRRT items (that are conform with the Rasch model assumptions) showed significant differences in the choices of solutions between the two response formats: multiple choice “1 out of 5” and sequential response format “1 out of 5”.

Table 4-33
Crosstabulations of the solution choices separated by the two different response formats: multiple choice (short: mc) "1 out of 5" and sequential (sort: seq) "1 out of 5" of the items of the Syllogisms 2009 (α= 0.05) of the FRRT. Fields surrounded by a thick frame represent solution choices that were not used by any of the test takers. Fields marked in grey show the most preferred solution position of the particular item (within the formats).

Crosstabulations

		Choice of solution position									
item names	format	**0** absolute number/ percentage	**1** absolute number/ percentage	**2** absolute number/ percentage	**3** absolute number/ percentage	**4** absolute number/ percentage	**5** absolute number/ percentage	correct solution position	n	p α= 0.05	df
frrt 5	mc	0/ 0%	0/ 0%	1/ 3.6%	1/ 3.6%	0/ 0%	26/ 92.8%	5	28		
	seq	0/ 0%	0/ 0%	0/ 0%	7/ 25%	0/ 0%	21/ 75%	5	28	**0.049**	2
frrt 16	mc	0/ 0%	6/ 21.4%	16/ 57.1%	5/ 17.9%	0/ 0%	1/ 3.6%	3	28		
	seq	9/ 32.1%	1/ 3.6%	14/ 50%	2/ 7.1%	0/ 0%	1/ 3.6%	3	28	**0.007**	4
frrt 25	mc	0/ 0%	5/ 8.9%	4/ 7.1%	1/ 1.8%	46/ 82.1%	0/ 0%	4	56		
	seq	6/ 10.9%	1/ 1.8%	1/ 1.8%	1/ 1.8%	46/ 83.6%	0/ 0%	4	55	**0.033**	4
frrt 27	mc	0/ 0%	0/ 0%	14/ 85.7%	0/ 0%	4/ 14.3%	0/ 0%	2	28		
	seq	1/ 3.6%	0/ 0%	26/ 92.9%	0/ 0%	0/ 0%	1/ 3.6%	2	28	**0.000**	3
frrt 42	mc	0/ 0%	33/ 58.9%	0/ 0%	0/ 0%	23/ 41.1%	0/ 0%	1	56		
	seq	6/ 10.9%	23/ 41.8%	0/ 0%	0/ 0%	25/ 45.5%	1/ 1.8%	1	55	**0.031**	3
frrt 48	mc	0/ 0%	4/ 3.8%	7/ 6.3%	41/ 36.6%	53/ 47.3%	7/ 6.3%	4	112		
	seq	10/ 8.8%	5/ 4.4%	3/ 2.7%	39/ 34.5%	55/ 84.7%	1/ 0.9%	4	113	**0.006**	5
frrt 52	mc	0/ 0%	2/ 7.1%	6/ 21.4%	20/ 71.4%	0/ 0%	0/ 0%	3	28		
	seq	9/ 32.1%	3/ 10.7%	4/ 14.3%	12/ 42.9%	0/ 0%	0/ 0%	3	28	**0.009**	3
frrt 63	mc	0/ 0%	0/ 0%	5/ 17.9%	5/ 17.9%	15/ 53.6%	3/ 10.7%	4	28		
	seq	7/ 25%	3/ 10.7%	3/ 10.7%	0/ 0%	12/ 42.9%	3/ 10.7%	4	28	**0.007**	5
frrt 64	mc	0/ 0%	2/ 7.1%	4/ 14.3%	0/ 0%	3/ 10.7%	19/ 67.9%	5	28		
	seq	8/ 29.6%	2/ 7.4%	3/ 11.1%	1/ 3.7%	0/ 0%	13/ 48.1%	5	27	**0.021**	5

The crosstabulations of the FRRT solution decisions within the Austrian sample show that there are significant differences between 9 of 57 FRRT items ($0.00 \leq p \leq 0.049$). However, there are no differences between the two formats concerning the most preferred answers (fields marked in grey). On the other hand, it can be seen that the response option *zero answers are correct* is used in 8 of those 9 items with the sequential response format "1 out of 5". Like within the Syllogisms 2009 items, the sequential response option seems more difficult because of this sixth possible response option. However, it has to be mentioned that the FRRT was quite easy to solve for the test takers (the results can be seen in chapter 4.4.4). This might be the reason why only 9 of 57 items showed significant differences in the choices of solutions. However, it has to be considered that if an item difficulty parameter is higher than the skill of a person, then the guessing effects can be reduced with the sequential response format "1 out of 5" compared to the multiple choice format "1 out of 5".

Concerning the U.S. American sample, it was not possible to statistically prove (*via* crosstabulations) if there are differences in the choices of solutions between the two response formats. Because of the fact that no Rasch model analyses had been possible, it was not feasible to prove if the items measure in one dimension and because of this fact, it was not possible to see if differences in the choices of solutions follow from other – unrecognized – interfering variables.

However, to answer the third study hypothesis, it has to be said that there are differences between the two countries concerning the editing times. U.S. American students needed less time to execute the test: 47 students needed less than about 14 minutes to finish the 18 items of the FRRT, while 65 students needed less than 3 to 4 minutes to finish the 10 Syllogisms 2009 items. A "censorship" of the U.S. American sample with the lower 5 % percentile limit of the Austrian editing times lead significant differences concerning the Syllogisms 2009 but not within the FRRT.

Concerning the choices of solutions within the response formats, no comparison of the two countries was possible but within the Austrian sample, significant differences between the response formats showed up in all (!) ten items of the Syllogisms 2009 and nine of fifty-seven items of the FRRT.

So, it can be said that H_0 (3) has to be accepted in parts (for the editing times of the Syllogisms 2009, the choices of solutions within all Syllogisms 2009 items and 9 of 57 FRRT items).

4.4.4 Results of the fourth study hypothesis

The fourth study hypothesis focuses on possible differences in the scores of the students of the Cypress College in Anaheim and the University of Vienna in Austria:

H_0 (4): There is no difference in the scores between psychology students at the Cypress College in California and psychology students at the University of Vienna.

H_1 (4): There is a difference in the scores between psychology students at the Cypress College in California and psychology students at the University of Vienna.

Due to the fact that no test of equivalency was possible (see chapter 4.1), the scores achieved by the test takers of the two countries were not possible to be compared. However, to still get information about which country scored higher, the scores (of the items that were conform with the Rasch model analyses within the Austrian sample) of the test takers were listed up in four tables. It was of interest to see how many students solved 0 %, 25 %, 50 %, 75 % or even more of the items and which country achieved better results. Again, the relative solution frequencies were used instead of the raw scores (explanation can be found in chapter 4.4.1.1), to create fair conditions. Concerning the U.S. American sample, only the "censored sample" was taken (as it is explained in chapter 4.1). To start with the FRRT results, the following tables show the distributions of the relative solution frequencies of the Austrian sample (table 4-34) and the U.S. American sample (table 4-35).

Table 4-34
Overview of the distribution of the FRRT relative solution frequencies within the Austrian sample (n= 225).

Distribution of the (FRRT) relative solution frequencies within the Austrian sample

Relative solution frequencies (in percent)	**Frequency**	**Percent**	**Cumulative percent**
44.44	4	1.8	1.8
50.00	1	0.4	2.2
53.33	4	1.8	4.0
55.56	3	1.3	5.3
60.00	2	0.9	6.2
61.11	2	0.9	7.1
66.67	10	4.4	11.6
72.22	13	5.8	17.3
73.33	8	3.6	20.9
77.78	16	7.1	28.0
80.00	11	4.9	32.9
83.33	33	14.7	47.6
86.67	5	2.2	49.8
88.89	29	12.9	62.7
93.33	11	4.9	67.6
94.44	29	12.9	80.4
100.00	44	19.6	100.0

Table 4-35
Overview of the distribution of the FRRT relative solution frequencies within the U.S. American sample (n= 57).

Distribution of the (FRRT) relative solution frequencies within the U.S. American sample

Relative solution frequencies (in percent)	**Frequency**	**Percent**	**Cumulative Percent**
27.78	1	1.8	1.8
33.33	4	7.0	8.8
38.89	2	3.5	12.3
40.00	1	1.8	14.0
44.44	2	3.5	17.5
55.56	6	10.5	28.1
60.00	2	3.5	31.6
61.11	2	3.5	35.1
66.67	8	14.0	49.1
72.22	7	12.3	61.4
73.33	1	1.8	63.2
77.78	3	5.3	68.4
80.00	1	1.8	70.2
83.33	2	3.5	73.7
86.67	2	3.5	77.2
88.89	2	3.5	80.7
93.33	1	1.8	82.5
94.44	5	8.8	91.2
100.00	5	8.8	100.0

The distributions of the FRRT relative solution frequencies show that Austrian students scored higher than the U.S. American students. While 1.8 % of the Austrian students achieved a relative solution frequency of 44.4 % (which was also the minimum score frequency), 17.6 % of the U.S. American students achieved 44.4 % and lower results. Concerning the relative solution frequency of 100 %, it can be said, that 19.6 % of the Austrian students achieved this result, while only 8.8 % of the U.S. American test takers achieved 100 %.

Now, focusing on the Syllogisms 2009, the distributions of the relative solution frequencies are listed up in table 4-36 (Austrian sample) and 4-37 (U.S. American sample). Only the six items that remained after the Austrian Rasch model analyses were used to have an overview of the scores achieved within the two countries.

Table 4-36
Overview of the distribution of the Syllogisms 2009 relative solution frequencies within the Austrian sample (n= 222).

Distribution of the (Syllogisms 2009) relative solution frequencies within the Austrian sample

Relative solution frequencies (in percent)	**Frequency**	**Percent**	**Cumulative Percent**
0.00	23	10.4	10.4
16.67	18	8.1	18.6
33.33	18	8.1	26.7
50.00	25	11.3	38.0
66.67	30	13.6	51.6
83.33	45	20.4	71.9
100.00	62	28.1	100.0

Table 4-37
Overview of the distribution of the Syllogisms 2009 relative solution frequencies within the U.S. American sample (n= 57).

Distribution of the (Syllogisms 2009) relative solution frequencies within the U.S. American sample

Relative solution frequencies (in percent)	**Frequency**	**Percent**	**Cumulative Percent**
0.00	27	26.5	26.5
16.67	54	52.9	79.4
33.33	9	8.8	88.2
50.00	5	4.9	93.1
66.67	4	3.9	97.1
83.33	2	2.0	99.0
100.00	1	1.0	100.0
0.00	27	26.5	26.5
16.67	54	52.9	79.4
33.33	9	8.8	88.2

The results in table 4-36 and table 4-37 show that also within the second sub-test of the verbal test battery, the Austrian test takers scored higher than the U.S. American test takers. While 93.1 % of the students in the USA achieved a relative solution frequency of 50 % and under, only 38 % of the Austrian test takers achieved a maximum score of 50 %. 62 Austrian students solved all of the items, while within the U.S. American sample, only 1 test taker got a score of 100 %.

Because of the fact that no statistical comparison of the students’ scores was feasible, the fourth study hypothesis cannot be answered. However, the results allow the assumption of a tendency into the direction of a better test performance of the Austrian sample.

4.4.5 Additional results of the study and a critical retrospective

In this chapter the primary focus is put on additionally results – which are not part of the hypotheses tests but are interesting to discuss.

As described in chapter 3.1.2 new item construction rules were developed for the FRRT. It was not the topic of this study to actually prove if those new item construction rules lead to more difficult items and this is why analyses which can check the item homogeneity like the Martin-Löf test were not calculated in this study. However, it would be interesting for future studies which focus more on the FRRT and particularly on the item construction rules of this test to analyze the composition of the FRRT item difficulties. To get a first impression if the newly developed item construction rules in this study lead to more difficult items, a histogram with the relative solution frequencies of the old and the new items was designed (figure 4-19). The items are lined up on the X- axis, beginning with the most difficult items on the left side and ending with the three items which were solved by every Austrian test taker on the right side: *item 1, item 9 and item 13*. (Only the Austrian data was taken.)

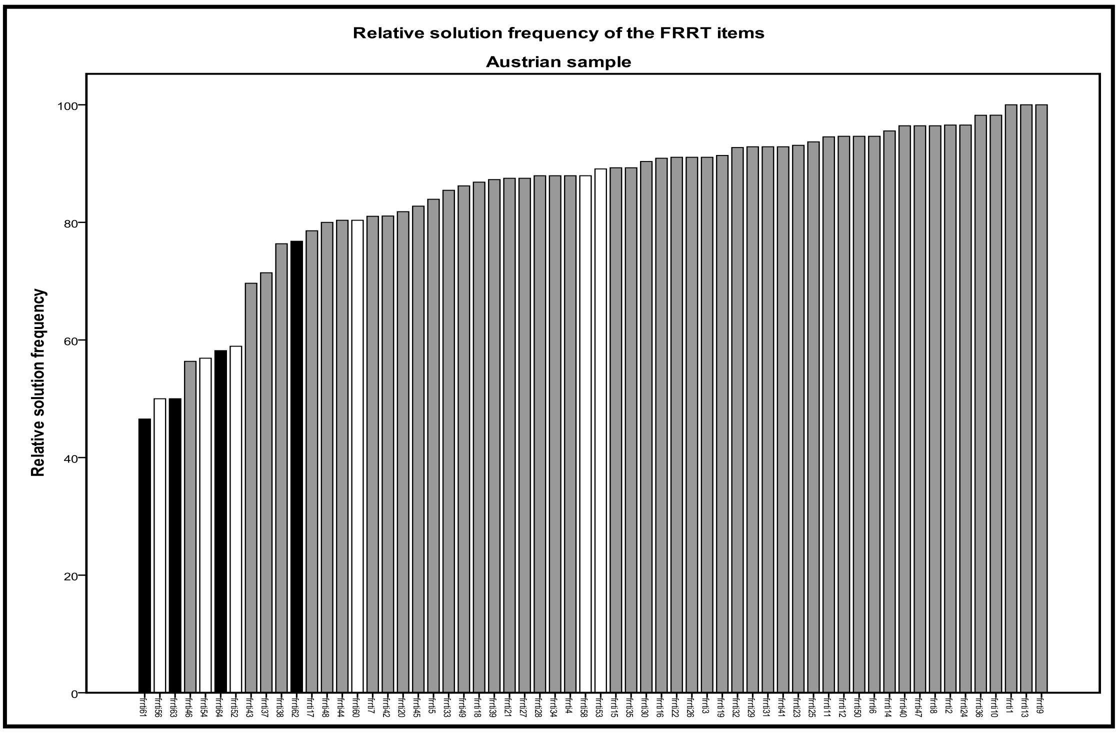


Figure 4-19: Relative solution frequencies of the FRRT items within the Austrian sample. The items are lined up, beginning with the most difficult ones on the very left side and ending with the easiest ones on the very right side (Items 1,9 and 13 were solved by every test taker of the Austrian sample). Bars which are colored in black show the item solution frequencies of those items that were constructed with the new item generation rules. Items that are colored in white are based on Schechtner's item construction rules but have a higher number of relations and redundancies than the "old" items. Items colored in grey are the original items from Schechtner (2009).

As it can be seen in figure 4-19, those items that were constructed with the new item construction rules are among the most difficult items of the FRRT. This could indicate that the new rules also have an influence on the item difficulties.

Though the main focus of this thesis was put on the psychometric qualities of various multiple choice formats, special interest was also put on the distractors that are used in the Family Relations Reasoning Test. As it is mentioned in this thesis, it turned out that the items of the FRRT were easy to solve for the test takers – especially the Austrian test takers. Therefore considerations on how to create better distractors which are similar to the solution were made. In this study, a random distractor generation *via* computer was established. The distractor pool out of which the distractors were taken can be seen in table 4-38. It has to be discussed, if it is better to use a systematic distractor construction rule with the help of a so called "distractor-tree" where each particular solution gets adequate distractors.

Table 4-38
Overview of the distractor pool out of which the four distractors for the FRRT were randomly chosen.

female	**male**	**female**	**male**	**female**	**male**	**female**	**male**
grandmother	grandfather	mother	father	aunt	uncle	sister	brother
granddaughter	grandson	daughter	son	cousin	cousin	niece	nephew

However, though a random allocation is mostly a good way to accomplish equipartition of effects, in this case it might have lead to a higher solution frequency of the items because some distractors were just not perfectly chosen for an item. To illustrate this consideration, an example shall be given: Within the FRRT, there are items, which story texts lead down from the third generation to the first (for example: starting with the grandmother and then going down to the grandchildren). So, the random distractor *grandmother*, which was chosen *via* computer, seems obviously "wrong" in this context. If it is still used, then it can be assumed that the *a-priori* guessing probability of an item with the multiple choice format "1 out of 5" increases from $^1/_5$ to $^1/_4$. For further studies, it would be interesting to see, if the item difficulty increases if distractors are chosen that are similar to the solution (like same generation, similar branch of the family tree diagram and so on).

5 Interpretation and summary of the study results

Although the data of the U.S. American sample was omitted (about 50 % of the data had to be excluded due to extreme response styles; see chapter 4.4.2), interesting results can be taken from this thesis. Unfortunately, because of the loss of a great amount of the American data, no Rasch model analyses could have been calculated and therefore the actual hypotheses testings were not possible for this sample. This lead to the consequences, that the planed comparisons between Austria and the United States of America were not always possible. However, the four hypotheses which were tested in this study show - as far as they can be answered - absorbing results.

The first study hypothesis was adjusted to the topic if items vary in their difficulty if they are designed with the multiple choice format "1 out of 5" or in the sequential format "1 out of 5". The study was designed in that way, that the content of the items remained the same, while the response format differed. This asserted that differences in item difficulty can be ascribed to the response format and not to the content of the items. Calculations with the Welch-Test showed, that within both sub-tests there are significant differences between the two response formats. Test takers who edited the test in the sequential response format "1 out of 5" achieved significantly lower results than test takers who executed the same items but with the multiple choice format "1 out of 5". (p_{FRRT}= 0.000 and $p_{Syllogisms\ 2009}$= 0.000; Alpha level of 0.05, two-sided).
For the U.S. American sample no statistical analyses could have been calculated, but graphical demonstrations (bar charts) showed, that also within the U.S. American sample, a tendency into the direction of the alternative hypothesis was noticeable.

Concerning the second study hypothesis it was the aim to explore if people who score high on the FRRT also score high on the Syllogisms 2009 and if people who score low on the FRRT also score low on the Syllogisms 2009. This hypothesis was designed in according to get a first impression of the concordance validity of the FRRT. Though only 6 out of 10 Syllogisms items were conform with the Rasch model, a significant Spearman correlation coefficient (r= 0.23; p= 0.001; α= 0.05) was established.

The reasons why the correlation coefficient is rather low could be the difference in difficulty between the two sub-test. (The items of the Syllogisms 2009 are compared to the items of the FRRT more difficult. A list of all the item parameters can be found in chapter 4.2.4.) Also, because of the fact that just 6 of 10 Syllogisms 2009 items fitted the Rasch model assumptions, a correlation with the long version probably would have brought more precise results.

Further it was considered if a change in the response formats has an influence on the correlation. A significant Spearman-Correlation was calculated for the combination: FRRT (in the sequential response format "1 out of 5") and the Syllogisms 2009 in both response formats (α= 0.0125 and α=0.316 (Bonferoni correction), p= 0.000 and p = 0.020). This indicates that there is a significant coherence between the solution frequency on the FRRT (only in the sequential response format) and the solution frequency on the Syllogisms items (in both formats). However, though the correlations are significant, the effects are rather small (22% and 10%).

Based on the results, it can be said that there is a coherence between the score on the FRRT and the score on the short version of the Syllogisms but it is very small and therefore not connotatively. (Validity studies with different – probably more connatural tests to the FRRT – might show more insightful results.)

Concerning the U.S. American data, it has to be mentioned that the sample was explored *via* scattergram which also showed a slight connection between the scores of the two sub-tests.

The third study hypothesis dealt with the response style (declared through the editing times and the choices of solution) of the test takers. It was of interest to see, if the students' decisions about the solution positions vary - depending on their country. Concerning the editing times, it turned out that about 50 % of the U.S. American test takers had completed the editing portion of the test in a very short time frame (see chapter 4.4.2) and therefore their results had to be excluded. To still make it possible to compare the two countries a 5 % percentile level of the Austrian editing time was established. The results of the American test takers who needed more time to execute the tests than the lowest 5 % percentile of the Austrian test takers were kept for further investigations. All the other results were excluded. Calculations with the Welch-Test showed significant results between the editing times concerning the sub-test Syllogisms 2009 but not for the FRRT. This results can be derived from the fact, that the Syllogisms items - which are more difficult than the FRRT items - had to be executed after the FRRT and it seems like that the U.S. American students who were mostly undergraduate students with less experience in psychological assessment were exhausted and less motivated to finish the test battery successfully.

Concerning the choices of solutions, it can be said, that analyses *via* crosstabulations and χ^2- Tests show significant differences between the two response formats. Test takers who used the sequential response format “1 out of 5” used different distractors (if they did not score) than test takers who executed the same item but with the multiple choice format “1 out of 5”. Such differences between the two formats can be primarily found for the Syllogisms 2009, where differences within all six items occurred. Pertaining to the FRRT, the test takers decided for different distractors in 9 of 57 items (that were conform with the Rasch model) between the two response formats. This could be derived from the fact, that the FRRT was easier for the test takers to solve and so the guessing effect was reduced.

The last study hypothesis applies to differences in scores between the two countries. Because of the fact that no test of equivalency *via* Rasch model was possible, no direct comparison with statistical analyzing methods was possible. This is the reason why only frequency tables with the relative solution frequencies achieved by the test takers were established. A closer look at these frequency tables show, that the Austrian test takers achieved higher scores ion both sub-test of the verbal test battery than the U.S. American test takers.

Conclusion: This study indicates a strong guessing effect within the multiple choice format “1 out of 5” compared to the sequential response format “1 out of 5”: If items are designed in the multiple choice format “1 out of 5”, they are significantly easier than the same items but in the sequential response format “1 out of 5”. Concerning the comparison of the two countries, unlucky circumstances did not allow calculating a test of equivalency *via* Rasch model and so the two countries could not be compared in an accurate way. For previous studies in foreign countries, it is to be advised to have a test assessor of the “home(-research)-team” *in situ* who guides the data collection. This could prevent misunderstandings and confine gratuitous confounding variables.

6 Discussion and focus on further research

This research study was designed to investigate not only differences between two different multiple choice response formats but also between the countries Austria and the USA. Due to the fact that about half of the U.S. American data had to be excluded (a lot of students executed the test in a very short time frame and showed extreme response behaviour as it can be seen in chapter 4.4.3.1), neither Rasch model analyses nor further statistical analyses were possible to calculate for the U.S. American sample. So, a test of equivalency *via* Rasch model between the two computer test batteries (English version and German version) was not possible. However, this would have been an obligated condition to be able to compare the two countries with each other. So, only comparisons concerning the editing times were drawn but compared to the big amount of data and variables within the sample, this was a rather meagre prey. A possible reason for the responding behaviour of the U.S. American students could have been the lack of financial compensation in this study. At American universities it is common that students are offered money for participating in scientific studies. For this thesis, it was not possible to allure and support the students financially. They "only" got offered a certificate about taking part in this research study, a feedback about their results and Austrian chocolate. So, this could have been a reason why the motivational aspects were probably rather low.

Because of the circumstances named above, the focus was primary put on the Austrian sample. It was the aim to see, if there are differences between the two response formats. This study could indicate that if the items were designed in the multiple choice format "1 out of 5", they were significantly easier than the same items (same content) but with the sequential response format "1 out of 5". Because of the fact that the content of the items remained the same but only the response format varied, those differences can be lead back to relevant guessing effects concerning the multiple choice format "1 out of 5": If an item is designed in this format, the test taker is able to see all possible answering alternatives at once, while if using the sequential response format the test taker has to decide step by step if one given answer is correct or not (without seeing the alternatives that will come up if he/she rejects an answer). Additionally, it also has to be mentioned that within the sequential response format, the test takers are not able to see the number of solution suggestions that will be given to them, while if using for example the multiple choice format “1 out of 5” they know, that there is one solution among four distractors. The test takers see all the solution alternatives at once and know that no other alternatives will follow. So, they know that the solution must be among these five suggestions. Concerning the sequential response format “1 out of 5”, the test takers do not know how many solution suggestions will follow. They have to decide step by step if a given solution is correct or if something “better” will follow. Thereby the person does not know if any other solution suggestion will even come up or if this was the last given alternative. It is not possible for the test takers to switch back and forth between the solution alternatives. If one possible solution is rejected then the person is not able to go back to change his/her decision if all the other given alternatives are not satisfying.

The conclusion that can be arrived from this study is, that if a computer test situation with a multiple choice response format is considered (for example for a selection assessment) and considerations are made if it is better to use the multiple choice format “1 out of 5” or the sequential response format “1 out of 5”, it should be considered that there are less guessing effects within the sequential response format. The likelihood that an item is solved will increase if the multiple choice format “1 out of 5” is used instead of the sequential response format “1 out of 5”. The explanation shall be given: If a test taker with an average skill to solve a particular item (0.00) has to solve this item, which has an average difficulty of 0.00, then the solution frequency is – in regards to the formula of the Rasch model 0.50. Figure 6-1 shows the calculating process:

$$p(+|\xi_\upsilon, \sigma_i) = e^{\xi\upsilon - \sigma i} / 1 + = e^{\xi\upsilon - \sigma i} = [e^{0.00-0.00}/(1 + e^{0.00-0.00})] = 1/(1+1) = ½ = 0.50$$

Figure 6-1: Rasch model calculation of a 50 % likelihood to solve an item with an item difficulty of 0.00 and a person skill of 0.00.

So, the chance for this person to solve the item is 50 %. If the difficulty of the particular item is not 0.00 but – depending on the chosen response format – higher for the sequential response format “1 out of 5” than for the multiple choice format “1 out of 5”, then the likelihood to solve the item will decrease or increase.

However, not only the response option can provide information about the difficulty of an item, but also the content. Concerning the FRRT, one problem was that this sub-test seemed too easy for the student test takers (especially the Austrians). It would be interesting for further studies to create FRRT items that describe two families who have no familial connection to each other. Then the test takers have to find out that there is more than one family described and have to chose the response option: "X and Y are not familial related to each other".

Though this study could not come up with important results concerning the cultural differences between Austria and the USA, the findings of this thesis still have a certain relevance for psychological assessment, especially for the selection assessment. In general it can be said that with the sequential response format "1 out of 5" the guessing effects can be massive abridged in comparison to the multiple choice format "1 out of 5". This result itself is a big step into the wide study field of psychological assessment.

7 Summary

One problem within multiple choice tests is that even if a test taker does not know the answer of an item he or she can solve it, simply through lucky guessing. Within the Division of Psychological Assessment of the University of Vienna, one of the major study fields is to do research about these problems and to find solutions how the guessing effects can be confined. The aim of the presented study was to investigate differences between the two response formats multiple choice formats “1 out of 5” and the sequential response format “1 out of 5” and also to illuminate, if there are cultural differences concerning the psychometric qualities of the two response formats. Therefore, a verbal test battery was designed in German and English and was provided to 225 psychology students at the University of Vienna, Austria and 104 students at the Cypress College in California. The verbal test battery, which consisted out of the two sub-tests: Family Relations Reasoning Test (60 items) and Syllogisms 2009 (10 items) had to be executed *via* computer. The 60 items of the Family Relations Reasoning Test (short: FRRT) were divided into four parallel groups with each 18 items (9 items were designed to be linking items and therefore appeared more than once). Each test taker had to execute one of the FRRT parallel groups and then ten Syllogisms 2009 items. All items of a sub-test were either designed in the multiple choice format “1 out of 5” or in the sequential response format “1 out of 5”. Students were allocated randomly to one of the four FRRT parallel groups and to the response formats. It is important to mention that only the format of the items changed between the various test forms, while the content remained always the same. This is a necessary fact to ascribe possible differences in the solution frequencies to the formats and not to the contents of the items. Concerning the cultural differences within the multiple choice formats, it turned out that because of the fact that about 50 % of the U.S. American sample had to be deleted due to extreme response behaviour, no test of equivalency between the two countries *via* Rasch model was possible and so further statistical analyses could not be calculated. Although the major topic of this thesis could not be investigated, it was still possible to investigate differences between the two response formats within the Austrian sample. Rasch model analyses disclosed a fitting item pool of 6 items after the deletion of four items concerning the sub-test Syllogisms 2009. Within the FRRT not all item easiness parameters could have been estimated because three items had been solved by all test takers. All the remaining FRRT items fitted the assumption of the Rasch model. The investigations showed that if the items are designed with the multiple choice format "1 out of 5", they are significantly easier than if they are in the sequential response format "1 out of 5". This leads to the result that there are massive guessing effects within the multiple choice format "1 out of 5" compared to the sequential alternative.

Keywords: multiple choice response formats, sequential response format, guessing effect, Rasch model

8 Indices

8.1 Bibliography

Andersen, E. B. (1973). *A Goodness of fit tes for the Rasch model.* Psychometrika, 38 (1), 123-140.

Attali, Y. & Bar-Hillel, M. (2003). Guess Where: The Positions of Correct Answers in Multiple-Choice Test items as a Psychometric Variable. *Journal of Educational Measurement*, 40(2), 109-128.

Bortz, J. & Döring, N. (2005). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*. 3. Überarbeitete Auflage. Springer Verlag.

Bühner, M. (2006). *Introduction toconstruction of tests ans questionnairs.* Pearson Studium.

Field, A. (2005). *Discovering Statistics Using SPSS*. SAGE Publications.

Fischer, G. (1974). *Introduction to the theory of psychological tests. Basics and appliances*. Verlag Hans Huber.

Fischer, G. (1995). *Derivations of the Rasch Model*. In G.H. Fischer & I. W. Molenaar (Hrsg.), *Rasch models* (S. 15-38). New York: Springer.

Kubinger, K. D. (2003). *Probabilistic Testtheory*. In K. D. Kubinger & R. S. Jäger (Hrsg.), Schlüsselbegriffe der Psychologischen Diagnostik, (S. 415-423). Weinheim: Beltz PVU.

Kubinger, K. D. (2006). *Psychological Assessment – Theory and Practice of psychological Assessment*. Göttingen: Hogrefe.

Kubinger, K.D. (2009). Psychological Assessment – Theory and Practice of psychological Assessment. Göttingen: Hogerefe.

Kubinger, K. & Gottschall, C. (2007). *Item difficulty of multiple choice tests dependant on different item response formats – An experiment in fundament research on psychological assessment*. Psychology Science, 49(4), 361-374.

Kubinger, K. D., Rasch, D. & Moder, K. (2009). *About the legend of the assumptions of the Student's t- Test for independant samples.* Psychologische Rundschau, 60(1), 26-27.

Liedlbauer, M. (2009). *About the image of multiple-choice tests in high-schools and at the University: Acceptance and Reasonability.* Unpublished Diploma Thesis of the Division of Psychological Assessment, University of Vienna.

Litzenberger, M., Gnambs, T. & Punter, J. (2005). *Guidlines about the construction of multiple- choice items examplified on the lecture "Psychology as Science" at the Department of Psychology*. Unveröff. Forschungsbericht der Fakultät für Psychologie, Wien: Universität Wien.

Mittring, G. & Rost, D. (2008). *Accursed Distractors – about the advantage of theoretical distractoranalyses for matrices tests (for good and highly skilled test takers).* Diagnostica, 54(4), 193-201.

Nortmann, U. (1996). *Modal Syllogisms, possible spaces, essentialism.* deGruyter.

PASW Statistics 18 for Windows, Vers. 18 2009. SPSS IBM Company.

Poinstingl, H. (2001). *The checkout of a computer test battery for crystallized and fluid intelligence*. Diploma Thesis of the Division of Psychological Assessment, University of Vienna.

Poinstingl, H. (2009). *The Linear Logistic Test Model (LLTM) as the methodological foundation of item generating rules for a new verbal reasoning test.* Psychology Science Quaterly, Volume 51, 2009 (2), pp. 123 – 134.

Poinstingl, H., Mair, P. & Hatzinger, R (2007). *Manual to the Softwarepackage eRm (extended Rasch modeling)*. Pabst Science Publishers.

R for Windows, Vers. 2.10.0. 2009. Wirtschaftsuniversität Wien.

Rasch, D. & Kubinger, K. (2006). *Statistik für das Psychologiestudium*. Elsevier: Spektrum.

Reber, A. & Reber, E. (2001). *The Penguin Dictionary of Psychology*. Third Edition. Penguin Reference.

Rost, J. (2004). *Hornbook Testtheory and Testconstruction.* Verlag Hans Huber.

Schechtner, C. (2009). *Developement of rational itemconstruction principles on the basis of a verbal reasoning test.* Unpublished Diploma Thesis of the Division of Psychological Assessment, University of Vienna.

Skoda, S. (2005). *The Family- Reasoning- Test.* Unpublished Diploma Thesis of the Division of Psychological Assessment, University of Vienna.

Zöfel, P. (2003). *Statistik for Psychologists*. Pearson Studium.

http://www.statistics.gov.uk. English FRRT names.

http://www.statistics4u.info/fundstat_germ/ee_welch_test.html. Welch-Test.

8.2 List of figures

Figure 3-1 ... 29
Figure 3-2 ... 31
Figure 3-3 ... 35
Figure 3-4 ... 35
Figure 3-5 ... 41
Figure 3-6 ... 42
Figure 4-1 ... 45
Figure 4-2 ... 45
Figure 4-3 ... 48
Figure 4-4 ... 49
Figure 4-5 ... 50
Figure 4-6 ... 51
Figure 4-7 ... 52
Figure 4-8 ... 52
Figure 4-9 ... 53
Figure 4-10 ... 53
Figure 4-11 ... 54
Figure 4-12 ... 54
Figure 4-13 ... 64
Figure 4-14 ... 65
Figure 4-15 ... 66
Figure 4-16 ... 67
Figure 4-17 ... 69
Figure 4-18 ... 71
Figure 4-19 ... 82
Figure 6-1 ... 89

8.3 List of tables

Table 2-1……..14
Table 3-1……..19
Table 3-2……22
Table 3-3……23
Table 3-4……25
Table 3-5……...26
Table 3-6……27
Table 3-7……32
Table 3-8……..32
Table 3-9……...33
Table 3-10……40
Table 4-1……..46
Table 4-2……48
Table 4-3……...49
Table 4-4……50
Table 4-5……51
Table 4-6……..52
Table 4-7……..52
Table 4-8……..53
Table 4-9……..54
Table 4-10……56
Table 4-11……56
Table 4-12……57
Table 4-13……58
Table 4-14……58
Table 4-15……59
Table 4-16……59
Table 4-17……60
Table 4-18……60
Table 4-19……61
Table 4-20……61
Table 4-21……63
Table 4-22……64
Table 4-23……64
Table 4-24……68
Table 4-25……69
Table 4-26……72
Table 4-27……73
Table 4-28……73
Table 4-29……73
Table 4-30……74
Table 4-31……74
Table 4-32……75
Table4-33……77
Table 4-34……79
Table 4-35……80
Table 4-36……81
Table 4-37……81
Table 4-38……83

9 Appendix

Appendix 9.1: The verbal test battery: Access to the verbal test battery

Figure 9.1-1: Screenshot of the homepage “www.univie.ac.at/psydiag” which offered an access to the verbal computer test battery called “Verbat 2009”.

Figure 9.1-2: Entrance page to the English version of the verbal test battery.

Figure 9.1-3: Introduction and instruction to the verbal test battery.

Your next step

Verbal test battery "VerBat 2009"

When you click on the link above, a window will pop up in which you will find a button labeled ***[Connect]***. If you click on this button, the remote desktop software will be started and a connection to the terminal server will be established. Verbat 2009 is best viewed *with a resolution of 1024* x *768.*

(It is possible that you may receive a warning message such as "The identity of the remote computer cannot be verified. Do you want to connect anyway?" If this occurs, then click on the button named ***[Yes].*** *This message is caused by a non-configured server authentification, but the security of your computer is of course not endangered.)*

In the following window you will be asked to enter your username and password, which you should have received by E-Mail or from your instructor. After pushing the button **[OK],** you will be logged on to the server and the test administration will automatically be started. After finishing the test, you will be automatically logged off.

Please complete the entire test in one sitting!

Advice: If a connection cannot be established with the terminal server and a window with many characters is opened by your internet browser, then save the link "Verbal test battery …" on your computer. After clicking twice on the saved file, a connection should be established and you will be logged on to the server.

Figure 9.1-4: Final instruction about the login into the verbal test battery.

Appendix 9.2: The Verbal Test battery: Handout and soziodemographic questionnaire

A scientific cooperation of
← The University of Vienna
and the
Cypress College,
California →

Dear student,

Thank you for participating in this research study, of the University of California, Los Angeles and the University of Vienna, Austria.

If you open the web side: www.univie.ac.at/psydiag and click on

"VerBAT (English) "

you will be connected to the test system.

To log yourself into the test system, please click "open" and then use the following password and username:

Username (german "Benutzername"):	english
Password: (german "Kennwort")	test

Important information

1) Please do not use any helping utilities except of your brain! (Do not use paper and pencils nor mobile phones.)

2) This test consists of two subtests:
   → The "Family Relation Reasoning Test": where you will read about family relations and then find a correct connection between two family members.
   → The "Syllogisms 2009" will give you two statements in each item and you have to find a correct conclusion.

3) Before each subtest, you will be asked about your name/ identification. Please use the following code instead of your name:
   → First two letters of your family name and the
   → First two letters of your first name and the
   → First two letters of your student identification number

   You will get a personal (anonymous) online feedback through this code.

Thank you very much for helping this research cooperation of the University of Vienna and the Cypress College.

Figure 9.2-1: Handout with information for the participants. This letter was handed out to the students of the Cypress College. The Austrian students got the same letter but in German.

Questions about the person			
Name		Profession	
Degree in progress		Nationality	
Native tongue		Age	
Sex			

Figure 9.2-2: Soziographic questionnaire at the beginning of each subtest (English version).

Appendix 9.3: Family Relations Reasoning Test – instruction, items, group distribution and unpublished logo draft.

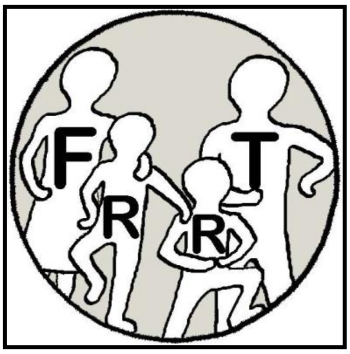


Figure 9.3-1: Unpublished logo draft of the Family Relations Reasoning Test (designed by Brigitte Hansmann).

Family Relations Reasoning Test (FRRT)

is a computer based test, where you have to determine the correct familial relationship based on the information provided.

Example

Claire is the daughter of Sophia. Sophia is the daughter of Maria. Maria is the __________ of Claire.
In this case, the right answer would be: Maria is the **grandmother** of Claire.

Test

Following the description of a family, a series of potential relationships will be offered and you have to find the correct relation between two particular people. Thereby five possible family relations will be offered and you have to decide which conclusion is the correct relation between the two family members. First click the right family relation and then click **[NEXT]**. The computer will automatically switch to the next task.

The completion of the tasks has no time limit. You may not change your answer once it has been saved.

Some provided help for you:

Read the following information very carefully. You can access this help also during the testing, if you click the question mark.

Accessory Information

Married couples in this test only have children together.
If Anna is the mother of Simon and Peter is the husband of Anna, then Simon is automatically the son of Peter as well.
The tasks are simply solvable through logical reasoning. The story text contains all the information which you will need to solve the tasks.

Good luck!
☺

Figure 9.3-1: Screenshot of theFRRT instruction page 1.

Family Relations Reasoning Test

(FRRT)

The Family Relations Reasoning Test is a computer based test, where you have to determine the correct familial relationship based on the information provided.

Example

Claire is the daughter of Sophia. Sophia is the daughter of Maria. Maria is the __________ of Claire.
In this case, the right answer would be: Maria is the **grandmother** of Claire.

Test

Following the description of a family, a series of potential relationships will be offered and you have to find out whether a given alternative is true or false. If you find an answer to be correct, then click the **[TRUE]**-option, followed by the **[NEXT]**-button.

If you think, that a conclusion is not correct, then choose the **[FALSE]**–option, followed by the **[NEXT]**-button**.** The computer will then give you more possible family relationships. If you find all possible alternatives to be false, then the computer will automatically switch to the next task.

The completion of the tasks has no time limit. You may not change your answer once it has been saved.

Some provided help for you:

Read the following information very carefully. You can access this help also during the testing, if you click the question mark.

Figure 9.3-2: Screenshot of the FRRT instruction page 2.

Accessory Information

Married couples in this test only have children together.
If Anna is the mother of Simon and Peter is the husband of Anna, then Simon is automatically the son of Peter as well.
The tasks are simply solvable through logical reasoning. The story text contains all the information which you will need to solve the tasks.

Good luck!

☺

Figure 9.3-3: Screenshot of the FRRT instruction page 3.

Comment: Table 9.3-1 shows all the items of the FRRT lined up with their numbers. The first 50 items are the old items of Schechtner, 2009 while the items 51 to 60 are the new items that were established for this study. The items are shown with their item number, the story text, the solution, the distractors, the solution position and the item easiness parameters (if an estimation was possible). All the solution suggestions were presented in full sentences in the test version.

Table 9.3-1
Overview of the FRRT items. The items are shown with their particular item number, the story text, their solution, the distractors, the solution position and the item easiness parameters (from left to right).

Item number	Story text	Solution	Distractors	Solution position	Item easiness parameter
Instruction item 1	Karen has a daughter called Tina.Felix is the brother of Tina. Karen is ____ of Felix.	Karen is the mother of Felix.	grandmother aunt cousin sister	2	instruction item
Instruction item 2	Harry is the son of Ella. Ella has a daughter called Linda. Linda is ____ of Harry.	Linda is the sister of Harry.	cousin niece daughter mother	4	instruction item
frrt 1	Lara is the daughter of Kurt. Iris is the wife of Kurt. Lara is ______ of Iris.	Lara is the daughter of Iris	aunt grandmother niece sister	2	estimation was not possible
frrt 2	Robin is the son of Bruno. Bruno is the husband of Layla. Robin is ______ of Layla.	Robin is the son of Layla	grandson nephew brother grandfather	5	1.671
frrt 3	Irvin is the husband of Doris and has a son called Alan. Viola is the daughter of Doris. Irvin is ______ of Viola.	Irvin is the father of Viola	uncle grandson brother grandfather	2	0.339
frrt 4	Ivan is the father of Edwin. Paula is the daughter of Ivan. Kathy is the mother of Paula. Edwin is ______ of Paula.	Edwin is the brother of Paula	grandson uncle nephew father	1	0.225
frrt 5	Leo is the father of Emma. Emma is the mother of Mark. Mark is ______ of Leo.	Mark is the grandson of Leo	cousin son grandfather brother	5	-0.627
frrt 6	Caleb is the father of Jacob. Amber is the daughter of Jacob. Amber is ______ of Caleb.	Amber is the granddaughter of Caleb	sister mother aunt daughter	1	0.929
frrt 7	Peter is the grandson of Mary. Gaby is the mother of Peter. Peter has a sister called Anna. Mary is ______ of Anna.	Mary is the grandmother of Anna	granddaughter cousin aunt mother	3	-0.377
frrt 8	Rose is the grandmother of Kevin. Fred is the husband of Rose and the father of Hanna. Fred is ______ of Kevin.	Fred is the grandfather of Kevin	cousin grandson brother father	4	1.374
frrt 9	Tom has a sister called Iris. Alena is the daughter of Iris. Tom is ______ of Alena.	Tom is the uncle of Alena	brother son grandfather cousin	3	estimation was not possible
frrt 10	Phil is the brother of Aidan. Aidan is the father of Billy. Billy is ______ of Phil.	Billy is the nephew of Phil	brother grandfather cousin nephew	5	2.384

frrt 11	Alice is the aunt of Simon. Tina has a brother called Simon. Rudy is the son of Alice. Simon is ______ of Rudy.	Simon is the cousin of Rudy	grandfather grandson nephew son	5	1.203
frrt 12	Mia has a sister called Diana and a husband called Sam. Ella is the daughter of Diana. Mia is ______ of Ella.	Mia is the aunt of Ella	grandmother sister daughter granddaughter	1	0.740
frrt 13	Trudy is the daughter of Randy. Maria is the wife of Randy. John is the brother of Trudy. John is ______ of Maria.	John is the son of Maria	grandson nephew uncle grandfather	2	estimation was not possible
frrt 14	Felix has a son called Chris and a wife called Lena. Lena is the mother of Cathy. Chris is ______ of Cathy.	Chris is the brother of Cathy	cousin father grandfather grandson	4	1.244
frrt 15	Lola is the mother of Alec and the sister of Edgar. Aaron is the husband of Lola. Alec has a sister called Erin. Erin is ______ of Aaron.	Erin is the daughter of Aaron	aunt granddaughter grandmother cousin	1	-0.087
frrt 16	Oscar has a daughter called Sarah and a wife called Linda. Linda has a brother called Kurt. John is the son of Linda. Sarah is ______ of John.	Sarah is the sister of John	grandmother aunt daughter niece	3	0.625
frrt 17	Edith has a husband called Gavin and a daughter called Hanna. Ben is the brother of Gavin and the father of Lucia. Max is the brother of Hanna. Max is ______ of Gavin.	Max is the son of Gavin	grandfather father cousin brother	3	-0.807
frrt 18	Owen has a sister called Keira and a wife called Emma. Keira is the wife of Hugo. Lee has a father called Hugo. Mia is the sister of Lee. Keira is ______ of Mia.	Keira is the mother of Mia	sister niece daughter granddaughter	4	-0.096
frrt 19	Bruce is the father of Tom and the son of Aidan. Suzy is the sister of Tom. Suzy is ______ of Aidan.	Suzy is the granddaughter of Aidan	aunt mother niece sister	4	0.637
frrt 20	#Megan is the daughter of Abby and has a son called Toby. Abby is the wife of Frank. Frank is ______ of Toby.	Frank is the grandfather of Toby	brother father cousin grandson	4	-0.251
frrt 21	Erica is the grandmother of Ralph. Ralph has a sister called Lucy. Diana is the mother of Lucy. Erica has a husband called Carl. Lucy is ______ of Carl.	Lucy is the granddaughter of Carl	sister aunt cousin niece	5	-0.286
frrt 22	Eric is the son of Paula and has a wife called Carla. Paula has a husband called Rick. Bruce is the son of Carla. Bruce is ______ of Paula.	Bruce is the grandson of Paula	cousin uncle nephew brother	4	0.244
frrt 23	Gina is the mother of Flora and Jan. Edwin is the father of Lucas. Megan is the wife of Edwin. Lucas has a daughter called Flora. Megan is ____________ of Flora.	Megan is the grandmother of Flora	mother cousin granddaughter sister	4	0.899
frrt 24	Tanya is the mother of Jenny and the daughter of Nelly. Mario is the husband of Tanya and the brother of Chad. Jenny has a brother called Edgar. Edgar is _______ of Nelly.	Edgar is the grandson of Nelly	father grandfather cousin nephew	1	1.671
frrt 25	Paul is the father of Jim. Molly is the daughter of Paul and has a daughter called Agnes. Agnes is ____________ of Jim.	Agnes is the niece of Jim	granddaughter mother grandmother daughter	4	0.903
frrt 26	Lilly has a mother called Eva. Eva is the mother of Mark. Neil is the son of Mark. Lilly is ______ of Neil.	Lilly is the aunt of Neil	daughter cousin granddaughter mother	5	0.140
frrt 27	Ruth is the mother of Kenny. Laura is the daughter of Ruth and has a husband called Seth. Vince has a mother called Laura. Vince is ______ of Kenny.	Vince is the nephew of Kenny	uncle cousin son grandfather	1	-0.075

frrt 28	Jonas is the son of Elise. Louis has a sister called Elise and a dauhter called Nina. Ivy is the wife of Louis. Nina is ______ of Jonas.	Nina is the cousin of Jonas	aunt granddaughter grandmother niece	3	0.225
frrt 29	David has a niece called Sofia. Cora is the wife of David. Sofia has a father called Caleb and a brother called Lucas. Mark is the son of Cora. Mark is ______ of Sofia.	Mark is the cousin of Sofia	uncle son nephew grandson	5	0.408
frrt 30	Edwin has an uncle called Joe and a mother called Doris. Amber is the wife of Joe and the mother of Mitch. Linda is the sister of Edwin. Amber is ______ of Linda.	Amber is the aunt of Linda	sister niece cousin mother	1	0.382
frrt 31	Noah is the son of Anna. Jacob is the husband of Anna. Rosy has a husband called Noah. Maya is the daughter of Rosy. Maya is ______ of Jacob.	Maya is the granddaughter of Jacob	aunt grandmother sister daughter	4	0.602
frrt 32	Betty is the mother of Lisa and the daughter of Willy. Willy has a wife called Helen. Jacob is the brother of Lisa. Helen is _____ of Jacob.	Helen is the grandmother of Jacob	sister cousin mother granddaughter	3	0.882
frrt 33	Julia is the daughter of Evan and the wife of Nick. Evan is the husband of Sandy. Nick is the father of Axel. Axel has a sister called Keira. Axel is ______ of Sandy.	Axel is the grandson of Sandy	grandfather father cousin brother	5	0.048
frrt 34	Romeo is the father of Sonya. Wilma has a husband called Heath and a son called Romeo. Bella is the daughter of Heath. Sonya is the sister of Brady. Heath is ______ of Brady.	Heath is the grandfather of Brady	father uncle brother nephew	3	0.225
frrt 35	Chad is the father of Adam and has a wife called Cindy. Vicky has a son called Simon and she is the daughter of Cindy. Lara is the wife of Adam and she has a daughter called Lea. Cindy is ______ of Lea.	Cindy is the grandmother of Lea	sister cousin granddaughter niece	4	-0.087
frrt 36	Mary is the mother of Aaron and the daughter of Henry. Henry has a wife called Lydia and a son called Max. Aaron has a sister called Julie and a father called Peter. Lydia is ______ of Julie.	Lydia is the grandmother of Julie	mother aunt daughter granddaughter	1	1.927
frrt 37	Tim is the father of Hugo and the son of Fiona. Irene has a brother called Hugo. Fiona is the mother of Lucy. Lucy is ______ of Irene.	Lucy is the aunt of Irene	daughter sister grandmother cousin	1	-1.251
frrt 38	Amy is the mother of David and the wife of Peter. Peter is the father of Aidan. Brian has a father called Aidan. David is ______ of Brian.	David is the uncle of Brian	nephew brother son grandfather	1	-0.633
frrt 39	Katie is the mother of Phil and the sister of Elena. Elena is the mother of Chad and the wife of Carl. Phil has a sister called Laura. Laura is ______ of Chad.	Laura is the cousin of Chad	mother niece aunt granddaughter	4	0.218
frrt 40	Bella has a son called Jan and a daughter called Ella. Niles is the son of Ella. Jan has a son called Joe and a wife called Vicky. Joe is _____ of Niles.	Joe is the cousin of Niles	nephew brother father uncle	4	1.374
frrt 41	Henry is the husband of Gina and the father of Keith. Gina has a father called Josef. Macy is the wife of Josef. Haley is the daughter of Macy and has a daughter called Abby. Gina is ______ of Abby.	Gina is the aunt of Abby	daughter cousin mother granddaughter	3	0.602
frrt 42	Ivan has a father called David. Layla is the wife of David. Pam is the daughter of Layla and has a daughter called Emily. Rita is the wife of Ivan and she has a son called Tim. Ivan is ______ of Emily.	Ivan is the uncle of Emily	son brother cousin father	1	-0.562
frrt 43	Tara is the mother of Paige. Hazel is the mother of Alan. Dan has a wife called Hazel. Elias has a sister called Paige. Alan is the husband of Tara. Elias is ______ of Dan.	Elias is the grandson of Dan	cousin father grandfather brother	5	-1.626
frrt 44	Jim is the father of Alena. Dana is the daughter of Edgar. Jenny has a husband called Edgar. Lucas is the brother of Alena. Dana is the wife of Jim. Jenny is ______ of Lucas.	Jenny is the grandmother of Lucas	mother sister cousin aunt	1	-0.682

frrt 45	Naomi is the mother of Brian and the sister of Ken. Dave is the father of Roger and he has a wife called Elisa. Brian has a sister called Selma. Roger is the husband of Naomi. Elisa is ______ of Selma.	Elisa is the grandmother of Selma	granddaughter daughter mother cousin	4	-0.245
frrt 46	Leon is the father of Owen. Nadia is the daughter of Mira and the sister of Aidan. Mira has a husband called Tom. Heidi has a brother called Owen. Nadia is the wife of Leon. Tom is ______ of Heidi.	Tom is the grandfather of Heidi	cousin nephew brother father	2	-1.772
frrt 47	Bob is the father of Ralph. Ralph has a sister called Lisa. Iris is the wife of Bob and the sister of Mia. Henry is the husband of Mia. Henry is ______ of Lisa.	Henry is the uncle of Lisa	cousin father son grandson	2	1.190
frrt 48	Bella is the daughter of Tessa and has a brother called Bruno. Simon is the husband of Tessa and has a nephew called Alvin. Ivy is the mother of Alvin. Ivy is ____ of Bruno.	Ivy is the aunt of Bruno	cousin granddaughter grandmother daughter	4	-0.608
frrt 49	Milo is the father of Nora and the brother of Alina. Julia is the daughter of Alina. Brian is the brother of Nora. Brad is the father of Julia. Julia is the sister of Alec. Brian is ______ of Alec.	Brian is the cousin of Alec	son father nephew grandfather	2	0.054
frrt 50	Emily is the mother of Toby and the sister of Hugo. Hugo has a wife called Gwen and a daughter called Wendy. Toby has a sister called Ada. Frank is the brother of Wendy. Ada is ______ of Frank.	Ada is the cousin of Frank	granddaughter niece aunt mother	4	0.740
frrt 51	Gaby is the daughter of Alec. Linda is the wife of Alec and has a father called Edwin. Cindy is the sister of Emma and has a husband called Edwin. Keith is the son of Cindy and the husband of Megan. Gaby is ______ of Keith.	Gaby is the sister of Keith.	cousin niece mother daughter	3	-2.227
frrt 52	Luna has a husband called Alden and a daughter called Paula. Sarah is the sister of Nora. Max is the father of Luna. Wendy has a husband called Fred and a daughter called Nora. Fred has a sister called Luna. Max is ______ of Sarah.	Max is the brother of Sarah.	grandson grandfather father son	3	0.408
frrt 53	Peter is the husband of Janet and he has a grandson called John. Willy is the husband of Clara. John has a brother called Paul and he is the son of Willy. Ella is the aunt of John. Ben has a son called Willy. Clara is _____ of Janet.	Clara is the daughter of Janet	niece mother grandmother aunt	3	-1.782
frrt 54	Heath is the husband of Anna. Emma has a mother called Tina. Rudy is the brother of Oscar. Layla has a granddaughter called Emma. Tina is the wife of Oscar. Rudy has a father called Heath and a wife called Nyla. Emma is ______ of Anna.	Emma is the granddaughter of Anna	daughter mother niece grandmother	5	-2.322
frrt 55	Owen is the father of Bella. Nina has a son called Tim and a husband called Felix. Mia is the sister of Lance. Bella has a brother called Felix. Lance is the father of Ralph and has a wife called Bella. Tim is ______ of Ralph.	Tim is the cousin of Ralph	father grandson son grandfather	3	0.225
frrt 56	Karen is the mother of Mandy. Hanna has a husband called Ethan and a daughter called Lara. Alvin is the father of Sonya and the husband of Mandy. Hanna has a mother called Karen. Iris is the sister of Sonya. Ethan is ______ of Sonya.	Ethan is the uncle of Sonya	cousin grandfather son brother	3	-0.682
frrt 57	Eric's mother Karen is the niece of Noah's brother Noel. Cora is the daughter of Tina. Sandy and Cora are sisters. Tina has a husband called Noel. Ben is the brother of Sandy. Ben is ______ of Karen.	Ben is the cousin of Karen	nephew uncle father brother	3	-2.322
frrt 58	Carl's sister Megan is the wife of Ken's son Nick. Alice is the wife of Ken and the mother of Suzy. Fiona's father Theo is the brother of Suzy. Nick is ______ of Fiona.	Nick is the uncle of Fiona	cousin brother uncle father	3	-0.926

frrt 59	Tanya is the niece of Caleb's father Tom. Luis is the uncle of Anna's granddaughter Lydia. Lee is the brother of Tanya. Lydia is the sister of Caleb. Piper has a son called Lee. Lydia is ______ of Piper.	Lydia is the niece of Piper	grandmother cousin sister mother	4	-2.730
frrt 60	Flora is the wife of Seth's brother Keith. Mia is the sister of Oscar's son Bruno. Lucia is the wife of Rudy's brother Oscar. Mia has a son called Keith. Lucia is ______ of Seth.	Lucia is the grandmother of Seth	mother cousin sister niece	5	-1.674

Appendix 9.4: Syllogisms 2009 – instruction, items and group distribution

Syllogisms 2009

is a computer based test, which shows you two statements (premises) in each task. With the help of these two statements you have to find a logical conclusion.

For example

The premises: **All As are Bs. All Bs are Cs.**
to the conclusion that: **All As are Cs.**
Since **all** As are Bs - AND - **all** Bs are Cs, **all** As must be Cs as well.

In this test, it is possible, that you may find some logical conclusions which seem absurd or conceptual incorrect to you. This is based on the fact that you can educe correct conclusions from absurd statements without paying attention to the actual meaning of the words.

The Testing

Each task offers two statements and possible conclusions in multiple choice format. You have to decide which of the conclusions is correct. First choose the conclusion which seems right to you and then click **[NEXT]**. The computer will switch to the next task then.

The completion of the tasks has no time limit. You may not change your answer once it has been saved.

Some provided help for you:

Read the following information very carefully. You can access this help also during the testing, if you click the question mark.

Figure 9.4-1: Screenshot of the first Syllogisms 2009 instruction page.

LOGICAL COMBINING

1. If it is assumed that "All As are Bs", it is not always necessarily valid that "All Bs are As" (for example: All poets are humans.)

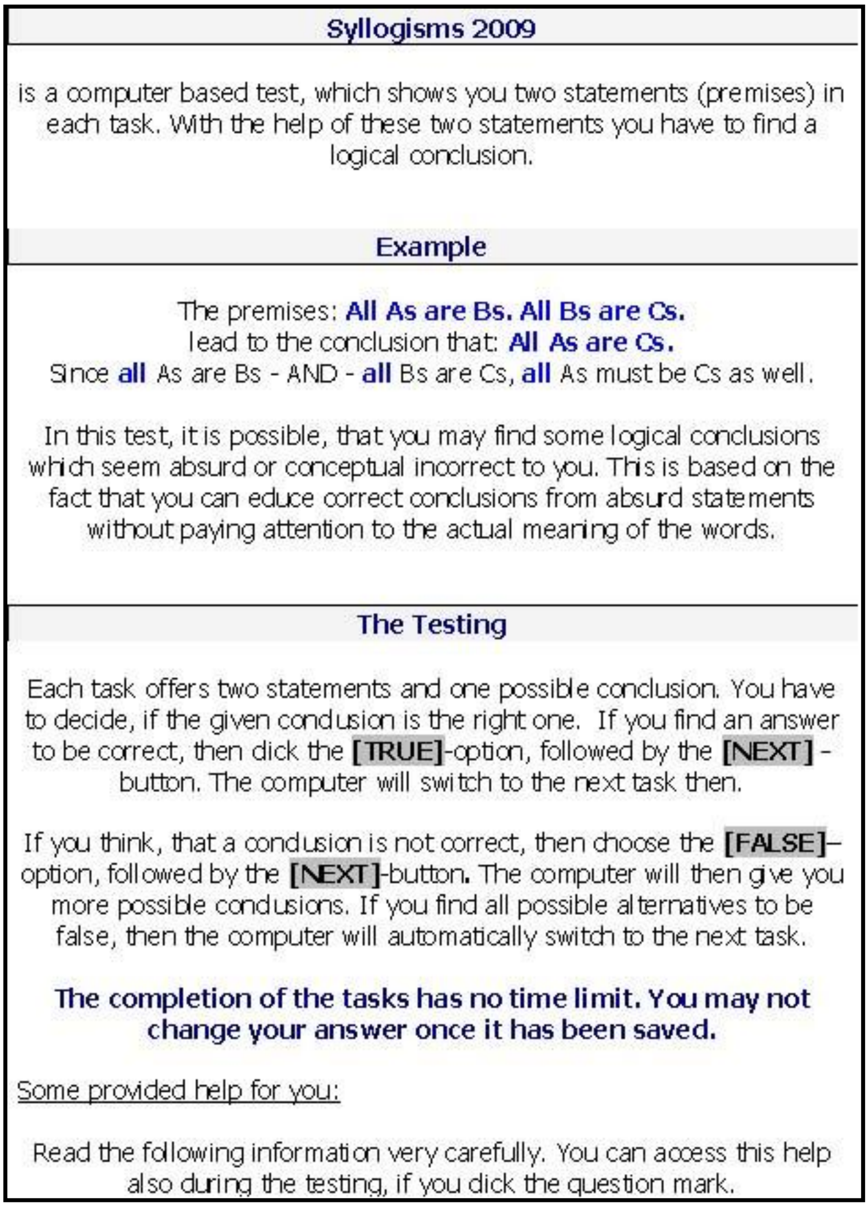
2. If it is assumed that "Some As are Bs", then "Some Bs are As" is valid as well. (for ex.: Some architects are technicians.)

3. If it is assumed that "All As are no Bs", then "All Bs are no As" is valid as well. (for ex.: All fish are no reptiles.)

4. If it is assumed that "Some As are no Bs", then it is not necessarily valid that "Some Bs are no As". (for ex.: Some US- Americans are no Californians.)

Good luck!
☺

Figure 9.4-2: Screenshot of the second Syllogisms 2009 instruction page. This information was also given if the help button was clicked.



Syllogisms 2009

is a computer based test, which shows you two statements (premises) in each task. With the help of these two statements you have to find a logical conclusion.

Example

The premises: **All As are Bs. All Bs are Cs.**
lead to the conclusion that: **All As are Cs.**
Since **all** As are Bs - AND - **all** Bs are Cs, **all** As must be Cs as well.

In this test, it is possible, that you may find some logical conclusions which seem absurd or conceptual incorrect to you. This is based on the fact that you can educe correct conclusions from absurd statements without paying attention to the actual meaning of the words.

The Testing

Each task offers two statements and one possible conclusion. You have to decide, if the given conclusion is the right one. If you find an answer to be correct, then click the **[TRUE]**-option, followed by the **[NEXT]** - button. The computer will switch to the next task then.

If you think, that a conclusion is not correct, then choose the **[FALSE]**–option, followed by the **[NEXT]**-button. The computer will then give you more possible conclusions. If you find all possible alternatives to be false, then the computer will automatically switch to the next task.

The completion of the tasks has no time limit. You may not change your answer once it has been saved.

Some provided help for you:

Read the following information very carefully. You can access this help also during the testing, if you click the question mark.

Figure 9.4-3: Screenshot of the third Syllogisms 2009 instruction page.

Comment: Table 9.4-1 shows all the items of the Syllogisms 2009 lined up with their item numbers, the premises, the solution, the distractors, the solution position and the item easiness parameters (if an estimation was possible). Items that had to be excluded are marked in grey.

Table 9.4-1
Overview of the Syllogisms 2009 items. The items are orders with their particular item number, the premises, their solution, the distractors, the solution position and the item easiness parameters (if an estimation was possible).

Item number	Story text	Solution	Distractors	Solution position	Item easiness parameter
Instruction item 1	Premise 1: All domestic cats are predators Premise 2: All nibblers are domestic cats	✓ All nibblers are predators	× All nibblers are not predators × No nibblers are predators × Some nibblers are not predators × Some nibblers are predators	2	instruction item
Instruction item 2	Premise 1: All Quambs are Hambs Premise 2: All Mambs are Quambs	✓ All Mambs are Hambs	× Some Mambs are not Hambs × Some Mambs are Hambs × All Mambs are not Hambs × No Mambs are Hambs	4	instruction item
syl 1	Premise 1: All odd numbers are natural numbers Premise 2: All prime numbers bigger than two are odd numbers	✓ All prime numbers bigger than two are natural numbers	× Some prime numbers bigger than two are natural numbers × All prime numbers bigger than two are not natural numbers × Some prime numbers bigger than two are not natural numbers × No prime numbers bigger than two are natural numbers	1	2.037
syl 3	Premise 1: All steamships are no sailing ships Premise 2: Some freighters are sailing ships	✓ Some freighters are not steamships	× All freighters are steamships × Some freighters are steamships × All freighters are not steamships × No freighters are steamships	4	-1.371
syl 4	Premise 1: All automorphisms are isomorphisms Premise 2: All embeddings are not isomorphisms	✓ All embeddings are not automorphisms	× All embedings are automorphisms × Some embeddings are automorphisms × Some embeddings are not automorphisms × Some automorphisms are embeddings	3	1.402
syl 6	Premise 1: Some A are not B Premise 2: All A are C	✓ Some C are not B	× All C are B × Some C are B × All C are not B × No C are B	3	-1.483
syl 8	Premise 1: All Texans are not Californians Premise 2: Some Virginians are Texans	✓ Some Virginians are not Californians	× All Virginians are Californians × No Virginians are Californians × All Virginians are not Californians × Some Virginians are Californians	5	-0.876
syl 10	Premise 1: If all clams were not animals Premise 2: and some snails were clams	✓ then some snails would not be animals	× then all snails would be animals × then some snails would be animals × then all snails would not be animals × then no snails would be animals	4	-0.339
syl 11	Premise 1: All drills are not home appliances Premise 2: Some drills are electronic appliances	✓ Some electronic appliances are not home appliances	× All electronic appliances are home appliances × Some electronic appliances are home appliances × All electronic appliances are not home appliances × No electronic appliances are home appliances	5	-0.438
syl 13	Premise 1: All flutes are not wind instruments Premise 2: Some flutes are string instruments	✓ Some string instruments are not wind instruments	× Some string instruments are wind instruments × All string instruments are not wind instruments × All string instruments are wind instruments × No string instruments are wind instruments	1	0.420
syl 14	Premise 1: If some snakes were not vipers Premise 2: and all snakes were reptiles	✓ then some reptiles would not be vipers	× then all reptiles would be vipers × then some reptiles would be vipers × then all reptiles would not be vipers × then no reptiles would be vipers	4	-1.200
syl 16	Premise 1: If some ministers were not members of a political party Premise 2: and all ministers were politicians	✓ then some politicians would not be memebers of a political party	× then all politicians would be members of a political party × then all politicians would not be members of a political party × then some politicians would be members of a political party × then no politicians would be members of a political party	3	1.849

Appendix 9.5: Continuative sample descriptions

Comment: Appendix 9.5 gives additional information about the samples. Table 9.5-1 gives an overview of the excluded FRRT items in each sub-sample, table 9.5-2 shows the distribution of the test takers to the sub-samples and table 9.5-3 shows the four parallel groups of the FRRT.

Table 9.5-1
Excluded items of the FRRT due to an extreme answering behaviour of the test takers within the sub- samples. Only the items 1, 9 and 13 had to be excluded within the Austrian sample because they were solved by all test takers. The other items were kept for further analyses.

Excluded FRRT items

Reason for exclusion	Austrian sample n= 225	Mixed sample n= 282
Item(s) solved by all test takers	Items: 1, 9, 13	Item: 9
Items solved by all test takers within one subgroup of the partition criterion: median	Items: 5, 6, 8, 10, 12, 15, 16, 21, 24, 25, 26, 28, 29, 31, 32, 35, 36, 38, 39, 40, 41, 43, 45, 47, 50, 53, 58, 62	Items: 1, 6, 8, 10, 12, 13, 15, 16, 21, 24, 26, 28, 29, 31, 32, 35, 36, 40, 41, 43, 45, 47, 50, 53, 58, 62
Items solved by all test takers within one subgroup of the partition criterion: sex	Items: 2, 6, 8, 10, 11, 35, 36, 40, 47	1, 8, 10, 11, 40, 47
Items solved by all test takers within one subgroup of the partition criterion: format	Items: 10, 15, 23, 32, 36, 40, 41, 50, 53.	Items: 1, 10, 13, 53
Items solved by all test takers within one of the subgroups of the partition criterion: country	no item exclusions	Items: 1, 13

Table 9.5-2
Overview of the distribution of the test takers to the various sub-samples.

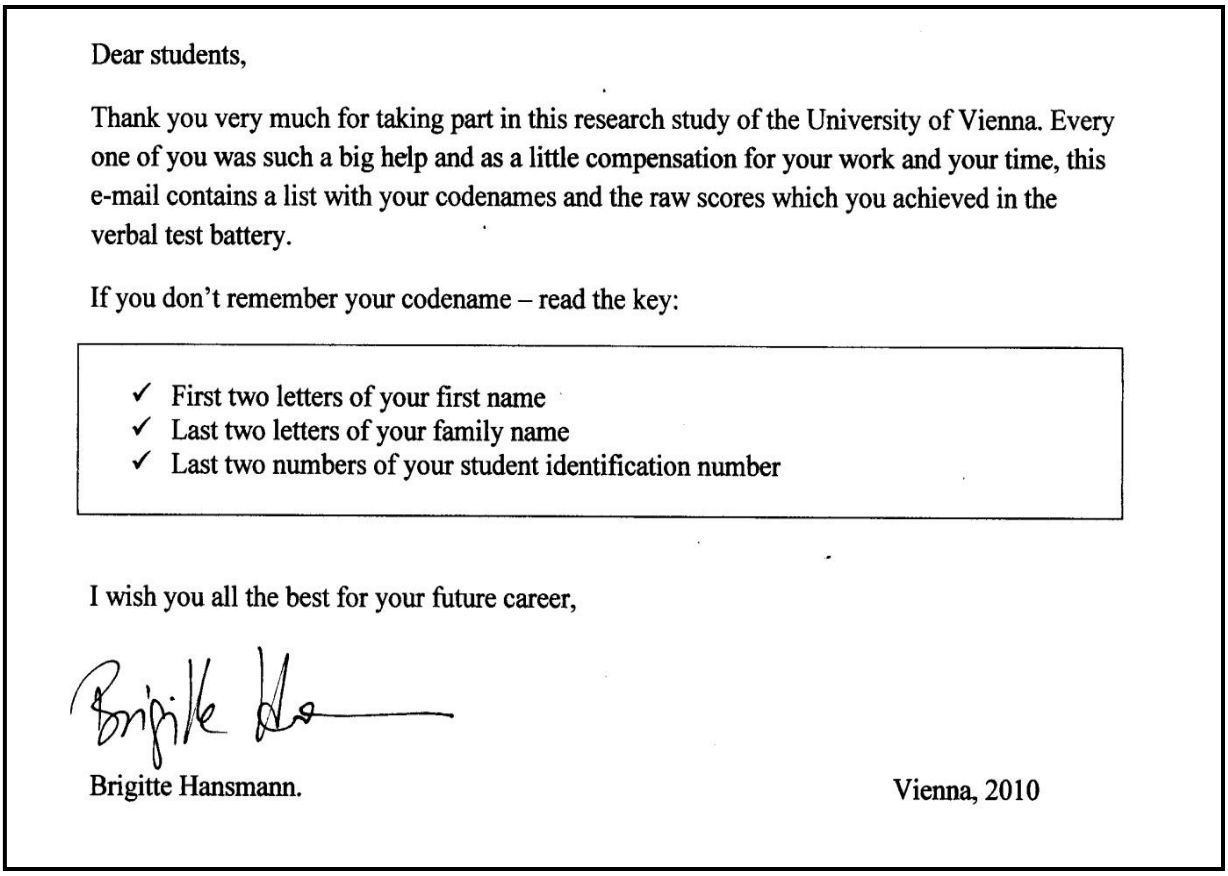
Sub-test	External Partition Criteria		Frequency	Percentage
FRRT	sex	men	170	75.6
		women	55	24.4
	format	MC	112	49.8
		SEQ	113	50.2
Syllogisms 2009	sex	men	52	76.5
		women	169	23.5
	format	MC	109	49.3
		SEQ	112	50.7

Table 9.5-3
Overview of the items to the four parallel groups of the FRRT and their item easiness parameters. Items written in thick letters are linking items and items which are marked in grey are excluded items. The items are ordered as they were offered to the students.

FRRT group 1		FRRT group 2		FRRT group 3		FRRT group 4	
Item number	easiness parameter	Item number	easiness parameter	Item number	easiness parameter	Item number	easiness parameter
3	0.339	2	1.671	1	deleted	5	-0.627
6	0.929	4	0.225	9	deleted	12	0.740
8	1.374	7	-0.377	**10**	**2.384**	**14**	**0.244**
14	**0.244**	**10**	**2.384**	11	1.203	15	-0.087
17	-0.807	**14**	**0.244**	13	deleted	**18**	**-0.096**
22	**0.244**	**18**	**-0.096**	**14**	**0.244**	21	0.244
25	**0.903**	19	0.637	16	0.625	**22**	**0.244**
27	-0.075	23	0.899	20	-0.251	26	0.140
30	0.382	24	1.67	**25**	**0.903**	29	0.408
31	0.602	28	0.225	32	0.882	35	-0.087
37	-1.251	**30**	**0.382**	33	0.048	36	1.927
40	1.374	34	0.225	38	-0.633	**42**	**-0.562**
41	0.602	45	-0.245	39	0.218	43	-1.626
44	-0.682	**48**	**-0.608**	**42**	**-0.562**	47	1.190
48	**-0.608**	49	0.054	46	-1.772	**48**	**-0.608**
56	-2.322	54	-1.782	**48**	**-0.608**	50	0.740
60	-0.682	58	0.225	53	0.408	52	-2.227
62	-0.926	61	-2.322	64	-1.674	63	-1.674

Appendix 9.6: Feedback about the results (to the students of the University of Vienna and the Cypress College in Anaheim, California)

Comment: The feedback was held in English for both samples. It was the aim to give those students who executed the verbal test battery a little “service” to show them how much their participation was appreciated. The feedback included a general presentation of the results of the study and a list with the raw scores that were achieved by the students. The students were informed that they could get a certificate that they took part in this study, if they would send an e-mail with their names and student identification number to the Austrian test assessor. Figure 9.6-1 shows a screenshot of the the feedback letter which was sent to the students after the data collection was complete.



Dear students,

Thank you very much for taking part in this research study of the University of Vienna. Every one of you was such a big help and as a little compensation for your work and your time, this e-mail contains a list with your codenames and the raw scores which you achieved in the verbal test battery.

If you don’t remember your codename – read the key:

- ✓ First two letters of your first name
- ✓ Last two letters of your family name
- ✓ Last two numbers of your student identification number

I wish you all the best for your future career,

Brigitte Hansmann.

Vienna, 2010

Figure 9.6-1: Screenshot of the feedback letter that was sent to the students after the data collection was complete.

C U R R I C U L U M V I T A E

P E R S O N A L D E T A I L S

Brigitte Christine Hansmann

Gatterholzgasse 18/17
1120 Wien

0660/ 464 353 2
a0200535@unet.univie.ac.at
farfallina399@hotmail.com

E D U C A T I O N & Q U A L I F I C A T I O N S

Since 2002	University of Vienna, Austria. Psychology studies.
1994 to 2002	High school „Frauengasse“, Baden (Austria). Final exams with extraordinary achievement.
1990 to 1994	Primary School „Santa Christiana“, Vienna (Austria)

P R O F E S S I O N A L E X P E R I E N C E

Since Jan.2008	Employment at Caritas, Vienna. Social worker for people with mental and/or psychological handicaps.
Nov.2007- Dec.2008	Internship at the Institute of Violence Studies (“IGF”) in Vienna, Austria.
Oct.2007 - Dec. 2008	Internship at the psychosomatic ambulance of the Santa Anna Children Hospital.

Aug.2007 and 2008	"Kinderfreunde Austria"; leadership of English summer camps for children in Austria
July.2007 - Aug.2007	Internship at the Juvenile Detention Center in Gerasdorf, Austria.
Sept. 2006 - Sept. 2006	Internship at the Los Angeles Police Department, Juvenile Division.
Feb.2006 - July 2006	Internship at "Jugend am Werk"; day center for people with mental handicaps.
Feb.2006 - June 2006	Internship at "Jung und Alt"; social work for people with Alzheimer disease.
Sept. 2004 - July 2005	Internship at a children crisis centre of the MA11 in Vienna.
2003 - 2007	Hilfswerk Traiskirchen; tutor for English, Mathematics and German.

A D D T I O N A L E X P E R I E N C E

Since Feb.2009	School for Psychotherapy at HOPP Vienna, Austria.
June 2005 - Sept.2005	Foreign quarter at the University of California, Los Angeles

A D D T I O N A L S K I L L S

Language skills	English: fluent German: native language Italian, French: basic knowledge
Computer skills	eRm, extended Rasch models Lisrel, linear structure equation modelling Microsoft Office: Power Point, Word, Excel PASW 18, Statistic Software